SEMI-THUE SYSTEMS WITH AN INHIBITOR

Robert McNaughton*
Department of Computer Science
Rensselaer Polytechnic Institute
Troy, NY 12180-3590, U.S.A.
mcnaught@cs.rpi.edu

April, 1998 (revised)

1. Introduction, notation and terminology. Semi-Thue systems constitute a universal model of computation, in the sense that any decision problem about computations is reducible to a problem about semi-Thue systems. Every recursively enumerable language is the language of some semi-Thue system. (E.g., see Chapter 7 of [3].)

And yet the concept is simple: A *semi-Thue system* is an ordered pair (Σ, Π), where Σ is a finite alphabet of characters and Π a finite set of rules (u, v), where in each case u and v are words over Σ. These systems are called "semi-Thue systems" because the rules are not necessarily reversible, u being referred to as "the left side" and v "the right side" of the rule (u, v). Semi-Thue systems are distinguished from Thue systems (named for their originator Axel Thue [15]) whose rules operate in both directions.

The important notion in the study of semi-Thue system is that of a derivation from one string to another. We write $w_1 \to w_2$ to mean that w_2 is *derived in one step* from w_1, i.e., that there exist words x and y such that $w_1 = xuy$ and $w_2 = xvy$, where (u, v) is a rule of the system. We say that z is *derivable* from w or that there is a *derivation* from w to z, if there exist strings $w_0, w_1, \ldots, w_p$ for some $p \geq 0$ such that $w_0 = w$, $w_p = z$ and $w_i \to w_{i+1}$ holds for each i, $0 \leq i \leq p-1$. Each w_i is a *line* of the derivation, which has $p+1$ lines, and has *p steps;* p is the *length* of the derivation. An *infinite derivation* is an infinite sequence $w_0, w_1, \ldots$, where $w_i \to w_{i+1}$ for all nonnegative integers i.

Although the class of semi-Thue systems is quite powerful, there have been some applications in which not all this power is appropriate. An example is the context-free

*Supported by Grant No. CCR-9500182 from the National Science Foundation.

grammar, which had as its origin the phrase-structure grammar put forth tentatively and critically by Chomsky [2] for natural languages, but which by now has its chief application in the domain of formal languages (see [11]). A *context-free grammar* is a semi-Thue system that (1) distinguishes between "terminal" and "nonterminal" characters of the alphabet, (2) restricts rules to those having single nonterminals as their left sides, and (3) restricts derivations to those whose first lines all consist of a single occurrence of a particular nonterminal designated as the start symbol (usually S in the literature) and whose last lines have only terminal characters.

Another application has arisen in the area of computer science concerned with theorem proving by machine, where semi-Thue systems are used and thought about extensively. Typically, these semi-Thue systems, often called "rewrite systems," are used to reduce words to simpler equivalent words (see, e.g., [7] or [1]). Rewrite systems and context-free grammars are quite different both in their purpose and their operation. Nevertheless, these enterprises do have in common their use of computationally weak semi-Thue systems.

The research program into which the present paper fits is not concerned with particular applications of semi-Thue systems such as grammars and rewrite systems. Rather it considers various subclasses of semi-Thue systems that appear to be usefully weaker than the entire class (without regard to any particular application), and then attempts to assess how useful and weak they are. One method of assessing a subclass is to determine whether certain decision problems are decidable for it. If so then that is evidence that the subclass is weak, since most significant decision problems are undecidable for the full class of all semi-Thue systems. And, if the decision problem is a computationally important one, it is evidence also that the subclass may be useful.

There are several decision problems that can be used as criteria in this way. Let us focus on three of them:

(1) The *halting problem.* Given a semi-Thue system (Σ, Π) of the subclass, and given $x \in \Sigma^*$, is every derivation whose first line is x finite?

(2) The *uniform halting problem.* Is every derivation in a given semi-Thue system of the subclass finite?

(3) The *derivability problem.* Given a semi-Thue system of the subclass, and given $x, y \in \Sigma^*$, does there exist a derivation of y from x?

For example, for both the class of context-free grammars and the class of rewrite systems used by many theoreticians of machine theorem proving, all three of these problems are decidable. Moreover, any subclass of the semi-Thue systems for which any of these three problems is solvable is weaker and possibly more useful than the class of all semi-Thue systems. Weaker because all three problems are undecidable for the class of all semi-Thue systems. And more useful, albeit in a restricted set of possible applications, because the ability to tell whether or not derivations will terminate gives the user an advantage. The importance of termination in the practical use of rewrite systems is described in detail in [6].

The uniform halting problem seems more difficult than the other two, although there seems to be no general proof that if that problem is solvable for any subclass of semi-Thue systems then the other two problems are solvable also. At any rate, several research workers have decided to focus on the uniform halting problem; in particular, the question of whether or not that problem is decidable for the subclass of one-rule semi-Thue systems has received some attention, so far without an answer, in [4], [5], [8], [9], [10], [12], [13], [14] and [16]. This subclass will be discussed in Section 4.

The subclass that is the main focus of this paper is the subclass of semi-Thue systems with an *inhibitor,* i.e., an alphabetic character that occurs at least once on the right side of every rule, but does not occur on the left side of any rule. The main result is that the uniform halting problem for this subclass is decidable, proved in Section 3, where it is also proved that the halting problem and the derivability problem are decidable. Derivations in semi-Thue systems with an inhibitor will be analyzed in detail in Section 2.

The subclass of semi-Thue systems with an inhibitor was noticed in the course of studying the subclass of one-rule systems [13]. The results of the present paper on semi-Thue systems with an inhibitor lead to a distinction that promises to be fruitful in analyzing derivations in one-rule systems without an inhibitor. That distinction, between *well behaved* derivations and *ill behaved* derivations, is explained in Section 4.

In this paper lower-case Greek iota (ι) will be used as an inhibitor. So, in a semi-Thue system with an inhibitor, $\iota \in \Sigma$, every right side has at least one occurrence of ι, and no left side has any occurrence of ι. Our convention will be that a semi-Thue system has k rules: (u_i, v_i) for each i, $1 \leq i \leq k$.

We shall write $x \rightarrow^n y$ to mean that there is a derivation of n steps ($n \geq 0$) from x to y; $x \rightarrow^* y$ to mean that there is a derivation from x to y; and $x \rightarrow^+ y$ to mean that there is such a derivation of at least one step.

For $x, y \in \Sigma^*$, $x \triangleright y$ will mean that there exist $z_1, z_2 \in \Sigma^*$ such that $x \rightarrow z_1 y z_2$. We shall write $x \triangleright^+ y$ to mean that there exist $z_1, z_2 \in \Sigma^*$ such that $x \rightarrow^+ z_1 y z_2$.

The well known fan theorem states that an infinite rooted tree in which each node has only finitely many children nodes has an infinite path. The following generalization will be useful in Section 2; the proof is left to the reader:

Theorem 1.1. If a forest of finitely many rooted trees has infinitely many nodes, each of which has only finitely many children nodes, then the forest has an infinite path.

Definition (the y). If x, y and z are words and xyz is a line in a derivation then it may be that y occurs several times as a factor (i.e., substring) of that line. For example, if $x = bbcc$, $y = bccb$ and $z = ccb$, so that $xyz = bbccbccbccb$, then there are three occurrences of y in xyz, the second of which we shall refer to as "the apparent occurrence of y in xyz," or "the y in xyz," or (frequently, when the designation xyz is understood) "the y." The article "the" will indicate that we are referring to a particular occurrence of y and not to the word y itself. When we wish to talk about the word apart from any occurrence in a line we shall simply say "y."

Definition (the same occurrence). If $x_1bx_2u_hx_3$ $(x_1u_hx_2bx_3)$ is the j^{th} line of a derivation where b is a letter and $x_1bx_2v_hx_3$ $(x_1v_hx_2bx_3)$ is the $(j+1)^{st}$ line of that derivation, then we talk of the apparent occurrence of b in the j^{th} line as being *the same* as the apparent occurrence of b in the $(j+1)^{st}$ line. On the other hand, no occurrence of a character in the apparent occurrences of u_h or v_h exists in the other line: any character occurrence in the apparent u_h is *destroyed,* and any character occurrence in the apparent v_h is *created,* in going from the j^{th} line to the $(j+1)^{st}$ line.

2. Analysis of derivations. Section 3 will present the decision procedure to determine whether a given semi-Thue system with an inhibitor is uniformly terminating. In preparation, this section will develop a method of analyzing derivations in such systems. To begin, it is stipulated that there be no occurrence of the inhibitor ι in the first line of a derivation. This stipulation will help us establish important structural concepts. It is justified by the fact that (since ι does not occur on the left side of any rule) if there is an infinite derivation from $x_1\iota x_2\iota\cdots\iota x_n$ then there is an infinite derivation from one of the x_i's. All semi-Thue systems discussed in this section and Section 3 will be systems with the inhibitor ι.

