Concrete Semantics
with Isabelle/HOL

Tobias Nipkow

Fakultät für Informatik
Technische Universität München

2017-10-16
Part II

Semantics
Chapter 7

IMP:
A Simple Imperative Language
1 IMP Commands

2 Big-Step Semantics

3 Small-Step Semantics
1. IMP Commands

2. Big-Step Semantics

3. Small-Step Semantics
Terminology

Statement: declaration of fact or claim

Semantics is easy.

Command: order to do something

Study the book until you have understood it.

Expressions are evaluated, commands are executed
Concrete syntax:

\[
\text{com} \ ::= \ \text{SKIP} \\
\|
\text{string} \ ::= \ aexp \\
\|
\text{com} \ ;; \ \text{com} \\
\|
\text{IF } bexp \ \text{THEN } \text{com} \ \text{ELSE} \ \text{com} \\
\|
\text{WHILE } bexp \ \text{DO } \text{com}
\]
Commands

Abstract syntax:

\[
\text{datatype } \text{com} = \text{SKIP} \\
\text{ Assign string aexp } \\
\text{ Seq com com } \\
\text{ If bexp com com } \\
\text{ While bexp com }
\]
Com.thy
1 IMP Commands

2 Big-Step Semantics

3 Small-Step Semantics
Big-step semantics

Concrete syntax:

\[(\text{com}, \text{initial-state}) \Rightarrow \text{final-state}\]

Intended meaning of \((c, s) \Rightarrow t\):
Command \(c\) started in state \(s\) terminates in state \(t\)

“\(\Rightarrow\)” here not type!
Big-step rules

\[(SKIP, s) \Rightarrow s\]

\[(x ::= a, s) \Rightarrow s(x ::= aval a s)\]

\[
\frac{(c_1, s_1) \Rightarrow s_2 \quad (c_2, s_2) \Rightarrow s_3}{(c_1;; c_2, s_1) \Rightarrow s_3}\]


Big-step rules

\[
\begin{align*}
  & bval\ b\ s \quad (c_1,\ s) \Rightarrow t \\
  & \quad \frac{}{(IF\ b\ THEN\ c_1\ ELSE\ c_2,\ s) \Rightarrow t} \\
  \quad & \quad \frac{\neg bval\ b\ s \quad (c_2,\ s) \Rightarrow t}{(IF\ b\ THEN\ c_1\ ELSE\ c_2,\ s) \Rightarrow t}
\end{align*}
\]
Big-step rules

\[
\neg \text{bval } b \; s \\
\frac{\text{(WHILE } b \; \text{DO } c, \; s) \Rightarrow s}{(c, \; s_1) \Rightarrow s_2 \quad (\text{WHILE } b \; \text{DO } c, \; s_2) \Rightarrow s_3}
\]

\[
\frac{\text{bval } b \; s_1}{(\text{WHILE } b \; \text{DO } c, \; s_1) \Rightarrow s_3}
\]
Examples: derivation trees

\[ \vdash \]

\[ \vdash \]

where \( w = \) \( \text{WHILE} \ b \ DO \ c \)
\[ b = \text{NotEq} \ (V \ "x") \ (N \ 2) \]
\[ c = "x" ::= \text{Plus} \ (V \ "x") \ (N \ 1) \]
\[ s_i = s("x" ::= i) \]

\[ \text{NotEq} \ a_1 \ a_2 = \]
\[ \text{Not}(\text{And} \ (\text{Not}(\text{Less} \ a_1 \ a_2)) \ (\text{Not}(\text{Less} \ a_2 \ a_1))) \]
Logically speaking

\[(c, s) \Rightarrow t\]

is just infix syntax for

\[\textit{big\_step} \ (c,s) \ t\]

where

\[\textit{big\_step} :: \ com \times state \Rightarrow state \Rightarrow bool\]

is an inductively defined predicate.
Big_Step.thy

Semantics
Rule inversion

What can we deduce from

- \((\text{SKIP}, s) \Rightarrow t\) ?
- \((x ::= a, s) \Rightarrow t\) ?
- \((c_1;; c_2, s_1) \Rightarrow s_3\) ?
- \((\text{IF}\ b\ \text{THEN}\ c_1\ \text{ELSE}\ c_2, s) \Rightarrow t\) ?
- \((w, s) \Rightarrow t\) where \(w = \text{WHILE}\ b\ \text{DO}\ c\) ?
Automating rule inversion

Isabelle command \texttt{inductive\_cases} produces theorems that perform rule inversions automatically.
We reformulate the inverted rules. Example:

\[
\begin{align*}
(c_1, c_2, s_1) \Rightarrow s_3 \\
\exists s_2. (c_1, s_1) \Rightarrow s_2 \land (c_2, s_2) \Rightarrow s_3
\end{align*}
\]

is logically equivalent to

\[
\begin{align*}
(c_1, c_2, s_1) \Rightarrow s_3 \\
\land s_2. [(c_1, s_1) \Rightarrow s_2; (c_2, s_2) \Rightarrow s_3] \implies P
\end{align*}
\]

Replaces assm \((c_1, c_2, s_1) \Rightarrow s_3\) by two assms \((c_1, s_1) \Rightarrow s_2\) and \((c_2, s_2) \Rightarrow s_3\) (with a new fixed \(s_2\)).

No \(\exists\) and \(\land\)! 
The general format: elimination rules

\[
\begin{align*}
asm &\quad asm_1 \implies P \\
& \vdots \\
&\quad asm_n \implies P \\
\hline 
\implies & \\
\end{align*}
\]

(possibly with $\bigwedge x$ in front of the $asm_i \implies P$)

Reading:

To prove a goal $P$ with assumption $asm$, prove all $asm_i \implies P$

Example:

\[
\begin{align*}
F \lor G &\quad F \implies P \\
& \implies P \\
G &\implies P \\
\hline 
P & \\
\end{align*}
\]
Theorems with \textit{elim} attribute are used automatically by \textit{blast}, \textit{fastforce} and \textit{auto}

Can also be added locally, eg \texttt{(blast elim: ... )}

Variant: \textit{elim!} applies elim-rules eagerly.
Big_Step.thy

Rule inversion
Command equivalence

Two commands have the same input/output behaviour:

\[ c \sim c' \equiv (\forall s \ t. \ (c,s) \Rightarrow t \iff (c',s) \Rightarrow t) \]

Example

\[ w \sim w' \]

where

\[ w = \text{WHILE } b \ \text{DO } c \]
\[ w' = \text{IF } b \ \text{THEN } c; ;; \ w \ \text{ELSE SKIP} \]
Equivalence proof

\[(w, s) \Rightarrow t\]

\[\iff\]

\[\land \ bval \ b \ s \land (\exists s'. (c, s) \Rightarrow s' \land (w, s') \Rightarrow t)\]

\[\lor\]

\[\neg \ bval \ b \ s \land t = s\]

\[\iff\]

\[(w', s) \Rightarrow t\]

Using the rules and rule inversions for \(\Rightarrow\).
Big Step.thy

Command equivalence
Execution is deterministic

Any two executions of the same command in the same start state lead to the same final state:

\[(c, s) \Rightarrow t \implies (c, s) \Rightarrow t' \implies t = t'\]

Proof by rule induction, for arbitrary \(t'\).
Big_Step.thy

Execution is deterministic
The boon and bane of big steps

We cannot observe intermediate states/steps

Example problem:

\((c, s)\) does not terminate iff \(\not\exists t. \ (c, s) \Rightarrow t\) ?

Needs a formal notion of nontermination to prove it. Could be wrong if we have forgotten a \(\Rightarrow\) rule.
Big-step semantics cannot directly describe

- nonterminating computations,
- parallel computations.

We need a finer grained semantics!
1. IMP Commands

2. Big-Step Semantics

3. Small-Step Semantics
Small-step semantics

Concrete syntax:

\[(com, \text{state}) \rightarrow (com, \text{state})\]

Intended meaning of \((c, s) \rightarrow (c', s')\):

The first step in the execution of \(c\) in state \(s\) leaves a “remainder” command \(c'\) to be executed in state \(s'\).

Execution as finite or infinite reduction:

\[(c_1, s_1) \rightarrow (c_2, s_2) \rightarrow (c_3, s_3) \rightarrow \ldots\]
Terminology

- A pair \((c, s)\) is called a configuration.
- If \(cs \rightarrow cs'\) we say that \(cs\) reduces to \(cs'\).
- A configuration \(cs\) is final iff \(\not\exists cs'. cs \rightarrow cs'\).
The intention:

\((\text{SKIP}, s)\) is final

Why?

\textit{SKIP} is the empty program. Nothing more to be done.
Small-step rules

\[(x ::= a, \ s) \rightarrow (\text{SKIP}, \ s(x ::= \text{aval} \ a \ s))\]

\[(\text{SKIP};; \ c, \ s) \rightarrow (c, \ s)\]

\[
\begin{align*}
(c_1, s) & \rightarrow (c'_1, s') \\
(c_1; ; c_2, s) & \rightarrow (c'_1; ; c_2, s')
\end{align*}
\]
Small-step rules

\[
\begin{align*}
\text{bval} \ b \ s \\
(\text{IF} \ b \ \text{THEN} \ c_1 \ \text{ELSE} \ c_2, s) & \rightarrow (c_1, s) \\
\neg \ \text{bval} \ b \ s \\
(\text{IF} \ b \ \text{THEN} \ c_1 \ \text{ELSE} \ c_2, s) & \rightarrow (c_2, s)
\end{align*}
\]

\[
(\text{WHILE} \ b \ \text{DO} \ c, s) \rightarrow (\text{IF} \ b \ \text{THEN} \ c_{;;} \ \text{WHILE} \ b \ \text{DO} \ c \ \text{ELSE} \ \text{SKIP}, s)
\]

Fact \ (\text{SKIP}, s) \ is \ a \ final \ configuration.
Small-step examples

\(("z" ::= V "x";; "x" ::= V "y";; "y" ::= V "z", s) \rightarrow \ldots\)

where \(s = <"x" := 3, "y" := 7, "z" := 5>\).

