The (Generalized) Post Correspondence Problem With Lists Consisting of Two Words is Decidable

A. Enhrenfeucht, G. Rozenberg and J. Karhumaki

CU-CS-204-81 February 1981

University of Colorado at Boulder
DEPARTMENT OF COMPUTER SCIENCE

ANY OPINIONS, FINDINGS, AND CONCLUSIONS OR RECOMMENDATIONS EXPRESSED IN THIS PUBLICATION ARE THOSE OF THE AUTHOR(S) AND DO NOT NECESSARILY REFLECT THE VIEWS OF THE AGENCIES NAMED IN THE ACKNOWLEDGMENTS SECTION.

INTRODUCTION

The Post Correspondence Problem, considered first by E. Post in [P], is perhaps the most useful problem as far as undecidable properties of formal languages are concerned (see, e.g., [H], [HU] and [S1]).

It can be formulated as follows. Let Σ be an alphabet and let h,g be two homomorphisms of Σ^*. The *Post Correspondence Problem* (PCP for short) is to determine whether or not there exists a word w in Σ^+ such that $h(w) = g(w)$. If $\#\Sigma = n$ then we say that we deal with the *Post Correspondence Problem of Length* n (PCP(n) for short).

The set of solutions of an instance of PCP (that is the set of all words satisfying the equation $h(w) = g(w)$) is referred to as an *equality language*. The "descriptional power" of PCP stems from the fact that it is able to code computations by arbitrary Turing machines. This is reflected in the fact that equality languages form a natural base in several characterizations of the class of recursively enumerable languages and its various subclasses (see, e.g., [BB], [C], [ER] and [S2]).

One particular aspect of PCP attracted quite a lot of attention. Since it is such a simply formulated problem of such a strong descriptional power, it forms an excellent framework for an attempt to formulate a boundary between "decidable" and "undecidable" (or "computable" and "noncomputable"). In other words one would like to establish as small as possible u such that PCP(u) is undecidable and as big as possible bound ℓ such that PCP(ℓ) is decidable. The smallest possible u so far is 10, which is derivable from a result of Matijasevic (see [C1]). As far as ℓ is concerned the only available (trivial) observation until now was the fact that PCP(1) is decidable. To establish whether or not PCP(2) is decidable turned out to be a challenging open problem.

There are also several results available which establish the decidability or undecidability of PCP not depending on the length but rather on other, more structural properties of the homomorphisms involved. For example, in [Le] it is proved that PCP remains undecidable when the involved homomorphisms are codes. Several interesting results related to PCP can be found in [CK] and [KS].

In this paper we consider a more general version of PCP(2) which is defined as follows. Let Σ, Δ be alphabets, h, g be two homomorphisms from Σ^* into Δ^* and let a_1,b_1,a_2,b_2 be words over Δ. The *Generalized Post Correspon-*

dence Problem (GPCP for short) is to determine whether or not there exists a word w in Σ^+ such that $a_1h(w)b_1 = a_2g(w)b_2$. If $\#\Sigma = n$ then we say that we deal with the ***Generalized Post Correspondence Problem of length*** n (GPCP(n) for short).

Note that if we set $a_1 = a_2 = b_1 = b_2 = \lambda$ then GPCP(n) reduces to PCP(n).

In this paper we prove that GPCP(2) is decidable (and so PCP(2) is decidable). Our proof involves several new techniques to deal with homomorphisms. In particular the construct called the *equality collector* (of two homomorphisms) plays a crucial role in this paper; we believe that the theory of equality collectors is worth to be investigated on its own.

Finally we want to remark that the solution of GPCP(2) that we present in this paper is a simplified version of the solution presented in [EhR].

2. PRELIMINARIES

In this paper only very basic notions of the formal language theory are needed. To fix the notation we specify the following. For other standard notions and notation we refer the reader to, e.g., [H] or [S].

We consider only finite alphabets, normally denoted by Σ. Moreover, in this paper Σ will be binary, say $\Sigma = \{0,1\}$, unless explicitly stated otherwise. A free monoid generated by Σ is denoted by Σ^* and its identity, the empty word, by λ. Let $\Sigma^+ = \Sigma^* - \{\lambda\}$. Elements of Σ^* are called words. For the length of a word x we use the notation $|x|$. For an integer n, $|n|$ denotes its absolute value. The number of occurrences of a letter c in a word x is denoted by $\#_c(x)$. For a finite set A, $\# A$ denotes its cardinality. Let $\Sigma = \{a_1, \dots , a_t\}$. The <u>Parikh-mapping</u> $\psi : \Sigma^* \to \mathbb{N}^{\#\Sigma}$ is defined as usual: for $1 \le i \le t$, the ith component of $\psi(x)$ equals to $\#_{a_i}(x)$.

For two words x and y, $x^{-1}y$ (resp. yx^{-1}) denotes the left (resp. right) difference of y by x. Consequently, if $y = xz$ then $z = x^{-1}y$, and if x is not a prefix of y, then $x^{-1}y$ is undefined (or, using another formulation, is the empty set of words). Certainly, the notions of the differences can be defined for languages as well. For instance, for languages L_1 and L_2, $L_1^{-1}L_2 = \{x^{-1}y \mid x \in L_1, y \in L_2\}$.

If a word x is a prefix (resp. proper prefix) of a word y we write x <u>pref</u> y (resp. x <u>p-pref</u> y). If either x <u>pref</u> y or y <u>pref</u> x (reps. x <u>p-pref</u> y or y <u>p-pref</u> x) then we write x <u>Pref</u> y (resp. x <u>p-Pref</u> y). The notation <u>pref</u>(x) denotes the set of all prefixes of x, while the notation $\underline{\text{pref}}_n(x)$ is used to specify the prefix of x of the length n. By definition, if $|x| < n$ then $\underline{\text{pref}}_n(x) = x$. For a language L, <u>pref</u>(L) (resp. <u>p - pref</u>(L)) denotes the set of all prefixes (resp. proper prefixes) of words in L. All the notions defined above for prefixes can be defined for suffixes as well. Then the notations of <u>pref</u> or <u>Pref</u> are replaced by <u>suf</u> or <u>Suf</u>.

The notion of a homomorphism from Σ^* into Δ^* is central for this paper. With the exception of sections 2 and 4 we consider λ-free homomorphisms only, i.e. homomorphisms for which $h(a) \neq \lambda$, for all a in Σ. The following two classes of homomorphisms over $\Sigma = \{0,1\}$ are important for us. We call a homomorphism $h : \Sigma^* \to \Sigma^*$ periodic if there exists a word p such that $h(\Sigma) \subseteq p^*$. By a marked homomorphism we mean a λ-free homomorphism h satisfying $\underline{pref}_1(h(0)) \neq \underline{pref}_1(h(1))$.

Certainly, the above notions can be defined for arbitrary alphabets as well (provided that the cardinality of the range alphabet is at least as large as that of the domain alphabet). It is well-known that a binary homomorphism h is nonperiodic if and only if $h(01) \neq h(10)$ if and only if h is injective.

We state now the central problem studied in this paper. This problem was introduced and first studied by Post [P]. Later on the problem has turned out to be one of the most useful decision problems within the formal language theory.

Definition 2.1. Let h and g be two homomorphisms from Σ^* into Δ^*. The Post Correspondence Problem (PCP for short) is to determine whether or not there exists a word w in Σ^+ such that $h(w) = g(w)$. If $\#\Sigma = n$ then we say that we deal with the Post Correspondence Problem of length n (PCP(n) for short).

In this paper we shall show that PCP(2) is decidable. In fact, we shall show that even a more general problem than PCP(2) is decidable. The generalization, for which the motivation becomes evident in the next section, is as follows.

Definition 2.2. Let h and g be two homomorphisms from Σ^* into Δ^* and let a_1,b_1,a_2,b_2 be words over Δ. The Generalized Post Correspondence Problem

(GPCP for short) is to determine whether or not there exists a word w in Σ^+ such that $a_1h(w)b_1 = a_2g(w)b_2$. If $\#\Sigma = n$ then we say that we deal with the Generalized Post Correspondence Problem of length n (GPCP(n) for short).

Let h,g,a_1,b_1,a_2 and b_2 be as in Definition 2.2. Then $I = (h,g,a_1,b_1,a_2,b_2)$ is an instance of GPCP. We shall show that GPCP(2) is decidable, i.e. that there exists an algorithm which decides for a given instance I of GPCP(2) whether or not there exists a word w in Σ^+ satisfying $a_1h(w)b_1 = a_2g(w)b_2$ or, in other words, whether or not I has a solution. When studying the decidability status of GPCP(n) we can certainly restrict the considerations to the case when $\Sigma = \Delta$. Consequently, in the sequel $\Sigma = \Delta = \{0,1\}$.

Let I be an instance of GPCP(2) as above. We say that I is periodic if either h or g is periodic, and that I is marked if h and g are marked. It will turn out that it suffices to show that the problem is decidable for periodic and marked instances of GPCP(2).

Finally, we say that I, or a pair (h,g) of homomorphisms, is unbalanced if either $|h(i)| \geq |g(i)|$, for $i = 0,1$, or $|g(i)| \geq |h(i)|$, for $i = 0,1$. Otherwise I or (h,g) is called balanced.

3. REDUCTION LEMMA

In this section we show that in order to solve PCP(2) it suffices to consider two kinds of homomorphisms: periodic and marked.

Reduction Lemma 3.1. For an instance $I = (h,g,\lambda,\lambda,\lambda,\lambda)$ of PCP(2), where h and g are nonperiodic, one can effectively construct a marked instance $I' = (h',g',a_1',b_1',a_2',b_2')$ of GPCP(2) such that I has a solution if and only if I' has a solution.