Definition (vital). For $b \in \Sigma$, if xby is a line other than the first in a derivation, then the apparent occurrence of b is *vital* if it is not ι and has been created either in that line or in some preceding line of the derivation (in other words, if that occurrence of b does not exist in the first line).

Definition (S-occurrence, the set S). (1) If $w \neq \lambda$ and xwy is a line of a derivation then the w is an *S-occurrence* in that line if all its character occurrences are vital, the x does not end in a vital character occurrence, and the y does not begin with a vital character occurrence. (That is to say, the x must either be null or end in a nonvital character occurrence. And the y must either be null or begin with a nonvital character occurrence.) In short, a nonnull S-occurrence is a maximal factor of vital character occurrences in a line of a derivation. (2) If a line of a derivation begins with (ends in) ι then the occurrence of λ at the left end (right end) of the line is an *S-occurrence.* And if bc is a factor of a line where both the b and the c are nonvital character occurrences, at least one of the two being ι, then the λ between the b and the c is an *S-occurrence.* (3) S is the set of all words having an S-occurrence in some line of some derivation.

Note that if $x\iota s\iota y \to^* z$ and s has no ι then $z = x'\iota\tau\iota y'$, where $x \to^* x'$, $s \to^* \tau$ and $y \to^* y'$. This follows from the segregating power of the inhibitor ι, and shows the importance of S-occurrences in lines all of whose non-iota character occurrences are vital. Even more revealing of the importance of S-occurrences is the next theorem.

Definition (S-sequence). The sequence $s_1, s_2, \cdots$ is an *S-sequence* if, for each i, $s_i \in S$ and $s_i \triangleright s_{i+1}$.

Theorem 2.1. In a semi-Thue system with an inhibitor, there exists an infinite derivation $w_1, w_2, \ldots$ if and only if there is an infinite S-sequence.

Proof: If there is an infinite S-sequence $s_1, s_2, \ldots$, then an infinite derivation $w_1, w_2, \ldots$ can be defined as follows: For each i, define z_i and z_i' so that $s_i \rightarrow z_i s_{i+1} z_i'$ (guaranteed by the definition of $s_i \triangleright s_{i+1}$). Then put $w_1 = s_1$ and, for each $i \geq 2$, put

$$w_i = z_1 z_2 \cdots z_{i-1} s_i z_{i-1}' \cdots z_2' z_1'$$

Clearly, $w_1, w_2, \ldots$ is an infinite derivation.

Now assume there is an infinite derivation $w_1, w_2, \ldots$. Let

$$w_1 = y_0 z_{1,1} y_1 z_{1,2} y_2 \cdots y_{m-1} z_{1,m} y_m$$

where the character occurrences of w_1 that are never destroyed in the derivation are precisely those that occur in the apparent factors $y_0, \ldots, y_m$. So, for every i,

$$w_i = y_0 z_{i,1} y_1 z_{i,2} y_2 \cdots y_{m-1} z_{i,m} y_m,$$

for some $z_{i,1}, \ldots, z_{i,m}$. Consequently, for some j, $1 \leq j \leq m$, the sequence

$$z_{1,j}, z_{2,j}, \ldots$$

is an infinite derivation possibly with repetitions; that is to say, there is an infinite sequence $i_1 = 1, i_2, i_3, \ldots$ where for all h

$$z_{i_h,j} = z_{i_h+1,j} = \cdots = z_{i_{h+1}-1,j} \rightarrow z_{i_{h+1},j}.$$

For convenience we alter the notation and set $z_h = z_{i_h,j}$ for all h. Then

$$z_1, z_2, \ldots$$

is an infinite derivation in which every character occurrence of z_1 is eventually destroyed. Since z_1 is only finitely long, there exists a p such that, for all $i \geq p$, the line z_i has no character occurrences in common with z_1.

Thus for all $i \geq p$,

$$z_i = s_{i,1} \iota s_{i,2} \iota \ldots \iota s_{i,q_i}$$

where each $s_{i,h} \in S$, and $q_i - 1$ is the number of ι's in z_i.

For each line $i \geq p$, there is exactly one h such that the $s_{i,h}$ is rewritten by a rule, becoming (for some $j \geq 1$) $s_{i+1,h} \iota \cdots \iota s_{i+1,h+j}$ in the $(i+1)^{st}$ line; for all

$h' \neq h, 1 \leq h' \leq q_i$, the $s_{i,h'}$ is copied to become the very same S-occurrence in the $(i+1)^{st}$ line (but with a different subscript for $h' > h$). Define $C'(< i, h' >, < i+1, h'' >)$ to mean that the $s_{i,h'}$ is copied to become the $s_{i+1,h''}$ in the $(i+1)^{st}$ line. Then define C to be the transitive, symmetric and reflexive closure of the relation C'. Thus $C(< i, h >, < i', h' >)$ holds if either (1) $i = i'$ and $h = h'$, (2) $i' > i$ and the $s_{i,h}$ is copied, recopied, etc., until it becomes the $s_{i',h'}$ in the $(i')^{th}$ line, or (3) $i > i'$ and the $s_{i',h'}$ is copied, recopied, etc., to become the $s_{i,h}$.

Clearly, C is an equivalence relation; furthermore if $C(< i, h >, < i', h' >)$ then the word $s_{i,h} = s_{i',h'}$. Note also that, for any i and $h \leq q_i$, if there is no $h' \leq q_{i+1}$ such that $C(< i, h >, < i+1, h' >)$ then, for finitely many values of h', $s_{i,h} \triangleright s_{i+1,h'}$.

Consider now the graph H whose nodes are the equivalence classes of the C relation: The roots of H are the classes having the respective pairs $< p, 1 >, \ldots, < p, q_p >$ as members; there is an edge from a node N_1 to a node N_2 of H if and only if there are i, h, h' such that $< i, h > \in N_1$, $< i+1, h' > \in N_2$ and the $s_{i+1,h'}$ in the $(i+1)^{st}$ line is part of what is obtained by rewriting from the $s_{i,h}$ in the i^{th} line. Clearly, H is a forest of rooted trees. Define $s(N) = s_{i,h}$ for any $< i, h > \in N$.

Since the z-derivation is infinite, H has infinitely many nodes, each of which has only finitely many children nodes. Hence, by Theorem 1.1, there is an infinite path through the graph: $N_1, N_2, \ldots$. Putting $s_j = s(N_j)$ for all j, we have $s_j \in S$ and $s_j \triangleright s_{j+1}$ for all j.□

Definition (v-section, $L(v_h), R(v_h)$). The word w is a *v-section* if ι does not occur in w and there is a v_h such that, for some x and y, either $v_h = x\iota w\iota y$, $v_h = w\iota y$ or $v_h = x\iota w$. Thus the null word is a v-section if and only if some v_h either begins in ι, ends in ι, or has $\iota\iota$ as a factor. If $v_h = w_1\iota y = x\iota w_2$, where w_1 and w_2 have no ι, then $w_1 = L(v_h)$ and $w_2 = R(v_h)$. In particular, $L(v_h)$ $(R(v_h)) = \lambda$ if and only if v_h begins with (ends in) ι.

Note that every v-section has an S-occurrence in some derivation. Whenever a rule (u_h, v_h) is applied in a derivation where $v_h = L(v_h)\iota \cdots \iota R(v_h)$, the apparent occurrences of the internal v-sections always become S-occurrences. The apparent occurrences of $L(v_h)$ and $R(v_h)$ become either S-occurrences or parts of S-occurrences. E.g., for the system

$$\{(a, a\iota bc\iota d), (c, e\iota f\iota g)\}$$

in the third line of the derivation

$$a$$

$$a\iota bc\iota d$$

$$a\iota be\iota f\iota g\iota d$$

the occurrence of g is an S-occurrence, but the occurrence of $e = L(v_2)$ is part of the S-occurrence be.

An S-occurrence of λ must, in its first line, be the result of an application of a rule having lambda as a v-section. However, if $L(v_h)$ or $R(v_h) = \lambda$, the application of the rule (u_h, v_h) need not result in an S-occurrence of λ. E.g., in the system

$$\{(a, a\iota bc\iota d), (c, \iota f \iota)\}$$

$L(v_2) = R(v_2) = \lambda$. In the derivation

$$a$$

$$a\iota bc\iota d$$

$$a\iota b\iota f\iota\iota d$$

the application of the second rule results in the third line in which $R(v_2)$ becomes an S-occurrence of λ but $L(v_2)$ does not, being absorbed into the S-occurrence b.

Generally, we can think of a nonnull S-occurrence in a line as a maximal substring occurrence consisting of consecutive vital character occurrences; it is brought about by one or more rule applications each of which is responsible for creating some of its vital character occurrences. The earliest S-occurrences are v-sections, which come about as a result of rule applications. The more complex S-occurrences are brought about by the modification of simpler S-occurrences. The theorems that follow will describe this process in detail.

