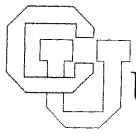
Generalized Post Correspondence Problem of Length 2
Part I:
Some special cases and the basic transformation

A. Ehrenfeucht
G. Rozenberg

CU-CS-188-80



University of Colorado at Boulder
DEPARTMENT OF COMPUTER SCIENCE

GENERALIZED POST CORRESPONDENCE PROBLEM OF LENGTH 2
PART I:
Some special cases and the basic transformation

by

A. Ehrenfeucht*
and
G. Rozenberg†

CU-CS-188-80 September, 1980

*A. Ehrenfeucht, Dept. of Computer Science, University of Colorado, Boulder, Colorado 80309 USA

†G. Rozenberg, Institute of Applied Mathematics and Computer Science University of Leiden, 2300 RA Leiden, The Netherlands

All correspondence to G. Rozenberg.

ABSTRACT

Let Σ, Δ be finite alphabets with the cardinality of Σ equal two, let h, g be homomorphisms from Σ^* into Δ^* and let $a_1, a_2, b_1, b_2 \in \Delta^*$. The *Generalized Post Correspondence Problem of length* 2 (GPCP(2) for short) is to determine whether or not there exists a word w in Σ^+ such that $a_1 h(w) a_2 = b_1 g(w) b_2$. This paper is the first one in the sequence of three papers which together demonstrate that GPCP(2) is decidable. As a special case of this result one gets that the celebrated Post Correspondence Problem of length 2 is decidable. In this paper we discuss several special cases of GPCP(2) and we introduce and study a transformation of GPCP(2) which turns out to be very fundamental in our solution of the GPCP(2) problem.

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.

Definition. 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 very interesting results concerning 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.

Definition. Let Σ, Δ be alphabets, h, g be two homomorphisms from Σ^* into Δ^* and let a_1, a_2, b_1, 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 $a_1 h(w) a_2 = b_1 g(w) b_2$.
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 $a_1 = a_2 = b_1 = b_2 = \Delta$ then GPCP(n) reduces to PCP(n).

This paper is the first one of three papers which together prove that GPCP(2) is decidable. In the present paper we consider several special cases of GPCP(2) as well as introduce the basic construct of our solution: the equality collector of a pair of marked homomorphisms.

0. PRELIMINARIES

In this paper we use mostly standard language-theoretic notation and terminology. Perhaps the following points require additional comments.

We consider finite alphabets only.

For an integer n, $abs(n)$ denotes the absolute value of n.

For a finite set Z, #Z denotes its cardinality.

Λ denotes the empty word. For a word x, $pref(x)$ denotes the set of prefixes of x and, for a positive integer n, $pref_n(x)$ denotes the prefix of x of the length n. If $x \in pref(y)$ then we write $x\ pref\ y$; if either $x\ pref\ y$ or $y\ pref\ x$ then we write x PREF y. For a language K, $pref(K) = \bigcup_{w \in K} pref(w)$. We have analogous notation for suffixes replacing everywhere above $pref$ by suf and PREF by SUF. For a letter c, $\#_c x$ denotes the number of occurrences of c in x and c-$pref(x)$ denotes the maximal prefix of x consisting of c's only. For a nonempty word x, $first(x)$ denotes the first letter of x and $last(x)$ denotes the last letter of x. For words x, y we say that they are *cyclic conjugates*, denoted $x \sim y$, if there exist words z_1, z_2 such that $x = z_1 z_2$ and $y = z_2 z_1$. We define $tail(x,y) = (u,w)$ if z is the longest common prefix of x and y, $x = z u$ and $y = z w$. If $y = x z$ then $x\backslash y = z$ and $y/z = x$.

As usual to avoid a very cumbersome notation and terminology we will sometimes use terms "letter," "word" and "subword" when we really mean an *occurrence* of a letter, an occurrence of a word or an occurrence of a subword respectively. However, as usual, it

should not lead to a confusion.

An infinite sequence $\{x_n\}_{n\geq 1}$ of nonnegative integers is *ultimately periodic* if there exist integers $t \geq 0$ and $p \geq 1$ such that $x_{n+p} = x_n$ for each $n > t$. The smallest t satisfying the above is called the *threshold* of $\{x_n\}$ and the smallest p satisfying the above is called the *period* of $\{x_n\}$.

Let $\{A_n\}$ be an infinite sequence of objects that is ultimately periodic; let t be its threshold and p its period. We say that this sequence can be effectively constructed if there exists an algorithm that constructs the first (t+p) objects of it.

In this paper we consider propagating homomorphisms only, that is homomorphisms h with the property that for no letter c, $h(c) = \Lambda$. Thus whenever we write a homomorphism we mean a propagating one.

For a homomorphism h of Σ^*, $maxr(h) = \max\{|h(c)| : c \in \Sigma\}$.

For a pair of homomorphisms h, g of Σ^*, $maxr(h,g) = \max\{maxr(h), maxr(g)\}$.

The following notion and the result (from [EhR1]) are quite basic in the study of homomorphisms on free monoids.

Definition 0.1. A homomorphism $h : \Sigma^* \to \Delta^*$ is *simplifiable* if there is an alphabet Θ with $\#\Theta < \#\Sigma$ and homomorphisms $f : \Sigma^* \to \Theta^*$, $g : \Theta^* \to \Delta^*$ such that $h = g\,f$. Otherwise h is called *elementary*. □

Theorem 0.1. Let $h : \Sigma^* \to \Delta^*$ be an elementary homomorphism with $\Sigma = \{c_1,\ldots,c_k\}$, $k \geq 1$. Consider $U = \{h(c_1),\ldots,h(c_k)\}$. Assume that $h(c_i)\,x\,\gamma = h(c_j)y$ for $i \neq j$, $\gamma \in \Delta^*$ and $x, y \in U^*$. Then $|h(c_i)x| \leq |h(c_1)h(c_2)\ldots h(c_k)| - k$. □

In this paper we will often consider equations in a free monoid. In particular the following type of equations will turn out to be useful. The following definition and result are from [EhR1].

Definition 0.2. Let Σ, Δ be alphabets with $\#\Sigma = 1$. Let f_1, f_2, g_1, g_2 be homomorphisms from Σ^* into Δ^* and let $a_1, a_2, a_3, b_1, b_2, b_3 \in \Delta^*$. Then a *unary 2-fold equation* is an equation of the form

$$a_1 f_1(x) a_2 f_2(y) a_3 = b_1 g_1(x) b_2 g_2(y) b_3 \quad \ldots\ldots (0.1)$$

in variables x, y. A *solution* of (0.1) is an ordered pair (α,β) with $\alpha, \beta \in \Sigma^*$ such that

$$a_1 f_1(\alpha) a_2 f_2(\beta) a_3 = b_1 g_1(\alpha) b_2 g_2(\beta) b_3.$$

If (0.1) has a solution then we say that it is *solvable*. □

The equation

$$a_1 f_1(x) a_2 = b_1 g_1(x) b_2 \quad \ldots\ldots (0.2)$$

which is a special case of the equation (0.1) (set $a_3 = b_3 = \Lambda$ and f_2 equal g_2 equal the identity on Σ^*) is referred to as a *unary 1-fold equation*.

Theorem 0.2. It is decidable whether or not an arbitrary unary 2-fold equation is solvable. □

Perhaps the most useful problem in considering decision problems within formal language theory is the following problem studied first by E. Post [P].

Definition 0.3. 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). □

In this and in two follow-up papers ([EhR3] and [EhR4]) we will demonstrate that even a more general problem than PCP(2) is decidable. The generalization that we study is defined as follows.

Definition 0.4. Let Σ, Δ be alphabets, h, g be two homomorphisms from Σ^* into Δ^* and let a_1,a_2,b_1,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 $a_1h(w)a_2 = b_1g(w)b_2$. 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 will be concerned with GPCP(2). It is clear that as far as the decidability of GPCP(2) is concerned one can restrict oneself to range alphabets of cardinality two and moreover, one can assume that the domain alphabet (Σ) and the range alphabet (Δ) are identical. Consequently, *unless explicitly stated otherwise, in this paper we consider homomorphisms of* Σ^* *into* Σ^* (*endomorphisms of* Σ^*) *where* $\#\Sigma = 2$. *We also set for this paper* $\Sigma = \{0,1\}$. Moreover we assume that given an instance $I = (h,g,a_1,a_2,b_1,b_2)$ of GPCP(2), a_1 PREF b_1 and a_2SUF b_2 since otherwise I has no solution.

We would like to finish this section with the following comment. Very often in this paper we will have a situation of the following kind. First, we compute explicitly a positive integer constant C and then we have a statement like this: "If a word w is such that $|w| \geq C$ then a certain property P of w holds." We do not intend here to provide the best upper bounds, and so the only meaning of such a statement is that for every positive integer $D \geq C$ we have:

"If a word w is such that $|w| > D$ then P of w holds."

