On D0L systems with finite axiom sets

Juha Honkala*

Abstract

We give a new solution for the language equivalence problem of D0L systems with finite axiom sets by using the decidability of the equivalence problem of finite valued transducers on HDT0L languages proved by Culik II and Karhumäki.

1 Introduction

The language equivalence problem for D0L systems with finite axiom sets was solved in [4]. The problem turns out to be much more difficult for DF0L systems than for D0L systems. The main idea in [4] is to decompose a given DF0L language in a canonical way into finitely many parts such that no part contains two words with equal Parikh vectors. This makes it possible to use ideas from [8]. The resulting algorithm gives a lot of information concerning the structures of the languages generated by two equivalent DF0L systems. Also the equivalence problem for DF0L power series over a computable field is solved in [4].

The purpose of this paper is to give a new solution of the DF0L language equivalence problem. The new proof for the decidability of the problem avoids many difficulties in [4] but fails to give precise information about language equivalent DF0L systems. In that respect it resembles the solutions of the D0L equivalence problem based on Hilbert's basis theorem which also are short but do not, for example, give any bounds for the problem (see [3]).

Our new solution again uses methods from [8] which in turn use ideas from [1]. In addition, we use the decidability of the equivalence problem of finite valued transducers proved by Culik II and Karhumäki [2]. In this way we obtain a solution of the DF0L language equivalence problem which is essentially based on commutative methods (see [5]).

For further background and motivation we refer to [6, 7, 8, 9, 10, 4]. It is assumed that the reader is familiar with the basics concerning D0L systems and their generalizations such as HDT0L systems, see [6, 7].

*Department of Mathematics, University of Turku, FIN-20014 Turku, Finland. Email: juha.honkala@utu.fi
Research supported by the Academy of Finland.

2 Definitions and earlier results

Let $X = \{x_1, \ldots, x_k\}$ be an alphabet with $k \geq 1$ letters. The *Parikh mapping* $\psi : X^* \longrightarrow \mathbf{N}^k$ is defined by

$$\psi(w) = (\#_{x_1}(w), \ldots, \#_{x_k}(w)),$$

for $w \in X^*$. Here $\#_{x_i}(w)$ is the number of occurrences of the letter x_i in the word w. The *length* of a word w is denoted by $|w|$. The length of the empty word ε equals zero.

A *D0L system* is a triple $G = (X, h, w)$ where X is a finite alphabet, $h : X^* \longrightarrow X^*$ is a morphism and $w \in X^*$ is a word. A DF0L system is obtained from a D0L system by replacing the word w by a finite set F. Hence, a *DF0L system* is a triple $G = (X, h, F)$ where X is a finite alphabet, $h : X^* \longrightarrow X^*$ is a morphism and $F \subseteq X^*$ is a finite set.

The *sequence* $S(G)$ and the *language* $L(G)$ of the D0L system $G = (X, h, w)$ are given by

$$S(G) = (h^n(w))_{n \geq 0}$$

and

$$L(G) = \{h^n(w) \mid n \geq 0\}.$$

The *language* $L(G)$ of the DF0L system $G = (X, h, F)$ is defined by

$$L(G) = \{h^n(w) \mid w \in F,\ n \geq 0\}.$$

Below we will discuss also DT0L and HDT0L systems. By definition, a *DT0L system* is a construct $(X, h_1, \ldots, h_n, w)$ such that $n \geq 1$ is an integer and (X, h_i, w) is a D0L system for $1 \leq i \leq n$. An *HDT0L system* is a construct $G = (X, Y, h_1, \ldots, h_n, h, w)$ such that $(X, h_1, \ldots, h_n, w)$ is a DT0L system (called the *underlying DT0L system* of G), Y is a finite alphabet and $h : X^* \longrightarrow Y^*$ is a morphism.

