Conditions for underflow and overflow of an arithmetic stack

D. T. Goodwin

Department of Computer Science, University of Keele, Keele, Staffordshire ST5 5BG

The behaviour of an arithmetic stack is formally described as the loading of an arbitrary string of a context free language symbol by symbol on to a stack. Instances of a special symbol in the string being loaded invoke an operation which removes the top cell of the stack in some undefined way. Necessary and sufficient conditions for stack length boundedness are stated and proved. One application of the results concerns the choice between compile time and run time checks for underflow and overflow. Another concerns the testing for applicability of a certain algorithm for inverting Metcalfe-Reeves translators.

(Received June 1975)

The formal device studied by this paper is a push-down store or stack on to which a string of symbols is being loaded, one symbol at a time. One special symbol C causes the top two cells of the stack to be replaced by one cell, in some undefined way. Similarly, other special symbols \( P_1, P_2, \ldots, P_r \) may be used. Any instance of \( P_r \) in the string is not loaded, but instead causes a particular (non-identity) permutation of a fixed number \( r_0 \) of the top cells of the stack. The only restriction on the input is that it is an arbitrary string of a fixed context free (CF) language, as defined briefly in the section on notation below.

**Example**

Let the only \( P_r \) be denoted by X, which interchanges the top two cells of the stack, and let C concatenate the top two cells of the stack. Then the string \( cbXCaXC \) is processed from left to right as follows:

<table>
<thead>
<tr>
<th>Stack contents (top cell at right)</th>
<th>String to be loaded</th>
</tr>
</thead>
<tbody>
<tr>
<td>—</td>
<td>cbXCaXC</td>
</tr>
<tr>
<td>c</td>
<td>bXCaXC</td>
</tr>
<tr>
<td>c b</td>
<td>XCaX</td>
</tr>
<tr>
<td>b c</td>
<td>CaX</td>
</tr>
<tr>
<td>bc</td>
<td>aX</td>
</tr>
<tr>
<td>bc a</td>
<td>X</td>
</tr>
<tr>
<td>a bc</td>
<td>C</td>
</tr>
</tbody>
</table>

This device and the special symbols C and X of this example were used by Metcalfe (1964) and Reeves (1967) to edit the output from a syntax driven translation system. In Goodwin (1975) a translation system on the same lines was developed which allowed the input/output grammars or 'translators' to be inverted automatically under certain given conditions. These conditions were restrictions on the ways in which translators were able to form items on the edit stack, using C. However, no proofs were given of stated algorithms for determining for an arbitrary translator whether the conditions were true. These proofs are given here.

Another application of more general interest is the provision of compile time and run time checks on whether certain high level language programs will cause underflow or overflow of a run time arithmetic stack. FORTRAN programs whose translated form uses a hardware or software stack could be processed, and also more generally, ALGOL 60 and certain POP2 programs, where the user can access the stack directly, and where recursive function calls are allowed. These possibilities are discussed further at the end of the paper.

In what follows only the number of cells on the stack is studied, not their contents. This explains how it is possible to leave C undefined except in so far as it removes the top cell. As seen above, it could have the additional effect of concatenating cell contents, or in the second application it might be an arithmetic operator, or a transfer to store instruction.

The theorems below are on the following loosely described topics:

- **Theorem 1** When the stack length is unbounded
- **Theorem 2** Uniqueness of the stack length
- **Theorem 3** Conditions for boundedness of the stack length after loading of the string
- **Theorem 4** Conditions for boundedness of the stack length during loading of the string
- **Theorem 5** Restrictions on stack length bounds during loading
- **Theorem 6** Disturbances caused by permutations of the stack cells.

The next two sections give the notation to be used in the discussion of context free languages and graphs, and include some simple lemmas. The treatment of the number of cells then proceeds.

1. **Notation**

A context free language is a language defined by a grammar \( G \) as follows. \( G \) consists of:

1. A finite alphabet of 'terminal symbols'. Here the unsuffixed letters \( a, b, c \ldots \) are used for these. A general terminal is denoted by \( t \).
2. A finite alphabet of 'nonterminal symbols'. A general nonterminal is denoted by \( N \), which is often suffixed. Sometimes \( N_1, N_2, \ldots, N_0, N_0, \ldots, N_n \) are used to denote all the nonterminals of \( G \).
3. A finite number of 'production rules' each of which is of the form

\[ N \rightarrow C_1C_2 \ldots C_j \ldots C_p \]

where each of the \( C_j \) may be either a single terminal or a single nonterminal. There may be a number of rules with \( N_k \), say, on the left hand side. These are called 'the rules of \( N_k \'). Any rule of \( G \) is identified as \( R(N) \), the \( i \)th rule of \( N \), where the rules of \( N \) are numbered in some arbitrary order. When all the rules of \( N \) are being considered at once, the \( i \)th rule is written

\[ N \rightarrow C_{i1}C_{i2} \ldots C_{ij} \ldots C_{ip} \]

(It is understood that in general the value of \( p \), the index of the last symbol of a rule, will vary from rule to rule).

4. A special nonterminal \( S \) which is one of \( \{ N_1, N_2, \ldots, N_n \} \).

The word 'string' is now restricted to meaning an arbitrary concatenation of symbols, possibly empty, which unless other-
wise stated may be any out of the alphabets 1 and 2. A general string is denoted by \( u, v, \) or \( w, \) possibly suffixed. A general string of terminals only is denoted by \( s. \) \( x \) is used for a single symbol which is either a terminal or a nonterminal. \( u^n \) denotes the string \( uu \ldots u \) where \( u \) is repeated \( n \) times.