Theorem 2.2. If the s is an S-occurrence in a line xsy of a derivation, then either x ends in ι or y begins with ι or both.

Proof: Let s be any word having an S-occurrence in the derivation. Let xsy be the first line that has that occurrence of s as an S-occurrence. Then $xsy = \alpha v_i \beta$, for some α, i, β, the preceding line being $\alpha u_i \beta$. Let us make three observations.

(1) The v_i must have either the rightmost character occurrence of the x, a character occurrence of the s or the leftmost character occurrence of the y. (Otherwise s would have an S-occurrence in the preceding line.)

(2) The v_i must have a character occurrence in common with either the x or the y. (Since s has no ι but v_i does, s cannot account for all of v_i.)

(3) If the v_i has any two character occurrences of the line xsy then it has every character occurrence between.

From (1), (2) and (3) we infer that the v_i contains either the rightmost character occurrence of the x or the leftmost character occurrence of the y.

Case I: $x = x'b$, where b is a single character, and the v_i contains the b. Then since the s is an S-occurrence, the b is not vital. And since it is part of the v_i, it is not a character occurrence of the first line. Hence $b = \iota$.

Case II: $y = cy'$, where c is a single character, and the v_i contains the c. Then, by similar reasoning, $c = \iota$.

We have proved that our theorem is true of the first line having that s as an S-occurrence. But since ι's are indestructible in a derivation, our theorem is true of all the lines having that s as an S-occurrence.□

We consider two examples. The first is the system with the two rules $(ab, aac\iota)$ and $(c, b\iota)$, and the derivation

$$ab$$

$$\underline{aac}\iota\underline{\ \ }$$

$$\underline{aab}\iota\underline{\ \ }\iota\underline{\ \ }$$

$$\underline{aaac}\iota\underline{\ \ }\iota\underline{\ \ }\iota\underline{\ \ }$$

whose S-occurrences are underlined. These are the first four lines of an infinite derivation, which is obtained by applying the two rules alternately. The lines of this derivation have increasingly long S-occurrences at their left ends, followed by several S-occurrences of λ.

Our second example is the system with the two rules $(ab, b\iota ac)$ and $(c, b\iota a)$ and the derivation

$$ab$$

$$\underline{b}\iota\underline{ac}$$

$$\underline{b}\iota\underline{ab}\iota\underline{a}$$

$$\underline{b}\iota\underline{b}\iota\underline{ac}\iota\underline{a}$$

$$\underline{b}\iota\underline{b}\iota\underline{ab}\iota\underline{a}\iota\underline{a}$$

Each line of the resulting infinite derivation will have several S-occurrences of b, followed by an S-occurrence of either ab or ac, and then several S-occurrences of a.

Definition (the sets S_L and S_R). $S_L = \{x | \text{some derivation has a line } x\iota y \text{ in which the } x \text{ has no nonvital character occurrences}\}$. $S_R = \{y | \text{some derivation has a line } x\iota y \text{ in which the } y \text{ has no nonvital character occurrences}\}$. It is not difficult to see that $\lambda \in S_L$ (S_R) if and only if $\lambda \in L(v_j)$ $(R(v_j))$ for some $j, 1 \leq j \leq k$. (k is the number of rules.)

It is clear from the definitions that $S_L \cup S_R \subseteq S$.

Theorem 2.3. If $xbsy$ $(xsby)$ is a line of a derivation in which the s is an S-occurrence and the apparent letter b is an occurrence from the first line, then $s \in S_L$ $(s \in S_R)$.

Proof for $xbsy$: By Theorem 2.2, $y = \iota y'$ for some y'. The result of deleting all lines of the derivation after the line $xbsy$, and then deleting the noted b and all characters to its left from all lines, is a derivation possibly with repeated lines, having the line $sy = s\iota y'$ in which the apparent s is an S-occurrence, and hence is vital. Thus $s \in S_L$ by definition.

The proof that $s \in S_R$ when the line is $xsby$ is symmetric.□

Theorem 2.4. (a) $S_L =$ the smallest set C such that, for all j, (1) $L(v_j) \in C$, and (2) $xu_jy \in C$ implies $xL(v_j) \in C$. (b) $S_R =$ the smallest C' such that, for all j, (1) $R(v_j) \in C'$, and (2) $xu_jy \in C'$ implies $R(v_j)y \in C'$.

Proof that $C \subseteq S_L$: We get $L(v_j) \in S_L$ from the two-line derivation u_j, v_j. If $xu_jy \in S_L$ then xu_jy has no ι and there is a derivation of $xu_jy\iota z$, for some z. This derivation can be extended to become a derivation of $xv_jy\iota z$, which shows that $xL(v_j) \in S_L$.

Proof that $S_L \subseteq C$: In any derivation D, put $J(D) = \{j|$ the j^{th} line does not begin with a character occurrence of the first line$\}$. If $j' > j \in J(D)$ and D has a j'^{th} line then $j' \in J(D)$ also. For each $j \in J(D)$, let the j^{th} line be $x_j\iota y_j$, where x_j is iota-free. Clearly, if $J(D) \neq \emptyset$ and j_0 is the smallest member of $J(D)$ then $x_{j_0} = L(v_h)$ for some h. Also, if $j > j_0$ and $x_j \neq x_{j-1}$ then, for some x', y', h' where x' has no ι, $x_{j-1} = x'u_{h'}y'$ and $x_j = x'L(v_{h'})$. Let $L(D) = \{x_j|j \in j(D)\}$. We have proved that $L(D) \subseteq C$. Since $S_L = \bigcup L(D)$ where the union is taken over all derivations D in the system, we have $S_L \subseteq C$.

Thus $S_L = C$. The proof that $S_R = C'$ is symmetric.□

We shall generally focus on S_L knowing that, whatever we prove about S_L, an appropriate similar assertion can be proved about S_R.

Definition (The set S_{Li}). For each i, $1 \leq i \leq k$,

$$S_{Li} = \{x|x \text{ has no } \iota \text{ and } L(v_i)\iota \rightarrow^* x\iota y \text{ for some } y\}$$

The proofs of the following two theorems are straightforward:

Theorem 2.5. $S_L = S_{L1} \cup \cdots \cup S_{Lk}$.

Theorem 2.6. If $xu_jy \in S_{Lj'}$ then $xS_{Lj} \subseteq S_{Lj'}$.

Continuing the examples given above, in the system whose rules are $(ab, aac\iota)$ and $(c, b\iota)$ we have $S_L = \{b\} \cup \{a^ib, a^ic|i \geq 2\}$ and $S_R = \{\lambda\}$. In the system whose rules are $(ab, b\iota ac)$ and $(c, b\iota a)$, we have $S_L = \{b\}$ and $S_R = \{ac, a\}$.

Theorem 2.7. $S =$ the smallest set S' such that:

(1) every v-section is in S';
(2) $S_L \cup S_R \subseteq S'$;
(3) if $u_h = xy$ and $yz \in S_L$ then $R(v_h)z \in S'$;
(4) if $xy \in S_R$ and $u_h = yz$ then $xL(v_h) \in S'$; and
(5) if $xu_hy \in S'$ then $xL(v_h) \in S'$ and $R(v_h)y \in S'$.

Proof: To prove that $S' \subseteq S$ we verify that (1)—(5) are all true of S. Items (1) and (2) are clear.

The verification for item (3) is as follows: Assume $u_h = xy$ and $yz \in S_L$. Then for some z' there is a derivation whose last line is $yz\iota z'$ where the yz is vital. The result of appending x to the left end of every line is a derivation whose last line is $xyz\iota z' = u_hz\iota z'$. We can extend this derivation by adding the line $v_hz\iota z'$, in which the word $R(v_h)z$ will have an S-occurrence, showing that $R(v_h)z \in S$. The verification for item (4) is similar.

To verify that (5) is true of S, assume $xu_hy \in S$. Then, for some z and z', there is a derivation whose last line is zxu_hyz' in which the xu_hy is an S-occurrence. If we append to this derivation the line zxv_hyz' the result is a derivation whose last line has both $xL(v_h)$ and $R(v_h)y$ as S-occurrences, showing that both these words are in S. (Examples illustrating (3) and (5) are provided after this proof.)

To prove that $S \subseteq S'$ we assume an arbitrary derivation D and prove that every S-occurrence in D is in S'. We do so by proving the following proposition by mathematical induction on i: If D has at least i lines then every S-occurrence in the i^{th} line is in S'.

This proposition is true for $i = 1$, since the first line has no S-occurrences. Assume it is true for the i^{th} line and let the $(i+1)^{st}$ line be xsy in which the s is an S-occurrence. Assume also that s does not have an S-occurrence in the i^{th} line, and that $u_h \to v_h$ is the rule by means of which the i^{th} line is rewritten as the $(i+1)^{st}$ line.

