FIXED POINT LANGUAGES OF DETERMINISTIC GENERALIZED SEQUENTIAL MACHINES ARE CONTEXT SENSITIVE

by

A. Van der Mude and
Adrian Walker

DCS-TR-66

Department of Computer Science
Rutgers, The State University of NJ
New Brunswick, New Jersey

February 2, 1978

1. INTRODUCTION

Fixed points of operators are widely studied, and they have been used to describe, for example, the semantics of recursive function definitions (Manna & Shamir [1977]), stability in discrete models for biological development (Herman and Walker [1976]), and to obtain results about cellular spaces (Takahashi [1976]). The regular languages have been characterized by Van Leewen [1975] as homomorphic images of the fixed points of monogenic functions. Herman and Walker [1976] have shown that the set of fixed points of an L-scheme, that is the set of all strings over an alphabet such that each string derives only itself under parallel replacement rules, is a regular language. This result was obtained by showing that, in a self derivation step, a descendant substring could only be located a bounded distance to the left or right of its parent substring. Since the parallel replacement rules of an L-scheme can easily be encoded as a deterministic generalized sequential machine (DGSM) transduction, - see e.g. Ginsburg [1966] for a description of a DGSM, - the question arises whether the fixed point language of a DSGM is also a regular language. A related question is whether, during a DGSM transduction of a string into itself, the amount by which the output lags or leads the input is bounded by a constant.

In this paper we show that such a constant bound does not exist, and that the device of defining a language as the set of fixed points of a DGSM has surprising power, namely, some of the languages so generated are not context-free. Since we show that the fixed point languages of DGSMs are accepted by deterministic LBAs, and since it has been pointed out by Ginsburg [1966] that the complement of the fixed point language of a DGSM is a context-free language, our results may suggest some new ways of studying the well known LBA problem, - whether or not there exists a language accepted by an LBA but not by any deterministic LBA.

2. DEFINITIONS

We follow the notation of Hopcroft & Ullman [1969], to which we add the following items.

A DGSM is a 4-tuple $M=\langle Q,V,\delta,q_0\rangle$ where Q is a finite set of states, V is a finite alphabet, $q_0 \in Q$ is the start state, and δ: $Q \times V \to Q \times V^*$. δ is extended to domain $Q \times V^*$ by $\delta(q,\lambda)=\langle q,\lambda\rangle$ and for $x \in V^*$ and $a \in V$, $\delta(q,xa) = \langle s,yz\rangle$ where $\delta(q,x) = \langle r,y\rangle$ and $\delta(r,a) = \langle s,z\rangle$. If $u \in V^*$ and $\delta(q_0,u) = \langle q,v\rangle$, then we write M(u) for v. We define the fixed point language of M as $F(M) = \{u \in V^+ | M(u)=u\}$. We define the class of fixed point languages of DGSMs as $F(DGSM)=\{F(M)|M$ is a DGSM$\} \cup \{F(M) \cup \{\lambda\} | M$ is a DGSM$\}$, where λ denotes the empty string.*

If x, $y \in V^*$ are such that there is a $z \in V^*$ for which xz=y, we say that x is a prefix of y, and that z is the remainder of $\langle x,y\rangle$. A deterministic modified Post correspondence problem (DMPCP) is as follows. Given lists $A=\{w_1,\ldots,w_k\}$ and $B=\{x_1,\ldots,x_k\}$ of strings in V^* such that no w_i is a prefix of any w_j, for $i \neq j$ and $i,j>1$, determine whether or not there exist integers $r,i_1,\ldots,i_r$ such that $w_1w_{i_1}\ldots w_{i_r}=x_1x_{i_1}\ldots x_{i_r}$. We say that $\langle x,y\rangle$ is a partial solution of a DMPCP if $x=w_1w_{i_1}\ldots w_{i_s}$, $y=x_1x_{i_1}\ldots x_{i_s}$ and x is a prefix of y.

We denote the class of regular languages as L(RG), of context-free languages as L(CF), and of languages accepted by DLBAs as T(DLBA). We denote the length of a string x by $|x|$, and the absolute difference of two integers i and j by $|i-j|$.

