

REPETITION OF SUBWORDS  
IN DOL LANGUAGES

by

A. Ehrenfeucht\*  
and  
G. Rozenberg\*\*

CU-CS-224-82

May, 1982

\*Department of Computer Science, University of Colorado, Boulder, Colorado,  
80309

\*\*Institute of Applied Math. and Computer Science, University of Leiden, Leiden,  
The Netherlands

This research was supported by NSF grant MCS 79-03838.

**ABSTRACT**

A language  $K \subset \Sigma^*$  is *repetitive* if for each positive integer  $n$  there exists a word  $w \in \Sigma^+$  such that  $w^n$  is a subword of  $K$ .  $K$  is called *strongly repetitive* if there exists a word  $w \in \Sigma^+$ , such that, for each positive integer  $n$ ,  $w^n$  is a subword of  $K$ . It is shown that it is decidable whether or not an arbitrary DOL language is repetitive. It is also shown that if a DOL language is repetitive then it is strongly repetitive.

## INTRODUCTION

The investigation of the combinatorial structure of languages forms an important part of formal language theory. One of the most basic combinatorial structures of languages is the repetition of subwords (in words of a language). Roughly speaking, the investigation of repetitions of subwords can be divided into two (certainly not disjoint!) directions.

- (1) The investigation of languages where repetitions of subwords (in the words of the language) are forbidden. This area was initiated by Thue ([T]) in 1906 and since then this area was a subject of active investigation in numerous areas of mathematics and in formal language theory (see, e.g., [BEM], [C], [MH], and [S1]).
- (2) The investigation of languages where repetitions of subwords (must) occur. The most classical example here is the class of context-free languages where the celebrated "pumping lemma" forces arbitrary long repetitions to be present in an infinite context-free language.

Recently one notices a revival of interest in area (1) ("Thue problems") among formal language theorists (see, e.g., [B], [H], [K], [S2]). In particular it was discovered that the theory of nonrepetitive sequences of Thue [T] is strongly related to the theory of DOL systems (see, e.g., [RS]). As a matter of fact it was pointed out in [B] that most (if not all) examples of the so called square-free sequences constructed in the literature are either DOL sequences or their codings. Thus by now quite a lot is known about DOL languages (sequences) not containing repetitions of subwords (see also [ER3]).

On the other hand very little is known on DOL languages containing repetitive structures. The pumping-like properties do not hold for DOL languages and "detecting" repetitiveness in a DOL language becomes a challenging problem.

This paper is devoted to the study of repetitiveness in DOL languages. Let us first make the notion of repetitiveness (of subwords in a language) more precise. We say that a language  $K \subseteq \Sigma^*$  is *repetitive* if for each  $n \geq 1$  there exists a word  $w \in \Sigma^+$  such that  $w^n$  is a subword of  $K$ . We say that  $K$  is *strongly repetitive* if there exists a word  $w \in \Sigma^+$  such that  $w^n$  is a subword of a word of  $K$  for each  $n \geq 1$ . It is easily seen that there exist repetitive languages that are not strongly repetitive, while on the other hand each strongly repetitive language is obviously repetitive. By the pumping lemma infinite context-free languages are strongly repetitive.

We demonstrate that

- (1) a DOL language is repetitive if and only if it is strongly repetitive and
- (2) it is decidable whether or not an arbitrary DOL system generates a repetitive language.

## 1. PRELIMINARIES

We assume the reader to be familiar with the basic theory of DOL systems (see, e.g., [RS]). We will use the standard notation and terminology concerning DOL systems (as used in [RS]).

Perhaps recalling the following notational matters will make the reading of this paper easier.

$\mathbf{N}$  denotes the set of nonnegative integers and  $\mathbf{N}^+$  denotes the set of positive integers. For a set  $A$ ,  $\#A$  denotes its cardinality.  $\Lambda$  denotes the empty word. For a nonempty word  $w$ ,  $first(w)$  denotes its first letter and  $last(w)$  denotes its last letter; for  $n \in \mathbf{N}$ ,  $pref_n(w)$  denotes the prefix of  $w$  of length  $n$  and  $sub_n(w)$  denotes the set of subwords (segments) of  $w$  of length  $n$ . Then  $sub(w)$  denotes the set of all subwords of  $w$  and for a language  $K$ ,  $sub(K) = \bigcup_{x \in K} sub(x)$ . For a DOL system  $G = (\Sigma, h, \omega)$ ,  $E(G)$  denotes its sequence,  $L(G)$  its language and  $maxr(G) = \max\{|z| : h(a) = z \text{ for some } a \in \Sigma\}$ . A letter  $a \in \Sigma$  is called *alive* if  $h^n(a) \neq \Lambda$  for all  $n \in \mathbf{N}^+$ . We will use  $T(G)$  to denote the (infinite) derivation tree corresponding to  $E(G)$ . For a node  $x$  in  $T(G)$ ,  $lb(x)$  denotes its label,  $anc(x)$  its direct ancestor and  $anc^2(x)$  the direct ancestor of  $anc(x)$ . Let  $E(G) = \omega_0, \omega_1, \dots$ . For a node  $x$  on the level  $r \geq 0$  of  $T(G)$  (counted top-down) and (an occurrence of) a subword  $z$  of  $\omega_s$  where  $s \geq r$  we use  $contr_z(x)$  to denote the contribution of  $x$  to  $z$ ; similarly if  $u$  is (an occurrence of) a subword in  $\omega_r$  then we use  $contr_z(u)$  to denote the contribution of  $u$  to  $z$ .

*In order not to overburden the (already involved) notation:*

(1) *we will often not distinguish notationally between a (sub)word and its occurrence, and*

(2) *we will often not distinguish in our notation between nodes and their*

labels;

as the precise meaning should be clear from the context, these conventions should not lead to a confusion.

We will recall now two useful notions concerning DOL systems. Let  $G = (\Sigma, h, \omega)$  be a DOL system.

A letter  $a \in \Sigma$  has *rank* 0 (in  $G$ ) (see, e.g., [ER2]) if  $L(G_a)$  is finite, where  $G_a = (\Sigma, h, a)$ . Let for  $i \geq 1, \Sigma_{(i)} = \Sigma - \{a \in \Sigma : a \text{ is of rank smaller than } i\}$  and let  $f_{(i)}$  be the homomorphism of  $\Sigma^*$  defined by :  $f_{(i)}(a) = a$  for  $a \in \Sigma_{(i)}$  and  $f_{(i)}(a) = \Lambda$  for  $a \in \Sigma - \Sigma_{(i)}$ . Then let  $h_{(i)}$  be the homomorphism of  $\Sigma_{(i)}^*$  defined by  $h_{(i)}(a) = f_{(i)}(h(a))$ . If a letter  $a \in \Sigma_{(i)}$  is such that the language of the DOL system  $(\Sigma_{(i)}, h_{(i)}, a)$  is finite then  $a$  has *rank*  $i$  (in  $G$ ). For  $i \geq 0$ , we use  $\Sigma_i$  to denote the set of all letters from  $\Sigma$  of rank  $i$ .

Let  $G = (\Sigma, h, \omega)$  and  $\bar{G} = (\bar{\Sigma}, \bar{h}, \bar{\omega})$  be DOL systems.  $\bar{G}$  is called a *simplification* of  $G$  if  $\#\bar{\Sigma} < \#\Sigma$  and there exist homomorphisms  $f : \Sigma^* \rightarrow \bar{\Sigma}^*$ ,  $g : \bar{\Sigma}^* \rightarrow \Sigma^*$  such that  $h = g f$ ,  $\bar{h} = f g$  and  $\bar{\omega} = f(\omega)$ . If  $G$  does not have a simplification it is called *elementary*. It is known ([ER1]) that if  $G$  is elementary then  $h$  is injective. If  $G_0, G_1, \dots, G_n, n \geq 0$ , is the sequence of DOL systems such that  $G_0 = G$ ,  $G_i$  is a simplification of  $G_{i-1}$  for  $1 \leq i \leq n$  and  $G_n$  is elementary, then  $G_n$  is called an *elementary version* of  $G$ .

## 2. BASIC DEFINITIONS AND RESULTS

In this section we define some basic notions (and some basic results concerning them) to be investigated in this paper. These include the main notion of (strong) repetitiveness of a language as well as several more technical notions which will be useful for proving the main results of this paper.

*Definition.* Let  $K$  be a language,  $K \subseteq \Sigma^*$ .

- (1)  $K$  is *repetitive* if for each  $n \in \mathbf{N}^+$  there exists a word  $w \in \Sigma^+$  such that  $w^n \in \text{sub}(K)$ .
- (2)  $K$  is *strongly repetitive* if there exists a word  $w \in \Sigma^+$  such that  $w^n \in \text{sub}(K)$  for each  $n \in \mathbf{N}^+$ . ■

Obviously, if  $K$  is strongly repetitive then  $K$  is repetitive, but there exist repetitive languages, that are not strongly repetitive. Consider, e.g., the language  $K_0 \subseteq \{a, b, c, d\}^+$  defined by

$$K_0 = \{(wd)^n : n \in \mathbf{N}^+, w \in \{a, b, c\}^+, |w| = n \text{ and for no } x, y \in \{a, b, c\}^*, z \in \{a, b, c\}^+, w = xzzy\}.$$

Clearly  $K_0$  is repetitive but not strongly repetitive language (notice that  $K$  is a context-sensitive language).

*Definition.* A DOL system  $G$  is called *(strongly) repetitive* if  $L(G)$  is *(strongly) repetitive*. ■

The following special subclass of DOL systems will be useful in the considerations of the next section.

*Definition.* A DOL system  $G = (\Sigma, h, \omega)$  is *pushy* if  $\text{sub}(L(G)) \cap \Sigma_0^*$  is infinite; otherwise  $G$  is *not pushy*. ■

If a DOL system  $G$  is not pushy then  $q(G)$  denotes  $\max\{|w| : w \in \text{sub}(L(G)) \cap \Sigma_0^*\}$ .

*Lemma 2.1*

- (1) It is decidable whether or not an arbitrary DOL system is pushy.
- (2) If a DOL system  $G = (\Sigma, h, \omega)$  is not pushy then  $\Sigma_i = \phi$  for all  $i > 0$ .
- (3) If a DOL system  $G$  is not pushy then  $q(G)$  is effectively computable.

*Proof.*

(1) Let  $G = (\Sigma, h, \omega)$  be a DOL system. We say that  $G$  satisfies the *edge condition* if the following holds:

there exist  $x \in \Sigma$ ,  $k \in \mathbf{N}^+$ ,  $w \in \Sigma^*$  and  $u \in \Sigma_0^+$  such that  $\text{alph}(u)$  contains an alive letter and either  $h^k(x) = wxu$  or  $h^k(x) = uxw$ .

We observe that  $G$  is pushy if and only if  $G$  satisfies the edge condition. This is seen as follows. Obviously, if  $G$  satisfies the edge condition then  $G$  is pushy.

Assume now that  $G$  is pushy. Then  $\text{sub}(L(G))$  contains arbitrary long words over  $\Sigma_0$ . Consider now a word  $z \in \Sigma_0^+ \cap \text{sub}(L(G))$ , it appears as a subword of  $\omega_r$  for some  $r \in \mathbf{N}^+$ , where  $E(G) = \omega_0, \omega_1, \dots$ . Thus we have the following situation:

Figure 1.

where  $l_z$  is the first to the left of  $z$  occurrence of a letter not in  $\Sigma_0$  and  $r_z$  is the first to the right of  $z$  occurrence of a letter not in  $\Sigma_0$ . Clearly  $z$  can be chosen so that at least one of  $l_z, r_z$  must exist as otherwise  $L(G)$  would be finite and so  $G$  could not be pushy. Assume that both  $l_z$  and  $r_z$  exist; if only one of them exist, the reasoning is even simpler. Then  $\rho_l(\rho_r)$  is the path leading from a node in  $\omega_0$  to  $l_z(r_z)$ .



Since  $\mathcal{Z}$  can be chosen arbitrarily long (at least) one of the following conditions must hold.

