ON THE DECIDABILITY OF THE SEQUENCE EQUIVALENCE PROBLEM FOR DOL-SYSTEMS*

K. Culik II
Department of Computer Science
University of Waterloo
Waterloo, Ontario

CS-75-24

September 1975

ON THE DECIDABILITY OF THE SEQUENCE EQUIVALENCE
PROBLEM FOR DOL-SYSTEMS*

by

K. Culik II
Department of Computer Science
University of Waterloo
Waterloo, Ontario

Abstract. A property, called smoothness, of a family of DOL-systems is introduced. It is shown that the sequence equivalence problem is decidable for every smooth family of DOL-systems. Then a large sub-family of DOL-systems, called simple DOL-systems is shown to be smooth.

0. Introduction

Shortly after the introduction of OL-systems by A. Lindenmayer in [7], the question was asked whether the equivalence problem is decidable for these systems [12]. The undecidability of the equivalence problem for (nondeterministic) OL-systems was shown, e.g. in [1]. The same question for deterministic OL-systems (DOL-systems) is conjectured to be decidable but remains open; according to the survey paper [11] it is "without any doubt, the most intriguing open mathematical problem around L-systems".

The equivalence problem was shown decidable for some special sub-classes of DOL-systems, e.g. [6]. The growth-equivalence problem for DOL-systems was shown to be decidable in [9] as well as the equivalence problem for other types of weak equivalences [8]. It was also shown in [8] that the language equivalence problem for DOL-systems is recursively decidable iff the sequence equivalence problem is recursively decidable and that to resolve the latter it is enough to consider DOL-systems in certain normal form.

* Part of this work was done during author's stay at the University of Utrecht in April 1975. It was also supported by National Research Council of Canada, Grant No.A7403.

The main goal of this paper is to show a sufficient condition for a subfamily of DOL-systems to have recursively decidable the sequence equivalence problem. Our approach is based on the notion of the balance of a string. Consider two DOL-systems $G_i = (\Sigma, h_i, \sigma)$ for i = 1,2. The balance of a string w in Σ^* is the difference of lengths of $h_1(w)$ and $h_2(w)$. We say that a pair of DOL-systems has bounded balance if there exists a constant $c > 0$ such that no prefix of any string generated by these systems has the balance larger than c. Finally, a subfamily of DOL-systems is called smooth if every pair of sequence equivalent systems from the subfamily has bounded balance. We close section 2 by showing that smoothness is a sufficient condition for decidability of the sequence equivalence problem.

In the next section we exhibit an example of a smooth subfamily of DOL-systems, called simple DOL-systems. Intuitively, a DOL-system is simple if every symbol of its alphabet can be obtained (possibly in several steps) from every other symbol of the alphabet. To show that the family of simple DOL-systems is smooth we first demonstrate that for every pair of sequence equivalent simple DOL-systems the balance of a long prefix of a string generated by such systems is "very small" compared with the length of the prefix. Then we strengthen this result by showing that the pair has bounded balance.

We have strong reasons to conjecture that the above approach can be extended to show the decidability of sequence equivalence problem for the family of (all) DOL-systems.

1. Prerequisites

The set of non-negative integers is denoted by N. Given an alphabet Σ, Σ^* is the free monoid generated by Σ with the unit ε (empty string); $\Sigma^+ = \Sigma^* - \{\varepsilon\}$.

For $w \in \Sigma^*$ and $a \in \Sigma$, $\#_a(w)$ is the number of occurrences of symbol a in the string w. If $\Sigma = \{a_1,\ldots,a_n\}$ then the vector $(\#_{a_1}(w),\#_{a_2}(w),\ldots,\#_{a_n}(w))$ is called the Parikh vector of w and is denoted by $[w]$.

For an integer i let $|i|$ denote the absolute value of i. For w in Σ^*, $|w|$ denotes the length of w; in particular $|\varepsilon| = 0$. For $\alpha \in N^k$, $\alpha = (\alpha_1,\ldots,\alpha_k)$, let $|\alpha| = \sum_{i=1}^{k} |\alpha_i|$, thus $|[w]| = |w|$ for $w \in \Sigma^*$. For a set Σ let $|\Sigma|$ be the cardinality of Σ.

For $w \in \Sigma^*$, $\min(w) = \{a \in \Sigma\colon a \text{ occurs in } w\}$.