3. RESULTS

In this section we shall characterize the class of fixed point languages of DGSMs by showing that every regular language is in the class, and that some, but not all of the languages of deterministic LBAs are in the class. We shall further distinguish the class from the context-free languages by

---

* We could endow M with accepting states. We would then define F(M) as the set of strings each of which maps into itself and takes M to an accepting state. However, we shall demonstrate context-sensitivity without this device.

showing that the emptiness problem for the fixed point languages of DGSMs is unsolvable.

<u>Theorem 1</u> $L(RG) \subseteq F(DGSM)$

<u>Proof</u> Let $K \in L(RG)$ be such that $\lambda \notin K$. Then, wlg, we may assume there is a deterministic finite automaton $M = \langle Q, V, \delta, q_0, F \rangle$ such that $K = T(M)$, where δ is everywhere defined.

Let $Q' = \{\langle q_0, \lambda \rangle\} \cup \{\langle q, \lambda \rangle \mid q \in F\} \cup \{\langle q, a \rangle \mid q \in Q - F$, there exists a $p \in Q$ such that $\delta(p,a) = q\}$, let $\delta' : Q' \times V \to Q' \times V^*$ be defined by

$$\delta'(\langle p,a \rangle, b) = \begin{cases} \langle\langle q,b \rangle, a \rangle, & \text{if } \delta(p,b) = q \notin F \\ \langle\langle q,\lambda \rangle, ab \rangle, & \text{if } \delta(p,b) = q \in F, \end{cases}$$

and let $M' = \langle Q', V, \delta', \langle q_0, \lambda \rangle\rangle$. Then M' is a DGSM, and it can be proved, by induction on the length of x, that $x \in T(M)$ iff $x \in F(M')$ □

Thus every regular language is a fixed point language of some DGSM. We shall show that the converse does not hold. First we need a preliminary result.

<u>Lemma 1</u> $F(DGSM) \subseteq T(DLBA)$

<u>Proof</u> Let M be a DGSM. Construct from M a DLBA M', having two tracks. M' reads the input to M from its first track, writes the output from M on its second track, rejects the input if the output is longer, and otherwise accepts the input iff the two tracks contain the same string □

<u>Theorem 2</u> There exists a DGSM M such that $F(M) \in T(DLBA) - L(CF)$.

<u>Proof</u> Let M be the DGSM shown in Figure 1, let

$L_1 = \{1^{2m+1}0^{2m}1^{2m-1}\ldots.0010 \mid m \geq 1\}$ and

$L_2 = \{1^{2m}0^{2m-1}1^{2m-2}\ldots.1101 \mid m \geq 1\}$. We shall show that $F(M) = L_1 \cup L_2$.

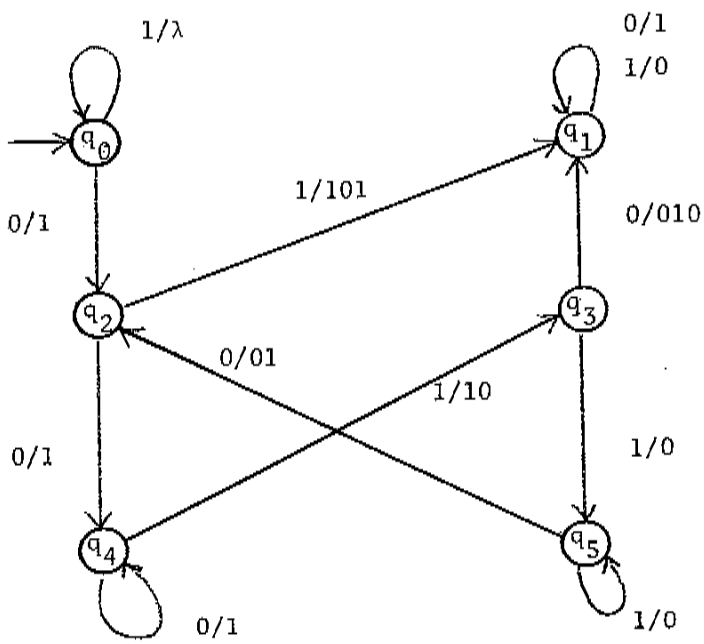

Figure 1 A DGSM

To show that $L_2 \subseteq F(M)$ we proceed as follows.