Proof. Let $\underline{cyc}_1$ be a mapping $\{0,1\}^* \to \{0,1\}^*$ defined as follows. For words $w = cu$, with $c \in \{0,1\}$ and $u \in \{0,1\}^*$, $\underline{cyc}_1(w) = uc$ and $\underline{cyc}_1(\lambda) = \lambda$. Let $\underline{cyc}_k = (\underline{cyc}_1)^k$. Clearly, for any mapping $f : \{0,1\}^* \to \{0,1\}^*$ and any word x in $\{0,1\}^+$

$$(1) \qquad \underline{cyc}_k(f(x)) = (\underline{pref}_{k_1}(f(x)))^{-1} f(x)\ \underline{pref}_{k_1}(f(x)),$$

where $0 \le k_1 < |f(x)|$ and $k_1 \equiv k \bmod (|f(x)|)$.

Now we start constructing I'. Since h and g are nonperiodic $h(01) \neq h(10)$ and $g(01) \neq g(10)$. Let z (resp. v) be the maximal common prefix of $h(01)$ and $h(10)$ (resp. $g(01)$ and $g(10)$). By symmetry, we may assume that $|z| \geq |v|$. We define

$$h' = \underline{cyc}_{|z|} \circ h \quad \text{and} \quad g' = \underline{cyc}_{|v|} \circ g.$$

By the choice of z and v, h' and g' are homomorphisms and moreover they are marked. Further we set

$$a_1' = \underline{suf}_{|z|-|v|}(z) = b_2' ,$$

$$a_2' = \lambda = b_1'.$$

Then, by (1), it is immediate that I has a solution if and only if I' has a solution.

The above gives a motivation to study the Generalized Post Correspondence Problem. Indeed, the replacement of arbitrary homomorphisms by marked ones, which is very essential in our later considerations, can be done via this generalization. The above reduction lemma can also be formulated for instances of GPCP(2). The reason why we took a nongeneralized case separately is that we want to have as simple as possible proof of the decicability of PCP(2).

Lemma 3.2. For an instance $I = (h,g,a_1,b_1,a_2,b_2)$ of GPCP(2), where h and g are nonperiodic, one can effectively construct a finite set MAR(I) of marked instances of GPCP(2) and a constant q such that I has a solution of the length at least q if and only if some instance in MAR(I) has a solution.

Proof. We use the notations from the proof of Reduction Lemma. The homomorphisms h' and g' are defined in the same way. The words a_1' and a_2' are now

$$a_1' = a_1 z \text{ and } a_2' = a_2 v.$$

To define the b-words let r be the minimal integer such that for every word $x \in \{0,1\}^*$, with $|x| = r$, $|h(x)| \geq |h(01)|$ and $|g(x)| \geq |g(01)|$. Then, for each u in $\{0,1\}^*$ with $|u| = r$, we define

$$b_{1,u}' = z^{-1}h(u)b_1 \text{ and } b_{2,u}' = v^{-1}g(u)b_2.$$

Finally, let

$$\text{MAR}(I) = \{(h',g',a_1',b_{1,u}',a_2',b_{2,u}') | u \in \{0,1\}^r\}$$

and

$$q = 2r.$$

Then the lemma follows. Indeed, for words $w,u \in \{0,1\}^*$, with $|w| \geq r$ and $|u| = r$,

$$a_i h(wu) b_i = a_i' h'(w) b_{i,u}', \quad \text{for } i = 1,2.$$

4. PERIODIC INSTANCES

In this section we settle the case when at least one of the homomorphisms involved in an instance of GPCP(2) is periodic. This also shows why we can restrict our attention to λ-free homomorphisms elsewhere. Indeed, a homomorphism over a binary alphabet which is not λ-free is periodic. Basically, the solution is similar to that presented in [KS] for PCP(n), see also [CK].

Theorem 4.1. It is decidable whether or not an instance $I = (h,g,a_1,b_1,a_2,b_2)$ of GPCP, with h periodic, has a solution.

Proof. Let $h(\Sigma) \subseteq p^*$. Define

$$L_1 = g^{-1}(a_2^{-1}a_1p^*b_1b_2^{-1})$$

$$L_2 = \{x \in \Sigma^* \mid |h(x)| - |g(x)| = |a_2b_2| - |a_1b_1|\}.$$

and

$$L = L_1 \cap L_2 .$$

Then, clearly,

$$y \in L$$

iff

$$y \in g^{-1}(a_2^{-1}a_1p^*b_1b_2^{-1}) \text{ and } |h(y)| - |g(y)| = |a_2b_2| - |a_1b_1|$$

iff

$$a_2g(y)b_2 \in a_1p^*b_1 \qquad \text{and } |a_1\ h(y)\ b_1| = |a_2\ g(y)\ b_2|$$

iff

$$a_2g(y)b_2,\ \ a_1\ h(y)\ b_1 \in a_1p^*b_1 \text{ and } |a_1\ h(y)\ b_1| = |a_2\ g(y)\ b_2|$$

iff

$$a_2\ g(y)\ b_2 = a_1\,h(y)\ b_1 .$$

Hence I has a solution if and only if L is nonempty.

The emptiness of L is seen to be decidable as follows. Let $\psi : \Sigma^* \to \mathbb{N}^{\#\Sigma}$ be the Parikh-mapping. Then

$$L_1 \cap L_2 = \emptyset \text{ iff } \psi(L_1 \cap L_2) = \emptyset \text{ iff } \psi(L_1) \cap \psi(L_2) = \emptyset,$$

where the last equality follows since $L_2 = \psi^{-1}(\psi(L_2))$. Now observe that L_1 is regular and hence $\psi(L_1)$ is semi-linear in the sense of [G]. The set $\psi(L_2)$, in turn, consists of nonnegative solutions of a linear equation and so it is effectively semi-linear. Finally, the intersection of two semi-linear sets is also effectively semi-linear. Hence the result follows. If necessary, the reader may consult [G].

We want to emphasize that the assumption of a binary alphabet is not at all needed in the above proof.

5. SOME SPECIAL INSTANCES

In this section we deal with some relatively simple cases which turn out to be important for the general solution and which we must settle separately.

First we consider unbalanced instances of GPCP(2).

Theorem 5.1. It is decidable whether an unbalanced instance I of GPCP(2) has a solution.

Proof. Let $I = (h,g,a_1,b_1,a_2,b_2)$ with $|h(i)| \geq |g(i)|$, for $i = 0,1$. We define recursively the sets U_i as follows:

$$U_0 = \{(a_2^{-1}a_1,h),\ (a_1^{-1}a_2,g)\}$$

$$U_{i+1} = \{(h(c)^{-1}\,xg(c),g),\ ((xg(c))^{-1}\,h(c),\ h) \mid (x,g) \in U_i,\ c \in \{0,1\}\}$$

$$\cup\ \{(g(c)^{-1}\,xh(c),\ h),((xh(c))^{-1}\,g(c),\ g) \mid (x,h) \in U_i,\ c \in \{0,1\}\}$$

By definition, (undefined,h) and (undefined,g) are undefined. Intuitively, U_i gives all words u such that either $a_1\ h(y) = a_2\ g(y)\ u$ or $a_1\ h(y)\ u = a_2\ g(y)$ for some word y with the length i. Moreover, the second component indicates which homomorphism is "ahead".

Clearly, I has a solution if and only if some U_j either contains an element (u,h) such that $b_1 = ub_2$ or it contains an element (u,g) such that $ub_1 = b_2$. But, by the form of (h,g) and by the recursive definition of the sets U_i, for each natural number q, there effectively exists a constant n_q such that any (u,h) or (u,g), with $|u| = q$, occurs in $\bigcup_{i=0}^{\infty} U_i$ if and only if it occurs in $\bigcup_{i=0}^{n_q} U_i$.

So the theorem follows, since we must only check whether $(b_1b_2^{-1},g)$ or $(b_2b_1^{-1},h)$ is in $\bigcup_{i=0}^{\infty} U_i$.

The instances of GPCP(2) considered in the following four lemmas are called special. Our general techniques do not apply to them. The reader may skip these four lemmas now and return to them after getting a motivation from section 7.

In what follows μ denotes a mapping of $\{0,1\}$ onto $\{0,1\}$ which is either the identity or the cyclic permutation, i.e. $\mu(0) = 1$ and $\mu(1) = 0$.

Lemma 5.1. It is decidable whether or not an instance $I = (h,g,a_1,b_1,a_2,b_2)$ of GPCP(2) with h and g of the form

$$h(i) \in (j(1-j))^*j \quad , \quad g(\mu(i)) \in (j(1-j))^*j \quad ,$$
$$h(1-i) \in ((1-j)j)^*(1-j), \quad g(\mu(1-i)) \in ((1-j)j)^*(1-j) \quad ,$$

for some i and j in $\{0,1\}$, has a solution.

Proof. Clearly, we may fix i and j, say $i = 0$ and $j = 0$. Then $h(0), g(\mu(0)) \in (01)^*0$ and $h(1), g(\mu(1)) \in (10)^*1$.
We start to "chase" a solution of I by generating the sequences

$$(a_1,a_2), (a_1\ h(i_1), a_2\ g(i_1)), (a_1\ h(i_1i_2), a_2\ g(i_1i_2)), \ldots$$

and

$$\ldots, (h(j_2j_1)\ b_1, g(j_2j_1)b_2), (h(j_1)b_1, g(j_1)b_2), (b_1,b_2)$$

such that $a_1\ h(i_1 \ldots i_t)$ p-Pref $a_2\ g(i_1 \ldots i_t)$ and $h(j_s \ldots j_1)b_1$ p-Suf $g(j_s \ldots j_1)b_2$ for $t \geq 1$ and $s \geq 1$.
If for instance $a_1 = a_2$, then we actually generate two sequences from left to right.

The basic observation is that as long as the sequences can be generated, 0's and 1's occur alternatively, i.e. $i_1 \ldots i_t$ and $j_s \ldots j_1$ are in pref$(01)^* \cup$ pref$(10)^*$. Hence, it can be decided whether or not at least one of these sequences is infinite and does not go into a cycle, i.e. at least one of the sequences of the differences is unbounded. If this is the case it suffices, by the above periodicity, to check whether or not $a_1\, h(i_1 \dots i_q)b_1 = a_2\, g(i_1 \dots i_q)b_2$ for some q smaller than an effectively computable constant.

