STOC(Milwaukee 1981), 1-6.

The ω-Sequence Equivalence Problem for DOL Systems is Decidable†

Karel Culik II
Department of Computer Science
University of Waterloo
Waterloo, Ontario, Canada

Tero Harju
Department of Mathematics
University of Turku
Turku, Finland

Abstract

The following problem is shown to be decidable. Given are homomorphisms h_1 and h_2 from Σ^* to Σ^* and strings σ_1 and σ_2 over Σ such that $h_i^n(\sigma_i)$ is a proper prefix of $h_i^{n+1}(\sigma_i)$ for $i = 1, 2$ and all $n \geq 0$, i.e. for $i = 1, 2$, h_i generates from σ_i an infinite string α_i with prefixes $h_i^n(\sigma_i)$ for all $n \geq 0$. Test whether $\alpha_1 = \alpha_2$. From this result easily follows the decidability of limit language equivalence (ω-equivalence) for DOL systems.

1. Introduction

Since the old work of Thue [15], infinite words (ω-words) have been investigated. Apart from being of interest in its own right, the theory of infinite words has often been able to shed light on some problems concerning ordinary finite words and languages of them. As regards infinite words associated to finite automata, the reader is referred to [8], and as regards those associated to context-free grammars, the reader is referred to [11].

When proving the existence of an infinite cubeless string over a binary alphabet and infinite squareless string over a three-letter alphabet Thue [15] used the following mechanism to generate infinite strings over alphabet Σ. Let $w \in \Sigma^*$ and h be a homomorphism from Σ^* to Σ^* such that $h^n(w)$ is a proper prefix of $h^{n+1}(w)$ for all $n \geq 0$. Then h generates from w an infinite string with prefixes $h^n(w)$ for all $n \geq 0$. Our main result is that it is decidable whether two infinite strings defined as above are equal (ω-DOL equivalence).

Without considering the prefix property we have the simplest (and most important) model for developmental (genetic) programs in cellular biology introduced by Lindenmayer and called DOL system. The problem whether two DOL systems generate an identical sequence of (finite) strings (DOL equivalence problem) was for a long time the best known open problem in the area of L systems [4]. It has been shown in [6] that the decidability of ω-DOL equivalence implies the decidability of ordinary DOL sequence equivalence which indicates that our main result is a hard one especially when the attempts to reduce it to DOL equivalence have not succeeded. However, we are using some auxiliary results and refinements of techniques from [4].

We consider both words (strings) and infinite words also referred to as <u>ω-words</u>, over a finite alphabet Σ. An <u>ω-language</u> is a set of ω-words. If $L \subseteq \Sigma^*$, then $\mathrm{Lim}(L)$ is the ω-language consisting of the ω-words with arbitrary long prefixes belonging to L.

The limit language equivalence problem (or ω-equivalence problem) for a family of languages is

† This research was supported by the Natural Sciences and Engineering Research Council of Canada, Grant No. A 7403.

the decision problem of whether $Lim(L_1) = Lim(L_2)$ for any two effectively given languages L_1 and L_2 from the family. We show that this problem is decidable for DOL languages, given by DOL systems. It was conjectured to be decidable in [6], where it was reduced to DOL system with the initial prefix property, i.e. to the equivalence problem for Thue's mechanism for generating infinite strings discussed above.

Our approach generalizes and extends the techniques used in [4] to prove the decidability of the DOL-sequence equivalence problem. Similar notions of normal systems, simple systems, common subalphabets and combinations of morphisms as in [4] are used, however the situation at a number of places is more difficult and new techniques need to be deviced. The basic strategy remains the same, we show that for every pair of ω-equivalent DOL systems we can construct a finite number of pairs of DOL systems each of them ω-equivalent with "bounded balance".

The crucial property of pairs of DOL systems for the proof of ordinary DOL equivalence is that each pair of sequence equivalent systems has "bounded balance". This is no more true in the case of ω-equivalence, however we will be able to overcome this difficulty by using the compositions hg and gh (or more complicated compositions) instead of homomorphisms g and h, and the result from [6] that the ω-equivalence of hg and gh (from some common starting string w) implies the ω-equivalence of g and h themselves.