\((w, s_0) \rightarrow \ldots\)

where \(w = \text{WHILE } b \text{ DO } c\)
\(b = \text{Less } (V "x") (N 1)\)
\(c = "x" ::= \text{Plus } (V "x") (N 1)\)
\(s_n = <"x" := n>\)
Small_Step.thy

Semantics
Are big and small-step semantics equivalent?
From $\Rightarrow$ to $\rightarrow^*$

**Theorem** $cs \Rightarrow t \implies cs \rightarrow^* (SKIP, t)$

Proof by rule induction (of course on $cs \Rightarrow t$)

In two cases a lemma is needed:

**Lemma**

$(c_1, s) \rightarrow^* (c_1', s') \implies (c_1;; c_2, s) \rightarrow^* (c_1';; c_2, s')$

Proof by rule induction.
From $\rightarrow^*$ to $\Rightarrow$

**Theorem** $c s \rightarrow^* (\text{SKIP}, t) \implies c s \Rightarrow t$

Proof by rule induction on $c s \rightarrow^* (\text{SKIP}, t)$. In the induction step a lemma is needed:

**Lemma** $c s \rightarrow c s' \implies c s' \Rightarrow t \implies c s \Rightarrow t$

Proof by rule induction on $c s \rightarrow c s'$. 
Equivalence

Corollary $cs \Rightarrow t \iff cs \rightarrow^* (SKIP, t)$
Small_Step.thy

Equivalence of big and small
Can execution stop prematurely?

That is, are there any final configs except \((\text{SKIP}, s)\)?

**Lemma** \(\text{final} \ (c, s) \implies c = \text{SKIP}\)

We prove the contrapositive

\[ c \neq \text{SKIP} \implies \neg \text{final}(c, s) \]

by induction on \(c\).

- **Case** \(c_1 ;; c_2\): by case distinction:
  - \(c_1 = \text{SKIP} \implies \neg \text{final} \ (c_1 ;; c_2, s)\)
  - \(c_1 \neq \text{SKIP} \implies \neg \text{final} \ (c_1, s) \) (by IH)
    \[ \implies \neg \text{final} \ (c_1 ;; c_2, s) \]

- Remaining cases: trivial or easy
By rule inversion: \((\text{SKIP}, s) \rightarrow ct \implies \text{False}\)

Together:

**Corollary** \(\text{final } (c, s) = (c = \text{SKIP})\)
Infinite executions

⇒ yields final state iff → terminates

Lemma \((\exists t. \ cs \Rightarrow t) = (\exists cs'. \ cs \rightarrow^* cs' \land \text{final} \ cs')\)

Proof: \((\exists t. \ cs \Rightarrow t)\)

\[= (\exists t. \ cs \rightarrow^* (\text{SKIP},t))\]
(by big = small)

\[= (\exists cs'. \ cs \rightarrow^* cs' \land \text{final} \ cs')\]
(by final = SKIP)

Equivalent:

⇒ does not yield final state iff → does not terminate
May versus Must

→ is deterministic:

**Lemma**  \( cs \to cs' \implies cs \to cs'' \implies cs'' = cs' \)

(Proof by rule induction)

Therefore: no difference between

- **may** terminate (there is a terminating \( \to \) path)
- **must** terminate (all \( \to \) paths terminate)

Therefore: \( \implies \) correctly reflects termination behaviour.

With nondeterminism: may have both \( cs \Rightarrow t \) and a nonterminating reduction \( cs \to cs' \to \ldots \)
Chapter 8

Compiler
4 Stack Machine

5 Compiler
Stack Machine

Compiler
Stack Machine

Instructions:

**datatype** \( instr = \)

- \( LOADI \ int \) : load value
- \( LOAD \ vname \) : load var
- \( ADD \) : add top of stack
- \( STORE \ vname \) : store var
- \( JMP \ int \) : jump
- \( JMPLESS \ int \) : jump if <
- \( JMPGE \ int \) : jump if ≥
Semantics

Type synonyms:

\[
\begin{align*}
\text{stack} & \quad = \quad \text{int list} \\
\text{config} & \quad = \quad \text{int} \times \text{state} \times \text{stack}
\end{align*}
\]

Execution of 1 instruction:

\[
iexec :: \text{instr} \Rightarrow \text{config} \Rightarrow \text{config}
\]
Instruction execution

\[
\text{\texttt{iexec instr}} (i, s, stk) =
\]

(\texttt{case instr of LOADI n \Rightarrow (i + 1, s, n \# stk)}
| \texttt{LOAD x \Rightarrow (i + 1, s, s x \# stk)}
| \texttt{ADD \Rightarrow (i + 1, s, (hd2 stk + hd stk) \# tl2 stk)}
| \texttt{STORE x \Rightarrow (i + 1, s(x := hd stk), tl stk)}
| \texttt{JMP n \Rightarrow (i + 1 + n, s, stk)}
| \texttt{JMPLESS n \Rightarrow}
  \quad (\texttt{if hd2 stk < hd stk then i + 1 + n else i + 1, s, tl2 stk})
| \texttt{JMPGE n \Rightarrow}
  \quad (\texttt{if hd stk \leq hd2 stk then i + 1 + n else i + 1, s, tl2 stk}))
Program execution (1 step)

Programs are instruction lists.

Executing one program step:

\[ \text{instr list} \vdash \text{config} \rightarrow \text{config} \]

\[
P \vdash \ c \rightarrow c' =
\]

\[
(\exists i \ s \ stk.
\]

\[
\begin{align*}
& \quad c = (i, s, stk) \land \\
& \quad c' = \text{iexec} (P \text{ !! } i) (i, s, stk) \land \\
& \quad 0 \leq i \land i < \text{size } P)
\end{align*}
\]

where

\[
'\text{a list !! int} = \text{nth instruction of list}
\]

\[
\text{size :: 'a list } \Rightarrow \text{int} = \text{list size as integer}
\]
Program execution (∗ steps)

Defined in the usual manner:

\[ P \vdash (pc, s, stk) \rightarrow^* (pc', s', stk') \]
Compiler.thy

Stack Machine
4 Stack Machine

5 Compiler
Compiling \( aexp \)

Same as before:

\[
\begin{align*}
    acomp (N n) &= [LOADI n] \\
    acomp (V x) &= [LOAD x] \\
    acomp (Plus a_1 a_2) &= acomp a_1 @ acomp a_2 @ [ADD]
\end{align*}
\]

Correctness theorem:

\[
\begin{align*}
    acomp a \\
    \vdash (0, s, stk) \rightarrow^* (size (acomp a), s, aval a s \neq stk)
\end{align*}
\]

Proof by induction on \( a \) (with arbitrary \( stk \)).

Needs lemmas!
$P \vdash c \rightarrow^* c' \implies P @ P' \vdash c \rightarrow^* c'$

$P \vdash (i, s, stk) \rightarrow^* (i', s', stk') \implies P' @ P \vdash (size P' + i, s, stk) \rightarrow^* (size P' + i', s', stk')$

Proofs by rule induction on $\rightarrow^*$, using the corresponding single step lemmas:

$P \vdash c \rightarrow c' \implies P @ P' \vdash c \rightarrow c'$

$P \vdash (i, s, stk) \rightarrow (i', s', stk') \implies P' @ P \vdash (size P' + i, s, stk) \rightarrow (size P' + i', s', stk')$

Proofs by cases.
Compiling $b_{exp}$

Let $ins$ be the compilation of $b$:

*Do not put value of $b$ on the stack but let value of $b$ determine where execution of $ins$ ends.*

Principle:

- Either execution leads to the end of $ins$
- or it jumps to offset $+n$ beyond $ins$.

Parameters: when to jump (if $b$ is True or False) where to jump to ($n$)

\[ b_{comp} :: b_{exp} \Rightarrow \text{bool} \Rightarrow \text{int} \Rightarrow \text{instr list} \]
Example

\[ b = \text{And} \ (\text{Less} \ (V \ "x") \ (V \ "y")) \\
\quad (\text{Not} \ (\text{Less} \ (V \ "z") \ (V \ "a"))). \]

\[ \text{bcomp} \ b \ False \ 3 = \]

\[ [\text{LOAD} \ "x", \]
\[ \text{LOAD} \ "y", \]
\[ \text{LOAD} \ "z", \]
\[ \text{LOAD} \ "a", \]
\[ ] \]
\textit{bcomp :: bexp \Rightarrow bool \Rightarrow int \Rightarrow instr list}

\begin{align*}
\text{bcomp } (Bc \ v) \ f \ n & = (\text{if } v = f \ \text{then} \ [\text{JMP } n] \ \text{else} \ []) \\
\text{bcomp } (\text{Not } b) \ f \ n & = \text{bcomp } b \ (\neg f) \ n \\
\text{bcomp } (\text{Less } a_1 \ a_2) \ f \ n & = \\
\text{acomp } a_1 @ \\
\text{acomp } a_2 @ (\text{if } f \ \text{then} \ [\text{JMPLESS } n] \ \text{else} \ [\text{JMPGE } n]) \\
\text{bcomp } (\text{And } b_1 \ b_2) \ f \ n & = \\
\text{let } cb_2 = \text{bcomp } b_2 \ f \ n; \\
& \quad m = \text{if } f \ \text{then} \ \text{size } cb_2 \ \text{else} \ \text{size } cb_2 + n; \\
& \quad cb_1 = \text{bcomp } b_1 \ False \ m \\
\text{in } cb_1 @ cb_2
\end{align*}
Correctness of $bcomp$

