

# Optimal reduction in replacement systems

John Staples

High level programming languages can generally be compiled in many different ways. Thus it is natural to ask if there is a best way, for example in the sense of achieving complete execution of the program in a minimum number of steps; such a complete, minimal execution procedure will be called optimal.

In recent years this question has been studied and answered for several simple models of programming languages, and a technique for proving procedures to be optimal has gradually emerged. Even for model languages the proofs of optimality become intricate; thus it is natural to emphasize the simplicity of the underlying technique by generalising it to an abstract system. That is the purpose of this paper. The general method to be given applies to prove all theorems (on optimal executions for model languages) which are known to the author. None of the previously known proofs has explicitly used the method, but in no case is it particularly difficult to modify the known proof so as to conform with the method.

## 1. Introduction

1.1. The abstract system with which this paper is concerned is Rosen's notion of general replacement system [5], which has become standard for purposes such as ours; see for example [2, 6]. A general replacement system is nothing but a set  $B$  together with a binary relation  $\rightarrow$  on  $B$ ; however the name general replacement system indicates a concern with a

---

Received 10 December 1976.

particular class of problems, for example as follows.

1.2. Given a general replacement system  $(B, \rightarrow)$ , we write  $a \rightarrow b$  if  $(a, b) \in \rightarrow$ . The problems we consider are motivated by the intuitive idea that each element represents a program at some state of execution; if  $a \rightarrow b$  then a possible next step in the execution of  $a$  is to transform  $a$  to  $b$ . Elements of  $B$  may be called terms.

1.3. We write  $\rightarrow^*$  for the reflexive, transitive closure of  $\rightarrow$ . If there is no  $b \in B$  such that  $a \rightarrow b$ , then we say that  $a$  is in normal form. If  $a \rightarrow^* b$  and  $b$  is in normal form then we say that  $b$  is a normal form of  $a$ . If each  $a \in B$  has at most one normal form then we say that the system has the normal form property. A standard way of proving the normal form property, for systems which have it, is to deduce it as a corollary of the Church-Rosser property, which is defined as follows.

A system  $(B, \rightarrow)$  is said to have the Church-Rosser property if for all  $a, b, c \in B$  such that  $a \rightarrow^* b, a \rightarrow^* c$ , there is  $d \in B$  such that  $b \rightarrow^* d, c \rightarrow^* d$ .

The Church-Rosser property is named after the authors of the original proof [1] of a particular instance of the property. The property in the generality considered here was first studied by Newman [3].

Intuitively a term in normal form represents a completely executed program. Since we generally require that programs should execute to the same conclusion no matter which of the alternative compilations is chosen, we are generally interested in systems which have the normal form property.

1.4. If  $a \rightarrow b$  then we say that  $a$  contracts to  $b$ . If

$$a = a_1, a_1 \rightarrow a_2, \dots, a_{n-1} \rightarrow a_n, a_n = b,$$

then we say that the displayed sequence of contractions is a reduction of  $a$  to  $b$ , of length  $n$ . We may loosely denote, in context, a particular reduction of  $a$  to  $b$  by  $a \rightarrow^* b$ . This paper gives a method for proving for appropriate systems that certain reductions of terms with normal form to normal form are optimal in the sense that they are both

- (i) complete: that is, do reduce the terms in question to normal form; and

- (ii) minimal: that is, all reductions to normal form are of minimal length.

Our proof does not assume the normal form property. On the contrary, the normal form property is established by our hypotheses, though the Church-Rosser property is not.

## 2. The optimality result

2.1. We write  $B_N$  for the set of terms of  $B$  which have a normal form.

2.2. We consider a subset  $\rightarrow_o$  of  $\rightarrow$  consisting entirely of contractions of terms of  $B_N$ ; but we do not assume that all such contractions are included, or that  $b \in B_N$  when  $a \rightarrow_o b$  and  $a \in B_N$ . We call elements of  $\rightarrow_o$   $o$ -contractions, and we call reductions which comprise solely  $o$ -contractions,  $o$ -reductions. Our object is to give conditions on  $\rightarrow_o$  which ensure that repeated  $o$ -contraction of a term with normal form is always possible until a normal form is reached, that a normal form is always reached in that way, and that the  $o$ -reduction so obtained is a reduction of the term to normal form of minimal length.