We now outline our proof of the decidability of ω-DOL equivalence, for a detailed proof see [5].

The main goal of section 2 is to show that without loss of generality we can restrict ourselves to normal 1-systems. In the next section 1-simple systems are introduced and it is shown, using linear algebraic arguments, that ω-equivalent 1-simple systems have combinations with bounded balance. The last section contains the most crucial arguments showing essentially that the general case can be reduced to the case of 1-simple normal systems.

2. Preliminaries

For notations and definitions in language theory not explained here we refer to [12]. We shall also assume familiarity with the results in [4].

The entity $|x|$ denotes (i) the absolute value of a complex number x ; (ii) the length of a word x ; (iii) the vector $(|x_1|, \ldots, |x_k|)$ if x is a real-valued vector $(x_1, \ldots, x_k)$.

Let x and y be two words over a finite alphabet. If x is a prefix (a postfix, resp.) of y then we denote $x <_{pr} y$ ($x <_{po} y$, resp.). A word x is periodic if it is of the form $x = y^n y_1$, where $n \geq 2$ and $y_1 <_{pr} y$. The words x and y are comparable if either $x <_{pr} y$ or $y <_{pr} x$. The empty word is denoted by e and the free monoid generated by a set Σ is denoted by Σ^* .

If $h_1, \ldots, h_k$ are endomorphisms on Σ^* then $\langle h_1, \ldots, h_k \rangle$ denotes the monoid generated by $h_1, \ldots, h_k$ under the operation of composition of morphisms.

An infinite word is called an ω-word and a set of ω-words is said to be an ω-language. To each language L (of finite words) we associate an ω-language $Lim(L)$, the limit language of L , which consists of the ω-words α having arbitrarily long prefixes belonging to L . Clearly if L is finite then $Lim(L) = \phi$.

A language L is semi-convergent if $Lim(L) \neq \phi$, convergent if each word in L is a prefix of some ω-word in $Lim(L)$. Furthermore L is said to be uniformly convergent if $\#Lim(L) = 1$, i.e. L has an unique limit word.

The limit language equivalence problem (or ω-equivalence problem) for a family of languages means the decision problem $Lim(L_1) =? Lim(L_2)$ for any (effectively given) L_1 and L_2 from the family.

We shall prove that the limit language equivalence problem is decidable for DOL languages.

For the proof of this result we shall first reduce the problem to a simplified form in this section. A DOL system is a construct $G = (\Sigma, h, \sigma)$, where Σ is a finite alphabet, h is an endomorphism on Σ^* and $\sigma \in \Sigma^*$. Denote

$$L(G) = \{h^n(\sigma) : n \geq 0\} ,$$

the language generated by G .

The system G (as defined above) is <u>prefix-preserving</u> if $\sigma <_{pr} h(\sigma)$. The following was shown in [6].

<u>Theorem 2.1</u>

The limit language equivalence problem is decidable for DOL systems iff it is decidable for prefix-preserving DOL systems. □

In fact it was shown in [6] that

$$\mathrm{Lim}(L(G)) = \bigcup_{i=1}^{n} \mathrm{Lim}(L(G_i)) ,$$

where G_i , $i = 1, 2, \ldots, n$, are subsystems of G and G_i is prefix-preserving for each i . From [6] we take also

<u>Theorem 2.2</u>

A prefix-preserving DOL system is uniformly convergent or its limit language is empty. □

By Theorem 2.2 $\#\mathrm{Lim}(L(G)) = 1$ if G is an prefix-preserving and $L(G)$ is infinite.

A DOL system $G = (\Sigma, h, \sigma)$ is a <u>1-system</u> if it is prefix-preserving and furthermore

(i) $\sigma \in \Sigma$, (denote $\Sigma_c = \Sigma - \{\sigma\}$) ,

(ii) $h(\sigma) \in \sigma\Sigma_c^+$ and $h(\Sigma_c) \subseteq \Sigma_c^*$,

(iii) if $a \in \Sigma_c$ then a occurs infinitely many times in the unique limit word of G .

The subset Σ_c of Σ is called a <u>core</u> (core alphabet) of G .