$0 \leq n \implies bcomp \ b \ f \ n$

$\vdash (0, \ s, \ stk) \rightarrow^* (size (bcomp \ b \ f \ n) + (if \ f = bval \ b \ s \ then \ n \ else \ 0), s, stk)$
Compiling \textit{com}

\[
\text{ccomp} :: \text{com} \rightarrow \text{instr list}
\]

\[
\text{ccomp} \ SKIP = []
\]

\[
\text{ccomp} \ (x ::= a) = \text{acomp} \ a \ @ \ [\text{STORE} \ x]
\]

\[
\text{ccomp} \ (c_1;\ c_2) = \text{ccomp} \ c_1 \ @ \ \text{ccomp} \ c_2
\]
\[ \text{ccomp} \left( \text{IF} \ b \ \text{THEN} \ c_1 \ \text{ELSE} \ c_2 \right) = \]

\[
\begin{align*}
\text{let} \ & \ c_{c_1} = \text{ccomp} \ c_1; \ c_{c_2} = \text{ccomp} \ c_2; \\
& \ cb = \text{bcomp} \ b \ False \ (\text{size} \ cc_{c_1} + 1) \\
\text{in} \ & \ cb \ @ \ cc_{c_1} \ @ \ \text{JMP} \ (\text{size} \ cc_{c_2}) \ # \ cc_{c_2}
\end{align*}
\]
\[
\text{ccomp (WHILE b DO c) } = \\
\text{let } cc = \text{ccomp } c; \ cb = \text{bcomp } b \ False \ (size \ cc + 1) \in \ cb \@ \ cc \@ [\text{JMP } (\neg (size \ cb + size \ cc + 1))] \\
\]
Correctness of $ccomp$

If the source code produces a certain result, so should the compiled code:

$$(c, s) \Rightarrow t \quad \Rightarrow$$

$ccomp\ c \vdash (0, s, stk) \rightarrow^* (\text{size} \ ccomp\ c, t, stk)$

Proof by rule induction.
The other direction

We have only shown “⇒”:

\[ \text{compiled code simulates source code.} \]

How about “⇐”:

\[ \text{source code simulates compiled code?} \]

If \( \text{ccomp} \ c \) with start state \( s \) produces result \( t \), and if(!) \((c, s) \rightarrow t'\), then “⇒” implies that \( \text{ccomp} \ c \) with start state \( s \) must also produce \( t' \) and thus \( t' = t \) (why?).

But we have \textit{not} ruled out this potential error:

\[ c \text{ does not terminate but } \text{ccomp} \ c \text{ does.} \]
The other direction

Two approaches:

- In the absence of nondeterminism:
  Prove that $ccomp$ preserves nontermination. A nice proof of this fact requires coinduction. Isabelle supports coinduction, this course avoids it.

- A direct proof: theory $Compiler2$

\[
ccomp c \vdash (0, s, stk) \rightarrow^* (\text{size} (ccomp c), t, stk') \implies (c, s) \Rightarrow t
\]
Chapter 9

Types
6 A Typed Version of IMP

7 Security Type Systems
6 A Typed Version of IMP

7 Security Type Systems
A Typed Version of IMP

Remarks on Type Systems

Typed IMP: Semantics
Typed IMP: Type System
Type Safety of Typed IMP
Why Types?

*To prevent mistakes, dummy!*
There are 3 kinds of types

**The Good**  Static types that *guarantee* absence of certain runtime faults.
Example: no memory access errors in Java.

**The Bad**  Static types that have mostly decorative value but do not guarantee anything at runtime.
Example: C, C++

**The Ugly**  Dynamic types that detect errors when it can be too late.
Example: “TypeError: . . .” in Python.
The ideal

Well-typed programs cannot go wrong.


The most influential slogan and one of the most influential papers in programming language theory.
What could go wrong?

1. Corruption of data
2. Null pointer exception
3. Nontermination
4. Run out of memory
5. Secret leaked
6. and many more . . .

There are type systems for *everything* (and more) but in practice (Java, C#) only 1 is covered.
A programming language is *type safe* if the execution of a well-typed program cannot lead to certain errors.

Java and the JVM have been *proved* to be type safe. (Note: Java exceptions are not errors!)
Correctness and completeness

Type soundness means that the type system is *sound/correct* w.r.t. the semantics:

*If the type system says yes, the semantics does not lead to an error.*

The semantics is the primary definition, the type system must be justified w.r.t. it.

How about *completeness*? Remember Rice:

*Nontrivial semantic properties of programs (e.g. termination) are undecidable.*

Hence there is no (decidable) type system that accepts *all* programs that have a certain semantic property.
Automatic analysis of semantic program properties is necessarily incomplete.
A Typed Version of IMP

Remarks on Type Systems

Typed IMP: Semantics

Typed IMP: Type System

Type Safety of Typed IMP
Arithmetic

Values:

\textbf{datatype} \quad val = Iv \text{ int} \mid Rv \text{ real}

The state:

\textit{state} = vname \rightarrow val

Arithmetic expressions:

\textbf{datatype} \quad aexp =

\begin{align*}
& Ic \text{ int} \mid Rc \text{ real} \mid V \text{ vname} \mid \text{ Plus } aexp \ aexp
\end{align*}
Why tagged values?

Because we want to detect if things “go wrong”.

What can go wrong? Adding integer and real!
No automatic coercions.

Does this mean any implementation of IMP also needs to tag values?
No! Compilers compile only well-typed programs, and well-typed programs do not need tags.

Tags are only used to detect certain errors and to prove that the type system avoids those errors.
Evaluation of $aexp$

Not recursive function but inductive predicate:

$taval :: aexp \Rightarrow state \Rightarrow val \Rightarrow bool$

$taval \ (Ic \ i) \ s \ (Iv \ i)$

$taval \ (Rc \ r) \ s \ (Rv \ r)$

$taval \ (V \ x) \ s \ (s \ x)$

\[
\begin{align*}
taval \ a_1 \ s \ (Iv \ i_1) & \quad taval \ a_2 \ s \ (Iv \ i_2) \\
& \quad \quad \Rightarrow taval \ (Plus \ a_1 \ a_2) \ s \ (Iv \ (i_1 + i_2))
\end{align*}
\]

\[
\begin{align*}
taval \ a_1 \ s \ (Rv \ r_1) & \quad taval \ a_2 \ s \ (Rv \ r_2) \\
& \quad \quad \Rightarrow taval \ (Plus \ a_1 \ a_2) \ s \ (Rv \ (r_1 + r_2))
\end{align*}
\]
Example: evaluation of \( \text{Plus} \left( V \ ''x'' \right) \ (Ic \ 1) \)

If \( s \ ''x'' = Iv \ i \):

\[
\text{taval} \left( V \ ''x'' \right) \ s \ (Iv \ i) \quad \text{taval} \ (Ic \ 1) \ s \ (Iv \ 1)
\]

\[
\text{taval} \ (\text{Plus} \left( V \ ''x'' \right) \ (Ic \ 1)) \ s \ (Iv(i + 1))
\]

If \( s \ ''x'' = Rv \ r \): then there is no value \( v \) such that

\[
\text{taval} \ (\text{Plus} \left( V \ ''x'' \right) \ (Ic \ 1)) \ s \ v.
\]
The functional alternative

\[ \text{taval} :: \text{aexp} \Rightarrow \text{state} \Rightarrow \text{val option} \]

Exercise!
Boolean expressions

Syntax as before. Semantics:

\[
\begin{align*}
\text{tbval} :: \text{bexp} \Rightarrow \text{state} \Rightarrow \text{bool} \Rightarrow \text{bool}
\end{align*}
\]

\[
\begin{align*}
\text{tbval (} Bc \ v \text{)} \ s \ v & \quad \text{tbval b s bv} \\
\text{tbval b} \text{1} \ s \ bv_1 & \quad \text{tbval b} \text{2} \ s \ bv_2 \\
\text{tbval (} \text{Not} \ b \text{)} \ s \ (\neg bv) & \quad \text{tbval (}\text{And} \ b_1 \ b_2\text{)} \ s \ (bv_1 \land bv_2) \\
\text{taval a} \text{1} \ s \ (Iv \ i_1) & \quad \text{taval a} \text{2} \ s \ (Iv \ i_2) \\
\text{tbval (} \text{Less} \ a_1 \ a_2 \text{)} \ s \ (i_1 < i_2) & \quad \text{tbval (}\text{Less} \ a_1 \ a_2\text{)} \ s \ (r_1 < r_2)
\end{align*}
\]
com: big or small steps?

We need to detect if things “go wrong”.

- **Big step semantics:**
  Cannot model error by absence of final state.
  Would confuse error and nontermination.
  Could introduce an extra error-element, e.g.
  \[
  \text{big_step} :: \text{com} \times \text{state} \Rightarrow \text{state option} \Rightarrow \text{bool}
  \]
  Complicates formalization.

- **Small step semantics:**
  \[
  \text{error} = \text{semantics gets stuck}
  \]
Small step semantics

\[ taval \ a \ s \ v \]
\[ (x ::= a, \ s) \rightarrow (\text{SKIP}, \ s(x ::= v)) \]

\[ tbval \ b \ s \ True \]
\[ (\text{IF} \ b \ \text{THEN} \ c_1 \ \text{ELSE} \ c_2, \ s) \rightarrow (c_1, \ s) \]

\[ tbval \ b \ s \ False \]
\[ (\text{IF} \ b \ \text{THEN} \ c_1 \ \text{ELSE} \ c_2, \ s) \rightarrow (c_2, \ s) \]

The other rules remain unchanged.
Example

Let \( c = ("x" ::= Plus (V "x") (Ic 1)) \).

