stichting mathematisch centrum



AFDELING INFORMATICA (DEPARTMENT OF COMPUTER SCIENCE)

IW 84/77

SEPTEMBER

K.R. APT & L.G.L.T. MEERTENS

COMPLETENESS WITH FINITE SYSTEMS OF INTERMEDIATE ASSERTIONS FOR RECURSIVE PROGRAM SCHEMES

Preprint

2e boerhaavestraat 49 amsterdam

6777 862

Printed at the Mathematical Centre, 49, 2e Boerhaavestraat, Amsterdam.

The Mathematical Centre, founded the 11-th of February 1946, is a non-profit institution aiming at the promotion of pure mathematics and its applications. It is sponsored by the Netherlands Government through the Netherlands Organization for the Advancement of Pure Research $(\mathsf{Z.W.O})$.

AMS(MOS) subject classification scheme (1970): 68A05

Completeness with finite systems of intermediate assertions for recursive program schemes *)

bу

K.R. Apt & L.G.L.T. Meertens

ABSTRACT

It is proved that in the general case of arbitrary context-free schemes a program is (partially) correct with respect to given initial and final assertions if and only if a suitable finite system of intermediate assertions can be found. Assertions are allowed from an extended state space. This result contrasts with the results of DE BAKKER & MEERTENS [1], where it is proved that if assertions are taken from the original state space V, then in the general case an infinite system of intermediate assertions is needed. In the case of functional schemes (where any deterministic scheme is a functional one) one can take $V \times V$ for the extended state space, thus obtaining a semantical counterpart of the use of auxiliary variables.

KEY WORDS & PHRASES: partial correctness, intermediate assertions, relational framework, extended state space, recursive program schemes.

^{*)} This report will be submitted for publication elsewhere.

1. INTRODUCTION

In DE BAKKER & MEERTENS [1] it was shown that an *infinite* system of intermediate assertions is needed to prove the completeness of the inductive assertion method in the case of an arbitrary system of (mutually) recursive parameterless procedures. On the other hand, GORELICK [4] extended the results of COOK [2] and obtained a completeness result for a Hoare-like axiomatic system (see HOARE [5]) for a fragment of ALGOL 60 in which (deterministic) systems of recursive procedures are allowed. Thus any true asserted statement is provable. (Observe, however, that the axiomatic system uses an oracle determining the truth of formulas from the underlying assertion language.) From the proof we can extract all intermediate assertions about atomic substatements of the original program. Since proofs are finite, we obtain a *finite* system of intermediate assertions. The purpose of this paper is to investigate this apparent contradiction.

The solution lies in the fact that Gorelick uses auxiliary variables (to store theinitial values of variables) which have no semantical counterpart in the relational framework of [1]. Semantically, the use of auxiliary variables corresponds to the use of states which have an additional coordinate (from a space W) inaccessible to a program. We shall call the domain of such states an extended state space.

We prove that if one allows intermediate assertions from an extended state space, then one can always find a finite system of intermediate assertions. More precisely, a program is partially correct with respect to given initial and final assertions if and only if a suitable finite system of assertions from an extended state space can be found. For the space W one can take the powerset of the original state space V. (Theorem 4.4 of [1] shows that for W one could also take the set of all so called index-triple sequences.) In the case of "functional" declaration schemes one can take for W the original state space V. Informally speaking, in the general case programs are viewed as predicate transformers, whereas in the case of functional schemes they are viewed as state transformers.

In [1] it is proved that in the case of regular declaration schemes (corresponding to flow-chart programs) one can always find a finite system of intermediate assertions. In more syntactical terms this can be interpreted

as a statement that auxiliary variables are not needed for correctness proofs in the case of flow-chart programs. They are needed in the general case of arbitrary systems of (parameterless) procedure declarations.

In the relational framework any subset of the state space can be taken as an assertion. This is clearly not the case with Hoare-like axiomatic systems, where assertions are formulas from an assertion language. Thus one should be cautious in translating results from one framework into the other, because there can exist subsets of the state space which do not correspond to (are not defined by) any formula from the assertion language. This problem within the relational framework could be resolved by defining a language over the state space in which assertions could be expressed. However, a natural question then arises which formulas (subsets) should be accepted as assertions. We shall not pursue here this matter further, leaving it as a subject of another paper.