We note here that if $G = (\Sigma, h, \sigma)$ is prefix-preserving and if $h(\sigma) = \sigma x$ then

$$h^{n+1}(\sigma) = h^n(\sigma)h^n(x) ,$$

for all $n \geqq 0$.

The next lemma reveals that we may restrict ourselves to 1-systems.

<u>Lemma 2.3</u>

The limit language equivalence problem is decidable for DOL systems iff it is decidable for 1-systems.

<u>Idea of Proof</u>.

Let $G_i = (\Sigma, h_i, \sigma_i)$, $i = 1, 2$ be two prefix-preserving DOL systems. We construct DOL systems G_i' , $i = 1, 2$ by choosing a new single symbol σ as their starting string and modifying the morphism so that σ will replace the prefix of the infinite words generated by G_1 and G_2 containing all "mortal letters" i.e. letters which occur only finitely often in the generated words.

The following important result was proved in [6] (Thm. 6 in [6]).

<u>Theorem 2.4</u>

Let $G_i = (\Sigma, h_i, \sigma)$ be two 1-systems for $i = 1, 2$ and $h \in \langle h_1, h_2\rangle$. Then

$$\mathrm{Lim}(L(G_1)) = \mathrm{Lim}(L(G_2)) = \{\alpha\}$$

iff

$$\mathrm{Lim}(L(G_1')) = \mathrm{Lim}(L(G_2')) = \{\alpha\} ,$$

where $G_i' = (\Sigma, h_i h, \sigma)$ for $i = 1, 2$. □

This result yields immediately to the following:

<u>Lemma 2.5</u>

Let G_i be as above and $g_i \in \langle h_1, h_2\rangle$ for $i = 1, 2$. The systems G_1 and G_2 define a common limit word α iff the 1-systems $G_i' = (\Sigma, h_i g_i, \sigma)$ define the limit word α for both $i = 1, 2$. □

For the next reduction we need some notations and facts from [4]. Let x be a word in Σ^* and define

$$\min(x) = \{a : a \text{ occurs in } x , a \in \Sigma\}.$$

Let $G = (\Sigma, h, \sigma)$ be a 1-system and $m : P(\Sigma) \to P(\Sigma)$ a function, where $P(\Sigma)$ is the set of subsets of Σ , such that

$$m(\phi) = \phi ,$$
$$m(\{a\}) = \min(h(a)) \quad \text{for } a \in \Sigma ,$$
$$m(A \cup B) = m(A) \cup m(B) .$$

The 1-system G is said to be <u>normal</u> if

$$a \in m^j(b) , \ j > 0 \ \text{ implies } \ a \in m(b)$$

holds for every $a,b \in \Sigma_c$. The following result immediately follows from [4, Lemma 2].

<u>Lemma 2.6</u>

(i) For each 1-system $G = (\Sigma, h, \sigma)$ one can find effectively an integer k such that the 1-system $G^k = (\Sigma, h^k, \sigma)$ is normal.

(ii) For each pair of normal 1-systems $G_i = (\Sigma, h_i, \sigma)$, $i = 1, 2$ one can find effectively an integer k such that the 1-systems $G_i^k = (\Sigma, h_i(h_1h_2)^k, \sigma)$, $i = 1, 2$ are normal. □

The morphism $(h_1h_2)^k$ in (ii) was called a <u>normal combination</u> of (G_1, G_2) in [4].

Combining Lemmas 2.3, 2.5 and 2.6(i) we obtain

Lemma 2.7

The limit language equivalence problem is decidable for DOL systems iff it is decidable for normal 1-systems.

3. 1-Simple Systems

A 1-system $G = (\Sigma, h, \sigma)$ is called 1-simple if there exists $m > 0$ such that for all $a,b \in \Sigma_c$ $a \in \min(h^m(b))$, i.e. from each core letter any other core letter can be obtained in certain fixed number of steps.

In the linear-algebraic terminology this means that the growth matrix of G restricted to Σ_c is primitive. We use this in the detailed proofs.

Using results from linear algebra and linear algebraic methods we obtain the following, for details see [5].

