THE SEQUENCE EQUIVALENCE PROBLEM FOR DOL ·SYSTEMS IS DECIDABLE *

K. ČULIK II
University of Waterloo,
Waterloo, Ont., Canada

I. FRIŠ
University of New England,
Armidale, N.S.W., Australia

Abstract

The sequence equivalence problem for DOL systems is shown to be decidable. In an algebraic formulation the sequence equivalence problem for DOL systems can be stated as follows: Given homomorphisms h_1 and h_2 on a free monoid Σ^* and a word σ from Σ^*, is $h_1^n(\sigma) = h_2^n(\sigma)$ for all $n > 0$?

0. Introduction

Given two homomorphisms h_1 and h_2 on a free monoid Σ^* and a word σ in Σ^*, can we determine whether $h_1^n(\sigma) = h_2^n(\sigma)$ for all $n > 0$? This is an algebraic formulation of the well-known DOL sequence equivalence problem. The problem originated in Lindenmayer systems which are mathematical models of cellular development. In that context it can be restated as the problem of the developmental equivalence of two genetic encodings in filamental organisms developing deterministically without interaction. Lindenmayer systems without interaction (DOL systems) were introduced in [5] and the equivalence problem for them was posed shortly afterwards in [8] .

Here we summarize [1] and [2] which together show that this problem is decidable. The aim of the present paper is to give an intuitive overview to the proof techniques used in [1] and [2]. We give only outlines of most proofs, readers interested in the technical details are referred to the two above papers. However, it may be useful to read this paper first since the informal outlines of some of the proofs will make the reading of [1] and [2] easier.

First we explain intuitively the basic ideas of our solution of the DOL equivalence
problem. Henceforth, by equivalence we mean sequence equivalence. The technical terms
which are used in the introduction but defined only later are enclosed in quotation
marks on firt use.

We start by showing that, without loss of generality, the testing for equivalence may
be restricted to "normal" systems. Then we introduce an essential auxiliary notion,
namely, the "balance" of a string w in Σ^* with respect to two homomorphisms h_1, h_2
on Σ^* (or with respect to two DOL system based on these homomorphisms). The balance
of w is the difference of the length of $h_1(w)$ and $h_2(w)$. The central intermediate
result is Theorem 4 showing that there exists a fixed bound on the balance of all
prefixes of all strings generated by two equivalent normal DOL systems. In section 3
and 4 this result is proved. First "simple" systems are considered. A normal system
is simple iff it has no "subsystem" in the sense of general algebra. If a system has
a subsystem, then the underlying set of the subsystem is called a "subalphabet".

For simple systems it is shown first, using the methods of linear algebra, that the
balance of every "long" prefix is "very small" compared to its length. Then, using
combinatorial methods, this result is strengthened to show that the balance of two
equivalent systems is bounded if at least one of the systems is simple.

For two equivalent systems neither of which is simple we find a common subalphabet and
show that either

1) all substrings of the language generated by the systems which are entirely in this
subalphabet are "short" (such a subalphabet is called "limited") or

2) that every two systems forming a pair "induced" by this subalphabet are equivalent.
Another pair of normal systems is obtained by "removing" the subsystem (i.e. by
omitting the symbols from the common subalphabet). As before, these "remainder"-systems
are equivalent because the original systems are equivalent. Since both the subsystem
and the remainder system are systems over a smaller alphabet we can use the bounded-
ness as an induction hypothesis. The base of the induction deals (essentially) with
systems over one letter so the claim is easy to verify. This allows us to assume that
the remainder-pair and (in the case of a subalphabet which is not limited) also every
induced-pair have bounded balance. As the case of limited subalphabets causes no
problem, this allows us to construct a bound on the balance for the original pair.

One of the more important technical details which has been omitted above is the
following. In every step of the induction we have to consider the non-propagating
systems and another singular case separately. Since a propagating system may have a
non-propagating remainder system we cannot include the propagating property into the
requirements for normality.

In the last section we design a decision procedure which works for every family of
DOL systems for which equivalence implies bounded balance. However, we have shown that

we need, without loss of generality, only to give the decision procedure for normal systems. Therefore it is sufficient that equivalence of normal systems implies bounded balance, this in fact we show in Section 5.

The normal systems are used for technical rather than essential reasons and we conjecture that every pair of equivalent DOL systems has bounded balance. It also seems feasible to sharpen the proof of Theorem 3 so that it would give a bound on the balance of a pair of DOL systems rather then to show only that such bound exists. This will then allow the complexity of the whole procedure to be determined.

1. Notation

Given an alphabet Σ, Σ^* denotes the free monoid generated by Σ, with unit (empty string) ε. A *DOL system* is a 3-tuple $G = (\Sigma, h, \sigma)$ consisting of *alphabet* Σ, *homomorphism* h and a *starting string* $\sigma \in \Sigma^*$. $L(G)$, the language generated by G, is defined as $\{h^n(\sigma) : n \geq 0\}$. G is said to be *reduced*, if every symbol from Σ occurs in at least one $h^n(\sigma)$, $n \geq 0$. To *reduce* G means to omit from Σ all the symbols which do not have this property.

For $w \in \Sigma^*$ and $a \in \Sigma$, $\#_a w$ denotes the number of occurrences of a in w. If $(a_1, \ldots, a_n)$ is an ordering of Σ, then $(\#_{a_1} w, \ldots, \#_{a_n} w)$ is called the *Parikh vector* of w and is denoted by $[w]$. The matrix $M = (m_{ij})_{\substack{1 \leq i \leq n \\ 1 \leq j \leq n}}$, where $m_{ij} = \#_{a_j} h(a_i)$ is called the *growth matrix* for G.

If i is a number, $|i|$ denotes the absolute value of i; if w is a string $|w|$ denotes the length of w; later on $|A|$ is also used for length of a vector A or maximum characteristic value of a matrix A. For $w \in \Sigma^*$, let $\min(w) = \{a : \#_a w > 0\}$. Given $G = (\Sigma, h, \sigma)$, we say that w is a G-prefix (G-substring, G-suffix) if w is a prefix (substring, suffix) of $h^n(\sigma)$ for some $n \geq 0$. Two DOL systems $G_i = (\Sigma, h_i, \sigma_i)$, $i = 1,2$ are called *(sequence) equivalent* if $h_1^n(\sigma_1) = h_2^n(\sigma_2)$ for all $n = 0, 1, \ldots$. Two DOL-systems G_1, G_2 are called *Parikh equivalent* if $[h_1^n(\sigma_1)] = [h_2^n(\sigma_2)]$ for all $n = 0, 1, \ldots$. The *balance* (with respect to G_1, G_2) of a string w in Σ^* is defined by $\beta(w) = ||h_1(w)| - |h_2(w)||$. If there exists $c \geq 0$ so that $\beta(x) \leq c$ for all G_1-prefixes x, then the pair (G_1, G_2) is said to have *bounded balance*. In this case the smallest such c is called the *balance* of the pair (G_1, G_2).