- (i)  $\rho_l$  contains different nodes  $n_1, n_2$  such that  $lb(n_1) = lb(n_2)$ , and both  $contr_u(n_1)$  and  $contr_u(n_2)$  contain (an occurrence of) an alive letter.
- (ii)  $\rho_r$  contains different nodes  $m_1, m_2$  such that  $lb(m_1) = lb(m_2)$  and both  $contr_u(m_1)$  and  $contr_u(m_2)$  contain (an occurrence of) an alive letter.

Then it is easily seen that  $G$  must satisfy the edge condition.

Now (1) follows from an easy observation that the edge condition is decidable (it is well known that it is decidable whether an arbitrary letter is in  $\Sigma_0$  and whether an arbitrary letter is alive).

(2) This follows directly from the definition of a letter with rank  $i > 0$ .

(3) Assume that  $G = (\Sigma, h, \omega)$  is not pushy, then  $q(G)$  exists. Clearly  $q(G) = \min\{n \in \mathbf{N}^+ : \Sigma_0^* \cap sub_{n+1}(L(G)) = \emptyset\}$ .

Thus to find  $q(G)$  it suffices to construct in succession sets  $\Sigma_0^* \cap sub_i(L(G))$ ,  $i = 1, 2, \dots$ , until one of these sets becomes empty - if this happens for the index  $i_0$  then  $q(G) = i_0 - 1$ . The existence of  $q(G)$  guarantees the termination of this algorithm. ■

Our next notion is the fundamental technical notion of this paper.

*Definition.* A DOL system  $G = (\Sigma, h, \omega)$  is called *special*, abbreviated a SDOL system, if it satisfies the following conditions.

- (0)  $G$  is reduced.
- (1)  $G$  is sliced meaning that
  - (1.1) for each  $a \in \Sigma$ , and each  $n \in \mathbf{N}^+$ ,  $alph(h^n(a)) = alph(h(a))$ ,
  - (1.2) for each  $a \in \Sigma$ , the length sequence  $\{|h^n(a)|\}_{n \geq 0}$  is either strictly increasing or constant and

- (1.3)  $\omega \in \Sigma$ .
- (2)  $G$  is strongly growing meaning that
- (2.1)  $G$  is propagating and
- (2.2) no letter in  $G$  has a rank (including the zero rank).
- (3)  $G$  is elementary. ■

The next few results bind the notion of repetitiveness with several subclasses of DOL systems as well as they indicate how this notion carries over through some operations on languages and DOL systems.

*Lemma 2.2.* Let  $G$  be a DOL system.

- (1) If  $G$  is pushy then  $G$  is strongly repetitive.
- (2) If  $G$  is finite then  $G$  is not repetitive.

Proof. (1) This follows easily from the observation made in the proof of Lemma 2.1 that the edge condition is equivalent to the pushy property.

(2) Obvious. ■

*Definition.* Let  $K$  be a language and let  $(K_1, \dots, K_n)$ ,  $n \geq 1$ , be a  $n$ -tuple of languages. Then  $K < (K_1, \dots, K_n)$  if  $K \subseteq K_1 K_2 \dots K_n$  and  $K_i \subseteq \text{sub}(K)$  for each  $1 \leq i \leq n$ . ■

*Lemma 2.3.* Let  $K, K_1, \dots, K_n$ ,  $n \geq 1$ , be languages. (1) Let

$$K = \bigcup_{i=1}^n K_i. \text{ Then}$$

$K$  is (strongly) repetitive if and only if there exists a  $1 \leq i \leq n$

such that  $K_i$  is (strongly) repetitive.

(2) Let  $K < (K_1, \dots, K_n)$ . Then

$K$  is (strongly) repetitive if and only if there exists a  $1 \leq i \leq n$  such that  $K_i$  is (strongly) repetitive.

*Proof.*

Obvious. ■

*Lemma 2.4.* Let  $G$  be a DOL system and let  $G'$  be its simplification. Then  $G$  is (strongly) repetitive if and only if  $G'$  is (strongly repetitive).

*Proof.*

Follows immediately from the fact that one can homomorphically "translate" from  $G$  to  $G'$  and from  $G'$  to  $G$ . ■

### 3. MAIN RESULTS

In this section we state two main results of this paper and indicate the strategy of their proofs.

The following two results are the main results of this paper.

*Theorem 1.* It is decidable whether or not an arbitrary DOL system  $G$  is repetitive. ■

*Theorem 2.* Every repetitive DOL system is strongly repetitive. ■

In order to prove these results we will prove the following two (more technical) theorems. They allow us to concentrate on SDOL systems (rather than consider arbitrary DOL systems).

*Theorem 3.*

- (1) It is decidable whether or not an arbitrary DOL system is repetitive if and only if it is decidable whether or not an arbitrary SDOL system is repetitive.
- (2) If every repetitive SDOL system is strongly repetitive, then every repetitive DOL system is strongly repetitive. ■

*Theorem 4.*

- (1) It is decidable whether or not an arbitrary SDOL system is repetitive.
- (2) Every repetitive SDOL system is strongly repetitive. ■

Clearly Theorem 3 and Theorem 4 together imply Theorem 1 and Theorem 2. Thus the rest of this paper is devoted to proofs of Theorem 3 and Theorem 4.

In the next section we prove Theorem 3. In Section 5 we consider closed and strongly closed subalphabets of the alphabet of a SDOL system. Considerations of this section form important technical tools for Section 6 where Theorem 4 is proved.

#### 4. PROOF OF THEOREM 3

In this section Theorem 3 is proved.

*Theorem 3.*

- (i) It is decidable whether or not an arbitrary DOL system is repetitive if and only if it is decidable whether or not an arbitrary SDOL system is repetitive.
- (ii) If every repetitive SDOL system is strongly repetitive then every repetitive DOL system is strongly repetitive. ■

*Proof.* (i) Clearly it suffices to prove the if part of the statement only. To this aim we proceed as follows.

Let  $G = (\Sigma, h, \omega)$  be an arbitrary DOL system.

First we decide whether or not  $G$  is finite (it is well known that finiteness is decidable for DOL systems). If  $G$  is finite then (see Lemma 2.2(2))  $G$  is not repetitive and we are done. If  $G$  is infinite then (see Lemma 2.1(1)) we decide whether or not  $G$  is pushy. If it is, then (by Lemma 2.2(1))  $G$  is strongly repetitive and we are done.

Thus let us assume that  $G$  is not pushy.

Let  $G^c$  be the "coded version of  $G$ " defined as follows:  $G^c = (\Sigma^c, h^c, \omega^c)$

where

$$\Sigma^c = \{ (\alpha, x, \beta) : x \in \Sigma - \Sigma_0, \alpha, \beta \in \Sigma_0^* \text{ and } |\alpha|, |\beta| \leq q(G) \},$$

$$\omega^c = (\alpha_1, y_1, \alpha_2)(\alpha_2, y_2, \alpha_3) \dots (\alpha_{n-1}, y_{n-1}, \alpha_n) \text{ where}$$

$$\omega = \alpha_1 y_1 \alpha_2 y_2 \dots \alpha_{n-1} y_{n-1} \alpha_n, y_i \in \Sigma - \Sigma_0 \text{ and } \alpha_j \in \Sigma_0^* \text{ for } 1 \leq i \leq n-1$$

and  $1 \leq j \leq n$ ,

for  $(\alpha, x, \beta) \in \Sigma^c$ ,

$$h^c((\alpha, x, \beta)) = (h(\alpha)\alpha_1, y_1, \alpha_2)(\alpha_2, y_2, \alpha_3) \dots (\alpha_{n-1}, y_{n-1}, \alpha_n h(\beta))$$

where

$$h(x) = \alpha_1 y_1 \alpha_2 y_2 \dots \alpha_{n-1} y_{n-1} \alpha_n, y_i \in \Sigma - \Sigma_0 \text{ and } \alpha_j \in \Sigma_0^* \text{ for } l \leq i \leq n-1$$

and  $1 \leq j \leq n$ .

By Lemma 2.1  $G^c$  is effectively constructible.

*Claim 4.1.*

- (1)  $G^c$  is (strongly) repetitive if and only if  $G$  is (strongly) repetitive.
- (2)  $G^c$  is strongly growing.

*Proof of Claim 4.1*

- (1) This follows directly from the following obvious observation.

If  $E(G) = \omega_0, \omega_1, \dots$  and  $E(G^c) = \omega_0^c, \omega_1^c, \dots$ , then, for every  $m \geq 0$ ,

$\omega_m = \alpha_1 y_1 \cdots \alpha_{n-1} y_{n-1} \alpha_n$  if and only if

$\omega_m^c = (\alpha_1, y_1, \alpha_2)(\alpha_2, y_2, \alpha_3) \dots (\alpha_{n-1}, y_n, \alpha_n)$  where  $y_i \in \Sigma - \Sigma_0$  and  $\alpha_j \in \Sigma_0^*$  for  $1 \leq i \leq n-1$  and  $1 \leq j \leq n$ .

- (2) Since  $G$  is not pushy, no letter outside  $\Sigma_0$  has a rank. Consequently no letter in  $G^c$  has a rank and so  $G^c$  is strongly growing. ■

*Claim 4.2.* There exists an algorithm which given a strongly growing DOL system  $H$  produces a finite set  $H_1, \dots, H_t$ ,  $t \geq 1$ , of DOL systems such that

- (1)  $H$  is (strongly) repetitive if and only if  $H_i$  is (strongly) repetitive for some  $1 \leq i \leq t$ ,
- (2)  $H_i$  is special for each  $1 \leq i \leq t$ .

*Proof of Claim 4.2:*

Consider the algorithm **A** defined by the following diagram

Figure 2.

where inputs are strongly growing DOL systems and the operations are defined as follows.

## SLICE

Let  $H = (\Theta, g, \rho)$  be a strongly growing DOL system. It is well known that, for each  $\alpha \in \Theta$ ,

$$\text{alph}(g(\alpha)), \text{alph}(g^2(\alpha)), \text{alph}(g^3(\alpha)), \dots$$

is an ultimately periodic sequence; let  $p_\alpha$  be a fixed positive integer which is a multiplicity of a period of this sequence and is bigger than a threshold of this sequence.

Let, for each  $\alpha \in \Theta$ ,  $r_\alpha$  be a positive integer such that the sequence  $|g^{r_\alpha}(\alpha)|, |g^{2r_\alpha}(\alpha)|, \dots$  is a strictly growing sequence of positive integers; it is well-known that such an  $r_\alpha$  exists.

Let  $s$  be the least common multiple of all the integers  $p_\alpha, r_\alpha$ . Then  $\text{SLICE}(H) = \{H_0, H_1, \dots, H_{s-1}\}$  where  $H_i = (\Theta, g^i, g^i(\rho))$  for  $0 \leq i \leq s-1$ .

## SPLIT

Let  $\mathbf{H}$  be a set of (strongly growing) DOL systems. Then  $\text{SPLIT}(\mathbf{H}) = \bigcup_{H \in \mathbf{H}} \overline{\text{SPLIT}}(H)$  where for  $H = (\Theta, g, \rho)$ ,

$\overline{\text{SPLIT}}(H) = \{H_1, \dots, H_{|\rho|}\}$  with  $H_i = (\Theta, g, a_i)$  for each  $1 \leq i \leq |\rho|$  where  $a_i$  is the  $i$ 'th letter of  $\rho$ .

## REMOVE UNACCESSIBLE (RU)

Let  $\mathbf{H}$  be a set of (strongly growing) DOL systems. Then  $\text{RU}(\mathbf{H}) = \bigcup_{H \in \mathbf{H}} \overline{\text{RU}}(H)$  where for  $H = (\Theta, g, \rho)$ ,  $\overline{\text{RU}}(H) = (\overline{\Theta}, \bar{g}, \rho)$  where

$$\overline{\Theta} = \bigcup_{i=0}^{\infty} \text{alph}(g^i(\rho)) \text{ and } \bar{g} \text{ equals } g \text{ restricted to } \overline{\Theta}^*.$$

## ULTIMATELY SIMPLIFY (US)

Let  $\mathbf{H}$  be a set of (strongly growing) DOL systems. Then  $\text{US}(\mathbf{H}) = \bigcup_{H \in \mathbf{H}} \overline{\text{US}}(H)$  where  $\overline{\text{US}}(H)$  is an elementary version of  $H$ .