Theorem 3.1

If $G_1 = (\Sigma, h, \sigma)$ and $G_2 = (\Sigma, g, \sigma)$ are limit language equivalent and 1-simple and there are morphisms h_1 and h_2 such that $h = h_1h_2$ and $g = h_2h_1$, then the maximal characteristic values and vectors of the two systems are equal. □

Let $G_i = (\Sigma, h_i, \sigma)$ be two DOL systems for $i = 1, 2$. Define a mapping $\beta : \Sigma^* \to \mathbb{Z}$ by setting

$$\beta(x) = |h_1(x)| - |h_2(x)| .$$

The integer $\beta(x)$ is called the balance of the word x (with respect to h_1 and h_2). The systems G_1 and G_2 are of bounded balance if there exists an integer k such that

$$|\beta(x)| \leq k$$

whenever x is a prefix of a word in $L(G_1)$.

In the proof of the decidability of ordinary sequence equivalence for DOL systems in [2] the crucial result was that every two equivalent normal DOL systems have bounded balance. This cannot be extended to ω-equivalence, even for systems generating aperiodic words, as is shown by the following example:

The axiom of both systems is c. The productions in the first system are

$$g : a \to aa,\ b \to b,\ c \to cab$$

and in the second:

$$f : a \to aaaa,\ b \to b,\ c \to cabaab .$$

Clearly, $f = g^2$, thus the two systems are ω-sequence equivalent, but they do not have bounded balance. Note, however gf and fg are ω-equivalent with bounded balance. They are even equal in this example.

On our way to show that any given pair of ω-equivalent DOL systems can be transformed into such a pair which has bounded balance we start with result showing that balance growth much slower than length of generated words for such systems.

Lemma 3.2

Let $G_i = (\Sigma, h_i, \sigma)$ be two limit language equivalent 1-simple systems with a common maximal characteristic value and let α be their common limit word. If α_n denotes the prefix of α of length n then

$$\lim_{n\to\infty} \frac{|\beta(\alpha_n)|}{n} = 0 ,$$

where β is the balance function of the pair (G_1, G_2). □

We now proceed to the crucial result of this section:

Lemma 3.3

If $G_{12} = (\Sigma, h_1h_2, \sigma)$ and $G_{21} = (\Sigma, h_2h_1, \sigma)$ are 1-simple (and corresponding to the limit language equivalent 1-systems G_1, G_2) then their balance is bounded.

Idea of Proof.

First we show that the systems are either exponentially growing or DOL-equivalent with balance zero.

For the case of exponentially growing systems we first use the "shifting technique" from [2]. The unboundeness of balance implies that there is infinitely many of local strict maxima of balance in the generated infinite string. The shifting technique shows that there is a non-empty substring of the form v^2 at the point of each sufficiently distant maxima. However, unlike in [4], this does not immediately yield a contradiction. A new technique is used to show that in this case there would have to exist substrings with large enough balance to contradict Lemma 3.2. □

4. The General Case

Given a 1-system $G = (\Sigma, h, \sigma)$ a set $\pi \subsetneq \Sigma_c$ is called a subalphabet of G if $h(\pi) \subseteq \pi^*$.

Denote $\Omega = \Sigma - \pi$ and let x^Ω be a word in Ω^* obtained by deleting the symbols from π in x. Furthermore set $h^\Omega(x) = h(x)^\Omega$ and $G^\Omega = (\Omega, h^\Omega, \sigma)$. A set π is called a <u>common subalphabet</u> of the 1-systems G_1 and G_2 if it is a subalphabet of both of them. Note, that in distinction with [4] we are not requiring that $\pi \neq \phi$. The following is obvious.

<u>Lemma 4.1</u>

If G_1 and G_2 are limit language equivalent then so are G_1^Ω and G_2^Ω for any common subalphabet π. Moreover, if G_1 is normal, then so is G_1^Ω. □

Let us fix for Lemmas 4.2-4.5 two normal 1-systems $G_i = (\Sigma, h_i, \sigma)$, $i = 1, 2$ which are limit language equivalent. It is not difficult to prove the following

<u>Lemma 4.2</u>

There is a morphism $h \in \langle h_1, h_2 \rangle$ and a common subalphabet π of $G_{1,i} = (\Sigma, h_i h, \sigma)$, $i = 1, 2$, such that $G_{1,1}$ and $G_{1,2}$ are normal and $G_{1,1}^\Omega$ and $G_{1,2}^\Omega$ are propagating for $\Omega = \Sigma - \pi$. Furthermore, π and h can be found effectively. □

The next lemma appears already in [4] for DOL equivalence. The proof for our case is almost the same.