Moreover, after we have computed explicitly constants like these in a number of similar situations we will switch to the statement of the kind: "One can effectively compute a positive integer constant C such that if a word w is such that .....", leaving the explicit computation of C to the reader.

1. SOME SPECIAL CASES

In this section we demonstrate the decidability of GPCP(2) in several special (rather "easy") cases. The results of this section will be used later on to settle more involved cases.

Theorem 1.1. It is decidable whether or not an arbitrary instance $I = (h,g,a_1,a_2,b_1,b_2)$ of GPCP(2) such that $|h(0)| \leq |g(0)|$ and $|h(1)| \leq |g(1)|$ has a solution.

Proof.

Consider the following algorithm. Let us generate all the words over {0,1} in their "natural" order (that is first according to the length and within the given length the words are ordered lexicographically; thus first few words in this order are 0,1,00,01,10,11,000,...). Each time a word, say w, is generated we check whether or not w satisfies one of the following three conditions:

(a). it is not true that $a_1h(w)$ PREF $b_1g(w)$,

(b). $|b_1g(w)| - |a_1h(w)| > abs(|a_2| - |b_2|)$,

(c). $w = w_1w_2$, $w_1 \neq \Lambda$, $w_2 \neq \Lambda$ and $tail(a_1h(w_1), b_1g(w_1)) = tail(a_1h(w), b_1g(w))$.

If w satisfies one of the conditions (a) through (c) then we say that w is a *stop word* and in our generating process we discord words which have a stop word as a prefix.

(i). Clearly, one can effectively compute a positive integer constant C such that no word longer than C will be generated by our process.

(ii). Now we claim that if I has a solution y then I has also a solution z such that z is a prefix of one of the stop words produced by our algorithm.

This is proved as follows.

Assume that y is a solution of I such that y is not a prefix of one of the stop words produced by the algorithm. Then, clearly, no prefix w of y can satisfy either (a) or (b) and y must be of the form wu where w satisfies (c). Consequently w_1u must also be a solution of I. Iterating this

process we arrive at a solution z that is a prefix of one of the stop words produced by the algorithm.

The theorem follows now from (i) and (ii). □

Theorem 1.2. Given a positive integer k, it is decidable whether or not an arbitrary instance I of GPCP(2) has a solution w such that $\#_1 w \leq k$.

Proof.

Let $I = (h,g,a_1,a_2,b_1,b_2)$.

We consider separately the following two cases.

(a). $|h(0)| \neq |g(0)|$.

Note that if x is a word such that $\#_1 x \leq k$ then

$abs(|h(x)| - |g(x)|) \geq t(x)$, where

$t(x) = (|x| - k)\, abs\,(|h(0)|-|g(0)|) - k\, abs\,(|h(1)|-|g(1)|)$.

But if w is a solution of I such that $\#_1 w \leq k$ then we have

$abs(|h(w)| - |g(w)|) \leq abs(|a_1| - |b_1|) + abs(|a_2| - |b_2|)$

and consequently we get that

$$|w| \leq \frac{|a_1a_2b_1b_2| + k\, abs\,(|h(1)|-|g(1)|)}{abs(|h(0)|-|g(0)|)} + k$$

Hence in this case to find a solution of I it suffices to check all the words over {0,1} shorter than certain effectively computable positive integer.

(b). $|h(0)| = |g(0)|$.

Here we will consider two subcases.

(b.1). It is not true that $h(0) \sim g(0)$.

Now let for a word x

$t(x) = \max\{|h(x)|,|g(x)|\} - (k|h(1)| + k|g(1)| + (2k+1)3|h(0)|)$.

Assume that w is a solution of I such that $\#_1 w \leq k$ and let us assume that $t(w) > 0$. Let $\alpha = a_1 h(w) a_2 = b_1 g(w) b_2$.

We will divide all occurrences of letters in α into two categories. An occurrence of a letter in α belongs to the *first category* if either it is "contributed" to α through h from an occurrence of 1 in w, or it is contributed to α through g from an occurrence of 1 in w or it belongs to the prefix a_1 or it belongs to the prefix b_1 or it belongs to the suffix a_2 or it belongs to the suffix b_2.

Otherwise an occurrence of a letter in α belongs to the *second category*.

(i). α contains an occurrence of a subword that is longer than $3|h(0)|$ and consists only of occurrences of letters of the second category.

This is seen as follows.

Let a *1-group in* α be a maximal sequence of occurrences of letters of the first category (maximal meaning that it cannot be prolongated neither to the left nor to the right without including an occurrence of a letter of the second category). Similarily we define the notion of a *2-group in* α. Since $\#_1 w \leq k$, the number of 1-groups in α (not counting groups resulting from prefixes a_1, b_1 and suffixes a_2, b_2) cannot be bigger that 2k. Consequently, the number of 2-groups in α cannot exceed (2k+1). From the definition of t(w) it follows that the total combined length of all the 2-groups is longer than $(2k+1)\,3|h(0)|$ and consequently there exists a 2-group that is longer than $3|h(0)|$.

(ii). From (i) it follows that either α contains a subword of the form h(0)h(0)h(0) which in turn contains a subword of the form g(0)g(0), or α contains a subword of the form g(0)g(0)g(0) which in turn contains a subword of the form h(0)h(0). It is easily seen that either of these cases implies that $h(0) \sim g(0)$; a contradiction.

Consequently it must be that $t(w) \leq 0$, and this implies that

$$|w| \leq (k|h(1)| + k|g(1)| + (2k+1)3|h(0)|).$$

(b.2). $h(0) \sim g(0)$.

In this case we prove the theorem by induction on k.

Basis. $k = 0$.

Then I has a solution if and only if the equation

$$a_1\hat{h}(x)a_2 = b_1\hat{g}(x)b_2 \quad \ldots\ldots(1.1)$$

has a solution where $\hat{h}$ and $\hat{g}$ are the restrictions of h and g respectively to the alphabet {0}. However, (1.1) is a 1-fold equation and so the theorem follows now from Theorem 0.2.

Induction step. We will show that if the theorem is true for k then it is also true for (k+1).

Let $q = abs(|a_1|-|b_1|) + 3|h(0)|$.

(iii). If I has a solution, then I has a solution w such that $|h(0\text{-}pref(w))| < q$.

This is rather obvious, because if z is a solution of I such that $|h(0\text{-}pref\, z)| \geq q$ then by removing the first occurrence of 0 in z one gets another solution of I.

Now for every $\ell < q$, let $I_\ell = (h,g,a_1h(0^\ell 1),a_2,b_1g(0^\ell 1),b_2)$.

(iv). I has a solution w such that $\#_1 w = k + 1$ if and only if, for some $\ell < q$, I_ℓ has a solution z such that $\#_1 z = k$.

This follows easily from the construction of I_ℓ for $\ell < q$.

Thus, in the case of $h(0) \sim g(0)$, the theorem follows from the inductive assumption.

Since the division into cases (a) and (b) is exhaustive, the theorem holds. □

Theorem 1.3. Given $w \in \{0,1\}^*$, it is decidable whether or not an arbitrary instance I of GPCP(2) has a solution belonging to the set $\{w\}^+$.

Proof.

The problem is trivial if $w = \Lambda$.

Hence assume that $w \neq \Lambda$.

Let $I = (h,g,a_1,a_2,b_1,b_2)$ be an instance of GPCP(2) and let $\hat{h}$, $\hat{g}$ be homomorphisms from $\{0\}^*$ into $\{0,1\}^*$ defined by $\hat{h}(0) = h(w)$ and $\hat{g}(0) = g(w)$. Then, obviously, I has a solution belonging to $\{w\}^+$ if and only if the equation

$$a_1\hat{h}(x)a_2 = b_1\hat{g}(x)b_2 \qquad (1.2)$$

has a solution. However, (1.2) is a 1-fold equation and so the theorem follows from Theorem 0.2. □

2. PERIODIC INSTANCES OF GPCP(2)

In this section we study the case when (at least) one of the homomorphisms involved in an instance of GPCP(2) is periodic, that is $h(0)h(1) = h(1)h(0)$. Although this case is not difficult to settle it plays an important role in the reduction of the general problem (see [ ]). Also the material of this and the previous section should help the reader not very familiar with the Post Correspondence Problem to get the "feeling" for this problem.

Definition 2.1. A homomorphism f on $\{0,1\}^*$ is called *periodic* if $f(0)f(1) = f(1)f(0)$. An instance $I = (h,g,a_1,a_2,b_1,b_2)$ of GPCP(2) is called *periodic* if either h or g is periodic. □

Obviously, if f is periodic then there exist a word p and positive integers k_0, k_1 such that $f(0) = p^{k_0}$ and $f(1) = p^{k_1}$.

Theorem 2.1. It is decidable whether or not an arbitrary periodic instance of GPCP(2) has a solution.

Proof.

Let $I = (h,g,a_1,a_2,b_1,b_2)$ be a periodic instance of GPCP(2).

We will consider separately several cases.