It is easily seen that when  $\mathbf{A}$  is given a strongly growing DOL system  $H$ , it produces a finite set  $H_1, \dots, H_t$ ,  $t \geq 1$ , of DOL systems which are special. Hence (2) of the statement of the claim holds. Then part (1) of the statement follows directly from Lemma 2.3 and Lemma 2.4. ■

Now we complete the proof of Theorem 3.(i) as follows.

Let us consider the algorithm  $\mathbf{R}$  given by the following diagram.

Figure 3.

Clearly, if it is decidable whether or not an arbitrary SDOL system is repetitive, then (from Claim 4.1 and Claim 4.2 it follows that) the algorithm  $\mathbf{R}$  decides whether or not an arbitrary DOL system is repetitive.

Hence (i) holds.

(ii) To prove (ii) let us assume that every repetitive SDOL system is strongly repetitive. Let us analyze the algorithm  $\mathbf{R}$  and in particular the cases when it decides that a DOL system in question is repetitive. There are two such cases.

(1) The answer "repetitive" given on the exit YES from the test "Is  $G$  pushy?". In this case, by Lemma 2.2(1),  $G$  is also strongly repetitive.

(2) The answer "repetitive" given on the exit YES from the test "Is one of  $G_i^c$  repetitive?". In this case we know that (at least) one of the "component systems"  $G_1^c, \dots, G_t^c$  is repetitive; since all these systems are special, our assump-



tion implies that (at least) one of the systems  $G_1^c, \dots, G_t^c$  is strongly repetitive.

Then, by Lemma 2.3 and Lemma 2.4,  $G$  is strongly repetitive.

Hence, whenever  $G$  is repetitive it is also strongly repetitive and (ii) holds.

Consequently Theorem 1 holds. ■

## 5. CLOSED AND STRONGLY CLOSED SUBSETS OF $\Sigma$

In this section we define and investigate closed and strongly closed subsets of (the alphabet of) a DOL system. The results of this section form an important tool in proving Theorem 4 in the next section.

Let  $G = (\Sigma, h, S)$  be a DOL system and let  $\Theta$  be a nonempty subset of  $\Sigma$ . We say that  $\Theta$  is *closed* (with respect to  $G$ ) if  $h(a) \in \Theta^*$  for each  $a \in \Theta$  and we say that  $\Theta$  is *strongly closed* (with respect to  $G$ ) if  $\text{alph}(h(a)) = \Theta$  for each  $a \in \Theta$ . Note that if  $\Theta$  is strongly closed then it is also closed.

Let  $w \in \text{sub}(L(G))$ , let  $\Theta \subseteq \Sigma$  and let  $u \in \Theta^+$ . We say that  $u$  is a  $\Theta$ -block (of  $w$ ) if  $w = \alpha a u b \beta$  where  $a, b \in \Sigma - \Theta$ . A  $\Theta$ -block  $u$  is *maximal* in  $w$  if no other  $\Theta$ -block in  $w$  is longer than  $u$ ; then  $B_{\Theta}(w)$  denotes the number of different maximal  $\Theta$ -blocks in  $w$ . (E.g., if  $w = a^3 c a^2 c^2 a^2 c a c$  and  $\Theta = \{a\}$  then  $B_{\Theta}(w) = 2$ ).

Now let  $G = (\Sigma, h, \omega)$  be an arbitrary (but fixed) special DOL system with  $E(G) = \omega_0, \omega_1, \dots$  and let  $m = \text{maxr}(G)$ . We will investigate several useful properties of closed and strongly closed subsets of  $\Sigma$ .

*Lemma 5.5*  $B_{\Theta}(w) \leq \#\Sigma m^4$  for each closed subset  $\Theta$  of  $\Sigma$  and each  $w \in \text{sub}(L(G))$ .

*Proof.*

Let  $\Theta$  be a closed subset of  $\Sigma$  and let  $w \in \text{sub}(L(G))$ . If  $B_{\Theta}(w) \neq 0$  then  $w$  contains a  $\Theta$ -block; let us consider a maximal  $\Theta$ -block  $u$  in  $w$ , assume that  $w = w_1 a u b w_2$  where  $a, b \in \Sigma - \Theta$  and consider the depicted occurrences of  $u$ ,  $a$  and  $b$  (these occurrences of  $a$  and  $b$  are referred to as the *left* and the *right border* of the given occurrence of  $u$  respectively). Since  $w \in \text{sub}(L(G))$ , there exists a  $r \in \mathbf{N}^+$  such that  $w$  is on the  $r$ -th level of  $T(G)$  ( $w \in \text{sub}(\omega_r)$ ). The situation can be illustrated as follows:

Figure 4.

where  $X$  is the first (bottom-up) common ancestor node of  $\mathit{first}(w)$  and  $\mathit{last}(w)$  while  $p(u)$  is the first (bottom-up) common ancestor node of the left and the right border of  $u$  ( $p(u)$  is called the *real ancestor* of  $u$ ; the (sub)tree rooted at  $X$  and ending at the  $r$ -th level (of  $T(G)$ ) is denoted  $T_w$ , the (sub)tree rooted at  $p(u)$  and ending at the  $r$ -th level is denoted  $T_u$ . The leftmost (rightmost) path in a tree leading from its root to a leaf is referred to as its *leftmost* (*rightmost*) *boundary*; both paths together constitute the *boundary* of the tree.

Let us consider  $T_u$  separately:

Figure 5.

where  $h(\mathit{lb}(p(u))) = \mathit{lb}(d_1) \dots \mathit{lb}(d_{|h(\mathit{lb}(p(u)))|})$ ,  $d_k$  falls on the left boundary of  $T_u$  and  $d_l$  falls on the right boundary of  $T_u$  where  $1 \leq k < l \leq |h(\mathit{lb}(p(u)))|$ .

We define the *type* of  $u$ , denoted as  $\mathit{type}(u)$  to be the triple  $(\mathit{lb}(p(u)), k, l)$ .

*Claim 5.3.* Either  $p(u)$  or  $\mathit{anc}(p(u))$  or  $\mathit{anc}^2(p(u))$  lie on the boundary of  $T_w$ .

*Proof of Claim 5.3:*

Assume that  $p(u)$  does not lie on the boundary of  $T_w$ . We show that then either  $\mathit{anc}(p(u))$  or  $\mathit{anc}^2(p(u))$  lie on the boundary of  $T_w$ . This is proved by contradiction as follows.

Assume that neither  $\mathit{anc}(p(u))$  nor  $\mathit{anc}^2(p(u))$  lie on the boundary of  $T_w$  - that is they are both nodes within  $T_w$  but outside of  $T_u$ . Since  $G$  is special, (the label of)  $\mathit{anc}^2(p(u))$  derives an occurrence of  $p(u)$  in one step and

consequently  $\omega_{r-1}$  contains an occurrence of  $aub$  that will contribute to the (given occurrence of) subword  $w$  a  $\Theta$ -block longer than  $u$  (remember that  $G$  is special and  $\Theta$  is closed); a contradiction because  $u$  is a maximal  $\Theta$ -block in  $w$ .

Hence Claim 5.3 holds. ■

Claim 5.4. For each  $s < r$  the number of real ancestors of maximal  $\Theta$ -blocks from  $w$  is bounded by  $2m^2$ .

*Proof of Claim 5.4.*

This follows directly from Claim 5.3. ■

Claim 5.5. If  $u$  and  $u'$  are maximal blocks of  $w$  such that  $\text{type}(u) = \text{type}(u')$  then both  $p(u)$  and  $p(u')$  are on the same level of  $T(G)$ .

*Proof of Claim 5.5.*

Assume to the contrary that  $p(u)$  is on the level  $\omega_{s_1}$  and  $p(u')$  is on the level  $\omega_{s_2}$  for some  $s_1 \neq s_2$  (say  $s_1 < s_2$ ).

Let  $a$  and  $b$  be the left and the right border of  $u$  in  $w$  and let  $a', b'$  be the left and the right border of  $u'$  in  $w$ . Let  $a_0 = a$ ,  $a_1$  be the ancestor of  $a$ , ... and  $a_i$  be the ancestor of  $a_{i-1}$  for  $1 \leq i \leq (s_1 - r)$ ; analogously let  $b_1, b_2, \dots, b_{s_1 - r}$  be the line of direct ancestors starting with the direct ancestor  $b_1$  of  $b_0 = b$ .

Hence we have the following situation:

Figure 6.

and an analogous situation for  $u'$ .

We observe that the position of  $a_{i-1}$  among the direct descendant nodes of  $a_i$

and the position of  $b_{i-1}$  among the direct descendant nodes of  $b_i$ ,  $1 \leq i \leq (s_1-r)-1$ , are uniquely determined by *type* ( $p(u)$ ). (Since  $\Theta$  is closed,  $a_{i-1}$  is the rightmost direct descendant  $x$  of  $a_i$  such that  $h(x) \notin \Theta^+$  and  $b_{i-1}$  is the leftmost direct descendant  $x$  of  $b_i$  such that  $h(x) \notin \Theta^+$  while positions of  $a_{(s_1-r)-1}$  and  $b_{(s_1-r)-1}$  in  $h(p(u))$  are determined by the type of  $p(u)$ ). An analogous observation holds for the subtree rooted at  $p(u')$  and ending at  $a'u'b'$ .

This implies that if  $s_1 \neq s_2$  (say  $s_1 < s_2$ ) then  $T_u$  is isomorphic to a strict subtree of  $T_{u'}$ , rooted at the root of  $T_{u'}$ , and consequently (because  $G$  is special and  $\Theta$  is closed)  $|u'| > |u|$  which contradicts the fact that both  $u$  and  $u'$  are maximal  $\Theta$ -blocks in  $w$ . ■

Now we complete the proof of Lemma 5.5 as follows.

By Claim 5.5 the number of maximal  $\Theta$ -blocks in  $w$  is bounded by the product of the number of different types of maximal  $\Theta$ -blocks in  $w$  by the number of real ancestors of maximal  $\Theta$ -blocks on one level of  $T(G)$ . Consequently, by Claim 5.4 the number of maximal  $\Theta$ -blocks in  $w$  is not bigger than  $(\#\Sigma \cdot \binom{m}{2}) \cdot 2m^2 \leq \#\Sigma m^4$ .

Thus Lemma 5.5 holds. ■

Now we move to investigate strongly closed subsets of  $\Sigma$ . We start by noting the following property.

*Lemma 5.6* Let  $\Theta$  be a strongly closed subset of  $\Sigma$ . For every  $n \in \mathbf{N}^+$  and every  $a, b \in \Theta$ ,  $|h^n(b)| \leq m \cdot |h^n(a)|$ .

*Proof.*

Since  $\Theta$  is strongly closed,  $b \in \text{alph}(h(a))$  and so  $|h^{n+1}(a)| \geq |h^n(b)|$  for each  $n \in \mathbf{N}^+$ . But  $m |h^n(a)| \geq |h^{n+1}(a)|$  and

so  $m |h^n(a)| \geq |h^n(b)|$  which proves the lemma. ▀

The relevance of strongly closed subsets of  $\Sigma$  to the investigation of repetitive properties of  $G$  stems from the following result.

*Lemma 5.7.* If  $n > \#\Sigma m^4 + 4$  then for each  $w \neq \Lambda$  holds: if  $w^n \in \text{sub}(L(G))$  then  $\text{alph}(w)$  is strongly closed.

*Proof.*

Let  $n > 2m^2 + 4$ ,  $w \neq \Lambda$  and let  $w^n \in \text{sub}(L(G))$ . Hence for some  $r > 1$ ,  $w^n$  is a subword of  $\omega_r$ . Thus we have the following situation:

Figure 7.

