Note

Easy cases of the D0L sequence equivalence problem

Juha Honkala\textsuperscript{a,b,*}\textsuperscript{1}

\textsuperscript{a}Department of Mathematics, University of Turku, FIN-20014 Turku, Finland
\textsuperscript{b}Turku Centre for Computer Science (TUCS), Lemminkäisenkatu 14, FIN-20520 Turku, Finland

Received 10 August 2000; received in revised form 16 October 2000; accepted 6 November 2000

Abstract

To test the equivalence of two binary D0L sequences it suffices to compare the first four terms of the sequences. We introduce a larger class of D0L systems for which sequence equivalence can be decided by considering the first ten initial terms.

\textsuperscript{*}Correspondence address: Department of Mathematics, University of Turku, 20014 Turku, Finland.
\textsuperscript{E-mail address:} juha.honkala@utu.fi (J. Honkala).
\textsuperscript{1}Research supported by the Academy of Finland.

Keywords: D0L systems; Equivalence problem; Decidability

1. Introduction

The D0L sequence equivalence problem was solved by Culik II and Fris [2] in the 1970s. Later different other solutions have been found [3,4,6,11]. However, the D0L sequence equivalence problem is still one of the most intriguing problems in the theory of free monoids. This is illustrated by the fact that in the general case all known algorithms to guarantee the equivalence of two given D0L sequences by comparing initial terms of the sequences require a prohibitive amount of work (see[5]). On the other hand it is possible that it suffices to test $2n$ initial terms of the sequences where $n$ is the cardinality of the alphabet (see [12]). At least, no counterexamples are known. The claim that this simple algorithm always gives the correct answer is called the $2n$-conjecture. The validity of the $2n$-conjecture has been shown by Karhumäki [7] if $n = 2$. All other cases remain open.

In this note, we give new simple cases of the D0L sequence equivalence problem.

If $G = (X,h,w)$ is a D0L system we define the \textit{rank} of $G$ to be the minimal number of words required to build up all the words $h(x)$, $x \in X$. More precisely, the rank of $G$ is the smallest number $k$ such that there is a $k$-element set $L$ with $\{h(x) | x \in X \} \subseteq L^*$. 
Below we will always use this notion of rank. In the previous literature concerning L
systems also different definitions of rank are used (see [9]). We show that if G and
H are D0L systems and G or H has rank at most two then G and H are sequence
equivalent if and only if the first ten words in S(G) and S(H) are equal. In particular,
if G and H are D0L systems over a three letter alphabet and at least one of G and
H is not elementary, then S(G) = S(H) if and only if the first ten terms in S(G) and
S(H) coincide.

It is assumed that the reader is familiar with D0L systems and elementary morphisms
(see [4,9,10]). In the proofs we will use results from Karhumäki [7].

2. The main result

A D0L system is a triple G = (X, h, w) where X is a finite alphabet, h : X* → X*
is a morphism and w ∈ X* is a word. The sequence S(G) generated by G consists of the
words
w, h(w), h2(w), h3(w), ... .
The language L(G) generated by G is defined by
\[ L(G) = \{ h^i(w) | i \geq 0 \} \]
Two D0L systems G and H are sequence equivalent if S(G) = S(H). The following
important result is due to Karhumäki [7].

Theorem 1. Suppose G = (\{0, 1\}, g, w) and H = (\{0, 1\}, h, w) are D0L systems. Then
the following conditions are equivalent:
(i) G and H are sequence equivalent,
(ii) g^i(w) = h^i(w) for i = 0, 1, 2, 3.

We will also need the following closely related result from Karhumäki [7].

Theorem 2. Suppose G = (\{0, 1\}, f, w) is a D0L system. If X is an arbitrary alphabet and
g : \{0, 1\}^* → X* and h : \{0, 1\}^* → X* are morphisms, then the following
conditions are equivalent:
(i) g and h agree on the language generated by G,
(ii) g(f^i(w)) = h(f^i(w)) for i = 0, 1, 2, 3.