(a). Both h and g are periodic.

Consider the equation

$$a_1h_1(x)h_2(y)a_2 = b_1g_1(x)g_2(y)b_2 \qquad (2.1)$$

where variables x and y range over $\{0\}^+$ and h_1, h_2, g_1, g_2 are homomorphisms of $\{0\}^*$ defined by $h_1(0) = h(0)$, $g_1(0) = g(0)$, $h_2(0) = h(1)$ and $g_2(0) = g(1)$.

(i). A pair $(x,y) = (0^n,0^m)$ is a solution of (2.1) if and only if $0^n 1^m$ is a solution of I.

This is proved as follows.

Assume that $(0^n,0^m)$ is a solution of (2.1).

Then we have

$$a_1h(0^n1^m)a_2 = a_1h(0^n)h(1^m)a_2 = a_1h_1(0^n)h_2(0^m)a_2 =$$
$$= b_1g_1(0^n)g_2(0^m)b_2 = b_1g(0^n)g(1^m)b_2 =$$
$$= b_1g(0^n1^m)b_2.$$

Similarily we show that if $0^n 1^m$ is a solution of I then $(0^n,0^m)$ is a solution of (2.1).

(ii). If w is a solution of I then also the word $0^n 1^m$ where $n = \#_0w$ and $m = \#_1w$ is a solution of I.

This follows from the obvious fact that a periodic homomorphism f is *permutational* in the sense that if u, z are words such that z is a permutation of u then f(u) = f(w).

Now the theorem follows from (i), (ii) and from Theorem 0.2.

(b). h is periodic and g is not periodic.

Let p be a word of the minimal length such that $h(0) = p^{k_0}$ and $h(1) = p^{k_1}$ for some positive integers k_0 and k_1. Since g is not periodic, it is easily seen that it must be elementary.

The form of h implies that solutions of the equation

$$a_1h(x)a_2 = b_1g(x)b_2 \quad \ldots\ldots(2.2)$$

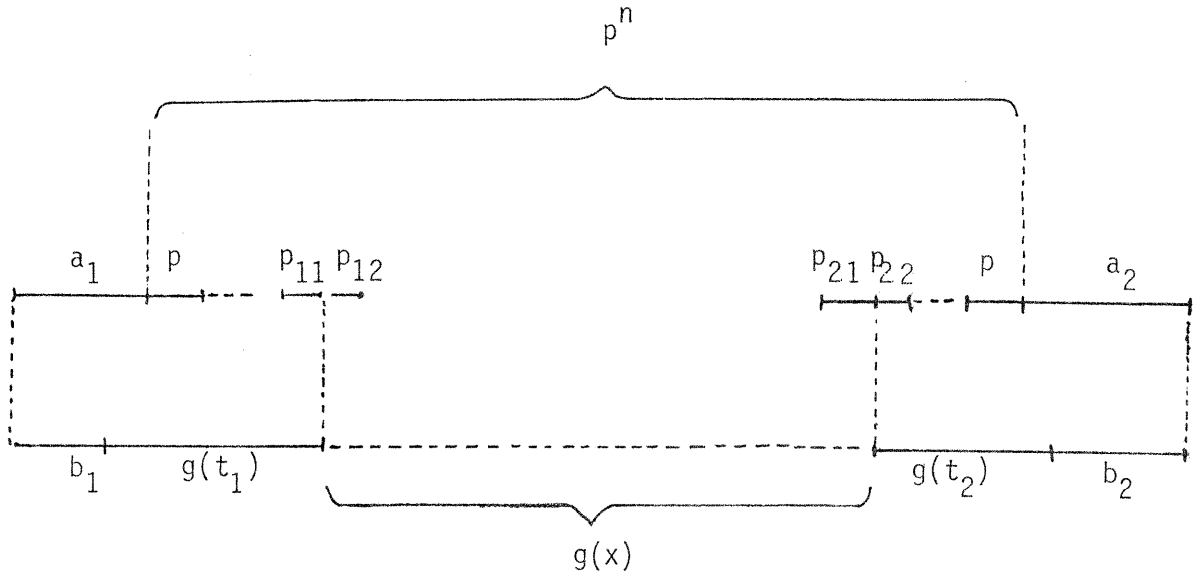
in variable x are included in solutions of the equation

$$a_1p^na_2 = b_1g(x)b_2 \quad \ldots\ldots(2.3)$$

in variables n and x where n ranges over positive integers.

Let t_1 be a word of the minimal length such that $a_1 \mathit{pref} b_1 g(t_1)$ and let t_2 be a word of the minimal length such that $a_2 \mathit{suf} g(t_2) b_2$. (If either t_1 or t_2 does not exist, then solutions of I must be shorter than certain effectively computable positive integer constant).

We have the following situation:



where $p = p_{11}p_{12} = p_{21}p_{22}$ for some words p_{11}, p_{12}, p_{21}, p_{22}.

Let $d = p_{12}p_{11}$ and let p_1 be the word such that $p_{12}p_{21} = dp_1$.

Consequently we are led to the equation

$d^m p_1 = g(x)$.....(2.4)

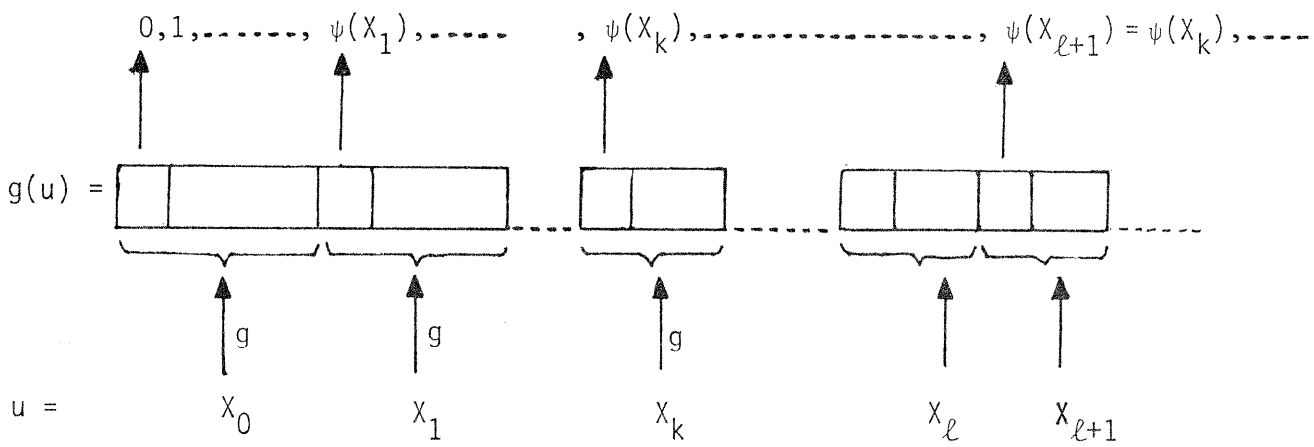
in variables m and x where m ranges over the positive integers.

Claim 2.1. There exists effectively a positive integer C_1 such that whenever $(m,x) = (q,u)$ is a solution of (2.4) such that $q > C_1$ then one can effectively find finite sets of words U_0, U_1 and W such that $u = u_1 y^i u_2$ for some $i \geq 1$, $u_1 \in U_1$, $u_2 \in U_2$ and $y \in W$.

Proof of Claim 2.1.

Let $|p| = r$. Let (q,u) be a solution of (2.4), so $d^q p_1 = g(u)$, and let $u = X_0X_1\ldots X_{|u|}$, where $X_0, X_1, \ldots, X_{|u|} \in \{0,1\}$. Let ψ be a function from $\{X_0, X_1, \ldots, X_{|u|}\}$ into $\{0,1,\ldots,r-1\}$ such that it assigns to each X_i the position of *first*$(g(X_i))$ modulo r in the string g(u).
Let k be the minimal positive integer such that the value of $\psi(X_k)$ repeats in the sequence $\psi(X_0), \ldots, \psi(X_{|u|})$ and let ℓ be the minimal integer, $\ell \geq k$, such that $\psi(X_{\ell+1}) = \psi(X_k)$. We assume that u is "large enough" so that such k and ℓ exist.

Thus we have the following situation


0,1,....., ψ(X1),.....
, ψ(Xk),.....
ψ(Xℓ+1) = ψ(Xk),....
g(u) =
g
g
g
u =
X0
X1
Xk
Xℓ
Xℓ+1

Let $t = |g(X_k \ldots X_\ell)|$ and let α be the subword of $g(u)$ starting at $first(g(X_{\ell+1}))$ and such that $|\alpha| = t$. Let α'' be the remaining suffix of $g(u)$.

Let $y = X_k \ldots X_\ell$ and let y' be the minimal subword of u such that it starts at $X_{\ell+1}$ and such that $g(y')$ contains α as its prefix. Let y'' be the remaining suffix of u.

First we prove the following.

(i). $y' = y$.

This is proved by contradiction as follows.

