The ultimate equivalence problem for DOL system*

by

Karel Culik II

Department of Computer Science
University of Waterloo

CS-77-23

September 1977

* This research was supported by the National Research Council of Canada, Grant No. A7403 and most of it has been done during the author's stay at the University of Karlsruhe, Karlsruhe, Germany.

Abstract

The ultimate equivalence problem for DOL systems is shown to be recursively decidable. In algebraic formulation this problem can be stated as follows: Given finite alphabet Σ, two homomorphisms h_1 and h_2 on the free monoid Σ^* and two words w_1, w_2 in Σ^*, does there exist $m \geq 0$ so that $h_1^n(w_1) = h_2^n(w_2)$ for all $n \geq m$?

1. Introduction

In [2] the DOL sequence equivalence problem has been shown to be decidable, i.e. given a finite alphabet Σ, two homomorphisms h_1, h_2 on Σ^* and w_1, w_2 in Σ^*, it is decidable whether $h_1^n(w_1) = h_2^n(w_2)$ for all $n \geq 0$. Since the positive answer is possible only if $w_1 = w_2$ we could simplify the formulation by giving only one starting word. However, this is not the case for the more general ultimate equivalence problem stated as an open problem in [9]. Here we are asking whether the two sequences $w_1, h_1(w_1), h_1^2(w_1), \ldots$ and $w_2, h_2(w_2), h^2(w_2), \ldots$ are ultimately identical, i.e. whether they agree after some arbitrary long initial period. More precisely, we are asking whether there exists $m \geq 0$ so that $h_1^n(w_1) = h_2^n(w_2)$ for all $n \geq m$. A very special case of the problem, namely the case $h_1 = h_2$ has been shown decidable in [3] and for nonerasing homomorphisms in [5].

Both the sequence equivalence and the ultimate equivalence problems for DOL systems are very important for biological applications. DOL systems are the most important of Lindenmayer systems as a useful mathematical model of cellular development. In that context the decidability of our problems means that it is possible to check mechanically whether two developmental programs (genetic encodings) in filamental organisms developing without interaction are equivalent or ultimately equivalent, i.e. whether they determine identical or ultimately identical organisms.

We base our solution on the results and techniques of [1] and [2] showing the decidability of DOL equivalence problem, and on a recent result in [7], [8] or [9] showing that it is decidable whether a semigroup generated by a finite number of matrices is finite. This later result allows us to check whether, in terminology of [2], a pair of normal DOL systems has bounded balance and if so compute the lowest bound. This not only allows a better algorithm for testing DOL-equivalence (see Section 5) but also enables us to give a procedure which terminates if two DOL systems are not ultimately equivalent. To detect ultimate equivalence, i.e. to assure termination if the given systems are ultimately equivalent, is easy when an algorithm for DOL equivalence test is available. However, the termination in the negative case, which is trivial for DOL equivalence, is not easy to assure in the case of ultimate equivalence.

2. Preliminaries

Given an alphabet Σ , Σ^* denotes the free monoid generated by Σ with unit (empty string) ε . $\Sigma^+ = \Sigma^* - \{\varepsilon\}$. For integer w , $|w|$ is its absolute value, for $w \in \Sigma^*$, $|w|$ is the length of w , specifically $|\varepsilon| = 0$.

A *DOL system* is a 3-tuple $G = (\Sigma, h, \sigma)$ consisting of *alphabet* Σ, *homomorphism* h on Σ^*, and *starting string* $\sigma \in \Sigma^+$. DOL system G generates the sequence $s(G) = \sigma, h(\sigma), h^2(\sigma), \ldots$ and the language $L(G) = \{h^n(\sigma) : n \geq 0\}$. System G is *reduced* if there is no $\Delta \subsetneq \Sigma$ so that $L(G) \subseteq \Delta^*$.

Two DOL systems $G_i = (\Sigma_i, h_i, \sigma_i)$ for $i = 1,2$ are *sequence equivalent* (or just *equivalent*) if $s(G_1) = s(G_2)$, they are *language equivalent* if $L(G_1) = L(G_2)$; G_1 and G_2 are *ultimately* (sequence) *equivalent* if there exists $N \geq 0$ so that $h_1^n(\sigma_1) = h_2^n(\sigma_2)$ for all $n \geq N$.

