Partial non-underflow and non-overflow of an arithmetic stack

D. T. Goodwin

Department of Computer Science, University of Keele, Newcastle, Staffs., ST5 5BG

This paper continues analysis of the model of arithmetic stack behaviour set up in Goodwin (1977). There conditions computable at high level language compile time were given under which run time stack overflow or underflow could not occur. Because the conditions are strict, it is worth considering the question 'Is it possible for no overflow or underflow to occur, even if the conditions do not apply?' A number of results are presented here, including an algorithm for finding the smallest stack size, if any, which permits an infinity of different program runs to take place.

(Received August 1976; revised July 1978)

This paper is a sequel to Goodwin (1977) in which stack non-overflow and non-underflow conditions are given when strings of a CF language are loaded symbol by symbol on to a stack. A particular symbol exists whose instances each cause the deletion of the top cell of the stack, in some undefined way. This model can easily be applied to the operation of a run time arithmetic stack, where the particular source program acts as the grammar of the CF language, and transfer-to-store instructions or arithmetic operations correspond to cell removal. A disadvantage of the conditions is that they are strict and only answer the question 'Is the program sure to run?' The author suspects that many useful programs would not satisfy this test. This paper considers the question 'Might the program run?' in the sense of whether it can be decided for any CF language whether some finite or infinite number of its strings cause no overflow nor underflow. The answer to this question could conceivably justify a compile time decision either to reject the program outright, or to run it with code inserted for over- and underflow checks at execution time, if no suitable interrupt were available. The former article discussed what high level language programs could be analysed in this way. Certainly all ALGOL 60 and FORTRAN programs are included.

It must be stressed that the case where only a finite number of strings cause no overflow nor underflow is not unduly restrictive and would not by itself justify rejection of the program. Each language string is a stream of variable references, constants and operators which is generated at run time and corresponds to the complete execution of the program. However the language string disappears symbol by symbol as it is generated. Each variable reference at once signals either a fetch to the stack or a transfer-to-store from the stack. Similarly a number input function would place its result on the stack. The operators immediately stimulate operations on the top cells of the stack. It may now be seen that a single particular language string could be used in an 'almost infinite' number of program runs, depending on the particular values of data supplied. Here the word 'almost' is only necessary because in practice a numeric variable can only take one of a finite, though very large, number of possible values, depending on its internal representation.

The question 'Might the program run?' is carefully phrased, and cannot be equated with 'Does the program have a good chance of running?' or 'Is it likely to run?'. These are matters of probability which can only be settled by considering the meanings, and hence the patterns of usage, of strings in the language. This difficult problem is not dealt with at all.

The algorithms of the sections below are often laborious and are given just to demonstrate computability. No claims for efficiency are made.

Notation and assumptions

This paper relies so heavily on the concepts, terminology and notation of the previous article that no further introduction is made here. For ease of reference the numbering of the theorems and lemmas is continued from that paper. Unless stated to the contrary no assumptions are made in what follows as to the existence of $I^+(S)$, $I^-(S)$, $m^+(S)$ and $m^-(S)$.

Shortest stack length

First considered is the question 'What is the shortest stack length necessary to accommodate some string of G throughout the whole of its deposition on the stack?' This problem turns out to be computable and hence answers the simpler question 'Can any string of G be accommodated using a stack of given length L?' At the moment only the upper bound is considered —thus strings of G which use only a few cells of the stack are deemed acceptable even though they may make the stack length negative.

Definitions

For an arbitrary string $s$ of terminals, let $m_U(s)$ be the maximum number of stack cells used at any time during the deposition of $s$. Then $m_U(s) \geq 0$ for all $s$.

For an arbitrary non-terminal $N$, let $N \Rightarrow s_N$ where $s_N$ is a terminal string. Then let

$$m_U(N) = \min_{s_N} (m_U(s_N)).$$

$m_U(N)$ clearly exists whatever the nature of $G$, and whether there are a finite or infinite number of $s_N$.

Similarly for any string $u$, where $u \Rightarrow s_u$, let

$$m_U(u) = \min_{s_u} (m_U(s_u)).$$

Now let $m_U(N)$ be the least maximum stack length used by all strings $s_N$ where the derivation $N \Rightarrow s_N$ begins with the $i$'th rule of $N$. (Let this rule be $N \Rightarrow e_1 \ldots e_j \ldots e_a$). Then

$$m_U(N) = \min_{i} (m_U(e_i)).$$

Consider how this least maximum can arise in the deposition of terminal strings derived from $e_1 \ldots e_j \ldots e_a$. From an argument similar to that used to derive the formula for the $m^+(N)$,

$$m_U(N) = \min_{\text{all} \ e_i} \left\{ \begin{array}{l}
m_U(e_i), \\
1(e_i) + m_U(e_i), \\
\vdots \\
1(e_i) + \ldots + 1(e_{j-1}) + m_U(e_j), \\
\vdots \\
1(e_1) + \ldots + 1(e_{a-1}) + 1(e_a) + m_U(e_a) \\
\end{array} \right\}.$$
This minimum is harder to compute than \( m^*(N) \). In that case the \( l^*(c_j) \) are maxima which are known beforehand, while here
the minimum may arise for any length \( l(c_j) \) of a terminal
string of \( c_j \). It is desired to vary all the lengths involved so that
the maximum stack length arising from any line of the formula
is minimised.