Let $G = (X, Y, h_1, \ldots, h_n, h, w)$ be an HDT0L system and let $Z_n = \{z_1, \ldots, z_n\}$ be an alphabet with n letters. Then the *sequence* of G is the mapping $S(G) : Z_n^* \longrightarrow Y^*$ defined by

$$S(G)(z_{i_1} \ldots z_{i_m}) = h h_{i_m} \ldots h_{i_1}(w)$$

for $m \geq 0$, $1 \leq i_1, \ldots, i_m \leq n$. The *sequence* of a DT0L system $(X, h_1, \ldots, h_n, w)$ equals the sequence of the HDT0L system $(X, X, h_1, \ldots, h_n, g, w)$ where the morphism $g : X^* \longrightarrow X^*$ is defined by $g(x) = x$ for all $x \in X$.

A *finite transducer* is a construct $\tau = (Q, \Sigma, \Delta, s_0, F, E)$ where Q is the finite set of states, Σ and Δ are the input and output alphabets, respectively, $s_0 \in Q$ is the initial state, $F \subseteq Q$ is the set of final states and $E \subseteq Q \times \Sigma^* \times \Delta^* \times Q$ is the finite set consisting of the transitions of τ. If $u \in \Sigma^*$ and $v \in \Delta^*$ we write $v \in \tau(u)$ if there is an accepting computation of τ having input u and output v. Let k be a nonnegative integer. A transducer τ is called *k-valued* if for all $u \in \Sigma^*$ the set

$\tau(u)$ contains at most k words. Finally, a transducer τ is called *finite valued* if it is k-valued for some k.

The following important result is due to Culik II and Karhumäki, [2]. Here two transducers τ_1 and τ_2 are called equivalent on a language L if $\tau_1(u) = \tau_2(u)$ for all $u \in L$.

Theorem 1. *It is decidable whether two finite valued finite transducers are equivalent on a given HDT0L language.*

3 The HDT0L covering problem

In this section we discuss the HDT0L covering problem which is a useful tool in the study of the DF0L language equivalence problem. It would suffice to consider the D0L covering problem but this would not simplify the discussion.

Let $H_i = (X_i, Y_i, h_{i1}, \dots, h_{in}, h_i, w_i)$, $1 \le i \le k+1$, be HDT0L systems. Then we say that the first k sequences $S(H_i)$ *cover* the last sequence $S(H_{k+1})$ if

$$S(H_{k+1})(u) \in \{S(H_i)(u) \mid 1 \le i \le k\}$$

for all $u \in Z_n^*$. If $k = 1$, then $S(H_1)$ covers $S(H_2)$ if and only if H_1 and H_2 are sequence equivalent. If $k > 1$, the covering relation generalizes sequence equivalence by allowing finitely many alternatives for each term of $S(H_{k+1})$.

Let H_i, $1 \le i \le k+1$, be as above. By the *HDT0L covering problem* we understand the problem of deciding whether or not $S(H_i)$, $1 \le i \le k$, cover $S(H_{k+1})$. To reduce the covering problem to the equivalence problem of finite valued transducers one lemma is required.

Lemma 2. *Let $H_i = (X_i, Y_i, h_{i1}, \dots, h_{in}, h_i, w_i)$, $1 \le i \le k$, be HDT0L systems. Then there is a DT0L system $H = (X, f_1, \dots, f_n, w)$ and finite valued finite transducers τ_I for $I \subseteq \{1, \dots, k\}$ such that*

$$\tau_I(S(H)(u)) = \{S(H_i)(u) \mid i \in I\} \tag{1}$$

for all $u \in Z_n^*$.

Proof. We may assume that the alphabets X_i, $1 \le i \le k$, are pairwise disjoint. Denote $X = X_1 \cup \ldots \cup X_k$, $Y = Y_1 \cup \ldots \cup Y_k$ and let $f_j : X^* \longrightarrow X^*$ be the morphism such that

$$f_j(x) = h_{ij}(x)$$

whenever $x \in X_i$, $1 \le i \le k$, $1 \le j \le n$. Denote $w = w_1 \ldots w_k$ and consider the DT0L system $H = (X, f_1, \dots, f_n, w)$.