- If \( s "x" = Iv i \):
  \[(c, s) \rightarrow (\text{SKIP}, s("x" := Iv (i + 1)))\]

- If \( s "x" = Rv r \):
  \[(c, s) \not\rightarrow\]
6 A Typed Version of IMP
Remarks on Type Systems
Typed IMP: Semantics
Typed IMP: Type System
Type Safety of Typed IMP
Type system

There are two types:

\textbf{datatype} \quad ty = \text{Ity} \mid \text{Rty}

What is the type of \texttt{Plus } (V "x") (V "y") ?

Depends on the type of "x" and "y"!

A \textit{type environment} maps variable names to their types:

\texttt{tyenv} = \texttt{vname} \Rightarrow \texttt{ty}

The type of an expression is always relative to a type environment \(\Gamma\). Standard notation:

\[ \Gamma \vdash e : \tau \]

Read: \textit{In the context of } \(\Gamma\), \(e\) has type \(\tau\)
The type of an $aexp$

$$
\Gamma \vdash a : \tau
$$

$$
tyenv \vdash aexp : ty
$$

The rules:

$$
\Gamma \vdash Ic \ i : Ity
$$

$$
\Gamma \vdash Rc \ r : Rty
$$

$$
\Gamma \vdash V \ x : \Gamma \ x
$$

$$
\begin{array}{c}
\Gamma \vdash a_1 : \tau \\
\Gamma \vdash a_2 : \tau
\end{array}
$$

$$
\Gamma \vdash Plus\ a_1\ a_2 : \tau
$$
Example

\[ \Gamma \vdash \text{Plus} \left( V \ "x" \right) \ (\text{Plus} \left( V \ "x" \right) \ (\text{Ic} \ 0)) : ? \]

where \( \Gamma "x" = Ity. \)
Well-typed \( bexp \)

Notation:

\[
\Gamma \vdash b \\
\text{tyenv} \vdash bexp
\]

Read: \emph{In context} \( \Gamma \), \( b \) \emph{is well-typed}. 
The rules:

\[ \Gamma \vdash Bc \ v \]
\[ \Gamma \vdash b \]
\[ \Gamma \vdash \text{Not} \ b \]
\[ \Gamma \vdash b_1 \quad \Gamma \vdash b_2 \]
\[ \Gamma \vdash \text{And} \ b_1 \ b_2 \]
\[ \Gamma \vdash a_1 : \tau \quad \Gamma \vdash a_2 : \tau \]
\[ \Gamma \vdash \text{Less} \ a_1 \ a_2 \]

Example: \[ \Gamma \vdash \text{Less} \ (Ic \ i) \ (Rc \ r) \] does not hold.
Well-typed commands

Notation:

\[ \Gamma \vdash c \]
\[ tyenv \vdash com \]

Read: *In context $\Gamma$, $c$ is well-typed.*
The rules:

\[
\Gamma \vdash SKIP \\
\frac{\Gamma \vdash a : \Gamma \ x}{\Gamma \vdash x ::= a}
\]

\[
\frac{\Gamma \vdash c_1 \quad \Gamma \vdash c_2}{\Gamma \vdash c_1 ;; c_2}
\]

\[
\frac{\Gamma \vdash b \quad \Gamma \vdash c_1 \quad \Gamma \vdash c_2}{\Gamma \vdash IF b THEN c_1 ELSE c_2}
\]

\[
\frac{\Gamma \vdash b \quad \Gamma \vdash c}{\Gamma \vdash WHILE b DO c}
\]
Syntax-directedness

All three sets of typing rules are syntax-directed:

- There is exactly one rule for each syntactic construct (\(SKIP\), ::=, \ldots).
- Well-typedness of a term \(C t_1 \ldots t_n\) depends only on the well-typedness of its subterms \(t_1, \ldots, t_n\).

A syntax-directed set of rules

- is executable by backchaining without backtracking and
- backchaining terminates and requires at most as many steps as the size of the term.
Syntax-directedness

The big-step semantics is not syntax-directed:

- more than one rule per construct and
- the execution of \textit{WHILE} depends on the execution of \textit{WHILE}.
6 A Typed Version of IMP
Remarks on Type Systems
Typed IMP: Semantics
Typed IMP: Type System
Type Safety of Typed IMP
Well-typed states

Even well-typed programs can get stuck . . .
. . . if they start in an unsuitable state.

Remember:

If \( s "x" = Ru r \)
then \(("x" ::= \text{Plus} (V "x") (Ic 1), s) \not\rightarrow\)

The state must be well-typed w.r.t. \( \Gamma \).
The type of a value:

\[
\begin{align*}
type (Iv i) &= Ity \\
type (Rv r) &= Rty
\end{align*}
\]

Well-typed state:

\[
\Gamma \vdash s \leftrightarrow (\forall x. \ type (s \ x) = \Gamma \ x)
\]
Type soundness

Reduction cannot get stuck:

- If everything is ok \( (\Gamma \vdash s, \Gamma \vdash c) \),
- and you take a finite number of steps,
- and you have not reached SKIP,
- then you can take one more step.

Follows from \textit{progress}:

- If everything is ok and you have not reached SKIP,
- then you can take one more step.

and \textit{preservation}:

- If everything is ok and you take a step,
- then everything is ok again.
The slogan

Progress \land Preservation \implies Type safety

Progress  Well-typed programs do not get stuck.
Preservation  Well-typedness is preserved by reduction.

Preservation: Well-typedness is an *invariant*. 
Progress:

\[ \Gamma \vdash c; \Gamma \vdash s; c \neq \text{SKIP} \] \implies \exists cs'. \ (c, s) \to cs'

Preservation:

\[ ((c, s) \to (c', s'); \Gamma \vdash c; \Gamma \vdash s) \implies \Gamma \vdash s' \]

\[ ((c, s) \to (c', s'); \Gamma \vdash c) \implies \Gamma \vdash c' \]

Type soundness:

\[ ((c, s) \to^\ast (c', s'); \Gamma \vdash c; \Gamma \vdash s; c' \neq \text{SKIP}) \]

\[ \implies \exists cs''. \ (c', s') \to cs'' \]
Progress:

$$[[\Gamma \vdash b; \Gamma \vdash s]] \Rightarrow \exists v. \ tbval b s v$$
Progress:

\[
[\Gamma \vdash a : \tau; \Gamma \vdash s] \implies \exists v. \ \text{taval} \ a \ s \ v
\]

Preservation:

\[
[\Gamma \vdash a : \tau; \text{taval} \ a \ s \ v; \Gamma \vdash s] \implies \text{type} \ v = \tau
\]
All proofs by rule induction.
Types.thy
The mantra

Type systems have a purpose:

*The static analysis of programs in order to predict their runtime behaviour.*

The correctness of the prediction must be provable.
6 A Typed Version of IMP

7 Security Type Systems
The aim:

*Ensure that programs protect private data like passwords, bank details, or medical records. There should be no information flow from private data into public channels.*

This is known as *information flow control.*
Language based security is an approach to information flow control where data flow analysis is used to determine whether a program is free of illicit information flows.

LBS guarantees confidentiality by program analysis, not by cryptography.

These analyses are often expressed as type systems.
Security levels

- Program variables have security/confidentiality levels.
- Security levels are partially ordered: \( l < l' \) means that \( l \) is less confidential than \( l' \).
- We identify security levels with \( \text{nat} \).
  Level 0 is public.
- Other popular choices for security levels:
  - only two levels, \( \text{high} \) and \( \text{low} \).
  - the set of security levels is a lattice.
Two kinds of illicit flows

Explicit: \( \text{low} := \text{high} \)

Implicit: \( \text{if high1} = \text{high2} \) \text{then low} := 1 \text{else low} := 0 \)
Noninterference

High variables do not interfere with low ones.

A variation of confidential input does not cause a variation of public output.

Program $c$ guarantees noninterference iff for all $s_1, s_2$:

If $s_1$ and $s_2$ agree on low variables (but may differ on high variables!), then the states resulting from executing $(c, s_1)$ and $(c, s_2)$ must also agree on low variables.
7 Security Type Systems

Secure IMP

A Security Type System
A Type System with Subsumption
A Bottom-Up Type System
Beyond
Security Levels

Security levels:

**type_synonym** \( \text{level} = \text{nat} \)

Every variable has a security level:

\( \text{sec} :: \text{vname} \rightarrow \text{level} \)

No definition is needed. Except for examples. Hence we define (arbitrarily)

\( \text{sec} \ x = \text{length} \ x \)
The security level of an expression is the maximal security level of any of its variables.

\[
\text{sec} :: \text{aexp} \Rightarrow \text{level}
\]

\[
\text{sec} \ (N \ n) = 0
\]

\[
\text{sec} \ (V \ x) = \text{sec} \ x
\]

\[
\text{sec} \ (\text{Plus} \ a \ b) = \text{max} \ (\text{sec} \ a) \ (\text{sec} \ b)
\]
Security Levels on $bexp$

$sec :: bexp \Rightarrow level$

$sec \ (Bc \ v) = 0$
$sec \ (Not \ b) = sec \ b$
$sec \ (And \ b_1 \ b_2) = max \ (sec \ b_1) \ (sec \ b_2)$
$sec \ (Less \ a \ b) = max \ (sec \ a) \ (sec \ b)$
Security Levels on States

Agreement of states up to a certain level:

\[ s_1 = s_2 \ (\leq l) \ \equiv \ \forall x. \ sec \ x \leq l \rightarrow s_1 \ x = s_2 \ x \]

\[ s_1 = s_2 \ (< l) \ \equiv \ \forall x. \ sec \ x < l \rightarrow s_1 \ x = s_2 \ x \]

Noninterference lemmas for expressions:

\[
\begin{align*}
& s_1 = s_2 \ (\leq l) \quad sec \ a \leq l \\
& \frac{}{aval \ a \ s_1 = aval \ a \ s_2}
\end{align*}
\]

\[
\begin{align*}
& s_1 = s_2 \ (\leq l) \quad sec \ b \leq l \\
& \frac{}{bval \ b \ s_1 = bval \ b \ s_2}
\end{align*}
\]
Security Type Systems

Secure IMP

A Security Type System

A Type System with Subsumption

A Bottom-Up Type System

Beyond
Security Type System

Explicit flows are easy. How to check for implicit flows:

Carry the security level of the boolean expressions around that guard the current command.