If a string \( w_0 \) contains a nonterminal \( N_0 \) it can be 'expanded' by an application of a rule \( R_i \( N_0 \). \) Let \( w_0 = u_0 N_0 v_0. \) Having chosen \( i, \) this instance of \( N_0 \) is replaced in \( w_0 \) by

\[
C_{i1} C_{i2} \ldots C_{i j} \ldots C_{i p}.
\]

This operation is written

\[
w_0 = u_0 N_0 v_0 \Rightarrow u_0 C_{i1} \ldots C_{i j} \ldots C_{i p} N_0 v_0 = w_1 \text{ (say)}.
\]

Similarly if \( w_1 \) contains an instance of \( N_1 \) (say), not necessarily distinct from \( N_0, \) a rule of \( N_1 \) can be used to expand \( w_1 \) into \( w_2. \) This operation \( w_0 \Rightarrow w_{k+1} \) can be continued as long as \( w_k \) contains at least one nonterminal. From now on the mark \( > \) is used more generally to show that \( w_k \) has been expanded from \( w_0 \) in one or more steps, for any \( k > 0. \) \( w_0 \Rightarrow w_k \) is a 'derivation' or a 'derivation of \( w_0 \) from \( w_0. \)

Notice that \( w_{k+1} \) contains symbols of two distinct origins. It contains \( C_{i1} \ldots C_{i p} \) which appear because of the application of the \( N_0 \) rule, and also other symbols which were present in \( w_0. \) In the derivation of \( w_{k+1} \) the nonterminal \( N_{k+1} \) may be chosen from either group of symbols in \( w_{k+1}. \) However, it will be useful to discuss \( w_0 \Rightarrow w_1 \Rightarrow w_2 \Rightarrow \ldots \Rightarrow w_k \) in which for each \( k, N_{k+1} \) is chosen only out of the \( C_{i1} \ldots C_{i p} \) in \( w_{k+1} \) which arise from the expansion of \( N_0. \) Such a derivation \( w_0 \Rightarrow w_k \) is here called a 'chained derivation' and is denoted by \( w_0 \Rightarrow w_k. \) It follows that there exist \( u_0, v_0, \ldots, u_k, v_k \) such that

\[
w_0 = u_0 N_0 v_0 \Rightarrow u_0 u_1 N_1 v_1 v_0 = u_0 u_1 u_2 N_2 v_2 v_2 v_1 v_0 = u_0 u_1 u_2 u_3 N_3 v_3 v_2 v_1 v_0 = \ldots = u_0 u_1 u_2 u_3 \ldots u_k N_k v_k v_{k-1} \ldots v_1 v_0 = w_k.
\]

A derivation \( N = w_0 \Rightarrow w_k \) can be expressed as a 'generation tree', which is a tree whose nodes are (all the instances of) the symbols of \( w_0, w_1, \ldots, w_k. \) Branches leave a node \( N_k \) to arise at the symbols into which \( N_k \) is expanded by a rule application.

**Example:**
Given rules \( N_0 \rightarrow N_1 N_2, N_1 \rightarrow ab, N_2 \rightarrow cN_3, N_3 \rightarrow \alpha N_4, \) then \( N_0 \rightarrow abcN_3 \) has the generation tree shown in Fig. 1. The chained derivation \( N_0 \Rightarrow N_1 \cdots N_k \) has the generation tree in Fig. 2, which shows its characteristic linear sequence of rule applications.

The strings of the CF language determined by \( G \) are now defined as those (finite) terminal strings \( s \) such that \( S \Rightarrow s. \) \( G \) is 'admissible' if for each \( N \) not the same as \( S, \) there exist some \( u, v \) such that \( S \Rightarrow u N v \) and there exists \( s \) such that \( u N v \Rightarrow s. \) Only admissible grammars are considered below.

A 'recursive derivation' is a derivation of the form \( N \Rightarrow u N v \) for arbitrary words \( u, v. \) The generation tree of a recursive derivation is also called recursive.

A 'chained recursive derivation' is a derivation \( N \Rightarrow u N v \) which is both chained and recursive, corresponding to a generation tree which is a linear sequence of rule applications beginning and ending at \( N \)-nodes.

A 'cycle' \( C \) of \( G \) is defined by a sequence

\[
N_i(i_1, i_2) \ldots N_i(i_j, j_2) \ldots N_m(i_m, j_m)
\]

of alternating nonterminal instances \( N_r \) and number pairs \( (i_r, j_r) \) such that

1. For each \( r < m, \) \( C_{ir}, r = r \in L_i, \) is the \( i_r \)th rule of \( N_r, \)
2. \( C_{im} = N_1 \) in the \( i_m \)th rule of \( N_m, \)

Given such a sequence, any cyclic permutation such as

\[
N_2(i_2, j_2) \ldots N_i(i_j, j_2) \ldots N_m(i_m, j_m) N_1(i_1, j_1)
\]

identifies the same cycle.