Let $\overline{H}_i = (X_i, h_{i1}, \dots, h_{in}, w_i)$ be the underlying DT0L system of H_i, $1 \le i \le k$. Then we have

$$S(H)(u) = S(\overline{H}_1)(u) \ldots S(\overline{H}_k)(u) \tag{2}$$

for $u \in Z_n^*$.

Let now $I \subseteq \{1, \ldots, k\}$ be a nonempty set and let τ_I be a transducer defined as follows. The input alphabet of τ_I is X and the output alphabet of τ_I is Y. The state set of τ_I is $\{q_0\} \cup \{q_i \mid i \in I\}$ where q_0 is the initial state and $\{q_i \mid i \in I\}$ is the final state set. The set E of transitions is defined by

$$\begin{aligned} E &= \{(q_0, \varepsilon, \varepsilon, q_i) \mid i \in I\} \cup \\ &\quad \{(q_i, x, h_i(x), q_i) \mid i \in I \text{ and } x \in X_i\} \cup \\ &\quad \{(q_i, x, \varepsilon, q_i) \mid i \in I \text{ and } x \notin X_i\}. \end{aligned}$$

Then τ_I is finite valued and (2) implies (1) for all $u \in Z_n^*$. □

Theorem 3. *The HDT0L covering problem is decidable.*

Proof. Let $H_i = (X_i, Y_i, h_{i1}, \ldots, h_{in}, h_i, w_i)$, $1 \le i \le k+1$, be HDT0L systems. Denote $I = \{1, \ldots, k\}$ and $J = \{1, \ldots, k+1\}$. By Lemma 2 there exist a DT0L system $H = (X, f_1, \ldots, f_n, w)$ and finite valued finite transducers τ_I and τ_J such that

$$\tau_I(S(H)(u)) = \{S(H_i)(u) \mid i \in I\}$$

and

$$\tau_J(S(H)(u)) = \{S(H_j)(u) \mid j \in J\}$$

for all $u \in Z_n^*$. Now

$$\tau_I(S(H)(u)) = \tau_J(S(H)(u)) \text{ for all } u \in Z_n^* \qquad (3)$$

if and only if

$$S(H_{k+1})(u) \in \{S(H_i)(u) \mid 1 \le i \le k\} \text{ for all } u \in Z_n^*.$$

The claim follows because by Theorem 1 we can decide the validity of (3). (Here we use Theorem 1 for DT0L languages.) □

4 The DF0L language equivalence problem

Let X be an alphabet with $k \ge 1$ letters and let $\psi : X^* \longrightarrow \mathbb{N}^k$ be the Parikh mapping. If $K \subseteq \mathbb{N}^k$ we denote

$$\psi^{-1}(K) = \{w \in X^* \mid \psi(w) \in K\}.$$

Lemma 4. *Let $G = (X, h, F)$ be a DF0L system and let $u \in F$. Assume that $\{h^i(u) \mid i \ge 0\}$ is an infinite set. Then there exist an integer $s \ge 0$, integers $n_1, \ldots, n_s$ and words $u_1, \ldots, u_s \in F$ such that*

$$\psi^{-1}(\psi h^n(u)) \cap L(G) = \{h^{n+n_1}(u_1), \ldots, h^{n+n_s}(u_s)\}$$

for almost all $n \ge 0$.

Proof. We will show that if $v \in F$ then either

$$\psi^{-1}(\psi h^n(u)) \cap \{h^i(v) \mid i \geq 0\} = \emptyset \tag{4}$$

for almost all $n \geq 0$ or, otherwise, there exists an integer m such that

$$\psi^{-1}(\psi h^n(u)) \cap \{h^i(v) \mid i \geq 0\} = \{h^{n+m}(v)\} \tag{5}$$

for almost all $n \geq 0$. (Here and in the sequel we say that a property holds for almost all n if there is an integer n_0 such that the property holds for all $n \geq n_0$.)