The well-typedness predicate:

\[ l \vdash c \]

Intended meaning:
“In the context of boolean expressions of level \( \leq l \), command \( c \) is well-typed.”

Hence:
“Assignments to variables of level \( < l \) are forbidden.”
Well-typed or not?

Let \( c = \) \( \text{IF } \text{Less} \ (V \ "x1") \ (V \ "x") \ \text{THEN} \ "x1" ::= N \ 0 \ \text{ELSE} \ "x1" ::= N \ 1 \)

\[
\begin{align*}
1 \vdash c \ ? & \quad \text{Yes} \\
2 \vdash c \ ? & \quad \text{Yes} \\
3 \vdash c \ ? & \quad \text{No}
\end{align*}
\]
The type system

\[ l \vdash SKIP \]

\[ \text{sec} \ a \leq \text{sec} \ x \quad l \leq \text{sec} \ x \]

\[ l \vdash x ::= a \]

\[ l \vdash c_1 \quad l \vdash c_2 \]

\[ l \vdash c_1 ;; c_2 \]

\[ \text{max} \ (\text{sec} \ b) \quad l \vdash c_1 \quad \text{max} \ (\text{sec} \ b) \quad l \vdash c_2 \]

\[ l \vdash \text{IF} \ b \ \text{THEN} \ c_1 \ \text{ELSE} \ c_2 \]

\[ \text{max} \ (\text{sec} \ b) \quad l \vdash c \]

\[ l \vdash \text{WHILE} \ b \ \text{DO} \ c \]
Remark:

\[ l \vdash c \] is syntax-directed and executable.
Anti-monotonicity

\[ l \vdash c \quad l' \leq l \]

\[ l' \vdash c \]

Proof by \ldots as usual.

This is often called a \textit{subsumption rule} because it says that larger levels subsume smaller ones.
Confinement

If \( l \vdash c \) then \( c \) cannot modify variables of level \( < l \):

\[
\frac{(c, s) \Rightarrow t}{s = t \ (\ < l)} \quad l \vdash c
\]

The effect of \( c \) is *confined* to variables of level \( \geq l \).

Proof by ... as usual.
Noninterference

\[(c, s) \Rightarrow s' \quad (c, t) \Rightarrow t' \quad 0 \vdash c \quad s = t \ (\leq l)\]

\[s' = t' \ (\leq l)\]

Proof by ... as usual.
7 Security Type Systems

Secure IMP
A Security Type System
A Type System with Subsumption
A Bottom-Up Type System
Beyond
The \( l \vdash c \) system is intuitive and executable

- but in the literature a more elegant formulation is dominant
- which does not need \( max \)
- and works for arbitrary partial orders.

This alternative system \( l \vdash' c \) has an explicit subsumption rule

\[
\frac{\vdash' c \quad l' \leq l}{\vdash' c}
\]

together with one rule per construct:
\[
\begin{align*}
l \vdash \text{'' SKIP} \\
sec a \leq sec x \quad l \leq sec x \quad \frac{}{l \vdash \text{'' } x ::= a} \\
l \vdash \text{'' } c_1 \quad l \vdash \text{'' } c_2 \\
\frac{}{l \vdash \text{'' } c_1 ;; c_2} \\
sec b \leq l \quad l \vdash \text{'' } c_1 \quad l \vdash \text{'' } c_2 \\
\frac{}{l \vdash \text{'' IF } b \text{ THEN } c_1 \text{ ELSE } c_2} \\
sec b \leq l \quad l \vdash \text{'' } c \\
\frac{}{l \vdash \text{'' WHILE } b \text{ DO } c}
\end{align*}
\]
• The subsumption-based system $\vdash'$ is neither syntax-directed nor directly executable.
• Need to guess when to use the subsumption rule.
Equivalence of $\vdash$ and $\vdash'$

$$l \vdash c \iff l \vdash' c$$

Proof by induction.
Use subsumption directly below $IF$ and $WHILE$.

$$l \vdash' c \iff l \vdash c$$

Proof by induction. Subsumption already a lemma for $\vdash$. 
7 Security Type Systems

Secure IMP
A Security Type System
A Type System with Subsumption
A Bottom-Up Type System
Beyond
Systems \( l \vdash c \) and \( l \vdash' c \) are *top-down*: level \( l \) comes from the context and is checked at ::= commands.

System \( \vdash c : l \) is *bottom-up*: \( l \) is the minimal level of any variable assigned in \( c \) and is checked at IF and WHILE commands.
\[\vdash \text{SKIP} : l\]

\[sec \ a \leq sec \ x\]
\[\vdash x ::= a : sec \ x\]

\[\vdash c_1 : l_1 \quad \vdash c_2 : l_2\]
\[\vdash c_1 ; c_2 : \text{min} \ l_1 \ l_2\]

\[sec \ b \leq \text{min} \ l_1 \ l_2\]
\[\vdash c_1 : l_1 \quad \vdash c_2 : l_2\]
\[\vdash \text{IF} \ b \ \text{THEN} \ c_1 \ \text{ELSE} \ c_2 : \text{min} \ l_1 \ l_2\]

\[sec \ b \leq l\]
\[\vdash c : l\]
\[\vdash \text{WHILE} \ b \ \text{DO} \ c : l\]
Equivalence of $\vdash :$ and $\vdash '$

$$\vdash c : l \implies l \vdash ' c$$

Proof by induction.

$$l \vdash ' c \implies \vdash c : l$$

Nitpick: $0 \vdash ' "x" ::= N 1 \quad \text{but not} \quad \vdash "x" ::= N 1 : 0$

$$l \vdash ' c \implies \exists l' \geq l. \vdash c : l'$$

Proof by induction.
7 Security Type Systems

Secure IMP
A Security Type System
A Type System with Subsumption
A Bottom-Up Type System
Beyond
Does noninterference really guarantee absence of information flow?

\[
(c, s) \Rightarrow s' \quad (c, t) \Rightarrow t' \quad 0 \vdash c \quad s = t \ (\leq l)
\]

\[
s' = t' \ (\leq l)
\]

Beware of covert channels!

\[
0 \vdash WHILE \ Less \ (V "x") \ (N 1) \ DO \ SKIP
\]
A drastic solution:

**WHILE**-conditions must not depend on confidential data.

**New typing rule:**

\[
\text{sec } b = 0 \quad 0 \vdash c \\
\hline
0 \vdash \text{WHILE } b \text{ DO } c
\]

**Now provable:**

\[
(c, s) \Rightarrow s' \quad 0 \vdash c \quad s = t \ (\leq l) \\
\hline
\exists t'. (c, t) \Rightarrow t' \land s' = t' \ (\leq l)
\]
Further extensions

• Time
• Probability
• Quantitative analysis
• More programming language features:
  • exceptions
  • concurrency
  • OO
  • …
The inventors of security type systems are Volpano and Smith.

Chapter 10

Data-Flow Analyses and Optimization
8 Definite Initialization Analysis

9 Live Variable Analysis
8 Definite Initialization Analysis

9 Live Variable Analysis
Each local variable must have a definitely assigned value when any access of its value occurs. A compiler must carry out a specific conservative flow analysis to make sure that, for every access of a local variable \( x \), \( x \) is definitely assigned before the access; otherwise a compile-time error must occur.

Java Language Specification

Java was the first language to force programmers to initialize their variables.
Examples: ok or not?

Assume $x$ is initialized:

IF $x < 1$ THEN $y := x$ ELSE $y := x + 1$;
y := y + 1

IF $x < x$ THEN $y := y + 1$ ELSE $y := x$

Assume $x$ and $y$ are initialized:

WHILE $x < y$ DO $z := x$; $z := z + 1$
Simplifying principle

We do not analyze boolean expressions to determine program execution.
8 Definite Initialization Analysis

Prelude: Variables in Expressions

Definite Initialization Analysis
Initialization Sensitive Semantics
Theory \textit{Vars} provides an overloaded function $\texttt{vars}$:

$$\begin{align*}
\texttt{vars} :: \texttt{aexp} & \Rightarrow \texttt{vname set} \\
\texttt{vars} (N n) & = \{\} \\
\texttt{vars} (V x) & = \{x\} \\
\texttt{vars} (\text{Plus} \ a_1 \ a_2) & = \texttt{vars} \ a_1 \cup \texttt{vars} \ a_2 \\
\texttt{vars} :: \texttt{bexp} & \Rightarrow \texttt{vname set} \\
\texttt{vars} (Bc v) & = \{\} \\
\texttt{vars} (\text{Not} \ b) & = \texttt{vars} \ b \\
\texttt{vars} (\text{And} \ b_1 \ b_2) & = \texttt{vars} \ b_1 \cup \texttt{vars} \ b_2 \\
\texttt{vars} (\text{Less} \ a_1 \ a_2) & = \texttt{vars} \ a_1 \cup \texttt{vars} \ a_2
\end{align*}$$
Vars.thy
Definite Initialization Analysis

Prelude: Variables in Expressions

Definite Initialization Analysis

Initialization Sensitive Semantics
Modified example from the JLS:

Variable \( x \) is definitely initialized after \( \text{SKIP} \)
iff \( x \) is definitely initialized before \( \text{SKIP} \).