Let a cycle \( C' \) be \( N'_1(i'_1, j'_1) \ldots N'_l(i'_l, j'_l) \) and let \( C \) and \( C' \) have a common nonterminal \( N_1 = N'_1. \) Then the sequence

\[
N_1(i_1, j_1) \ldots N_l(i_m, j_m) N'_1(i'_1, j'_1) \ldots N'_l(i'_l, j'_l)
\]

also defines a cycle which is said to be 'composed' from \( C \) and \( C'. \)

Now in \( R_i(N_i) \) let \( u_r \equiv C_{i_r1} C_{i_r2} \ldots C_{i_r j_r - 1} \) and let

\[
v_r \equiv C_{i_r j_r - 1} \ldots C_{i_{r+1} p}.
\]

Choose a cycle \( C \) of \( G \) and \( N_i \) in it. Then these choices identify a chained recursive derivation

\[
N_r \Rightarrow u_r u_{r+1} \ldots u_p N_{r+1} u_{r+2} \ldots v_r \ldots v_{p-1} v_p v_{p-2} \ldots v_r \ldots v_{p-1} v_p \ldots v_{p-1} v_p
\]

Given a cycle \( C \) all such chained recursive derivations are called 'the chained recursive derivations of \( C. \)'

**Lemma 1**
In all the chained recursive derivations \( N \Rightarrow u N v \) of a given cycle \( C, \) the strings \( u, v \) are constant, except for permutations of symbols.

Further notation is introduced as required.

**2. A useful graph**
It is helpful to introduce below a directed graph \( G_R \) associated with \( G. \) The simple graph theory and terminology used here is adapted from Berge and Ghouila-Houri (1969).

A finite graph consists of a finite number of points or 'nodes' joined together by a finite number of directed lines or 'arcs'. A 'path' is a sequence of arcs such that the end node of one arc is the start node of the next. A path may pass through a particular node more than once, and use a particular arc more than once.

A 'circuit' is a path in which any node of the path can be considered as both the start node and end node of the path. An 'elementary circuit' is a circuit in which no node occurs more than once. There are clearly only a finite number of elementary circuits.

If two circuits \( A, B \) have a common node \( p \) then another circuit \( C \) can be 'composed' out of the 'components' \( A, B \) by joining them at \( p. \) Thus no composed circuit can be elementary. By examining multiple instances of nodes on a circuit it is easy to see that:

**Lemma 2**
Every (finite) circuit of a graph is either elementary or can be composed out of a finite number of elementary circuits, perhaps repeated.

The graph \( G_R \) is now constructed. Its nodes are single instances of the nonterminals and terminals of \( G. \) Draw from each \( N \) an arc to every \( C_i \) in each rule \( R_i(N). \) This arc may be labelled \((N, i, j). \) Add a special end node \( E, \) and draw an arc \( t \) from every terminal \( t \) to \( E. \)

**Example**
For the rules \( S \rightarrow NS \) \( R_1(S) \)
\( S \rightarrow b \) \( R_2(S) \)
\( N \rightarrow aNC \) \( R_3(N) \)
\( N \rightarrow b \) \( R_3(N) \)

the graph \( G_R \) is shown in Fig. 3. \( G_R \) exhibits some of the properties of \( G, \) and it will be a useful tool in establishing by graph theory the conditions of solution and methods of solution of certain systems of equations and recurrence relations.

**Lemma 3**
Every cycle of \( G \) is composed of one or more of a finite basis of
have a set \( \{l(u)\} \) of such \( l \)-values, possible infinite, and if \( v \rightarrow w \) then \( \{l(v)\} \subseteq \{l(w)\} \), since every terminal expansion of \( w \) is a terminal expansion of \( v \). If \( l(u) \) has a (least) upper bound, call it \( l^+(u) \). If \( l(u) \) has a (greatest) lower bound, then call it \( l^-(u) \). The string \( u \) is termed \( l^- \)-bounded, \( l^+ \)-bounded, or just \( l \)-bounded if respectively \( l^- \) (\( l \)) exists, \( l^+ \) (\( l \)) exists or both of these exist. A particular case of \( l \)-boundedness is 'l-uniqueness', when \( l^+(u) = l^-(u) \), and \( l(u) \) is unique.

**Examples**

1. \( S \rightarrow N_1N_2, N_1 \rightarrow aN_1, N_3 \rightarrow a, N_3 \rightarrow C_1N_2, \) \( N_2 \rightarrow C \). A terminal string of \( S \) is \( a^nC^n \), for any integers \( m, n \geq 1 \). Thus \( l(S) = m - n \) and is completely unbounded.

2. \( S \rightarrow aS, S \rightarrow a \). The terminal strings are \( a^m \), \( m \geq 1 \). \( l(S) = m \), so that \( l^- (S) = 1, l^+(S) \) does not exist.

3. \( S \rightarrow CS, S \rightarrow C \). Here \( l^- (S) \) does not exist, \( l^+(S) = -1 \).

4. \( S \rightarrow \lambda, N_1 \rightarrow \lambda, N_1 \rightarrow \lambda, \) \( N_2 \rightarrow \lambda \). \( l(N_1) = 1 \), \( l(N_2) = 1 \). Thus \( l(S) = \{2, 3, 1, 1\} \), \( l^-(S) = -2, l^+(S) = 2 \). The last rule of \( N_2 \), though recursive, adds no further lengths to \( l(N_2) \).

5. \( S \rightarrow N_1N_2C, N_1 \rightarrow \lambda, C \rightarrow aC, N_3 \rightarrow \lambda \). Here \( l(S), l(N_1) \), and \( l(N_2) \) are all unique.

Since the above definitions concerning \( l(u) \) can apply to any \( N \) and thus to \( S \), the terminology can be applied naturally to grammars as well.

Also for a chained derivation \( N \rightarrow uv \), define the 'left-hand length' \( lh(N) \rightarrow uv \equiv l(u) \).