Assume that $y' \neq y$. Since $|\alpha| = |g(y)|$ and we assume that u is large enough (so that we are still "within" the periodic part d^q), we have $\alpha = g(y)$ and consequently

$$\alpha\, \alpha'' \; pref \; \alpha\, \alpha\, \alpha'' \quad \ldots\ldots (2.5)$$

and

$$g(y)\, pref\, g(y') \quad \ldots\ldots (2.6)$$

Now we consider two possibilities.

(i.1). $y\, pref\, y'$.

Then from the "minimality condition" on y' it follows that $y = y'$; a contradiction.

(i.2). $y'\, pref\, y$.

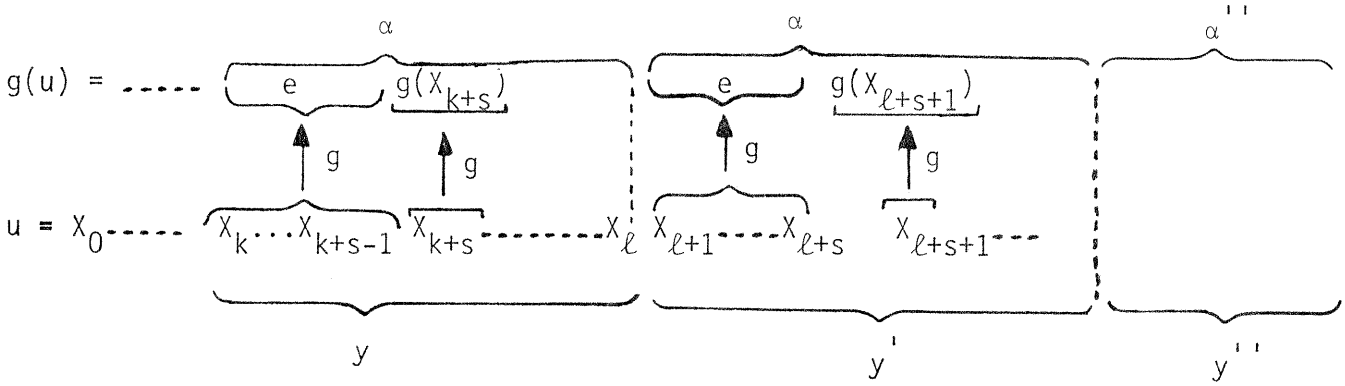
Then (2.6) implies that $g(y') = g(y)$ which (for the propagating homomorphism g) is possible only if $y' = y$; a contradiction.

Consequently it is not true that $y\, PREF\, y'$.

Thus there exists a $s \leq \ell$ such that $X_{k+s} \neq X_{\ell+1+s}$. Since g is not periodic, we have $g(X_{k+s}) \neq g(X_{\ell+1+s})$.

Let $e = g(X_k X_{k+1} \ldots X_{k+s-1}) = g(X_{\ell+1} X_{\ell+2} \ldots X_{\ell+s})$.

Then we have the following situation:


α
α
α''
g(u) =
e
g(X_{k+s})
e
g(X_{ℓ+s+1})
g
g
g
g
u = X_0
X_k ... X_{k+s-1}
X_{k+s}
X_ℓ
X_{ℓ+1}
X_{ℓ+s}
X_{ℓ+s+1}
y
y'
y''

Thus (2.5) implies that $(e\backslash(\alpha\alpha''))\ pref\ (e\backslash(\alpha\alpha\alpha''))$ which for m "big enough" contradicts the basic property of elementary homomorphisms, see Theorem 0.1.

Consequently it must be that $y' = y$ and so (i) holds.

Thus indeed u is of the form $u_1 y^i u_2$ for some $i \geq 0$ where $u_1 = X_0 \ldots X_{k-1}$ and u_2 is the appropriate suffix (which we obtain when we cannot "fit" anymore y into u).

The claim follows now easily. □

Now we check whether there exists a solution (q,u) of (2.4) such that $q \leq C_1$.

If such a solution (q,u) exists we check whether the equation $a_1 p^n a_2 = b_1 g(t_1) g(u) g(t_2) b_2$ in variable n ranging over positive integers has a solution.

If such a solution $n = n_0$ exists then we check whether n_0 can be expressed in the form $n_0 = k_0 \#_0 (t_1 \ u \ t_2) + k_1 \#_1 (t_1 \ u \ t_2)$. If it does than $x = t_1 \ u \ t_2$ is a solution of I.

Otherwise, that is if one of the above three steps provides a negative answer, we proceed as follows.

If each solution (q,u) of (2.4) is such that $q > C_1$ then consider again equation (2.3).

The claim implies that we can effectively find a finite number of words of the form $u_1\ y\ u_2$ such that x must be of the form $t_1\ u_1\ y^i\ u_2\ t_2$, $i \geq 1$, for one of them. We consider one by one each of those possibilities.

If $x = t_1\ u_1\ y^i\ u_2\ t_2$ is a solution of (2.3), then we have

$$a_1 h(t_1\ u_1) h(y^i) h(u_2\ t_2) a_2 = b_1\ g(t_1\ u_1) g(y^i) g(u_2\ t_2) b_2 \quad \ldots\ldots\ldots\ldots\ldots\ldots (2.7)$$

We set now $a_1' = a_1\, h(t_1\, u_1)$, $a_2' = h(u_2\, t_2)a_2$, $b_1' = b_1\, g(t_1\, u_1)$, $b_2' = g(u_2\, t_2)b_2$ and homomorphisms h', g' on $\{0\}^*$ be defined by $h'(0) = h(y)$ and $g'(0) = g(y)$.

Then clearly (2.7) has a solution if and only if the following 1-fold equation has a solution

$$a_1' h'(0^i)a_2' = b_1' g'(0^i)b_2' \quad \text{.....(2.8)}$$

Hence by Theorem 0.2 it is decidable whether or not (2.7) has a solution (in i). If it has, say $i = i_0$, then clearly I has a solution, namely $x = t_1\, u_1\, y^{i_0}\, u_2\, t_2$.

On the other hand, if, for none of the possibilities $u_1\, y\, u_2$, (2.7) has a solution then I does not have a solution.

Consequently it is decidable whether or not I has a solution and the theorem holds in case (b).

(c). g is periodic and h is not periodic.

This case is analogous to (b).

However (a), (b) and (c) together imply the theorem. □

3. THE BASIC TRANSFORMATION

In this section we will be concerned with *marked* homomorphisms, that is homomorphisms such that their images of 0 and of 1 differ on the first letter. For these homomorphisms and for instances of GPCP(2) involving such homomorphisms we present the transformation, called the *equality collector*, that will be the most crucial in our solution of the GPCP(2) problem. The basic properties of this transformation are studied in this section.

Definition 3.1. A homomorphism f of $\{0,1\}^*$ is *marked* if $first(f(0)) \neq first(f(1))$. □

Definition 3.2. If $f : \{0,1\}^* \to \{0,1\}^*$ is a marked homomorphism, $i, j \in \{0,1\}$ and $first(f(i)) = j$ then we say that i is the j-*index of* f and we write $i = j\text{-}ind(f)$. □

Definition 3.3. For a pair of marked homomorphisms h, g of $\{0,1\}^*$ we define the function $\mu_{h,g}$ from $\{0,1\}$ into $\{0,1\}$ by: for $i, j \in \{0,1\}$, $\mu_{h,g}(i) = j$ if and only if $first(h(i)) = first(g(j))$ and $first(g(i)) = first(h(j))$. We write μ rather than $\mu_{h,g}$ whenever h,g are understood. □

It is easily seen that $\mu_{h,g}$ is a well defined function for marked homomorphisms h, g.

Suppose that we are given a pair of words (α,β) such that one is a (proper) prefix of another one, say α is a proper prefix of β. Assume also that we are given two marked homomorphisms h and g and we want to find two words u and w such that $\alpha\, h(u) = \beta\, g(w)$. A very

natural way to proceed is as follows. Since β is "ahead" we look at the first letter of the "difference" between β and α, say c (thus αc is a prefix of β). Since h is marked, c uniquely defines a letter from $\{0,1\}$, say i_1, such that we must consider $\alpha\, h(i_1)$ if we want to "catch up" with β. We iterate this procedure until we get a word $i_1 \ldots i_{r_1}$ such that either $\alpha\, h(i_1 \ldots i_{r_1}) = \beta$ or β is a proper prefix $\alpha\, h(i_1 \ldots i_{r_1})$. If $\alpha\, h(i_1 \ldots i_{r_1}) = \beta$ we are done; otherwise we do now the same on the "β side," that is we construct a sequence of letters $j_1, j_2, \ldots, j_{s_1}$ such that either $\alpha\, h(i_1 \ldots i_{r_1}) = \beta\, g(j_1 \ldots j_{s_1})$ or $\alpha\, h(i_1 \ldots i_{r_1})$ is a proper prefix of $\beta\, g(j_1 \ldots j_{s_1})$. We iterate this procedure again, starting now on the "α side". Either this process ends successfully (that is we find sequences $i_1 \ldots i_r$ and $j_1 \ldots j_s$ such that $\alpha\, h(i_1 \ldots i_r) = \beta\, g(j_1 \ldots j_s)$), or it will continue "infinitely long" or it blocks (that is we perform a step the consequence of which is that the word on the α side and the word on β side are not anymore related by the prefix relation).