For two sets $A, B, A \cup B$ denotes their union. If A, B are disjoint, we stress this by writing $A + B$ for the union. Finally, we will often write a instead of $\{a\}$ for a one-element set.

2. The Normal Systems

Let $G = (\Sigma, h, \sigma)$ be a DOL system. We define the function m: $P(\Sigma) \to P(\Sigma)$, where $P(\Sigma)$ is the set of all subsets of Σ by putting

$m(\phi) = \phi$,

$m(\{a\}) = \min(h(a))$ for $a \in \Sigma$,

$m(A \cup B) = m(A) \cup m(B)$.

It is easy to see that $m^i(a) = \min(h^i(a))$ for all $i \geq 1$. We will write $m(a)$ for $m(\{a\})$ and use m_1, m_2, m_{12} etc. to denote similar functions based on h_1, h_2, $h_1 h_2$, etc.

Definition 1 A DOL system $G = (\Sigma, h, \sigma)$ is called an lr-system if $\Sigma = \Sigma_l + \Sigma_c + \Sigma_r$ is a decomposition of Σ into three non-empty disjoint sets such that $h(a) \in \Sigma_l \Sigma_c^*$ for $a \in \Sigma_l$, $h(a) \in \Sigma_c^*$ for $a \in \Sigma_c$, $h(a) \in \Sigma_c^* \Sigma_r$ for $a \in \Sigma_r$, and $\sigma \in \Sigma_l \Sigma_c^* \Sigma_r$. We call Σ_c the *core* of Σ, Σ_l is called the *left side*, Σ_r the *right side* of Σ. The number of symbols in the core Σ_c of Σ is called the *order* of G.

Definition 2 A DOL system $G = (\Sigma, h, \sigma)$ is called *normal* if

(1)   G is an lr-system,

(2)   G is reduced,

(3)   if $a \in m^j(b)$ for some $j > 0$ then $a \in m(b)$,

holds for every $a, b \in \Sigma_c$.

Now, we show that, without loss of generality, the testing for equivalence can be restricted to the normal systems.

Theorem 1 The sequence equivalence problem is decidable for DOL systems iff it is decidable for normal systems.

Proof Outline (for details see [2] Lemma 1,2 and Theorem 1). Given a DOL system $G = (\Sigma, h, \sigma)$, let $G^{j,k} = (\Sigma, h^j, h^k(\sigma))$ be the (j,k) speed-up of G. It is easy to see that two DOL systems G_1 and G_2 are equivalent iff all the j pairs of DOL systems $G_1^{j,k}$, $G_2^{j,k}$ for $k = 0, \ldots, j-1$ are equivalent.

Given any pair $G_i = (\Sigma, h_i, \sigma)$ i = 1,2 of DOL systems we can effectively find j so that the speed-up systems $G_i^{j,k}$ are normal for i = 1,2 and $k = 0, \ldots, \ldots, j - 1$. We first find h_1, h_2 for which $h_1^{k_1}$ and $h_2^{k_2}$ meet condition (3) of normality. Then we choose j as the least common multiple of h_1 and h_2.

Finally, we reduce the systems $G_i^{j,k}$ for i = 1,2 and $k = 0, \ldots, j - 1$, and if the systems are not lr-systems we create the sides "artificially".

□

3. The lr-Simple Systems

Definition 3 Let $G = \langle \Sigma_l + \Sigma_c + \Sigma_r, h, \sigma \rangle$ be an lr-system. A homomorphism h is called
lr-simple if for every $a, b \in \Sigma_c$ and every $k > 0$ there is $j > 0$ such that $a \in m^{kj}(b)$.
Equivalently, calling h *lr-irreducible* if for every $a, b \in \Sigma_c$ there is $j > 0$ such that
$a \in m^j(b)$, h is lr-simple iff h^k is lr-irreducible for all $k \geq 1$. We call G lr-simple
if h is lr-simple.

If G is lr-simple and normal, then from $a \in m^{kj}(b)$ we get $a \in m(b)$.

Putting $a = b$ we get $a \in m(a)$, which implies in turn that $a \in m^k(b)$ for all $i \geq 1$.
Thus if G is normal, G is lr-simple iff $m(b) = \Sigma_c$ for all $b \in \Sigma_c$. However, the follo-
wing lemma is needed for systems not necessarily normal.

Lemma 1 Let $G_i = (\Sigma, h_i, \sigma)$, $i = 1, 2$ be two Pariskh equivalent DOL systems, G_1 lr-simple
and of order at least two ($|\Sigma| \geq 2$). Then for every $\varepsilon > 0$ there is $n_0 > 0$ such that
for every $w \in \Sigma^*$, $w \notin (\Sigma_l \cup \Sigma_r)^*$, $\beta(h_1^n(w)) \leq \varepsilon |h_1^n(w)|$ for all $n > n_0$.

Proof Outline (for details see [2], Lemma 3). By [7], for $i = 1, 2$, the Parikh vector
v_i^k of $h_i^k(\sigma)$ can be written as $v_i = x M_i^k$ where x is the Parikh vector of w and M_i is
the growth matrix of M_i. Therefore the computation of $v_i^1, v_i^2, \ldots$ constitutes the power
method for computing the characteristic u_i vector of M_i corresponding to the largest
eigenvalue of M_i. From the observation that $\beta(u_1) = \beta(u_2) = 0$ then easily follows that

$$\lim_{n \to \infty} \frac{\beta(h_1^n(w))}{|h_1^n(w)|} = 0 .$$

□

Theorem 2 Let $G_i = (\Sigma, h_i, \sigma)$, $i = 1, 2$, be two Parikh equivalent DOL systems, G_1 l r-
simple and $|\Sigma| \geq 2$. Then for every $\varepsilon > 0$ there is $K > 0$ such that for every G_1-prefix
w, $|w| > K$ we have $\beta(w) \leq \varepsilon |w|$.