2. PRELIMINARIES

As in DE BAKKER & MEERTENS [1] we shall use binary relations over the state space to provide an interpretation for systems of mutually recursive procedures. More precisely, given a set $P = \{P_1, \dots, P_n\}$ of procedure symbols, we define a language of "statements" S(P) as follows: let $A = \{A_1, A_2, \dots\}$ be a set of "elementary action" symbols, $B = \{t_1, t_2, \dots\}$ a set of "Boolean expressions". S(P) is then the least set containing $A \cup B \cup P$ that is closed under the operations; (sequencing) and U (nondeterministic choice).

By a declaration scheme we mean a set $\mathcal{D} = \{P_1 \leftarrow S_1, \dots, P_n \leftarrow S_n\}$ where for $i = 1, \dots, n$, $P_i \in \mathcal{P}$, $S_i \in \mathcal{S}(\mathcal{P})$.

In [1] a theory of partial correctness and inductive assertions has been worked out in a relational framework. (The meaning of) a program is viewed as a binary relation over the state space, i.e., a set of pairs of initial and final states, whereas an assertion is viewed as a subset of the state space, i.e., the set of states satisfying the assertion. We recall some definitions from [1] which are used below.

Let V be the domain of states. Letters R, R_1, \ldots denote binary relations over V, p, q, r subsets of V, x, y, z elements of V.

$$R_{1};R_{2} = \{(x,y): \exists z[xR_{1}z \land zR_{2}y]\}$$

$$p_{+} = \{(x,x): x \in p\}$$

$$p \circ R = \{y: \exists x[x \in p \land xRy]\}$$

$$I = \{(x,x): x \in V\}$$

$$\Omega \text{ denotes the empty set.}$$

Throughout the paper we use the convention from [1] that in any expression involving programs and assertions built up by using ;, \cup or \subseteq we suppress the subscript " $_+$ ".

So, for example, if we write $p;R\subseteq R;q$ we actually mean $p_+;R\subseteq R;q_+$, i.e., $\forall x,y[(x\in p \land xRy) \rightarrow y\in q]$, or (informally speaking) that the program R is partially correct with respect to p and q. We shall need the following results proved in [1].

If X_1, \dots, X_n , Y_1, \dots, Y_n are subsets of $V \times V$ then by definition $(X_1, \dots, X_n) \leq (Y_1, \dots, Y_n)$ iff $X_i \subseteq Y_i$ for $i = 1, \dots, n$. \leq is a partial ordering.

Let $\mathcal{D} = \{P_1 \Leftarrow S_1, \dots, P_n \Leftarrow S_n\}$ be a declaration scheme. By an *interpretation* $i_{\mathcal{D}}$ into a state space V we mean a mapping from S into relations over V such that for each $A \in A$ $i_{\mathcal{D}}(A)$ is a binary relation over V, for each $t \in \mathcal{B}$ $i_{\mathcal{D}}(t)$ is a subset of V, for each $P \in P$ $i_{\mathcal{D}}(P)$ is a binary relation over V and

(a)
$$i_{\mathcal{D}}(S_1; S_2) = i_{\mathcal{D}}(S_1); i_{\mathcal{D}}(S_2)$$

(b)
$$i_{\mathcal{D}}(S_1 \cup S_2) = i_{\mathcal{D}}(S_1) \cup i_{\mathcal{D}}(S_2)$$

(c)
$$(i_{\mathcal{D}}(P_1), \dots, i_{\mathcal{D}}(P_n)) \text{ is the } \leq -\text{least } n\text{-tuple such that}$$

$$(i_{\mathcal{D}}(P_1), \dots, i_{\mathcal{D}}(P_n)) = (i_{\mathcal{D}}(S_1), \dots, i_{\mathcal{D}}(S_n)) \text{ holds}$$

The above definition is the usual denotational semantics of recursive program schemes. Its justification and equivalence with operational semantics is an immediate consequence of the results proved in [1].

Observe that if for example $\mathcal{D} = \{P \leftarrow t_1; t_2\}$, then due to the convention mentioned above $i_{\mathcal{D}}(P) = i_{\mathcal{D}}(t_1)_+; i_{\mathcal{D}}(t_2)_+$.

