

Interproc analyzer for recursive programs with numerical variables

Bertrand Jeannet

Contents

1	Invoking Interproc	1
2	The “Simple” language	2
2.1	Syntax and informal semantics	2
2.1.1	Program.	2
2.1.2	Instructions.	3
2.1.3	Expressions.	3
2.2	Formal semantics	6
2.2.1	Semantics domains.	6
2.2.2	Semantics of expressions.	6
2.2.3	Semantics of programs	7
3	What Interproc does for you	8

1 Invoking Interproc

The executable (`interproc` or `interproc.opt` is invoked as follows:

```
interproc <options> <inputfile>
```

The input file should be a valid “Simple” program. The options are:

`-debug <int>` debug level, from 0 (lowest) to 4 (highest). Default is 0.

`-domain {box|octagon|polka|polkastrict|polkaeq|ppl|pplstrict|pplgrid|polkagrid}`
abstract domain to use (default: `polka`). All domains supported by the Apron library can be specified:

box	intervals
octagon	octagons
polka,ppl	topologically closed convex polyhedra
polkastrict,pplstrict	possibly non-closed convex polyhedra
polkaeq	linear equalities
pplgrid	linear congruences
polkagrid	reduced product of linear congruences and Polka convex polyhedra

`-depth <int>` depth of recursive iterations (default 2, may only be more). See Fixpoint library documentation.

`-guided <bool>` if true, guided analysis of Gopand and Reps (default: false). See Analyzer

-widening <bool><int><int><int> specifies usage of widening first heuristics, delay and frequency of widening, and nb. of descending steps (default: false 1 1 2). See Fixpoint

-analysis <('f'|'b')+> sequence of forward and backward analyses to perform (default "f")

2 The “Simple” language

Program Sample:

```
/* Procedure definition */
proc MC(n:int) returns (r:int)
var t1:int, t2:int;
begin
  if (n>100) then
    r = n-10;
  else
    t1 = n+11;
    t2 = MC(t1);
    r = MC(t2);
  endif;
end
/* Main procedure */
var a:int, b:int;
begin
  b = MC(a);
end
```

2.1 Syntax and informal semantics

2.1.1 Program.

<code><program></code>	<code>::=</code>	<code><procedure_def>*</code> <code>[var <vars_decl> ;]</code> <code>begin <statement>+ end</code>	definition of procedures local variables of the main procedure body of the main procedure
<code><procedure_def></code>	<code>::=</code>	<code>proc id (<vars_decl>) returns (<vars_decl>)</code> <code>[var <vars_decl> ;]</code> <code>begin <statement>+ end</code>	signature, with names of input and output formal parameters (other) local variables body

A procedure definition first declares a list of formal input parameters, then a list of formal output parameters (which are local variables that defines the tuple of returned values), then a list of local variables, and then a body. A procedure ends with its last statement.

<code><vars_decl></code>	<code>::=</code>	<code>ϵ</code> <code>id : <type> (, id : <type>)*</code>
<code><type></code>	<code>::=</code>	<code>int</code> <code>real</code>

Variables are either of integer type, or of real type. Real variables can behave as floating-point variables by modifying the exact semantics of arithmetic operations (see below).

2.1.2 Instructions.

```
<statement> ::= skip ;
              | halt ;
              | fail ;
              | assume <bexpr> ;
              | id = random ;
              | id = <nexpr> ;
              | id = id ( <var_list> ) ;
              | ( <var_list> ) = id ( <var_list> ) ;
              | if <bexpr> then <statement>+ [else <statement>+ ] endif ;
              | while <bexpr> do <statement>+ done ;
<var_list> ::=  $\epsilon$  | id ( , id)*
```

The `skip` operation does nothing. The `halt` operation halts the program execution. The `fail` has exactly the same operational semantics, but, in addition, indicates a final control point for backward analysis (see below).

The `assume expr` instruction is equivalent to `if expr then skip; else halt; .` It may be used to abstract non-deterministic assignments. For instance, the non-deterministic assignment of a value between 0 and 2 to a variable `x` can be written as

```
x = random;
assume x>=0 and x<=2;
```

A call to a procedure with a single return value can be written either as `x = P(...)` or `(x)=P(...)`. A call to a procedure with multiple return values is written as `(x,y) = P(a,b)`. Input parameters are passed by value. Formal input parameters in a procedure are assumed immutable.