Proof Outline (for details see [2], Lemma 4). Every G_1-prefix w can be written as
$w = h_1^{n_0}(u_1)x$ for a "large" integer n_0 and "short" string $x \in \Sigma^*$. We have
$\beta(w) \leq \beta(h_1^{n_0}(u_1)) + \beta(x)$. Given $\varepsilon > 0$ we have $\beta(h_1^{n_0}(u_1)) \leq \varepsilon/2 |u_1| \leq \varepsilon/2 |w|$ by Lemma 1
and $\beta(x) \leq \varepsilon/2 |w|$ since x is snort.

We have shown that the balance of "long" strings is small with respect to their
length. However, our goal is to show a stronger result, namely, that the balance is
bounded. Before we can show this we will need a few auxiliary notions.

The notion of a *derivation forest* of a string w with respect to a DOL system G is
obtained by the obvious modification of the well-known notion of a derivation tree
for a context-free grammar. The difference is that in DOL systems we have an axiom
(string of symbols) rather than a single starting symbol of context-free grammars.

Let F be the derivation forest of $h^n(\sigma)$ for some $n > 0$ with respect to DOL system $G = (\Sigma, h, \sigma)$. A path in F from a node on the lowest level (of σ) to a node α on level n is called *the chain of* α. Formally, a chain q is a string in $(\Sigma \times N)^*$ such that if $q = (a_0, k_0) \ldots (a_n, k_n)$, then

(i) $1 \leq k_0 \leq |\sigma|$;

(ii) $1 \leq k_i \leq |h(a_{i-1})|$ for $1 \leq i \leq n$;

(iii) a_{i+1} is the k_{i+1}-th symbol in $h(a_i)$ for $0 \leq i \leq n-1$.

I.e., k_0 is the position of a_0 in σ, and for $i \geq 1$, k_i determines which of possibly several branches is taken. First components of pairs are clearly redundant but they allow us to state easily the condition (iii).

The string $a_0 \ldots a_n$ is called *the trace of chain q*. A chain q is said to be *periodic with period p and prefix (initial segment)* q_1 if $q = q_1 p^m q_2$ for some $m \geq 2$ and q_2 is a prefix of p. Chain q is *leftmost (rightmost)* on level i if $k_i = 1 (k_i = |h(a_{i-1})|)$. Chain q is *fully leftmost (fully rightmost)* if it is leftmost (rightmost) on all levels. For a node α of derivation forest F and a specific occurrence of G-substring w, we say that w *contains* α if α is one of the nodes labeled by a symbol from w.

Let q be a periodic chain with prefix r and period p. Then there are cyclically repeating (after each $|p|$ steps) common G-substrings on at least one side of q (see [4], Theorem 11.3 and 11.4).

<u>Theorem 3</u> Let $G_i = (\Sigma, h_i, \sigma)$ for $i = 1,2$ be two equivalent *lr*-systems. Let G_1 be *lr*-simple and of order at least two ($|\Sigma| \geq 2$). Then the pair (G_1, G_2) has bounded balance.

<u>Proof Outline</u> (for details see [1], Lemma 3.1, Theorem 3.2 and [2], Theorem 2. Assume that the pair (G_1, G_2) does not have bounded balance (Assumption 1). Therefore, for every $n_0 > 0$ there must exist n, $n \geq n_0$, and u,v in Σ^* so that $h_1^n(\sigma) = uv$ and the following conditions hold:

(i) $\beta(u) > \beta(w_1)$ for any prefix w_1 of $h_1^j(\sigma)$ where $0 \leq j < n$.

(ii) $\beta(u) \geq \beta(w_2)$ for any prefix w_2 of $h_1^n(\sigma)$.

(iii) $\beta(u) > \beta(w_3)$ for any prefix w_3 of u.

Let F be the derivation forest of G_1 and α be the node in F at the last symbol of prefix u at level n. Let q be the chain of α in F and let α_1 and α_2 be the first two nodes of chain q (from the top) such that the label of α_1 is the same as the label of α_2, the label of the left neighbor of α_1 is the same as that of the left neighbor of α_2 and the same holds for the right neighbors. Let these common labels be a,b,c from left to right; they, of course, are not necessarily different.

Let the levels of α_1 and α_2 in the derivation forest of G_1 (from top) be r and r+t, respectively. Clearly, since G_1 is *lr*-simple, there exists a constant C independent

of n_0 so that $r+t \leq C$. Note that we have the constant $C_1 = |\Sigma|+1$ with the property that on levels higher than C_1 each node of any chain has at least one neighbor both on the left and on the right. This is so since otherwise q would have a fully leftmost (rightmost) initial segment with some symbol occurring at least twice in its trace; therefore, u would be a prefix (v would be a suffix) of $h_1^j(\sigma)$ for some $j < n$, which would be in contradiction with condition (i) implied by Assumption 1.

Let $q = q_1 q_2 q_3$ where q_2 is the section of q between nodes α_1 and α_2. Let q' be the periodical chain defined by $q' = q_1 q_2^j q_4$ where $j > 0$ and q_4 is a proper prefix of q_2, and j and q_4 chosen so that the length of q' is the same as the length of q. Informally, we have chosen q' so that it coincides with q up to the second occurrence of abc and then continues periodically. Therefore, there are cyclically repeating longer and longer substrings on both sides of q'. Specifically $h_1^{n-t-r}(abc)$ is a common substring of $h_1^{n-t}(\sigma)$ and $h_1^n(\sigma)$ which on level n contains node α since chain q goes through node α_2. Moreover, node α is not close to either end of the common substring since α_2 is labeled by the middle symbol b in abc and both $|h_1^m(a)|$ and $|h_1^m(c)|$ are exponentially growing (for growing m) by Lemma 3.1.