In the sequel we shall always consider programs with respect to a given declaration scheme. We shall freely identify statements and their interpretations, hoping that no confusion will result from this.

3. EXTENDING THE STATE SPACE

For any two sets A and B a subset f of A×B is called a function if

$$\forall x,y,z[((x,y) \in f \land (x,z) \in f) \rightarrow y=z].$$

If f is a function then $dom(f) = \{x: \exists y((x,y) \in f)\}$. For $x \in dom(f)$ we denote the unique y such that $(x,y) \in f$ by f(x).

We write $f: A \xrightarrow{part} B$ to denote the fact that f is a subset of $A \times B$ which is a function. If $f: A \xrightarrow{part} B$ and dom(f) = A we write $f: A \rightarrow B$. We define

$$a(f) = \{(x,\sigma) : \sigma \in dom(f) \land (x,f(\sigma)) \in a\},$$

$$a[\sigma] = \{x : (x,\sigma) \in a\}.$$

Observe that $a; R \subseteq R; b$ means that $a_+; R^{\uparrow} \subseteq R^{\uparrow}; b_+, i.e.$, that $\forall x, y, \sigma[((x, \sigma) \in a \land xRy) \rightarrow (y, \sigma) \in b].$

Obviously we have

$$a; R \subseteq R; b \text{ iff for all } \sigma \in W \quad a[\sigma]; R \subseteq R; b[\sigma],$$
 (*)

so correctness of a program with respect to assertions from the extended state space can be proved by a simple reduction to the original state space.

In the proofs below we shall use Scott induction to prove inclusions between relations on $V \times W$.

<u>Scott induction</u>. Let $\mathcal{D} = \{P_1 \leftarrow S_1(P_1, \dots, P_n), \dots, P_n \leftarrow S_n(P_1, \dots, P_n)\}$ be a declaration scheme. Let $\mathcal{E}_{\ell}(X_1, \dots, X_n)$ and $\mathcal{E}_{r}(X_1, \dots, X_n)$ be two expressions built up from assertions from $V \times W$ and programs from $V \times V$ and formal (placeholding) variables X_1, \dots, X_n using ; and \cup and let the following two conditions be satisfied

(i)
$$E_{\ell}(\Omega,\ldots,\Omega) \subseteq E_{r}(\Omega,\ldots,\Omega)$$
, and

(ii) for each
$$R_1, \dots, R_n \subseteq V \times V$$

if $E_{\ell}(R_1, \dots, R_n) \subseteq E_{r}(R_1, \dots, R_n)$
then $E_{\ell}(S_1(R_1, \dots, R_n), \dots, S_n(R_1, \dots, R_n)) \subseteq E_{r}(S_1(R_1, \dots, R_n), \dots, S_n(R_1, \dots, R_n))$.
Then $E_{\ell}(P_1, \dots, P_n) \subseteq E_{r}(P_1, \dots, P_n)$.

The proof is analogous to the proof of the version formulated in [1].

4. COMPLETENESS RESULTS

(i) The general case.

The general form of the theorem we want to prove threatens to obscure its basic simplicity because of the heavy use of indices that is needed. We shall therefore state our theorem by means of an example. It should be clear how to extend it to the general case of an arbitrary context-free declaration scheme.

THEOREM 1. Assume the declaration $P \leftarrow A_1; P; A_2; P; A_3 \cup A_4$. For any two assertions $p, q \subseteq V$

$$p;P \subseteq P;q$$

iff there exists a set W, assertions $a,b \subseteq V \times W$, a set $u \subseteq W$ and functions $f \colon W \xrightarrow{part} W$, $g \colon W \xrightarrow{part} W$ and $\alpha \colon p \to u$ such that

$$a; A_1 \subseteq A_1; a(f),$$

 $b(f); A_2 \subseteq A_2; a(g),$
 $b(g); A_3 \subseteq A_3; b,$
 $a; A_4 \subseteq A_4; b$
(1)

and

$$\alpha \subseteq a$$
, (2) b $\cap (V \times u) \subseteq q \times u$.

PROOF. If part.