Consider now the  $\lfloor \frac{n}{2} \rfloor$ -th occurrence of  $w$  in (the depicted occurrence of)  $w^n$ . Let  $a$  be an arbitrary element of  $\text{alph}(w)$  and consider an arbitrary occurrence of  $a$  in the given occurrence of  $w$  (in  $w^n$ ). Let  $x = \text{anc}^2(a)$ :

Figure 8.

Since  $n > \#\Sigma m^4 + 4$ , all second generation descendants of  $x$  lie within (the given occurrence of)  $w^n$ . On the other hand,  $a \in \text{alph}(h^2(x))$  implies that (because  $G$  is special)  $a \in \text{alph}(h(x))$  and so by the above  $\text{alph}(h(a)) \subseteq \text{alph}(w)$ . Thus  $\text{alph}(w)$  is closed.

Assume now that  $\text{alph}(w)$  is not strongly closed. Thus  $\text{alph}(w)$  contains a letter, say  $b$ , such that  $\text{alph}(h(b)) \neq \text{alph}(w)$ ; let  $\text{alph}(h(b)) = \Theta$ . This means that  $\text{alph}(h(w))$  contains a letter not in  $\Theta$  and consequently  $h(w^n)$  contains at least  $n - 2 = \#\Sigma m^4 + 2$  maximal  $\Theta$ -blocks. This however contradicts Lemma 5.5 and so  $\text{alph}(w)$  must be strongly closed. ▀

We define now a concept important for our further considerations.

Let  $z \in L(G)$ ,  $\Theta$  be a strongly closed subset of  $\Sigma$  and let  $u \in \text{sub}(z) \cap \Theta^+$ . Thus we have the following situation

Figure 9.

where  $X$  is the first (bottom-up) common ancestor of  $\text{first}(u)$  and  $\text{last}(u)$ ;  $T_u$  is a subtree of  $T(G)$  rooted at  $X$  with  $u$  being its frontier.

The *cover* of  $u$ , denoted  $\text{cov}(u)$ , is the subgraph of  $T_u$  spanned on all nodes of  $T_u$  the contribution of which to  $\omega_r$  is totally included in  $u$  (this includes also nodes from  $u$ ). The *surface* of  $\text{cov}(u)$ , denoted  $\text{sur}(u)$ , consists of all nodes of  $\text{cov}(u)$  such that their direct ancestor in  $T(G)$  is not in  $\text{cov}(u)$ . Let  $s < r$  be the smallest integer such that some nodes of  $\text{cov}(u)$  are on the level  $s$  of  $T(G)$ . All nodes from  $\text{cov}(u)$  on the level  $s$  form the *level 0* of  $\text{cov}(u)$  - their set is denoted by  $\text{cov}_0(u)$ ; all nodes from  $\text{cov}(u)$  which are on level  $s+1$  of  $T(G)$  form *level 1* of  $\text{cov}(u)$  - their set is denoted by  $\text{cov}_1(u)$ , and so on up to  $i = (s-r)$  where  $(s-r)$  is called the *height* of  $\text{cov}(u)$  and denoted by  $ht(u)$ .

*Lemma 5.8.* The number of surface nodes on each level of  $\text{cov}(u)$  is bounded by  $2m^2$ .

*Proof.*

First we note that for each node  $x$  in  $\text{sur}(u)$  either  $x$  or  $\text{anc}(x)$  or  $\text{anc}^2(x)$  is on the boundary of  $T_u$ . This is seen as follows. Assume that  $x$  is not on a boundary of  $T_u$ . Now if neither  $\text{anc}(x)$  nor  $\text{anc}^2(x)$  are on the boundary of  $T_u$  then  $\text{anc}^2(x)$  is within  $T_u$  and its contribution to  $x$  lies within  $u$ . Since  $G$  is special this implies that  $lb(\text{anc}(x)) \in \Theta$  and so  $x$  is not in  $\text{sur}(u)$ ; a contradiction.

But if each node  $x$  in  $sur(u)$  is such that either  $x$  or  $anc(x)$  or  $anc^2(x)$  is on the boundary of  $T'_u$ , then the number of surface nodes on each level of  $cov(u)$  is bounded by  $2m^2$  (no more than  $m^2$  nodes coming from left boundary and no more than  $m^2$  nodes coming from the right boundary). ■

*Lemma 5.9* For each node  $b$  of  $cov_0(u)$ ,  $\frac{|contr_u(b)|}{|u|} > \frac{1}{4m^3}$ .

*Proof.*

*Claim 5.6.* For every node  $b \in cov_0(u)$ , every positive integer  $k \in \{0, \dots, ht(u)\}$  and every node  $a \in cov_k(u)$ ,

$$|contr_u(a)| \leq \frac{m \cdot |contr_u(b)|}{2^k}.$$

*Proof of Claim 5.6.*

Since  $G$  is special, for all positive integers  $r, s$  with  $s \leq r$  we have (remember that we identify nodes with their labels whenever it does not lead to a confusion)  $\frac{|h^r(a)|}{2^s} \geq |h^{r-s}(a)|$ . Thus, by Lemma 5.6,

$$m \cdot \frac{|h^r(b)|}{2^s} \geq |h^{r-s}(a)| \text{ and so for } r = ht(u) \text{ and } s = k \text{ we get}$$

$$|contr_u(a)| \leq \frac{m \cdot |contr_u(b)|}{2^k}. \quad \blacksquare$$

Clearly  $u$  can be expressed as the catenation (in proper order) of contributions to  $u$  from all letters from  $sur(u)$ . Thus  $|u| = \sum_{i=0}^{ht(u)-1} C_i$  where  $C_i$  is the joint length of contributions of all the letters from  $sur(u) \cap cov_i(u)$ . Assume now that  $b$  is a node in  $cov_0(u)$ . Hence, by Claim 5.6 and Lemma 5.8, for each  $1 \leq i \leq ht(u)-1$

$$C_i \leq \frac{2 \cdot m^3 \cdot |contr_u(b)|}{2^i} \text{ and so}$$



$$\begin{aligned}
|u| &\leq \sum_{i=0}^{ht(u)-1} \frac{2 \cdot m^3 \cdot |contr_u(b)|}{2^i} \\
&< 2 \cdot m^3 \cdot |contr_u(b)| \sum_{i=0}^{\infty} \frac{1}{2^i} \\
&< 4 \cdot m^3 \cdot |contr_u(b)|.
\end{aligned}$$

Thus  $\frac{|contr_u(b)|}{|u|} > \frac{1}{4m^3}$  and the lemma holds. ■

*Lemma 5.10.* Let  $0 \leq l \leq ht(u)-1$  and let  $a \in cov_l(u)$ . Then

$$\frac{|contr_u(a)|}{|u|} > \frac{1}{4m^{4+l}}.$$

*Proof.*

Let  $c \in \Sigma$ . Then obviously, for each  $r, s \geq 1$  such that  $r \geq s$ ,

$$|h^r(c)| \leq m^{r-s} \cdot |h^s(c)|.$$

Thus if we take  $b \in cov_0(u)$  and set  $r = ht(u)$ ,  $l = r-s$  we get

$$|h^{ht(u)}(b)| \leq m^l \cdot |h^{ht(u)-l}(b)|$$

and so by Lemma 5.6 we get

$$|contr_u(b)| \leq m^{l+1} \cdot |contr_u(a)|.$$

Consequently by Lemma 5.9 we get

$$\frac{|contr_u(a)|}{|u|} > \frac{1}{4m^{4+l}}. \quad \blacksquare$$

## 6. PROOF OF THEOREM 4

In this section we provide a proof of Theorem 4. We start by introducing the following useful notion.

Let  $\Theta$  be a strongly closed subset of  $\Sigma$ ; we assume some fixed order of elements of  $\Theta$ . Let  $\pi$  be a cyclic permutation of  $\Theta$ . We say that  $h$  is  $(\Theta, \pi)$ -cyclic if the following two conditions are satisfied:

- (1) for each  $x \in \Theta$ , if  $h(x) = x_1 \cdots x_m$  where  $x_1, \dots, x_m \in \Theta$ , then  $x_{i+1} = \pi(x_i)$  for each  $i \leq i \leq m-1$ ,
- (2) for each  $x, y \in \Theta$ , if  $\pi(x) = y$  then  $\pi(\text{last}(h(x))) = \text{first}(h(y))$ .

*Lemma 6.11.* If  $h$  is  $(\Theta, \pi)$ -cyclic, then for every  $x \in \Theta$  and every  $n \in \mathbf{N}^+$  there exists a  $w \in \Theta^+$  such that  $|w| = \#\Theta$  and  $w^n \in \text{sub}(h^n(x))$ .

*Proof.*

Assume that a word  $u$  is of the form  $u = a\pi(a) \cdots \pi^t(a)$  for some  $a \in \Theta$  and  $t \geq 0$ . Then  $h(u) = b\pi(b) \cdots \pi^{t'}(b)$  where  $(t'+1) \geq (t+1)\max\{2, \#\Theta\}$ .

This observation follows directly from the definition of a  $(\Theta, \pi)$ -cyclic homomorphism (and the fact that  $h$  is strictly growing).

Then the lemma follows from this observation: for each  $x \in \Theta$  and  $n \in \mathbf{N}^+$  it suffices to take  $w = \text{pref}_{\#\Theta}(h^n(x))$ . ■

*Lemma 6.12.* There exists a  $\rho \in \mathbf{N}^+$  such that, for each  $w \neq \Lambda$  and each  $n \geq \rho$ , if  $w^n \in \text{sub}(L(G))$ , then  $\text{alph}(w)$  is strongly closed and there exists a permutation  $\pi$  of  $\text{alph}(w)$  such that  $h$  is  $(\text{alph}(w), \pi)$ -cyclic.

*Proof.*

Let  $\rho = 12 \cdot \#\Sigma m^8$ . Since  $\rho > \#\Sigma m^4 + 4$ , Lemma 5.7 implies that if  $w \neq \Lambda$ ,  $n \geq \rho$  and  $w^n \in \text{sub}(L(G))$  then  $\text{alph}(w)$  is strongly closed.

So let  $n \geq \rho$  and let  $w \neq \Lambda$  be such that  $v = w^n \in \text{sub}(L(G))$ . Let  $\Theta = \text{alph}(w)$  - we know that  $\Theta$  is strongly closed. Thus we leave the following situation:

Figure 10.

Let  $q = \text{ht}(v)$ . Clearly we can assume that  $\text{cov}(v)$  has at least five levels.

*Claim 6.7.*

- (1) for each  $a \in \text{cov}_4(v)$ ,  $|\text{contr}_v(a)| \geq 3|w|$ ,
- (2) for each  $a \in \text{cov}_3(v)$ ,  $|\text{contr}_v(a)| \geq 3m|w|$ .

*Proof of Claim 6.7.*

- (1) Let  $a \in \text{cov}_4(v)$ . By Lemma 5.10,

$$\frac{|\text{contr}_v(a)|}{|v|} > \frac{1}{4m^8}.$$

Hence

$$|\text{contr}_v(a)| > \frac{n|w|}{4m^8}$$

and since  $n \geq \rho$

$$|\text{contr}_v(a)| \geq \frac{\#\Sigma \cdot 12|w|m^8}{4m^8} = \#\Sigma \cdot 3|w| \geq 3|w|.$$

- (2) Analogously for  $a \in \text{cov}_3(v)$  we get

$$|\text{contr}_v(a)| \geq \#\Sigma \cdot 3m|w| \geq 3m|w|. \quad \square$$

Let  $p \in \text{cov}_0(v)$  and let  $\gamma_3$  be the (occurrence of) the subword contributed by  $p$  to  $\text{cov}_3(v)$  and let  $\gamma_4$  be the (occurrence of) subword contributed by  $p$  to  $\text{cov}_4(v)$ . Since  $\Theta$  is strongly closed and  $G$  is special, each letter of  $\gamma_3$  occurs at least 3 times in  $\gamma_3$  and each letter of  $\gamma_4$  occurs at least 3 (even 4)

times in  $\gamma_4$ .