If $s = \lambda$ then s is either (a) at the left end of the line, which begins in ι, (b) at the right end of the line, which ends in ι, (c) flanked by two consecutive ι's, (d) flanked by nonvital $b \neq \iota$ on the left and ι on the right, or (e) flanked by ι on the left and nonvital $c \neq \iota$ on the right. Since λ is not an S-occurrence in the preceding line in the derivation, the s must be a factor of the v_h in the $(i+1)^{st}$ line. In each of the cases (a), (b), (d) and (e), the ι flanking the s must be part of the v_h and λ is a v-section of v_h. In case (c) either both ι's are part of the v_h and hence λ is an interior v-section of v_h, or only one of the two ι's is part of the v_h and λ is an end v-section of v_h.

Henceforth we assume $s \neq \lambda$. Let the noted occurrence of s in the $(i+1)^{st}$ line consist of the $B(s)^{th}$ through the $E(s)^{th}$ characters of that line, in left-to-right order. And let the noted occurrence of v_h in that line consist of the $B(v_h)^{th}$ through the

$E(v_h)^{th}$ characters. Note that $B(v_h) \leq E(v_h)$ and $B(s) \leq E(s)$. To carry through the proof, we divide into cases based on the relative positions of s and v_h in the $(i+1)^{st}$ line.

Case I: $B(v_h) \leq B(s)$ and $E(s) \leq E(v_h)$. Then s is a factor of v_h, and since s is an S-occurrence of the $(i+1)^{st}$ line, s must be a v-section of v_h. By (1), $s \in S'$.

In the remaining cases either $E(v_h) < E(s)$ or $B(s) < B(v_h)$. Note that $E(v_h) < E(s)$ implies $B(v_h) < B(s)$ and $B(s) < B(v_h)$ implies $E(s) < E(v_h)$: for, since v_h has an occurrence of ι but s does not, the v_h cannot be wholly inside the s. The possibility $E(v_h) < E(s)$ will give rise to Cases II and III, while the possibility $B(s) < B(v_h)$ will give rise to Cases IV and V, which are right-left symmetric to Cases II and III, respectively.

Case II: $E(v_h) < B(s)$. I.e., $B(v_h) \leq E(v_h) < B(s) \leq E(s)$. Then $E(v_h) = B(s) - 1$; for otherwise there would be no change in the vicinity of the s in going from the i^{th} line to the $(i+1)^{st}$ line, and the s being an S-occurrence in the $(i+1)^{st}$ line would also be an S-occurrence in the i^{th} line, contrary to our assumption. Furthermore, the rightmost character of v_h must be ι, otherwise that character occurrence would be vital and the s would not be an S-occurrence in the $(i+1)^{st}$ line. Thus $R(v_h) = \lambda$. And since s does not have an S-occurrence in the i^{th} line, the s must be a proper suffix of an S-occurrence rs in that line.

Case IIa: r is a proper suffix of u_h, i.e., $u_h = r'r$, $r' \neq \lambda$. Then rs is an S-occurrence in line i preceded by a character occurrence of the first line. (The rightmost character occurrence of r' cannot be ι since it is part of u_h, and cannot be vital since it is not part of the S-occurrence rs.) It follows by Theorem 2.3 that $rs \in S_L$. Taking $x = r', y = r, z = s$, we apply (3) getting $R(v_h)s \in S'$. Since $R(v_h) = \lambda$, we have $s \in S'$.

Case IIb: r is not a proper suffix of u_h. Then $r = r''u_h$, for some r''. Here we apply (5) to get $s = R(v_h)s \in S'$ from $r''u_hs \in S'$, completing Case II.

Case III: $E(v_h) < E(s)$ and $B(s) \leq E(v_h)$. Since v_h cannot be a substring of s, we get $B(v_h) < B(s)$. Thus Case III implies $B(v_h) < B(s) \leq E(v_h) < E(s)$. Consider the character occurrence b immediately to the left of the leftmost character occurrence of s. It must be an ι, for otherwise it would be vital and would have to be part of the S-occurrence s. Furthermore, it must be the rightmost ι in the v_h, for otherwise, there would be an ι inside of the S-occurrence s. Therefore, we can assume that the $(i+1)^{st}$ line is $x'z_1\iota z_2z_3y$, where $v_h = z_1\iota z_2$, $s = z_2z_3$ and $z_2 = R(v_h)$. The i^{th} line is $x'u_hz_3y$.

Case IIIa: all the character occurrences of the u_h are vital in the i^{th} line. Then, for some x'', x''' where $x' = x''x'''$, the $x'''u_hz_3$ is an S-occurrence in the i^{th} line. By the inductive hypothesis, $x'''u_hz_3 \in S'$, and by (5), $s = R(v_h)z_3 \in S'$.

Case IIIb: not all the character occurrences in the u_h in the i^{th} line $x'u_h z_3 y$ are vital. Then $u_h = u'u''$ where $u' \neq \lambda$, all the character occurrences in the u'' are vital but the rightmost character occurrence in the u' is not vital. This character occurrence is not an ι, so it must be a character occurrence of the first line of the derivation. It follows that $u''z_3$ is an S-occurrence, since all its character occurrences are vital but it is not adjacent to a vital character occurrence in the i^{th} line, either on the left or on the right. Moreover, by Theorem 2.3, $u''z_3 \in S_L$. Thus $s = R(v_h)z_3 \in S'$ by (3).

Case IV: $E(s) < B(v_h)$, i.e., $B(s) \leq E(s) < B(v_h) \leq E(v_h)$. This case is left-right symmetric to Case II. (It is obtained from Case II by simultaneously interchanging $<$ and $>$ and interchanging B and E.)

Case V: $B(s) < B(v_h)$ and $B(v_h) \leq E(s)$, i.e., $B(s) < B(v_h) \leq E(s) < E(v_h)$. This case is left-right symmetric to Case III.□

We illustrate (3) of Theorem 2.7 by the Thue system

$$\{(c, aa\iota e), (ba, h\iota g\}$$

Taking $x = b$, $y = a$ and $z = a$, we have $xy = ba = u_2$ and $yz = aa \in S_L \subseteq S'$. Hence by (3) we get $ga = R(v_2)z \in S'$. A derivation with an S-occurrence of ga is:

$$bc$$

$$baa\iota e$$

$$h\iota ga\iota e$$

We illustrate (5) by the Thue system

$$\{(e, h\iota ab), (f, cd\iota j), (bc, h\iota j)\}$$

Assume for the moment that $abcd \in S'$, and take $x = a$, $y = d$. Since $u_3 = bc$, by (5) we get $ah = xL(u_3) \in S'$ and $jd = R(v_3)y \in S'$. A derivation with S-occurrences of $abcd$, ah and jd is

$$ef$$

$$h\iota abf$$

$$h\iota abcd\iota j$$

$$h\iota ah\iota jd\iota j$$

Definition (ABC property). A word w has the *ABC property* if there exist words A, B, C such that $w = ABC$, $|B| \leq \max_{1\leq h\leq k}(|v_h|)$, $Ax_1 \in S_R$ for some x_1 and $x_2C \in S_L$ for some x_2.

Theorem 2.8. Every $s \in S$ has the ABC property.

Proof: Since $S = S'$ (see Theorem 2.7) we can complete our proof by showing that every $s \in S'$ has the ABC property.

To this end we note that (1) if s is a v-section then we can take $A = C = \lambda$ and $B = s$.

For (2), if $s \in S_L$ (S_R) then we can take $A = B = \lambda$ and $C = s$ ($B = C = \lambda$ and $A = s$).

For (3), if $s = R(v_h)z$ where $u_h = xy$ and $yz \in S_L$, then we can take $A = R(v_h)$, $x_1 = \lambda$, $B = \lambda$, $C = z$ and $x_2 = y$.

For (4), if $s = xL(v_h)$ where $xy \in S_R$ and $u_h = yz$, then we can take $A = x$, $x_1 = y$, $B = \lambda$, $C = L(v_h)$ and $x_2 = \lambda$.

For (5) we should demonstrate that if xu_hy has the ABC property then both $s_1 = xL(v_h)$ and $s_2 = R(v_h)y$ also have it. Accordingly, we assume that $xu_hy = A'B'C'$ where $|B'| \leq \max_{1 \leq h \leq k}(|v_h|)$, $A'x'_1 \in S_R$ and $x'_2C' \in S_L$. We shall prove that s_1 has the ABC property, leaving the similar proof for s_2 to the reader. A division into cases is required according to whether or not x is a prefix of A' and, if not, whether or not it is a prefix of $A'B'$.