A DOL-system is a 3-tuple $G = (\Sigma,h,\sigma)$ where Σ is an alphabet, h is a homomorphism on Σ^* and axiom σ is in Σ^+.

For DOL-system $G = (\Sigma,h,\sigma)$ the language generated by G is defined as $L(G) = \{h^n(\sigma) : n \geq 0\}$.

Two DOL-systems $G_i = (\Sigma_i,h_i,\sigma_i)$ for i = 1,2 are called (sequence) equivalent if $h_1^n(\sigma_1) = h_2^n(\sigma_2)$ for all n in N; we write $G_1 \equiv G_2$. They are language equivalent if $L(G_1) = L(G_2)$.

For the definition of the growth matrix of a DOL-system we refer to [9].

For DOL-system $G = (\Sigma,h,\sigma)$ we say that w in Σ^+ is a G-prefix (G-substring) if w is a prefix (substring) of $h^n(\sigma)$ for some $n \geq 0$.

2. A sufficient condition for decidability of sequence equivalence

Let $G_i = (\Sigma, h_i, \sigma)$ for $i = 1,2$, be two DOL-systems, and let w be in Σ^*. The balance of w with respect to (G_1, G_2) is denoted by $\beta(w)$ and defined as

$$\beta(w) = ||h_1(w)| - |h_2(w)||$$

We say shortly balance of w if a pair (G_1, G_2) is understood.

Let $G_i = (\Sigma, h_i, \sigma)$ for $i = 1,2$ be two equivalent DOL-systems and $c \geq 0$. We say that the pair (G_1, G_2) has c-bounded balance if $\beta(w) \leq c$ for every G_1-prefix w. We say that (G_1, G_2) has bounded balance if it has c-bounded balance for some $c \geq 0$.

We say that a family F of DOL-systems is smooth if every pair of sequence equivalent systems from F has bounded balance.

Theorem 2.1 The (sequence) equivalence problem is recursively decidable for every smooth family F of DOL-systems.

Proof Clearly, we can restrict ourselves to DOL-systems from F with identical alphabets and identical axioms.

We will exhibit two semidecision procedures, one for non-equivalence and the other for equivalence.

1. The semidecision procedure for non-equivalence is trivial, we compute $h_1^n(\sigma)$ and $h_2^n(\sigma)$ for $n = 0,1,2,\ldots$ and stop with answer "non-equivalent" if $h_1^n(\sigma) \neq h_2^n(\sigma)$ for some n.
2. Our semiprocedure for equivalence is based on the assumption that F is smooth, i.e. 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 $n \geq 0$ 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 [10].

Now we design a semiprocedure which will check successively for $k = 1,2,\ldots$ 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. [13] 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) If M_k finds that $h_1(v) \neq h_2(v)$ for some prefix v of w before its buffer overflows, it stops (in an accepting state) and produces "0".

(iii) Otherwise (buffer overflows) M_k stops (in an accepting state) and produces output "1".

<u>Note:</u> The different outputs in (ii) and (iii) are used to describe an alternative procedure below. Let T_k be the translation defined by M_k.

Clearly, $T_k(L(G_1)) = \phi$ iff the pair (G_1,G_2) has k-bounded balance and $h_1(w) = h_2(w)$ for all $w \in L(G_1)$. By [4] or [3] we can construct an EOL-system S_k such that $L(S_k) = T_k(G_1)$. Finally, it is recursively decidable [2,10] whether the EOL-language $L(S_k)$ is empty.

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

Alternative proof of Theorem 2.1. We can drop the first semi-procedure and modify the second into an algorithm (always halts) as follows:

We again construct the EOL-system S_k successively for $k = 1,2, \ldots$. For every k, if $L(S_k) = \phi$ then stop with answer "$G_1 \equiv G_2$". Otherwise, if $0 \in L(S_k)$, then stop with answer "$G_1 \not\equiv G_2$". Otherwise, increase k and repeat.

We are able to check whether $0 \in L(S_k)$ since the membership problem is decidable for EOL-languages [2,10].

If $G_1 \equiv G_2$ then the algorithm halts for the same reason as the second procedure above.

If $G_1 \not\equiv G_2$, even if the balance is not bounded, there exists a shortest G_1-prefix u such that $h_1(u) \neq h_2(u)$. We need at most buffer of length $|u|$ to establish $G_1 \not\equiv G_2$ so the algorithm would stop, at the latest, at system S_k.