Now, let $h_1^{n-t}(\sigma) = u_1 xy v_1$ and $h_2^n(\sigma) = u_2 xy v_2$ where $xy = h_1^{n-r-t}(abc)$, $u_2 x = u$ and $yv_2 = v$, i.e. the node α on level n is at the last symbol of x. Let $u_1 x = u'$. So we have the situation illustrated in Fig. 1.
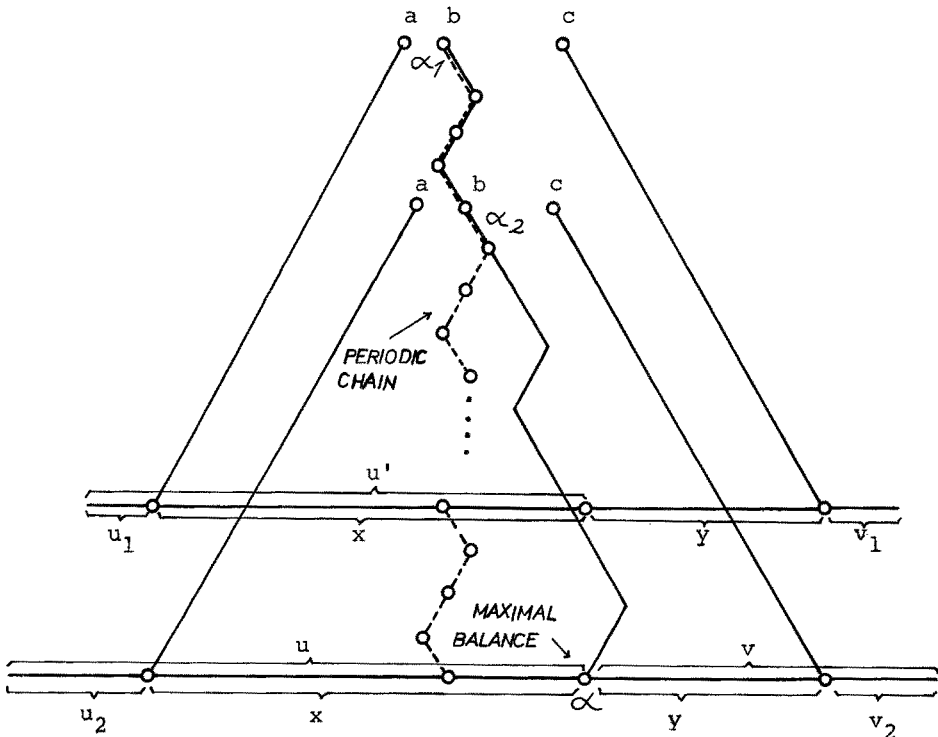

Fig.1

As the growth of all symbols in a simple system is exponential of the same power, it it easy to see that the length of both x and y is linearly proportional to the length of the whole string $h_1^n(\sigma)$. Therefore it follows, using Theorem 2, that for each $\epsilon > 0$ there exists n_0 so that $\beta(u) \leq \epsilon|x|$ and $\beta(u) \leq \epsilon|y|$ where u,x, and y are determined by n_0.

Now, both $h_1^n(\sigma)$ and $h_1^{n-t}(\sigma)$ have xy as a substring with node $\propto$ at the last symbol preceding y on level n. Since the two systems are equivalent both $h_1(y)$ and $h_2(y)$ are substrings of $h_1^{n+1}(\sigma)$ and of $h_1^{n-t+1}(\sigma)$. We recall that both $\beta(u')$ and $\beta(u)$ are "very small" with respect to $h_1(y)$. Assumption 1 gives $\beta(u') < \beta(u)$, therefore the relative position of $h_1(y)$ and $h_2(y)$ as substrings of $h_1^{n-1}(\sigma)$ differs by a "small" shift w (see Fig. 2) (small with respect to the length of $h_1^{n+1}(\sigma)$ and also of $h_1(y)$), however, the relative position of corresponding strings considered as substrings of $h^{n-t+1}(\sigma)$ differs by even smaller shift w'. Consider now the "shifts" z and z' of the longer substrings $h_1(xy)$ and $h_2(xy)$ on levels n+1 and n-t+1. Clearly, $|z| - |z'| = |w| - |w'|$ and thus also $|z| > |z'|$. Similar results are obtained also where the "shifts" are in opposite directions.
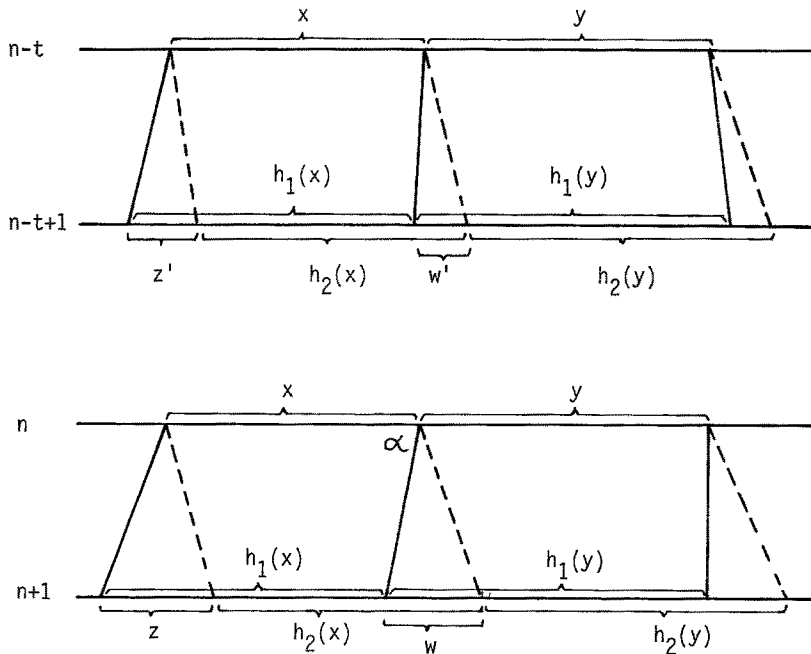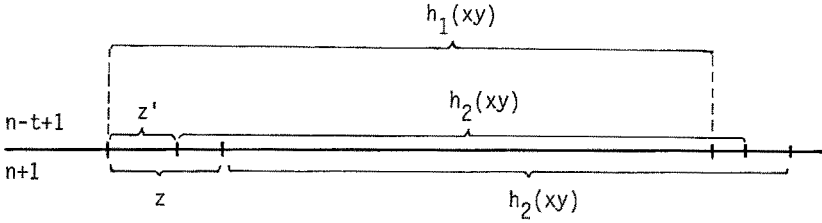

Fig.2

Fig.3

So if we compare levels n-t+1 and n+1, as shown on Fig. 3., by matching the common substring $h_1(xy)$ we see that $h_2(xy)w_1 = w_2h_2(xy)$ for some "very short" strings w_1 and w_2. Therefore the string $h_2(xy)$ is periodical with period much shorter than both x and y. Therefore there exists r such that r is a suffix of x and prefix of y as illustrated on Fig. 4.
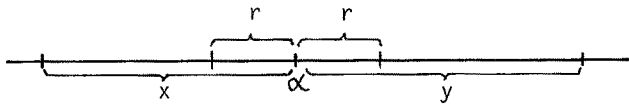
Fig.4

So far we have used only the fact $\beta(u)$ is strictly maximal up to level n (condition (i)), but not yet the conditions (ii) and (iii) implied by Assumption 1. Now we will exploit them by considering two cases:

Case I. Let $\beta(r) = 0$. A suffix of x is also a suffix of u so u = $\bar{w}r$ for some $\bar{w}$ in Σ^* and therefore $\beta(u) = \beta(\bar{w})$ which is in contradiction with condition (iii) implied by Assumption 1.