Conditional and iteration constructions have the usual semantics.

2.1.3 Expressions.

Boolean expressions.

```
<bexpr> ::= true | false | brandom
          | <constraint>
          | not <bexpr> | <bexpr> or <bexpr> | <bexpr> and <bexpr>
          | ( <bexpr> )
<constraint> ::= <nexpr> (== | >= | > | <= | <) <nexpr>
```

Atoms of Boolean expressions are Boolean constants, the unknown value denoted by `brandom`, and constraints over numerical variables.

Numerical expressions.

```

<nexpr> ::= <coeff> | random
         | id
         | <nexpr> <binop> <nexpr>
         | ( <nexpr> )
<binop> ::= (+|-|*|/|%)[_(<i|f|d|l|q>)[,(n|0|+oo|-oo|?)]
<unop>  ::= -
         | (cast|sqrt)[_(<i|f|d|l|q>)[,(n|0|+oo|-oo|?)]
<coeff> ::= <float> | <rational>
<float> ::= C/OCaml floating-point number syntax
<rational> ::= <integer> | <integer>/<integer>

```

The syntax of numerical expressions is the same as the one used in the APRON interactive toplevel (in OCaml language).

The priorities between (Boolean and arithmetic) operators is $\{ *, / \} > \{ +, - \} > \{ <, <=, =, >=, > \} > \{ \text{not} \} > \{ \text{and} \} > \{ \text{or} \}$

By default, numerical expressions are evaluated using exact arithmetic on real numbers (even if some involved variables are integer), and when assigned to a variable, filtered by the variable domain (when it is integers). For instance, the assignment

```
x = 3/4;
```

is equivalent to the `halt` instruction if `x` is declared as integer, because the rational `3/4` is not an integer value. The problem here is that the instruction is not well-typed (even if this is not checked). If one want to emulate the semantics of the integer division operation, with rounding to zero, one should write

```
x = 3 /_i,0 4;
```

Indeed, most arithmetic operators have optional qualifiers that allows to modify their semantics. The first qualifier indicates a numerical type, the second one the rounding mode.

- By default, operations have an exact arithmetic semantics in the real numbers (even if involved variables are of integer).

The type qualifiers modify this default semantics. Their meaning is as follows:

```

i   integer semantics
f   IEEE754 32 bits floating-point semantics
d   IEEE754 64 bits floating-point semantics
l   IEEE754 80 bits extended floating-point semantics
q   IEEE754 128 bits floating-point semantics

```

- By default, the rounding mode is “any” (this applies only in non-real semantics), which allows to emulate all the following rounding modes:

```

n   nearest
0   towards zero
+oo towards infinity
-oo towards minus infinity
?   emulates any rounding mode

```

Examples of numerical expressions.