In the other cases there are only a finite number of words to be checked through (including the words which are obtained by continuing the above sequences by one step in the case when both of the sequences terminate because the proper prefix or suffix requirement is not fulfilled).

Lemma 5.2. Let $I = (h,g,a_1,b_1,a_2,b_2)$ be an arbitrary instance of GPCP(2) such that h and g are of the form

$$h(i) \in j^* \quad , \quad g(\mu(i)) \in j^* \; ,$$
$$h(1-i) \in (1-j)^* j^k \; , \; g(\mu(1-i)) \in (1-j)^* j^\ell \; ,$$

where i and j are in $\{0,1\}$, $k\ell = 0$ and if $k \neq 0$ (resp. $\ell \neq 0$) then $|g(\mu(i))| > k$ (resp. $|h(i)| > \ell$). It is decidable whether or not such an arbitrary I has a solution.

Proof. By symmetry, we may set $i = 0$, $j = 0$ and $\ell = 0$. Consequently, $h(0) \in 0^*$, $h(1) \in 1^* 0^k$, $g(\mu(0)) \in 0^*$ and $g(\mu(1)) \in 1^*$, with $|g(\mu(0))| > k$. By Theorem 5.1., we may further assume that I is balanced.

Again we start to "chase" a solution by generating the sequence

$$(a_1,a_2),\ (a_1 h(i_1),\ a_2\, g(i_1)),\ (a_1\, h(i_1 i_2),\ a_2\, g(i_1 i_2)) \dots$$

such that $a_1\, h(i_1 \dots i_t)$ Pref $a_2\, g(i_1 \dots i_t)$. In the case of $a_1 = a_2$ we actually generate two such sequences and if in these sequences at some stage the components are equal, then the sequence may branch into two sequences. Observe, however, that because of the form of (h,g) the branching may happen only at the very beginning and so one can consider a finite number of sequences obtained in the above way. Let us consider such a sequence.

Now the following observation is crucial. Let (α_i,β_i) denote the ith element in the above sequence. The number of change points, i.e. the number of positions where the letter changes into another one in the words $\alpha_i^{-1}\beta_i$ or $\beta_i^{-1}\alpha_i$, does not decrease. This is certainly true if $|a_1| \geq |a_2|$ since in that case every time when a change point is "eaten up" (during the generation of the sequence), a new one is produced. And this is also true in the case $|a_1| < |a_2|$ with the possible exceptions occurring at the very beginning of the process, i.e. when $a_1^{-1}a_2$ is "eaten up". Here the inequality $|g(\mu(0))| > k$ is needed.

The above guarantees that we can effectively decide whether or not our chase of a solution leads to a solution.

Lemma 5.3. For an instance $I = (h,g,a_1,b_1,a_2,b_2)$ of GPCP(2) with h and g of the form

$$\begin{aligned} h(i) &= j \quad , & g(\mu(i)) &= j , \\ h(1-i) &= ((1-j)j^n)^N(1-j) \quad , & g(\mu(1-i)) &= ((1-j)j^n)^M (1-j) \quad , \end{aligned}$$

where i and j are in $\{0,1\}$, $n \geq 1$ and $N \neq M$, it is decidable whether or not it has a solution.

Proof. As before we may set $i = 0$ and $j = 0$. If μ is the identity, then (h,g) is unbalanced and we are done. Consequently, $h(0) = 0$, $h(1) = (10^n)^N 1$, $g(0) = (10^n)^M 1$ and $g(1) = 0$. If u is a solution of I, then necessarily $\#_i(a_1\ h(u)\ b_1) = \#_i(a_2\ g(u)\ b_2)$, for $i = 0,1$, i.e.

$$\tag{1} \begin{cases} (1-nM)\ \#_0(u) + (nN-1)\ \#_1(u) = k_0 \\ -(M+1)\ \#_0(u) + \ (N+1)\ \#_1(u) = k_1 \end{cases} ,$$

where $k_i = \#_i(a_2b_2) - \#_i(a_1b_1)$ for $i = 0,1$. Since

$$D = \begin{vmatrix} 1-nM & nN-1 \\ -M-1 & N+1 \end{vmatrix} = (n+1)(N-M) \neq 0$$

the system (1) has a unique solution, which proves the lemma.

Lemma 5.4. Let $I = (h,g,a_1,b_1,a_2,b_2)$ be an arbitrary instance of GPCP(2) with h and g of the form

$$\begin{aligned} h(i) &= jj^n, & g(\mu(i)) &= jj^{\ell}, \\ h(1-i) &= (1-j)j^m, & g(\mu(1-i)) &= (1-j)j^k, \end{aligned}$$

for some i and j in $\{0,1\}$ and $n,m,\ell,k \geq 1$ with $n+m \neq k+\ell$. It is decidable whether or not I has a solution.

Proof. We again set $i = 0$ and $j = 0$. The case when μ is the identity is clear. Indeed, when chasing a solution every time 1 is "eaten up" another is created and this is the only way how 1 may appear. We leave the details to the reader.

So assume that $h(0) = 00^n$, $h(1) = 10^m$, $g(0) = 10^k$ and $g(1) = 00^{\ell}$. We use the same argument as in the previous proof. Now we obtain

$$(2) \qquad \left\{ \begin{aligned} (n+1-k)\ \#_0(u)+(m-(\ell+1))\ \#_1(u) &= k_0 \\ -\#_0(u)+ \qquad\qquad \#_1(u) &= k_1 , \end{aligned} \right.$$

where $k_i = \#_i(a_2b_2) - \#_i(a_1b_1)$, for $i = 0,1$. Now the determinant of (2) is

$$D = \begin{vmatrix} n+1-k & m-\ell-1 \\ -1 & 1 \end{vmatrix} = (n+m)-(k+\ell) \neq 0 \quad .$$

Since $D \neq 0$, we are done: the possible solution can be found from the set $\{u \mid u \text{ satisfies } (2)\}$.

6. ECOL CONSTRUCTION

Now we have come to the central point of our solution for GPCP(2). As justified by results in section 3 we consider from now on marked homomorphisms only. For these homomorphisms and for instances of GPCP(2) involving such homomorphisms we present the transformation called the equality collector. This is the fundamental notion of our solution.

We start by an intuitive description of the basic idea behind this construction. Let h and g be marked homomorphisms and let α and β be two words for which αp-Pref β, say α is a proper prefix of β. We are interested in finding two words u and v such that $\alpha h(u) = \beta g(v)$. A natural way to start is as follows. Since β is "ahead" we look at the first letter of the difference $\alpha^{-1}\beta$, and since h is marked this letter defines uniquely a letter i_1 from $\{o,1\}$ such that (if the required u and v exist at all) then u must start with i_1. We iterate the process until a word $i_1 \ldots i_{r_1}$ is found such that either $\alpha\ h(i_1 \ldots i_{r_1}) = \beta$ or β is a proper prefix of $\alpha\ h(i_1 \ldots i_{r_1})$. In the second case we continue by changing the roles of h and g. Finally we get one of the following possibilities. Either the process ends successfully, i.e. we find words $i_1 \ldots i_r$ and $j_1 \ldots j_s$ such that $\alpha\ h(i_1 \ldots i_r) = \beta\ g(j_1 \ldots j_s)$, or it blocks (i.e. in some step the continuation is not possible) or it continues "infinitely long".

We formalize the above in the following way.

Definition 6.1. Let h and g be marked homomorphisms of $\{0,1\}^*$ and let $\alpha,\beta \in \{0,1\}^*$ with α Pref β.We define an (α,β)-sequence with respect to h and g, in symbols $(\alpha,\beta)_{h,g}$, inductively as follows:

(i) $(\alpha,\beta)^{(0)}_{h,g} = (\alpha,\beta)$.

(ii) For $j \geq 0$

$$(\alpha,\beta)_{h,g}^{(j+1)} = \begin{cases} (\alpha' h(i), \beta') & \text{if} \quad (\alpha,\beta)_{h,g}^{(j)} = (\alpha',\beta'), \\ & |\alpha'| < |\beta'| \quad \text{and } \alpha' h(i) \underline{\text{Pref}}\ \beta', \\ (\alpha',\beta' g(i)) & \text{if} \quad (\alpha,\beta)_{h,g}^{(j)} = (\alpha',\beta'), \\ & |\alpha'| > |\beta'| \quad \text{and } \alpha' \underline{\text{Pref}}\ \beta' g(i). \end{cases}$$

Let $(\alpha,\beta)_{h,g}^{(j)} = (\alpha_j,\beta_j)$ whenever it is defined. We say that $(\alpha,\beta)_{h,g}$

- is <u>successful</u> if,for some j $\alpha_j = \beta_j$ (i.e. the process terminates for this reason),

- <u>blocks</u> if for some j α_j <u>p-pref</u> β_j and it is not true that $\alpha_j\ h(i)$ <u>Pref</u> β_j, for i in {0,1}, or β_j <u>p-pref</u> α_j and it is not true that α_j <u>Pref</u> $\beta_j\ g(i)$, for i in {0,1},

- is <u>infinite</u> if (α_j,β_j) is defined for all j.

In the first case we write $s((\alpha,\beta)_{h,g}) = \alpha_j (= \beta_j)$.

Clearly, the above classification is exhaustive. Observe also that if $(\alpha,\beta)_{h,g}$ is successful, then for some words u and v

$$\alpha\ h(u) = s((\alpha,\beta)_{h,g}) = \beta\ g(v)$$

and, moreover, u and v are minimal, i.e. the above equation does not hold for any pair (u',v'), where $|u'| < |u|$ or $|v'| < v$. The following lemma is also obvious.

<u>Lemma 6.1</u>. Given an arbitrary pair (h,g) of homomorphisms and words α and β, it is decidable whether or not the sequence $(\alpha,\beta)_{h,g}$ is succesful, infinite or blocks.