Suppose now that L ⊆ X* is an arbitrary language. Then the combinatorial rank or
degree d(L) of L is defined by
\[ d(L) = \min \{ \text{card}(L_i) | L \subseteq L_1^i \} \]
(see [1]). Clearly \( d(L) \leq \text{card}(X) \). Motivated by this notion of rank we define the rank \( d(G) \) of the D0L system \( G = (X, h, w) \) by
\[
d(G) = d(\{ h(x) \mid x \in X \}).
\]

We will now state and prove the main result.

**Theorem 3.** Suppose \( G = (X, g, w) \) and \( H = (X, h, w) \) are D0L systems. If \( d(G) \leq 2 \) or \( d(H) \leq 2 \) then the following conditions are equivalent:

(i) \( S(G) = S(H) \),

(ii) \( g^i(w) = h^i(w) \) for \( i = 0, 1, 2, \ldots, 9 \).

**Proof.** Suppose \( d(G) \leq 2 \). If (i) holds, (ii) holds. Suppose that (ii) holds. Then we have
\[
g^2 g(w) = h^2 g(w)
\]
for \( i = 0, 1, 2, 3, 4 \). Let \( Y \) be a binary alphabet. Because \( d(G) \leq 2 \) there exist morphisms \( g_1 : X^* \to Y^* \) and \( g_2 : Y^* \to X^* \) such that
\[
g = g_2 g_1.
\]
By (1) and (3) we have
\[
(g_2 g_1) g(w) = h g_2 h g_1 g(w)
\]
for \( i = 0, 1, 2, 3, 4 \). By applying \( g_1 \) on both sides of (4) we obtain
\[
g_1 g_2 g_1 g_1 g_2 g_1(w) = g_1 h g_2 g_1 h g_2 g_1 g_1(w)
\]
for \( i = 0, 1, 2, 3. \) By (4) and (5) for \( i = 0, 1, 2, 3. \)
\[
g_1 g_2 g_1 g_1 g_2 g_1 g_1(w) = h g_2 g_1 g_1 g_2 g_1 g_1(w)
\]
for \( i = 0, 1, 2, 3. \) Now, by Theorems 1 and 2 we have (5) and (6) for all values of \( i. \) Therefore also (4) and (1) hold for all values of \( i. \)

By replacing in the above reasoning the word \( w \) by \( g(w) \) we see that also (2) holds for all values of \( i. \) Now the validity of (1) and (2) for all \( i \) and the equation \( g(w) = h(w) \) imply that \( S(G) = S(H). \)

**Theorem 4.** Suppose \( X \) is a three letter alphabet and \( G = (X, g, w) \) and \( H = (X, h, w) \) are D0L systems. If \( G \) or \( H \) is not elementary then \( S(G) = S(H) \) if and only if the first ten words in the sequences \( S(G) \) and \( S(H) \) are equal.

**Proof.** This is a special case of Theorem 3.
In the following section we discuss the ranks of D0L systems in more detail to see the wide range of applicability of Theorem 3.

3. Further results

If $G$ and $H$ are D0L systems and $S(G) = S(H)$, we have

$$d(L(G)) = d(L(H)).$$

On the other hand it does not follow that $d(G) = d(H)$.

**Example 1.** Let $X = \{a, b, c, d\}$ be an alphabet with four letters and define the morphisms $g : X^* \to X^*$ and $h : X^* \to X^*$ by

$$g(a) = g(b) = \varepsilon, \quad g(c) = abc, \quad g(d) = dc$$

and

$$h(a) = \varepsilon, \quad h(b) = a, \quad h(c) = bc, \quad h(d) = dca.$$

Consider the D0L systems $G = (X, g, dc)$ and $H = (X, h, dc)$. Then

$$g^n(dc) = h^n(dc) = dca^n$$

for $n \geq 0$. Hence $S(G) = S(H)$ although $d(G) = 2$ and $d(H) = 3$. Note also that $d(L(H)) < d(H)$.