We illustrate the syntax and (approximated) semantics of numerical expressions with the following program, annotated by Interproc using the command

```
interproc.opt -domain polka -colorize false examples/numerical.spl

proc integer () returns () var a : int, b : int, c : int;
begin
  /* (L5 C5) top */
  a = 5; /* (L6 C8) [|a-5=0|] */
  b = 2; /* (L7 C8) [|b-2=0; a-5=0|] */
  if brandom then
    /* (L8 C17) [|b-2=0; a-5=0|] */
    c = a / b; /* (L9 C14) bottom */
  else
    /* (L10 C6) [|b-2=0; a-5=0|] */
    c = a /_i,? b; /* (L11 C16) [|b-2=0; a-5=0; -c+3>=0; c-2>=0|] */
    c = a /_i,0 b; /* (L12 C18) [|c-2=0; b-2=0; a-5=0|] */
    c = a /_i,-oo b; /* (L13 C20) [|c-2=0; b-2=0; a-5=0|] */
    c = a /_i,+oo b; /* (L14 C20) [|c-3=0; b-2=0; a-5=0|] */
    c = a /_i,n b; /* (L15 C18) [|b-2=0; a-5=0; -c+3>=0; c-2>=0|] */
    c = a %_i,? b; /* (L16 C16) [|c-1=0; b-2=0; a-5=0|] */
  endif; /* (L17 C8) [|c-1=0; b-2=0; a-5=0|] */
end

proc exact () returns (z : real) var x : real, y : real;
begin
  /* (L22 C5) top */
  x = 5; /* (L23 C8) [|x-5=0|] */
  y = 2; /* (L24 C8) [|y-2=0; x-5=0|] */
  z = x / y; /* (L25 C10) [|2z-5=0; y-2=0; x-5=0|] */
  y = 0.1; /* (L26 C10)
    [|2z-5=0; 36028797018963968y-3602879701896397=0; x-5=0|] */
  z = x + y; /* (L27 C10)
    [|36028797018963968z-183746864796716237=0;
    36028797018963968y-3602879701896397=0; x-5=0|] */
  z = z - y; /* (L28 C10)
    [|z-5=0; 36028797018963968y-3602879701896397=0; x-5=0|] */
end

proc floating () returns (z : real) var x : real, y : real;
begin
  /* (L33 C5) top */
  x = 5; /* (L34 C8) [|x-5=0|] */
  y = 2; /* (L35 C8) [|y-2=0; x-5=0|] */
  z = x /_f,? y; /* (L36 C13)
    [|y-2=0; x-5=0;
    -713623846352979940529142984724747568191373312z
    +1784059828558929176909397126420998565333565441>=0;
    713623846352979940529142984724747568191373312z
    -1784059403205970525736317797202739275623301119>=0|] */
  y = 0.1; /* (L37 C10)
    [|36028797018963968y-3602879701896397=0; x-5=0;
    -713623846352979940529142984724747568191373312z
    +1784059828558929176909397126420998565333565441>=0;
    713623846352979940529142984724747568191373312z
    -1784059403205970525736317797202739275623301119>=0|] */
  z = x +_f,? y; /* (L38 C14)
    [|36028797018963968y-3602879701896397=0; x-5=0;
    -713623846352979940529142984724747568191373312z
```

```

+3639482050260215524856578735848702239922192385>=0;
713623846352979940529142984724747568191373312z
-3639481182540179876463495959770156714984210431>=0[] */
z = z -_f,? y; /* (L39 C14)
[|36028797018963968y-3602879701896397=0; x-5=0;
-5986310706507378352962293074805895248510699696029696z
+29931560882862943060522688904967984086376852368654337>=0;
5986310706507378352962293074805895248510699696029696z
-29931546182211708189135890236173744477275669529624577>=0[] */
end

var z : real
begin
/* (L43 C5) top */
() = integer(); /* (L44 C17) top */
z = exact(); /* (L45 C14) [z-5=0] */
z = floating(); /* (L46 C17)
[|-5986310706507378352962293074805895248510699696029696z
+29931560882862943060522688904967984086376852368654337>=0;
5986310706507378352962293074805895248510699696029696z
-29931546182211708189135890236173744477275669529624577>=0[] */
end

```

2.2 Formal semantics

We give here the standard operational semantics of the “Simple” language.

2.2.1 Semantics domains.

They are summarized by the following table:

$k \in K$	program’s control points
$\epsilon \in \mathbf{Env} = \mathbf{Var} \rightarrow \mathbb{R}$	environments
$(k, \epsilon) \in K \times \mathbf{Env}$	activation records
$s \in \mathbf{S} = (K \times \mathbf{Env})^+$	program states, which are non-empty stacks

An environment is defined by a function assigning a value to a set of variables \mathbf{Var} visible in the current scope.

A program state is a non-empty stack of activation records. A stack is viewed as a word where the last letter corresponds to the top of the stack.

2.2.2 Semantics of expressions.

The semantics of a numerical expression \mathbf{nexpr} is a function

$$\llbracket \mathbf{nexpr} \rrbracket^n : \mathbf{Env} \rightarrow \wp(\mathbb{R})$$

We refer to [1] for the precise definition of this semantic function.