Similar statements for each language construct.
\[ D :: \text{vname set} \Rightarrow \text{com} \Rightarrow \text{vname set} \Rightarrow \text{bool} \]

\[ D A c A' \] should imply:

*If all variables in \( A \) are initialized before \( c \) is executed, then no uninitialized variable is accessed during execution, and all variables in \( A' \) are initialized afterwards.*
\[ D A \ SKIP \ A \]

\[ \text{vars } a \subseteq A \]

\[ D A (x ::= a) \ (\text{insert } x \ A) \]

\[ D A_1 \ c_1 \ A_2 \ D A_2 \ c_2 \ A_3 \]

\[ D A_1 (c_1 ;; c_2) A_3 \]

\[ \text{vars } b \subseteq A \]

\[ D A \ c_1 \ A_1 \ D A \ c_2 \ A_2 \]

\[ D A (\text{IF } b \ \text{THEN } c_1 \ \text{ELSE } c_2) (A_1 \cap A_2) \]

\[ \text{vars } b \subseteq A \]

\[ D A \ c \ A' \]

\[ D A (\text{WHILE } b \ \text{DO } c) \ A \]
Correctness of $D$

- Things can go wrong: execution may access uninitialized variable.
  \[ \implies \text{We need a new, finer-grained semantics.} \]

- Big step semantics:
  semantics longer, correctness proof shorter

- Small step semantics:
  semantics shorter, correctness proof longer

For variety’s sake, we choose a big step semantics.
Definite Initialization Analysis

Prelude: Variables in Expressions
Definite Initialization Analysis
Initialization Sensitive Semantics
\[ \text{state} = \text{vname} \Rightarrow \text{val option} \]

where

\textbf{datatype} 'a option = None | Some 'a

Notation: \[ s(x \mapsto y) \text{ means } s(x := \text{Some } y) \]

Definition: \[ \text{dom } s = \{ a. s a \neq \text{None} \} \]
Expression evaluation

\[ \text{aval} :: \text{aexp} \Rightarrow \text{state} \Rightarrow \text{val option} \]

\[ \text{aval} (N \ i) \ s = \text{Some} \ i \]

\[ \text{aval} (V \ x) \ s = \ s \ x \]

\[ \text{aval} (\text{Plus} \ a_1 \ a_2) \ s = \]

\[ (\text{case} \ (\text{aval} \ a_1 \ s, \ \text{aval} \ a_2 \ s) \ of \]
\[ (\text{Some} \ i_1, \ \text{Some} \ i_2) \Rightarrow \text{Some}(i_1 + i_2) \]
\[ | \ \_ \Rightarrow \text{None}) \]
\textit{bval} :: \textit{bexp} \Rightarrow \textit{state} \Rightarrow \text{bool option}

\[\textit{bval} \ (Bc \ v) \ s = \text{Some} \ v\]

\[\textit{bval} \ (\text{Not} \ b) \ s =\]
\[
(\text{case} \ \textit{bval} \ b \ s \ \text{of} \ None \Rightarrow \text{None} \ |
\text{Some} \ bv \Rightarrow \text{Some} \ (\neg \ bv))
\]

\[\textit{bval} \ (\text{And} \ b_1 \ b_2) \ s =\]
\[
(\text{case} \ (\textit{bval} \ b_1 \ s, \textit{bval} \ b_2 \ s) \ \text{of} \\
\text{Some} \ bv_1, \text{Some} \ bv_2) \Rightarrow \text{Some}(bv_1 \land bv_2) \ |
_ \Rightarrow \text{None})
\]

\[\textit{bval} \ (\text{Less} \ a_1 \ a_2) \ s =\]
\[
(\text{case} \ (\textit{aval} \ a_1 \ s, \textit{aval} \ a_2 \ s) \ \text{of} \\
\text{Some} \ i_1, \text{Some} \ i_2) \Rightarrow \text{Some}(i_1 < i_2) \ |
_ \Rightarrow \text{None})\]
Big step semantics

\[(\text{com, state}) \Rightarrow \text{state option}\]

A small complication:

\[
\begin{align*}
(c_1, s_1) &\Rightarrow \text{Some } s_2 & (c_2, s_2) &\Rightarrow s \\
(c_1 ;; c_2, s_1) &\Rightarrow s \\
(c_1, s_1) &\Rightarrow \text{None} \\
(c_1 ;; c_2, s_1) &\Rightarrow \text{None}
\end{align*}
\]

More convenient, because compositional:

\[(\text{com, state option}) \Rightarrow \text{state option}\]
Error (*None*) propagates:

\[(c, \text{None}) \Rightarrow \text{None}\]

Execution starting in (mostly) normal states (*Some s*):

\[(\text{SKIP}, s) \Rightarrow s\]

\[\text{aval } a \ s = \text{Some } i\]

\[(x ::= a, \text{Some } s) \Rightarrow \text{Some } (s(x \mapsto i))\]

\[\text{aval } a \ s = \text{None}\]

\[(x ::= a, \text{Some } s) \Rightarrow \text{None}\]

\[(c_1, s_1) \Rightarrow s_2 \quad (c_2, s_2) \Rightarrow s_3\]

\[(c_1;\ c_2, s_1) \Rightarrow s_3\]
\[\text{bval } b \ s = \text{Some } \text{True} \quad (c_1, \text{Some } s) \Rightarrow s'\]

\[(\text{IF } b \ \text{THEN } c_1 \ \text{ELSE } c_2, \text{Some } s) \Rightarrow s'\]

\[\text{bval } b \ s = \text{Some } \text{False} \quad (c_2, \text{Some } s) \Rightarrow s'\]

\[(\text{IF } b \ \text{THEN } c_1 \ \text{ELSE } c_2, \text{Some } s) \Rightarrow s'\]

\[\text{bval } b \ s = \text{None} \quad (\text{IF } b \ \text{THEN } c_1 \ \text{ELSE } c_2, \text{Some } s) \Rightarrow \text{None}\]
\begin{align*}
\text{bval } b \ s &= \text{Some False} \\
\Rightarrow (\text{WHILE } b \text{ DO } c, \text{Some } s) &\Rightarrow \text{Some } s \\
\text{bval } b \ s &= \text{Some True} \\
(c, \text{Some } s) &\Rightarrow s' \quad (\text{WHILE } b \text{ DO } c, s') &\Rightarrow s'' \\
\Rightarrow (\text{WHILE } b \text{ DO } c, \text{Some } s) &\Rightarrow s'' \\
\text{bval } b \ s &= \text{None} \\
\Rightarrow (\text{WHILE } b \text{ DO } c, \text{Some } s) &\Rightarrow \text{None}
\end{align*}
Correctness of $D$ w.r.t. $\Rightarrow$

We want in the end:

Well-initialized programs cannot go wrong.

If $D(\text{dom } s) c A'$ and $(c, \text{Some } s) \Rightarrow s'$  
then $s' \neq \text{None}$.

We need to prove a generalized statement:

If $(c, \text{Some } s) \Rightarrow s'$ and $D A c A'$ and $A \subseteq \text{dom } s$  
then $\exists t. s' = \text{Some } t \land A' \subseteq \text{dom } t$.

By rule induction on $(c, \text{Some } s) \Rightarrow s'$.
Proof needs some easy lemmas:

\[\text{vars } a \subseteq \text{dom } s \implies \exists i. \text{aval } a \ s = \text{Some } i\]
\[\text{vars } b \subseteq \text{dom } s \implies \exists \text{bv}. \text{bval } b \ s = \text{Some } \text{bv}\]
\[D A c A' \implies A \subseteq A'\]
8 Definite Initialization Analysis

9 Live Variable Analysis
Motivation

Consider the following program:

\[
\begin{align*}
x & := y + 1; \\
y & := y + 2; \\
x & := y + 3
\end{align*}
\]

The first assignment is redundant and can be removed because \( x \) is dead at that point.
Semantically, a variable $x$ is live before command $c$ if the initial value of $x$ can influence the final state.

A weaker but easier to check condition:
We call $x$ live before $c$ if there is some potential execution of $c$ where $x$ is read before it can be overwritten.
Implicitly, every variable is read at the end of $c$.

Examples: Is $x$ initially dead or live?
- $x := 0$ 😞
- $y := x; y := 0; x := 0$ 😊
- WHILE $b$ DO $y := x; x := 1$ 😊
At the end of a command, we may be interested in the value of only some of the variables, e.g. only the global variables at the end of a procedure.