**4. Recurrence relations for the \( l(N) \)**

Consider a rule \( N \rightarrow C_1 \ldots C_p \) and let \( C_j \) be \( N_i \). A derivation of \( N \) which starts with the above rule is unrestricted as to which derivation of \( N_i \), \( C_j \) expands into. Thus all that can be said of \( l(C) \) is that \( l(C) \) is in \( \{l(N_i)\} \). For any such derivation

\[
l(N) = \sum_{j=1}^{p} l(C_j).
\]

Hence to find a \( l(N) \) value choose known values for each of the \( l(C) \) and add them. This is a kind of recurrence relation which may be written

\[
l(N) < \sum_{j=1}^{p} l(C_j).
\]

Moreover, by taking all the rules of \( G \) at once, the recurrence relations which arise determine all the \( l(N) \) values, for all \( N \), by applying a number of rules it is easy to see that if \( N \rightarrow u = w_1 \ldots w_p \ldots w_s \), then

\[
l(N) < \sum_{r=1}^{s} l(w_r).
\]

Out of this comes a trivial theorem which helps to determine the unboundedness (or non-uniqueness) of \( G \) — it is sufficient to find just one unbounded (or non-unique) nonterminal.

**Theorem 1**

1. \( G \) is \( l^- \)-bounded if and only if all \( N \) are \( l^- \)-bounded.

2. \( G \) is \( l^+ \)-bounded if and only if all \( N \) are \( l^+ \)-bounded.

3. \( G \) is \( l \)-bounded if and only if all \( N \) are \( l \)-bounded.

4. \( G \) is \( l \)-unique if and only if all \( N \) are \( l \)-unique.

**Proof**

**1. Necessity**

Since \( G \) is admissible, then for every \( N \) not the same as \( S \), there exist \( u, v, s \) such that \( S \rightarrow uNv \rightarrow s \), so that there exists at
least one \( I(S) \) value generated by \( I(S) = -I(u) + I(N) + I(v) \). Now \( u, N \) and \( v \) can be expanded independently into strings of terminals, and so their lengths cannot always compensate each other to keep \( I(S) \) \( l^- \)-bounded unless they are all \( l^- \)-bounded. Thus \( S \) and hence \( G \) is \( l^- \)-bounded only if every \( N \) is \( l^- \)-bounded.

**Sufficiency**

If all \( N \) are \( l^- \)-bounded then \( S \) is and so is \( G \).

2, 3 and 4: The proofs are analogous to 1.

**5. The ‘\( l^- \)-uniqueness’ of \( G \)**

\( l^- \)-uniqueness is a desirable property in the translator-inversion algorithm in particular and for clarity of understanding in general. When \( G \) is \( l^- \)unique the recurrence relations

\[
I(N) < - \sum_{j=1}^{n} I(C_j)
\]

become consistent equations in the \( I(N) \). It is interesting to consider the converse, i.e. whether the \( I(N) \) are unique if the equations are consistent. A set of linear equations does not in general have a unique solution (see, say, Griffiths, 1947). However the origin of these equations gives them a special form which does ensure uniqueness as the following theorem shows.

**Theorem 2**

Let \( \{ I(N_1), I(N_2), \ldots, I(N_h) \} \) be a solution of the equations \( \{ I(N_h) = \sum_{j=1}^{n} I(C_{ij}) \} \), for all possible \( h, i, j \) which arise from the production rules of \( G \). Then this solution is unique, so that \( G \) is \( l^- \)-unique.

**Proof**

The proof is by induction on the number \( r \) of rule-applications necessary to expand \( N \) into a string \( s_N \).

Consider a derivation \( N \Rightarrow s_N \) in which just one rule-application is used. This rule must be \( N \Rightarrow C_{i1} \cdots C_{ij} \cdots C_{ip} \) for some \( i \) in which each of the \( C_{ij} \) is a terminal. Because the length of the right-hand side is constant \( I(s_N) = I(N) \). Thus \( I(N) \) is unique for all \( N \Rightarrow s_N \) such that \( r = 1 \).

The induction step is as follows. Assume that for all \( r \) up to some \( r \geq 1 \), every derivation \( N \Rightarrow s_N \) which has \( r \) rule applications has the unique length \( I(N) \). Now consider \( N \Rightarrow s_N \), if any, with \( r + 1 \) rule applications, where the derivation starts with \( N \rightarrow C_{i1} \cdots C_{ij} \cdots C_{ip} \). Here some of the \( C_{ij} \) may be nonterminals. Let \( r_1, \ldots, r_p \) be the numbers of rule applications in the subtrees starting at \( C_{i1}, \ldots, C_{ij}, \ldots, C_{ip} \) respectively. Then

\[ r + 1 = 1 + \sum_{j=1}^{p} r_j \]

so that for each \( j \), \( r > r_j \). Hence the assumption of the induction step is applicable and the lengths \( I(C_{ij}) \) are unique. Thus the length of the whole right-hand side is unique and therefore must be \( I(N) \).

Thus by induction \( I(N) \) is unique however many rule applications are involved in a derivation \( N \Rightarrow s_N \).

To determine \( G \)'s \( l^- \)-uniqueness the steps are therefore:

1. Construct a tentative set of lengths \( \{ I(N_1), \ldots, I(N_h) \} \) by using the simplest rules of \( G \).
2. Substitute these in all the equations of \( G \). If all the equations are satisfied, then \( G \) is \( l^- \)-unique.

**6. \( l^- \)-boundedness conditions**

The theorem of this section (Theorem 3) proves necessary and sufficient conditions for the lower and upper \( l^- \)-boundedness of \( G \), and also gives a little more when these properties occur together. The proofs concerning upper and lower \( l^- \)-boundedness are analogous, and only \( l^- \)-boundedness is dealt with in detail. Lemma 4 which precedes the theorem is written in terms of the \( l^- \) proof only.