Before setting our central definition we need some notations. Let (h,g) be an ordered pair of marked homomorphisms. Then a mapping $\mu_{h,g}$

(or μ for short) of $\{0,1\}$ onto $\{0,1\}$ is defined by

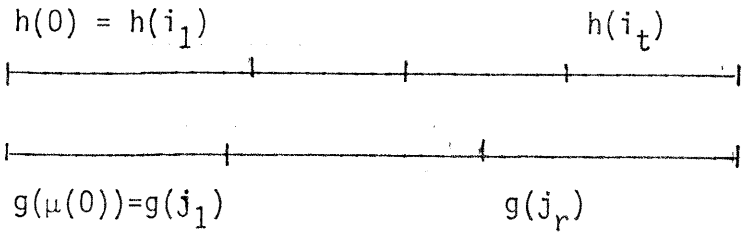
$$\mu(i) = i \quad \text{for } i = 0,1, \text{ if } \underline{\text{pref}_1}\,(h(0)) = \underline{\text{pref}_1}\,(g(0))$$
$$\mu(i) = 1-i \quad \text{for } i = 0,1, \text{ if } \underline{\text{pref}_1}\,(h(0)) \neq \underline{\text{pref}_1}\,(g(0)).$$

Consequently, $\underline{\text{pref}_1}\,(h(i)) = \underline{\text{pref}_1}\,g(\mu(i))$, for $i = 0,1$. We call a pair (h,g) of marked homomorphisms successful if both $(h(0), g(\mu(0)))_{h,g}$ and $(h(1), g(\mu(1)))_{h,g}$ are successful.

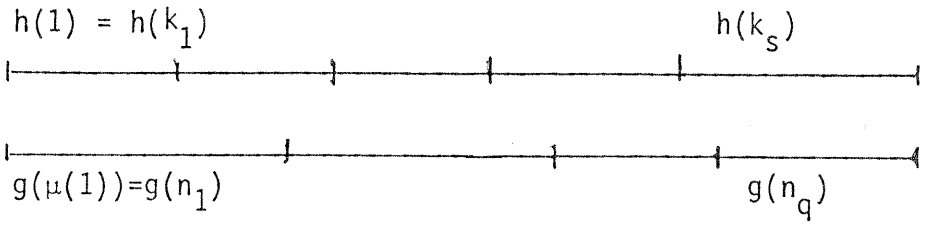
Definition 6.2. Let (h,g) be a successful pair of homomorphisms. Then the equality collector of (h,g), denoted as $\mathrm{ecol}(h,g)$, is the pair $(\overline{h},\overline{g})$ of homomorphisms of $\{0,1\}^*$ defined by

$$\overline{h}(0) = h^{-1}(s((h(0), g(\mu(0)))_{h,g})) \qquad \overline{h}(1) = h^{-1}(s((h(1), g(\mu(1)))_{h,g}))$$
$$\overline{g}(0) = g^{-1}(s((h(0), g(\mu(0)))_{h,g})) \qquad g(1) = g^{-1}(s((h(1), g(\mu(1)))_{h,g})).$$

The following remarks concerning the above definition are in order. Since h and g are marked they are nonperiodic and hence injective. So the values of $\overline{h}(i)$ and $\overline{g}(i)$ are well-defined. Observe also that $\overline{h}$ and $\overline{g}$ are marked. Finally, we want to emphasize that the above definition can be described in a very illustrative way. Indeed, since the pair (h,g) is successful we have something as follows:



and



With these notations

$$\bar{h}(0) = i_1 \ldots i_t \qquad \bar{h}(1) = k_1 \ldots k_s$$
$$\bar{g}(0) = j_1 \ldots j_r \qquad \bar{g}(1) = n_1 \ldots n_q \quad .$$

Observe also that

$$h(\bar{h}(i)) = g(\bar{g}(i)) \text{ , for } i = 0,1 \quad ,$$

i.e. the pairs $(\bar{h}(0), \bar{g}(0))$ and $(\bar{h}(1), \bar{g}(1))$ are solutions of the equation $h(u) = g(w)$. Moreover, they are the only minimal solutions, i.e. for any solution (u',w') either $\bar{h}(0)$ pref u' and $\bar{g}(0)$ pref w' or $\bar{h}(1)$ pref u' and $\bar{g}(1)$ pref w'.

The usefulness of the ecol construction relies on the fact that $(\bar{h},\bar{g})$, if it exists, is a "smaller" pair of homomorphisms than (h,g) (except for some special cases).
To be able to show this we now define what we mean by "smaller".

Definition 6.3. Let (h,g) be a pair of homomorphisms of $\{0,1\}^*$. We define the size of (h,g), in symbols $\sigma(h,g)$, to be

$$\sigma(h,g) = \#\underline{\text{p-suf}}\{h(0),h(1)\} + \#\underline{\text{p-suf}}\{g(0), g(1)\}.$$

We first show that the ecol construction never enlarges the size of a pair of homomorphisms.

Lemma 6.2. Let (h,g) be a successful pair of homomorphisms and let ecol $(h,g) = (\bar{h},\bar{g})$; Then

$$\sigma(\bar{h},\bar{g}) \leq \sigma(h,g).$$

Proof. Let us recall the figure following Definition 6.2, i.e. let $h(i_1 \ldots i_t) = g(j_1 \ldots j_r)$ and $h(k_1 \ldots k_s) = g(n_1 \ldots n_q)$, where $i_1 = 0$, $k_1 = 1$ and t,r,s and q are as small as possible, so that $\bar{h}(0) = i_1 \ldots i_t$,

$\overline{h}(1) = k_1 \ldots k_s$, $\overline{g}(0) = j_1 \ldots j_r$ and $\overline{g}(1) = n_1 \ldots n_q$. We consider the sequences

$$z_1, \ldots, z_{t+r-2} \quad \text{and} \quad y_1, \ldots, y_{s+q-2}$$

of nonempty labeled suffixes encountered in the step by step generation of the successful sequences $(h(0), g(\mu(0)))_{h,g}$ and $(h(1), g(\mu(1)))_{h,g}$, respectively. The suffixes are labeled in the sense that each of them contains also information whether it is obtained from $\{h(0), h(1)\}$ or $\{g(0), g(1)\}$ in the construction.

We first observe that neither the z-sequence nor the y-sequence can contain repetitions. This is because h and g are marked and the sequences $(h(0), g(\mu(0)))_{h,g}$ and $(h(1), g(\mu(1)))_{h,g}$ are successful. If the same holds true for the combined sequence $z_1, \ldots, z_{t+r-2}$, $y, \ldots, y_{s+q-2}$ we are done. Indeed, in that case

$$\sigma(h,g) \geq t + r + s + q - 4 = \sigma(\overline{h},\overline{g})$$

It remains the case when for some $m \leq t+r-2$ and $\ell \leq s+q-2$ $z_m = y_\ell$. Let ℓ and m be minimal. Then

$$\sigma(h,g) \geq t + r - 2 + \ell - 1 = \sigma(\overline{h},\overline{g})$$

which completes the proof of the lemma.

7. DETAILED ANALYSIS OF ECOL(h,g).

In this section we sharpen Lemma 6.2. We analyse the implications of the equality $\sigma(\overline{h},\overline{g}) = \sigma(h,g)$. It turns out that the equality is possible only in some special cases.

Lemma 7.1. Let (h,g) be a balanced and successful pair of homomorphisms and let ecol(h,g) = $(\overline{h},\overline{g})$. If

$$\sigma(\overline{h},\overline{g}) = \sigma(h,g)$$

then either

(i) $L = \{h(0), h(1), g(0), g(1)\} \subseteq \underline{pref}\ (01)^* \cup \underline{pref}\ (10)^*$

or

(ii) $\{h(i), g(\mu(i))\} \subseteq j^*$ for some i and j in $\{0,1\}$.

Proof. The basic argument of the proof goes as follows:
By the proof of Lemma 6.2, the equality means that all the suffixes from $\{g(0), g(1)\}$ and from $\{h(0), h(1)\}$ are encountered in the construction of $(\overline{h},\overline{g})$. In particular, it follows that for any proper suffix α from $\{h(0), h(1)\}$ (resp. from $\{g(0), g(1)\}$) either α Pref g(0) or α Pref g(1) (resp. α Pref h(0) or α Pref g(1)).

We have two cases.

I. $\rho = \min\{|x| \mid x \in L\} > 1$.
We assume that (i) is not satisfied, i.e. L contains a word 00 (or symmetrically 11) as a subword, say

$$h(0) = x_1 00 x_2 \qquad \text{for some words } x_1 \text{ and } x_2.$$

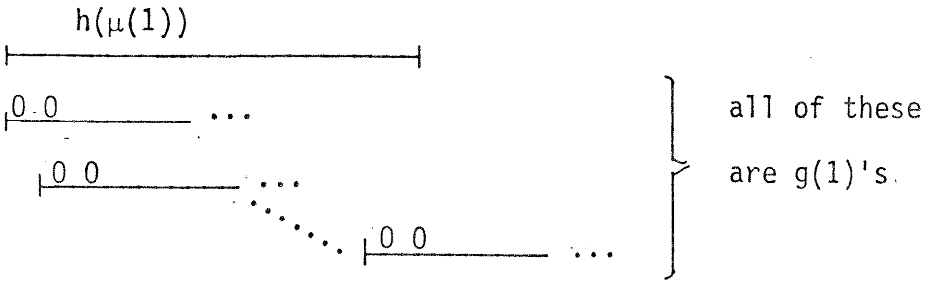
Then, by above, either g(0) Pref$(x_1^{-1}\ h(0))$ or g(1) Pref$(x_1^{-1}\ h(0))$ Let this be true for g(1). Then, since $|g(1)| \geq 2$,

$$g(1) = 00x_3 \qquad \text{for some } x_3 \text{ .}$$

Hence $h(\mu(1))$ starts with two 0's, say

$$h(\mu(1)) = 00x_4 \quad \text{for some } x_4.$$

We apply our basic argument and conclude that the situation can be illustrated as:



i.e. $h(\mu(1)) \in 0^*$. Symmetrically $g(1) \in 0^*$, and so (ii) is satisfied.