Case II. Let $\beta(r) > 0$. Since r is also a prefix of y we can write $h_1^n(\sigma) = \bar{u}r^2\bar{v}$ for

some $\bar{u},\bar{v}$ in Σ^* such that $\bar{u}r = u$ and $r\bar{v} = v$. Since $\beta(r) > 0$, clearly, either $\beta(\bar{u}) > \beta(u)$ or $\beta(\bar{u}r^2) > \beta(u)$ which is in contradiction with condition (ii) implied by Assumption 1.

4. Subalphabets, remainder systems and induced systems.

Given a DOL system $G = (\Sigma,h;\sigma)$, a nonempty set Π properly contained in Σ_c is called a *subalphabet* if $h(a) \in \Pi^*$ for each $a \in \Pi$. Let $\Omega = \Sigma - \Pi$. If G is an lr-system we will also use Ω_c for Σ_c-Π. For every $z \in \Sigma^*$ we denote by z^Ω the string z with all symbols from Π omitted, thus $z^\Omega \in \Omega^*$. We define G^Ω as $<\Omega,h^\Omega,\sigma^\Omega>$ where $h^\Omega(x) = (h(x))^\Omega$ for $x \in \Omega$. If for a sequence $s = s_1,s_2,\ldots$ we write $s^\Omega = s_1^\Omega,s_2^\Omega,\ldots$, then obviously

$$(4) \quad (s(G))^\Omega = S(G^\Omega)$$

where $s(G)$ is the sequence generated by DOL system G.

Given two DOL systems G_1,G_2,Π is called their *common* subalphabet if Π is a subalphabet of G_i for $i = 1,2$. From (4) we get immediately that if G_1,G_2 are equivalent and have a common subalphabet Π then also the remainder systems G_1^Ω and G_2^Ω are equivalent. It is also obvious that if G is normal, so is G^Ω.

<u>Lemma 2</u> Let $G_i = <\Sigma,h_i,\sigma>$, $i = 1,2$ be two normal propagating equivalent DOL systems. Then G_1 and G_2 have a common subalphabet Π, or the composite homomorphism h_1h_2 is lr-simple.

<u>Proof Outline</u> (for details see [2], Lemma 5). Assume that G_1 and G_2 have no common subalphabet. First we show that all symbols in Σ_c occur in infinitely many strings of $L(G_1)$. Let Γ be the set of symbols which occur only finitely many times. Since G_1 and G_2 are equivalent clearly either $\Sigma - \Gamma$ is a subalphabet or Γ is empty.

Now, write $a \underset{i}{\rightarrow} b$ for $a,b \in \Sigma_c$ and $i = 1,2$ if $b \in m_i(a)$, i.e. a directly derives b in G_i. Let $\Rightarrow = \underset{1}{\rightarrow} \cup \underset{2}{\rightarrow}$ and $\Rightarrow^*$ be the transitive and reflexive closure of $\Rightarrow$. Finally, let $\tilde{m}(a) = \{b \in \Sigma_c : a \Rightarrow^* b\}$. Obviously, $m_i(\tilde{m}(a)) \subsetneq \tilde{m}(a)$ for $i = 1,2$, so $\tilde{m}(a) = \Sigma_c$ since otherwise $\tilde{m}(a)$ would be a subalphabet. This means that $a \Rightarrow^* b$ for any $a,b \in \Sigma_c$. In more detail we can write $a \underset{i_1}{\rightarrow} c_1 \underset{i_2}{\rightarrow} c_2 \underset{i_3}{\rightarrow} \cdots \cdots \underset{i_{n-1}}{\rightarrow} c_n \underset{i_n}{\rightarrow} b$, for $n \geq 0$, $i_j \in \{1,2\}$ and $c_i \in \Sigma_c$. Because G_1 and G_2 are propagating and have no common subalphabet we can choose $i_1 = 1$ and $i_n = 2$. From the normality it also follows that we can "abbreviate" several successive transfers by either $\underset{1}{\rightarrow}$ or $\underset{2}{\rightarrow}$ into one such transfer, so we get $a \underset{1}{\rightarrow} d_1 \underset{2}{\rightarrow} d_2 \underset{1}{\rightarrow} \cdots \underset{1}{\rightarrow} d_m \underset{2}{\rightarrow} b$ for some $0 \leq m \leq n$ and $d_i \in \Sigma_c$. Therefore $a \underset{1,2}{\longrightarrow} d_2 \underset{1,2}{\longrightarrow} d_4 \cdots \underset{1,2}{\longrightarrow} b$ where $x \underset{1,2}{\longrightarrow} y$ means that $y \in m_{1,2}(x)$.

We have shown that for any $a,b \in \Sigma_c$ we have $a \underset{1,2}{\longrightarrow}^* b$. That means that the growth matrix $M_{1,2}$ associated with $h_{1,2}$ is in terminology of [4] irreducible. Now, we need only to show a stronger property, namely that a system is lr-simple if its growth matrix restricted to Σ_c is primitive (see [4]). From results in [4] it follows that

if the growth matrix is not primitive then there exist q > 1 and a partition P of Σ_c with q classes such that for every a,b $\in \Sigma_c$, if a $\in m_{1,2}^q$(b), then a and b belong to the same class of P. We use again the normality of G_1 and G_2 to show that if b $\Rightarrow$ a then a and b must belong to the same class of P. Since a $\Rightarrow^*$ b for any a,b in Σ_c there cannot exist a partition with at least two classes and that completes the proof that h_1h_2 is $\mathit{l}r$-simple.

□

<u>Definition 4</u> Given G = <Σ h,σ>. A subalphabet $\Pi \subsetneq \Sigma$ is called *limited* if there is a constant k such that for every substring u $\in \Pi^*$ of L(G) we have $|u| < k$. Note that Π is limited with respect to every DOL system equivalent to G.

<u>Lemma 3</u> Let G_1,G_2 be two equivalent systems, with a common subalphabet Π. If Π is limited and if the pair (G_1^Ω, G_2^Ω) has a bounded balance, then the pair (G_1,G_2) has bounded balance.