We first prove by Scott induction that

$$a;P \subseteq P;b.$$
 (3)

Assume that a; $X \subseteq X$;b for some $X \subseteq V \times V$, i.e., that

$$\forall x, y, \sigma [((x, \sigma) \in a \land xXy) \rightarrow (y, \sigma) \in b].$$

Thus for any function h: $W \xrightarrow{part} W$

$$\forall x, y, \sigma [(\sigma \in dom(h) \land (x, h(\sigma)) \in a \land xXy) \rightarrow (y, h(\sigma)) \in b],$$

i.e. according to our notation,

$$a(h); X \subseteq X; b(h).$$
 (4)

Now, due to the assumptions, Lemma 1 and (4),

$$\begin{array}{lll} a; (A_1; X; A_2; X; A_3) &=& (a; A_1); X; A_2; X; A_3 &\subseteq A_1; a(f); X; A_2; X; A_3 &\subseteq A_1; X; b(f); A_2; X; A_3 &\subseteq A_1; X; A_2; a(g); X; A_3 &\subseteq A_1; X; A_2; b(g); A_3 &\subseteq (A_1; X; A_2; X; A_3); b. \end{array}$$

a, by Lemma 1 and the assumptions,

$$a;(A_1;X;A_2;X;A_3 \cup A_4) \subseteq (A_1;X;A_2;X;A_3 \cup A_4);b.$$

 ϱ obviously $a; \Omega \subseteq \Omega; b$, by Scott induction (3) holds.

We are now ready to prove $p; P \subseteq P; q$. Suppose that $x \in p$ and x P y for $x, y \in V$. We have to show: $y \in q$. By the assumptions $x \in dom(\alpha)$, so (x) is defined and $(x, \sigma) \in a$. By (3), $(y, \sigma) \in b$. Since $\sigma \in u$, by the astions $(y, \sigma) \in q \times u$, so $y \in q$.

Only if part.

We have to establish the existence of W, a, b, u, f, g and α satisfying (1) (2).

 $W = \{r: r \subseteq V\}$, the powerset of V. Define

$$a = \{(x,r): x \in r, r \subseteq V\},$$

$$b = \{(x,r): x \in r \circ P, r \subseteq V\},$$

$$u = \{p\}$$

let

f = {(r,r
$$\circ$$
A₁): r $\subseteq V$ },
g = {(r,r \circ (A₁;P;A₂)): r $\subseteq V$ },
 $\alpha = p \times \{p\}$.

Frve that for any $r \subseteq V$ and h: $W \to W$ a[r]=r, b[r]=roP, [r]=h(r) and b(h)[r]=h(r)oP.

Hence, for any $r \subseteq V$:

$$a[r]; A_{1} = r; A_{1} \subseteq A_{1}; r \circ A_{1} = A_{1}; f(r) = A_{1}; a(f)[r],$$

$$b(f)[r]; A_{2} = f(r) \circ P; A_{2} = (r \circ A_{1}) \circ P; A_{2} = r \circ (A_{1}; P); A_{2} \subseteq$$

$$A_{2}; (r \circ (A_{1}; P) \circ A_{2}) = A_{2}; r \circ (A_{1}; P; A_{2}) = A_{2}; g(r) =$$

$$A_{2}; a(g)[r],$$

So by (*) (1) holds.

By the definition of a, $\alpha = p \times \{p\} \subseteq a$. By the definition of b, b $\cap (V \times \{p\}) = \{(x,p): x \in p \circ P\}$. But by the assumption $p; P \subseteq P; q$, so $p \circ P \subseteq q$. So $\{(x,p): x \in p \circ P\} \subseteq q \times \{p\}$.

This concludes the proof of Theorem 1.

The above theorem should not be too surprising to the diligent reader of DE BAKKER & MEERTENS [1]. A similar fact is there stated as 5.3 on page 354. In [1], however, assertions are always subsets of the original state space, thus the above system $\{a,a(f),a(g),b,b(f),b(g)\}$ from the extended state space corresponds to the infinite system $\{a[\sigma],b[\sigma]\}_{\sigma\in W}$ from the original state space. Observe also that the proof of the "only if part" of 5.3 is much more difficult—the required space W is the space T of the index-triple sequences. On the other hand, T is always a countable set, which is not the case if we take for D the set $\{r: r \subseteq V\}$ (unless V is finite), as we did in the proof of the "only if part". In some situations, however, we can simply take for W the state space V.