Consider the \( n \)'th line \( l(c_1) + \ldots + l(c_{n-1}) + mU(c_n) \). No other
line of the formula involves \( c_n \) so the minimum \( mU(c_n) \) can
be used, and regarded as a constant when considering other \( c_j \).
Now the only lines to involve \( c_{n-1} \) are the \((n-1)\)'th and \(n\)'th, viz.

\[
\sum_{1}^{n-2} l(c_j) + mU(c_{n-1}), \quad \sum_{1}^{n-2} l(c_j) + l(c_{n-1}) + mU(c_n).
\]

So it is necessary to choose \( s_{n-1}, c_{n-1} \) \( \Rightarrow \) \( s_{n-1}, \) so that max
\( (mU(s_{n-1}), l(s_{n-1}) + mU(c_n)) \) is minimised, where \( mU(c_n) \) is a constant.
For such \( s_{n-1} \) it is not always true that \( mU(s_{n-1}) = mU(c_n) \).
This suggests a new function defined as follows:

\[ mU(u, k) = \min \{ \max \{ mU(s_{n}), l(s_{n}) + k \} \} \]

It follows that \( mU(u, 0) = mU(u) \).

Now if \( u \Rightarrow s_{n} \) and \( v \Rightarrow s_{n} \), then

\[
mU(u, v, k) = \min \{ \max \{ mU(s_{n}, s_{n}), l(s_{n}, s_{n}) + k \} \}
= \min \{ \max \{ mU(s_{n}), l(s_{n}) + mU(u), l(s_{n}) + s_{n} \}
+ k \}\]

\[
= \min \{ \max \{ mU(s_{n}), l(s_{n}) + \max \{ mU(s_{n}), l(s_{n}) \}
\}
+ k) \}\]

\[
= \min \{ \max \{ mU(s_{n}), l(s_{n}) + \max \{ mU(s_{n}), l(s_{n}) \}
\}
(1, s_{n}) \}
+ k) \}\]

\[
\min(mU(u, \max(mU(s_{n}), l(s_{n}) + k))).
\]

From the definition of \( mU(u, k) \) it is clear that it increases
monotonically with \( k \), so that the above minimum is attained
by minimising the second parameter,

\text{i.e.} mU(u, k) = mU(u, \min \{ \max \{ mU(s_{n}), l(s_{n}) + k \} \}
all \( s_{n} \)

\[ = mU(u, mU(v, k)). \]

A third parameter \( r \) for the \( mU \) function is now introduced.
This restricts the terminal strings of \( u \) to those whose generation
trees are no more than \( r \) rule-applications deep. Then if
\( N = c_1 \ldots c_j \ldots c_n \) is the \( r \)'th rule of \( N \).

\[
mU(N, k, r + 1) = mU(c_1, c_j \ldots c_n, k, r)
\]

\[
= mU(c_1, c_j \ldots c_n, mU(c_n, k, r), r)
\]

\[
= mU(c_1, c_j \ldots c_{n-1}, mU(c_n, k, r), r, r)
\]

\[
\ldots
\]

\[
= mU(c_1, mU(c_2, mU \ldots mU(c_n, k, r), r) \ldots r),
\]

\[
\ldots, r)
\]

and

\[
mU(N, k, r + 1) = \min \{ mU(N, k, r), mU(N, k, r + 1) \}, r \gg 0.
\]

(For low \( r \) some of the \( mU \) function values may not exist since
they may eventually call for the evaluation of some \( mU(N, k, 0) \)
—impossible since \( N \) is a non-terminal.)

Now since the \( c_j \), for all \( j, i \) and \( N \) are either terminals or non-
terminals of \( G \), the above relations define a simple recursive
procedure for finding the \( mU(N, 0, r + 1) \), and as many of the
\( mU(N, k, r + 1) \) as are required for computing the \( mU(N, 0, r + 2) \).
The use of the \( r \) parameter guarantees termination of
the procedure, since it specifies that strings of a certain finite
measure of complexity are analysed in terms of simpler ones,
until eventually terminal strings are reached. No doubt a much
faster iterative method could be found which preserved all
intermediate \( mU(c_j) \ldots \) values found so that no duplicated
evaluations were made.

It still needs to be shown how to terminate the algorithm,
having arrived at values for the \( mU(N) \). Certainly

\[
mU(N) = \lim_{r \to \infty} [mU(N, 0, r)]
\]

because by taking sufficiently large \( r \) every desired terminal
string derived from \( N \) will be considered. Also as \( r \) increases,
\( mU(N, k, r) \) decreases monotonically and \( mU(N, k, r) \geq mU(N, 0, r) \geq mU(N) \).
Termination takes place at a certain \( r \) value when for each \( N, k \),

\[
mU(N, k, r + 1) = mU(N, k, r),
\]

because then the algorithm would give the same values for the
\( mU(N, k, r + 2) \) as it had for the \( mU(N, k, r + 1) \). However
for this condition to be useful the \( k \) values that can arise must
be shown to be bounded for all \( N, i \) and \( r \). This is true for each
\( N, i \) because as \( r \) increases \( mU(c_n, 0, r) \) is bounded, \( mU(c_n, 0, r, r) \) is bounded and so on. To avoid unnecessary
evaluations just to test for termination the condition could be expressed

\[
mU(N, 0, r + 1) = mU(N, 0, r)
\]

for all \( N \), and

\[
mU(N, k, r + 1) = mU(N, k, r)
\]

for all \( N \) and all \( k \) already evaluated.

It is also easy to see that by following the progress of the above
algorithm one may construct a string of \( G \) which uses only the
minimum stack length \( mU(S) \). This is by taking note of:

1. the \( i \) value used when evaluating each \( mU(N, k, r + 1) \);
2. the lowest \( r \) for which each of the \( mU(N, k, r) \) attains its
minimum value.

\textbf{Definition}

Let \( mL(N) \) for each \( N \) be the maximal lower bounds analogous
to the \( mU(N) \). The algorithm to compute them would clearly
be similar to the one above. Of course it would be unusual for a
single string \( s \) of \( G \) to have the property that \( mU(s) = mU(S) \)
and \( mL(S) = mL(S) \).

These results are summarised in the following theorem.