First, if $\{h^i(v) \mid i \geq 0\}$ is a finite set then (4) holds for almost all $n \geq 0$. Suppose $\{h^i(v) \mid i \geq 0\}$ is infinite. Then

$$\psi h^i(v) \neq \psi h^j(v) \quad \text{if} \quad i \neq j.$$

Now, if there exist integers m_1 and m_2 such that

$$\psi h^{m_1}(u) = \psi h^{m_2}(v) \tag{6}$$

then (5) holds for almost all $n \geq 0$ if we set $m = m_2 - m_1$. Finally, if (6) holds for no values of m_1 and m_2 then (4) holds for all $n \geq 0$. □

Let $G = (X, h, F)$ be a DF0L system. A word sequence $(w_n)_{n \geq 0}$ is called a *subsequence* of G if there exist $w \in L(G)$ and a positive integer a such that

$$w_n = h^{an}(w)$$

for all $n \geq 0$. In Section 3 we have explained what it means that a given D0L sequence is covered by finitely many given D0L sequences. We now define this notion for DF0L systems.

Let $G_i = (X, h_i, F_i)$, $i = 1, 2$, be DF0L systems. Then G_2 is said to *cover* G_1 if for all $u \in F_1$ there exist a nonnegative integer r and a positive integer k such that for all integers j, $0 \leq j < k$, the sequence $(h_1^{kn+j+r}(u))_{n \geq 0}$ is covered by finitely many subsequences of G_2.

Lemma 5. *Let $G_i = (X, h_i, F_i)$, $i = 1, 2$, be DF0L systems. Assume that $L(G_1) = L(G_2)$ and that alph$(w) = X$ for all $w \in L(G_1)$. Then G_1 and G_2 cover each other.*

Proof. Let $G_i = (X, h_i, F_i)$, $i = 1, 2$, be DF0L systems such that $L(G_1) = L(G_2)$ and $\text{alph}(w) = X$ for all $w \in L(G_1)$. If $L(G_1)$ is finite the claim holds. Assume that $L(G_1)$ is infinite. Without restriction assume also that $\text{card}(F_1) = \text{card}(F_2)$. (If necessary, we replace F_1 by the set $\{h_1^j(u) \mid u \in F_1, 0 \leq j < \text{card}(F_2)\}$ and F_2 by the set $\{h_2^j(v) \mid v \in F_2, 0 \leq j < \text{card}(F_1)\}$.) Denote $t = \text{card}(F_1)$,

$$F_1 = \{u_0, \ldots, u_{t-1}\}$$

and

$$F_2 = \{v_0, \ldots, v_{t-1}\}.$$

Further, denote $k = \text{card}(X)$ and let $P(x_1, \ldots, x_k)$ be a polynomial with nonnegative integer coefficients such that the mapping $P : \mathbf{N}^k \longrightarrow \mathbf{N}$ is injective (see [8]). Define the mappings $f : \mathbf{N} \longrightarrow \mathbf{N}$ and $g : \mathbf{N} \longrightarrow \mathbf{N}$ by

$$f(ti + j) = P(\psi h_1^i(u_j))$$

and

$$g(ti + j) = P(\psi h_2^i(v_j))$$

for $i \geq 0$ and $0 \leq j < t$. Then f and g are D0L growth functions (see [8]) and

$$\{f(n) \mid n \in \mathbf{N}\} = \{g(n) \mid n \in \mathbf{N}\}.$$

Hence there exist integers $a \geq 1$, $r \geq 0$, $x_k \geq 1$ and $y_k \geq 0$ for $0 \leq k < a$ such that

$$f(an + k + r) = g(x_k n + y_k)$$

for $n \geq 0$, $0 \leq k < a$ (see [1]). Without restriction we assume that t divides a and that t divides x_k for all $0 \leq k < a$. Denote $a = bt$. Fix $u \in F_1$. It follows that there is an integer $\beta \geq 0$ such that for all integers α, $0 \leq \alpha < b$, there exist $v_{j_\alpha} \in F_2$ and integers $q_\alpha \geq 1$, $p_\alpha \geq 0$ such that