A *DTOL system* is a tuple $G = (\Sigma, h_1, \ldots, h_m, \sigma)$ where Σ is an alphabet, h_i is a homomorphism on Σ^* for $i = 1,\ldots,m$ and $\sigma \in \Sigma^+$ is the starting string. DTOL system G generates the language $L(G) = \{h_{i_1}(h_{i_2}(\ldots h_{i_r}(\sigma) \ldots) : r \geq 0, 1 \leq i_j \leq m \text{ for } 1 \leq j \leq r\}$.

We will now remind the definitions of ℓr-systems and normal systems as given in [2]. For $w \in \Sigma^*$ let $alph(w) = \{a \in \Sigma : a \text{ occurs in } w\}$. Let $G = (\Sigma, h, \sigma)$ be a DOL system. We define the function $m : \mathcal{P}(\Sigma) \to \mathcal{P}(\Sigma)$, where $\mathcal{P}(\Sigma)$ is the set of all subsets of Σ by putting

$$m(\phi) = \phi ,$$

$$m(\{a\}) = alph(h(a)) \quad \text{for } a \in \Sigma ,$$

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

A DOL system $G = (\Sigma, h, \sigma)$ is called an *ℓr-system* if $\Sigma = \Sigma_\ell \cup \Sigma_c \cup \Sigma_r$ is a decomposition of Σ into three nonempty disjoint sets such that for

$$a \in \Sigma_\ell \; , \; h(a) \in \Sigma_\ell \Sigma_c^* \; ;$$
$$a \in \Sigma_c \; , \; h(a) \in \Sigma_c^* \; ;$$
$$a \in \Sigma_r \; , \; h(a) \in \Sigma_c^* \Sigma_r \; ;$$
$$\text{and} \quad \sigma \in \Sigma_\ell \Sigma_c^* \Sigma_r \; .$$

A reduced ℓr-system $G = (\Sigma, h, \sigma)$ is called *normal* if $a \in m^j(b)$ for some $j > 0$ implies $a \in m(b)$ for all $a,b \in \Sigma_c$.

For $L \subseteq \Sigma^+$, let $\mathrm{Pref}(L) = \{w \in \Sigma^+ : wu \in L$ for some $u \in \Sigma^*\}$.

3. The lowest bound on balance

In [2] it was shown that each pair of sequence equivalent normal systems has "bounded balance". We extend this result to ultimately equivalent systems.

Definition Given a pair of homomorphisms h_1, h_2 on Σ^*, the balance of a string w in Σ^* is defined by $B(w) = |h_1(w)| - |h_2(w)|$. Note that the balance β in [2] was defined as $\beta(w) = B(w)$. Whenever a pair of DOL systems is considered the balance is understood with respect to their homomorphisms. Note that B is an additive function.

Definition Let $G_i = (\Sigma, h_i, \sigma_i)$ for $i = 1,2$ be two DOL systems and B the balance with respect to h_1, h_2. We say that the pair G_1, G_2 has bounded balance if there exists $C \geq 0$ such that $|B(w)| \leq C$ for each $w \in \mathrm{Pref}(L(G_1))$. We say that C is the lowest bound if there is no $C' < C$ so that $|B(w)| \leq C'$ for each $w \in \mathrm{Pref}(L(G_1))$.

Theorem 1 Let $G_i = (\Sigma, h_i, \sigma_i)$ for $i = 1,2$ be a pair of normal ultimately equivalent systems. Then the pair G_1, G_2 has bounded balance.

Proof Since G_1, G_2 are ultimately equivalent there exists $m \geq 0$ so that G_1', G_2' are sequence equivalent where $G_i' = (\Sigma , h_i, h_i^m(\sigma_i))$ for $i = 1,2$. Therefore the balance is bounded on $\mathrm{Pref}(L(G_1'))$ by Theorem 3 in [2]. The balance is also bounded on finitely many prefixes of $\{h_1^k(\sigma_1) : 0 \leq k < m\}$ hence it is bounded on $\mathrm{Pref}(L(G_1))$.

□

Lemma 1 Let $G_i = (\Sigma, h_i, \sigma_i)$ for $i = 1,2$ be two DOL systems. The balance is bounded on all prefixes of $L(G_1)$ iff it is bounded on all substrings of $L(G_1)$.