We will consider now  $\gamma_3$  and  $\gamma_4$ . Let  $x$  be an arbitrary letter of  $\gamma_3$  and let  $y$  be an arbitrary letter of  $\gamma_4$ . Then we will consider arbitrary three consecutive occurrences of  $x$  in  $\gamma_3$  and arbitrary three consecutive occurrences of  $y$  in  $\gamma_4$ ; let then  $\gamma_3 = \cdots x\alpha_1x\alpha_2x \cdots$  and  $\gamma_4 = \cdots y\beta_1y\beta_2y \cdots$  where the depicted occurrences of  $x$  and  $y$  are the considered occurrences ( $x \notin \text{alph}(\alpha_1\alpha_2)$  and  $y \notin \text{alph}(\beta_1\beta_2)$ ).

*Claim 6.8.*  $\alpha_1 = \alpha_2$  and  $\beta_1 = \beta_2$ .

*Proof of Claim 6.8.*

We will prove that  $\beta_1 = \beta_2$ ; the proof of equality  $\alpha_1 = \alpha_2$  is analogous.

By *Claim 6.7*,  $\text{contr}_v(y) = w_1w^lw_2$  for some  $l \geq 1$ , a proper suffix  $w_1$  of  $w$  and a proper prefix  $w_2$  of  $w$ .

We have four possible cases to consider.

(1)  $\text{contr}_v(\beta_1) = \text{contr}_v(\beta_2) = \Lambda$ .

This can happen only if  $\beta_1 = \beta_2 = \Lambda$  - then indeed  $\beta_1 = \beta_2$ .

(2)  $\text{contr}_v(\beta_1) = \Lambda$  and  $\text{contr}(\beta_2) = \bar{w}_1w^{\bar{l}}\bar{w}_2$  for some  $\bar{l} \geq 1$ , a proper suffix  $\bar{w}_1$  of  $w$  and a proper prefix  $\bar{w}_2$  of  $w$ .

Then

$$\text{contr}_v(y\beta_1y\beta_2y) = w_1w^lw_2w_1w^lw_2\bar{w}_1w^{\bar{l}}\bar{w}_2w_1w^lw_2 = \delta.$$

Consequently  $w_2w_1 = w$ ,  $w_2\bar{w}_1 = w$  and  $\bar{w}_2w_1 = w$  and so  $w_1 = \bar{w}_1$  and  $w_2 = \bar{w}_2$ .

$$\text{Thus } \delta = w_1w^{3l+\bar{l}+3}w_2 = h^{q-4}(yy\beta_2y).$$

On the other hand

$$\begin{aligned} h^{q-4}(y\beta_2yy) &= w_1w^lw_2w_1w^{\bar{l}}w_2w_1w^lw_2w_1w^lw_2 \\ &= w_1w^{3l+\bar{l}+3}w_2 = \delta \end{aligned}$$

Thus  $h^{q-4}(y\beta_2yy) = h^{q-4}(yy\beta_2y)$  where  $y\beta_2yy \neq yy\beta_2y$ .

Since  $h$  is injective, this is a contradiction and consequently case (2) is impossible.

(3)  $contr_v(\beta_1) = \bar{w}_1 w^{\bar{l}} \bar{w}_2$  for some  $\bar{l} \geq 1$ , a proper suffix  $\bar{w}_1$  of  $w$  and a proper prefix  $\bar{w}_2$  of  $w$  and  $contr_v(\beta_2) = \Lambda$ .

This case is analogous to case (3).

(4)  $contr_v(\beta_1) = w_1' w^{l'} w_2'$  and  $contr(\beta_2) = \bar{w}_1 w^{\bar{l}} \bar{w}_2$  for some  $l', \bar{l} \geq 1$ , proper suffixes  $w_1', \bar{w}_1$  of  $w$  and proper prefixes  $w_2', \bar{w}_2$  of  $w$ .

Then

$$contr_v(y\beta_1 y\beta_2 y) = w_1' w^{l'} w_2' w_1' w^{l'} w_2' w_1' w^{l'} w_2' \bar{w}_1 w^{\bar{l}} \bar{w}_2 w_1' w^{l'} w_2' = \delta.$$

Consequently  $w_2 w_1' = w_2' w_1 = w_2 \bar{w}_1 = \bar{w}_2 w_1 = w$ . Hence  $w_1' = \bar{w}_1$  and  $w_2' = \bar{w}_2$ .

Thus

$$\delta = w_1 w^{3l+l'+4} w_2 = h^{q-4}(y\beta_1 y\beta_2 y).$$

On the other hand

$$h^{q-4}(y\beta_2 y\beta_1 y) = w_1 w^{3l+l'+4} w_2 = h^{q-4}(y\beta_1 y\beta_2 y).$$

If  $\beta_2 \neq \beta_1$  then  $y\beta_2 y\beta_1 y \neq y\beta_1 y\beta_2 y$  and, because  $h$  is elementary, we get a contradiction.

Thus  $\beta_1 = \beta_2$ .

Now Claim 6.8 follows from cases (1) through (4). ■

*Claim 6.9.* Let  $x, y \in \Theta$  and let  $\gamma_3 = \alpha_0 x \alpha_1 x \cdots \alpha_{m-1} x \alpha_m$  and  $\gamma_4 = \beta_0 y \beta_1 y \cdots \beta_{r-1} y \beta_r$  where  $m, r \geq 3$ ,  $x \notin \text{alph}(\alpha_0 \cdots \alpha_m)$  and  $y \notin \text{alph}(\beta_0 \cdots \beta_r)$ . Then  $\alpha_1 = \alpha_2 = \cdots = \alpha_{m-1}$  and  $\beta_1 = \beta_2 = \cdots = \beta_{r-1}$  and moreover if either  $\alpha_1 = \Lambda$  or  $\beta_1 = \Lambda$  then  $\#\Theta = 1$ .

*Proof of Claim 6.9.*

The first part of the conclusion follows directly by Claim 6.8. The second part follows from the fact that  $\Theta$  is strongly closed and so all occurrences of  $x$  in  $\gamma_3$  (all occurrences of  $y$  in  $\gamma_4$ ) are consecutive only if  $\#\Theta = 1$ . ■

If  $\#\Theta = 1$  then clearly  $h$  is  $(\Theta, id_{\Theta})$ -cyclic where  $id_{\Theta}$  is the identity mapping of  $\Theta$ .

Thus we will assume that  $\#\Theta > 1$ .

*Claim 6.10.* There exist a permutation  $\pi_3$  of  $\Theta$  and a permutation  $\pi_4$  of  $\Theta$  such that  $\gamma_3 = a\pi_3(a)\pi_3^2(a) \cdots \pi_3^r(a)$  and  $\gamma_4 = b\pi_4(b)\pi_4^2(b) \cdots \pi_4^s(b)$  for some  $s \geq r > 1$  and  $a, b \in \Theta$ .

*Proof of Claim 6.10.*

We will prove that  $\gamma_3$  is of the form  $a\pi_3(a) \cdots \pi_3^r(a)$ ; the proof for  $\gamma_4$  can be done analogously.

Let  $x, y \in \Theta$ .

First we note that there can be at most one occurrence of  $y$  between any two consecutive occurrences of  $x$  in  $\gamma_3$ . This follows directly from Claim 6.9, because otherwise there would be no occurrence of  $x$  between these (and hence any) two occurrences of  $y$  in  $\gamma_3$  which contradicts the fact that  $\Theta$  is strongly closed (and  $\gamma_3$  is an image by  $h$  of a word in  $\Theta^+$ ).

Similarly, Claim 6.9 implies that there is at least one occurrence of  $y$  between any two consecutive occurrences of  $x$  in  $\gamma_3$ .

Consequently, there is *precisely* one occurrence of  $y$  between any two consecutive occurrences of  $x$  in  $\gamma_3$ .

Now let  $\gamma_3 = c_1 \cdots c_t$  for some  $t \geq 3$ ,  $c_1, \dots, c_t \in \Theta$  and let  $\#\Theta = n$ . Then we have  $\gamma_3 = c_1 c_2 \cdots c_n c_1 \bar{\gamma}_3$  for some  $\bar{\gamma}_3 \in \Theta^+$  where  $c_1, c_2, \dots, c_n$  are all (occurrences of) different letters. If we set  $c_1 = a$  and  $\pi_3$  to be deter-

mined by  $\pi_3(c_i) = c_{i+1}$  for  $1 \leq i \leq n-1$  and  $\pi_3(c_n) = c_1$  then the claim holds. To see this we will show that  $\pi_3(c_j) = c_{j+1}$  for  $1 \leq j \leq t-1$ . This is certainly true for  $1 \leq j \leq n$ . Take  $j > n$  and consider the subword  $c_{j-n+2} \cdots c_{j-1} c_j c_{j+1}$  of  $\gamma_3$  (where  $\pi_3(c_{j-n+2}) = c_{j-n+3}, \dots, \pi_3(c_{j-1}) = c_j$ ). Clearly none of the letters  $c_{j-n+2}, \dots, c_{j-1}, c_j$  equals  $c_{j+1}$  (as otherwise in a subword of the length  $n-1$  the letter  $c_{j+1}$  would repeat and consequently between two occurrences of  $c_{j+1}$  we would miss a letter from  $\Theta$ ). Thus  $\{c_{j+1}\} = \Theta - \{c_j, c_{j-1}, \dots, c_{j-n+2}\}$  and consequently  $\pi_3(c_j) = c_{j+1}$ . ■

Let  $\gamma_2$  be the subword contributed by  $p$  to  $cov_2(v)$ ; then we have  $\gamma_3 = h(\gamma_2)$  and  $\gamma_4 = h(\gamma_3)$ . Since  $\Theta$  is strongly closed, for any letter  $a \in \Theta$   $h(a)$  is a subword of  $\gamma_3$  and a subword of  $\gamma_4$ . Consequently the sequence of letters in  $h(a)$  must "run" according to  $\pi_3$  and according to  $\pi_4$  (see Claim 6.10). But  $alph(h(a)) = \Theta$  and so  $\pi_3 = \pi_4$ .

Let  $\pi = \pi_3 = \pi_4$ . Then (because  $\gamma_4 = h(\gamma_3)$ ),  $h$  is  $(\Theta, \pi)$ -cyclic and so Lemma 6.12 holds. ■

Now Theorem 4 is proved as follows.

*Claim 6.11.*  $G$  is repetitive if and only if there exists a strongly closed  $\Theta \subseteq \Sigma$  and a permutation  $\pi$  of  $\Theta$  such that  $h$  is  $(\Theta, \pi)$ -cyclic.

*Proof of Claim 6.11.*

If  $G$  is repetitive then Lemma 6.12 implies that there exist a strongly closed  $\Theta \subseteq \Sigma$  and a permutation  $\pi$  of  $\Theta$  such that  $h$  is  $(\Theta, \pi)$ -cyclic.

On the other hand if there exist a strongly closed  $\Theta \subseteq \Sigma$  and a permutation  $\pi$  of  $\Theta$  such that  $h$  is  $(\Theta, \pi)$ -cyclic, then by Lemma 6.11, for each  $a \in \Theta$  and each  $n \geq 1$ ,  $(a\pi(a) \cdots \pi^{\Theta-1}(a))^n \in sub(L(G))$  and so  $L(G)$  is repeti-

tive. ■

Since (obviously) it is decidable whether or not there exists a strongly closed  $\Theta \subseteq \Sigma$  and a permutation  $\pi$  of  $\Theta$  such that  $h$  is  $(\Theta, \pi)$ -closed, Claim 6.11 implies (a) of Theorem 2.

Part (b) of Theorem 4 is seen as follows.

By Lemma 6.12, if  $G$  is repetitive then there exist a strongly closed  $\Theta \subseteq \Sigma$  and a permutation  $\pi$  of  $\Theta$  such that  $h$  is  $(\Theta, \pi)$ -cyclic. Then by Lemma 6.11, for each  $a \in \Theta$  and each  $n \in \mathbf{N}^+$   $(a\pi(a)\pi^2(a) \cdots \pi^{\#\Theta-1}(a))^n \in \text{sub}(L(G))$ . Consequently  $G$  is strongly repetitive.