Case I: $A' = xA''$. Then, for $s_1 = xL(v_h)$, take $A = x$, $x_1 = A''x'_1$, $B = \lambda$, $C = L(v_h)$, $x_2 = \lambda$, which satisfies the requirement, since $Ax_1 = xA''x'_1 = A'x'_1 \in S_R$, $x_2C = L(v_h) \in S_L$ and $|B| = 0$.

Case II: $B' = B''_1B''_2$, $x = A'B''_1$. For $s_1 = xL(v_h)$, take $A = A'$, $x_1 = x'_1$, $B = B''_1$, $C = L(v_h)$ and $x_2 = \lambda$, which satisfies the requirement, since A is the same as A', B is no longer than B' and $x_2C = L(v_h) \in S_L$.

Case III: $C' = C''C'''$, $x = A'B'C''$. Since $A'B'C' = xu_hy$, we have $C''' = u_hy$ and $C' = C''u_hy$. We now take $A = A'$, $x_1 = x'_1$, $B = B'$, $C = C''L(v_h)$ and $x_2 = x'_2$. Clearly A, x_1 and B satisfy the requirement, since they are the same as A', x'_1 and B'.

It remains to prove $x_2C \in S_L$. Note first that

$$x'_2C''u_hy = x'_2C''C''' = x'_2C' \in S_L$$

So, by Theorem 2.4, $x'_2C''L(v_h) \in S_L$. Since $x'_2 = x_2$ and $C''L(v_h) = C$, we have $x'_2C''L(v_h) = x_2C$, and our proof is complete.□

Theorem 2.9. If S is infinite then either S_L or S_R is infinite.

Proof: If S is infinite then $\{|s||s \in S\}$ is an unbounded set of lengths. By Theorem 2.8, each $s = A_sB_sC_s$ in accord with the ABC property. Since $\{|B_s||s \in S\}$ is bounded, either $\{|A_s|\}$ or $\{|C_s|\}$ is unbounded. But each A_s (C_s) is a factor of a member of S_R (S_L). It follows that either S_R or S_L is infinite.□

3. The algorithm. This section presents the algorithm to determine whether a given semi-Thue system with an inhibitor is uniformly terminating,[1] using the analysis of the preceding section. At the conclusion of this section, it is proved that the halting problem and the derivability problem are also decidable.

The algorithm will begin by constructing finite automata for the languages S_L and S_R.

Definitions related to finite automata. A *nondeterministic finite automaton (automaton* for short) is a finite directed graph each of whose arcs has either a letter or the symbol λ as a label; one node is designated as the *initial node* and any number of nodes are designated as *accepting nodes.* A *walk* through the graph is a sequence

$$N_0, A_1, N_1, \ldots, N_{p-1}, A_p, N_p$$

where the N_i's are nodes and the A_i's are arcs, such that for each i, A_i goes from N_{i-1} to N_i. This walk *goes from* N_0 to N_p. The word *spelled out* by this walk is the result of deleting all λ's from the word $a_1 a_2 \cdots a_p$ where, for each i, a_i is the label of N_i. In particular, the word spelled out is the null word itself if $a_i = \lambda$ for all i. The *language* of an automaton is the set of all words spelled out by walks from the initial node to an accepting node.

We begin by constructing, for the finite language $\{L(v_h) | 1 \leq h \leq k\}$, a loopless finite automaton G_0 with:

(1) exactly one initial node N_I with no arc entering it;
(2) exactly one accepting node N_T with no arc leaving it;
(3) exactly one path from N_I to N and exactly one path from N to N_T, for each node N other than N_I and N_T;
(4) nodes $N_{L1} \ldots, N_{Lk}$ such that, for each i, the path from N_{Li} to N_T spells out $L(v_i)$; in particular, if $L(v_i) = \lambda$ then there is simply a lambda arc, i.e., arc labeled λ, from N_{Li} to N_T; and
(5) for each i, a lambda arc from N_I to N_{Li}.

The construction is done so that

(6) there are no nodes in G_0 other than those required by (1)–(4), and there are no lambda arcs except those explicitly mentioned in (4) and (5).

From G_0 we construct a finite automaton G_L for the language S_L by repeating the following step as often as possible: for a pair of nodes N, N' and an integer $i, 1 \leq i \leq k$,

[1] I am grateful to Friedrich Otto for pointing out a defect in a previous version of this algorithm in [13].

Figure 1: G_0

Figure 2: G_L

if there is a walk from N to N' spelling out u_i, insert a lambda arc from N to N_{Li}, provided that there is not one already there.

Note that G_L and G_0, having the same set of nodes, differ only in that G_L has certain lambda arcs that G_0 does not have. It follows that the construction step is repeated only finitely many times and the graph G_L so constructed is a nondeterministic finite automaton.

As an example, G_0 and G_L for the semi-Thue system

$$\{(a, bc\iota b), (c, da\iota b)\}$$

are shown in Figures 1 and 2, respectively. Thus $S_L = b(db)^*(c \cup da) \ \cup \ d(bd)^*(a \cup bc)$.

The automaton G_R for the language L_R is like G_L except that N_I and N_T are interchanged and all arrows are reversed. More explicitly, we first construct G_{0R} for $\{R(v_h)|1 \leq h \leq k\}$, with

(1′),(2′),(3′) the same as (1),(2),(3);

($4'$) nodes $N_{R1}, \ldots, N_{Rk}$ such that, for each i, the path from N_I to N_{Ri} spells out $R(v_i)$; if $R(v_i) = \lambda$ this is simply a lambda arc; and
($5'$) for each i, a lambda arc from N_{Ri} to N_T.

The construction is done so that

($6'$) there are no nodes in G_{0R} other than those required by ($1'$)–($4'$), and there are no lambda arcs except those explicitly mentioned in ($4'$) and ($5'$).

From G_{0R}, G_R is constructed by repeatedly finding N, N' and i such that there is a walk from N to N' spelling out u_i; and then inserting a lambda arc from N_{Ri} to N'. G_R is similar enough to G_L so that we can carry through detailed reasoning about G_L knowing that corresponding things about G_R will also follow.

Theorem 3.1. The language of G_L (G_R) is S_L (S_R).

Proof for G_L: With Theorem 2.4, the proof that S_L is a subset of the language of G_L is straightforward and is left to the reader.

For the converse let $A_1, A_2, \cdots, A_q$ be the lambda arcs in G_L other than those of G_0, in the order in which they are added in the construction. For each h, $1 \leq h \leq q$, let G_h be the graph that results from G_0 by adding the lambda arcs $A_1, \cdots, A_h$. Thus for $h < q$, G_{h+1} is G_h with A_{h+1} added, and $G_q = G_L$. Where A_{h+1} goes from N to N_{Li}, there is a walk in G_h from N to some node N' spelling out u_i.

Let $P(h, n)$, for $q \geq h \geq 1$ and $n \geq 0$, be the following assertion: For all i, $1 \leq i \leq k$, if w is spelled out by a walk in G_h from N_{Li} to N_T in which the lambda arc A_h occurs at most n times then $w \in S_{Li}$.

Our objective will be to prove that $P(q, n)$ is true for all n. First we note that $P(1, 0)$ is true, since for each i there is only one walk in G_1 without the arc A_1 from N_{Li} to N_T, which is a walk in G_0, and that walk spells out the word $L(v_i) \in S_{Li}$.

Next we prove that, for each $h \leq q$ and n, $P(h, n)$ implies $P(h, n+1)$: Assume $P(h, n)$ and let w be spelled out by a walk W in G_h from N_{Li} to N_T in which A_h occurs $n+1$ times. Let $W = W_1 A_h W_2$ where A_h occurs n times in W_1 but does not occur in W_2. Let W_1 and W_2 spell out w_1 and w_2, respectively. Assume A_h goes from node N to N_{Lg}. By the construction of G_h from G_{h-1} there is a walk W_3 in G_{h-1} from N to N_T spelling out a word $u_g y$ for some y. The walk $W_1 W_3$ spelling out $w_1 u_g y$ has only n occurrences of the arc A_h. Thus $P(h, n)$ implies $w_1 u_g y \in S_{Li}$. The walk W_2 from N_{Lg} to N_T has no occurrences of the arc A_h. Hence $w_2 \in S_{Lg}$, which, by Theorem 2.6, implies $w_1 w_2 \in S_{Li}$. So $P(h, n)$ implies $P(h, n+1)$ for all $h \leq q$ and n.

From this it follows by mathematical induction that, for all n and $h \leq q$, $P(h, 0)$ implies $P(h, n)$. But, for all $h < q$, $P(h, n)$ for all n is equivalent to $P(h+1, 0)$.

Putting all this together we get the proposition

$$\text{For all } n, \ P(q,n)$$

which, by Theorem 2.5, clearly implies that the language of $G_q = G_L$ is included in S_L, concluding the proof that the language of G_L equals S_L.