Proof Assume that $B(w) \leq C$ for all $w \in \mathrm{Pref}(L(G_1))$ but the balance B is unbounded on substrings of $L(G_1)$. Then there must exist $x,y,z \in \Sigma^*$ so that $xyz \in L(G_1)$ and $|B(y)| > 2C$. By additivity of B we have $|B(xy)| \geq |B(y)| - |B(x)|$ and since $B(x) \leq C$ we get $|B(xy)| > 2C - C = C$ a contradiction.

□

We continue by showing that for every DOL system G a DTOL system $\tau(G)$ can be constructed generating all prefixes of $L(G)$.

Definition Let $G = (\Sigma, h, \sigma)$ be a DOL system with $m = \max(\{|h(a)| : a \in \Sigma\} \cup \{|\sigma|\})$. Let $\overline{\Sigma} = \{\overline{a} : a \in \Sigma\}$, $\Sigma' = \Sigma \cup \overline{\Sigma} \cup \{s\}$ for $s \notin \Sigma$. For $i = 1,\ldots,m$ and $w \in \Sigma^*$, $w = w_1w_2\ldots w_s$, $w_k \in \Sigma$, $1 \leq k \leq s$, let $\mu_i(w) = w_1w_2\ldots w_{s-1}\overline{w_s}$ if

$s \leq i$, $\mu_i(w) = w_1 w_2 \ldots w_{i-1}\overline{w}_i$ if $s \geq i$. Finally, let $\tau(G)$ be the DTOL system $(\Sigma', h_1, \ldots, h_m, s)$ where for all $i = 1,\ldots,m$

(i) $h_i(a) = a$ for $a \in \Sigma$,

(ii) $h_i(\overline{a}) = \mu_i(h(a))$ for $a \in \Sigma$,

(iii) $h_i(s) = \mu_i(\sigma)$.

<u>Example</u> Let $G = (\{a,b\}, h, ab)$ where $h(a) = aba$, $h(b) = b$. Then $\tau(G) = (\{a,b,\overline{a},\overline{b},s\}, h_1, h_2, h_3, s)$ where

$$h_i(a) = aba \quad \text{for} \quad i = 1,2,3$$
$$h_i(b) = b \quad \text{for} \quad i = 1,2,3$$
$$h_1(\overline{a}) = \overline{a} \ , \ h_2(\overline{a}) = a\overline{b} \ , \quad h_3(\overline{a}) = ab\overline{a}$$
$$h_i(\overline{b}) = \overline{b} \quad \text{for} \quad i = 1,2,3$$
$$h_1(s) = \overline{a} \ , \ h_i(s) = a\overline{b} \quad \text{for} \quad i = 2,3 \ .$$

<u>Lemma 2</u> Let G be a DOL system and $L(G) \subseteq \Sigma^+$. Let g be the homomorphism defined by $g(a) = g(\overline{a}) = a$ for all $a \in \Sigma$. Then $g(L(\tau(G))) = \text{Pref}(L(G))$.

<u>Proof</u> It is easy to verify by induction that the DTOL system $\tau(G)$ generates, when bars are ignored, in $k+1$ steps exactly all prefixes of $h^k(\sigma)$. □

Now , we are ready for the crucial auxiliary result.

<u>Theorem 2</u> Given two normal systems $G = (\Sigma, h, \sigma)$ and $G' = (\Sigma, h', \sigma')$, it is decidable whether the pair G_1, G' has bounded balance and if so, the lowest bound can be effectively computed.

Proof Consider $\tau(G) = (\Sigma', h_1, \ldots, h_m, s)$ as defined above. We choose a fixed ordering of Σ', let $\Sigma' = \{a_1, \ldots, a_t\}$, $a_1 = s$. We extend h and h' to Σ' be defining $h(\bar{a}) = h(a)$, $h'(\bar{a}) = h'(a)$ for $a \in \Sigma$, $h(s) = \sigma_1$, and $h'(s) = \sigma'$. Hence, $B(\bar{a}) = B(a) = |h(a)| - |h'(a)|$ for $a \in \Sigma$ and $B(s) = |\sigma| - |\sigma'|$, and let η be a column vector $\eta = (B(a_1), \ldots, B(a_t))$. Further, let M_i be the growth matrix (see [9]) of the DOL system (Σ', h_i, s) for $i = 1,\ldots,m$. As an abbreviation, for $v \in \{1,\ldots,m\}^+$, $v = v_1 \ldots v_r$, $1 \le v_j \le m$, for $j = 1,\ldots,r$, we define $H^v(x) = h_{v_r}(\ldots h_{v_1}(x) \ldots)$ for each $x \in \Sigma^*$ and $M^v = M_{v_1} M_{v_2} \ldots M_{v_r}$.