For $n \geq 2$ even, let $x_n = y_n \cdots y_1$ where, for $1 \leq j \leq n$,

$$y_j = \begin{cases} 1^n 0, \text{ if } j=n \\ 1^{j-1}0, \text{ if } j \text{ is even, } j \neq n \\ 0^{j-1}1, \text{ otherwise.} \end{cases}$$

Then $L_2 = \cup_{m=1}^{\infty}\{x_{2m}\}$. By induction on j, we have, that for $1<j<n$, $\delta(q_o, y_n \cdots y_j) = \langle q_k, y_n \cdots y_{j-1}\rangle$, where k=2 if j is even, k=3 otherwise. Hence $\delta(q_o, y_n \cdots y_2) = \langle q_2, y_n \cdots y_3 \rangle$, and it follows from Figure 1 that $\delta(q_o, y_n \cdots y_1) = \langle q_1, y_n \cdots y_1\rangle$. The proof that $L_1 \subseteq F(M)$ is similar.

To show that $F(M) \subseteq L_1 \cup L_2$ we let $x \in F(M)$. Then from Figure 1, the first symbol of x is not 0, and x contains at least one 0. So we can write $x = 1^n 0^m z$. For n even we can write $x = z_n \cdots z_1$ where $z_n = 1^n 0$. Then $z_n = y_n$. So from Figure 1, $x = y_n y_{n-1} z_{n-2} \cdots z_1$. Repeating this step yields $x = y_n y_{n-1} \cdots y_2 z_1$. But

$\delta(q_o, y_n \ldots y_2) = \langle q_2, y_n \ldots y_3 \rangle$ so, since $y_2=10$, it follows from Figure 1 that $z_1=1w$ for some w. So $x=y_n \ldots y_2 y_1 w$, hence from Figure 1, $w=\lambda$. Hence $x \in L_2$. A similar proof will establish that, if n is odd, then $x \in L_1$.

Thus $F(M)=L_1 \cup L_2=K$, say. It now follows from Lemma 1 that $K \in T(DLBA)$. To show that $K \notin L(CF)$, we use an argument similar to the proof of the uvwxy theorem (Hopcroft & Ullman [1969]) to show that for any context-free grammar G such that $K \subseteq L(G)$, there is an $x \in L(G)$ of the form $1^k 0^{2m-1} 1^{2m-2} \ldots 1101$ with $k<2m$ □

So far, we have shown that each regular language is the fixed point language of some DGSM, and that the fixed point languages extend into, but not beyond, the class of languages accepted by DLBAs. That they do not coincide with this class is shown as follows.

__Lemma 2__ $\{a^n b^n c^n \mid n \geq 1\} \in T(DLBA) - F(DGSM)$

__Proof__ Let $L=\{a^n b^n c^n \mid n \geq 1\}$. Obviously $L \in T(DLBA)$. Suppose there is a DGSM M such that $L=F(M)$. Then M contains a submachine of the form shown in Figure 2, where each $\bar{\alpha} \in \{a\}^*$, each $\bar{\beta} \in \{b\}^*$ and each $\bar{\gamma} \in \{c\}^*$.