□

3. Simple DOL-systems

Let G be a DOL-system over at least two letter alphabet with growth matrix M. We say that G is simple DOL-system (SDOL-system) if there exists $k \geq 1$ so that all elements of M^k are nonzero.

Lemma 3.1 Let $G = (\Sigma, h, \sigma)$ be an SDOL-system. Then G is exponentially growing [9]. Moreover, there exist $n_0, d, c_1, c_2 > 0$ so that for all $n \geq n_0$ and every w in Σ^*

$$c_1 d^n |w| \leq |h^n(w)| \leq c_2 d^n |w|.$$

Proof It follows from results in [9]. □

Lemma 3.2 Let $G_i = (\Sigma, h_i, \sigma)$ be two sequence equivalent SDOL-systems. For each a in Σ and each $\varepsilon > 0$ there exists $n_{a,\varepsilon}$ so that $\beta(h_1^n(a)) \leq \varepsilon |h_1^n(a)|$ for all $n \geq n_{a,\varepsilon}$.

Proof (version due to J. Hammerum). Let M_1 and M_2 be the growth matrices of G_1 and G_2, respectively. Let k be the smallest k such that M_i^k has all nonzero-elements, for $i = 1,2$. Such k exists since G_1 and G_2 are simple.

Then for all vectors v and all $\varepsilon > 0$ there exists m_0 so that for $m > m_0$ there is a vector t_m and a number d_m so that

$$vM_1^{km} = d_m u + t_m$$

where $|t_m| < \varepsilon\, |vM_1^{km}|$ and u is the characteristic vector with the largest eigenvalue for M_1^k.

It is easy to establish that M_1 has the same property. From this follows that for all $a \in \Sigma$ and $\varepsilon > 0$ there exists n_0, so that for all $n > n_0$ there exists a vector t_n and a number d_n so that

$$[h_1^n(a)] = [a]M_1^n$$

$$= d_n u + t_n,$$

where $|t_n| < \varepsilon\ |h_1^n(a)|$ and u is the characteristic vector with the largest eigenvalue for M_1.

We can prove that

$$\beta(w) \leq |[w](M_1\text{-}M_2)|$$

because

$$\beta(w) = ||h_1(w)| - |h_2(w)||$$

$$\leq \sum_{a\in\Sigma} |\#_a(h_1(w)) - \#_a(h_2(w))|$$

$$= |[w](M_1\text{-}M_2)|$$

(one may notice that equality occurs when one of words $h_1(w)$ and $h_2(w)$ is a subword of the other.)

Noting that u is a characteristic vector for M_2 as G_1 and G_2 are equivalent the following inequality holds

$$\beta(h_1^n(a)) \leq |(d_n u + t_n)(M_1\text{-}M_2)|$$

$$\leq |d_n u(M_1\text{-}M_2)| + |t_n(M_1\text{-}M_2)|$$

$$= |t_n(M_1\text{-}M_2)|$$

which proves the lemma. □

Lemma 3.3 Let $G_i = (\Sigma, h_i, \sigma)$ for $i = 1,2$ be two sequence equivalent SDOL-systems. For each $\varepsilon > 0$ there is $n_\varepsilon > 0$ so that for every w in Σ^+ and all $n \geq n_\varepsilon$ we have $\beta(h_1^n(w)) \leq \varepsilon |h_1^n(w)|$.

Proof By Lemma 3.2 for each a in Σ and each $\varepsilon > 0$ there is $n_{a,\varepsilon}$ so that for $n \geq n_{a,\varepsilon/2}$ we have $\beta(h_1^n(a)) \leq \varepsilon |h_1^n(a)|$. Let $n_\varepsilon = \max_{a\in\Sigma}\{n_{a,\varepsilon}\}$ and let $w = a_1 a_2 \ldots a_k$. For $n \geq n_\varepsilon$ we have

$$\beta(h_1^n(w)) \leq \sum_{i=1}^{k} \beta(h_1^n(a_i)) \leq \varepsilon \sum_{i=1}^{k} |h_1^n(a_i)| = \varepsilon |h_1^n(w)|.$$

□

Theorem 3.1 Let $G_i = (\Sigma, h_i, \sigma)$ for $i = 1,2$ be two sequence equivalent SDOL-systems. For each $\varepsilon > 0$ there exists $K_\varepsilon > 0$ so that $\beta(w) \leq \varepsilon |w|$ for every G_1-prefix w such that $|w| \geq K_\varepsilon$.