(ii) The case of functional declaration schemes.

The reader should be warned that if W=V, the construct a;R becomes ambiguous, its meaning being either $\{(x,y)\colon \exists z[(x,z)\in a \land zRy]\}$ or $a_+;R^{\uparrow}$. In the considerations below we always mean the latter by imposing the convention that letters a,b (with possible subscripts) always mean assertions.

Assume an arbitrary context-free declaration scheme

$$\{P_{i} \leftarrow S_{i,1} \cup S_{i,2} \cup ... \cup S_{i,M_{i}}\}_{i=1}^{n},$$
 (5)

with M_i some integer ≥ 1 , and each $S_{i,j}$, $j=1,\ldots,M_i$, of the form

with elementary actions A(i,j,k), $P(i,j,k) \in \{P_1,\ldots,P_n\}$, and $K_{i,j}$ an integer ≥ 0 (if $K_{i,j}=0$ then $S_{i,j}$ is simply A(i,j,0)).

DEFINITION 1. The scheme (5) is called functional if

i) for all i=1,...,n, j=1,...,M and k=0,...,K i,j A(i,j,k) is a function. ii) each P_i (i=1,...,n) is a function.

Observe that the condition ii) is clearly fulfilled if i) holds and for each i=1,...,n and j_1,j_2 such that $1 \le j_1 < j_2 \le M_i$

$$dom(A(i,j_1,0)) \cap dom(A(i,j_2,0)) = \emptyset$$

We shall now prove that in the case of functional declaration schemes we can take $V \times V$ as an extended state space. Again we prove it by means of an example, leaving the general case to the reader.

THEOREM 2. Suppose that $P \leftarrow A_2; P; A_2; P; A_3 \cup A_4$ is a functional declaration scheme. For any two assertions $p, q \subseteq V$

$$p;P \subseteq P;q$$

iff there exist assertions $a,b \subseteq V \times V$, a set $u \subseteq V$ and functions $f \colon V \xrightarrow{part} V$, $g \colon V \xrightarrow{part} V$ and $\alpha \colon p \to u$ such that (1) and (2) hold.

<u>PROOF.</u> The "if part" is an immediate consequence of Theorem 1. To prove the "only if part", put $a = \{(x,x): x \in V\}$, $b = \{(y,x): xPy\}$, u=p and let $f=A_1$, $g=A_1$; P; A_2 and $\alpha = \{(x,x): x \in p\}$.

Let x,y,σ be arbitrary elements of V.

- (i) We have to show: $a; A_1 \subseteq A_1; a(f)$, i.e., $(x, \sigma) \in a$ and xA_1y implies $(y, \sigma) \in a(f)$, which is equivalent to $\sigma \in dom(f)$ and $(y, f(\sigma)) \in a$. Suppose $(x, \sigma) \in a$ and xA_1y . By the definition of a $x=\sigma$, so, since $f = A_1$, $\sigma \in dom(f)$, $y = f(x) = f(\sigma)$ and $(y, f(\sigma)) \in a$.
- (ii) We have to show: $b(f); A_2 \subseteq A_2; a(g), i.e., \sigma \in dom(f), (x, f(\sigma)) \in b$ and xA_2y implies $\sigma \in dom(g)$ and $(y,g(\sigma)) \in a$.

Suppose $\sigma \in dom(f)$, $(x, f(\sigma)) \in b$ and xA_2y . Then $f(\sigma)Px$. Since $\sigma A_1f(\sigma)$

we find $\sigma(A_1;P;A_2)y$, so $\sigma \in dom(g)$ and $y = g(\sigma)$, i.e., $(y,g(\sigma)) \in a$.

(iii) We have to show: $b(g); A_3 \subseteq A_3; b$, i.e., $\sigma \in dom(g)$, $(x,g(\sigma)) \in b$ and xA_3y implies $(y,\sigma) \in b$. Suppose $\sigma \in dom(g)$, $(x,g(\sigma)) \in b$ and xA_3y . Then $g(\sigma)Px$, i.e., by the definition of $g(A_1;P;A_2;P)x$ and so $\sigma(A_1;P;A_2;P;A_3)y$. Thus, σPy , which means that $(y,\sigma) \in b$.