Essentially this situation is formalized now.

Definition 3.4. Let h, g be marked homomorphisms from $\{0,1\}^*$ into $\{0,1\}^*$ and let $\alpha, \beta \in \{0,1\}^*$. For a nonnegative integer i we define $(\alpha,\beta)^{(i)}_{h,g}$ inductively as follows.

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

The h-*projection* of $(\alpha,\beta)^{(0)}_{h,g}$, denoted by $((\alpha,\beta)^{(0)}_{h,g})_h$ or simply by $(\alpha,\beta)^{(0)}_{h}$ whenever g is understood, is defined by $(\alpha,\beta)^{(0)}_{h} = \Lambda$. The g-*projection* of $(\alpha,\beta)^{(0)}_{h,g}$, denoted by $((\alpha,\beta)^{(0)}_{h,g})_g$ or simply by $(\alpha,\beta)^{(0)}_{g}$ whenever h is understood, is defined by $(\alpha,\beta)^{(0)}_{g} = \Lambda$.

i + 1: $(\alpha,\beta)_{h,g}^{(i+1)}$ is defined if and only if

$\alpha\, h((\alpha,\beta)_h^{(i)})$ PREF $\beta\, g((\alpha,\beta)_g^{(i)})$ and $\alpha\, h((\alpha,\beta)_h^{(i)}) \neq \beta\, g((\alpha,\beta)_g^{(i)})$.

If $(\alpha,\beta)_{h,g}^{(i+1)}$ is defined and $c \in \{0,1\}$ then

(a). if $\alpha\, h((\alpha,\beta)_h^{(i)})c$ *pref* $\beta\, g((\alpha,\beta)_g^{(i)})$ then $(\alpha,\beta)_{h,g}^{(i+1)} = (\alpha,\beta)_{h,g}^{(i)}(h,c\text{-}ind(h))$,

and

(b). if $\beta\, g((\alpha,\beta)_g^{(i)})c$ *pref* $\alpha\, h((\alpha,\beta)_h^{(i)})$ then $(\alpha,\beta)_{h,g}^{(i+1)} = (\alpha,\beta)_{h,g}^{(i)}(g,c\text{-}ind(g))$.

If $(\alpha,\beta)_{h,g}^{(i+1)}$ is defined then the h-*projection* of it and the g-*projection* of it are defined by:

if (a) holds then $(\alpha,\beta)_h^{(i+1)} = (\alpha,\beta)_h^{(i)}c\text{-}ind(h)$ and $(\alpha,\beta)_g^{(i+1)} = (\alpha,\beta)_g^{(i)}$,

and

if (b) holds then $(\alpha,\beta)_h^{(i+1)} = (\alpha,\beta)_h^{(i)}$ and $(\alpha,\beta)_g^{(i+1)} = (\alpha,\beta)_g^{(i+1)}c\text{-}ind(g)$.

For $i \geq 0$ we say that

$(\alpha,\beta)^{(i)}$ is *successful* if $\alpha\, h((\alpha,\beta)_h^{(i)}) = \beta\, g((\alpha,\beta)_g^{(i)})$, and

$(\alpha,\beta)^{(i)}$ *blocks* if it is not true that $\alpha\, h((\alpha,\beta)_h^{(i)})$PREF $\beta\, g((\alpha,\beta)_g^{(i)})$. □

Definition 3.5. Let h, g be marked homomorphisms from $\{0,1\}^*$ into $\{0,1\}^*$ and let $\alpha, \beta \in \{0,1\}^*$. The (α,β)-*sequence* (*with respect to* h,g), denoted by $(\alpha,\beta)_{h,g}$, is defined as follows.

(a). Assume that $i \geq 0$ is such that $(\alpha,\beta)_{h,g}^{(i)}$ is successful (note that i is unique). Then $(\alpha,\beta)_{h,g} = (\alpha,\beta)_{h,g}^{(i)}$ and we say that $(\alpha,\beta)_{h,g}$ is *successful*.

(b). Assume that $i \geq 0$ is such that $(\alpha,\beta)_{h,g}^{(i)}$ blocks (note that i is unique). Then $(\alpha,\beta)_{h,g} = (\alpha,\beta)_{h,g}^{(i)}$ and we say that $(\alpha,\beta)_{h,g}$ *blocks*.

(c). If there is no i satisfying either (a) or (b) then $(\alpha,\beta)_{h,g}$ is the infinite (to the right) word over the alphabet

$\{(h,\Lambda), (h,0), (h,1), (g,\Lambda), (g,0), (g,1)\}$ such that for each $i \geq 0$, $(\alpha,\beta)^{(i)}_{h,g}$ is its prefix.

The h-*projection* of $(\alpha,\beta)_{h,g}$, denoted by $((\alpha,\beta)_{h,g})_h$ or simply by $(\alpha,\beta)_h$ whenever g is understood, is defined by:
if (a) holds then $(\alpha,\beta)_h = (\alpha,\beta)^{(i)}_h$,
if (b) holds then $(\alpha,\beta)_h = (\alpha,\beta)^{(i)}_h$, and
if (c) holds then $(\alpha,\beta)_h$ is the infinite (to the right) word over $\{0,1\}$ such that for each $i \geq 0$, $(\alpha,\beta)^{(i)}_h$ is its prefix.

The g-*projection* of $(\alpha,\beta)_{h,g}$, denoted by $((\alpha,\beta)_{h,g})_g$ or simply by $(\alpha,\beta)_g$ whenever h is understood, is defined by:
if (a) holds then $(\alpha,\beta)_g = (\alpha,\beta)^{(i)}_g$,
if (b) holds then $(\alpha,\beta)_g = (\alpha,\beta)^{(i)}_g$, and
if (c) holds then $(\alpha,\beta)_g$ is the infinite (to the right) word over $\{0,1\}$ such that for each $i \geq 0$, $(\alpha,\beta)^{(i)}_g$ is its prefix. □

Here are two rudimentary properties of (α,β) - sequences.

Lemma 3.1. Let h, g be marked homomorphisms from $\{0,1\}^*$ into $\{0,1\}^*$ and let $\alpha, \beta \in \{0,1\}^*$. If $(\alpha,\beta)_{h,g}$ is infinite then it is ultimately periodic.

Proof.

Obvious. □

Lemma 3.2. Let h, g be marked homomorphisms from $\{0,1\}^*$ into $\{0,1\}^*$. It is decidable whether or not $(\alpha,\beta)_{h,g}$ is successful or $(\alpha,\beta)_{h,g}$ blocks or $(\alpha,\beta)_{h,g}$ is infinite for arbitrary words $\alpha, \beta \in \{0,1\}^*$.

Proof.

Obvious. □

Now we define the construction that is perhaps the most important construction of our solution of GPCP(2).

Definition 3.6. Let (h,g) be an ordered pair of marked homomorphisms such that both the sequence $(h(0), g(\mu(0)))_{n,g}$ and the sequence $(h(1), g(\mu(1)))_{h,g}$ are successful. Then the *equality collector of* (h,g), denoted as *ecol* (h,g), is the pair of homomorphisms $(\overline{h},\overline{g})$ on $\{0,1\}^*$ defined by

$\overline{h}(0) = 0(h(0),g(\mu(0)))_h$, $\overline{h}(1) = 1(h(1), g(\mu(1)))_h$,

$\overline{g}(0) = \mu(0)(h(0),g(\mu(0)))_g$ and $\overline{g}(1) = \mu(1)(h(1),g(\mu(1)))_g$. □

Remark. Throughout the paper, given a pair of homomorphisms h, g we will use the "bar notation" $(\overline{h},\overline{g})$ to denote *ecol* (h,g). □

The usefulness of the (iterative) application of the *ecol* function relies on the following property. First we need a definition.

Definition 3.7. For a pair of homomorphisms (h,g) of $\{0,1\}^*$ we define its *suffix index* , denoted by σ(h,g), as

$\sigma(h,g) = \#\, suf\{h(0),h(1)\} + \#suf\{g(0),g(1)\}$. □

Lemma 3.3. Let (h,g) be an ordered pair of marked homomorphisms such that $ecol(h,g) = (\overline{h},\overline{g})$ exists. Then

(a). both $\overline{h}$ and $\overline{g}$ are marked,

(b). $\sigma(h,g) \geq \sigma(\overline{h},\overline{g})$.

Proof.