II. $\rho = 1$.

Let $|g(0)| = 1$, say $g(0) = 0$. If $h(\mu(0)) \in 0^*$ we are done: (ii) is satisfied. So assume that

$$h(\mu(0)) = 0x_5 1x_6 \quad \text{for some } x_5 \text{ and } x_6.$$

Now the basic argument, applied to the suffixes x_6 and $1^{-1}h(\mu(1))$ of $\{h(0),h(1)\}$, yields that

$$g(1) \in (10^*)^+1$$

(remember that $|g(1)|>1$, because $|g(0)| = 1$ and (h,g) is balanced).
If $g(1)= x_7 11x_8$, for some x_7 and x_8, then proceeding as in the case I we conclude that either (ii) is satisfied or $h(\mu(1)) = 1$.

Assume first that $h(\mu(1)) \neq 1$. If (ii) is not satisfied, then

$$g(1) \in (10^+)^+1.$$

Now we again apply our basic argument. To obtain all the suffixes of $g(1)$, necessarily

$$g(1) \in (10)^+1.$$

This is because $h(\mu(0)) = 0x_5 1x_6$ and $h(\mu(1)) = 1x_9$, for some x_9. Moreover, $h(\mu(0))$ must be of the form

$$h(\mu(0)) = 01x_{10} \qquad \text{for some } x_{10}.$$

Now we claim that $h(\mu(0))$ cannot contain two consecutive 0's or 1's. This follows, again by our basic argument, since $g(0) = 0$, $g(1) \in (10)^+1$ and $\underline{\text{pref}}_2\ (h(\mu(0)) = 01$. Consequently,

$$h(\mu(0)) \in \underline{\text{pref}}\ (01))^*.$$

Exactly the same argument shows that

$$h(\mu(1)) \in \underline{\text{pref}}\ (10)^* \quad .$$

Now instead of $\underline{\text{pref}}_2\ (h(\mu(0))) = 01$ we use the equality $\underline{\text{pref}}_1\ (h(\mu(1))) = 1$. So (i) is satisfied in this subcase.

The remaining possibility is $h(\mu(1)) = 1$, i.e. we have altogether

$$g(0) = 0 \quad , \qquad h(\mu(0)) \in (01^*)^+0 \ ,$$
$$g(1) \in (10^*)^+1 \ , \qquad h(\mu(1)) = 1 \qquad ,$$

where the formula for $h(\mu(0))$ follows by symmetry.

Again we apply our basic argument to $g(1)$: either $h(\mu(0)) \in 0^*$ or $g(1) \in 1^*$ or $g(1)$ is in $(1^+0)^+1$. The first two possibilities lead to (ii). Hence, by symmetry, the remaining case is:

$$g(0) = 0 \ , \qquad h(\mu(0)) \in (0^+1)^+0 \ ,$$
$$g(1) \in (1^+0)^+1 \ , \qquad h(\mu(1)) = 1 \qquad .$$

In this case our basic argument immediately yields that

$$g(1) \in (10)^+1 \qquad \text{and} \qquad h(\mu(0)) \in (01)^+0 \ .$$

Hence, (i) is satisfied and our proof is complete.

We will analyse now special cases of Lemma 7.1. In the case (i) we have the following 7 possibilities (remaining cases are symmetric versions of these):

I	$h(0) \in (01)^*$,	$g(\mu(0)) \in (01)^*$,
	$h(1) \in (10)^*$,	$g(\mu(1)) \in (10)^*$,
II	$h(0) \in (01)^*0$,	$g(\mu(0)) \in (01)^*$,
	$h(1) \in (10)^*$,	$g(\mu(1)) \in (10)^*$.
III	$h(0) \in (01)^*0$,	$g(\mu(0)) \in (01)^*$,
	$h(1) \in (10)^*1$,	$g(\mu(1)) \in (10)^*$.
IV	$h(0) \in (01)^*0$,	$g(\mu(0)) \in (01)^*0$,
	$h(1) \in (10)^*$,	$g(\mu(1)) \in (10)^*$.
V	$h(0) \in (01)^*0$,	$g(\mu(0)) \in (01)^*$,
	$h(1) \in (10)^*$,	$g(\mu(1)) \in (10)^*1$.
VI	$h(0) \in (01)^*0$,	$g(\mu(0)) \in (01)^*0$,
	$h(1) \in (10)^*1$,	$g(\mu(1)) \in (10)^*$.
VII	$h(0) \in (01)^*0$,	$g(\mu(0)) \in (01)^*0$,
	$h(1) \in (10)^*1$,	$g(\mu(1)) \in (10)^*1$.

We must use different techniques in different cases. First we show that the cases II and V are impossible.

Lemma 7.2. For pairs of homomorphisms of the form II or V, the sequence $(h(0),\ g(\mu(0)))_{h,g}$ is not successful.

Proof. In case II the relation $h(0x)$ Pref $g(\mu(0)y)$ implies that $x \in 1^*$ and $y \in (\mu(0))^*$. Hence the result follows because $\underline{suf}_1\ (h(1)) \neq \underline{suf}_1(g(\mu(0)))$.

The case V is even simpler. Indeed, the relation $\underline{suf}_1\ (h(0)) = \underline{suf}_1\ (h(1)) \neq \underline{suf}_1\ (g(\mu(0))) = \underline{suf}_1\ (g(\mu(1)))$ guarantees the result.

In cases I, IV and VI $(\overline{h},\overline{g})$, if it exists, is strictly "smaller" than (h,g).

Lemma 7.3. Let (h,g) be a balanced pair of homomorphisms of the form I, IV or VI. If (h,g) is successful, then $\sigma(\overline{h},\overline{g}) < \sigma(h,g)$.

Proof. In case I suffixes $(10^*)1$ are not encountered in the construction of $(\overline{h},\overline{g})$. In case IV the same holds true for the suffix 0 of $h(1)$.

In case VI we first conclude that, because (h,g) is balanced, either $h(0)$ or $h(1)$ is of the length not smaller than three. Consequently 0 or 01 is a proper suffix in $\{h(0), h(1)\}$ but it is not encountered in the construction of $(\overline{h},\overline{g})$. This is because $\underline{suf}_1(g(0)) = \underline{suf}_2(g(1)) = 0$ while $\underline{suf}_1(h(0)0^{-1}) = \underline{suf}_1(h(1)(01)^{-1}) = 1$.

Case III is dealt as follows.

Lemma 7.4. For pairs of homomoprhisms of the form III and VI either

(i) $\quad (\overline{h}(0), \overline{g}(\mu(0)))_{\overline{h},\overline{g}}$ or $(\overline{h}(1), \overline{g}(\mu(1)))_{\overline{h},\overline{g}}$ is not successful,

or

(ii) $\quad |\overline{h}(i)| \geq |\overline{g}(i)| \quad$ for $i = 0,1$

Proof. In this case the relation $h(0x) = g(\mu(0)y)$, with x and y minimal, implies that $x \in (10)^*1$ and $y \in (\mu(0))^*$. So by symmetry,

$$\overline{h}(0) \in (01)^+ , \qquad \overline{g}(0) \in (\mu(0))^+ ,$$
$$\overline{h}(1) \in (10)^+ , \qquad \overline{g}(1) \in (\mu(1))^+ .$$

Consequently, the result follows.

The case VII is one of our special cases from section 5 (cf. Lemma 5.1). So our analysis of case (i) of Lemma 7.1 is finished.

Now we consider case (ii) of Lemma 7.1.

Lemma 7.5. Let (h,g) be a successful and balanced pair of homomorphisms such that $\{h(i), g(\mu(i))\} \subseteq j^*$ for some i and j in $\{0,1\}$, and let $\mathrm{ecol}(h,g) = (\overline{h},\overline{g})$. If $\sigma(\overline{h},\overline{g}) = \sigma(h,g)$, then either the pair $(\overline{h},\overline{g})$ is not successful or h and g or $\overline{h}$ and $\overline{g}$ are in one of the following forms:

(i)
$$h(i) \in j^*, \qquad g(\mu(i)) \in j^*,$$
$$h(1-i) \in (1-j)^* j^k, \qquad g(\mu(1-i) \in (1-j)^* j^\ell,$$

where i and j are in $\{0,1\}$, $k\ell = 0$ and if $k \neq 0$ (resp. $\ell \neq 0$) then $|g(\mu(i))| > k$ (resp. $|h(i)| > \ell$); or

(ii)
$$h(i) = j, \qquad g(\mu(i)) = j,$$
$$h(1-i) = ((1-j)j^n)^N(1-j), g(\mu(1-i)) = ((1-j)j^n)^M(1-j),$$
where i and j are in $\{0,1\}$, $n > 0$ and $N \neq M$; or

(iii)
$$h(i) = j\, j^n, \qquad g(\mu(i)) = j\, j^\ell,$$
$$h(1-i) = (1-j)j^m, \qquad g(\mu(1-i)) = (1-j)j^k,$$
where i and j are in $\{0,1\}$, and $n,m,k,\ell \geq 1$ with $n+m \neq k+\ell$.

Proof. By symmetry, we may set $i=0$ and $j=0$. Hence our starting point is as follows:

$$h(0) \in 0^+, \qquad g(\mu(0)) \in 0^+,$$
$$h(1) \in 1\Sigma^*, \qquad g(\mu(1)) \in 1\Sigma^*.$$

As in the proof of Lemma 7.1. we iteratively apply the basic argument presented therein. We consider separately three different cases.

I. $h(0) = 0 = g(\mu(0))$.

Since (h,g) is balanced if $h(1)$ or $g(\mu(1))$ contains 11 as a subword then they are both in 1^+ (cf. case I of the proof of Lemma 7.1). Consequently, (i) is satisfied or otherwise $h(1), g(\mu(1)) \in \underline{\text{pref}}\ (10^+)^+$. In the latter case the suffixes of $1^{-1}h(1)$ and $1^{-1}g(\mu(1))$ may be encountered in the construction of $(\overline{h},\overline{g})$ only if $\underline{\text{suf}}_1\ (h(1)) = \underline{\text{suf}}_1(g(\mu(1))) = 1$. Consequently, $h(1), g(\mu(1)) \in (10^+)^+1$. Further the number of 0's between any two occurrences of 1 must be the same. Otherwise $h(1)$ or $g(\mu(1))$ would contain a suffix which is not met in contructing $(\overline{h},\overline{g})$. Hence, $h(1) = (10^n)^N1$ and $g(\mu(1)) = (10^n)^M1$ for some $n,N,M \geq 1$. If $N = M$, then $|\overline{h}(i)| = 1 = |\overline{g}(i)|$, for $i = 0,1$, and hence the equality $\sigma(\overline{h},\overline{g}) = \sigma(h,g)$ does not hold.