The semantics of a Boolean expression \mathbf{bexpr} is a function

$$\llbracket \mathbf{bexpr} \rrbracket^b : \mathbf{Env} \rightarrow \wp(\mathbb{B})$$

which is defined inductively as follows:

$$\begin{aligned}
\llbracket \text{true} \rrbracket^b &= \lambda \epsilon . \{ \text{true} \} \\
\llbracket \text{false} \rrbracket^b &= \lambda \epsilon . \{ \text{false} \} \\
\llbracket \text{brandom} \rrbracket^b &= \lambda \epsilon . \{ \text{true}, \text{false} \} \\
\llbracket \text{not } e \rrbracket^b &= \lambda \epsilon . \{ \neg b \mid b \in \llbracket e \rrbracket^b(\epsilon) \} \\
\llbracket e_1 \text{ and } e_2 \rrbracket^b &= \lambda \epsilon . \{ b_1 \wedge b_2 \mid b_1 \in \llbracket e_1 \rrbracket^b(\epsilon) \wedge b_2 \in \llbracket e_2 \rrbracket^b(\epsilon) \} \\
\llbracket e_1 \text{ or } e_2 \rrbracket^b &= \lambda \epsilon . \{ b_1 \vee b_2 \mid b_1 \in \llbracket e_1 \rrbracket^b(\epsilon) \wedge b_2 \in \llbracket e_2 \rrbracket^b(\epsilon) \} \\
\llbracket e_1 \text{ op } e_2 \rrbracket^b &= \llbracket (e_1 - e_2) \text{ op } 0 \rrbracket^b \\
\llbracket e = 0 \rrbracket^b &= \lambda \epsilon . (0 \in \llbracket e \rrbracket^n(\epsilon)) \\
\llbracket e \geq 0 \rrbracket^b &= \lambda \epsilon . (\llbracket e \rrbracket^n(\epsilon) \cap [0, \infty] \neq \emptyset) \\
\llbracket e \leq 0 \rrbracket^b &= \lambda \epsilon . (\llbracket e \rrbracket^n(\epsilon) \cap]-\infty, 0] \neq \emptyset) \\
\llbracket e > 0 \rrbracket^b &= \lambda \epsilon . (\llbracket e \rrbracket^n(\epsilon) \cap]0, \infty] \neq \emptyset) \\
\llbracket e < 0 \rrbracket^b &= \lambda \epsilon . (\llbracket e \rrbracket^n(\epsilon) \cap]-\infty, 0[\neq \emptyset)
\end{aligned}$$

The semantics of a Boolean expression can be extended as a predicate on environments as follows:

$$\begin{aligned}
\llbracket \text{bexpr} \rrbracket^b : \mathbf{Env} &\rightarrow \mathbb{B} \\
\epsilon &\mapsto (\text{true} \in \llbracket \text{bexpr} \rrbracket^b)
\end{aligned}$$

2.2.3 Semantics of programs

We define the semantics of a program as a discrete dynamical system $(\mathcal{S}, \mathcal{S}_0, \rightarrow)$ where \mathcal{S} is the state-space of the program (see above), \mathcal{S}_0 is the set of initial states, and \rightarrow is the transition relation between states.

If k_0 denotes the starting point of the main procedure, then \mathcal{S}_0 is defined as

$$\mathcal{S}_0 = \{k_0\} \times \mathbf{Env}$$

This reflects the hypothesis that variables are uninitialized (read: they may contain arbitrary values) at start.

The transition relation is defined as follows:

$$\begin{aligned}
&\frac{(k) \text{ skip } (k')}{\omega \cdot (k, \epsilon) \rightarrow \omega \cdot (k', \epsilon)} \quad \frac{(k) \text{ assume } \text{bexpr} (k') \quad \text{true} \in \llbracket \text{bexpr} \rrbracket^b(\epsilon)}{\omega \cdot (k, \epsilon) \rightarrow \omega \cdot (k', \epsilon)} \\
&\frac{(k) \text{ x=random } (k') \quad \forall y \in \mathbf{Var} \setminus \{x\} : \epsilon'(y) = \epsilon(y)}{\omega \cdot (k, \epsilon) \rightarrow \omega \cdot (k', \epsilon')} \\
&\frac{(k) \text{ x=nexpr } (k') \quad v \in \llbracket \text{nexpr} \rrbracket^n(\epsilon)}{\omega \cdot (k, \epsilon) \rightarrow \omega \cdot (k', \epsilon[x \mapsto v])} \\
&\frac{(k) \text{ if } \text{bexpr} \text{ then } (k_1) \dots (k'_1) \text{ else } (k_2) \dots (k'_2) ; (k')}{\begin{aligned} &(k, \epsilon) \rightarrow (k_1, \epsilon) \text{ if } \text{true} \in \llbracket \text{bexpr} \rrbracket^b(\epsilon) \\ &(k, \epsilon) \rightarrow (k_2, \epsilon) \text{ if } \text{false} \in \llbracket \text{bexpr} \rrbracket^b(\epsilon) \\ &(k'_1, \epsilon) \rightarrow (k', \epsilon) \\ &(k'_2, \epsilon) \rightarrow (k', \epsilon) \end{aligned}} \\
&\frac{(k) \text{ while } \text{bexpr} \text{ do } (k_1) \dots (k'_1) \text{ done } (k_2)}{\begin{aligned} &(k, \epsilon) \rightarrow (k_1, \epsilon) \text{ if } \text{true} \in \llbracket \text{bexpr} \rrbracket^b(\epsilon) \\ &(k, \epsilon) \rightarrow (k_2, \epsilon) \text{ if } \text{false} \in \llbracket \text{bexpr} \rrbracket^b(\epsilon) \\ &(k'_1, \epsilon) \rightarrow (k, \epsilon) \end{aligned}}
\end{aligned}$$

$$\begin{array}{c}
(k) (y_0, \dots, y_N) = f(x_0, \dots, x_M) (k') \\
\epsilon_{\text{start}}(\vec{pin}) = \epsilon(\vec{x}) \\
\epsilon' = \epsilon[\vec{y} \mapsto \epsilon_{\text{exit}}(\vec{pout}) \\
\hline
\omega \cdot (k, \epsilon) \rightarrow \omega \cdot (k', \epsilon) \cdot (k_{\text{start}}^f, \epsilon_{\text{start}}) \\
\omega \cdot (k', \epsilon) \cdot (k_{\text{exit}}^f, \epsilon_{\text{exit}}) \rightarrow \omega \cdot (k', \epsilon')
\end{array}$$

3 What Interproc does for you

Interproc takes as input a “Simple” program, annotates it with control points, perform forward and/or backward analysis, and print the annotated programs with invariants associated with each control point, like this:

Annotated program after forward analysis:

```

var x : int, y : int, z : int
begin
  /* (L5 C5) top */
  assume z >= 10 and z <= 20; /* (L6 C25) [| -z+20>=0; z-10>=0|] */
  x = 0; /* (L7 C8) [| x=0; -z+20>=0; z-10>=0|] */
  y = 0; /* (L7 C13) [| -3x+y=0; -x+z+1>=0; -z+20>=0; z-10>=0; x>=0|] */
  while x <= z do
    /* (L8 C17) [| -3x+y=0; -x+z>=0; -z+20>=0; z-10>=0; x>=0|] */
    x = x + 1; /* (L9 C12)
  [| -3x+y+3=0; -x+z+1>=0; -z+20>=0; z-10>=0; x-1>=0|] */
    y = y + 3; /* (L10 C12) [| -3x+y=0; -x+z+1>=0; -z+20>=0; z-10>=0; x-1>=0|] */
  done; /* (L11 C7) [| -3x+y=0; -x+z+1=0; -x+21>=0; x-11>=0|] */
  if y >= 42 then
    /* (L12 C15) [| -3x+y=0; -x+z+1=0; -x+21>=0; x-14>=0|] */
    fail; /* (L13 C9) bottom */
  endif; /* (L14 C8) [| -3x+y=0; -x+z+1=0; -3x+41>=0; x-11>=0|] */
end

```

Invariants

An *invariant* I is a set of environments:

$$I \in \wp(\text{Env})$$

The result of the analyses performed by Interproc is the projection $p(X)$ of sets of stacks on their top elements. The result of this projection is a function that maps control point to invariants:

$$\begin{array}{ccc}
p : \wp((K \times \text{Env})^+) & \longrightarrow & (K \rightarrow \wp(\text{Env})) \\
X & \longmapsto & \lambda k . \{\epsilon \mid \omega \cdot (k, \epsilon) \in X\}
\end{array}$$

Intuitively, the invariant on numerical variables associated with each program point, as returned by Interproc, indicates the possible values for these variables at each program point, for any stack tail.