If it should occur that  $o$ -contractions commute, in the sense that: whenever  $a \in B_N$ ,  $b, c \in B$ ,  $b \neq c$ , and  $a \rightarrow_o b$ ,  $a \rightarrow_o c$ , there is  $d \in B$  such that  $b \rightarrow_o d$ ,  $c \rightarrow_o d$ ,

then it is straightforward to show (see 3.1) that all  $o$ -reductions of  $a$  to normal form have the same length and reduce to the same normal form. For systems with the normal form property that is evidently a necessary condition for  $o$ -reductions to be optimal, but it is not usually sufficient. We shall see however that it is one of four conditions which taken together are sufficient (2.4).

2.3. In nontrivial programs and systems it is often possible to carry out computational steps which are completely irrelevant to the execution, and which therefore must be avoided in any optimal reduction. Consider a trivial example; at a branching point such as,

if  $2 = 0$  then  $y = 2 + 3$  else  $y = 4 + 5$ ,

there is no need to evaluate the sum  $2 + 3$ .

We shall assume in the hypotheses of our abstract result that we are able to describe the subset  $\rightarrow_{\delta}$  of  $\rightarrow$  comprising worthless contractions of terms of  $B_N$ . Such a subset will be called the set of dummy contractions, in accordance with previous practice in specific systems (see [7, §]). A sequence of dummy contractions is naturally called a dummy reduction and loosely denoted  $\rightarrow_{\delta^*}$ .

2.4. It is convenient to have a slightly different notation also for sets of contractions and sets of reductions, as follows. We write  $\gamma, \sigma, \delta$  for arbitrary contractions,  $\sigma$ -contractions, dummy contractions respectively. If  $r$  denotes a set of contractions or reductions then  $r^*$  denotes the set of reductions obtainable by iterating the contractions or reductions of  $r$  a finite number of times. If  $r, s$  denote sets of contractions or reductions then  $rs$  denotes the set of reductions obtainable by first applying to a term a contraction or reduction from  $r$  and then another from  $s$ . Also we denote by  $r - s$  a contraction or reduction which is in  $r$  but not in  $s$ , by  $r_{\leq}$  a reduction comprising either no contractions or else just one contraction or reduction from  $r$ , and by  $\bar{r}$  the set of contractions not in  $r$ .

Thus, for example,  $\delta^*\gamma\delta^*$  denotes a dummy reduction (perhaps comprising zero dummy contractions) followed by an arbitrary contraction followed by a dummy reduction.

We now state:

2.5 THE MAIN RESULT. *If  $(B, \rightarrow)$  is a system such that  $\rightarrow$  has subsets  $\rightarrow_{\sigma}, \rightarrow_{\delta}$  which satisfy the following conditions then  $B$  has the normal form property, every  $b \in B_N$  has an  $\sigma$ -reduction to normal form, and every  $\sigma$ -reduction of  $b$  to normal form is a reduction of  $b$  to normal form of minimal length. The conditions are:*

2.5.1. *if  $a \in B_N$ ,  $b \in B$ , and  $a \rightarrow b$  then there is  $c \in B$  such that  $a \rightarrow_{\sigma} c$ ;*

2.5.2. if  $a \in B_N$ ,  $b, c \in B$ ,  $b \neq c$ , and  $a \rightarrow_o b$ ,  $a \rightarrow_o c$ ,  
 then there is  $d \in B$  such that  $b \rightarrow_o d$ ,  $c \rightarrow_o d$ ;

2.5.3. (i) if  $a \in B_N$ ,  $b, c \in B$ , and  $a \rightarrow_o b$ ,  $a \rightarrow_\delta c$ ,  
 then there is  $d \in B$  such that  $c \rightarrow_o d$ ,  $b \rightarrow_{\delta^*} d$ ,

(ii) if  $a \in B_N$ ,  $b, d \in B$ ,  $a \rightarrow_\delta b$ , and  $b \rightarrow_o d$   
 then there is  $c \in B$  such that  $a \rightarrow_o c$  and  
 $c \rightarrow_{\delta^*} d$ ;

2.5.4. if  $a \in B_N$ ,  $b, c \in B$ , and  $a \rightarrow b$  is a contraction of  $\bar{\delta} - o$ , and if  $a \rightarrow_o c$ , then there is  $d \in B$  such that  
 $b \rightarrow^* d$  by a reduction of  $(o^*\delta^*)^*o(o^*\delta^*)^*$  and such that  
 $c \rightarrow^* d$  by a reduction of  $\delta^*\gamma_{\leq 1}\delta^*$ .