\textbf{Theorem 7}

1. It is possible to compute the shortest stack length necessary
to accommodate some string of \( G \) through the whole of its
deposition on the stack. A string which uses just this stack
length can be found.
2. Suppose the strings of \( G \) are to be deposited on a stack where
some cells are occupied already. Then it is possible to com-
pute the least inroad that a string of \( G \) can make upon the
cells already occupied. A string of \( G \) which does this can be
found.

\textbf{Corollary}

Given a positive integer \( L \) it is decidable whether there is any
string of \( G \) which can be deposited on a stack of length \( L \). If
there is, then such a string can be found.

A consequence of this theorem is that for many high level
language programs which at run time use an arithmetic stack,
whether known or accessible to the programmer or not, it is
decidable at compile time what the minimum stack length must
be for any program to run.

The above discussion is particularly relevant to FORTRAN
and ALGOL programs since for these languages \( m^*(S) = 0 \) so
that only upper bounds need be investigated.
Upper and lower bounds together

Definition
A string $s$ of $G$ is ‘stack-bounded in $\{L_1, L_2\}$’ if the length of the stack during the deposition of $s$ remains in the finite interval $[L_1, L_2]$. It is only necessary to consider intervals which include the integer zero, since it is normally assumed that the stack length is zero before deposition of a string of $G$.

Consider first a string $s$ of terminals, which may be part of a complete terminal string of $G$. Let the stack length before loading $s$ be $l_1$ and afterwards $l_2$. Let the Boolean function $sb(s, l_1, l_2)$ be true if $s$ is stack bounded in the interval $[L_1, L_2]$ and false otherwise. Then for any $s_1, s_2$,

$$sb(s_1s_2, l_1, l_2) = sb(s_1, l_1, l) \land sb(s_2, l, l_2)$$

where $l$ is the stack length after the loading of $s_1$. Now let

$$sb(s) = sb(s, l_1, l_2),$$

the Boolean matrix of $sb(s, l_1, l_2)$ values, where $l_1, l_2$ may each take all the values in $[L_1, L_2]$. Then

$$sb(s_1s_2) = \bigcup_{l_1, l_2} sb(s_1, l_1, l_2) \cup sb(s_2, l_2, l_2).$$

Furthermore, let $sb(N) = \bigcup_{s_N} sb(s_N)$. The $(l_1, l_2)$th element of this matrix is true if there is some $s_N$ such that if deposition of loading starts at stack length $l_1$, it finishes at length $l_2$ without exceeding the bounds $L_1, L_2$. Introduce now a second parameter $r$ which restricts the $s_N$ considered to those who generation trees are no more than $r$ rule-applications deep. Also denote by $sb(N, r)$ the matrix arising from the application of the $i$th rule of $N, N \rightarrow c_1 \ldots c_j \ldots c_n$. Then

$$sb(N, r + 1) = sb(N, r) \cup sb(N, r + 1),$$

and

$$sb(N, r + 1) = sb(c_1, r) \cup sb(c_2, r) \cup \ldots \cup sb(c_n, r).$$

The above rules define a simple though laborious procedure for finding the $sb(N, r)$ for all $N$, and so $sb(S, r)$ in particular. The procedure must terminate, i.e. there is an $r_t$ such that for every $N sb(N, r_t + 1) = sb(N, r_t)$. This is because the number of elements in each $sb(N, r)$ is fixed, and as $r$ increases the number of true elements can only increase. If the number of distinct non-terminals is $m$, then there can be at most $m(L_2 - L_1 + 1)^2$ iterations before termination.

Once termination has been reached the truth-values of the matrix elements $sb(S, 0, 0)$ for any $l$ in $[L_1, L_2]$ show whether any string of $G$ is stack-bounded in $[L_1, L_2]$. Furthermore if one of them is true then by working back through the computation one can construct the generation tree of a terminal string which performed in this way.

Theorem 8
It is decidable whether any string of $G$ is stack-bounded in a given interval $[L_1, L_2]$. Where it is shown to exist, such a string can be found.

Theorem 9
It is possible to find a least interval $[L_1, L_2]$ in which any string of $G$ can be stack-bounded.

Proof
Choose any string $s$ of $G$ and find $mU(s)$ and $mL(s)$. Then there are only a finite number of smaller intervals than $[mL(s), mU(s)]$ which include zero. By examining these in turn as in Theorem 8, a least interval can be found.

Definition
$G$ is ‘infinitely stack-bounded in $[L_1, L_2]$’ if there exist an infinity of strings of $G$ which are stack-bounded in $[L_1, L_2]$.

Theorem 10
It is decidable whether $G$ is infinitely stack-bounded in a given interval $[L_1, L_2]$.

Proof
If $G$ is infinitely stack-bounded in $[L_1, L_2]$ then choosing any integer $r$, there exists $r' > r$ such that a string of $G$ is stack-bounded in $[L_1, L_2]$ and has a generation tree which is $r'$ rule-applications deep. It might be hoped that examination of the $sb(S, r)$ for large $r$ would enable this infinite sequence of $r'$ values to be detected or proved absent, but this is not so. This is because the presence of a true element in $sb(S, r)$ only implies the existence of a string with a generation tree no more than $r$ rule-applications deep. However, it is possible to modify the $sb(S, r)$ so as to derive an acceptable test.

Consider the rule

$$sb_i(N, r + 1) = \Pi sb(c_j, r)$$

which arises from the $i$th rule of $N$, namely $N \rightarrow c_1 \ldots c_j \ldots c_n$. While it is not necessary to restrict $sb_i(N, r + 1)$ to $(r + 1)$-deep derivations only, it is now important to include within $sb_i(N, r + 1)$ a record of these new derivations, which have no mention in $sb(N, r)$. Such a derivation only arises if one or more of the $c_j$ components is $r$-deep. Hence it is necessary to record within each of the $sb(c_j, r)$ whether any derivations of $c_j$ are new, i.e. exactly $r$-deep. Therefore let each element—say the $(l_1, l_2)$th element—of the matrices $sb(N, r), sb_i(N, r)$ have one of the possible values:

1. ‘False’—no terminal derivation of $N$ which is no more than $r$-deep remains within $[L_1, L_2]$ for the starting length $l_1$ and the ending length $l_2$.
2. ‘Old’—such a terminal derivation of $N$ exists but is less than $r$-deep.
3. ‘New’—such an exactly $r$-deep terminal derivation of $N$ exists. Possibly ‘Old’ derivations exist too.