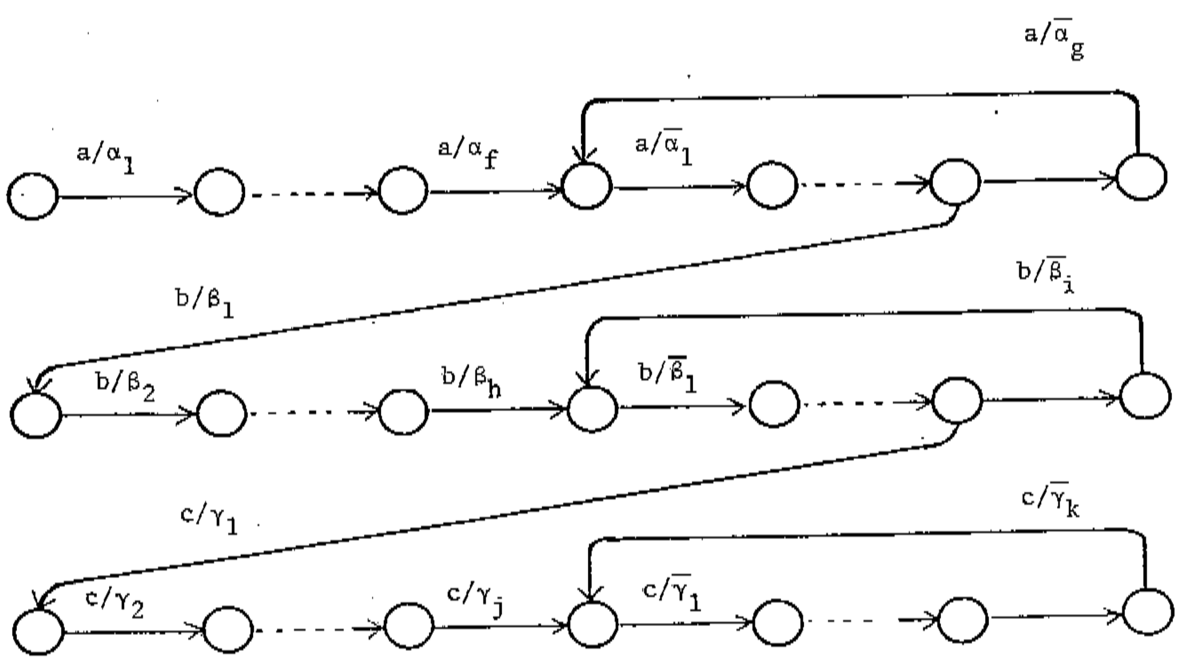

__Figure 2__ __Submachine of a DGSM__

Suppose that b does not appear in $\gamma_1 \cdots \gamma_j$. Then $M(a^n b^n) = a^n b^n c^m$ for some $m \geq 0$, since each $\overline{\gamma} \epsilon \{c\}^*$. If m=0 then $a^n b^n \epsilon F(M)$, a contradiction. If m>0 then, for large enough n, c occurs in some $\overline{\beta}$, again a contradiction. So b appears in $\gamma_1 \cdots \gamma_j$.

Let ℓ be the length of the longest string γ. Since (i) b appears in $\gamma_1 \cdots \gamma_j$ (ii) $|\gamma_1 \cdots \gamma_j| \leq \ell j$ (iii) $\overline{\gamma}_1 \cdots \overline{\gamma}_k \epsilon \{c\}^*$, and (iv) $a^n b^n c^n \epsilon F(M)$, we have $|(|a^n b^n c^j| - |M(a^n b^n c^j)|)| \leq \ell j$. Hence if $|\overline{\gamma}_1 \cdots \overline{\gamma}_k| \neq k$ then, for $n > k\ell j$, we have $M(a^n b^n c^n) = a^n b^n c^m$ for some $m \neq n$, a contradiction. On the other hand, if $|\gamma_1 \cdots \gamma_k| = k$ then, for $n > j$, $M(a^n b^n c^n) = a^n b^n c^n$ implies $M(a^n b^n c^{n+k}) = a^n b^n c^{n+k}$, again a contradiction. Thus there is no DGSM M such that L=F(M). □

So, although the fixed point languages of DGSMs extend into the class of languages of DLBAs, we have the following result.

<u>Theorem 3</u> $F(DGSM) \underset{\neq}{\subset} T(DLBA)$

<u>Proof</u> Immediate from Lemmas 1 and 2 □

Yet we shall establish that the emptiness problem for fixed point languages of DGSMs is unsolvable. We need a preliminary result.

<u>Lemma 3</u> DMPCP is unsolvable.

<u>Proof</u> Let M = <K, Γ, Σ, δ, q_0,F> be a deterministic Turing machine, and let B be the blank symbol. We construct from M a DMPCP as follows.