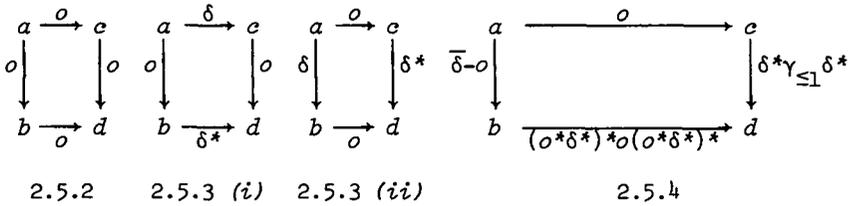


FIGURE 1

REMARK. The conditions of the above result are illustrated in Figure 1. In 2.5.4 the condition on the reduction of  $b$  to  $d$  is simply that it comprise  $o$ -contractions and dummy contractions, but including at least one  $o$ -contraction.

The proof of the result is given in the next section.

### 3. Proof of optimality

Throughout this section we consider a system  $(B, \rightarrow)$  which satisfies the hypotheses of 2.5; but in view of the discussion of 2.2 it should be noticed that the proof of 3.1 depends only on the hypothesis 2.5.2. We begin the section by stating and proving some needed lemmas, and conclude with the proof of 2.5.

3.1. All  $o$ -reductions of a term  $a \in B_{\mathbb{N}}$  to normal form have the same length and reduce  $a$  to the same normal form.

In order to prove 3.1 we first prove:

3.2. If  $a \in B_{\mathbb{N}}$  has an  $o$ -reduction  $R$  to normal form  $b$  which has length  $n$  and first contraction  $a \rightarrow_o a_1$ , and if  $a \rightarrow_o a_2$  where  $a_2 \neq a_1$ , and if  $a_1 \rightarrow_o a_3$ ,  $a_2 \rightarrow_o a_3$ , then  $a_2$  has an  $o$ -reduction to  $b$  of length  $n - 1$ .

Proof. By induction on  $n$ . The case  $n = 2$  is easily checked from 2.5.2 as follows. The second contraction of  $R$  must be  $a_1 \rightarrow_o a_3$ ; for if it were  $a_1 \rightarrow_o b$  where  $b \neq a_3$  then from 2.5.2 there would be  $a_4$  such that  $a_3 \rightarrow_o a_4$ ,  $b \rightarrow_o a_4$ , contradiction since  $b$  is in normal form.

Suppose then that  $n > 2$ . If  $a_1 \rightarrow_o a_3$  is the second contraction  $a_1 \rightarrow_o a_4$  of  $R$  then the assertion is trivially true. If not then 2.5.2 applies to  $a_1$  to show that there is  $a_5$  such that  $a_3 \rightarrow_o a_5$ ,  $a_4 \rightarrow_o a_5$ . Then the inductive hypothesis applies to  $a_1$  to show that  $a_5$  has an  $o$ -reduction to normal form  $b$  of length  $n - 3$ . Hence  $a_3$  has an  $o$ -reduction to  $b$  of length  $n - 2$  as required.

3.3 Proof of 3.1. Suppose that  $R, S$  are  $o$ -reductions of lengths  $m, n$  respectively to normal forms  $p$  and  $q$ . We prove by induction on  $k = \max\{m, n\}$  that  $m = n$  and  $p = q$ .

The case  $k = 0$  is trivially true. Also the case  $k = 1$  is easily seen to be true, for  $R, S$  comprise the contractions  $a \rightarrow_o p$ ,  $q \rightarrow_o q$  respectively and  $p = q$  since the assumption  $p \neq q$  leads from 2.5.2 to the conclusion that  $p, q$  are not in normal form, contradiction.