Proof Let w be a prefix of $h_1^n(\sigma)$. We define $t > 0$ and $u_t, v_t, u_{t-1}, v_{t-1}, \ldots, u_1, v_1$ in Σ^* as follows:

(i) Let t be the maximal integer such that $h_1^t(b)$ is a prefix of w where b is the first symbol of $h_1^{n-t}(\sigma)$.

(ii) Let $u_t = h_1^{n-t}(\sigma)$ and let v_t be the longest prefix of $h_1^{n-t}(\sigma)$ such that $h_1^t(v_t)$ is a prefix of w.

(iii) For $i = t-1, \ldots, 0$ u_i is obtained from $h_1^{n-i}(\sigma)$ by removing its prefix $h_1^{t-i}(v_t) h_1^{t-i-1}(v_{t-1}) \ldots h_1(v_{i+1})$, and v_i is the longest prefix of u_i such that $h_1^t(v_t) h^{t-1}(v_{t-1}) \ldots h^i(v_i)$ is a prefix of w ($h_1^0(x) = x$ for each x in Σ^*).

Let $w_i = h^i(v_i)$ for $i = 0,1,\ldots,t$. Clearly, $w = w_t w_{t-1} \ldots w_0$.

Note that w_k may be empty for some k.

Given $\varepsilon > 0$ there exists, by Lemma 3.3, $n_{\varepsilon/2}$ so that $\beta(h_1^i(v_i)) \leq \frac{\varepsilon}{2}|h_1^i(v_i)|$ for $n_{\varepsilon/2} \leq i \leq t$, i.e. $\beta(w_i) \leq \frac{\varepsilon}{2}|w_i|$ for $n_{\varepsilon/2} \leq i \leq t$ and, therefore, $\beta(w_t w_{t-1} \cdots w_{n_{\varepsilon/2}}) \leq \frac{\varepsilon}{2}|w_t w_{t-1} \cdots w_{n_{\varepsilon/2}}|$.
Let $Q = \max_{a \in \Sigma} \beta(a)$. We have $\beta(w_{n_{\varepsilon/2}-1} \cdots w_1) \leq Q|w_{n_{\varepsilon/2}-1} \cdots w_1|$ and we, clearly, can choose K_ε so that if $|w| \geq K_\varepsilon$, then $|w_{n_{\varepsilon/2}-1} \cdots w_1| \leq \frac{\varepsilon}{2Q}|w|$; and thus $\beta(w_{n_{\varepsilon/2}-1} \cdots w_1) \leq \frac{\varepsilon}{2}|w|$.

Together, we have for w such that $|w| \geq K_\varepsilon$

$$\beta(w) \leq \beta(w_t \cdots w_{n_{\varepsilon/2}}) + \beta(w_{n_{\varepsilon/2}-1} \cdots w_1) \leqq (\frac{\varepsilon}{2} + \frac{\varepsilon}{2})|w| = \varepsilon|w|.$$

□

Note that only a weaker equivalence than sequence equivalence is used in proofs of Lemma 3.2, Lemma 3.3 and Theorem 3.1, namely so called Parikh equivalence, see [8].

A derivation forest of a string w with respect to a DOL-system G is an obvious modification of the well-known notion of derivation tree for a context-free grammar, we have an axiom (string of symbols) rather than one starting symbol of a context-free grammar.

Let F be derivation forest of $h^n(\sigma)$ for some $n > 0$ with respect to SDOL-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

$$q = (a_0, k_0) \cdots (a_n, k_n), \text{ then}$$

(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 $0 \leq i \leq n-1$.

Intuitively, 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 $k_i = 1$ ($k_i = |h(a_i)|$). 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 symbols 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 [5; Theorem 11.3 and 11.4]).

Theorem 3.2 If $G_i = (\Sigma, h_i, \sigma)$ for $i = 1,2$ are two sequence equivalent SD0L-systems, then the pair (G_1, G_2) has bounded balance.