It is easy to see how the logical operators $\land$ and $\lor$ can be modified suitably. For consider $sb(c_j, r_1, l_1) \land sb(c_{j+1}, l_1, l_2)$. As before the result of the operator must be ‘False’ if either operand is ‘False’, and if both are ‘Old’ the result must be ‘Old’. However if either is ‘New’ and the other is ‘Old’ the result must be ‘New’ because the result expresses the fact that this combination of derivations of $c_j, c_{j+1}$ will help to form a new $(r + 1)$-deep derivation of $N$ if at least ‘Old’ derivations can be found for the other $c_1 \ldots c_{j-1}$. Thus $\land$ is given by the following table, and the discussion for the new $\lor$ is similar:

<table>
<thead>
<tr>
<th>\lor</th>
<th>F</th>
<th>O</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
</tr>
<tr>
<td>O</td>
<td>F</td>
<td>O</td>
<td>N</td>
</tr>
<tr>
<td>N</td>
<td>F</td>
<td>N</td>
<td>N</td>
</tr>
</tbody>
</table>

Now let the unary matrix operator $O[ ]$ have the effect of replacing all ‘New’ elements in the matrix by ‘Old’. The formulae for the $sb$ matrices now become:

$$sb(N, r + 1) = O[\Pi sb(N, r)] \cup \cup sb(N, r + 1) \geq 0, \text{ for all } N$$

where $sb(N, r + 1) = \Pi sb(c_j, r)$ as $i$ varies over all rules of $N$, for each $N$, and

$$sb(N, 0) = \text{('False')}$$

and for every terminal $t$, and $r > 0$

$$sb(t, r) = O[sb(t, 0)].$$
Example
Consider the simple grammar whose rules are $S \rightarrow aSC$, $S \rightarrow a$, and the interval $[0, 2]$.

\[
\begin{align*}
\text{sb}(a, 0) &= (F \ N \ F), \quad \text{sb}(C, 0) = (F \ F \ N), \\
\text{sb}(S, 0) &= (F \ F \ F).
\end{align*}
\]

Then $\text{sb}(S, 1) = \text{sb}_2(S, 1) = \text{sb}(a, 0) = (F \ N \ F)$.

Now $\text{sb}_1(S, 2) = \text{sb}(a, 1) \cdot \text{sb}(S, 1) \cdot \text{sb}(C, 1)

\[
\begin{align*}
&= (F \ F \ O) \cdot (F \ F \ N) \cdot (F \ F \ O) \\
&= (F \ N \ F).
\end{align*}
\]

and $\text{sb}_2(S, 2) = \text{sb}(a, 1) = O[\text{sb}(a, 0)] = (F \ O \ F)$.

Hence $\text{sb}(S, 2) = O[\text{sb}(S, 1), \text{sb}_1(S, 2), \text{sb}_2(S, 2)] = (F \ N \ F)$.

Similarly $\text{sb}(S, 3) = (F \ F \ F)$.

from which clearly if $r \geq 3$, then $\text{sb}(S, r)$ has no ‘new’ elements and there are no more stack-bounded strings.

Proof of Theorem 10 continued
As with the former $\text{sb}$ matrices let evaluation continue until $r = r_1$ (say), after which no more ‘false’ elements are overwritten with ‘old’ or ‘new’. For the ‘old’ and ‘new’ elements in the $\text{sb}(N, r_1)$, for all $N$, one of three possibilities remains:

1. The ‘new’ elements are all replaced by ‘old’ elements as $r$ increases, so that for some $r_2 > r_1$, no ‘new’ elements remain in any $\text{sb}(N, r)$ if $r \geq r_2$.

2. The ‘new’ and ‘old’ elements become fixed for $r \geq r_3 \geq 1$.

This is a special case of 3.

3. While the number $k$, (which is no more than $m(L_2 - L_1 + 1)^2$) and identity of the ‘true’ elements are fixed, they oscillate in possessing ‘old’ or ‘new’ values, the only restriction being that for any $r$, they cannot all be ‘old’, for otherwise they would continue to be ‘old’ for all higher $r$. Thus there are $2^r - 1$ possibilities for the set of ‘old’ and ‘new’ values. Cases 1 and 3 can now be distinguished by examining the $\text{sb}(N, r_1 + 2^2)$ matrices for all $N$. Either all elements will be ‘false’ or ‘old’, or there will exist distinct $r_4$, $r_5$, in the interval $[r_1, r_1 + 2^2]$ for which $\text{sb}(N, r_4) = \text{sb}(N, r_5)$ for all $N$. The former situation is case 1, showing that $G$ is not infinitely stack-bounded in $[L_1, L_2]$, and the latter is case 3. Then provided an element $\text{sb}(S, 0, l)$, for some $l$, has the ‘new’ value for some $r'$ in the oscillating sequence, then there will be an infinite number of such $r'$, and $G$ is infinitely stack-bounded in $[L_1, L_2]$. Hence the theorem is proved.

Theorem 11
If $G$ is known to be infinitely stack-bounded in some interval $[L_1, L_2]$ it is possible to find a least interval in which it is infinitely stack-bounded.

Proof
Since zero must be in any such interval, there are only a finite number of intervals shorter than $[L_1, L_2]$ which need to be examined. The procedure of Theorem 10 can be used for each such interval.