Then we say that \( x \) is live before \( c \) relative to the set of variables \( X \).
Liveness analysis

\[ L :: \text{com} \Rightarrow \text{vname set} \Rightarrow \text{vname set} \]

\[ L \ c \ X = \text{live before } c \text{ relative to } X \]

\[ L \ SKIP \ X = X \]
\[ L \ (x ::= a) \ X = \text{vars } a \cup (X - \{x\}) \]
\[ L \ (c_1;;c_2) \ X = L \ c_1 \ (L \ c_2 \ X) \]
\[ L \ (\text{IF } b \ \text{THEN } c_1 \ \text{ELSE } c_2) \ X = \]
\[ \text{vars } b \cup L \ c_1 \ X \cup L \ c_2 \ X \]

Example:

\[ L \ ("y" ::= V "z";;"x" ::= \text{Plus} (V "y") (V "z")) \]
\[ \{"x"\} = \{"z"\} \]
WHILE b DO c

$L \omega X$ must satisfy

- \text{vars } b \subseteq L \omega X \quad \text{(evaluation of } b)\)
- \text{X} \subseteq L \omega X \quad \text{(exit)}
- L \ c \ (L \omega X) \subseteq L \omega X \quad \text{(execution of } c)\)
We define

\[ L ( \text{WHILE } b \text{ DO } c) \ X = \text{vars } b \cup X \cup L c X \]

\[ \implies \]

\[ \text{vars } b \subseteq L w X \quad \checkmark \]
\[ X \subseteq L w X \quad \checkmark \]
\[ L c (L w X) \subseteq L w X \quad ? \]
\[ L \text{ SKIP } X = X \]
\[ L (x ::= a) X = \text{vars } a \cup (X - \{x\}) \]
\[ L (c_1 ;; c_2) X = L c_1 (L c_2 X) \]
\[ L (\text{IF } b \text{ THEN } c_1 \text{ ELSE } c_2) X = \text{vars } b \cup L c_1 X \cup L c_2 X \]
\[ L (\text{WHILE } b \text{ DO } c) X = \text{vars } b \cup X \cup L c X \]

Example:

\[ L (\text{WHILE } \text{Less } (V "x") (V "x") \text{ DO } "y" ::= V "z") \]
\[ \{"x"\} = \{"x","z"\} \]
Gen/kill analyses

A data-flow analysis $A :: \text{com} \Rightarrow \tau \text{ set} \Rightarrow \tau \text{ set}$
is called gen/kill analysis
if there are functions $\text{gen}$ and $\text{kill}$ such that

$$A c X = X - \text{kill} c \cup \text{gen} c$$

Gen/kill analyses are extremely well-behaved, e.g.

$$X_1 \subseteq X_2 \implies A c X_1 \subseteq A c X_2$$

$$A c (X_1 \cap X_2) = A c X_1 \cap A c X_2$$

Many standard data-flow analyses are gen/kill.
In particular liveness analysis.
**Liveness via gen/kill**

\[
\text{kill} :: \text{com} \Rightarrow \text{vname set}
\]

- \(\text{kill \ SKIP}\) = \(\{\}\)
- \(\text{kill \ (} x ::= a \text{)}\) = \(\{x\}\)
- \(\text{kill \ (} c_1 ;; c_2\text{)}\) = \(\text{kill } c_1 \cup \text{kill } c_2\)
- \(\text{kill \ (} \text{IF } b \text{ THEN } c_1 \text{ ELSE } c_2\text{)}\) = \(\text{kill } c_1 \cap \text{kill } c_2\)
- \(\text{kill \ (} \text{WHILE } b \text{ DO } c\text{)}\) = \(\{\}\)
\[ \text{gen :: com } \Rightarrow \text{ vname set} \]

\[ \text{gen } \text{SKIP} \quad = \quad \{\} \]

\[ \text{gen } (x ::= a) \quad = \quad \text{vars } a \]

\[ \text{gen } (c_1;; c_2) \quad = \quad \text{gen } c_1 \cup (\text{gen } c_2 - \text{kill } c_1) \]

\[ \text{gen } (\text{IF } b \text{ THEN } c_1 \text{ ELSE } c_2) \quad = \quad \text{vars } b \cup \text{gen } c_1 \cup \text{gen } c_2 \]

\[ \text{gen } (\text{WHILE } b \text{ DO } c) \quad = \quad \text{vars } b \cup \text{gen } c \]
Proof by induction on $c$.

$$L \ c \ X = \text{gen} \ c \cup (X - \text{kill} \ c)$$

$$L \ c \ (L \ w \ X) \subseteq L \ w \ X$$
Digression:
definite initialization via gen/kill

$A \ c \ X$: the set of variables initialized after $c$
  if $X$ was initialized before $c$

How to obtain $A \ c \ X = X - \text{kill } c \cup \text{gen } c$:

$\text{gen } \text{SKIP} = \{\}$

$\text{gen } (x ::= \ a)$
  $= \{x\}$

$\text{gen } (c_1 ;; c_2)$
  $= \text{gen } c_1 \cup \text{gen } c_2$

$\text{gen } (\text{IF } b \ \text{THEN } c_1 \ \text{ELSE } c_2)$
  $= \text{gen } c_1 \cap \text{gen } c_2$

$\text{gen } (\text{WHILE } b \ \text{DO } c)$
  $= \{\}$

$\text{kill } c = \{\}$
Live Variable Analysis

Correctness of $L$

Dead Variable Elimination

True Liveness

Comparisons
\((.,.) \Rightarrow \) . and \(L\) should roughly be related like this:

*The value of the final state on* \(X\)

*only depends on*

*the value of the initial state on* \(L \cap X\).

Put differently:

*If two initial states agree on* \(L \cap X\)

*then the corresponding final states agree on* \(X\).*
Equality on

An abbreviation:

\[ f = g \text{ on } X \equiv \forall x \in X. f(x) = g(x) \]

Two easy theorems (in theory \textit{Vars}):

\[ s_1 = s_2 \text{ on vars } a \implies \text{aval } a s_1 = \text{aval } a s_2 \]
\[ s_1 = s_2 \text{ on vars } b \implies \text{bval } b s_1 = \text{bval } b s_2 \]
Correctness of $L$

If $(c, s) \Rightarrow s'$ and $s = t$ on $L \ c \ X$ then $\exists \ t'. (c, t) \Rightarrow t' \land s' = t'$ on $X$.

Proof by rule induction.
For the two $WHILE$ cases we do not need the definition of $L \ w$ but only the characteristic property

$$\text{vars } b \cup X \cup L \ c \ (L \ w \ X) \subseteq L \ w \ X$$
Optimality of $L \ w$

The result of $L$ should be as small as possible: the more dead variables, the better (for program optimization).

$L \ w \ X$ should be the least set such that

\[ \text{vars } b \cup X \cup L \ c \ (L \ w \ X) \subseteq L \ w \ X. \]

Follows easily from $L \ c \ X = \text{gen } c \cup (X - \text{kill } c)$:

\[ \text{vars } b \cup X \cup L \ c \ P \subseteq P \implies L \ (\text{WHILE } b \ \text{DO } c) \ X \subseteq P \]
Live Variable Analysis
Correctness of $L$
Dead Variable Elimination
True Liveness
Comparisons
Bury all assignments to dead variables:

\( bury :: \text{com} \Rightarrow \text{uname set} \Rightarrow \text{com} \)

\[
\begin{align*}
\text{bury} \; \text{SKIP} \; X & = \; \text{SKIP} \\
\text{bury} \; (x ::= a) \; X & = \; \text{if} \; x \in X \; \text{then} \; x ::= a \; \text{else} \; \text{SKIP} \\
\text{bury} \; (c_1 ;; c_2) \; X & = \; \text{bury} \; c_1 \; (L \; c_2 \; X) ;; \; \text{bury} \; c_2 \; X \\
\text{bury} \; (\text{IF} \; b \; \text{THEN} \; c_1 \; \text{ELSE} \; c_2) \; X & = \\
\quad \text{IF} \; b \; \text{THEN} \; \text{bury} \; c_1 \; X \; \text{ELSE} \; \text{bury} \; c_2 \; X \\
\text{bury} \; (\text{WHILE} \; b \; \text{DO} \; c) \; X & = \\
\quad \text{WHILE} \; b \; \text{DO} \; \text{bury c} \; (L \; (\text{WHILE} \; b \; \text{DO} \; c) \; X)
\end{align*}
\]
Correctness of $bury$

$bury \ c \ UNIV \sim \ c$

where $UNIV$ is the set of all variables.

The two directions need to be proved separately.
\[(c, s) \Rightarrow s' \iff (\text{bury } c \text{ UNIV}, s) \Rightarrow s'\]

Follows from generalized statement:

\text{If } (c, s) \Rightarrow s' \text{ and } s = t \text{ on } L c X \text{ then } \exists t'. (\text{bury } c X, t) \Rightarrow t' \land s' = t' \text{ on } X.

Proof by rule induction, like for correctness of \( L \).
\[(bury \ c \ UNIV, \ s) \Rightarrow s' \quad \Rightarrow \quad (c, \ s) \Rightarrow s'\]

Follows from generalized statement:

\[
\text{If } (bury \ c \ X, \ s) \Rightarrow s' \quad \text{and} \quad s = t \quad \text{on} \quad L \ c \ X
\quad \text{then} \quad \exists \ t'. \ (c, \ t) \Rightarrow t' \quad \land \quad s' = t' \quad \text{on} \quad X.
\]

Proof very similar to other direction, but needs inversion lemmas for \textit{bury} for every kind of command, e.g.

\[
(bc_1 \;;; \ bc_2 = bury \ c \ X) =
\]

\[
(\exists \ c_1 \ c_2.
\quad c = c_1 \;;; \ c_2 \quad \land
\quad bc_2 = bury \ c_2 \ X \quad \land \quad bc_1 = bury \ c_1 \ (L \ c_2 \ X))
\]
Live Variable Analysis

Correctness of $L$

Dead Variable Elimination

True Liveness

Comparisons
Let $f :: \tau \Rightarrow \tau$ and $x :: \tau$.

