Program Analysis (70020)

While Language

Herbert Wiklicky

Department of Computing
Imperial College London

herbert@doc.ic.ac.uk
h.wiklicky@imperial.ac.uk

Autumn 2022

Syntactic Constructs

We use the following syntactic categories:

$$\begin{array}{lcll} a & \in & \textbf{AExp} & \text{arithmetic expressions} \\ b & \in & \textbf{BExp} & \text{boolean expressions} \\ S & \in & \textbf{Stmt} & \text{statements} \end{array}$$

Abstract Syntax of WHILE

The syntax of the language WHILE is given by the following **abstract syntax**:

$$
\begin{array}{lll}
a & ::= & x \mid n \mid a_1 \; op_a \; a_2 \\
b & ::= & true \mid false \mid not\; b \mid b_1 \; op_b \; b_2 \mid a_1 \; op_r \; a_2 \\
S & ::= & x := a \\
 & & \mid \mathbf{skip} \\
 & & \mid S_1 ; S_2 \\
 & & \mid \mathbf{if}\; b \;\mathbf{then}\; S_1 \;\mathbf{else}\; S_2 \\
 & & \mid \mathbf{while}\; b \;\mathbf{do}\; S
\end{array}
$$

Syntactical Categories

We assume some countable/finite set of variables is given;

$$
\begin{array}{rcll}
x, y, z, \ldots & \in & \mathbf{Var} & \text{variables} \\
n, m, \ldots & \in & \mathbf{Num} & \text{numerals} \\
\ell, \ldots & \in & \mathbf{Lab} & \text{labels}
\end{array}
$$

Numerals (integer constants) will not be further defined and neither will the operators:

$$
\begin{array}{rcll}
op_a & \in & \mathbf{Op}_a & \text{arithmetic operators, e.g. } +, -, \times, \ldots \\
op_b & \in & \mathbf{Op}_b & \text{boolean operators, e.g. } \wedge, \vee, \ldots \\
op_r & \in & \mathbf{Op}_r & \text{relational operators, e.g. } =, <, \leq, \ldots
\end{array}
$$

Labelled Syntax of WHILE

The labelled syntax of the language WHILE is given by the following **abstract syntax**:

$$
\begin{array}{lcl}
a & ::= & x \mid n \mid a_1 \; op_a \; a_2 \\
b & ::= & \mathit{true} \mid \mathit{false} \mid \mathit{not}\; b \mid b_1 \; op_b \; b_2 \mid a_1 \; op_r \; a_2 \\
S & ::= & [x := a]^{\ell} \\
 & & \mid [\mathbf{skip}]^{\ell} \\
 & & \mid S_1 ; S_2 \\
 & & \mid \mathbf{if}\; [b]^{\ell} \;\mathbf{then}\; S_1 \;\mathbf{else}\; S_2 \\
 & & \mid \mathbf{while}\; [b]^{\ell} \;\mathbf{do}\; S
\end{array}
$$

An Example in WHILE

An example of a program written in this WHILE language is the following one which computes the factorial of the number stored in x and leaves the result in z:

$$
\begin{array}{l}
[\, y := x \,]^1; \\
[\, z := 1 \,]^2; \\
\mathbf{while}\; [y > 1]^3 \;\mathbf{do}\; (\\
\qquad [\, z := z * y \,]^4; \\
\qquad [\, y := y - 1 \,]^5); \\
[\, y := 0 \,]^6
\end{array}
$$

Note the use of meta-symbols, brackets, to group statements.

Concrete Syntax of WHILE

To avoid using brackets (as meta-symbols) we could also use the **concrete syntax** of the language WHILE as follows:

$$
\begin{array}{lcl}
a & ::= & x \mid n \mid a_1 \; op_a \; a_2 \\
b & ::= & \mathit{true} \mid \mathit{false} \mid \mathit{not}\; b \mid b_1 \; op_b \; b_2 \mid a_1 \; op_r \; a_2 \\
S & ::= & x := a \\
 & & \mid \textbf{skip} \\
 & & \mid S_1 ; S_2 \\
 & & \mid \textbf{if}\; b \;\textbf{then}\; S_1 \;\textbf{else}\; S_2 \;\textbf{fi} \\
 & & \mid \textbf{while}\; b \;\textbf{do}\; S \;\textbf{od}
\end{array}
$$

Initial Label

When presenting examples of Data Flow Analyses we will use a number of operations on programs and labels. The first of these is

$$\mathit{init} : \textbf{Stmt} \to \textbf{Lab}$$

which returns the initial label of a statement:

$$
\begin{array}{rcl}
\mathit{init}([\, x := a \,]^\ell) & = & \ell \\
\mathit{init}([\, \textbf{skip} \,]^\ell) & = & \ell \\
\mathit{init}(S_1 ; S_2) & = & \mathit{init}(S_1) \\
\mathit{init}(\textbf{if}\; [b]^\ell \;\textbf{then}\; S_1 \;\textbf{else}\; S_2) & = & \ell \\
\mathit{init}(\textbf{while}\; [b]^\ell \;\textbf{do}\; S) & = & \ell
\end{array}
$$