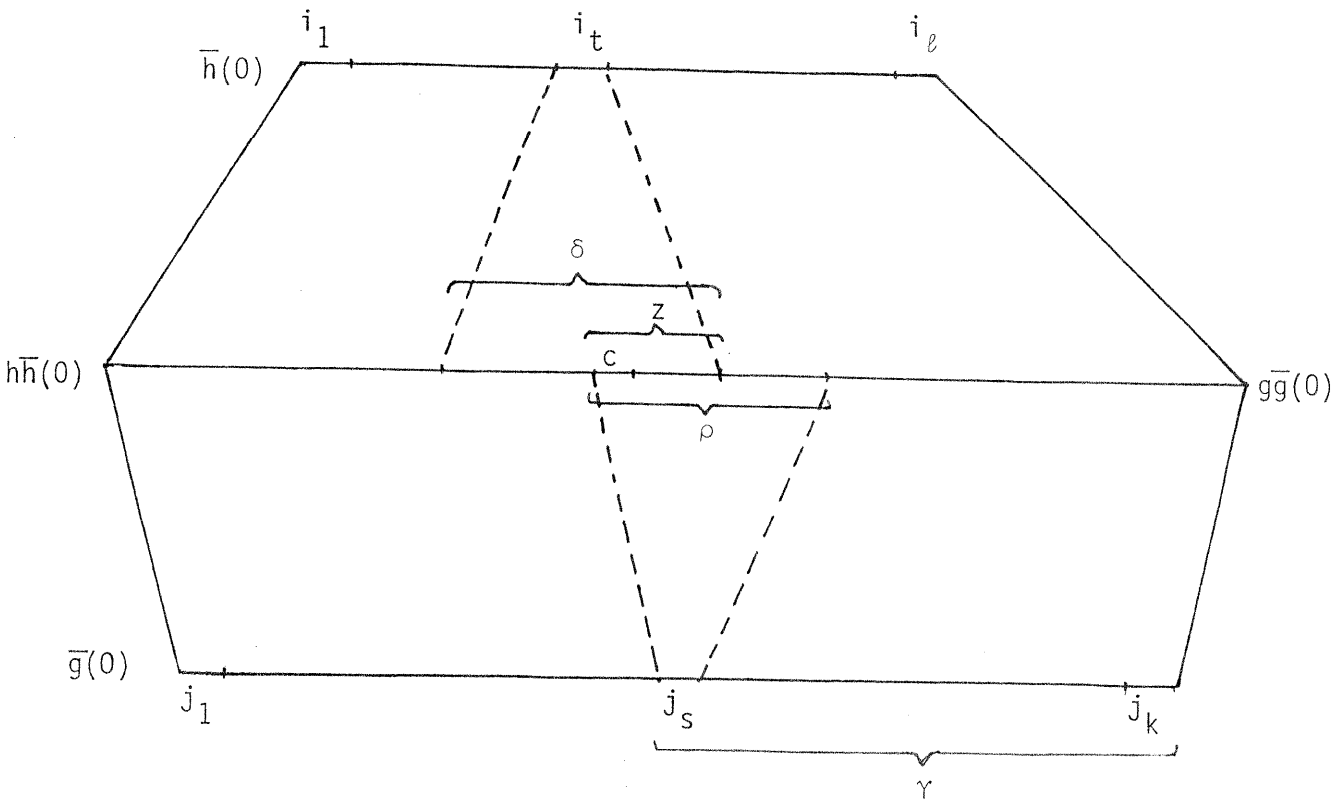
(a) is obvious.

To prove (b) we proceed as follows.

(b.1). We will demonstrate that there exists a function (perhaps partial) from $suf\{h(o), h(1)\}$ onto $suf\{\overline{g}(0), \overline{g}(1)\}$.

(b.1.1.). We construct a function f_1 from $suf\overline{g}(0)$ into $suf\{h(0), g(0)\}$. Assume that γ is a nonempty suffix of $\overline{g}(0)$. Let $\overline{g}(0) = j_1 \ldots j_k$, $k \geq 1$, $j_1, \ldots, j_k \in \{0,1\}$ and let $\gamma = j_s \ldots j_k$ for some $s \in \{1, \ldots, k\}$. Let $\overline{h}(0) = i_1 \ldots i_\ell$, $\ell \geq 1$, $i_1, \ldots, i_\ell \in \{0,1\}$. Consider now the contribution ρ from j_s through g to $g\overline{g}(0) = h\overline{h}(0)$. Let c be the leftmost (occurrence of a) letter in ρ and let i_t be the element of $\overline{h}(0)$ such that its contribution δ through h to $h\overline{h}(0) = g\overline{g}(0)$ includes c. Let z be the suffix of δ starting at c. Clearly z is well defined. We let z to be the value of our function applied to γ.

Thus we have the following situation:



Clearly, ρ PREF z ; since it is difficult to depict both possibilities in one figure we illustrate the possibility of z *pref* ρ only.

(b.1.2). Similarly we construct a function f_2 from $suf\overline{g}(1)$ into $suf\{h(0), h(1)\}$.

Combining inverses of functions f_1 and f_2 we get a binary relation f which is a subset of the Cartesian product $suf\{h(0), h(1)\} \times suf\{\overline{g}(0), \overline{g}(1)\}$. Note, hovewever, that (in the notation from above) z *uniquely* determines γ and so f must be a function (perhaps partial) which is onto.

(b.2). Analogously to (b.1) we show that there exists a function (perhaps partial) from $suf\{g(0), g(1)\}$ *onto* $suf\{\overline{h}(0), \overline{h}(1)\}$.

The part (b) of the lemma follows directly from (b.1) and (b.2). □

We will consider now instances of GPCP(2) involving marked homomorphisms only.

Definition 3.8. An instance $I = (h,g,a_1.a_2.b_1.b_2)$ of GPCP(2) is called *marked* if both h and g are marked. □

Remark. In the rest of this paper, even if not explicitly stated, we deal with marked instances of GPCP(2) *only.*

Definition 3.9. Let $I = (h,g,a_1,a_2,b_1,b_2)$ be a marked instance of GPCP(2) such that both the sequence $(h(0),g(\mu(0)))_{h,g}$ and the sequence $(h(1),g(\mu(1)))_{h,g}$ are succesful. The *tail equation of* I, denoted as $E_{Tail(1)}$, is the equation

$h(x)a_2 = g(y)b_2$

in variables x, y ranging over $\{0,1\}^*$. A pair of words (u,w) is called a *short solution* of $E_{Tail(I)}$ if $h(u)a_2 = g(w)b_2$ and moreover, $|h(u)a_2| \leq |a_2b_2| + |h\bar{h}(0)| + |h\bar{h}(1)|$. The set of all short solutions of $E_{Tail(I)}$ is denoted by $sol(E_{Tail(I)})$. □

The notion of the equality collector is extended now to instances of GPCP(2) as follows.

Definition 3.10. Let $I = (h,g,a_1,a_2,b_1,b_2)$ be a marked instance of GPCP(2) such that the sequence $(a_1,b_1)_{h,g}$ is successful, the sequence $(h(0),g(\mu(0)))_{h,g}$ is successful, the sequence $(h(1),g(\mu(1)))_{h,g}$ is successful and $sol(E_{Tail(I)}) \neq \emptyset$. Then an *equality collector of* I, denoted *ecol* I, is an instance $J = (\bar{h},\bar{g},\bar{a}_1,u,\bar{b}_1,w)$ of GPCP(2) such that $(\bar{h},\bar{g}) = ecol(h,g)$, $\bar{a}_1 = (a_1,b_1)_h$, $b_1 = (a_1,b_1)_g$ and $(u,w) \in sol(E_{Tail(I)})$. The set of all equality collectors of I is denoted by ECOL(I). □

Clearly we can iterate *ecol* and ECOL constructions; we will use $ecol^i$ and $ECOL^i$ (for a positive integer i) to denote the i'th iteration of *ecol* or ECOL function respectively. Note that, in general, $ecol^i$ or $ECOL^i$ may be undefined.

Definition 3.11. If I is a marked instance of GPCP(2) such that $ECOL(I) \neq \emptyset$ then we say that I is *successful*; otherwise we say that I is *unsuccessful*. □

The following result "justifies" the use of ECOL transformation as a tool in solving the GPCP(2).

Theorem 3.1. Let I be a marked instance of GPCP(2) such that ECOL(I) is not empty. One can effectively compute a positive integer constant C such that: I has a solution if and only if either I has

a solution not longer than C or there exists a J in ECOL(I) such that J has a solution.

Proof.

Let $I = (h,g,a_1,a_2,b_1,b_2)$ and let

$C = |h(\overline{a}_1)g(\overline{b}_1)| + |a_2a_2b_1b_2| + 2(|h(\overline{h}(0))| + |h(\overline{h}(1))|)$.

Assume that for two words $\alpha, \beta \in \{0,1\}^+$ we have

$a_1h(\alpha)a_2 = b_1g(\beta)b_2 = \gamma$

where $|\gamma| > C$.