Thus Theorem 4 holds. ■

#### ACKNOWLEDGEMENTS

The authors gratefully acknowledge the support of NSF grant MSC 79-03838. They are indebted to R. Verraedt for useful comments concerning the first draft of this paper.



## REFERENCES

- [BEM] D. R. Bean, A. Ehrenfeucht and G. F. McNulty, Avoidable patterns in strings of symbols, 1979, *Pacific Journal of Mathematics*, v. 85, no. 2, 261-293.
- [B] J. Berstel, Sur les mots sans carré définis par un morphisme, 1979, *Springer Lecture Notes in Computer Science*, v. 71, 16-25.
- [C] A. Cobham, Uniform tag sequences, *Mathematical Systems Theory*, 1972, v. 6, n. 2, 164-191.
- [ER1] A. Ehrenfeucht and G. Rozenberg, Simplifications of homomorphisms, *Information and Control*, 1978, v. 38, n. 3, 298-309.
- [ER2] A. Ehrenfeucht and G. Rozenberg, On the structure of polynomially bounded DOL systems, *Fundamenta Informaticae*, 1979, v. 2, 187-197.
- [ER3] A. Ehrenfeucht and G. Rozenberg, On the subword complexity of square-free DOL languages, 1981, *Theoretical Computer Science*, v. 16, 25-32.
- [H] M. Harrison, *Introduction to formal language theory*, 1978, Addison-Wesley, Reading, Massachusetts.
- [K] J. Karhumäki, On cubic-free  $\omega$ -words generated by binary morphisms, *Discrete Applied Mathematics*, to appear.
- [MH] M. Morse and G. Hedlund, Unending chess, symbolic dynamics and a problem of semigroups, 1944, *Duke Mathematical Journal*, v. 11, 1-7.
- [RS] G. Rozenberg and A. Salomaa, *The mathematical theory of L systems*, 1980, Academic-Press, London, New York.
- [S1] A. Salomaa, Morphisms on free monoids and language theory, in R. V. Book, ed., *Formal language theory, perspectives and open problems*, 1980, Academic Press, London, New York, 141-166.

- [S2] A. Salomaa, *Jewels of formal language theory*, 1981, Computer Science Press.
- [T] A. Thue, Über unendliche Zeichenreihen, 1906, *Norske Vid. Selsk. Skr., I Mat. Nat. Kl., Christiania*, v. 7, 1-22.

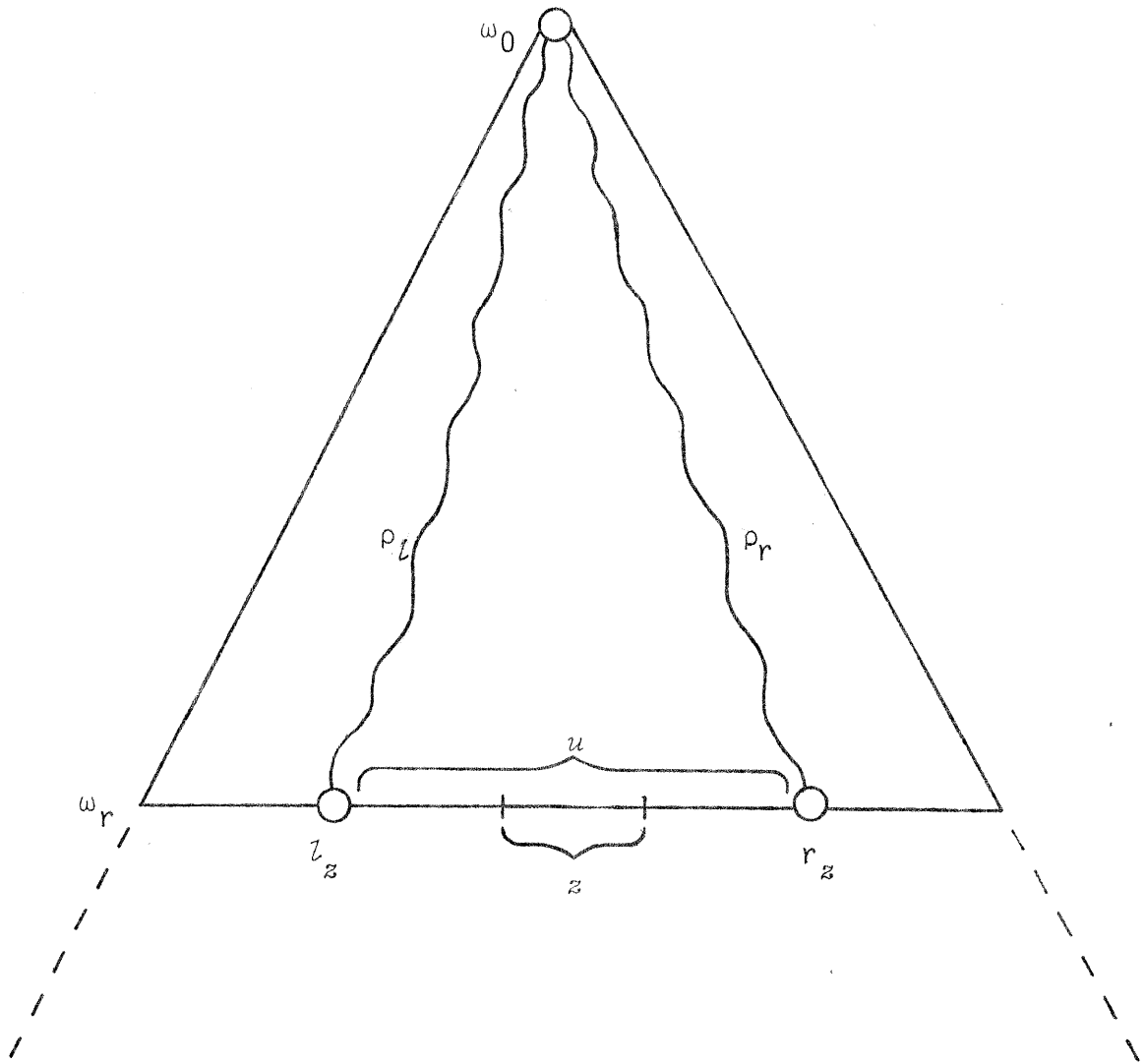


Figure 1.

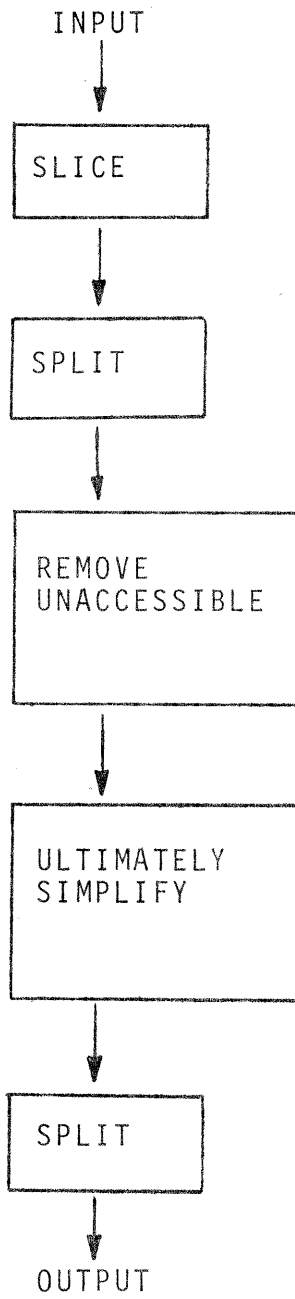


Figure 2.

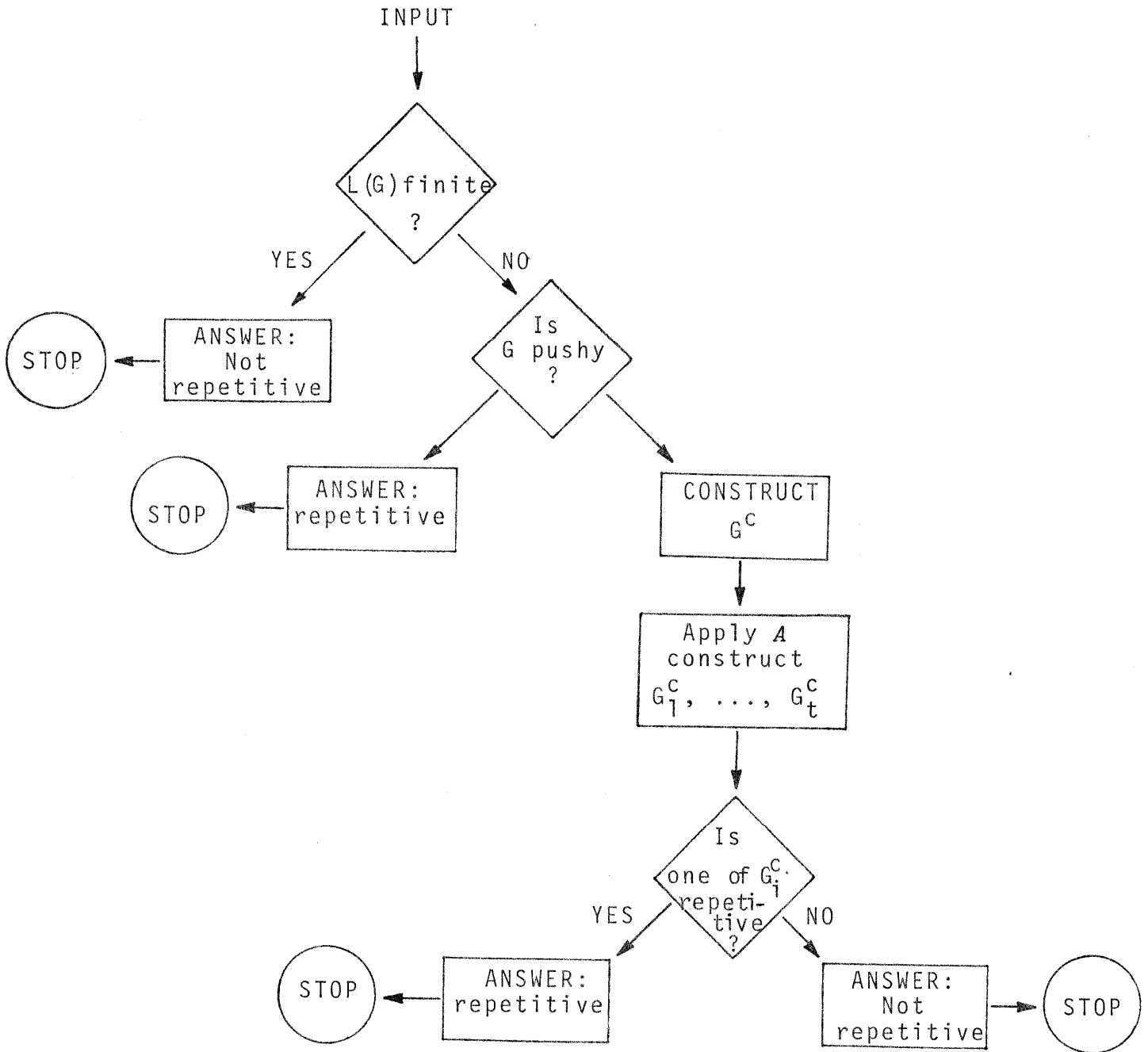


Figure 3.