The remainder of the article is devoted to proving computable conditions under which $G$ is infinitely stack-bounded in some interval, and determining a least such interval in the general case. Various trivial ‘theorems’ arise on the way which have some slight value of their own—it is useful to know that the $l(N)$ can be generated parametrically as well as with unwieldy recurrence relations.

Some Diophantine equations
Some Diophantine equations are used a number of times in the succeeding sections. Their relevant properties are therefore summarised here. One is:

\[
a_1 x_1 + a_2 x_2 + \ldots + a_i x_i + \ldots = a^* \tag{1}
\]

where the $a_i$ and $a^*$ are any given integers and solutions in the $x_i$ are sought for integers $x_i > 0$, for all $i$. The results given in Mordell (1969) may be summarised and trivially extended as follows:

1. A solution can only exist if the highest common factor of all the $a_i$ divides into $a^*$.

2. Given 1, then there are always an infinite number of solutions if the $a_i$ do not all have the same sign. The $x_i$ can then be expressed as linear functions of a finite number of parameters.

3. If all the $a_i$ do have the same sign then solutions exist only if $a^*$ also has the same sign. If so then there are either no solutions or a finite number of solutions.

4. Given fixed $a_i$ and integers $A+$ and $A-$ it is computable whether there exists any $a^* < A+$, or alternatively $a^* > A-$, for which there is a solution of (1), and if so what values it may respectively take.

Now let

\[
z_{ij} = b_{ij} x_1 + b_{ij} x_2 + \ldots + b_{ij} x_i + \ldots + b_{ij} x_n,
\]

for $i = 1$ or 2 and $1 \leq j \leq n$, and consider the two simultaneous equations

\[
\begin{align*}
&z_{11} y_1 + z_{12} y_2 + \ldots + z_{1i} y_i + \ldots + z_{1n} y_n = 0, \\
&z_{21} y_1 + z_{22} y_2 + \ldots + z_{2i} y_i + \ldots + z_{2n} y_n = 0.
\end{align*}
\]

(2)

to be solved for non-negative $y_i$ and $x_{ij}$ in terms of the arbitrary integers $b_{ij}$. The needed result is:

5. It is computable whether there is a solution to the equations (2) for given values of the $b_{ij}$. The author has a proof of computability, but for brevity it is omitted from this paper.

N-non-recursive subtrees
In subsequent sections it will be shown how to compute whether a particular subset of the basic cycles of $G$ can appear in the derivation of a single string of the language. These subsets will in turn be used to provide formulae for generating all the $l(N)$. As a first step $N$-non-recursive subtrees are defined and their maximum number of possible occurrences is shown to be computable.

Definitions
A generation tree is ‘$N$-non-recursive’ if its base-node is the non-terminal $N$ and it contains no other instance of $N$. The base-node of such a tree is termed ‘non-recursive’ for brevity in what follows. However, note that $N$-non-recursiveness does not necessarily imply zero-recursiveness.

The marks $N \Rightarrow u$ are defined to mean either that $N = u$, or that $N \Rightarrow u$.

For each pair of non-terminals $N'$ and $N$ of $G$, let the maxi-
num number of $N$-non-recursive subtrees which can exist in the 
tree of any expansion of $N'$ be $I(N', N)$. Trivially also 
define $I(t, t) = 0$ for any terminal $t$.

**Examples**

All zero $I(., .)$ values are omitted.

1. \[
\begin{align*}
S & \rightarrow NN \\
N & \rightarrow d \\
& \quad (S, N) = 2, \\
& \quad I(S, S) = 1, \\
& \quad I(N, N) = 1.
\end{align*}
\]
2. \[
\begin{align*}
S & \rightarrow NS \\
N & \rightarrow d \\
S & \rightarrow e \\
& \quad (S, S) = 1, \\
& \quad I(S, N) = \infty, \\
& \quad I(N, N) = 1.
\end{align*}
\]

3. \[
\begin{align*}
S & \rightarrow MSS \\
M & \rightarrow Nf \\
N & \rightarrow d \\
S & \rightarrow e \\
& \quad (S, S) = \infty, \\
& \quad I(S, M) = \infty, \\
& \quad I(S, N) = \infty, \\
& \quad I(N, N) = 1, \\
& \quad I(M, M) = 1.
\end{align*}
\]

4. \[
\begin{align*}
S & \rightarrow PS \\
P & \rightarrow p \\
S & \rightarrow Nk \\
N & \rightarrow d \\
N & \rightarrow fN \\
& \quad (S, S) = 1, \\
& \quad I(S, N) = 1, \\
& \quad I(N, N) = 1, \\
& \quad I(N, P) = \infty.
\end{align*}
\]

It is easy to see that for any $N$, either $I(N, N) = 1$ or $I(N, N) = \infty$. Also if $N_1 \Rightarrow uN_2v$ for some $u, v$ then $I(N, N_1) \leq I(N, N_2)$. Note that it does not follow that if $N'$ is in cycle then $I(N', N)$ will be infinite. The following lemma gives a precise condition for infiniteness.

**Lemma 5**

$I(N', N)$ is infinite if and only if there exists $N_1$ such that $N' \Rightarrow uN_1v$ for some $u, v$.

and $N_1 \Rightarrow u_1N_1v_1$ for some $u_1, v_1$, where $u_1v_1 \Rightarrow u_2N_2v_2$ for some $u_2, v_2$.

**Proof**

1. **Sufficiency**

   Given that $N$ exists with the above property, then taking any 
   integer $j$ and executing the $N_1$ cycle $j$ times,

   \[
   N' \Rightarrow uN_1v \Rightarrow uu_1N_1v_1v,
   \]

   and so ignoring the orders of symbols,

   \[
   N' \Rightarrow uu_1N_1v_1vN_1.
   \]

   Thus the number of derivable instances of $N$ in an expansion of

   $N'$ is unbounded. Each such $N$-instance either is or can be used 
   to derive a non-recursive instance of $N$.