The proof that the language of G_R is S_R is similar.□

Next it is proved that certain derivations in the semi-Thue system can be obtained from certain walks in G_L and G_R. In particular, loops in the automaton graphs will yield derivation loops in the semi-Thue system. We confine our attention to G_L, knowing that corresponding results about G_R will also be valid. We refine our consideration of the algorithm in obtaining the graph G_L from G_0, considering the sequence $G_0, G_1, \ldots, G_q = G_L$ as defined in the proof of Theorem 3.1.

Definition ($w_0(N)$). For any node N other than N_I, $w_0(N)$ is the word spelled out by the unique walk from N to N_T in G_0.

Theorem 3.2. For each i, $0 \le i \le q$, and for any two nodes N and N' in G_i other than N_I, if there is a walk from N to N' spelling out a word x, then $w_0(N) \to^* x w_0(N')z$, for some z. If the walk has at least one lambda arc then $w_0(N) \to^+ x w_0(N')z$, for some z. These derivations are obtainable effectively.

Proof: We begin by proving the first sentence by mathematical induction on i. That sentence is clearly true for $i = 0$: in this case $w_0(N) = x w_0(N')$, since the only relevant walks are segments of the walks from the N_{Li}'s to N_T, which are disjoint from one another. We now assume it is true for i, $0 \le i \le q-1$ (the i inductive hypothesis) and prove it is true for $i+1$. This proof is itself by mathematical induction on the length of the walk from N to N'. The proposition is clearly true when this length is 0. We assume it for walks of length e (the e inductive hypothesis) and we prove it for walks of length $e+1$. Thus let

$$N_0 = N, N_1, \ldots, N_e, N_{e+1} = N'$$

be this walk in G_{i+1}, and let the word $x_{e+1} = x_e a$ be the word spelled out by it, x_e being the word spelled out by the walk $N_1, \ldots, N_e$ of length e. By the e inductive hypothesis, $w_0(N_0) \to^* x_e w_0(N_e)z$, for some z.

Case I: $a \ne \lambda$. Then the arc from N_e to N_{e+1} labeled a is on the walk in G_0 from N_e to N_T. Consequently, $w_0(N_e) = a w_0(N_{e+1})$, so, for some z,

$$w_0(N_0) \to^* x_e w_0(N_e)z = x_e a w_0(N_{e+1})z = x_{e+1} w_0(N_{e+1})z$$

Case II: $a = \lambda$. Then $N_{e+1} = N_{Lh}$ for some h, and (whether the lambda arc from N_e to N_{e+1} is the new lambda arc of G_{i+1} or one already in G_i) there is a walk from N_e

to N_T in G_i spelling out a word $u_h z'$, for some z'. By the i inductive hypothesis, $w_0(N_e) \to^* u_h z''$, for some z''. Thus we have, for some z''':

$$w_0(N_0) \to^* x_e w_0(N_e) z \to^* x_e u_h z'' z \to x_e L(v_h) z''' = x_{e+1} w_0(N_{e+1}) z'''$$

since $x_e = x_{e+1}$ and $L(v_h) = w_0(N_{e+1})$.

This concludes the proof of the first sentence in the statement of our theorem. The second sentence follows from the fact that a lambda arc in the walk causes Case II to apply, insuring that the derivation has at least one step. Clearly, all these derivations are obtained effectively.□

Definition (loop derivation). If $w \rhd^* y$ and $y \rhd^+ y$ then we say the system has a *loop derivation on y from w*. Note that a loop derivation provides us with one kind of infinite derivation: from $w \to^* x'yz'$ and $y \to^+ xyz$, we get the infinite derivation

$$w, \ldots, x'yz', \ldots, x'xyzz', \ldots, x'xxyzzz', \ldots$$

From the proof of Theorem 3.2 we also get

Theorem 3.3. If node N is on a loop in G_L then there is a loop derivation on $w_0(N)$ in the semi-Thue system.

(It should be mentioned that a loop does not imply that S_L is an infinite set. It is possible for the loop to consist entirely of lambda arcs and for S_L to be finite. For example, this happens with the system with the two rules $(a, b\iota c)$, $(b, a\iota c)$, for which $S_L = \{a, b\}$. Following the constructions we get $a \rhd^+ a$, which we could have inferred in this simple example from the observation that $a \to^2 a\iota c\iota c$.)

Definition. $m = \max_{1 \le h \le k}(|v_h|)$.

Theorem 3.4. If $s \in S$ and $|s| > (2k+1)m$ then there is a loop derivation from s.

Proof: That s has the ABC property (by Theorem 2.8) and that $|s| > (2k+1)m$ together imply that $s = ABC$ where $Ax_1 \in S_R$ and $x_2 C \in S_L$, for some x_1, x_2, and either $|A| > km$ or $|C| > km$.

Case I: $|C| > km$. Since $x_2 C \in S_L$, there is a walk in G_L spelling out C ending at N_T. But, by the construction of G_L, all words spelled out by loop-free paths in G_L have length $\le km$. Thus there is a loop in the walk spelling out C, and so by Theorem 3.3 there is a loop derivation on some suffix of C, and hence a loop derivation from s.

Case II: $|A| > km$. The proof is similar, using the graph G_R for S_R.□

Theorem 3.5. There is an algorithm that determines whether both G_L and G_R are without loops and, if so, enumerates the finite set S.

Proof: By Theorem 3.1, $S_L = L(G_L)$ and $S_R = L(G_R)$. It is easy to tell whether both graphs are without loops.

Assume now that they are without loops. Then the sets S_L and S_R are finite and, by Theorem 2.9, so is S. S_L and S_R are enumerable from G_L and G_R. Let S_0 be the smallest set satisfying (1)–(4) of Theorem 2.7. From the finite enumeration of S_L and S_R we can enumerate S_0. Noting that S is the smallest class that contains S_0 and is closed under (5) of Theorem 2.7, let us recursively define the sets S_{i+1}, for all $i \geq 0$:

$$S_{i+1} = S_i \cup \{xL(v_h), R(v_h)y | xu_hy \in S_i, 1 \leq h \leq k\}$$

Clearly, S_{i+1} is computable from S_i; and $S = \bigcup_{i=1}^{\infty} S_i$. But since S is finite by Theorem 2.9, in computing the successive S_i's, eventually we shall reach an i such that $S_{i+1} = S_i$, which implies that, for this i, $S = S_i$.□

Theorem 3.6. If a semi-Thue system with an inhibitor in which S is finite has an infinite derivation then there exists an $s \in S$ such that $s \triangleright^+ s$.

Proof: Since there is an infinite derivation, there is an infinite S sequence $s_1, s_2, \cdots$ by Theorem 2.1. Because S is finite there must exist p and q, $q > p$, such that $s_p = s_q$. Thus $s_p \triangleright^+ s_p$.□

Theorem 3.7 (Main Theorem). There is an algorithm that produces either a loop derivation in a given semi-Thue system with an inhibitor, or the information that the system is uniformly terminating.

Proof: The algorithm begins by constructing the automaton G_L for S_L. If G_L has a loop then from that loop a loop derivation is effectively determined, by Theorem 3.3.

If G_L has no loop, the analogous automaton G_R for S_R is constructed. If G_R has a loop, analogously a loop derivation is effectively determined.

If neither G_L nor G_R has a loop then, by Theorem 3.5, S is finite and can be enumerated. The $\triangleright$ relation on S is computed and, from this, the $\triangleright^+$ relation on S. If there is an $s \in S$ such that $s \triangleright^+ s$ then we have a loop derivation, effectively. Otherwise, by Theorem 3.6, the system has no infinite derivation.□

Corollary 1. The uniform termination problem for semi-Thue systems with an inhibitor is decidable.

Corollary 2. If a semi-Thue system with an inhibitor has an infinite derivation then it has a loop derivation.

In studying the complexity of the algorithm of Theorem 3.7, we assume that the expression T naming the semi-Thue system is simply the list of its rules, $(u_1, v_1), \cdots (u_k, v_k)$. The following assertions should be clear to the reader: The automaton G_0 is constructible in polynomial time. Each G_{i+1} is constructible from G_i in polynomial time. Since all the automata $G_0, G_1, \ldots, G_q = G_L$ have the same set of nodes and each G_{i+1} is obtained from G_i by adding an arc, q is bounded by a polynomial in the number of these nodes. Thus the construction of the nondeterministic finite automaton G_L is accomplished in polynomial time, and similarly for G_R. It is possible to determine in polynomial time whether G_L (G_R) has a loop, and if so to produce the loop derivation in polynomial time.

However, I cannot prove that the enumeration of S_L (S_R), if it is finite, can be done in polynomial time, since $|S_L|$ ($|S_R|$) may be exponential in the size of G_L (G_R).