By Lemma 2, for each w in Pref(L(G)) there exists $v \in \{1,\ldots,m\}^+$ such that $\bar{w} = H^v(s)$ where $\bar{w}$ is w with bar over the last symbol. Therefore the Parikh vector of w with respect to Σ' is $(1, 0, \ldots, 0)M^v$, and thus $B(w) = (1, 0, \ldots, 0)M^v\eta$. In terminology of [9] we have exhibited a Z-rational function whose coefficients are exactly the balances of all the prefixes of $L(G)$. Therefore the pair G, G' has bounded balance iff the corresponding Z-rational function has finitely many distinct coefficients. This problem has shown to be decidable in [6].

It remains to show that we can effectively compute the lowest bound in the case the pair G_1, G_2 has bounded balance. Let $B^v = (B_1^v, \ldots, B_t^v)^T = M^v\eta$ for arbitrary $v \in \{1, \ldots, m\}^+$. We already know that $B_1^v = B(H^v(s))$. Similarly, $B_i^v = B(H^v(a_i))$ for $i = 2,\ldots,t$. By Lemma 1 the balance is bounded on all prefixes iff it is bounded on all substrings. Since G is reduced, $H^v(a)$ with the bars removed for each $a \in \Sigma$ is a substring of $L(G)$, therefore the set $\{B_i^v : 1 \le i \le t,\ v \in \{1,\dots,m\}^+\}$ is bounded and thus there is only finitely many distinct vectors B^v for all $v \in \{1,\dots,m\}^+$. We can easily find the finite set $B = \{B^v : v \in \{1,\dots,m\}^+\}$ as the closure of $\{\eta\}$ under multiplication from left by matrices $M_1, \dots, M_m$. Finally, the lowest bound C is obtained as $C = \max\{|B_1| : (B_1, \dots, B_t) \in B\}$.

□

Note that we could have omitted the argument using the Z-rational function and started by computing set B. We have shown that B is finite iff the balance is bounded and by [6] or [7] or [8] the finiteness of B is decidable.

4. Ultimate equivalence problem

Lemma 3 The ultimate equivalence problem is decidable for DOL systems iff it is decidable for normal systems.

Proof An obvious modification of the proof of Theorem 1 in [2].

Theorem 3 The ultimate (sequence) equivalence problem for DOL systems is decidable.

Proof In view of Lemma 3 we can restrict ourselves without loss of generality to normal systems, clearly we may also assume that they are over the same alphabet Σ. Given two normal systems $G_i = (\Sigma, h_i, \sigma_i)$, for $i = 1,2$, we test first whether the pairs (G_1, G_2) and (G_2, G_1) have bounded balance which is decidable by Theorem 2. If either pair has unbounded balance the systems are not ultimately equivalent by Theorem 1. If both pairs have bounded balance we compute the lowest bounds by Theorem 2, let C be their maximum.

Now, we construct the deterministic g.s.m. M_C , i.e. with buffer of length C , from the proof of Theorem 2.1 in [1]. Using further the notation from [1], we compute the language $T_C(L(G_1))$ where T_C is the translation defined by M_C . We note that G_1 and G_2 are equivalent iff $T_C(L(G_1)) = \phi$ which is decidable since $T_C(L(G_1))$ is a g.s.m. image of a DOL language and therefore it is generated by an EOL system which can be effectively constructed. We check $T_C(L(G_i))$, $i = 1,2$ for finiteness which is also decidable (see [4]). If $T_C(L(G_1))$ or $T_C(L(G_2))$ is infinite, then $h_1(x) \neq h_2(x)$ for infinitely many x in $L(G_1)$ or $L(G_2)$ and therefore G_1 and G_2 are, clearly, not ultimately equivalent. If $T_C(L(G_1))$ and $T_C(L(G_2))$ are both finite, then there is $p \geq 0$ such that $h_1(x) = h_2(x)$ for all $x \in \{h_1^k(\sigma_1) : k \geq p\} \cup \{h_2^k(\sigma_2) : k \geq p\}$. Since the DOL equivalence problem is decidable [2], we can effectively find the smallest such p , namely the smallest $p \geq 0$ such that the DOL systems $G_1^{p,i}$ and $G_2^{p,i}$ are equivalent for $i = 1,2$, where $G_j^{p,i} = (\Sigma, h_j, h_i^p(\sigma_i))$ for $1 \leq i,j \leq 2$.