Final Labels

We will also need a function which returns the set of final labels in a statement; whereas a sequence of statements has a single entry, it may have multiple exits (e.g. in the conditional):

$$\mathit{final} : \textbf{Stmt} \rightarrow \mathcal{P}(\textbf{Lab})$$

$$\begin{array}{rcl}
\mathit{final}([\, x := a \,]^{\ell}) & = & \{\ell\} \\
\mathit{final}([\, \textbf{skip} \,]^{\ell}) & = & \{\ell\} \\
\mathit{final}(S_1 ; S_2) & = & \mathit{final}(S_2) \\
\mathit{final}(\textbf{if } [b]^{\ell} \textbf{ then } S_1 \textbf{ else } S_2) & = & \mathit{final}(S_1) \cup \mathit{final}(S_2) \\
\mathit{final}(\textbf{while } [b]^{\ell} \textbf{ do } S) & = & \{\ell\}
\end{array}$$

The **while**-loop terminates immediately after the test fails.

Elementary Blocks

The building blocks of our analysis is given by **Block** is the set of statements, or elementary blocks, of the form:

- $[\, x := a \,]^{\ell}$, or
- $[\, \textbf{skip} \,]^{\ell}$, as well as
- tests of the form $[b]^{\ell}$.

Blocks

To access the statements or test associated with a label in a program we use the function

$$\mathit{blocks} : \mathbf{Stmt} \rightarrow \mathcal{P}(\mathbf{Block})$$

$$\begin{array}{rcl}
\mathit{blocks}([\ x := a\ ]^{\ell}) & = & \{[\ x := a\ ]^{\ell}\} \\
\mathit{blocks}([\ \mathbf{skip}\ ]^{\ell}) & = & \{[\ \mathbf{skip}\ ]^{\ell}\} \\
\mathit{blocks}(S_1 ; S_2) & = & \mathit{blocks}(S_1) \cup \mathit{blocks}(S_2) \\
\mathit{blocks}(\mathbf{if}\ [b]^{\ell}\ \mathbf{then}\ S_1\ \mathbf{else}\ S_2) & = & \{[b]^{\ell}\} \cup \\
 & & \mathit{blocks}(S_1) \cup \mathit{blocks}(S_2) \\
\mathit{blocks}(\mathbf{while}\ [b]^{\ell}\ \mathbf{do}\ S) & = & \{[b]^{\ell}\} \cup \mathit{blocks}(S)
\end{array}$$

Labels

Then the set of labels occurring in a program is given by

$$\mathit{labels} : \mathbf{Stmt} \rightarrow \mathcal{P}(\mathbf{Lab})$$

where

$$\mathit{labels}(S) = \{\ell \mid [B]^{\ell} \in \mathit{blocks}(S)\}$$

Clearly $\mathit{init}(S) \in \mathit{labels}(S)$ and $\mathit{final}(S) \subseteq \mathit{labels}(S)$.

Flow

$$\mathit{flow} : \mathbf{Stmt} \rightarrow \mathcal{P}(\mathbf{Lab} \times \mathbf{Lab})$$

which maps statements to sets of flows:

$$\begin{array}{rcl}
\mathit{flow}([\ x := a\ ]^\ell) & = & \emptyset \\
\mathit{flow}([\ \mathbf{skip}\ ]^\ell) & = & \emptyset \\
\mathit{flow}(S_1;S_2) & = & \mathit{flow}(S_1) \ \cup\ \mathit{flow}(S_2) \ \cup \\
 & & \{(\ell, \mathit{init}(S_2)) \mid \ell \in \mathit{final}(S_1)\} \\
\mathit{flow}(\mathbf{if}\ [b]^\ell\ \mathbf{then}\ S_1\ \mathbf{else}\ S_2) & = & \mathit{flow}(S_1) \cup \mathit{flow}(S_2) \ \cup \\
 & & \{(\ell, \mathit{init}(S_1)), (\ell, \mathit{init}(S_2))\} \\
\mathit{flow}(\mathbf{while}\ [b]^\ell\ \mathbf{do}\ S) & = & \mathit{flow}(S) \cup \{(\ell, \mathit{init}(S))\} \ \cup \\
 & & \{(\ell', \ell) \mid \ell' \in \mathit{final}(S)\}
\end{array}$$