Example 1 illustrates the wide range of applicability of Theorem 3. To apply Theorem 3 it is only required that one of the systems $G$ and $H$ has rank at most two. It may well turn out that $S(G) = S(H)$ even if the other has a much greater rank.

We do not know whether Theorem 3 is optimal. It might be sufficient to compare less than ten initial terms. However, the following example shows that four is not enough. Hence the bound in Theorem 3 is necessarily larger than the bound for binary D0L systems.

**Example 2.** We will modify the well known example due to Nielsen [8]. Let $X = \{a, b, A, B\}$ be a four letter alphabet. Define the morphisms $g : X^* \to X^*$ and $h : X^* \to X^*$ by

$$g(a) = aAbB, \quad g(A) = bB, \quad g(b) = aAaAbBbBaA, \quad g(B) = \varepsilon$$

and

$$h(a) = aAbB, \quad h(A) = bBaAaAbBbBbB, \quad h(b) = aA, \quad h(B) = \varepsilon.$$

Consider the D0L systems $G = (X, g, a)$ and $H = (X, h, a)$. Then we have

$$g^n(a) = h^n(a) \text{ for } n = 0, 1, 2, 3$$
but
\[ g^4(a) \neq h^4(a). \]

In fact
\[ g(a) = h(a) = aAbB, \]
\[ g^2(a) = h^2(a) = aAbBbaAabBBaA \]
and
\[ g^3(a) = h^3(a) = w_1w_2w_1, \]
where \( w_1 = g^2(a), \ w_2 = aAaAbBbBaAaAbBbB \). However, \( g^4(a) \) begins with
\[ g^3(a)aAbBbBaAbB \]
while \( h^4(a) \) begins with \( g^3(a)aAbBbBaAaA \).

We conclude by showing that for D0L systems \( G \) the numbers \( d(G) \) and \( d(L(G)) \) behave very differently. Denote
\[ d_{\min}(G) = \min\{d(H) \mid S(G) = S(H)\}. \]

**Example 3.** Let \( k \geq 2 \) be an integer and \( \Sigma_k = \{a_1, a_2, \ldots, a_k, c\} \) be an alphabet with \( k + 1 \) letters. Define the word \( w \) by
\[ w = \prod a_{i_1}a_{i_2}\ldots a_{i_k}, \]
where the product is over all permutations of \( 1, 2, \ldots, k \). Consider the D0L system \( G_k \) defined by
\[ G_k = (\Sigma_k, g, cw), \]
where
\[ g(c) = cw, \quad g(a_i) = a_i \quad \text{for} \quad 1 \leq i \leq k. \]

Then
\[ L(G_k) \subseteq cw^* \]
implies that \( d(L(G_k)) = 2 \).

Now, suppose \( H = (\Sigma_k, h, cw) \) is a D0L system such that \( S(H) = S(G_k) \). Because
\[ h(cw) = cw^2 \quad \text{and} \quad h(cw^2) = cw^3 \]
we have
\[ |h(c)| + |h(w)| = 1 + 2|w| \quad \text{and} \quad |h(c)| + 2|h(w)| = 1 + 3|w|. \]
Therefore \( |h(w)| = |w| \) and \( |h(c)| = 1 + |w| \). Hence
\[ h(c) = cw \quad \text{and} \quad h(w) = w. \]
Furthermore, \( h(a_i) \) is a nonempty word for all \( 1 \leq i \leq k \). Indeed, \( w \) is a product of \( k! \) different words of length \( k \). If we had \( h(a_i) = \varepsilon \) for some \( 1 \leq i \leq k \), the word \( h(w) \)
would be a product of \( k! \) words of length \( k \) at least two of which are equal. Putting all this together we see that \( h(\sigma) = g(\sigma) \) for all \( \sigma \in \Sigma_k \). Consequently
\[
d_{\min}(G_k) = d(G_k) = k + 1
\]
although \( d(L(G)) = 2 \).

References