<u>Proof</u> Let the balance of (G_1^Ω, G_2^Ω) be c and let k be such that $|u| \leq k$ for all G_1-sub-strings u from Π^*. Then the balance of the pair (G_1,G_2) is clearly smaller or equal to (c+1)k+c. □

<u>Definition 5</u> Let G_1,G_2 be a pair of DOL systems, $G_i = (\Sigma, h_i, \sigma)$. Given $\iota = i_1 i_2 \ldots i_n$ with n $\geq$ 1 and $i_1, \ldots, i_n \in \{1,2\}$, the set S = $\{G_1^j, G_2^j : 0 \leq j \leq n\}$ of pairs of DOL systems is called ι-*combination* of (G_1,G_2) where $G_i^j = (\Sigma, \bar{h}_i, \sigma_{ij})$, for i = 1,2, j = 0, ...,n, $\bar{h}_1 = h_1 h_{i_1} h_{i_2} \ldots h_{i_n}$, $\bar{h}_2 = h_2 h_{i_1} h_{i_2} \ldots h_{i_n}$, $\sigma_{i,0} = \sigma$ and $\sigma_{i,j} = h_{ij} \ldots h_{i_n}(\sigma)$ for i =1,2 and j = 1,...,n. Finally, we reduce each system G_i^j, if necessary.

Instead of 1-combination we will say just combination. If $\iota = (21)^k$ for the minimal k $\geq$ 0 such that each G_i^j is normal we call the ι-combination the *normal combination* of (G_1,G_2). We show that for normal systems G_1,G_2 such k always exists. We find k so that $m_{21}^k(a) = m_{21}^{ks}(a)$ for all a$\in \Sigma$ and s = 1,2,... (see [2], Lemma 2.). Therefore also $m_i(m_{21}^k(a)) = m_i(m_{21}^{ks}(a))$ for i = 1, 2, and s = 1,2, ... Now, to show that the homomorphisms of the normal combination satisfy condition (3) of normality we note that

$$m_2^{k_1} m_1^{k_2} \ldots m_2^{k_{2n-1}} m_1^{k_{2n}}(a) = m_{21}^n(a)$$

for each a$\in \Sigma$, n $\geq$ 1 and arbitrary $k_1, \ldots, k_{2n} \geq 1$; since, because of normality of G_1 and G_2, the repetitions of the same homomorphisms are irrelevant. Specifically,

$$[m_2 m_{21}^k]^s(a) = m_{21}^{ks}(a) = m_{21}^k(a) = m_2 m_{21}^k(a)$$

and
$$[m_1 m_{21}^k]^s(a) = m_1 m_{21}^{ks}(a) = m_1 m_{21}^k(a)$$

for each a$\in \Sigma$ and s $\geq$ 1, which shows that the systems of a normal combination satisfy condition (3) of normality.

We will say that the set S has bounded balance if each pair $(G_1^j, G_2^j) \in S$ has bounded balance.

<u>Lemma 4</u> Let (G_1, G_2) be a pair of DOL systems. Let S be their ι-combination for some $\iota \in \{1,2\}^+$. Then

(i) G_1, G_2 are equivalent iff for all $(G_1^j, G_2^j) \in S$, G_1^j, G_2^j are equivalent.

(ii)   Let G_1 and G_2 be equivalent. Then (G_1, G_2) has bounded balance iff their ι-combination S has bounded balance.

<u>Proof</u> Part (i) can be proved analogically as Theorem 1. Now, let $k = |\iota|$ and assume that (G_1^j, G_2^j) has bounded balance and let w be a G_1-prefix, say, $ww' = h_1^n(\sigma)$ for some $n \geq 0$ and some $w' \in \Sigma^*$. When proving that the balance is bounded on a set of strings we may neglect finitely many strings, so, let $n \geq k$. Let $\iota = i_1 i_2 \ldots i_k$ and $h = h_{i_1} h_{i_2} \ldots h_{i_k}$. Let ua with $u \in \Sigma^*$, $a \in \Sigma$ be a prefix of $h_1^{n-k}(\sigma)$ such that $h(u)$ is a prefix of w, but w is a proper prefix of $h(ua)$. (Such ua exists if w is a proper prefix, but if w is the whole string $h_1^n(\sigma)$ then $\beta(w) = 0$, so again we may ignore this), i.e. $h(u)x = w$ for some $x \in \Sigma^*$, and x is a prefix of $h(a)$, from which $|x| \leq H^k$ and $\beta(x) \leq BH^k$, i.e. $\beta(w) \leq \beta(h(u)) + BH^k$, where $H = \max_{a \in \Sigma} |h_1(a), h_2(a)|$ and $B = \max_{a \in \Sigma} \beta(a)$. The boundedness of $\beta(w)$ follows from the fact that $\beta(h(u)) = ||\bar{h}_1(u)| - |\bar{h}_2(u)|| = \beta_j(u)$, where we denoted by β_j the balance in (G_1^j, G_2^j) which is bounded, and j is chosen so that w is a G_1^j-prefix.

The converse, namely, that if (G_1, G_2) has bounded balance so has each (G_1^j, G_2^j) is obvious, and is not in fact needed in our proofs.   □

<u>Definition 6</u> Let $G = (\Sigma, h, \sigma)$ be a DOL system and let $\Pi \subset \Sigma$ be a subalphabet, and assume that h^Ω is propagating. For every $avb \in \Omega\Pi^*\Omega$ we define an *induced system* $G^{avb} = (\Sigma^a + \Pi' + {}^b\Sigma, \hat{h}, \bar{a}v\bar{\bar{b}})$ as follows.

For $a \in \Omega$, let $h(a) = xcv$, where $c \in \Omega$, $v \in \Pi^*$. (Note that such decomposition is possible because h^Ω is propagating, and is obviously unique.) We denote $\iota(a) = c, \iota'(a) = v$. Similarly, writing $h(a) = v'c'y$, where $c' \in \Omega$, $v' \in \Pi^*$, we define $r(a) = c'$, $r'(a) = v'$.