2. **Necessity**

   Let the largest number of non-terminals on the righthand side of
   any rule of $G$ be $q$. Consider an expansion of some non-
   terminal $N'$ of $G$ such that its tree contains no path between
   two non-terminals longer than $n$ arcs, where $n$ is the number of
   distinct non-terminals of $G$. Such a tree can have no more than
   $q^n$ zero-recursive non-terminal instances. Now if $I(N', N)$ is
   infinite there must be an expansion of $N'$ with more than $q^n$
   non-recursive instances of $N$, each of which has the property that
   the path from $N'$ to it passes through no other non-recursive
   instances of $N$. Thus these instances may be regarded as the
   leaves of a tree which is derived as follows: take the original
   tree of the expansion and at each node remove every subtree
   which does not contain a non-recursive instance of $N$. On this
   new tree there can still be no more than $q$ subtrees attached to
   any one node so that from the argument above there must be a
   path from $N'$ to a non-recursive instance of $N$ which is more
   than $n$ arcs long. Hence there exists a non-terminal $N_1$, not

   necessarily distinct from $N'$ or $N$, with more than one instance
   on a path from $N'$ to $N$.

   i.e. $N' \Rightarrow uN_1v$ for some $u, v$

   and $N_1 \Rightarrow u_1N_1v_1$ for some $u_1, v_1$.

   Now suppose that there are no $u_2, v_2$ such that $u_1v_1 \Rightarrow u_2N_2v_2$,

   so that use of the $N_1$ cycle introduces no new instances of $N$.

   Consider the expansion of $N'$ which is as before except that now
   the $N_1$ cycle (together with one instance of $N_1$) is
   removed. Then either the path considered still has more than $n$
   non-terminal instances on it, in which case the argument can be
   repeated, or no more than $n$ instances. Repeating for each such
   path, a tree can be derived which has more than $q^n$ non-
   recursive instances of $N$ with no path more than $n$ arcs long—
   an impossibility. Therefore there exist $u_2, v_2$ such that

   $u_1v_1 \Rightarrow u_2N_2v_2$.

   The condition of this lemma is clearly computable, for it is

   equivalent to determining whether certain paths exist between

   nodes of a subgraph of $G$, which is finite.

   Now let $R_i(N') = N' \rightarrow C_{i1}C_{i2} \ldots C_{ij}C_{ip}$.

   Then if $I(N', N)$ is finite,

   $I(N', N) = 1$ if $N' = N$, and

   $I(N', N) = \max \left[ \sum_{i=1}^{p} I(C_{ij}, N) \right]$ otherwise,

   from its definition. From Lemma 5 above the formula is never

   usefully recursive in any quantity $I(N_1, N)$, which makes its

   computation finite and straightforward.

**Co-existent basic cycle subsets**

**Definition**

A subset of the basic cycles of $G$ is 'co-existent' if there exists a

string of the language which uses each of the cycles of the subset

at least once, and no others at all.

In this section an algorithm is developed for determining

whether a basic cycle subset is co-existent. Hence all co-

existent subsets can be found.

There are three possible co-existence relationships between

two basic cycles:

1. The use of one cycle may prohibit the use of the other.

   e.g. \[
   \begin{align*}
   &S \rightarrow N \\
   &N \rightarrow aN \\
   &N \rightarrow M \\
   &M \rightarrow bM
   \end{align*}
   \]

   The strings are \{ $a^m \ | \ m \geq 1$ \} \{ $b^m \ | \ m \geq 1$ \}

2. The cycles may be used independently.

   e.g. \[
   \begin{align*}
   &S \rightarrow MN \\
   &N \rightarrow aN \\
   &N \rightarrow M \\
   &M \rightarrow bM
   \end{align*}
   \]

   The strings are \{ $b^m a^m \ | \ m, n \geq 1$ \}

3. One cycle may be used only if the other one is also used.

   e.g. \[
   \begin{align*}
   &S \rightarrow N \\
   &N \rightarrow a \\
   &M \rightarrow bM \\
   &N \rightarrow MNd
   \end{align*}
   \]

   The strings are \{ $a, b^m a^m d \ | \ m \geq 0, n \geq 1$ \}.

Note that in the simple grammar \{ $S \rightarrow SS, S \rightarrow a$ \}, the first

rule defines two basic cycles. A particular application of this

rule, however, may be regarded as the use of either cycle, but

not both. Thus these cycles may be used independently, and this

grammar formally has three co-existent basic cycle subsets,

according to the selection of either or both of these cycles.

Consider the tree of a string of a language in which two cycles

are used. Let $N_1$ and $N_2$ be non-terminal instances, one in each of

the two cycle occurrences. Consider the paths from the
base-node $S$ to $N_1$ and to $N_2$. These paths must diverge at some
node of the tree, perhaps at the base-node, and when they do go must be
where a particular rule
\[ N \to u'N'_1v'N'_2w' \]
where $t$ varies depending on $C_i^f_N$.

In the algorithm for determining the co-existence of the sub-
set $B_i^N$, terminal expansions were found which each included a
path selection but had no further cycles. Call them
\[ e_1^i, \ldots, e_{s_i}^i, \ldots, e_{s_i}^i, \]
and let the number of instances of
\[ C_i^f_N \in e_i^N \] be $n_i^{0,ij}$.

for any particular $j$, $1 \leq j \leq s_i$. The generating functions for the $l(N)$ can now be presented.

**Lemma 6**

1. Every terminal expansion of $N$ has a length
\[ f_i^N(x_{ijk}) = l(e_i^N) + \sum_{j=1}^{s_i} \sum_{k=1}^{r_i} x_{ijk} l_{ijk}^N \]
for some $i, h$ and a vector $x_{ijk}$ of non-negative integers $x_{ijk}$
provided that for fixed $i, h$ and $j$
\[ \sum_{k=1}^{r_i} x_{ijk} = 0 \text{ only if } n_i^{0,ij} > 0 . \]

2. Conversely, for every selection of $i, h$ and $x_{ijk}$ (according
to the rule just given), the above expression gives the length
of some expansion of $N$.