II. $h(0) = 0$ and $g(\mu(0)) \in 00^+$ (or symmetrically $h(0) \in 00^+$ and $g(\mu(0)) = 0$). Since (h,g) is balanced $|h(1)| > 1$. If $g(\mu(1)) = 1$, then clearly $h(1) = 1^+0$ and so (i) is satisfied. If, in turn, $|g(\mu(1))| > 1$ then either $h(1)$ and $g(\mu(1))$ are in 1^+ or they are in $\underline{\text{pref}}\ (10^+)^+$ as in case I. The first possibility leads to (i). In the second case we conclude, by the fact $g(\mu(0)) \in 00^+$, that $h(1) \in 10^+$. Furthermore the number of 0's in $h(1)$ is either 0 or 1. This is seen since $h(0) = 0$, $g(\mu(0)) \in 00^+$ and $g(\mu(0)) \in \underline{\text{pref}}\ (10^+)^+$. Consequently, $h(1) = 1$ or $h(1) = 10$ and so (h,g) is unbalanced, which completes the case II.

III. $h(0) \in 00^+$ and $g(\mu(0)) \in 00^+$. If one of the $h(1)$, $g(\mu(1))$ equals 1 then the other one belongs to 1^+0, say $h(1) = 1$ and $g(\mu(1)) \in 1^+0^\ell$. If $\ell > |g(\mu(0))|$ then $0^{\ell-1}$ appears in $\underline{\text{p-suf}}\{g(0),g(1)\}$ but is not encountered in the construction of $(\overline{h},\overline{g})$, because $|h(0)| \geq 2$ and $h(1) = 1$. Consequently, $|g(\mu(0))| \geq \ell$. Since (h,g) is balanced, $|h(0)| > \min\{|g(\mu(0))|, |g(\mu(1))|\} \geq \ell$. So (i) is satisfied in this case.

We still have to consider the case when $|h(1)| \geq 2$ and $|g(\mu(1))| \geq 2$. In that case both $h(1)$ and $g(\mu(1))$ are in 1^+ or 10^+. The first possibility yields (i).

The second possibility is handled as follows. If (h,g) does not satisfy (iii), i.e. the length requirement is not fulfilled, then $|h(01)| = |g(01)|$. Consequentely, $\overline{h}(0) \in 0^+$, $\overline{h}(1) \in$ pref (10), $\overline{g}(0) \in (\mu(0))^+$ and $\overline{g}(1) \in$ pref $(\mu(1)\ \mu(0))$. If $|\overline{h}(1)| = 1$ or $|\overline{g}(1)| = 1$ then we are done: $(\overline{h},\overline{g})$ is of the form (i). So let $\overline{h}(1) = 10$ and $\overline{g}(1) = \mu(1)\ \mu(0)$. Now μ must be the identity; otherwise $(\overline{h},\overline{g})$ is not successful. Finally we use the assumption $\sigma(h,g) = \sigma(\overline{h},\overline{g})$ to conclude that $(\overline{h},\overline{g})$ is of the form (iii). Indeed, by this assumption, $|\overline{h}(0)| = |g(\mu(0))|$ and $|\overline{g}(0)| = |h(0)|$ and so $|\overline{h}(01)| \neq |\overline{g}(01)|$.

Hence our proof of Lemma 7.5 is complete.

A pair (h,g) of homomorphisms is called <u>special</u> if it is either in one of the forms (i)-(iii) from the statement of Lemma 7.5. or it is of the form VII (see the listing of forms following Lemma 7.1).

An instance I of GPCP(2) is called <u>special</u> whenever the pair of homomorphisms in I is special. Using these notions we combine now the results of this section.

<u>Basic Lemma 7.6</u> . Let (h,g) be a pair of marked homomorphisms. Then at least one of the following conditions holds true:

a) (h,g) is not successful,

b) (h,g) is unbalanced,

c) (h,g) is special,

d) $\sigma(\overline{h},\overline{g}) < \sigma(h,g)$,

e) $(\overline{h},\overline{g})$ is either special, unbalanced or not successful,

where in the last two cases $(\overline{h},\overline{g}) = \text{ecol}(h,g)$.

We conclude this section by observing that if I is a successful instance of GPCP(2), then ecol(I), with few exceptions, is strictly smaller than I. Moreover, in the exceptional cases either ecol(I) is not successful or either I or ecol(I) is of the form dealt with in section 5. Consequently the base for induction has been laid.

8. BASIC INDUCTION STEP

In this section we show how to use the ecol construction in solving GPCP(2). First we extend the notion of the ecol transformation to instances of GPCP(2).

Definition 8.1. Let $I = (h,g,a_1,b_1,a_2,b_2)$ be a marked instance of GPCP(2) such that the sequence $(a_1,a_2)_{h,g}$ and the pair (h,g) of homomorphisms are successful and the equation $h(u)b_1 = g(w)b_2$ has a solution. An *equality collector* of I, in symbols ecol(I), is any instance $J = (\overline{h},\overline{g},\overline{a}_1,\overline{b}_1,\overline{a}_2,\overline{b}_2)$ of GPCP(2) such that

$$(\overline{h},\overline{g}) = \mathrm{ecol}(h,g),$$
$$\overline{a}_1 = h^{-1}(a_1^{-1}\ s((a_1,a_2)_{h,g})),$$
$$\overline{a}_2 = g^{-1}(a_2^{-1}\ s((a_1,a_2)_{h,g})),$$

and $(\overline{b}_1,\overline{b}_2)$ is a minimal solution of the equation $h(u)b_1 = g(w)b_2$, i.e. a solution such that the equation does not have any solution (u',w') satisfying u' *p-pref* $\overline{b}_1$ or w' *p-pref* $\overline{b}_2$. The set of all equality collectors of I is denoted by ECOL(I).

Clearly, $\overline{a}_1$ and $\overline{a}_2$ are unique, while the pair $(\overline{b}_1,\overline{b}_2)$ need not be unique. Indeed, there may be one or two minimal solutions, one of the form $(0u_1,\mu(0)w_1)$ and the other of the form $(1u_2,\mu(1)w_2)$. Consequently, $\#\mathrm{ECOL}(I) \le 2$. Observe also that since h and g are marked

$$|\overline{a}_1|, |\overline{a}_2| \le \max\{|h(01)|,|g(01)|\} + |a_1a_2|$$

and

$$|\overline{b}_1|, |\overline{b}_2| \le \max\{|h(01)|,|g(01)|\} + |b_1b_2|.$$

Definition 8.2. Let I be a marked instance of GPCP(2) such that $\mathrm{ECOL}(I) \ne \emptyset$. Then we say that I is *successful*. Otherwise I is called *unsuccessful*.

Basically because of Lemma 6.1., it is decidable whether a given instance I is successful.

The following result underlies the use of the ecol transformation in solving GPCP(2).

Theorem 8.1. Let $I = (h,g,a_1,b_1,a_2,b_2)$ be a successful instance of GPCP(2). Then I has a solution if and only if it has a solution no longer than $k = 2\max\{|h(01)|,|g(01)|\} + |a_1a_2b_1b_2|$ or for some J in ECOL(I) J has a solution.

Proof. Assume first that I has a solution γ with $|\gamma| \geq k$, i.e. $a_1h(\gamma)b_1 = a_2g(\gamma)b_2$. Since $|\gamma| \geq k$ and I is successful we may decompose γ in two ways

(1) $$\gamma = \gamma_{0,1}\,\gamma_{1,1}\cdots\gamma_{t,1} = \gamma_{0,2}\,\gamma_{1,2}\cdots\gamma_{t,2}\;,\ \text{for some } t \geq 1,$$

where

$$\begin{aligned} &a_1h(\gamma_{0,1}) = a_2g(\gamma_{0,2}),\\ &h(\gamma_{i,1}) = g(\gamma_{i,2})\;,\ \text{for}\quad i = 1,\dots,t\text{-}1,\\ &h(\gamma_{t,1})b_1 = g(\gamma_{t,2})b_2\;, \end{aligned}$$

and moreover none of these equations holds true for a pair of words where at least one of the components is a proper prefix of the given one. Consequently, $(\gamma_{t,1},\gamma_{t,2})$ is a minimal solution in the sense of Definition 8.1. Let J be the ecol version of I associated with this minimal solution and let

$$\sigma = \underline{\text{pref}}_1(\gamma_{1,1})\cdots\underline{\text{pref}}_1(\gamma_{t-1,1}).$$

Then, by the definition of J and by (1), we obtain

$$\begin{aligned} \overline{a}_1\,\overline{h}(\sigma)\,\overline{b}_1 &= \gamma_{0,1}\,\gamma_{1,1}\cdots\gamma_{t,1} = \gamma\\ &= \gamma_{0,2}\,\gamma_{1,2}\cdots\gamma_{t,2} = \overline{a}_2\,\overline{g}(\sigma)\,\overline{b}_2 \end{aligned}$$

Hence, σ is a solution of J.