$$\psi h_1^{bn+\alpha+\beta}(u) = \psi h_2^{q_\alpha n + p_\alpha}(v_{j_\alpha})$$

for $n \geq 0$. Because $L(G_1) = L(G_2)$ we have

$$h_1^{bn+\alpha+\beta}(u) \in \psi^{-1}(\psi h_2^{q_\alpha n+p_\alpha}(v_{j_\alpha})) \cap L(G_2)$$

for $n \geq 0$.

Next, fix α, $0 \leq \alpha < b$. Because $\text{alph}(v_{j_\alpha}) = X$, the set $\{h_2^i(v_{j_\alpha}) \mid i \geq 0\}$ is infinite. By Lemma 4 there exist an integer $s \geq 0$, integers $n_1, \ldots, n_s$ and words $w_1, \ldots, w_s \in F_2$ such that

$$\psi^{-1}(\psi h_2^{q_\alpha n+p_\alpha}(v_{j_\alpha})) \cap L(G_2) = \{h_2^{q_\alpha n+p_\alpha+n_1}(w_1), \ldots, h_2^{q_\alpha n+p_\alpha+n_s}(w_s)\}$$

for almost all $n \geq 0$. Hence

$$h_1^{bn+\alpha+\beta}(u) \in \{h_2^{q_\alpha n+p_\alpha+n_1}(w_1), \ldots, h_2^{q_\alpha n+p_\alpha+n_s}(w_s)\}$$

for almost all $n \geq 0$. In other words, G_2 covers G_1. It is seen similarly that G_1 covers G_2. □

Theorem 6. *It is decidable whether or not two given DF0L systems are language equivalent.*

Proof. It suffices to consider DF0L systems $G = (X, h, F)$ such that $\text{alph}(w) = X$ for all $w \in L(G)$ (see [8]). The claim follows because there exists a semialgorithm for equivalence and there exists a semialgorithm for nonequivalence. The existence of a semialgorithm for equivalence follows by Theorem 3 and Lemma 5. (Here we use Theorem 3 for D0L systems.) The existence of a semialgorithm for nonequivalence is clear. □

References

[1] J. Berstel and M. Nielsen, The growth range equivalence problem for D0L systems is decidable, in A. Lindenmayer and G. Rozenberg (eds.): *Automata, Languages, Development* (North-Holland, Amsterdam, 1976) 161-178.

[2] K. Culik II and J. Karhumäki, The equivalence of finite valued transducers (on HDT0L languages) is decidable, *Theoret. Comput. Sci.* **47** (1986) 71-84.

[3] J. Honkala, A short solution for the HDT0L sequence equivalence problem, *Theoret. Comput. Sci.* **244** (2000) 267-270.

[4] J. Honkala, The equivalence problem for DF0L languages and power series, *J. Comput. System Sci.* **65** (2002) 377-392.

[5] J. Honkala, A note on systems of alternative word equations, *Bull. EATCS* **78** (2002) 237-240.

[6] G. Rozenberg and A. Salomaa, *The Mathematical Theory of L Systems* (Academic Press, New York, 1980).

[7] G. Rozenberg and A. Salomaa (eds.): *Handbook of Formal Languages*, Vol. 1-3 (Springer, Berlin, 1997).

[8] K. Ruohonen, The decidability of the F0L-D0L equivalence problem, *Inform. Process. Letters* **8** (1979) 257-260.

[9] K. Ruohonen, On a variant of a method of Berstel's and Nielsen's, *Fund. Inform.* **4** (1981) 369-400.

[10] K. Ruohonen, Equivalence problems for regular sets of word morphisms, in G. Rozenberg and A. Salomaa (eds.), *The Book of L* (Springer, Berlin, 1986) 393-401.

Received August, 2002