<u>Lemma 4.3</u>

If G_1 and G_2 are propagating then they have a nonempty common subalphabet or the morphisms h_1h_2 and h_2h_1 are 1-simple. □

Using the Lemmas 4.1-4.3 we show the following

<u>Lemma 4.4</u>

There is a morphism $h \in \langle h_1, h_2 \rangle$ and a common subalphabet π of normal

$$\hat{G}_i = (\Sigma, h_i h, \sigma), \quad i = 1, 2$$

such that $(h_1hh_2h)^\Omega$ and $(h_2hh_1h)^\Omega$ are 1-simple for $\Omega = \Sigma - \pi$. Furthermore π and h can be found effectively. □

<u>Lemma 4.5</u>

The limit language equivalence problem is decidable for DOL systems iff it is decidable for pairs of normal 1-systems

$$G_i = (\Sigma, h_i, \sigma), \quad i = 1, 2$$

with the following property (P) : G_1 and G_2 have a common subalphabet π such that $(h_1h_2)^\Omega$ and $(h_2h_1)^\Omega$ are 1-simple for $\Omega = \Sigma - \pi$.

<u>Proof</u>.

The Lemmas 4.2-4.4 are clearly effective in the construction of the common subalphabets and of the morphisms. The lemma follows now by Theorem 2.4 and Lemma 2.7. □

In the next lemma we prove that the property P implies bounded balance of the "commutator systems".

<u>Lemma 4.6</u>

Let $G_i = (\Sigma, h_i, \sigma)$ be two normal limit language equivalent 1-systems with the property (P). Then the 1-systems $G_{12} = (\Sigma, h_1h_2, \sigma)$ and $G_{21} = (\Sigma, h_2h_1, \sigma)$ have bounded balance.

<u>Idea of Proof</u>.

Let π be the common subalphabet of G_1 and G_2 such that G_{12}^Ω and G_{21}^Ω are 1-simple for $\Omega = \Sigma - \pi$. The 1-simple systems G_{12}^Ω and G_{21}^Ω are limit language equivalent and thus by Lemmas 3.6 and 3.7 their balance is bounded. Define $g_1 = (h_1h_2)^k$ and $g_2 = (h_2h_1)(h_1h_2)^{k-1}$, clearly, the balance of g_1^Ω and g_2^Ω is also bounded by B.

Hence, one of the words $(g_1^n(x))^\Omega$ and $(g_2g_1^{n-1}(x))^\Omega$ is prefix of the other for each $n \geq 0$ and the length of their "difference" is uniformly bounded for all n. We show that these "differences" are cyclicly repeating for growing n. We use this fact to transform 1-systems $\hat{G}_i = (\Sigma, g_i, \sigma)$, $i = 1, 2$, into two sequence equivalent DOL systems. Their balance is bounded by [7] and from the construction it follows that also the pair $\hat{G}_1$, $\hat{G}_2$ has bounded balance. □

Let h and g be two endomorphisms on Σ^*. The <u>compatibility language</u> of h and g is defined by

$$\mathrm{Com}(h, g) = \{x : h(x) <_{pr} g(x) \text{ or } g(x) <_{pr} h(x)\}.$$

Note the resemblance of $\mathrm{Com}(h, g)$ and $\mathrm{Eq}(h, g)$ as defined in [12]. In fact the equality language $\mathrm{Eq}(h, q)$ is included in $\mathrm{Com}(h, g)$. Denote by $\mathrm{Com}_k(h, g)$ the subset of $\mathrm{Com}(h, g)$ which has balance at most k, i.e.

$$\mathrm{Com}_k(h, g) = \{x : x \in \mathrm{Com}(h, g),\ x_1 <_{pr} x \to |\beta(x_1)| \leq k\}.$$

The following lemma is an analogy of the result for bounded equality languages. Since the proof is obvious and similar to that given for $\mathrm{Eq}_k(h, g)$, see [2] or [12], we just state the result.