Conversely, assume that an ecol version $J = (\overline{h},\overline{g},\overline{a}_1,\overline{b}_1,\overline{a}_2,\overline{b}_2)$ has a solution ρ, i.e.

$$\overline{a}_1\ \overline{h}(\rho)\ \overline{b}_1 = \overline{a}_2\ \overline{g}(\rho)\ \overline{b}_2\ .$$

Recalling the definition of the ecol version we see that

$$\begin{aligned}
&a_1\ h(\overline{a}_1\ \overline{h}(\rho)\ \overline{b}_1)b_1 = \\
&a_1\ h(\overline{a}_1)\ h(\overline{h}(\rho))\ h(\overline{b}_1)b_1 = \\
&a_2\ g(\overline{a}_2)\ g(\overline{g}(\rho))\ g(\overline{b}_2)b_2 = \\
&a_2\ g(\overline{a}_2\ \overline{g}(\rho)\overline{b}_2)b_2\ .
\end{aligned}$$

Consequently, $\tau = \overline{a}_1\ \overline{h}(\rho)\ \overline{b}_1 = \overline{a}_2\ \overline{g}(\rho)\ \overline{b}_2$ is a solution of I.

9. UNSUCCESSFUL INSTANCES

Here we settle the case of unsuccessful instances of GPCP(2).

Theorem 9.1. It is decidable whether or not an unsuccessful instance $I = (h,g,a_1,b_1,a_2,b_2)$ of GPCP(2) has a solution.

Proof. We have to consider quite a large number of different cases depending on the way in which I is unsuccessful.

I. $(a_1,a_2)_{h,g}$ blocks.
Now the equation $a_1h(x_1)\ \underline{\text{Pref}}\ a_2h(x_2)$ holds true for a finite number of pairs (x_1,x_2) only, and consequently if I has a solution then it is found by checking through a finite set of words.

II. $(a_1,a_2)_{h,g}$ is infinite.
In this case we first search words t_1,t_2,p_1 and p_2 such that the equation

$$a_1h(x_1)\ \underline{\text{Pref}}\ a_2g(x_2)$$

implies

$$x_i \in \underline{\text{pref}}\ (t_ip_i^*) \quad , \quad \text{for } i = 1,2.$$

Clearly, possible solutions of I are among the common prefixes of $t_1p_1^*$ and $t_2p_2^*$. If $\underline{\text{pref}}\ (t_1p_1^*) \neq \underline{\text{pref}}\ (t_2p_2^*)$, then there are only a finite number of candidates to be checked and hence we are done. So let t and p be words such that

$$(1) \qquad \underline{\text{pref}}\ (t_1p_1^*) = \underline{\text{pref}}\ (tp^*) = \underline{\text{pref}}\ (t_2p_2^*).$$

The case $|h(p)| \neq |g(p)|$, is not difficult to settle.

Indeed, in this case we can effectively find a constant k such that for words x in (1), with $|x| \geq k$,

$$\big||a_1h(x)| - |a_2g(x)|\big| > \max\{|b_1^{-1}b_2|, |b_2^{-1}b_1|\}$$

which guarantees that if I has a solution it has such one shorter than k.

If, in turn, $|h(p)| = |g(p)|$ then the sets

$$L_1 = \{(a_1h(x))^{-1} a_2g(x) \mid x \in \underline{\text{pref}}\ (tp^*)\}$$

and

$$L_2 = \{(a_2g(x))^{-1} a_1h(x) \mid x \in \underline{\text{pref}}\ (tp^*)\}$$

are finite and hence to test whether or not I has a solution it suffices to test whether there exists a word w in L_1 such that $wb_1 = b_2$ or whether there exists a word u in L_2 such that $b_1 = ub_2$.

This completes the proof of case II.

III. $(a_1,a_2)_{h,g}$ is successful.

Here we have several subcases.

(i) both $(h(0), g(\mu(0)))_{h,g}$ and $(h(1),g(\mu(1)))_{h,g}$ blocks.

Here we can apply the reasoning from case I.

(ii) $(h(0), g(\mu(0)))_{h,g}$ is infinite and $(h(1), g(\mu(1)))_{h,g}$ blocks, or the other way around.

This case can be settled analogously to the case II.

Indeed, if a solution exists then it can be found from an effectively constructible set $F \cup \underline{\text{pref}}\ (tp^*)$, where F is a finite set and $t,p \in \{0,1\}^*$.

(iii) $(h(0), g(\mu(0)))_{h,g}$ is successful and $(h(1),g(\mu(1)))_{h,g}$ blocks, or vice versa.

This is essentially similar to case III (ii). Now, if a solution exists then it can be found in a set F ∪ pref (tp^*s), where F is a finite set and $t,p,s \in \{0,1\}^*$.

(iv) both $(h(0), g(\mu(0)))_{h,g}$ and $(h(1), g(\mu(1)))_{h,g}$ are successful. Since I is unsuccessful the equation $h(u)b_1 = g(w)b_2$ has no solution (u,w). Consequently, if I has solutions at all they must be among the prefixes of the word $\bar{a}_1 = h^{-1}(a_1^{-1}(s((a_1,a_2)_{h,g}))$.

(v) both $(h(0), g(\mu(0)))_{h,g}$ and $(h(1), g(\mu(1)))_{h,g}$ are infinite. This is again essentially similar to case II. Now we can effectively construct words t,p,q and s in $\{0,1\}^*$ such that if I has a solution then it has a solution in the set pref (tp^*) ∪ pref (sq^*). Consequently, a solution of I can be effectively found (if it exists at all).

(vi) $(h(0), g(\mu(0)))_{h,g}$ is successful and $(h(1), h(\mu(1)))_{h,g}$ is infinite, or vice versa. This is the most complicated case. We first look for words t_i,p_i,r_i,q_i,o_i and s_i, for $i = 1,2$, such that p_i and q_i are nonempty and the equation

(2) $$a_1h(x_1) \underline{\text{Pref}}\ a_2g(x_2)$$

implies that, for both i = 1 and i = 2, either

(3) $$x_i \in \underline{\text{pref}}\ (o_iq_i^*)$$

or

(4) $$x_i \in \underline{\text{pref}}(t_ip_i^ns_iq_i^*) \quad \text{for some } n \geq 0.$$

Whether the first case leads to a solution can be decided as in case II.

So it remains to be shown that we can decide whether or not I has a solution satisfying (4). We have two subcases.

a) $|p_1| \neq |p_2|$. First we look for a constant n_0 such that if the languages pref $(t_1p_1^ns_1q_1^*)$ and pref $(t_2p_2^ns_2q_2^*)$, for each $n \geq n_0$, have arbitrarily long common prefixes, then necessarily, for $n \geq n_0$ and $i = 1,2$,

$$(5) \qquad t_i p_i{}^n r_i q_i{}^* \subseteq \underline{\text{pref}}\ (tp^*)$$

where t and p are some words in $\{0,1\}^*$. If the above common prefix condition is not satisfied, then possible solutions can be found applying, possibly several times, arguments from case II. If this condition is satisfied, then again the existence of a solution of I satisfying (4), with $n \geq n_0$, can be decided as in case II. The same holds true also for each $n < n_0$.

b) $|p_1| = |p_2|$. Now if the languages $\underline{\text{pref}}\ (t_1 p_1{}^n s_1 q_1{}^*)$ and $\underline{\text{pref}}\ (t_2 p_2{}^n s_2 q_2{}^*)$ have arbitrarily long common prefixes only for small values of n, say $n \leq n_1$, then the existence of a solution of I can be decided using several times arguments from II. In the other case we can find words t, p, r_1' and r_2' in $\{0,1\}^*$ such that p, r_1' and r_2' are nonempty and

$$t_i p_i{}^n r_i' q_i{}^* = tp^n r_i' q_i{}^* \quad \text{for } i = 1,2 \quad \text{and} \quad n \geq 0.$$

We should be able to decide whether there exists a solution of I in $\underline{\text{pref}}\ (tp^n r_1' q_1{}^*) \cap \underline{\text{pref}}\ (tp^n r_2' q_2{}^*)$, for some $n \geq 0$. Let first $|h(p)| = |g(p)|$. Assuming that $|a_1 h(t)| \geq |a_2 g(t)|$ we may find a constant n_2 such that

$$(a_2 g(tp^{n_2}))^{-1}\ a_1 h(tp^{n_2}) = (a_2 g(tp^{n_2+1}))^{-1}\ a_1 h(tp^{n_2+1}).$$

Consequently, if I has a solution at all then it has a solution in

$$\underline{\text{pref}}\ (tp^{n_2} r_1' q_1{}^*) \cap \underline{\text{pref}}\ (tp^{n_2} r_2' q_2{}^*) \text{ for some } n \leq n_2+1.$$

Hence the methods of II becomes applicable.

We still have to consider the case when $|h(p)| \neq |g(p)|$, say $|g(p)| > |h(p)|$. Let first $|g(q_2)| \geq |h(q_2)|$. We look for a constant n_3 such that

$$|a_2\ g(tp^n p')| - |a_1\ h(tp^n p')| > \max\ \{|b_1^{-1}b_2|\ ,\ |b_2^{-1}b_1|\}$$

for all $n \geq n_3$ and $p' \in \underline{\mathrm{pref}}\ (pr_2'q_2{}^*)$. Consequently, I cannot have solutions in $tp^n\ \underline{\mathrm{pref}}\ (pr_2'q_2{}^*)$, for $n \geq n_3$. Whether or not I has solutions in $\underline{\mathrm{pref}}\ (tp^n r_2' q_2{}^*)$ for some fixed $n < n_3$ can be decided as in case II.

Secondly, let $|g(q_2)| < |h(q_2)|$. Let m and k be natural numbers satisfying $m(|g(p)| - |h(p)|) = k(|h(q_2)| - |g(q_2)|) > |g(pr_2')| + |h(q_2)|$ and $k|g(q_2)| > |h(q_2)| + \max\ \{|b_1^{-1}b_2|\ ,\ |b_2^{-1}b_1|\}$.
Now we choose a constant n_4 such that

(5) $$|a_1\ h(tp^n)| \geq |a_1 a_2| + |h(tpr_2')| + |g(tq_2)|$$

and

(6) $$|a_2\ g(tp^n p')| - |a_1\ h(tp^n p')| > \max\ \{|b_1^{-1}b_2|, |b_2^{-1}b_1|\} + \\ + k(|h(q_2)| - |g(q_2)|)$$

for all $n \geq n_4$ and $p' \in \underline{\mathrm{pref}}\ (pr_2'q_2)$. Assume that I has a solution $w = tp^N x$, with $N \geq n_4 + m$ and $x \in \underline{\mathrm{pref}}\ (r_2'q_2{}^*)$. We claim that I has a solution $tp^{N-m}\overline{x}$ for some $\overline{x} \in \underline{\mathrm{pref}}\ (r_2'q_2{}^*)$.