Thus if G_L or G_R has a loop, the main algorithm produces a loop derivation in polynomial time and terminates. But if neither G_L nor G_R has a loop then the main algorithm has to enumerate the finite set S. Since there is no polynomial bound on $|S|$ for those S that are finite, the algorithm as written in the proof of Theorem 3.7 is not a polynomial-time algorithm. However, this does not imply that the following has a negative answer:

Open question 1. Is there a polynomial-time algorithm for the problem of whether a given semi-Thue system with an inhibitor is uniformly terminating?

This section closes by settling the two remaining problems of Section 1, the halting problem and the derivability problem for semi-Thue systems with an inhibitor.

Theorem 3.8. The halting problem for semi-Thue systems with an inhibitor is decidable. If such a system has an infinite derivation from a word w then it has a loop derivation from w.

Proof: For a given w and $T = \{(u_1, v_1), \cdots, (u_k, v_k)\}$ let Δ_i be the set of derivations from w of length i, and let Ω_i be the set of all S-sequences that can be taken from those derivations (as in the proof of Theorem 2.1). Note that $\Delta_1 = \Omega_1 = \emptyset$, and each Δ_{i+1} and Ω_{i+1} are readily computable from Δ_i and Ω_i. Consider three possibilities:

(1) For some $i > 1$, $\Delta_i = \emptyset$. Then there is no infinite derivation from w.

(2) There is an i and an S-sequence in Ω_i with a repeated S-expression. Then there is an loop derivation from w.

(3) There is an i and an S-sequence in Ω_i with an S-expression s such that $|s| > m$. Then by Theorem 3.4 there is a loop derivation from s, and hence a loop derivation from w.

If any of these possibilities occurs, then we have the answer to the question after a finite amount of time. It remains to prove that one of them must occur. Theorem 1.1 can be used to prove that if there is no infinite derivation from w then the set of lengths of the derivations from w has an upper bound and possibility (1) will occur. If there is an infinite derivation from w then by Theorem 2.1 there is an infinite S-sequence $s_1, s_2, \cdots$. The proof of that theorem makes it clear that $w \rhd^* s_1$, and hence $w \rhd^* s_i$, for all i. If that S-sequence has a repeated element then possibility (2) will occur. If not there will be no bound on the length of the elements occurring in that S-sequence and possibility (3) will occur.□

Theorem 3.9. The derivability problem for semi-Thue systems with an inhibitor is decidable.

Proof: We define $I(w)$ to be the number of iotas in the word w. Where $w_0, w_1, \cdots, w_p$ is a derivation and $0 \leq i \leq p-1$, let $I(w_{i+1}) - I(w_i)$ be the *weight* of the $(i+1)^{st}$ step, which equals $I(v)$, (u, v) being the rule used. Let the weighted length of a derivation be the sum of the weights of all the steps of the derivation. Given x and y, any derivation of y from x must have a weighted length of $I(y) - I(x)$.

The algorithm that decides whether y is derivable from x simply enumerates all derivations from x whose weighted length equals $I(y) - I(x)$. Because all weights are positive, no line z such that $I(z) > I(y)$ can be part of a such a derivation. Consequently, the list of such derivations can be enumerated readily. Finally, y is derivable from x if and only if the last line of one of these derivations is y.□

Open question 2. Does there exist an algorithm for the following problem: Given a semi-Thue system with an inhibitor and words x and y, does $x \triangleright^* y$ hold?

4. Well behaved derivations.[2] We now turn our attention to semi-Thue systems without an inhibitor, with an emphasis on those having only one rule. Some derivations in these systems turn out to be like those in systems with an inhibitor.

Definition (inhibited rule, inhibition system). If $u_i \rightarrow v'v''$ is a rule of a semi-Thue system T without ι then $u \rightarrow v'\iota v''$ is an *inhibited rule* of T. (v' or v'' can be the null string.) The *inhibition system* of T is the semi-Thue system whose rules are all the inhibited rules of T. An immediate consequence of this definition is

Theorem 4.1. If $x_1, x_2, \ldots$ is a finite or infinite derivation in the inhibition system of the semi-Thue system T then, where each x_i' is x_i with all ι's erased, $x_1', x_2', \ldots$ is a derivation in T.

Definition (well behaved, ill behaved). A derivation D in a semi-Thue system T without ι is *well behaved* if there is a derivation in the inhibition system of T from which D is the result of deleting all ι's. Otherwise D is *ill behaved.*

From Theorems 4.1 and 3.7 we get

Theorem 4.2. There is an algorithm that produces, given a semi-Thue system without ι, either a well behaved loop derivation in the system or the information that the system has no well behaved infinite derivation.

Example 1. The inhibition system T' of the system T whose one rule is $(cb, bbcc)$ has five rules:

$$(cb, \iota bbcc)$$

[2]This section is based on material from [13].

$$(cb, b\iota bcc)$$

$$(cb, bb\iota cc)$$

$$(cb, bbc\iota c)$$

$$(cb, bbcc\iota)$$

This system T' has an infinite derivation. In fact, all we need for this infinite derivation is the one rule $(cb, bb\iota cc)$. The infinite derivation is based on the following loop of length 2:

$$c\underline{cb}$$

$$\underline{c\overline{b}}\overline{b\iota cc}$$

$$\overline{bb\iota cc}b\iota cc$$

(In this and the examples to follow, the part of the line with an underscore is the occurrence of u that is rewritten as the v in the next line, which has an overscore.) The $(2n)^{th}$ line in the infinite derivation is $(bb\iota)^{n-1}cbb(\iota cc)^n$; the $(2n+1)^{st}$ line is $(bb\iota)^n ccb(\iota cc)^n$.

Accordingly, the original system also has an infinite well behaved derivation based on the loop

$$c\underline{cb}$$

$$\underline{c\overline{b}}\overline{bcc}$$

$$\overline{bbcc}bcc$$

The $(2n)^{th}$ line of this infinite derivation is $(bb)^{n-1}cbb(cc)^n$; the $(2n+1)^{st}$ line is $(bb)^n ccb(cc)^n$.

Example 2. The following is an ill behaved derivation in the system with the one rule $(ccb, bbccc)$:

$$cc\underline{ccb}b$$

$$\underline{cc\overline{b}}\overline{bccc}b$$

$$\overline{bbc\underline{cc}}\underline{b}cccb$$

$$bbc\overline{bbccc}c\underline{ccb}$$

$$bbcbbcc\underline{cc\overline{b}}\overline{bccc}$$

$$bbcbbcc\overline{bbccc}bccc$$

To prove that this derivation is ill behaved we note that the inhibition system has six rules, whose right sides are, respectively, $\iota bbccc$, $b\iota bccc$, $bb\iota ccc$, $bbc\iota cc$, $bbcc\iota c$ and $bbccc\iota$. Thus the second line in the corresponding derivation in the inhibition system has six possibilities. It is left to the reader to verify that in each of these six cases the ι will inhibit the replacement of the occurrence of bbc in one of the lines below. (For example, if the *ccb* in the first line is rewritten as *bbιccc* in the second line then the *ccb* in the fifth line cannot be rewritten.)

However, the first five lines of the above derivation form a well behaved derivation, which can be verified by considering the following derivation in the inhibition system:

$$cc\underline{ccb}b$$

$$\underline{cc\overline{b}}\overline{bc\iota cc}b$$

$$\overline{\iota bbc\underline{cc}}\underline{b}c\iota ccb$$

$$\iota bbc\overline{\iota bbccc}c\iota\underline{ccb}$$

$$\iota bbc\iota bbcccc\iota\overline{bbccc\iota}$$

If all the ι's from this derivation are deleted, what remains are the first five lines of the above derivation in T.

The system with the single rule $(ccb, bbccc)$ has an infinite derivation. This is clear from the first five lines of our original derivation, which shows that

$$ccccbb \;\rhd^4\; ccccbb$$

The sixth line of this infinite derivation is the sixth line of the above derivation, which shows that the infinite derivation is ill behaved.

This system has no infinite well behaved derivation. To verify this fact we can refer to Theorem 4.1 and prove that its inhibition system has no infinite derivation. Using the algorithm of Section 3 for this is tedious as it involves enumerating S. Rather than do this we work with a superset of S:

Put $T = \{b^i c^j | 0 \le i \le 2; 0 \le j \le 3\}$. Then it is rather simple to verify using Theorem 2.4 that

$$S_L = \{\lambda, b, bb, bbc, bbcc, bbccc\} \text{ and}$$
$$S_R = \{\lambda, c, cc, ccc, bccc, bbccc\}$$

Thus $S_L \cup S_R \subseteq T$; in particular, for each rule (u, v), we have $L(v), R(v) \in T$.