Lemma 4.7

For each $k \geqq 0$ and endomorphisms h and g, the set $\mathrm{Com}_k(h, g)$ is regular and can be effectively constructed. □

Finally we collect the above results into the main theorem. □

Theorem 4.8

The limit language equivalence problem is decidable for DOL systems.

Proof.

By Lemma 4.5 we may restrict ourselves to normal 1-systems which have the property P. Let $G_i = (\Sigma, h_i, \sigma)$ be two of such systems for $i = 1, 2$.

We use two semidecision procedures, one for non-equivalence and one for equivalence.

(A) The first semidecision procedure computes $h_1^n(\sigma)$ and $h_2^n(\sigma)$ for $n = 0, 1, \ldots$ and checks for each n if these words are compatible. Thus if G_1 and G_2 are not limit language equivalent then the procedure finds an integer n such that $h_1^n(\sigma)$ and $h_2^n(\sigma)$ are incomparable.

(B) The procedure for equivalence constructs the regular languages $\mathrm{Com}_k(h_1h_2, h_2h_1) = C_k$ for $k = 0, 1, \ldots$ inductively. For each k one checks if

(1) $$L(G_1) \subseteq C_k .$$

The above inclusion can be checked effectively since $L(G_1)$ is a DOL language and C_k is a regular set by Lemma 4.7. If G_1 and G_2 are limit language equivalent, then by Lemma 4.6 the system $G_{12} = (\Sigma, h_1h_2, \sigma)$ and $G_{21} = (\Sigma, h_2h_1, \sigma)$ have bounded balance and thus (1) holds for some $k \geqq 0$. On the other hand if (1) holds for some k then G_1 and G_2 are limit language equivalent since $L(G_1)$ is an infinite prefix-preserving language. □

References

[1] R. Bellman, "Introduction to Matrix Analysis", McGraw-Hill, New York, 1960.

[2] K. Culik II, On the decidability of the sequence equivalence problem for DOL-systems, Theoret. Comput. Sci. 3 (1976), 75-84.

[3] K. Culik II, The ultimate equivalence problem for DOL systems, Acta Inform. 10 (1978),79-84.

[4] K. Culik II and I. Fris, The decidability of the equivalence problem for DOL systems, Inform. and Control, 35 (1977), 20-39.

[5] K. Culik II and T. Harju, "The ω-sequence equivalence problem for DOL-systems is decidable," submitted to JACM, also University of Waterloo, Department of Computer Science Research Report CS-81-02.

[6] K. Culik II and A. Salomaa, "On infinite words obtained by iterating morphisms," submitted to Theoret. Comput. Sci., also University of Waterloo, Department of Computer Science Research Report CS-80-29,revised Feb.81.

[7] A. Ehrenfeucht and G. Rozenberg, "Every two equivalent DOL systems have a regular true envelope," Theoret. Comput. Sci.

[8] S. Eilenberg, "Automata, Languages and Machines", Vol. A, Academic Press, New York, 1974.

[9] F.R. Gantmacher, "The Theory of Matrices," Vol. II, Chelsea, New York, 1959.

[10] M. Linna, The decidability of the DOL prefix problem, Internat. J. of Comput. Math. 6 (1977), 127-142.

[11] M. Nivat, "Infinite words, infinite trees, infinite computations," in J.W. Barleer and J. van Leeuwen (Ed), "Foundations of Computer Science," III. 2, 3-52, Mathematisch Centrum, Amsterdam, 1979.

[12] G. Rozenberg and A. Salomaa, "The Mathematical Theory of L Systems," Academic Press, New York, 1980.

[13] A. Salomaa, "Morphisms of free monoids and language theory," Proc. Conf. on Formal Languages in Santa Barbara, December 1979, Academic Press, in preparation.

[14] A. Salomaa, Jewels in formal language theory, Computer Science Press, in preparation.

[15] A. Thue, "Über unendliche Zeichenreichen," Videnskapsselsk Skrifter I, Kristiania (1906), 1-22.