Since w is a solution of I,(6) implies that $x = r_2'q_2^k x'$ for some x'. Let $w' = tp^{N-m}r_2'x'$. Then, by the choice of k and m,

(7) $$|a_2\ g(w)| - |a_1\ h(w)| = |a_2\ g(w')| - |a_1\ h(w')|.$$

Further, from the choice of k, from (6) and from the fact that $tp^N x$ is a solution of I, it follows that

$$(a_1\ h(tp^N r_2'))^{-1}\ a_2\ g(tp^N r_2')\ \underline{\mathrm{pref}}_{|h(q_2)| + |g(q_2)|}\ (g(q_2)^*) \in \underline{\mathrm{pref}}\ (h(q_2)^*).$$

This together with (6), the choice of k and (5) implies that

$$\underline{\mathrm{suf}}\ (g(p^* r_2'q_2{}^*)) = \underline{\mathrm{suf}}\ (h(q_2)^* q')\ \text{and}\ \underline{\mathrm{suf}}(h(p^*)r_2') = \underline{\mathrm{suf}}(h(q_2)^*)$$

for some $q' \in \underline{\mathrm{pref}}\ (h(q_2))$. Consequently, by (7) showing that w' is also a solution of I. In the above we have used three times the following well known result (see, e.g., [H]): if two words α^k and β^m have a common prefix of the length $|\alpha| + |\beta|$ then α and β are powers of the same word. So we conclude that if I has solutions at all, then it has solutions in $\underline{\text{pref}}(tp^n r_2' q_2^*)$, for some $n \leq n_4 + m$. Hence the methods of case II become applicable, which completes our proof of case III and hence the proof of the theorem.

10. MAIN RESULT

Finally, we are ready to establish our main result.

Theorem 10.1. It is decidable whether or not an arbitrary instance of PCP(2) has a solution.

Proof. The algorithm is described in the flowchart below. The following comments are in order. The block of the form $\langle \emptyset \rangle$ denotes that at this point we decide whether or not an instance of GPCP(2) in question has a solution. Furthermore the word "special" refers to special instances defined in section 7.

That our algorithm terminates is seen as follows.

Certainly, it is decidable whether or not a given instance I of GPCP(2) is periodic, unbalanced, special or successful.For each of these the existence of a solution can be decided by Theorem 4.1., Theorem 5.1., Lemmas 5.1 - 5.4. and Theorem 9.1., respectively. Moreover, by the Basic Lemma, if I is in none of the above forms, then the instances in ECOL(I) are strictly smaller than I, i.e. $(\overline{h},\overline{g}) < (h,g)$, with the possible exception of the case (e) of the Basic Lemma. And in that case the ecol(I)'s are either special or unbalanced or unsuccessful which guarantees the termination during the next cycle. Consequently, our algorithm will always terminate.

That the answer obtained is correct follows from Theorem 8.1. and the Reduction Lemma. Theorem 8.1 also gives an upper bound for the length of solutions needed to be considered separately. A bound is $K = 2\max\{|h(01)|, |g(01)|\} + |a_1a_2b_1b_2|$, depending on an instance used to define the ecol version in question.

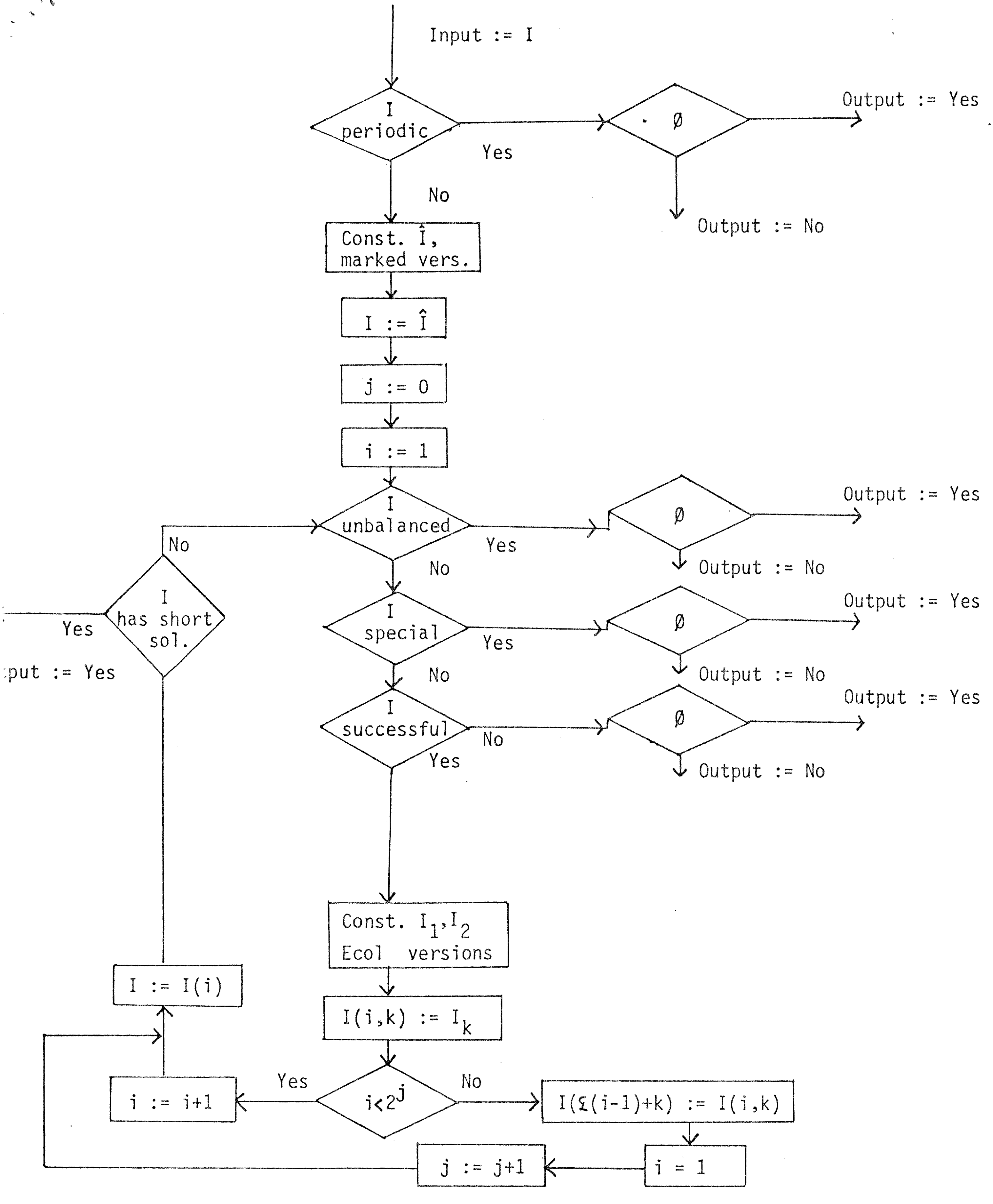


Figure 1: PCP(2)-algorithm.

Actually we have proved even a stronger result.

Theorem 10.2. It is decidable whether or not an arbitrary instance of GPCP(2) has a solution.

Proof. The only difference from above is that now when constructing marked versions we obtain a finite set of new instances instead of one only, c.f. Lemma 3.2.

ACKNOWLEDGMENTS. The authors are indebted to D. Janssens, A. Salomaa and R. Verraedt for useful comments. The first and the third author gratefully acknowledge support under National Science Foundation grant number MCS 79-03838. The second author is grateful to the Finnish Academy for providing excellent working conditions.

REFERENCES.

[BB] Book, R.V. and Brandenburg, F.J., Equality sets and complexity classes *SIAM J. of Comp.,* to appear.

[C] Culik, K., II, A purely homomorphic characterization of recursively enumerable sets, *J. of the ACM* 26, 345-350, 1979.

[Cl] Claus, V., Die Grenze zwischen Entscheidbarkeit und Nichtentscheidbarkeit, Fernstudienkurs für die Fernuniversität Hagen, Open University, Hagen, 1979.

[CK] Culik, K., II and Karhumaki, J., On the equality sets for homomorphisms on free monoids with two generators, *R.A.I.R.O., Inf. Theorique,* to appear.

[ER] Engelfriet, J. and Rozenberg, G., Fixed point languages, equality languages and representations of recursively enumerable languages, *J. of the ACM,* to appear.

[EhR] Ehrenfeucht, A. and Rozenberg, G., Generalized Post Correspondence Problem of length 2; Parts I, II and III. Technical Reports of the Computer Science Department, University of Colorado at Boulder, 1980.

[G] Ginsburg, S., *The Mathematical Theory of Context-Free Languages,* McGraw-Hill, New York 1966.

[H] Harrison, M.A., *Introduction to formal language theory,* Addison-Wesley Publ., 1978.

[HU] Hopcroft, J.E. and Ullman, J.D., *Introduction to Automata Theory, Languages and Computation,* Addison-Wesley Publ., 1979.

[KS] Karhumaki, J. and Simon, I., A note on elementary homomorphisms and the regularity of equality sets, *Bulletin of the EATCS* 9, 1979.

[Le] Lecerf, Y., Recursive insolubilitē de l'equation generale de diagonalisation de deux monomorphisms de monoides libres $\Psi x = \Psi x$, *Comptes rendus* 257, 2940-2943, 1963.

[P] Post, E.L., A variant of a recursively unsolvable problem, *Bull. of the Am. Math. Soc.,* 52, 264-268, 1946.

[S1] Salomaa, A., *Formal Languages,* Academic Press, 1973.

[S2] Salomaa, A., Equality sets for homomorphisms on free monoids, *Acta Cybernetica*-4,127-239, 1978.