Then clearly γ is of the form

$\gamma =$

$$\left[\begin{array}{c} a_1h(\overline{a}_1) \\ b_1g(\overline{b}_1) \end{array}\right] \left[\begin{array}{c} h(i_{11}i_{12}\cdots i_{1r_1}) \\ g(j_{11}j_{12}\cdots j_{1s_1}) \end{array}\right] \cdots \left[\begin{array}{c} h(i_{t1}\cdots i_{tr_t}) \\ g(j_{t1}\cdots j_{ts_t}) \end{array}\right] \left[\begin{array}{c} h(u)a_2 \\ g(w)b_2 \end{array}\right]$$

for some $t \geq 1$ where $(u,w) \in sol(E_{Tail(I)})$.

Let $\sigma = i_{11}i_{21}\cdots i_{t1}$.

Then clearly we have

$\alpha = \overline{a}_1\overline{h}(\sigma)u$ and $\beta = \overline{b}_1\overline{g}(\sigma)w$.....(3.1)

(a). Assume that $\alpha = \beta$ is a solution of I such that

$|a_1h(\alpha)a_2| = |b_1g(\alpha)b_2| > C$.

Then from (3.1) it follows that σ is a solution of the instance $J = (\overline{h},\overline{g},\overline{a}_1,u,\overline{b}_1,w) \in ECOL(I)$.

(b). Assume now that $J = (\overline{h},\overline{g},\overline{a}_1,u,\overline{b}_1,w) \in ECOL(I)$ has a solution δ; hence

$\overline{a}_1\overline{h}(\delta)u = \overline{b}_1\overline{g}(\delta)w$.....(3.2).

However we have

$$a_1h(\overline{a}_1\overline{h}(\delta)u)a_2 = a_1h(\overline{a}_1)h\overline{h}(\delta)h(u)a_2$$

$$= b_1g(\overline{b}_1)g\overline{g}(\delta)g(w)b_2$$

$$= b_1g(\overline{b}_1\overline{g}(\delta)w)b_2 \ .$$

Thus (3.2) implies that $\tau = \overline{a}_1\overline{h}(\delta)u = \overline{b}_1\overline{g}(\delta)w$ is a solution of I.

The theorem follows now from (a) and (b). □

We will consider now the case when the *ecol* transformation can be applied "ad infinitum."

Lemma 3.4. Let (h,g) be an ordered pair of marked homomorphisms such that for each $i \geq 1$, $ecol^i(h,g)$ is defined. Then the sequence (h,g), $ecol(h,g)$, $ecol^2(h,g)$, ... is ultimately periodic and it can be effectively constructed.

Proof.

This is a straightforward consequence of Lemma 3.3. □

We will use the notation *trace* (h,g) to denote the sequence (h,g), $ecol(h,g)$, $ecol^2(h,g)$, ... and if this sequence is infinite then we use *thres* (h,g) to denote the length of its threshold part and *per* (h,g) to denote the length of its period part.

In our solution of the GPCP(2) (in the case of marked instances) we will iteratively apply the ECOL transformation until we reach the "stable situation" which is formally defined as follows.

Definition 3.12. Let $I = (h,g,a_1,a_2,b_1,b_2)$ be an instance of marked GPCP(2) such that *trace* (h,g) is infinite and let *thres* (h,g) = r. Then $ecol^{r+1}(h,g)$ is called *stable*.

We say that $J = (\hat{h},\hat{g},\hat{a}_1,\hat{a}_2,\hat{b}_1,\hat{b}_2)$ is a *stable version* of I whenever $J \in ECOL^{r+1}(I)$; the set of all stable versions of I is denoted by STABLE(I). We also say then that J is a *stable instance* of GPCP(2) (with respect to I). □

The next three results describe some basic properties of stable instances of GPCP(2).

Lemma 3.5.

(a). Let (h,g) be a pair of marked homomorphisms such that *trace* (h,g) is infinite. Let *thres* (h,g) = r. Then, for $i \geq r + 1$, $\sigma(ecol^i(h,g)) = \sigma(ecol^{i+1}(h,g))$.

(b). Let $I = (h,g,a_1,a_2,b_1,b_2)$ be a stable instance of GPCP(2). Then $\sigma(h,g) = \sigma(ecol(h,g))$.

Proof.

(a). Since *trace* (h,g) is ultimately periodic, for $i \geq r + 1$, $ecol^i(h,g) = ecol^{i+p}(h,g)$ where $p = per(h,g)$.

Thus Lemma 3.3 implies that for $i \geq r + 1$,

$\sigma(ecol^i(h,g)) = \sigma(ecol^{i+1}(h,g))$.

(b). This follows directly from (a). □

Definition 3.13. Two languages K_1 and K_2 are said to be *prefix compatible*, denoted K_1 *pcom* K_2, if for every $x \in K_1$ there exists a $y \in K_2$ such that x PREF y and for every $x \in K_2$ there exists a $y \in K_1$ such that x PREF y. □

Lemma 3.6.

(a). Let $(\hat{h},\hat{g})$ be a pair of marked homomorphisms such that *trace* $(\hat{h},\hat{g})$ is infinite. Let *thres* $(\hat{h},\hat{g}) = r$. Let $i \geq r + 1$ and let *ecol*^{i} $(h,g) = (h,g)$. Then

$\{g(0), g(1)\}$ *pcom*$(suf(\{h(0), h(1)\}))$ and

$\{h(0), h(1)\}$ *pcom*$(suf(\{g(0), g(1)\}))$.

(b). Let $I = (h,g,a_1,a_2,b_1,b_2)$ be a stable instance of GPCP(2). Then

$\{g(0), g(1)\}$ *pcom*$(suf(\{h(0), h(1)\}))$ and

$\{h(0), h(1)\}$ *pcom*$(suf(\{g(0), g(1)\}))$.

Proof.

(a). We refer the reader to the proof of Lemma 3.3 (especially to the (b.1) part of it where the basic construction is described in detail); we will use the notation from there. By Lemma 3.5 it follows that the function f is a bijection from $suf\{h(0), h(1)\}$ onto $suf\{\bar{g}(0), \bar{g}(1)\}$.
Note that the basic construction from the proof of Lemma 3.3 assigns to each element x from $\{g(0), g(1)\}$ a nonempty subset of the set $suf\{h(0), h(1)\}$ such that x PREF z for each z in this subset. Since f is total we also know that, by the same construction, for every element from the set $suf\{h(0), h(1)\}$ there exists an element x in the set $\{g(0), g(1)\}$ such that z PREF x. Consequently $\{g(0), g(1)\}$ *pcom*$(suf\{h(0), h(1)\}$.

Analogously we show that $\{h(0), h(1)\}$ *pcom*$(suf\{g(0), g(1)\})$.

Thus (a) holds.

(b). This is a simple corollary of (a). □

Lemma 3.7.

(a). Let (h,g) be a pair of marked homomorphisms such that *ecol* (h,g)

exists. Then it cannot be that $last(h(0)) = last(h(1)) \neq last(g(0)) = last(g(1))$.

(b). Let $I = (h,g,a_1,a_2,b_1,b_2)$ be a stable instance of GPCP(2). Then it cannot be that $last(h(0) = last(h(1)) \neq last(g(0)) = last(g(1))$.

Proof.

(a). We prove it by contradiction.

If $last(h(0)) = last(h(1)) \neq last(g(0)) = last(g(1))$ then, obviously, neither the sequence $(h(0), g(\mu(0)))_{h,g}$ is successful nor the sequence $(h(1),g(\mu(1)))_{h,g}$ is successful. Consequently, $ecol(h,g)$ is not defined; a contradiction.

(b). This is a simple corollary of (a). □

We end this section with the following result on the effectiveness of the STABLE(I) operation.

Lemma 3.8. Given an arbitrary marked instance I of GPCP(2) one can effectively decide whether or not STABLE(I) = $\emptyset$; moreover, if STABLE(I) $\neq \emptyset$ it can be effectively constructed.

Proof.

It follows essentially from Lemma 3.2 and Lemma 3.4. □

4. MORE SPECIAL CASES

In this section we solve more special cases of GPCP(2). The cases considered in this section (or their solutions) are connected to the ECOL transformation.

Theorem 4.1. It is decidable whether or not an arbitrary unsuccessful instance of GPCP(2) has a solution.

Proof.

This proof consists of considering quite a number of cases. Although in all the cases the intuitive idea of the proof is rather clear the formal proofs become quite tedious. For this reason we try to explain the intuitive idea of the proof in several typical ("crucial") cases, leaving the rest of the proof to the reader. We hope that after reading our outline the interested reader can construct (if necessary) the formal proof of the theorem.

Let $I = (h,g,a_1,a_2,b_1,b_2)$ be an unsuccessful instance of GPCP(2).

(a). If $sol(E_{Tail(I)}) = \emptyset$ then, clearly, if I has a solution then it is not longer than certain effectively computable positive integer constant. Hence in this case one can decide whether or not I has a solution.

(b). Assume that $(a_1,b_1)_{h,g}$ is not successful.