The proof is by induction on the ‘\( l^- \)-recursiveness’ of a derivation \( N \Rightarrow s_N \), which is a measure of its complexity defined as follows:

A general tree of the derivation \( N \Rightarrow s_N \) of a terminal string \( s_N \) is ‘\( q^- \)-recursive’ where \( q \) is the number of recursive subtrees it contains, including itself if it is recursive.

A general string \( u \) is said to be \( q^- \)-recursive if attention is restricted to the subset of derivations of \( u \) in which no symbol of \( u \) is more than \( q^- \)-recursive, but in which at least one symbol of \( u \) is a \( q^- \)-recursive.

Now define \( I_q(u), I_q^+(u), I_q^-(u) \) to be analogous to \( I(u), I^+(u), I^-(u) \) but where only \( q^- \)-recursive derivations of \( u \) are considered.

From the above definitions it follows that if \( u \) is \( q^- \)-recursive then the symbols of \( u \) can be rearranged arbitrarily without affecting its \( q^- \)-recursiveness or, of course, its length. From Lemma 1 it is therefore reasonable to refer to the length of a cycle \( C \) as \( I(C) \equiv I(u_1) + I(u_2) \), where there is a chained recursive derivation \( N \equiv > u_1 N u_2 \) of \( C \). Similarly \( I_q(C) \equiv I_q(u_1 u_2) \). Also needed later is the left-hand-length \( lhl(C) = I(u_1) \).

The proof of Theorem 3 relates all \( I(N) \) values to \( I_q(N) \), the lower bound of the finite set \( \{ I_q(N) \} \). Lemma 4 now provides the induction step.

**Lemma 4**

For any \( N, I_q^-(N) > I_q^-(N), q > 0 \), provided that for all \( r, 0 < r < q \),

\[
I_q^-(N) > I_q^-(N), \quad \text{and} \quad I(N) = I_q^-(N), \quad \text{for all cycles } C, I_q^-(C) = 0. \tag{C1}
\]

**Proof**

Consider a \( q^- \)-recursive derivation \( N \Rightarrow s_N \) of \( N \). The aim is to construct from this another derivation \( N \Rightarrow s_N \) which is at most \( (q - 1) \)-recursive and which has no greater in length. Then \( I_q(N \Rightarrow s_N) > I_q(N \Rightarrow s_N) > I_q^-(N) \) where \( r < q - 1 \), from (C1). By choosing \( N \Rightarrow s_N \) so that \( I_q^-(N) \) is minimal the relation becomes \( I_q^-(N) > I_q^-(N) \) as required.

\( N \Rightarrow s_N \) is constructed as follows. Let \( N' \) be the nonterminal in the base-node of one of the recursive subtrees of \( N \Rightarrow s_N \), so that \( N \Rightarrow u N' v \) for some strings \( u, v \). Let \( N'' \) be an embedded instance of the same nonterminal \( N' \), so that \( N' \Rightarrow u' N' v' \), say. Now form the tree for \( N \Rightarrow s_N \) by replacing the \( N' \) subtree by the \( N'' \) subtree. Because one recursive use of \( N'' \) has been removed, \( N \Rightarrow s_N \) cannot be more than \( (q - 1) \)-recursive, and similarly the string \( u' v' \) cannot have any symbol which is more than \( (q - 1) \)-recursive, so that

\[ I(u' v') > I_q^-(u' v'), \]

from (C1),

\[ = 0, \]

from (C2).

Hence \( I(N \Rightarrow s_N) = I(u N' v') = I(u N'' v') + I(u' v') \)

\[ = I(N \Rightarrow s_N), \quad \text{as required} . \]

**Theorem 3**

1. \( G \) is \( l^- \)-bounded and for every \( N l^-(N) = I_q^-(N) \) if and only if \( I_q(N) = 0 \) for each basic (i.e. elementary) cycle of \( G \).

2. \( G \) is \( l^- \)-bounded and for every \( N l^+(N) = I_q^+(N) \) if and only if \( I_q(N) < 0 \) for each basic cycle of \( G \).

3. \( G \) is \( l^- \)-bounded and for every \( N \{ I(N) \} = \{ I_q(N) \} \) if and only if \( I_q(N) = 0 \) for each basic cycle of \( G \).

**Proof**

**Necessity**

1. Suppose \( G \) is \( l^- \)-bounded, and \( l^-(N) = I_q(N), \forall N \), but...
there is an elementary cycle \( C \) for which \( l_0^-(C) > 0 \) is false. Then there exists a derivation \( N \overset{\alpha}{\Rightarrow} u_1N\alpha_1 \ldots \alpha_n \) of \( C \) such that \( I(u_1\alpha_1) < 0 \). Choose some derivation \( N \overset{\alpha}{\Rightarrow} s_N \), and let it have length \( q^N \). Then \( N \overset{\alpha}{\Rightarrow} u_1N\alpha_1 \ldots \alpha_n \) defines a recurrence relation:
\[
i_{i+1}^N = l(u_{i+1}) + i^N + l_0(u_i), \quad i > 0
\]
\[
= l(u_{i+1}) + i^N < l^N.
\]
Hence the integer sequence \( q^N, i^N, \ldots, i^N \ldots \) has no lower bound so that \( N \) is not \( I^- \)-bounded, contrary to hypothesis. Hence \( I_0^-(C) > 0 \) for all \( C \) of \( G \).