(iv) We have to show: $a; A_4 \subseteq A_4; b$, i.e., $(x,\sigma) \in a$ and xA_4y implies $(y,\sigma) \in b$. Suppose $(x,\sigma) \in a$ and xA_4y . Then $x=\sigma$ and xPy, i.e., $(y,\sigma) \in b$.

- (v) Obviously $\alpha \subseteq a$.
- (vi) We have to show: b \cap ($V \times u$) \subseteq $q \times u$, i.e., $(x,y) \in b$ and $y \in p$ implies $x \in q$. Suppose $(x,y) \in b$ and $y \in p$. Then, $(y,x) \in p$, and since $p;P \subseteq p;q$, we find $x \in q$. This concludes the proof.

<u>REMARK.</u> In most cases one deals with deterministic programs which can be represented by means of functional declaration schemes. A notable exception is Dijkstra's guarded command construct (see DIJKSTRA [3]), which gives rise to programs of bounded determinacy. It is easy to see that in the case of declaration schemes of bounded determinacy we can take for W the set of all finite subsets of V. The proof is analogous to the proof of the "only if part" of Theorem 1.

5. AN APPLICATION

Having obtained specific forms of completeness results we shall illustrate their usefulness by the following example.

Let the state space V be the set of natural numbers N. Consider the following declaration

$$P \leftarrow [n \le 100]; [n := n+11]; P; P \cup [n > 100]; [n := n-10]$$
 (6)

where, of course, $[n \le 100] = \{x: x \le 100\}$, $[n := n+11] = \{(x,y): y = x+11\}$ and so on. P is of course McCarthy's well-known 91 function defined in a

relational framework. We want to prove that

$$[n \le 100]; P \subseteq P; [n=91].$$
 (7)

Observe that the above declaration is of the form $P \leftarrow A_1; P; A_2; P; A_3 \cup A_4$ where

$$A_1 = [n \le 100]; [n := n+11]$$
 $A_2 = I$
 $A_3 = I$
 $A_4 = [n > 100]; [n := n-10]$

Since (6) is a functional declaration scheme, we can use Theorem 2 to prove (7). The easiest way to proceed is to define the required relations and functions as in the proof of Theorem 2, taking for P $[n \le 100]$; [n := 91] U [n > 100]; [n := n-10], and to check that (1) and (2) hold.

Thus we define:

```
a = \{(x,x): x \in N\};
b = \{(x,y): (x=91 \land y \le 100) \lor (x=y-10 \land y>100)\};
u = \{x: x \le 100\};
f = \{(x,y): x \le 100 \land y = x+11\};
g = [n \le 100]; [n := n+11]; ([n \le 100]; [n := 91] \cup [n > 100]; [n := n-10];
= \{(x,y): (90 \le x \le 100 \land y=x+1) \lor (x<90 \land y=91)\};
\alpha = \{(x,x): x \le 100\}.
```

We leave the task of checking that (1) and (2) indeed hold to the reader. Now, by Theorem 2, (7) holds.

ACKNOWLEDGEMENT

We are grateful to J.W. de Bakker for critical comments on an earlier version.

REFERENCES

- [1] DE BAKKER, J.W. & L.G.L.T. MEERTENS, On the Completeness of the Inductive Assertion Method, Journal of Computer and System Sciences, vol. 11, No. 3, pp. 323-357 (1975).
- [2] COOK, S.A., Axiomatic and Interpretive Semantics for an Algol Fragment, Technical Report no. 79, University of Toronto (1975).
- [3] DIJKSTRA, E.W., A Discipline of Programming, Prentice-Hall, Inc., London (1976).
- [4] GORELICK, G.A., A Complete Axiomatic System for Proving Assertions about Recursive and Non-Recursive Programs, Technical Report no. 75, University of Toronto (1975).
- [5] HOARE, C.A.R., An Axiomatic Basis for Programming Language Constructs, C. ACM 12, pp. 576-580 (1969).