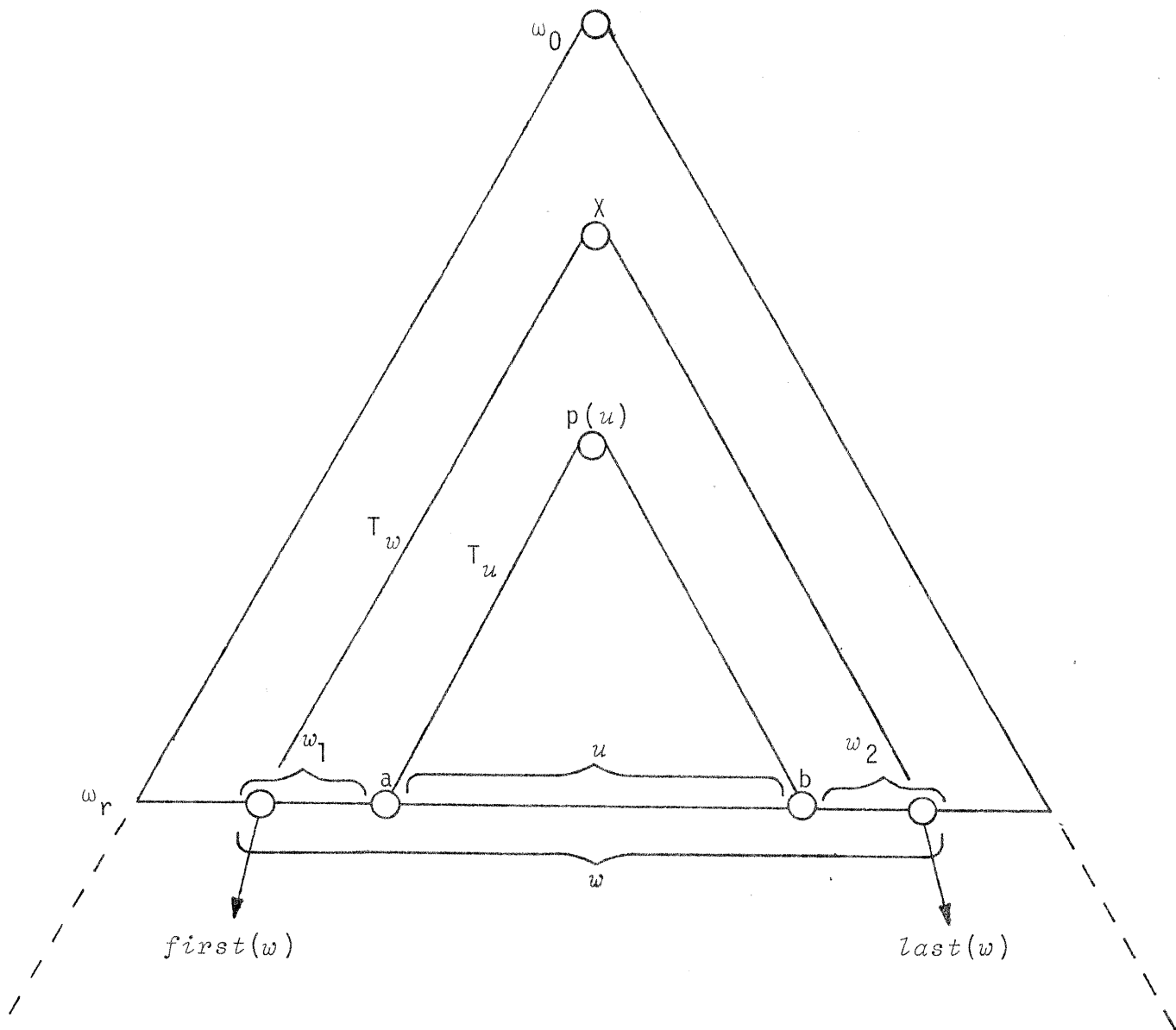


Figure 4.

$T(u)$ :

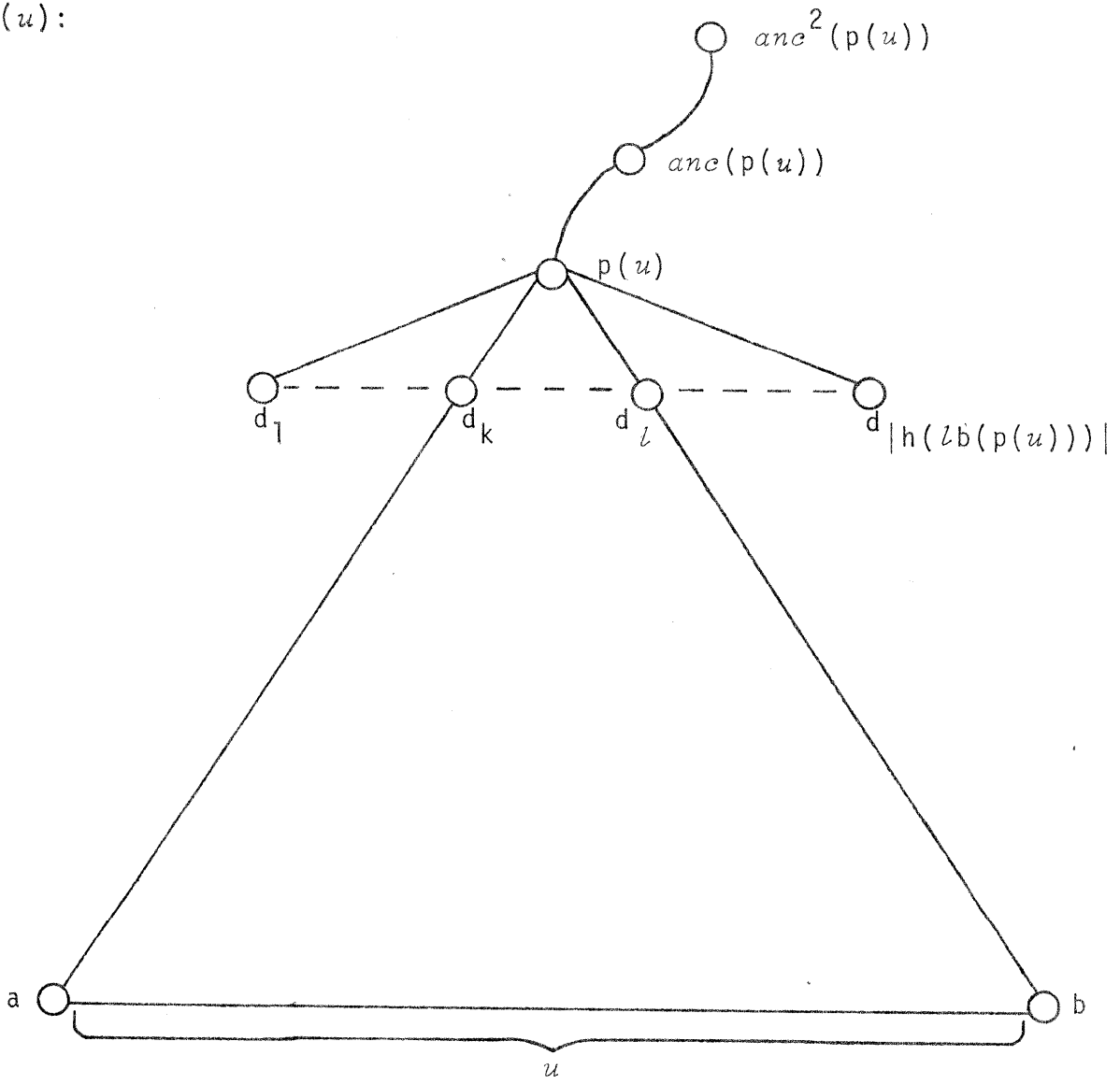


Figure 5.

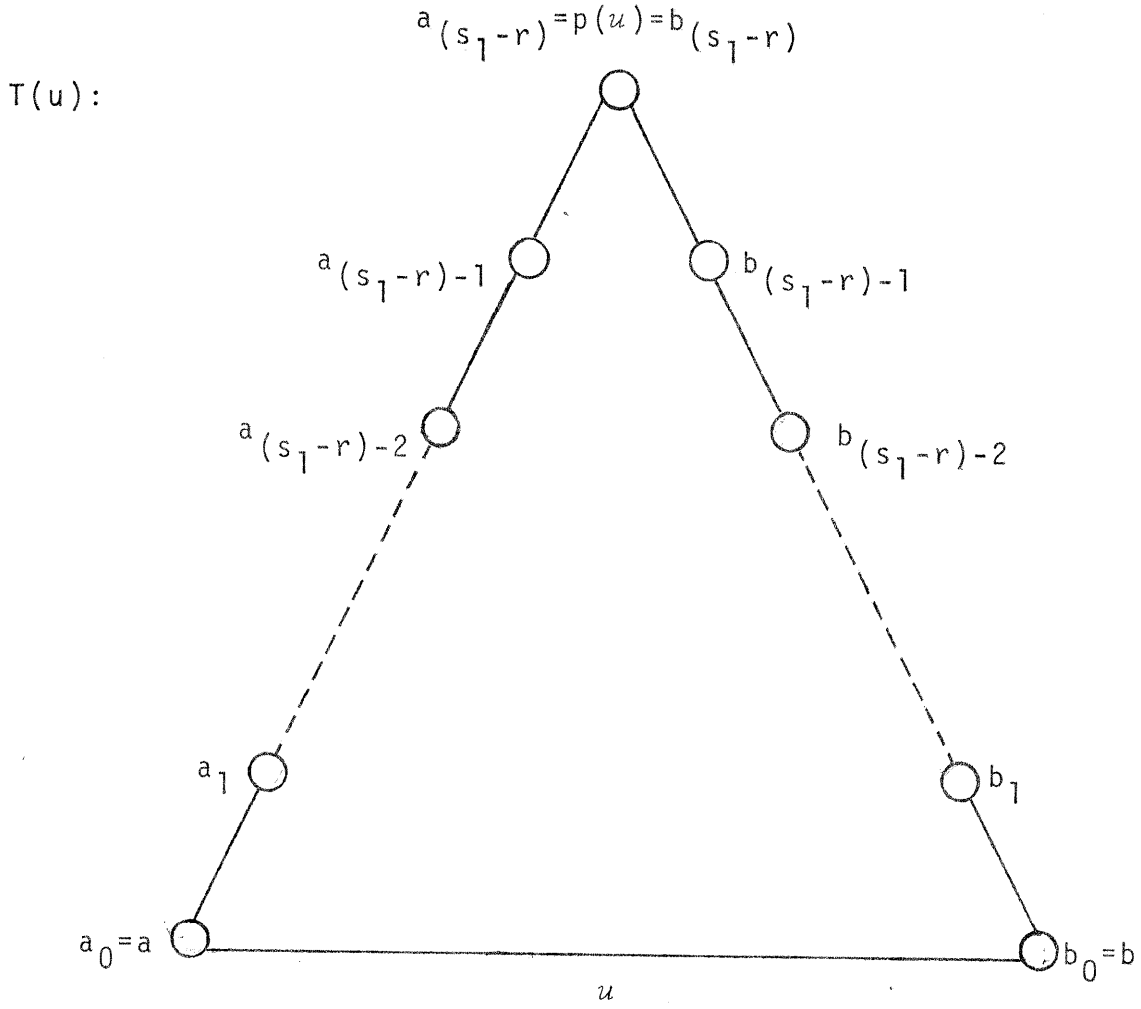


Figure 6.