2. The proof is analogous to 1.

3. Apply 1 and 2 together.

** Sufficiency **

1. The proof is inductive using Lemma 4. It remains to prove (C2) and to show that (C1) holds for \( q = 1, r = 0 \). (C1) reduces to showing that \( I_0^-(N) \) exists, which is true because \( \{l_0(N)\} \) is finite. (C2) follows from the conditions of the theorem by Lemma 3.

2. The proof is analogous to 1.

3. Apply 1 and 2 together to show that \( G \) is \( l \)-bounded. The conditions of 1 and 2 also give \( I^-(u_1) + I^-(u_0) > 0 \) \( = I^+(u_1) + I^+(u_0) \). Hence \( I_0^-(u_1) = I^+(u_1) = l(u_1) = 0 \), and \( 0 = I^-(u_0) = I^+(u_0) = l(u_0) \) by the definition of upper and lower bounds.

Now Lemma 4 can be rephrased to show that if \( N \overset{\alpha}{\Rightarrow} s_N \) is \( q \)-recursive, \( q > 0 \), then one can find a \( q'-\)recursive \( s_N \) such that \( q' < q \) and \( I(s_N) = I(s_N) \). By applying this as many times as is necessary it follows that for any \( q \)-recursive \( s_N \) there is a \( 0 \)-recursive \( s_N \) with the same length. Hence
\[
\{l(N)\} \subseteq \{l_0(N)\}
\]
by definition of \( l_0(N) \). Therefore \( \{l(N)\} = \{l_0(N)\} \).

7. ** Determination of the \( I^- \) boundedness of \( G \)**

The following computable steps can therefore be used to determine whether \( G \) is \( I^- \)-bounded. (It is only worth doing this if an application of Theorem 2 has shown \( G \) is not \( I^- \)-unique.)

1. Inspect the rules of \( G \) to see if any \( N \) is obviously unbounded (Theorem 1).

2. If no unbounded \( N \) is apparent draw the graph \( G_R \) and find its elementary circuits. These identify a set of basic cycles \( C \) of \( G \). (An algorithm for finding the elementary circuits of \( G \) is given in Weinblatt (1972).)

3. For each \( C \) take one of its chained recursive derivations \( N \overset{\alpha}{\Rightarrow} u_1N\alpha_1 \ldots \alpha_n \) and for each symbol \( x_i \), in \( u_1 \alpha_1 \ldots \alpha_n \), find \( l_0^-(x_i) \) and \( I_0^-(x_i) \). Hence determine \( I_0^+(x_i) \) and \( l_0^+(x_i) \) and apply Theorem 3. If \( G \) is completely bounded then the \( I(N) \) are the \( l_0(N) \), by Theorem 3.

8. ** The gross number of items yielded by a nonterminal **

The preceding sections have dealt with the effects of depositing on the stack complete terminal strings derived from non-terminals. In this section the effect on the stack is considered at all stages during the deposition of a nonterminal's string. As an example let \( E \) have the extra concatenate function mentioned in the introduction and consider the rules \( S \to CSa \), \( S \to b \), which yield the strings \( Cba, CCBaa, \ldots, Cba^n \), for all integers \( n \). Then although all of these strings have length \((n + 1) \times n \), they successively combine more and more of the items already on the stack before depositing more. In contrast, the rules \( S \to aSC, S \to b \) yield strings \( a^n b^n C^n \), all having length \( 1 \), but which successively add more and more items to the stack before matching concatenations take place.

These effects might be of real concern to the implementor of a stack handling grammars of this kind, because words of the language might overemply or overfill the stack. This section deals with conditions for grammars to be 'well behaved' in this way. However, a more severe effect than overemplying is discussed under 'Disturbance measurements' below.

It is useful to define \( m^-(u) \) to be the 'gloss' minimum length (in the generalised sense of the last section) which any terminal string derived from \( u \) can take on the stack at any time during or after its deposition. Similarly define \( m^+(u) \) to be the 'gloss' maximum length of any terminal string of \( u \). Since \( I^-(u) \) and \( I^+(u) \) are the minimum and maximum lengths just after the deposition of \( u \), \( m^-(u) \) exists only if \( I^-(u) \) exists, and \( m^+(u) \) exists only if \( I^+(u) \) exists. Also \( m^-(u) = < I^-(u) \) and \( m^+(u) >= I^+(u) \). In the remainder of this section the relevant \( l \)-bouds are always assumed to exist.

In order to evaluate \( m^-(N) \) for all \( N \) in \( G \), consider any rule \( N \overset{\alpha}{\Rightarrow} C_1 \ldots C_j \ldots C_p = > s_N = s(C_1)s(C_2) \ldots s(C_p) \) and consider the process of depositing \( s_N \) on the stack. It may be that \( s(C_1) \) causes the length of \( s_N \) to be a minimum, so that certainly \( m^-(N) = < l^-(C_1) \). However it may be \( s(C_2) \) which causes the minimal length of \( s_N \). In this case the whole of \( s(C_1) \) is deposited before \( s(C_2) \) is begun and so
\[
m^-(N) = < l^-(C_1) + m^-(C_2) \]

Similarly \( m^-(N) = < l^-(C_1) + l^-(C_2) + m^-(C_3) \)

\[
m^-(N) = < \sum_{j=1}^{p-1} l^-(C_j) + m^-(C_p) \]

and
\[
m^-(N) = < \sum_{j=1}^{p-1} l^-(C_j) + l^-(C_p) = \sum_{j=1}^{p} l^-(C_j) \]