Hence we assume that  $k \geq 2$ . Denote the first contractions of  $R, S$  by  $a \rightarrow_o a_1$ ,  $a \rightarrow_o a_2$  respectively. If  $a_1 = a_2$  then the inductive hypothesis applies to  $a_1$  to give the result immediately. If  $a_1 \neq a_2$  then from 2.5.2 there is  $a_3$  such that  $a_1 \rightarrow_o a_3$ ,  $a_2 \rightarrow_o a_3$ . Then 3.2 applies to both  $a_1, a_2$  to show that  $a_3$  has  $o$ -reductions of lengths

$m - 2, n - 2$  to normal forms  $p, q$  respectively. Thus the inductive hypothesis applies to  $a_3$  to show that  $p = q$  and  $m - 2 = n - 2$ .

3.4. If  $a \in B_N$  has an  $o$ -reduction  $R$  to normal form  $b$  of length  $n$ , and if  $a \rightarrow_{\delta^*} a_1$  then  $a_1$  has an  $o$ -reduction of length  $n$  to  $b$ .

Proof. By induction on  $n$ . Then, for a proof by induction on the number of contractions comprising  $a \rightarrow_{\delta^*} a_1$  it is evidently enough to suppose that  $a \rightarrow_{\delta} a_1$ . Then the case  $n = 0$  is vacuously true and the case  $n = 1$  is evident from 2.5.3 (i). Hence we suppose that  $n \geq 2$ .

Say that the first contraction of  $R$  is  $a \rightarrow_o a_2$ . Then from 2.5.3 (i) there is  $a_3$  such that  $a_2 \rightarrow_{\delta^*} a_3, a_1 \rightarrow_o a_3$ . The inductive hypothesis applies to  $a_2$  to show that  $a_3$  has an  $o$ -reduction to  $b$  of length  $n - 1$ , so as  $a_1 \rightarrow_o a_3$ ,  $a_1$  has an  $o$ -reduction to  $b$  of length  $n$ .

3.5. If  $a \in B_N$  has an  $o$ -reduction  $R$  to normal form  $b$  of length  $n$  and if  $a \rightarrow^* c$  by a reduction  $S$  of type  $(o^*\delta^*)^*$  which includes just  $m$   $o$ -contractions, then  $c$  has an  $o$ -reduction to  $b$  of length  $n - m$ .

Proof. For a proof by induction on the number of contractions in  $S$  it is evidently enough to treat the two cases when  $S$  is a single contraction  $a \rightarrow_o a_1$ , or  $a \rightarrow_{\delta} a_1$ . These cases are given by 3.2 and 3.4 respectively.

3.6. If  $a \rightarrow_{\delta^*} a_1$  and if  $a_1$  has an  $o$ -reduction to normal form  $b$  of length  $n$  then  $a$  has an  $o$ -reduction to normal form  $b$  of length  $n$ .

Proof. By induction on  $n$ . Then for a proof by induction on the number of contractions comprising  $a \rightarrow_{\delta^*} a_1$  it is enough to suppose that  $a \rightarrow_{\delta} a_1$ . The case  $n = 0$  is trivial so we suppose that  $n > 0$ . Say that the first contraction of  $R$  is  $a_1 \rightarrow_o a_2$ . Then from 2.5.3 (ii) there is  $a_3$  such that  $a \rightarrow_o a_3$  and  $a_3 \rightarrow_{\delta^*} a_2$ . Then by the inductive hypothesis  $a_3$  has an  $o$ -reduction to  $b$  of length  $n - 1$ , so that  $a$  has an  $o$ -reduction to  $b$  of length  $n$  as required.

3.7. Proof of 2.5. We prove by induction on  $m$  that if  $a \in B$  has a reduction  $R$  to normal form  $b$  of length  $m$  then  $a$  has an  $o$ -reduction to  $b$  of length at most  $m$ . The result then follows from 3.1.