An Example Flow

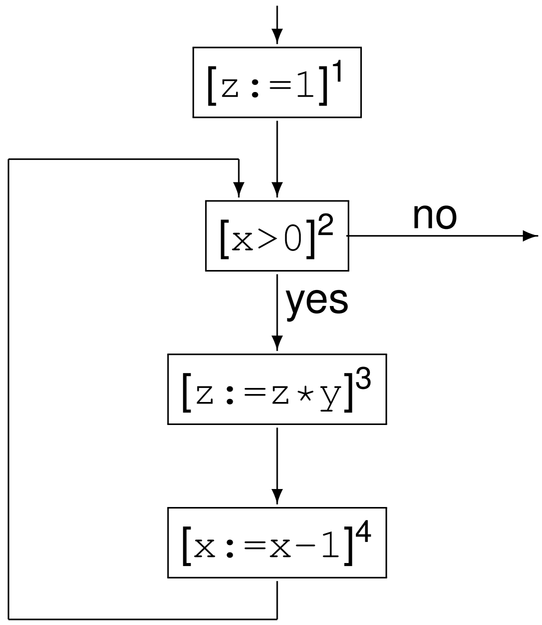
Consider the following program, power, computing the x-th power of the number stored in y:

$$\begin{array}{l}
[\ z := 1\ ]^1; \\
\mathbf{while}\ [x > 1]^2\ \mathbf{do}\ (\\
\qquad [\ z := z * y\ ]^3; \\
\qquad [\ x := x - 1\ ]^4);
\end{array}$$

We have $\mathit{labels}(\text{power}) = \{1, 2, 3, 4\}$, $\mathit{init}(\text{power}) = 1$, and $\mathit{final}(\text{power}) = \{2\}$. The function *flow* produces the set:

$$\mathit{flow}(\text{power}) = \{(1, 2), (2, 3), (3, 4), (4, 2)\}$$

Flow Graph



Forward Analysis

The function *flow* is used in the formulation of *forward analyses*. Clearly *init*(S) is the (unique) entry node for the flow graph with nodes *labels*(S) and edges *flow*(S). Also

$$\begin{aligned} \mathit{labels}(S) &= \{\mathit{init}(S)\} \cup \\ &\quad \{\ell \mid (\ell, \ell') \in \mathit{flow}(S)\} \cup \\ &\quad \{\ell' \mid (\ell, \ell') \in \mathit{flow}(S)\} \end{aligned}$$

and for composite statements (meaning those not simply of the form $[B]^\ell$) the equation remains true when removing the $\{\mathit{init}(S)\}$ component.

Reverse Flow

In order to formulate *backward analyses* we require a function that computes reverse flows:

$$\mathit{flow}^R : \mathbf{Stmt} \to \mathcal{P}(\mathbf{Lab} \times \mathbf{Lab})$$

$$\mathit{flow}^R(S) = \{(\ell, \ell') \mid (\ell', \ell) \in \mathit{flow}(S)\}$$

For the power program, flow^R produces

$$\{(2, 1), (2, 4), (3, 2), (4, 3)\}$$

Backward Analysis

In case $\mathit{final}(S)$ contains just one element that will be the unique entry node for the flow graph with nodes $\mathit{labels}(S)$ and edges $\mathit{flow}^R(S)$. Also

$$\begin{aligned}\mathit{labels}(S) \;=\; & \mathit{final}(S) \;\cup \\ & \{\ell \mid (\ell, \ell') \in \mathit{flow}^R(S)\} \;\cup \\ & \{\ell' \mid (\ell, \ell') \in \mathit{flow}^R(S)\}\end{aligned}$$

Notation

We will use the notation $S_\star$ to represent the program we are analysing (the “top-level" statement) and furthermore:

- **Lab**$_\star$ to represent the labels (*labels*$(S_\star)$) appearing in $S_\star$,
- **Var**$_\star$ to represent the variables ($FV(S_\star)$) appearing in $S_\star$,
- **Block**$_\star$ to represent the elementary blocks (*blocks*$(S_\star)$) occurring in $S_\star$, and
- **AExp**$_\star$ to represent the set of *non-trivial* arithmetic subexpressions in $S_\star$ as well as
- **AExp**(a) and **AExp**(b) to refer to the set of non-trivial arithmetic subexpressions of a given arithmetic, respectively boolean, expression.

An expression is trivial if it is a single variable or constant.

Isolated Entries & Exits

Program $S_\star$ has *isolated entries* if:

$$\forall \ell \in \textbf{Lab} : (\ell, init(S_\star)) \notin flow(S_\star)$$

This is the case whenever $S_\star$ does not start with a **while**-loop.

Similarly, we shall frequently assume that the program $S_\star$ has *isolated exits*; this means that:

$$\forall \ell_1 \in final(S_\star)\ \forall \ell_2 \in \textbf{Lab} : (\ell_1, \ell_2) \notin flow(S_\star)$$

Label Consistency

A statement, S, is label consistent if and only if:

$$[B_1]^\ell, [B_2]^\ell \in \mathit{blocks}(S) \text{ implies } B_1 = B_2$$

Clearly, if all blocks in S are uniquely labelled (meaning that each label occurs only once), then S is label consistent.

When S is label consistent the statement or clause "where $[B]^\ell \in \mathit{blocks}(S)$" is unambiguous in defining a partial function from labels to elementary blocks; we shall then say that ℓ labels the block B.