However this last inequality can be disregarded since \( m^-(C_p) = < l^-(C_p) \). Putting these inequalities together
\[
m^-(N) = < \min_{1 \leq j \leq p} \left[ \sum_{k=0}^{j-1} l^-(C_k) + m^-(C_p) \right] \]

where for convenience \( l^-(C_j) \equiv 0 \). One of these expressions arises for each production rule of \( N \), so that \( m^-(N) \) is the minimum of all these expressions:
\[
m^-(N) = \min_{R \in R(G), 1 \leq j \leq p} \left[ \sum_{k=0}^{j-1} l^-(C_k) + m^-(C_p) \right] \]

The author has an algebraic algorithm for the solution of this set of equations, one for each \( N \), together with proof of necessity and sufficient conditions for solution. However this approach does not give any understanding of what is happening. Given below is a more illuminating method, based on mapping the problem on to the graph \( G_R \).

Assign to each arc \( (N, i, j) \) of \( G_R \) the arc length
\[
\sum_{k=0}^{j-1} l^-(C_k)
\]
which is the minimum lefthand length of the trivial chained derivation \( N \overset{\alpha}{\Rightarrow} u_1x_{ij} \), where \( u_j = C_1C_2 \ldots C_{i-j+1} \), \( x = C_i \), \( v_i = C_{i+1} \ldots C_p \). Then (as in the proof of Lemma 3) for any chained derivation \( N \overset{\alpha}{\Rightarrow} u_1x_{ij} \), the minimum lefthand length \( = l^-(u_1) \) is the sum of the arc lengths which make up the corresponding path in \( G_R \). Also assign the arc length \( m^-(i) = l^-(i) \) to each arc \( i \). Now consider again the cause of \( N \) having a gross minimum length \( m^-(N) \). This minimum is attained by the deposition of a particular terminal \( i \) of a particular terminal expansion of \( N \), i.e. there exist \( u_1, v_i \) such that \( N \overset{\alpha}{\Rightarrow} u_1v_i \). Hence by the argument used before
\[
m^-(N) = l^-(u_1) + m^-(i) \]

which is the length of an arc from \( N \) to \( E \) on \( G_R \). Hence \( m^-(N) \)
is the minimum path length from $N$ to $E$. Berge and Ghoulia-Houri (1965) give the well known result that such a minimum path exists for every $N$ if and only if there is no circuit with a negative arc length. This is equivalent to the necessary and sufficient condition that \( l h^{-1}(C) = l h^{-1}(C^G) = 0 \) for every cycle $C$ of $G$. This justifies the following theorem:

**Theorem 4**

1. $m^{-}(N)$ exists for every $N$ of $G$ if and only if $l h^{-1}(C) > 0$ and $l h^{-1}(C^G) > 0$ for all basic cycles $C$ of $G$.
2. $m^+(N)$ exists for every $N$ of $G$ if and only if $l h^{-1}(C) < 0$ and $l h^{-1}(C^G) < 0$ for all basic cycles $C$ of $G$.

Evaluation of the $m^-(N)$ and $m^+(N)$ is straightforward since the arc lengths are functions of the $l^-$ and $l^+$ respectively which are all known beforehand. For some possible methods see Berge and Ghoulia-Houri, (pp. 180-182) and Iri (1969). Furthermore, the minimum path length problem has a unique solution (although more than one path may attain that length). Hence the original system of equations has a unique solution, because the grammar, the graph and the equations are in (1,1)-correspondence. It follows that if a tentative solution, say \( \{ M^{-}(N_1) \ldots M^{-}(N_N) \ldots \} \) does satisfy the equations, then the $m^{-}(N)$ exist and $M^{+}(N_k) = m^{-}(N_k)$ for each $k$. However the analogue of Theorem 2, when $m^{-}(N_k) = m^{-}(N_k)$ for each $k$, is not interesting because this equality is true only for a trivial subset of grammars.

**Theorem 5**

1. For every $N$, $m^{-}(N) = < 1.$
2. For every $N$, $m^{+}(N) > = -1.$

**Proof**

1. For any rule $N -> C_1 \ldots C_p$, $m^{-}(N) = < m^{-}(C_i).$ Thus as the lefthand branch of a generation tree of $N$ is followed the $m^-$ value for every subtree encountered cannot decrease. Finally the last nonterminal $N_0$ (say) is encountered where $N_0 -> C_0C_p \ldots C_2$ (say).

Then $m^{-}(N) = < m^{-}(N_0) = < m^{-}(t_0) = < \max_{N_i} m^{-}(t) = \max_{t \leq N_i} |l(t)| = 1.$

2. The argument is analogous to 1.

9. Disturbance measurements

The $P_R$ permutation symbols are now discussed. Consider a single rule grammar $S -> P_1ab$, where $P_1$ is an interchange. Here $l(S) = 2$, $m^-(S) = 1$, but before any symbols at all are deposited on the stack the top two cells are interchanged, thus interfering with material not deposited by this grammar. Thus the necessary condition for no underflow is that no previously deposited material should be disturbed. This is developed as follows.

Let $d(u)$ be the number of previously deposited cells disturbed at any stage during the loading of $s$, where $u = s$. Then for any derivation $N -> s_N$, the maximum disturbance

$$d^+(N = > s_N) = \max_{u \leq N} \left\{ \sum_{j=1}^{N} l^{-1}(C_i) \right\}$$

by an argument similar to that used in obtaining $m^{-}(N)$. So the maximum disturbance any derivation of $N$ could make is

$$d^+(N) = \max_{u \leq N} \left\{ \sum_{j=1}^{N} l^{-1}(C_i) \right\}$$

An analogous formula applies for $d^{-}(N)$, which is the minimum disturbance which can take place at any stage during the deposition of any terminal string derived from $N$. $d^{-}(N)$ can be negative.