Next, we use Theorem 2.7 to verify that

$$S \subseteq TT = \{b^i c^j b^k c^m | 0 \le i, k \le 2; 0 \le j, m \le 3\}$$

by proving the following:

(1) every v-section is in TT;
(2) $S_L \cup S_R \subseteq TT$;
(3) (since all rules have the same left side ccb) if $xy = ccb$ and $yz \in S_L$ then $R(v)z \in TT$ (for all right sides v);
(4) if $xy \in S_R$ and $yz = ccb$ then $xL(v) \in TT$; and
(5) if $xccby \in TT$ then $xL(v) \in TT$ and $R(v)y \in TT$.

Parts (1) and (2) are clear. For (3), $yz \in S_L$ implies $yz \in T$; furthermore, $R(v) \in T$ for all rules (u, v); hence $R(v)z \in TT$. The reasoning for (4) is similar to the reasoning for (3).

For (5), $xccby = b^i c^j b^k c^m$ implies that $xcc = b^i c^j$ and $by = b^k c^m$. So, a fortiori, $x \in T$ and $y \in T$. From $L(v), R(v) \in T$, we then get $xL(v), R(v)y \in TT$.

Having proved that $S \subseteq TT$, we complete the proof that the inhibition system has no infinite derivation by proving there is no infinite S sequence (invoking Theorem 2.1).

The proof is by contradiction. Assume there is an infinite S sequence $s_1, s_2, \ldots$. Then each $s_h \in TT$; $s_h \rhd s_{h+1}$; $s_h = b^i c^j b^k c^m$; and $s_{h+1} = b^{i'} c^{j'} b^{k'} c^{m'}$ with $i, i', k, k' \leq 2$ and $j, j', m, m' \leq 3$. The following must be true:

(a) $j \geq 2$ and $k \geq 1$. (Otherwise, nothing could be derived from u_h.)

(b) Either $j' = j - 2$ or $k' = k - 1$.

(Part (b) can be verified by first noting that there is only one occurrence of ccb in $s_h = b^i c^j b^k c^m$. Thus if $s_h \to w$, then w is a word resulting from

$$b^i c^{j-2} / bbccc / b^{k-1} c^m$$

by placing a single ι anywhere between the two slashes and deleting the slashes; s_{h+1} is either the word to the left of the ι or the word to its right. If to the left then $j' = j - 2$; if to the right, $k' = k - 1$.)

Taking $s_1 = b^i c^j b^k c^m$ and $s_2 = b^{i'} c^{j'} b^{k'} c^{m'}$, we have by (b) either $j' = j - 2 \leq 1$ or $k' = k - 1 \leq 1$. If s_3 exists then $j' > 1$ by (a), so $k' \leq 1$. By (a) again we get $k' = 1$.

Where $s_3 = b^{i''} c^{j''} b^{k''} c^{m''}$ we get by (b) either $j'' = j' - 2 \leq 1$ or $k'' = k' - 1 = 0$. By (a) we then see that s_4 cannot exist, which completes our proof that the system with the one rule $(ccb, bbccc)$ has no infinite well behaved derivation, ending our discussion of Example 2.

In [13] there is a much more expeditious algorithm for the problem of whether a given one-rule semi-Thue system has an infinite well behaved derivation. That algorithm does not require consideration of the inhibition system of the given system, but involves a structural analysis of one-rule systems that is well outside the purview of this paper. That structural analysis having been established, the proof in [13] that the one-rule system $(ccb, bbccc)$ has no infinite well behaved derivation takes one quarter the space used in the proof given above as part of Example 2.

This example generalizes. Zantema and Geser [16] prove that a system with one rule

$$(c^m b^n, b^p c^q)$$

in which (1) $p > n$, (2) $q > m$, and (3) either p is a multiple of n or q is a multiple of m, has an infinite derivation. In [13] it is proved that such a system in which either $p < 2n$ or $q < 2m$ has no infinite well behaved derivation. From these two results it follows that any such system in which either $2n > p > n$ and $q = in$ for i an integer ≥ 2 or else $2m > q > m$ and $p = in$ has an infinite ill behaved derivation but no infinite well behaved derivation.

Interestingly, it is proved in [16] that there is no infinite derivation at all in any of the following cases: (1) $p \leq n$, (2) $q \leq m$, (3) $p < 2n$ and q is not a multiple of m, (4) $q < 2m$ and p is not a multiple of n. (Senizergues [14] has extended these results of Zantema and Geser.) In [13] it is proved that if $p \geq 2n$ and $q \geq 2m$ then there is an infinite well behaved derivation.

The two examples discussed in this section illustrate the distinction between well behaved infinite derivations and ill behaved infinite derivations. They are intended to suggest the importance of this distinction to the question of whether the uniform halting problem for one-rule semi-Thue systems is decidable. It is generally conjectured that this problem is decidable, and some progress has been made in proving partial results along that line. However, the question of whether the uniform halting problem for one-rule semi-Thue systems is decidable is very much open. It seems to me that further progress on this question will come only if research workers achieve a structural understanding of ill behaved derivations. There are partial results towards this end in [13].

The halting problem is open for one-rule semi-Thue systems. On the other hand, we have

Theorem 4.3. The derivability problem is decidable for one-rule semi-Thue systems.

Proof: Given x, y and a semi-Thue system whose one rule is (u, v) our algorithm to determine whether y is derivable from x divides into three cases according to the relative lengths of u and v: Case I, $|v| < |u|$; Case II, $|v| = |u|$; and Case III, $|v| > |u|$. In Cases I and II, the finite set of words derivable from x can be enumerated, and the presence or absence of y in the set easily determined. In Case III, the finite set of words derivable from x whose length does not exceed that of y can be enumerated, again yielding an answer to the question.□

For a class of semi-Thue systems an interesting question is, does every semi-Thue system in the class with an infinite derivation have a loop derivation? The result in Section 3 shows that this question has an affirmative answer for the class of semi-Thue systems with an inhibitor. When restricted to well behaved infinite derivations, it has an affirmative answer for all semi-Thue systems. However, it is an open question for one-rule semi-Thue systems; which means that the question restricted to infinite ill behaved derivations is open for one-rule semi-Thue systems.

References

[1] Ronald V. Book and Friedrich Otto, *String-rewriting systems,* Springer-Verlag, 1993.

[2] Noam Chomsky, "Three models for the description of language," *IRE Trans. Inform. Theory,* IT2 (1956), pp. 113–124.

[3] Martin D. Davis, Ron Sigal and Elaine J. Weyuker, *Computability, complexity, and languages,* Academic Press, 1994.

[4] Nahum Dershowitz, "Termination of rewriting," *J. Symbolic Computation,* vol. 3 (1987), pp. 69–115, and vol. 4 (1987), pp. 409–410.

[5] Nahum Dershowitz and Charles Hoot, "Topics in termination," in *Lecture Notes in Computer Science,* no. 690 *(Rewriting techniques and applications,* Claude Kirchner, ed.), Springer-Verlag, 1993.

[6] John V. Guttag, Deepak Kapur and David R. Musser, "On proving uniform termination and restricted termination of rewriting systems," *Siam J. Computing,* vol. 12 (1983), pp. 189–214.

[7] Matthias Jantzen, *Confluent String Rewriting,* Monographs on Theoretical Computer Science, vol. 14, Springer-Verlag, 1988.

[8] Winfried Kurth, *Termination und Konfluenz von Semi-Thue-systemen mit nur einer Regel,* Dissertation, Technischen Universität Clausthal, 1990.

[9] Winfried Kurth, "Explanations to the text," unpublished paper of five pages distributed by the author, summarizing [8].

[10] Winfried Kurth, "One-rule semi-Thue systems with loops of length one, two or three," Report, Abteillung für forstliche Biometrie und Informatik, Universität Göttingen, 1995.

[11] Robert McNaughton, "The development of formal language theory since 1956," *International Jour. Foundations of Computer Science,* Vol. I (1990), pp. 355–368.

[12] Robert McNaughton, "The uniform halting problem for one-rule semi-Thue systems: progress report," Report 94-18, Department of Computer Science, Rensselaer Polytechnic Institute, 1994.

[13] Robert McNaughton, "Well behaved derivations in one-rule semi-Thue systems," Report 95-15, Department of Computer Science, Rensselaer Polytechnic Institute, 1995

[14] Gerard Sénizergues, "On the termination problem for one-rule semi-Thue systems", *Proceedings of R.T.A. 96, Lecture Notes in Computer Science 1103,* 1996, pp. 302–316.

[15] Axel Thue, "Probleme über Veränderungen von Zeichenreihen nach gegeben Regeln," *Skr. Vid. Kritianaia, I. Mat. Naturv. Klasse,* No. 10(1914).

[16] Hans Zantema and Alfons Geser, "A complete characterization of termination of $0^p1^q \to 1^r0^s$," *Proceedings of the 6th Conference on Rewriting Techniques and Applications of 1995,* Vol. 914 of *Lecture Notes in Computer Science*, (J. Hsiang, ed.), pp.41–45, Springer, 1955.