We define $\Sigma^a = \{\bar{c}$: there is $n \geq 0$ and a sequence $c_0 = a, c_1, \ldots, c_{n-1}, c_n = c, c_j \in \Omega$ such that $c_j = \iota(c_{j-1}), j = 1, 2, \ldots, n\}$, where $\bar{c}$ is one new symbol for each $c \in \Omega$. Similarly, we define ${}^b\Sigma$ starting with $c_0 = b$ and using r instead of ι: ${}^b\Sigma = \{\bar{\bar{c}}$: there is $m \geq 0$ and a sequence $c_0 = b, c_1, \ldots, c_m = c, c_j \in \Omega$ and $c_j = r(c_{j-1})$ for $j = 1, 2, \ldots, m\}$, and where $\bar{\bar{c}}$ is another new symbol, one for each $c \in \Omega$. Let $\hat{h}(\bar{a}) = \overline{\iota(a)}\iota'(a)$ for $a \in \Omega$, $\hat{h}(\bar{\bar{a}}) = r'(a)\overline{\overline{r(a)}}$ for $a \in \Omega$, $\hat{h}(d) = h(d)$ for $d \in \Pi$. Finally, Π' is the subset of Π of symbols actually used when the homomorphism $\hat{h}$ is repeatedly applied to v. That completes the definition of G^{avb}. When starting with G_1 or G_2 we will, as

usual, talk about $\hat{h}_1, \hat{h}_2, G_1^{avb}$, and G_2^{avb}.

__Lemma 5__ Let G_1, G_2 be two equivalent DOL systems with a common subalphabet Π. Assume both h_1 and h_1^Ω are propagating and there exists a constant k such that for every G_1-prefix of the form xav, where $a \in \Omega$, $x \in \Sigma^*$, and $v \in \Pi^*$ we have

(5)   if $|v| > k$, then $h_1^\Omega(xa) = h_2^\Omega(xa)$.

Then for every $avb \in \Omega\Pi^*\Omega$, $|v| > k$, avb a substring of $L(G_1)$ the systems G_1^{avb}, G_2^{avb} are equivalent.

__Proof__ As avb is a G_1-substring, we can write $xavby = h_1^j(\sigma)$ for some $x,y \in \Sigma^*$ and some $j \geq 0$. From (5) we have $h_1(xa) = x'l_1(a)l_1'(a)$, $h_2(xa) = x'l_2(a)l_2'(a)$, where l_1, l_1' and l_2, l_2' are functions from definition 6 based here on h_1 and h_2. Similarly, $h_i(xavb) = x'l_i(a)l_i'(a)h_i(v)r_i'(b)r_i(b)x_i''$, for some $x', x_i'' \in \Sigma^*$, $i = 1,2$. Strings $h_1(xa)$, $h_2(xa)$ and $h_1(xavb)$, $h_2(xavb)$ are prefixes of the same string $h_1^{j+1}(\sigma) = h_2^{j+1}(\sigma)$, so $l_1(a) = l_2(a) \in \Omega$; $l_1'(a)h_1(v)r_1'(b)$ and $l_2'(a)h_2(v)r_2'(b) \in \Pi^*$, but they are equal as the next symbol $r_1(b) = r_2(b) \in \Omega$. That is, $|v| > k$ implies (through $h_1^\Omega(xa) = h_2^\Omega(xa)$) that $\hat{h}_1(\bar{a}v\bar{b}) = \hat{h}_2(\bar{a}v\bar{b})$. As h_1 - and thus $\hat{h}_1$ - are propagating, also $|h_1(v)| \geq |v| > k$. This proves that G_1^{avb}, G_2^{avb} are equivalent.   □

5. The main theorem

__Theorem 4__ Every pair of normal equivalent DOL systems has bounded balance.

__Proof Outline__ (for details see [2], Theorem 3.)
Let $G_i = (\Sigma, h_i, \sigma)$ for $i = 1,2$; and let r be the order of G_1 (same as G_2), i. e. $r = |\Sigma_c|$. The proof is by the induction on r. It is easy to verify that the assertion holds for systems with one-letter core ($|\Sigma_c| = 1$) which serves as base, $r = 1$, of the induction.

The induction hypothesis states that the assertion holds for systems of order smaller than $r \geq 1$. To prove the induction step let us consider two normal equivalent systems of order r, i.e. $|\Sigma_c| = r \geq 2$.

Case I: Assume that $h_1(a) = h_2(a) = \epsilon$ for some $a \in \Sigma_c$. Then $\Pi = \{a\}$ is a common subalphabet. Let $\Omega = \Sigma - \Pi$. Since G_1 and G_2 are equivalent also G_1^Ω and G_2^Ω are equivalent and since $|\Omega_c| < |\Sigma_c|$ the pair (G_1^Ω, G_2^Ω) has bounded balance by the induction hypothesis. Subalphabet Π is clearly limited and therefore the pair (G_1, G_2) has bounded balance by Lemma 3.

Case II: Assume that $h_1(a) = \epsilon$ for some $a \in \Sigma_c$ but not necessarily $h_2(a) = \epsilon$. Consider the normal combination of (G_1, G_2). Clearly, we have $\bar{h}_1(a) = \epsilon$, $\bar{h}_2(a) = \epsilon$, so by case I, (G_1^i, G_2^i) has bounded balance for $i = 1,2$ and so has (G_1, G_2) by Lemma 4.

Case III. This is the most difficult case and we will give only a very rough outline

of the proof, which omits many difficult details (see [2], Lemma 9. claim 1, cases A-D and Subcases III A - III D).

We may now assume that both G_1 and G_2 are propagating. By Lemma 2 either the combination of (G_1, G_2) is simple (this implies, using Theorem 3 and Lemma 4, that (G_1, G_2) has bounded balance), or there is a common subalphabet Π. Let $\Omega = \Sigma - \Pi$ and $\Omega_c = \Sigma_c - \Pi$. We may assume that Π is maximal, i.e. there is no subalphabet Π' so that $\Pi \subsetneqq \Pi' \subsetneqq \Sigma_c$.

Since G_1 and G_2 are equivalent the remainder systems G_1 and G_2 are also equivalent, and since they are of order smaller than r, and normal, the pair (G_1, G_2) has bounded balance by the induction hypothesis.

Now, we consider any G_1-prefix w. If all substrings of w in Π^* are "short", we bound $\beta(w)$ easily. To take care of the G_1-prefixes with long substrings in Π^+ we first show that there is only a finite number of induced systems of G_1 or G_2 generating all such "long" substrings in Π^+ with "sides" from Ω.

The next difficult step is to show that for every G_1-prefix uv with $u \in \Sigma^* \Omega$ and $v \in \Pi^+$ where v is "long" we have $\beta(u) = 0$, i.e. that the balance is zero at every point where a "long" substring in Π^+ starts. Therefore, using Lemma 5, we show that the corresponding induced systems G_1^{avb} and G_2^{avb} for each "long" v in Π^+ are equivalent. Since G_1^{avb} and G_2^{avb} are of order less than r they have bounded balance by the induction hypothesis.