Now let $r_M$ be the maximum number of cells rearranged by any of the $P_R$ operations of $G$. The disturbances made by individual symbols are zero for normal symbols, 1 or 2 for $C$ and $r_R$ for $P_R$, where $2 = r_R = r_M$.

**Theorem 6**

1. The $d^+(N)$ all exist if and only if the $m^{-}(N)$ exist. Moreover, for all $N$, $0 = < d^+(N) + m^{-}(N) = < r_M$ and $0 = < d^+(N)$.
2. The $d^-(N)$ all exist if and only if the $m^+(N)$ exist. Moreover, for all $N$, $0 = < d^-(N) + m^+(N) = < r_M$ and $d^-(N) = < r_M$.

**Proof**

1. Using the graphical method to solve the $d^+(N)$ equations, solutions exist if and only if the circuits of $G_N$ have arc lengths not greater than zero, since maxima are being sought. But each arc length involved in a circuit has the same magnitude but opposite sign compared with the arc lengths in the $m^{-}(N)$ network, where the necessary and sufficient condition was 'circuit arc length not less than zero'. This proves the equivalence of the existence conditions.

Call the $d^+(N)$ network $D$. Now consider the network $M$ which differs from $D$ in that the arcs $t$ have lengths $-m^{-}(t)$. Then every arc length has the same magnitude but is opposite in sign compared with those in the $m^{-}(N)$ network.

The values of $d(t) + l(t) = d^+(t) + m^{-}(t)$ are as follows:

$$t \quad d \quad l \quad d + l$$

<table>
<thead>
<tr>
<th>$P_R$</th>
<th>$r_R$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Thus in every case $0 = < d^+(t) + m^{-}(t) = < r_M$. Now let $t_1$ be the terminal at which $d^+(N)$ is attained, where $t_2$ is reached from $N_1$. Let $t_2$ be the terminal at which $m^{-}(N)$ is attained, where $t_2$ is reached from $N_2$. The proof is analogous to the one just given, except that $N_1$ and $N_2$ need not be distinct.) Using the maximal property of $t_1$ in network $D$,

$$d^+(N) = \max_{t_1 \leq N} \left\{ d^+(N) + m^{-}(N) \right\}$$

Also using the maximal property of $t_2$ in network $M$,

$$d^{-}(N) = \max_{t_2 \leq N} \left\{ d^+(N) - m^{-}(N) \right\}$$

The proof of the relation $d^+(N) = > 0$ is on the lines of Theorem 5.

2. The proof is analogous to 1 above.

As has been seen, evaluation of the $d^+(N)$ is analogous to the evaluation of the $m^{-}(N)$, and the testing of a tentative solution $[D(N_1) \ldots D(N_n)]$ by substitution is also valid.

10. Application to Metcalfe-Reeves translators

In Goodwin (1975) it was shown that sufficient conditions for a certain translator-inversion algorithm to work were that for each $N$ the $l(N)$ and $d(N)$ values were unique, that $l(N) = 1$, and that $d(N) = d^+(N) = 0$. Proofs have been given in Theorem 2 and following Theorem 4 that these conditions can be verified by substitution of the desirable values in the $l$ and $d$ equations.

11. Application to compile time data stack checking

Only a sketch of the method is given. A high level language program can be regarded as defining the grammar of a generator.
Each assignment statement would correspond to one rule of the grammar and would there be expressed in Reverse Polish form, ending with a store operator which is another special case of C. Again function calls could be included. The treatment of the $l$, $m$ and $d$ quantities above is sufficient to allow functions which take from the stack an arbitrary fixed number of parameters and place on it any fixed number of results. So-called variadic functions in which the number of results or parameters varies at run time could not be allowed.

Loops as defined by backward GOTO statements or DO-type statements are allowed so long as their stack length is zero. This is always true in FORTRAN since the elementary stack altering operation is the assignment statement whose length is zero. Forward GOTO statements, if part of a condition, lead to the function in which they occur having more than one rule in the grammar.

Let the finite allowable stack length be $L$. Then analysis of this derived grammar at compile time could answer the overflow questions according to the table in Figs. 4 and 5.

The 'Certain or possible' cases in Fig. 4 need explanation. The maximum gross length $m^+(S)$ may be attained during the deposition of all strings $s_v$ of the language, in which case the corresponding program is bound to fail; on the other hand strings may exist whose individual gross length is always far short of $m^+(S)$, i.e. depending on its data the execution of the program may well not demand the use of the whole physical stack. These two different types of grammars can be distinguished by an algorithm when the number of relevant net and gross lengths which strings $s_v$ can take is finite. Conditions are given when this is true, but for brevity here the proof is deferred to a later article. When these conditions do not hold, it is not known whether an algorithm exists, although the author conjectures that it does. A similar discussion applies to Fig. 5.

These overflow and underflow results could be used simply to reject or accept the program at compile time. Alternatively they might be used to set the value of $L$, or as an automatic method of determining when to insert coding to check for stack overflow or underflow. Of course, these basic ideas are well known, and originality is only claimed for the systematic treatment above.

Acknowledgement

The author is indebted to Professor C. M. Reeves for reading a draft of this article and making valuable constructive criticisms.

References