Single forward/reachability analysis

A forward analysis compute the set of reachable states of the program, which is defined as:

$$\text{Reach} = \{s \in S \mid s_0 \in S_0 \wedge s_0 \rightarrow^* s\}$$

Interproc returns an overapproximation of $p(\text{Reach})$.

Single backward/coreachability analysis

A backward analysis computes the set of states from which one can reach a `fail` instruction.

The set of states coreachable from a `fail` instruction is defined as:

$$\mathit{Coreach} = \{s \in S \mid s \rightarrow^* s_f \wedge s_f \in F\}$$

with the set of final states

$$F = \{\omega \cdot (k, \epsilon) \mid k \in K_{\text{fail}} \wedge \epsilon \in \text{Env}\}$$

where K_{fail} is the set of control points just preceding `fail` instructions. `Interproc` returns an overapproximation of $p(\mathit{Coreach})$.

Combined forward/backward analyses

`Interproc` allows to combine forward and backward analysis any number of time.

If X denotes the invariant computed by `Interproc` at the previous step, at the current step:

- a forward analysis will consider the set of initial states

$$S_0 \cap p^{-1}(X)$$

- a backward analysis will consider the set of final states

$$F \cap p^{-1}(X)$$

For instance, after executing the command

```
interproc -analysis fb essai.spl
```

on the previous program, we get

Annotated program after forward analysis

```
var x : int, y : int, z : int
begin
  /* (L5 C5) top */
  assume z >= 10 and z <= 20; /* (L6 C25) [| -z+20>=0; z-10>=0|] */
  x = 0; /* (L7 C8) [| x=0; -z+20>=0; z-10>=0|] */
  y = 0; /* (L7 C13) [| -3x+y=0; -x+z+1>=0; -z+20>=0; z-10>=0; x>=0|] */
  while x <= z do
    /* (L8 C17) [| -3x+y=0; -x+z>=0; -z+20>=0; z-10>=0; x>=0|] */
    x = x + 1; /* (L9 C12)
    [| -3x+y+3=0; -x+z+1>=0; -z+20>=0; z-10>=0; x-1>=0|] */
    y = y + 3; /* (L10 C12) [| -3x+y=0; -x+z+1>=0; -z+20>=0; z-10>=0; x-1>=0|] */
  done; /* (L11 C7) [| -3x+y=0; -x+z+1=0; -x+21>=0; x-11>=0|] */
  if y >= 42 then
    /* (L12 C15) [| -3x+y=0; -x+z+1=0; -x+21>=0; x-14>=0|] */
    fail; /* (L13 C9) bottom */
  endif; /* (L14 C8) [| -3x+y=0; -x+z+1=0; -3x+41>=0; x-11>=0|] */
end
```

Annotated program after backward analysis

```
var x : int, y : int, z : int
begin
  /* (L5 C5) [| -z+20>=0; z-13>=0|] */
  assume z >= 10 and z <= 20; /* (L6 C25) [| -z+20>=0; z-13>=0|] */
  x = 0; /* (L7 C8) [| x=0; -z+20>=0; z-13>=0|] */
end
```

```

y = 0; /* (L7 C13) [| -3x+y=0; -x+z+1>=0; -z+20>=0; z-13>=0; x>=0|] */
while x <= z do
  /* (L8 C17) [| -3x+y=0; -x+z>=0; -z+20>=0; z-13>=0; x>=0|] */
  x = x + 1; /* (L9 C12)
  [| -3x+y+3=0; -x+z+1>=0; -z+20>=0; z-13>=0; x-1>=0|] */
  y = y + 3; /* (L10 C12) [| -3x+y=0; -x+z+1>=0; -z+20>=0; z-13>=0; x-1>=0|] */
done; /* (L11 C7) [| -3x+y=0; -x+z+1=0; -x+21>=0; x-14>=0|] */
if y >= 42 then
  /* (L12 C15) [| -3x+y=0; -x+z+1=0; -x+21>=0; x-14>=0|] */
  fail; /* (L13 C9) bottom */
endif; /* (L14 C8) bottom */
end

```

References

- [1] A. Miné. *Domaines numériques abstraits faiblement relationnels (Weakly relational numerical abstract domains)*. PhD thesis, École Normale Supérieure de Paris, 2004. PDF.