Proof 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 top) such that the label of α_1 is the same as that 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 also holds for the right neighbors. Let the common labels be a, b, c from the left, 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, given an SDOL system G, there exists a constant C so that $r+t \leq C$ independently on n_0. We only note that first we have the constant $C_1 = |\Sigma|+1$ with the property that on levels higher than C_1 there is at least one neighbor both to the left and to the right of the node of chain q. 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 $q = q_1q_2q_3$ where q_2 is the section of q between nodes α_1 and α_2. Let q' be the periodical chain defined by $q' = q_1q_2^jq_4$ where $j > 0$ and q_4 is a proper prefix of q_2, 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 $h_1^{n-t}(\sigma) = u_1xyv_1$ and $h_2^n(\sigma) = u_2xyv_2$ where $xy = h_1^{n-r-t}(abc)$, $u_2x = u$ and $yv_2 = v$, i.e. the node α on level n is at the last symbol of x. We write $u' = u_1x$ and $v' = yv_1$. Clearly, $h_1^{n-r-t}(a)$ is a prefix of x and $h_1^{n-r-t}(c)$ is a suffix of y, therefore from Lemma 3.1. and discussion above it follows that the length of both x and y for growing n is linearly proportional to the length of the whole string $h_1^n(\sigma)$, i.e. there exists constant K, dependent on G_1 only, such that $K|x| \geq |h_1^n(\sigma)|$ and $K|y| \geq |h_1^n(\sigma)|$. Therefore, it follows from Theorem 3.1 that for each $\varepsilon > 0$ there exists n_0 so that $\beta(u) \leq \varepsilon\ |x|$ and $\beta(u) \leq \varepsilon\ |y|$ where u,x and y are determined by n_0.

Now, we explain first the following step in the proof informally and then we will give the details. Both $h_1^n(\sigma)$ and $h_1^{n-t}(\sigma)$ have y as a substring with node α 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 $|h_1(y)|$. By Assumption 1 $\beta(u') < \beta(u)$, and therefore the relative position of $h_1(y)$ and $h_2(y)$ as substrings of $h_1^{n-t+1}(\sigma)$ is by a "small" shift (with respect to the length of $h_1^{n+1}(\sigma)$ and also of $h_1(y)$) different than the relative position of the same strings as substrings of $h^{n+1}(\sigma)$. Therefore $h_1(y)$ has to have "long" identical prefix ans suffix and consequently must be periodic with a period arbitrarily short with respect to its length for large enough n.

Formally, using the notation introduced above, we have.

(1) $h_1^{n-t}(\sigma) = u'yv_1$ and $h_1^n(\sigma) = uyv_2$

where $\beta(u)$ is strictly maximal up to the level n. Since the systems G_1 and G_2 are equivalent we obtain from (1)

(2) $h_1(u')h_1(y)h_1(v_1) = h_2(u')h_2(y)h_2(v_1)$

and

(3) $h_1(u)h_1(y)h_1(v) = h_2(u)h_2(y)h_2(v_2)$.

Without loss of generality we may assume that $|h_1(u')| \geq |h_2(u')|$,i.e. $h_1(u') = h_2(u')\ z'$ for some z' in Σ^*. Therefore by removing prefix $h_2(u')$ on both sides of (2) we have

(4) $z'h_1(y)h_1(v_1) = h_2(y)h_2(v_1)$.

Now we have to consider two cases.

Case A. Let $|h_1(u)| \geq |h_2(u)|$, i.e. $h_1(u) = h_2(u)z$ for some z in Σ^*. By Assumption 1, $\beta(u) > \beta(u')$ and thus $|z| > |z'|$. By removing prefix $h_2(u)$ on both sides of (3) we get

(5) $zh_1(y)h_1(v_1) = h_2(y)\ h_2(v_2)$.

Since $|z| > |z'|$ and z is "very small" with respect to $h_1(y)$ it follows by comparing (4) and (5) that there exists p in Σ^* so that $z=z'p$ and $h_1(y) = p^j\ d$ where both p and d are "very small" with respect to $h_1(y)$.

Case B. Let $|h_1(u)| < |h_2(u)|$, i.e. $h_1(u)z = h_2(u)$ for some z in Σ^*, whereagain $|z|$ is "very small" with respect to $|h_1(y)|$. By removing the prefix $h_2(u)$ from both sides of (3) we obtain

(6) $\delta h_1(v_2) = h_2(y)h_2(v_2)$

where δ is obtained from $h_1(y)$ by removing prefix z, i.e. $h_1(y) = z\delta$.
By comparing (4) and (6), we see that δ is prefix of $z'h_1(y)$ and therefore the string $h_1(y)$ has an identical "very long" prefix and suffix and thus must be periodic, i.e. the form $h_1(y) = p^j d$, where both $|p|$ and $|d|$ are "very small" with respect to $|h_1(y)|$.