Before giving the proof and an explanation of the $f_i^N$ is useful: $f_i$
fixes the cycle subset and $h$ specifies a particular minimal
expansion of $N$ which will permit the cycles of $B_i^N$ to be used.
The $x_{ijk}$ fix the number of additional times a particular cycle
is to appear, with its result yielding a particular zero-reducible
tree structure. Recursiveness in the expansion of the result of
the cycle is allowed for by varying some of the $x_{ijk}$. The
condition of 1 specifies that each cycle in the subset must
appear at least once. Only when the $e_i^N$ concerned contains a
particular cycle $C_i^f_N$ can
\[ \sum_{k=1}^{r_i} x_{ijk} = 0 \text{ are } \]

**Proof of Lemma 6**

1. Consider a terminal expansion $e_i^N$ of $N$. Then either it is
0-reducing or its tree contains a basic cycle whose result is
0-reducing. In the former case the lemma is proved. In
the latter case it means that a node $N'$ has a proper subtree
with base-node $N'$ such that
\[ N' \Rightarrow uN'v \Rightarrow uwv \]
where $uw$ is 0-reducing. Now form a new tree and a corre-
sponding new expansion $e_{ij}^N$ by putting the latter instance of
$N'$, with its subtree ($N' \Rightarrow w$), in place of the former instance
of $N'$. This has the effect of removing the basic cycle with the
0-reducing result:
\[ l(e_i^N) = l(e_{ij}^N) + l(uw) \]
\[ = l(e_{ij}^N) + l_{ijk}^N \]
for some $i, j, k$.

The next step is to reduce $e_{ij}^N$ in the same way, ensuring
when removing cycles that for each non-terminal in the
removed cycles some instance is retained in the new
reduced subtree. Eventually no further cycle instances can
be removed, bearing the above restriction in mind, and the
resultant tree's expansion is one of the $e_i^N$. This establishes
the formula.

2. From the construction of any expansion $e_i^N$, it is possible to
add as many instances of cycles $C_i^j$ with 0-reducing results
as desired. Hence any particular formula certainly has a
corresponding expansion of $N$. 

158

The Computer Journal Volume 23 Number 2
Computability of the $l(N)$

**Theorem 12**

1. It is computable whether any expansion of $N$ has a given length $L$, and also whether there is a finite or an infinite number of such expansions. In the former case they can all be determined. In the latter case formulae can be derived from which as many such expansions can be generated as required.

2. As 1, but with the expansions required to be in a finite interval $[L_1, L_2]$.

3. As 2 but with the expansions in infinite intervals $[-\infty, L_2]$ or $[L_1, +\infty]$.

**Proof**

1. From Lemma 6 the problem is equivalent to solving all the equations

$$f_N(x_{ijk}) = L$$

in the $x_{ijk}$, for all $i$ and $h$. These are not simultaneous equations. Each separately may contribute solutions for the $x_{ijk}$ and hence expansions of $N$. Such an equation is an example of the Diophantine equation (1) given above. A solution in the $x_{ijk}$ specifies the number of instances of each basic cycle that each corresponding satisfactory expansion may have. There are only a finite number of terminal expansions for each such solution. They can be found exhaustively by developing generation trees in which all combinations of alternative expansions of the non-terminals are taken.

2. Repeat 1 for all $L$ in $[L_1, L_2]$.

3. Derive all suitable $a^*$ in $[-\infty, L_2]$ or in $[L_1, +\infty]$ using Diophantine equation result 4 above (Mordell, 1969). Then continue as just above.

**Theorem 13**

As Theorem 12, but considering any string $u$ rather than $N$.

**Proof**

1. Permute the symbols of $u$ until $u = wv$, where $v$ consists of all the terminals of $u$, and

$$w = N_1 \ldots N_y \ldots N_b,$$

being all the non-terminals of $u$. Then the problem is equivalent to solving separately all the equations

$$L = l(v) + \sum_{y = 1}^{b} f_N(x_{ijk})$$

for all possible combinations of the $i$ and $h$ varying separately for each $N$. These equations have the same Diophantine form as those of Theorem 12 above.

2. As for Theorem 12.

**Is there any infinite stack-boundedness?**

**Theorem 14**

$G$ is infinitely stack-bounded in some interval $[L_1, L_2]$ if and only if there exists a cycle $C$, not necessarily basic, such that

$$l(C) = 0 \text{ and } llh(C) = 0.$$

**Proof**

1. **Sufficiency**: take a string of $G$ whose derivation includes a non-terminal of the cycle for which the conditions are true. Then apply the cycle to generate an infinity of stack-bounded strings.

2. **Necessity**: suppose it is known that $L_1, L_2$ exist for which an infinity of strings are stack-bounded in $[L_1, L_2]$. The idea of the proof is to show that there is a suitable cycle with zero left-hand length, and then that there are sufficient of these to guarantee that one of them must have zero righthand length as well.