We consider separately two cases.

(b.1). $(a_1,b_1)_{h,g}$ blocks.

Then clearly a solution of I cannot be longer than an effectively computable constant and so one can effectively decide whether or not I has a solution.

(b.2). $(a_1,b_1)_{h,g}$ is infinite.

Thus by Lemma 3.1 $(a_1,b_1)_{h,g}$ is ultimately periodic; let α be the threshold part of $(a_1,b_1)_{h,g}$ and let β be the period part of $(a_1,b_1)_{h,g}$. Then let α_h be the h-projection of α (that is the sequence of second components from the subsequence of α consisting of letters whose first component is h), β_h be the h-projection of β, α_g be the g-projection of α and β_g be the g-projection of β.

Let Z be the set of all ordered pairs (u,w) of words over {0,1} such that either $|u| < |\beta_h|$ or $|w| < |\beta_g|$ and $a_1 h(\alpha_h)\, h(\beta_h)\, h(u)\, a_2 = b_1 g(\alpha_g) g(\beta_g) h(w)\, b_2$. If Z is empty then a solution of I cannot be longer than $2|\alpha\beta|$; so one can effectively decide whether or not I has a solution when $Z = \emptyset$.

Hence assume that $Z \neq \emptyset$; clearly Z can be effectively constructed.

Let h_1, g_1 be homomorphisms of $\{0\}^*$ defined by $h_1(0) = \beta_h$ and $g_1(0) = \beta_g$. For every $(u,w) \in Z$ consider the equation

$$\alpha_h h_1(x)\, u = \alpha_g g_1(x)\, w \qquad (4.1)$$

Assume that I has a solution γ longer than $2|\alpha\beta|$. Clearly then there exists a $m \geq 1$ such that for some $(u,w) \in Z$ we have

$$\gamma = \alpha_h \beta_h^m u = \alpha_g \beta_g^m w.$$

Then clearly $x = 0^m$ is a solution of (4.1) for the proper choice of (u, w).

On the other hand if $x = 0^m$, $m \geq 1$, is a solution of one of (the finite number of) equations of type (4.1), then $\alpha_h \beta_h^m u = \alpha_g \beta_g^m w$ is a solution of I.

Since (4.1) is a 1-fold equation, Theorem 0.2 implies that one can effectively decide whether or not I has a solution longer than $2|\alpha\beta|$.

This ends the proof of case (b.2).

(c). Assume that $(a_1,b_1)_{h,g}$ is successful.

We consider separately several cases.

(c.1). Both $(h(0), g(\mu(0)))_{h,g}$ and $(h(1), g(\mu(1)))_{h,g}$ block.

Clearly in this case one can effectively compute a positive integer constant C such that if I has a solution then it is shorter than C.

(c.2). $(h(0), g(\mu(0)))_{h,g}$ is infinite and $(h(1), g(\mu(1)))$ blocks.

This case can be solved analogously to the case (b.2). Now a "long enough" solution of I will look as follows:

first one successfully completes the (a_1,b_1) sequence,

then one runs the threshold part of $(h(0), g(\mu(0)))_{h,g}$,

then one runs (a number of times) the period part of $(h(0), g(\mu(0)))_{h,g}$ and

finally one matches the "suffix part" (a_2,b_2).

Hence one can construct a 1-fold equation analogous to (4.1) that solves the "I problem" for long enough solutions.

(c.3). $(h(1), g(\mu(1)))_{h,g}$ is infinite and $(h(0),g(\mu(0)))_{h,g}$ blocks.

This case is symmetric to the case (c.2).

(c.4). $(a_1,b_1)_{h,g}$ is successful, $(h(0), g(\mu(0)))_{h,g}$ is successful and $(h(1), g(\mu(1)))_{h,g}$ is infinite.

One has a number of cases here; we will consider only one of them (the most involved one) leaving all other to the reader.

(c.4.1). Consider a solution which looks like this.

First one completes the $(a_1,b_1)_{h,g}$ sequence,

then one completes (perhaps several times) the $(h(0), g(\mu(0)))_{h,g}$ sequence,

then one starts the $(h(1), g(\mu(1)))_{h,g}$ sequence, runs (perhaps several times) its period and then one matches it (using an appropriate (u,w)) with the suffix part (a_2,b_2).

To decide whether or not I has a solution of this form one proceeds as follows.

Let $(a_1,b_1)_h = \overline{a}_1$, $(a_1,b_1)_g = \overline{b}_1$.

Let h_1, g_1 be homomorphisms of $\{0\}^*$ defined by

$h_1(0) = 0(h(0),g(\mu(0)))_h$ and $g_1(0) = \mu(0)(h(0),g(\mu(0)))_g$.

Let α be the threshold part of $(h(1), g(\mu(1)))_{h,g}$ and β be the period part of this sequence. Let α_h, β_h be the h-projections of α and β respectively and let α_g, β_g be the g-projections of α and β respectively.

Let Z be the set of all ordered pairs (u,w) of words over $\{0,1\}$ such that either $|u| < |\beta_h|$ or $|w| < |\beta_g|$ and $h(\alpha_h)\, h(\beta_h)h(u)\, a_2 = g(\alpha_g)g(\beta_g)h(w)b_2$. Clearly if Z is empty then I has a solution of the type that we consider in (b.4.1) only if it has a solution of this type shorter than certain effectively computable positive integer constant; so one can effectively decide whether or not I has a solution when $Z = \emptyset$.

Hence assume that $Z \neq \emptyset$; clearly Z can be effectively constructed. Let h_2, g_2 be homomorphisms of $\{0\}^*$ defined by $h_2(0) = \beta_h$ and $g_2(0) = \beta_g$.

Now for every $(u,w) \in Z$ consider the equation

$$\overline{a}_1 h_1(x)\, \alpha_h\, h_2(y)\, u = \overline{b}_1 g_1(x)\, \alpha_g\, g_2(y)\, w \qquad (4.2).$$

Similarly to the case (b.2) we can see that I has a "long enough" solution if and only one of the (finite number of) equations (4.2) has a solution. Since (4.2) is a 2-fold equation, Theorem 0.2 implies that

one can effectively decide whether or not I has a long enough solution.

This settles case (c.4.1).

(c.5). $(a_1,b_1)_{h,g}$ is successful, $(h(0),g(\mu(0)))_{h,g}$ is successful and $(h(1),g(\mu(1)))_{h,g}$ blocks.

This case is easy to handle; it is similar to the case (b.2).

(c.6). $(a_1,b_1)_{h,g}$ is successful, $(h(0),g(\mu(0)))_{h,g}$ is infinite and $(h(1),g(\mu(1)))_{h,g}$ is successful.

This case is symmetric to (c.4).

(c.7). $(a_1,b_1)_{h,g}$ is successful, $(h(0),g(\mu(0)))_{h,g}$ blocks and $(h(1),g(\mu(1)))_{h,g}$ is successful.

This case is symmetric to (c.5).

Since (a), (b) and (c) exhaust all possibilities, the theorem holds. □

Theorem 4.2. It is decidable whether or not an arbitrary marked instance $I = (h,g,a_1,a_2,b_1,b_2)$ of GPCP(2) such that $h(i) = g(j)$ for some $i, j \in \{0,1\}$ has a solution.

Proof.

If ECOL(I) = ∅ then the theorem follows from Theorem 4.1.

If ECOL(I) ≠ ∅ then let us consider $ecol(h,g) = (\overline{h},\overline{g})$. Since for some $i, j \in \{0,1\}$ we have that $h(i) = g(j)$, it must be that either $|\overline{h}(0)| = |\overline{g}(0)| = 1$ or $|\overline{h}(1)| = |\overline{g}(1)| = 1$ and consequently, by Theorem 1.1 we can decide whether or not there exists an element of ECOL(I) that has a solution. Hence the theorem follows from Theorem 3.1. □

ACKNOWLEDGMENTS

The authors gratefully acknowledge support under National Science Foundation grant number MCS 79-03838.

[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, University of Waterloo, Techn. Rep. CS-79-17, 1979.

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

[EhR1] Ehrenfeucht, A., and Rozenberg, G., Simplification of homomorphisms, *Information and Control* 38, 298-309, 1978.

[EhR2] Ehrenfeucht, A. and Rozenberg, G., On unary 2-fold equations, Dept. of Comp. Science, University of Colorado at Boulder, Techn. Report No.

[EhR3] Ehrenfeucht, A. and Rozenberg, G., Generalized Post Correspondence Problem of length 2; Part II. Cases distinguished by patterns, Dept. of Comp. Science, University of Colorado at Boulder, Techn. Report No. CU-CS-189-80

[EhR4] Ehrenfeucht, A. and Rozenberg, G., Generalized Post Correspondence Problem of length 2; Part III. Decidability, Institute of Applied Dept. of Comp. Science, University of Colorado at Boulder, Techn. Report No. CU-CS-190-80

[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 insolubilité 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, 127-139, 1978.