Thus in both case A and case B $h_1(y)$ has to be of the form $p^j d$ where by choosing n_0 large enough we can make p arbitrarily short with respect to $h_1(y)$ and therefore j arbitrarily large.

So far we have used only the fact that $\beta(u)$ is strictly maximal up to the level n (condition (i)) not yet the condition (ii) and (iii). So we can use the above arguments also symetrically for the common substring x. Therefore also x must be periodical and by taking the common period we see that whole substring xy must be periodical. In the following p will denote the period of xy, i.e. $x = d_1 p^{k_1}$ and $y = p^{k_2} d_2$ where p, d_1 and d_2 are "short" and k_1, k_2 "large" in the sense used above.

Now, we will consider again two cases.

Case I: Let $\beta(p) = 0$. Since p is arbitrarily shorter than $h_1(y)$ but nonempty and since $K|x| \geq h_1^n(\sigma) \geq |x|$ and $K|y| \geq h_1^n(\sigma) \geq |y|$ we also have $|p| < |h_1(x)|$.

So we can write $u = \bar{u}p$ for some $\bar{u}$ in Σ^* and since $\beta(p) = 0$, we have $\beta(\bar{u}) = \beta(u)$ which is in contradiction with condition (iii) implied by Assumption 1.

Case II: Let $\beta(p) > 0$. We already know that $|p| < |h_1(x)|$ and $|p| < |h_1(x)|$. Therefore, we can write $h_1^n(\sigma) = \bar{u}p^2\bar{v}$ for $\bar{u},\bar{v}$ in Σ^* such that $\bar{u}p= u$ and $p\bar{v} = v$. Since $\beta(p) > 0$, clearly either $\beta(\bar{u}) > \beta(u)$ or $\beta(\bar{u}p^2) > \beta(u)$ which is in contradiction with condition (ii) implied by Assumption 1. □

Corollary The sequence equivalence problem is decidable for the family of SDOL-systems.

Proof By Theorem 3.2 and Theorem 2.1. □

ACKNOWLEDGEMENTS

The author is grateful to P. Johansen, A. Lindenmayer, T. Maibaum, J. Opatrný, A. Salomaa and S. Skyum for discussions and remarks to this work.

References

[1] M. Blattner, The unsolvability of the equality problem for sentential forms of context-free grammars, J. of Computer and System Sciences 7, 463-468 (1973).

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

[3] K. Culik II and J. Opatrný, Macro OL-systems, Inter. J. of Computer Math. 5, Section A, to appear.

[4] G.T. Herman, Closure properties of some families of languages associated with biological systems, Information and Control.

[5] G.T. Herman and G. Rozenberg, Developmental systems and languages, North-Holland/American Elsevier, 1974.

[6] P. Johansen and E. Meiling, Free groups in Lindenmayer systems, in L-systems edited by G. Rozenberg and A. Salomaa, Lecture Notes in Computer Science, Springer-Verlag, vol.15 (1974).

[7] A. Lindenmayer, Developmental systems without cellular interaction, their languages and grammars, J. of Theoretical Biology 30, 455-484 (1971).

[8] M. Nielsen, On the decidability of some equivalence problems for DOL-systems, Inf. and Control 25, 166-193 (1974).

[9] A. Paz and A. Salomaa, Integral sequential word functions and growth equivalence of Lindenmayer systems, Inf. and Control 23, 313-343 (1973).

[10] G. Rozenberg, Extensions of tabled OL-systems and languages, Inter. J. of Computer and Information Sciences, vol.2, 311-334, (1973).

[11] G. Rozenberg and A. Salomaa, The mathematical theory of L-systems, in Progress in Information Science edited by J. Tou, Academic Press, to appear.

[12] Problem book - Unusual Automata Theory January 1972, Dep. of Computer Science, University of Aarhus Techn. rep. DAIMI PB-15, pp.14-26, June 1973.

[13] J.E. Hopcroft and J.D. Ullman, Formal languages and their relation to automata, Addison-Wesley, Reading, Mass. (1969).