The case  $m = 0$  is trivially true. The case  $m = 1$  is easy, for if  $a \rightarrow b$  then  $a \rightarrow_o b$ , since  $a \rightarrow_o c$  for some  $c$  by 2.5.1; thus  $a \rightarrow_\delta b$  implies (2.5.3 (i)) that  $b$  is not in normal form, contradiction, and  $a \rightarrow_{\delta-o} b$  implies (2.5.4) that  $b$  is not in normal form, contradiction. Since  $a \rightarrow_o b, a \rightarrow_o c$ , then  $b = c$ , for  $b \neq c$  implies (2.5.2) that  $b$  is not in normal form, contradiction.

Hence we assume that  $m \geq 2$ . Say that the first contraction of  $R$  is  $a \rightarrow a_1$ . We assume, as we may by inductive hypothesis, that the remainder of  $R$  which reduces  $a_1$  to  $b$  is an  $o$ -reduction. We then assume that  $a \rightarrow a_1$  is not an  $o$ -contraction, for if it is then the previous assumption shows that  $R$  itself is an  $o$ -reduction of  $a$  to  $b$  of length  $m$ . We also assume that  $a \rightarrow a_1$  is not a dummy contraction, for otherwise the result follows from 3.6. Now there is (2.5.1) an  $o$ -contraction  $a \rightarrow_o a_2$  of  $a$ , so (2.5.4) there is  $a_3 \rightarrow B$  such that  $a_2 \rightarrow^* a_3$  by a reduction of  $\delta^* \gamma \delta^*$ , and  $a_1 \rightarrow^* a_3$  by a reduction of  $(o^* \delta^*)^* o (o^* \delta^*)^*$ . As  $a_1$  has an  $o$ -reduction to  $b$  of length  $m - 1$ , and as the reduction  $a_1 \rightarrow^* a_3$  is an  $(o^* \delta^*)^*$  which includes at least one  $o$ -contraction, then (3.5)  $a_3$  has an  $o$ -reduction to  $b$  of length at most  $m - 2$ .

Now say that the reduction  $a_2 \rightarrow^* a_3$  is

$$a_2 \rightarrow_{\delta^*} a_4 \rightarrow a_5 \rightarrow_{\delta^*} a_3 .$$

From 3.6,  $a_5$  has an  $o$ -reduction to  $b$  of length at most  $m - 2$ , so by inductive hypothesis  $a_4$  has an  $o$ -reduction to  $b$  of length at most  $m - 1$ , so by 3.6 again  $a_2$  has an  $o$ -reduction to  $b$  of length at most  $m - 1$ . Since  $a \rightarrow_o a_2$  we conclude that  $a$  has an  $o$ -reduction to  $b$  of

length at most  $m$ , as required.

#### 4. Discussion

4.1. If  $\sigma$  can be chosen to be  $\bar{\delta}$  then evidently Condition 2.5.4 is vacuously true.

4.2. The simplified hypotheses indicated in 4.1 are enough for the very simple, and suitably chosen, system of [4], but for the more complex systems of [7, 8] the hypotheses as stated in 2.5 are needed.

#### References

- [1] Alonzo Church and J.B. Rosser, "Some properties of conversion", *Trans. Amer. Math. Soc.* 39 (1936), 472-482.
- [2] Robert M. Keller, "A fundamental theorem of asynchronous parallel computation", *Parallel processing*, 102-112 (Proceedings of the Sagamore Computer Conference, 1974. Lecture Notes in Computer Science, 24. Springer-Verlag, Berlin, Heidelberg, New York, 1975).
- [3] M.H.A. Newman, "On theories with a combinatorial definition of 'equivalence'", *Ann. of Math.* (2) 43 (1942), 223-243.
- [4] G. Pacini, "An optimal fix-point computation rule for a simple recursive language" (Nota Interna B73-10, Pisa, 1973).
- [5] Barry K. Rosen, "Tree-manipulating systems and Church-Rosser theorems", *J. Assoc. Comput. Mach.* 20 (1973), 160-187.
- [6] John Staples, "Church-Rosser theorems for replacement systems", *Algebra and logic*, 291-306 (Lecture Notes in Mathematics, 450. Springer-Verlag, Berlin, Heidelberg, New York, 1975).
- [7] John Staples, "Efficient combinatory reduction I", submitted.
- [8] John Staples, "Efficient combinatory reduction II", submitted.

Department of Mathematics and Computer Science,  
Queensland Institute of Technology,  
North Quay, Queensland.