List A	List B
\$\$	\$\$\$$q_0$w¢
XY	XY
\$X	X
X\$\$	X¢
X¢	X\$\$\$
\$\$	¢
¢	\$\$\$

For $q \in K-F$, $p \in K$, and $X,Y,Z \in \Gamma-\{B\}$:

$\$qX$	Yp	if $\delta(q,X) = \langle p,Y,R\rangle$
ZqX	ZYp	
ZqX	pZY	if $\delta(q,X) = \langle p,Y,L\rangle$
$Zq\$\$$	$ZYp¢$	if $\delta(q,B) = \langle p,Y,R\rangle$
$Zq¢$	$ZYp\$\$\$$	
$Zq\$\$$	$pZY¢$	if $\delta(q,B) = \langle p,Y,L\rangle$
$Zq¢$	$pZY\$\$\$$	

For $q \in F$, and $X,Y \in \Gamma-\{B\}$:

XqY	q
$Xq\$\$$	$q¢$
$Xq¢$	$q\$\$\$$
$\$qY$	q
$\$q¢\$\$\$$	$\$\$\$$
$q\$\$\$¢$	$¢$
qX	q

In a manner similar to the proof by Hopcroft & Ullman [1969] that MPCP is unsolvable, we can show that the DMPCP described above has a solution iff the Turing machine M, given input w, halts in an accepting state. □

We use Lemma 3 as follows.

<u>Theorem 4</u> It is undecidable, for an arbitrary DGSM M, whether or not $F(M) = \emptyset$.

<u>Proof</u> Let P be a DMPCP with lists $A=\{w_1, \ldots, w_k\}$ and $B=\{x_1, \ldots, x_k\}$, and let $M = \langle Q, V, \delta, q_0\rangle$ be a DGSM constructed from P as follows. $\delta(q_0,\$) = \langle q_1,\$\$\$\rangle$, $\delta(q_1,\$) = \langle q_2,q_0 w¢\rangle$. For each $w_i = a_{i1} \cdots a_{ii}$ with $i>1$, there exist $q_2 = q_{i0}, q_{i1}, \ldots, q_{i(n-1)} \in Q$ such that $\delta(q_{i(j-1)}, a_{ij}) = \langle q_{ij}, \lambda\rangle$ for $1<j<n_i$, and $\delta(q_{i(n-1)}, a_{in_i}) = \langle q_2, x_i\rangle$. For all $z \notin A-\{w_1\}$, δ is such that $\delta(q_2,z) = \langle r,¢¢\rangle$ and for all $a \in V$, $\delta(r,a) = \lambda$. Since P is a DMPCP, M is a DGSM. Clearly $F(M)=\emptyset$ iff P has no solution.□

4. CONCLUSIONS

We have shown that the class of fixed point languages of DGSMs contains all regular languages, contains languages which are not context-free, and forms a proper subclass of the languages accepted by deterministic LBAs.

We have also shown that it is unsolvable to determine, for an arbitrary DGSM, whether or not its fixed point language is empty. These properties place in a new light the previously known result that the complement of a fixed point language of a DGSM is a context-free language. In establishing the properties, we have shown that, during the transduction of a string into itself, the output of a DGSM can lag arbitrarily far behind the input. Intuitively, this arbitrary lag seems to be associated with the context-sensitivity of the fixed point language. Thus it could be interesting to extend the present work to nondeterministic generalized sequential machines.

5. REFERENCES

Ginsburg, S. [1966], The Mathematical Theory of Context-free Languages, McGraw-Hill, New York.

Herman, G.T. and Adrian Walker [1976], On the stability of some biological schemes with cellular interactions, Theor. Computer Sci 2, 115-130.

Hopcroft, J.E. and J.D. Ullman [1969], Formal Languages and Their Relation to Automata, Addison-Wesley, Reading, Massachusetts.

Manna Z. and A. Shamir [1977], The optimal approach to recursive programs, Comm. ACM 20, 824-831.

Takahashi, H. [1976], The maximum invariant set of an automaton system, Info. Contr. 32, 307-354.

Van Leewen, J. [1975], On the fixpoints of monogenic functions in free monoids, Semigroup Forum 10, 315-328.