T(G):

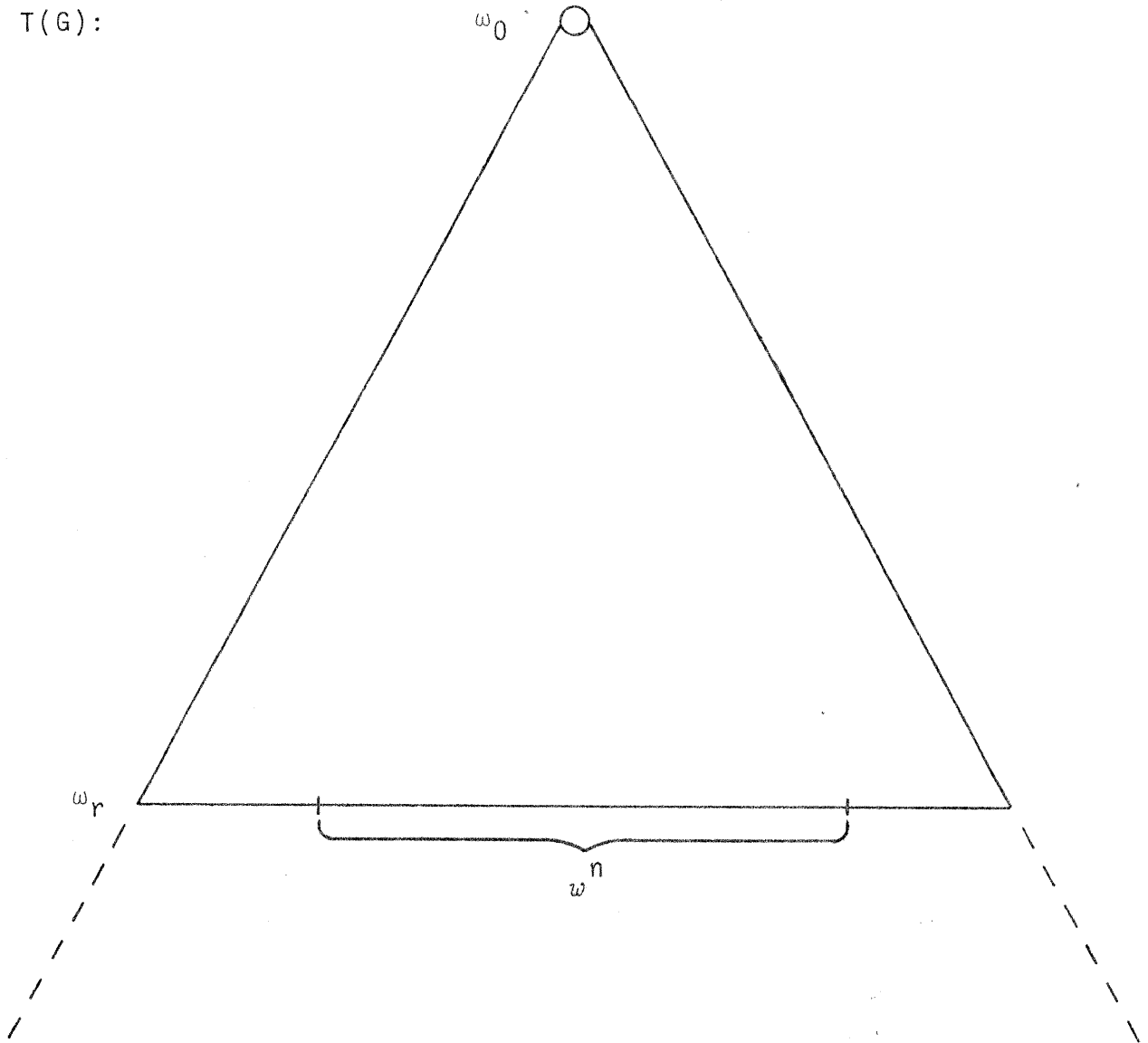


Figure 7.

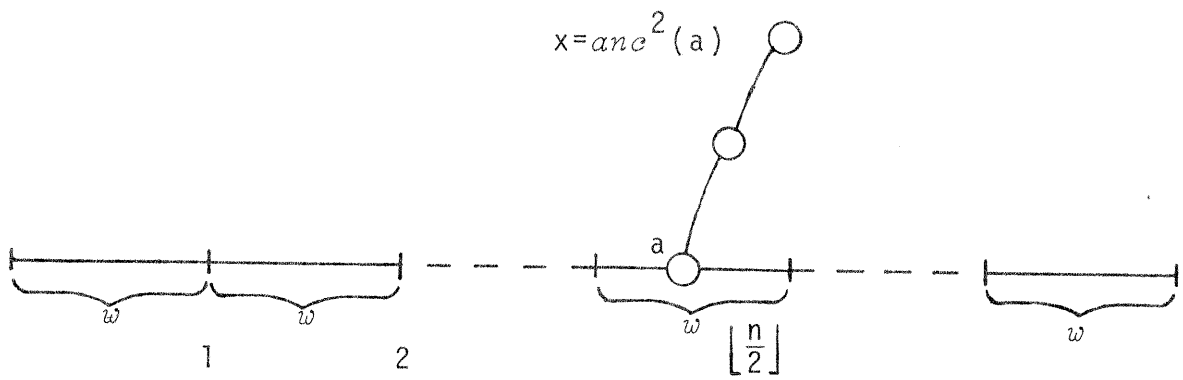


Figure 8.

$T(G):$

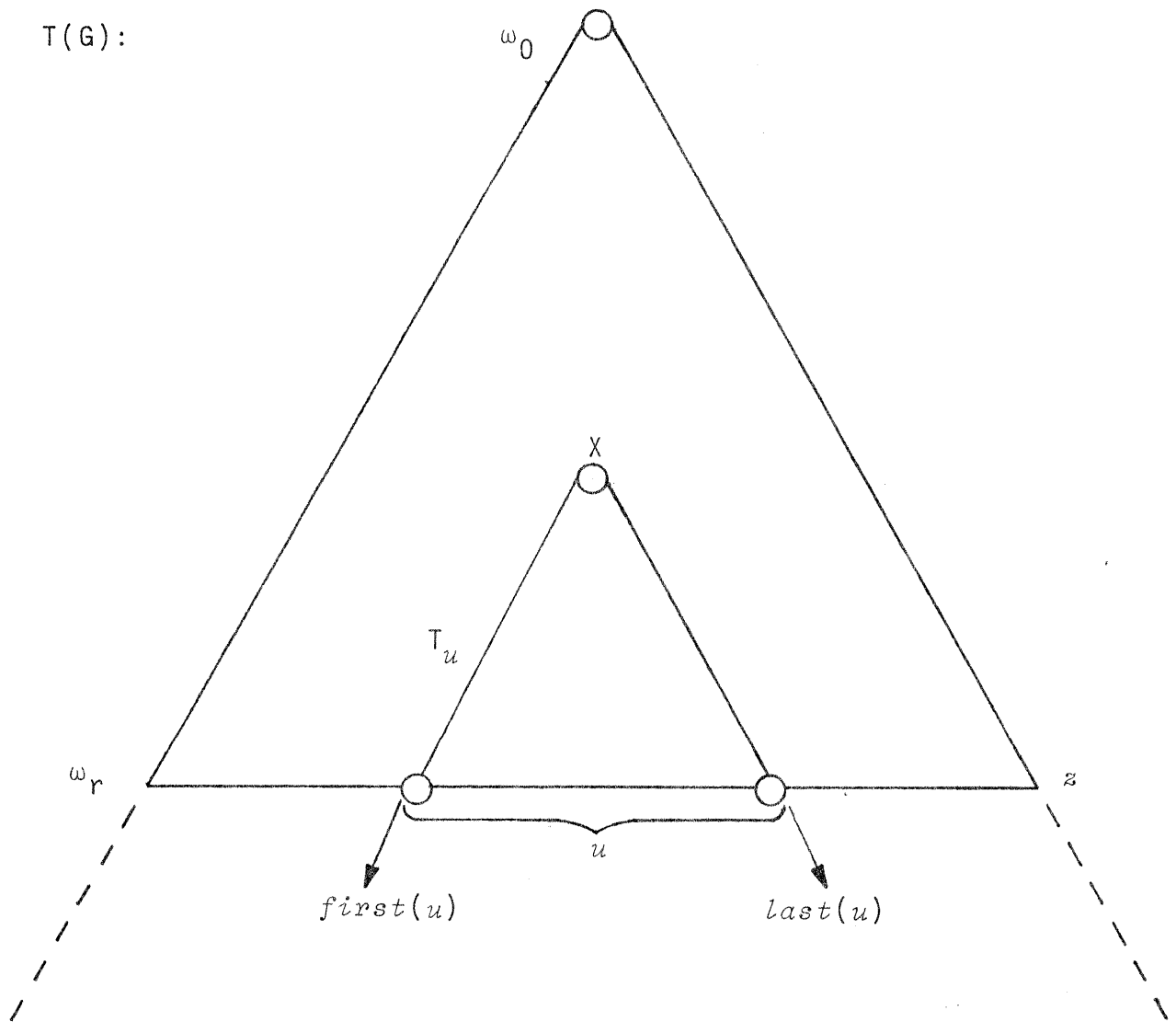


Figure 9.

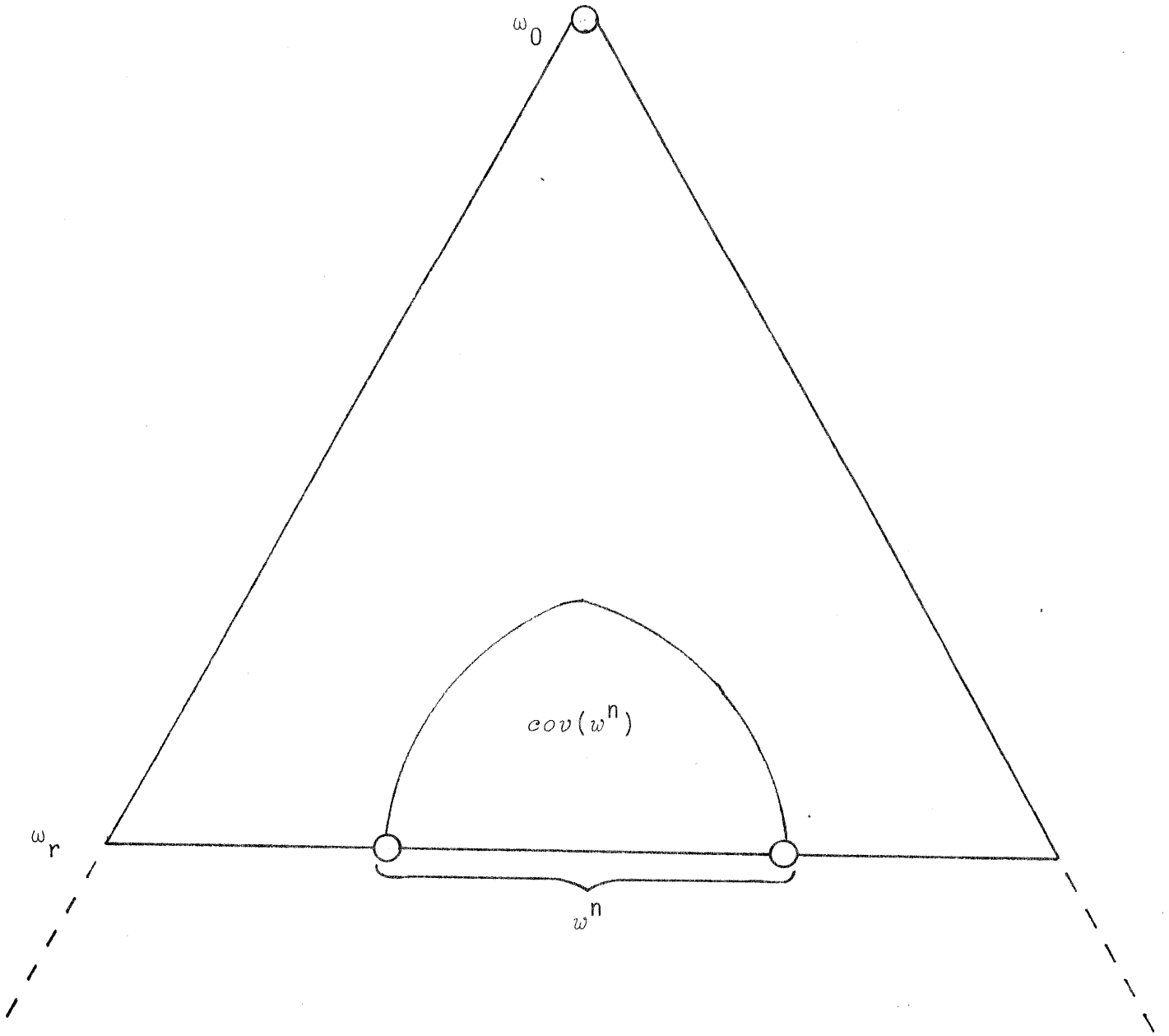


Figure 10.