Now, clearly G_1 and G_2 are ultimately equivalent iff there exists $m \geq p$ such that $h_1^m(\sigma_1) = h_2^m(\sigma_2)$. For $m \geq p$, $h_2^m(\sigma_2) = h_2^{m-p}(h_2^p(\sigma_2)) = h_1^{m-p}(h_2^p(\sigma_2))$, therefore the required m exists iff there exists $k \geq 0$ such that $h_1^k(h_1^p(\sigma_1)) = h_1^k(h_2^p(\sigma_1))$. The existence of such k is decidable by Theorem 1 of [3].

□

<u>Theorem 4</u> Given alphabet Σ , homomorphisms h_1, h_2 on Σ^* and strings σ_1, σ_2 in Σ_1^* it is decidable whether there exists $m,n \geq 0$ so that $h_1^{m+k}(\sigma_1) = h_2^{n+k}(\sigma_2)$ for all $k \geq 0$.

<u>Proof</u> Similarly like for the ultimate equivalence problem (Lemma 3) we can also here restrict ourselves without loss of generality to normal systems. Then we proceed exactly as in the proof of Theorem 3 only the last paragraph need to be modified as follows.

Now, there exist $m,n \geq 0$ so that $h_1^{m+k}(\sigma_1) = h_2^{n+k}(\sigma_2)$ for all $k \geq 0$ iff there exist $m,n \geq p$ so that $h_1^m(\sigma_1) = h_2^n(\sigma_2)$. For $n \geq p$, $h_2^n(\sigma_2) = h_2^{n-p}(h_2^p(\sigma_2)) = h_1^{n-p}(h_2^p(\sigma_2))$, therefore we are asking whether there exist $m,n \geq p$ such that $h_1^{m-p}(h_1^p(\sigma_1)) = h_1^{n-p}(h_2^p(\sigma_2))$. This is decidable by Theorem 2 of [3].

□

5. DOL-equivalence problem

As it is clear from the previous section, the possibility to compute the best bound on balance gives also a simpler algorithm than the one given in [2] for testing equivalence of DOL systems G_1, G_2. For each pair of normal systems we compute the lowest bound C, construct g.s.m. machine M_C and then test $T_C(L(G_1))$ for emptiness.

However, this does not simplify essentially the proof of decidability of the DOL-equivalence problem since the very difficult result that equivalence implies bounded balance (Theorem 3 in [2]) is still needed.

Acknowledgements

The author is grateful to A. Salomaa for remarks to the draft version of this paper.

References

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

2. Culik II, K., Fris, I.: The sequence equivalence problem for DOL systems is decidable, Inf. and Control, in print.

3. Culik II, K.: The decidability of v-local catenativity and of other properties of DOL systems, Information Processing Letters, in print.

4. Culik II, K., Maurer, H.A.: Propagating chain-free normal forms for EOL systems, Inform. and Control, in print.

5. Ehrenfeucht, A., Rozenberg, G.: On Simplifications of PDOL Systems, Proceedings of a Conference on Theoretical Computer Science, University of Waterloo, Waterloo, Ontario, Canada, August 1977.

6. Jacob, G.: Decidabilite de la finitude des demi-groupes de Matrices, Proceedings of the third G.I. Conference on Theoretical Computer Science, Darmstadt, March 1977, Lecture Notes in Computer Science, Springer Verlag.

7. Jacob, G.: Un algorithme calculant le cardinal, fini ou infini, des demi-groupes de matrices, Theoretical Computer Science, in print.

8. Mandel, A., Simon, I.: On finite semigroups of matrices, Research Report, Instituto de Matematica e Statistika, Universidade de Sao Paulo, Brasil.

9. Salomaa, A., Soittola, M.: Automata-Theoretic Aspects of Formal Power Series, Springer Verlag, in print.