Consider the infinite set

$$B = \{ b_1 \ldots b_i \ldots \}$$

which consists of all the finite strings $b_i$ of $G$ which are stack-bounded in $[L_1, L_2]$. Since there are only a finite number of rules $S \rightarrow w$, then there must exist one of them, say $S \rightarrow w_1$, which is the first rule applied in the derivations of an infinite subset $B_1$ of $B$. Similarly there must exist a rule $N_i \rightarrow w_2$, where $S \rightarrow w_1 = u_1N_i, u_2$, for some $u_1, v_1$, such that an infinite subset $B_2$ of $B_1$ start off their derivations with $S \rightarrow w_1, N_i \rightarrow w_2$, so that $S \Rightarrow u_1w_2v_1$. Similarly for any integer $k$, there exists an infinite set $B_{k+1}$ each of whose members has a derivation which begins

$$S \Rightarrow u_1u_2 \ldots u_kw_ka_{k+1} \ldots v_2v_1,$$

with a corresponding non-terminal sequence $N_1 \ldots N_f \ldots N_k$, each of which is derived from the last. By choosing $k$ greater than the number of distinct non-terminals of $G$, the later $N_f$ must coincide with earlier ones in the sequence. Hence there exists a non-terminal $N$ such that, for any given integer $f$, $N$ occurs $f + 1$ times in the sequence $N_1 \ldots N_f \ldots N_k$, for sufficiently large $k$. Call these instances of $N, N_{j_1}, \ldots, N_{j_f}, N_{j_r+1}$, Then for each $i$

$$S \Rightarrow u_1u_2 \ldots u_kN_{j_1}v_{j_1} \ldots v_{i+1}v_i \Rightarrow u_1u_2 \ldots u_kN_{j_1}v_{j_1} \ldots v_{i+1}v_i \equiv y_i^1 \text{ (say)}$$

Let $u_1u_2 \ldots u_k = y_i^0, v_{j_1} \ldots v_{i+1} = v_i^0$, and let $u_{j_1+1} \ldots u_{j_r+1} \equiv u_{j_1} \ldots v_{j_r+1} \equiv v_{j_r+1}^0$. Then

$$y_{i+1}^1 = y_i^0 \equiv u_1u_2 \ldots u_kN_{j_1}v_{j_1} \ldots v_{j_r+1}v_{j_r+1} \ldots v_i$$

and $S \Rightarrow u_1u_2 \ldots u_kN_{j_1}v_{j_1} \ldots v_{j_r+1}v_{j_r+1} = y_{i+1}^1$.

Consider the loading of a particular terminal derivation of $v_{j_r+1}$ on to the stack, and let the stack-length after the deposition of the terminal string derived from $u_i$ be $l$, for $0 \leq l \leq f$. Choose $f > L_2 - L_1 + 1$. Then because the terminal string does not exceed the bounds $L_1$ and $L_2$, there must be distinct integers $r$ and $s$ such that

$$l_r = l_s = l \text{, i.e. } l(u_1 \ldots u_k) = 0,$$

and there exists a chained recursive derivation

$$N_{j_{r+1}} \Rightarrow u_{r+1} \ldots u_kN_{j_{r+1}}v_{r+1} \ldots v_{r+1}^0$$

whose left-hand length is zero. More generally, for sufficiently large $f$ there exists $l, L_1 \leq l \leq L_2$, and distinct integers $r_1, r_2, \ldots, r_{L_2-L_1+2}$, all $\leq f$, such that

$$l = l_{r_1} = l_{r_2} = \ldots = l_{r_{L_2-L_1+2}}.$$

Consider now the loading of a terminal derivation of

$$v_{r_1} \ldots v_{r_2} \ldots v_{r_{L_2-L_1+2}} \ldots v_i$$

on to the stack. Let $l_r$ be the stack length just before $v_i$ is loaded. Consider the stack lengths $l_r$, for $0 \leq l \leq L_2 - L_1 + 2$. Then there exist integers $a, b, a > b$, such that

$$l_r = l_s = l \text{, i.e. } l(v_{r_1} \ldots v_{r_{a-1}}) = 0,$$

and from above

$$l(v_{r_{a+1}} \ldots v_{r_{b+1}}) = 0.$$

Hence the cycle corresponding to the chained recursive derivation

$$N \Rightarrow u_{r_{a+1}} \ldots u_kN_{r_{a+1}} \ldots v_{r_{b+1}},$$

has a zero left-hand length and zero righthand length, and hence zero overall length, as required.

**Theorem 15**

It is computable whether there exist integers $L_1, L_2$ such that $G$ is infinitely stack-bounded in $[L_1, L_2]$. If so such an interval can be found.
Proof

From Theorem 14, it is sufficient to show that it is computable whether there exists a cycle $C$ for which $l(C) = 0$ and $lh(C) = 0$. Let $\{C_1, \ldots, C_t\}$ be a composite subset of the basic cycles of $G$. Then consider a cycle $C$ composed from $y_j$ instances of $C_j$, for each $j$, $1 \leq j \leq t$. Then

$$l(C) = \sum_{j=1}^{t} y_j \cdot l(C_j)$$

$$lh(C) = \sum_{j=1}^{t} y_j \cdot lh(C_j).$$

In general there are a finite number of alternative expressions for $l(C)$ and $lh(C)$, each on the lines of the proof of Theorem 13. For each combination of alternatives the simultaneous equations

$$l(C) = 0$$

$$lh(C) = 0$$

take the form of Diophantine equations (2) and the Diophantine result 5 shows it to be computable whether a solution exists.

If one of these sets of equations proves to be solvable, then the algorithm for deciding that also leads readily to one solution at least. This means that one cycle with the desired properties is known. Suppose that $N$ is a non-terminal in the cycle, say $N \Rightarrow u_iNv_i$, where $u_i$ and $v_i$ are terminal strings. Then choose any derivation $S \Rightarrow u'Nv'$ for any terminal $u'$ and $v'$, and a non-cyclic derivation $N \Rightarrow s_N$. One can then show that an infinity of terminal strings generated by the cycle is $u'u_1s_Nv'n'$, $n \geq 1$, whose upper and lower bounds are the same as those for $u'u_1s_Nv'n'$, where $n = 1$. Thus one suitable interval $[L_1, L_2]$ can be found if any exist at all.

Theorem 16

It is possible to find a smallest interval, if any, in which $G$ is infinitely stack-bounded.

Proof

This theorem just brings together the results of Theorems 11 and 15.

References