Now, we are finally in the position to show that the balance is bounded also on G_1-prefixes with "long" substrings in Π^+. We can write each such G_1-prefix w in the form w = xyz where $x \in \Sigma^* \Omega$, $y \in \Pi^+$, $z \in \Sigma^*$ where y is the last "long" substring in Π^+ from left, i.e. z does not contain any "long" substring in Π^+ (possibly z = ϵ). It is essential that we already know that $\beta(x) = 0$. Let $H = \max_{i=1,2} (\max_{a \in \Sigma} h_i(a))$, c be the bound on balance of G_1, G_2 and c_π be the maximum of bounds on balance of the finitely many pairs of induced systems. From $\beta^\Omega(x) = 0$, it follows that $\beta(x) \leq H$. Also $\beta(y) \leq c_\pi + 2H$ and finally $\beta(z) \leq c(H^c + 1)H$. So together we have $\beta(w) \leq \beta(x) + \beta(y) + \beta(z) \leq H + c_\pi + 2H + c(H^c + 1)H$ and therefore the balance of the pair G_1, G_2 is bounded.

6. Equivalent systems with bounded balance

Now, we show that when we consider only DOL systems for which equivalence implies bounded balance, we can design a decision procedure for equivalence. We conjecture that this implication holds for all DOL systems, however in view of Theorem 1 it is sufficient that it holds for normal systems as proved in section 5.

Theorem 5 Let F be a family of DOL systems with the property:(*) If G_1 and G_2 are two equivalent systems in F, then the pair (G_1, G_2) has bounded balance. Then the sequence equivalence problem for F is decidable.

Proof (see also [1], Theorem 2.1).Clearly, we can restrict ourselves to a pair of

reduced DOL systems from F with identical alphabets and identical axioms. We will exhibit two semidecision procedures, one for nonequivalence and the other for equivalence.

The semidecision procedure for nonequivalence is trivial, we compute $h_1^n(\sigma)$ and $h_2^n(\sigma)$ for $n = 0,1,2,...$ and stop with answer "nonequivalent" if $h_1^n(\sigma) \neq h_2^n(\sigma)$ for some n.

Our semiprocedure for equivalence is based on the condition (*) for F, namely that a pair of equivalent systems from F has bounded balance.

Clearly, $h_1^n(\sigma) = h_2^n(\sigma)$ for $n \geq 0$ iff $h_1^n(\sigma) = h_2(h_1^{n-1}(\sigma))$ for $n > 0$ iff $h_1(w) = h_2(w)$ for each w in $L(G_1)$. $L(G_1)$ is a DOL language and therefore also an EOL-language [9].

Now we design a semiprocedure which will check successively for $k = 1,2,...$ whether the pair (G_1,G_2) has k-bounded balance and whether G_1 and G_2 are sequence equivalent. We already know that to check the sequence equivalence it is enough to check whether $h_1(w) = h_2(w)$ for each $w \in L(G_1)$. The checking of these two properties for a particular k is done as follows:

Let M_k be a deterministic g.s.m [10] with a "buffer" of length k in its finite control which for any input string w in Σ attempts to check (from left to right when reading w) whether $h_1(w) = h_2(w)$. It is obviously possible to do this if G_1 and G_2 have k-bounded balance since we have available a "buffer" of length k (i.e. a buffer able to contain k symbols from Σ). Given input w, our g.s.m. M_k will produce its output as follows:

(i) If the buffer of M_k does not overflow and $h_1(w) = h_2(w)$, then no output is produced (M_k goes into a non-accepting state).

(ii) Otherwise, (M_k finds that $h_1(v) \neq h_2(v)$ for some prefix v of w or the buffer overflows), M_k stops in an accepting state and therefore produces some output (what output is produced is irrelevant).

Clearly, $T_k(L(G_1)) = \emptyset$ iff the pair (G_1,G_2) has k-bounded balance and $h_1(w) = h_2(w)$ for all $w \in L(G_1)$. Since EOL-systems are effectively closed under g.s.m. mappings [9] we can construct an EOL-system S_k such that $L(S_k) = T_k(G_1)$. Finally, it is recursively decidable [3,9] whether the EOL-language $L(S_k)$ is empty. Therefore, simply enumerate $S_1,S_2,...$ and test each S_k for $L(S_k) = \emptyset$. The systems G_1 and G_2 are equivalent iff there is k so that $L(S_k) = 0$.

Our semiprocedure must eventually stop if $G_1 \equiv G_2$ since, because condition (*) for F, there exists $c > 0$ so that G_1 and G_2 have c-bounded balance. □

Corollary 1 The sequence equivalence problem for DOL systems is decidable.

Proof Theorem 4 shows that the family of normal systems satisfies the condition (*) of Theorem 5, therefore, by Theorem 5, the sequence equivalence problem is decidable for this family. Thus, by Theorem 1, the sequence equivalence problem is decidable for all DOL systems. □

Corollary 2 The language equivalence problem for DOL systems is decidable, i.e. given two DOL systems G_1,G_2, it is decidable whether $L(G_1) = L(G_2)$.

Proof By Corollary 1 and Corollary 4.6 in [6].

□

References

1. Culik II, K., On the decidability of the sequence equivalence problem for DOL-systems, Theoretical Computer Science 3, 75-84 (1977).

2. Culik II, K. and Fris, I., The decidability of the equivalence problem for DOL-systems, Inf. and Control, in print, also Res. Rep. CS-76-25, Dept. of Computer Science, University of Waterloo (1976).

3. Culik II, K., On some families of languages related to developmental systems, Inter. J. of Computer Math. 4, Section A, 31-42 (1974).

4. Gantmacher, F. R., The Theory of Matrices, vol. 2, Chelsea, New York (1960).

5. Lindenmayer, A., Developmental systems without cellular interaction, their languages and grammar, J. of Theoretical Biology 30, 455-484 (1971).

6. Nielsen, M., On the decidability of some equivalence problem for DOL-systems, Information and Control 25, 166-193 (1974).

7. Paz, A. and Salomaa, A., Integral sequential word functions and growth equivalence of Lindenmayer systems, Information and Control 23, 313-343 (1973).

8. Problem Book, Unusual Automata Theory, January 1972, Dept. of Computer Science, University of Aarhus Techn. rep. DAIMI PB-15, 14-26 (1973).

9. Rozenberg, G., Extensions of tabled OL-systems and languages, Inter.J. of Computer and Information Sciences, Vol. 2, 311-334 (1973).

10. Salomaa, A., Formal Languages, Academic Press, New York, (1973).