If $f \ x = x$ then $x$ is a fixpoint of $f$.

Let $\leq$ be a partial order on $\tau$, eg $\subseteq$ on sets.

If $f \ x \leq x$ then $x$ is a pre-fixpoint (pfp) of $f$.

If $x \leq y \implies f \ x \leq f \ y$ for all $x,y$, then $f$ is monotone.
Application to $L \ w$

Remember the specification of $L \ w$:

$$\text{vars } b \cup X \cup L \ c \ (L \ w \ X) \subseteq L \ w \ X$$

This is the same as saying that $L \ w \ X$ should be a pfp of

$$\lambda P. \ \text{vars } b \cup X \cup L \ c \ P$$

and in particular of $L \ c$. 
True liveness

\[ L \left( "x" ::= V "y" \right) \{\} = \{ "y" \} \]

But "y" is not truly live: it is assigned to a dead variable.

Problem:

\[ L \left( x ::= a \right) X = \text{vars } a \cup (X - \{x\}) \]

Better:

\[
L \left( x ::= a \right) X = \\
(\text{if } x \in X \text{ then vars } a \cup (X - \{x\}) \text{ else } X)
\]

But then

\[ L \left( \text{WHILE } b \text{ DO } c \right) X = \text{vars } b \cup X \cup L \ c \ X \]

is not correct anymore.
\( L(x ::= a)\ X = \)

\((\text{if} \ x \in X \ \text{then} \ \text{vars} \ a \cup (X - \{x\}) \ \text{else} \ X)\)

\( L(\text{WHILE} \ b \ \text{DO} \ c)\ X = \text{vars} \ b \cup X \cup L\ c\ X \)

Let \( w = \text{WHILE} \ b \ \text{DO} \ c \)

where \( b = Less\ (N\ 0)\ (V\ y) \)

and \( c = y ::= V\ x;\ x ::= V\ z \)

and \( \text{distinct} \ [x,\ y,\ z] \)

Then \( L\ w\ \{y\} = \{x,\ y\}, \) but \( z \) is live before \( w \)!

\( \{x\}\ y ::= V\ x \ \{y\}\ x ::= V\ z \ \{y\} \)

\( \implies L\ w\ \{y\} = \{y\} \cup \{y\} \cup \{x\} \)
\begin{align*}
b & = \text{Less} (N \, 0) \, (V \, y) \\
c & = y ::= \ V \ x; \ x ::= \ V \ z \\
L \ w \ {\{y\}} & = \{x, \ y\} \text{ is not a pfp of } L \ c: \\
\{x, \ z\} & y ::= \ V \ x \ {\{y, \ z\}} \ x ::= \ V \ z \ {\{x, \ y\}} \\
L \ c \ {\{x, \ y\}} & = \{x, \ z\} \not\subseteq \{x, \ y\}
\end{align*}
Define $L \ w X$ as the least pfp of $
abla P. \ \vars b \cup X \cup L c P$
Existence of least fixpoints

**Theorem** (Knaster-Tarski) Let $f :: \tau \text{ set} \Rightarrow \tau \text{ set}$. If $f$ is monotone ($X \subseteq Y \implies f(X) \subseteq f(Y)$) then

$$lfp(f) := \bigcap \{P \mid f(P) \subseteq P\}$$

is the least pre-fixpoint and least fixpoint of $f$. 
Proof of Knaster-Tarski

**Theorem** If $f : \tau \text{ set} \Rightarrow \tau \text{ set}$ is monotone then $lfp(f) := \bigcap \{P \mid f(P) \subseteq P\}$ is the least pre-fixpoint.

**Proof**
- $f(lfp\ f) \subseteq lfp\ f$
- $lfp\ f$ is the least pre-fixpoint of $f$

**Lemma** Let $f$ be a monotone function on a partial order $\leq$. Then a least pre-fixpoint of $f$ is also a least fixpoint.

**Proof**
- $fp \leq p \implies fp = p$
- $p$ is the least fixpoint
Definition of $L$

$L \ (x ::= a) \ X =$
$(\text{if } x \in X \text{ then } \text{vars } a \cup (X - \{x\}) \text{ else } X)$

$L \ (\text{WHILE } b \ DO \ c) \ X = \text{lfp } f_w$
where $f_w = (\lambda P. \text{vars } b \cup X \cup L \ c \ P)$

Lemma $L \ c$ is monotone.

Proof by induction on $c$ using that $\text{lfp}$ is monotone:
$lfp \ f \subseteq lfp \ g$ if for all $X$, $f \ X \subseteq g \ X$

Corollary $f_w$ is monotone.
Computation of $lfp$

**Theorem** Let $f : \tau \text{ set} \Rightarrow \tau \text{ set}$. If

- $f$ is monotone: $X \subseteq Y \implies f(X) \subseteq f(Y)$
- and the chain $\{\} \subseteq f(\{\}) \subseteq f(f(\{\})) \subseteq \ldots$ stabilizes after a finite number of steps, i.e. $f^{k+1}(\{\}) = f^k(\{\})$ for some $k$,

then $lfp(f) = f^k(\{\})$.

**Proof** Show $f^i(\{\}) \subseteq p$ for any pfp $p$ of $f$ (by induction on $i$).
Computation of \( \text{lfp} \ f_w \)

\[ f_w = (\lambda P. \text{vars} \ b \cup X \cup L \ c \ P) \]

The chain \( \{\} \subseteq f_w \{\} \subseteq f_w^2 \{\} \subseteq \ldots \) must stabilize:

Let \( \text{vars} \ c \) be the variables in \( c \).

**Lemma** \( L \ c \ X \subseteq \text{vars} \ c \cup X \)

**Proof** by induction on \( c \)

Let \( V_w = \text{vars} \ b \cup \text{vars} \ c \cup X \)

**Corollary** \( P \subseteq V_w \implies f_w \ P \subseteq V_w \)

Hence \( f_w^k \{\} \) stabilizes for some \( k \leq |V_w| \).

More precisely: \( k \leq |\text{vars} \ c| + 1 \)

because \( f_w \{\} \supseteq \text{vars} \ b \cup X \).
Example

Let \( w = \text{WHILE } b \text{ DO } c \)
where \( b = \text{Less } (N \, 0) \, (V \, y) \)
and \( c = y ::= V \, x; \, x ::= V \, z \)

To compute \( L_w \{ y \} \) we iterate \( f_w \, P = \{ y \} \cup L_c \, P \):

\[
\begin{align*}
f_w \{ \} &= \{ y \} \cup L_c \{ \} = \{ y \} : \\
\{ \} \quad y ::= V \, x \quad \{ \} \quad x ::= V \, z \quad \{ \} \\
f_w \{ y \} &= \{ y \} \cup L_c \{ y \} = \{ x, \, y \} : \\
\{ x \} \quad y ::= V \, x \quad \{ y \} \quad x ::= V \, z \quad \{ y \} \\
f_w \{ x, \, y \} &= \{ y \} \cup L_c \{ x, y \} = \{ x, \, y, \, z \} : \\
\{ x, \, z \} \quad y ::= V \, x \quad \{ y, \, z \} \quad x ::= V \, z \quad \{ x, \, y \}
\end{align*}
\]
Computation of $lfp$ in Isabelle

From the library theory While_Combinator:

\[
\text{while} :: (\tau 
\Rightarrow \text{bool}) \Rightarrow (\tau 
\Rightarrow \tau) \Rightarrow \tau 
\Rightarrow \tau \\
\text{while } b \ f \ s = (\text{if } b \ s \text{ then while } b \ f \ (f \ s) \text{ else } s)
\]

**Lemma** Let \( f :: \tau \ set \Rightarrow \tau \ set \). If

- \( f \) is monotone: \( X \subseteq Y \implies f(X) \subseteq f(Y) \)
- and bounded by some finite set \( C \): \( X \subseteq C \implies f X \subseteq C \)

then \( lfp \ f = \text{while } (\lambda X. f X \neq X) \ f \ \{ \} \)
Limiting the number of iterations

Fix some small $k$ (eg 2) and define $Lb$ like $L$ except

$$Lb \ w \ X = \begin{cases} g^i_w \ {} & \text{if } g^{i+1}_w {} = g^i_w {} \text{ for some } i < k \\ V_w & \text{otherwise} \end{cases}$$

where $g_w \ P = vars \ b \cup X \cup Lb \ c \ P$

**Theorem** $L \ c \ X \subseteq Lb \ c \ X$

**Proof** by induction on $c$. In the **WHILE** case:

If $Lb \ w \ X = g^i_w {}$: $\forall P. \ L \ c \ P \subseteq Lb \ c \ P \ (IH) \implies \\
\forall P. \ f_w \ P \subseteq g_w \ P \implies f_w(g^i_w {}) = g_w (g^i_w {}) = g^i_w {} \implies \ L \ w \ X = lfp f_w \subseteq g^i_w {} = Lb \ w \ X$

If $Lb \ w \ X = V_w$: $L \ w \ X \subseteq V_w$ (by **Lemma**)
9 Live Variable Analysis

Correctness of $L$

Dead Variable Elimination

True Liveness

Comparisons
Comparison of analyses

• Definite initialization analysis is a *forward must analysis*:
  • it analyses the executions starting from some point,
  • variables *must* be assigned (on every program path) before they are used.

• Live variable analysis is a *backward may analysis*:
  • it analyses the executions ending in some point,
  • live variables *may* be used (on some program path) before they are assigned.
Comparison of DFA frameworks

Program representation:

- Traditionally (e.g. Aho/Sethi/Ullman), DFA is performed on control flow graphs (CFGs). Application: optimization of intermediate or low-level code.
- We analyse structured programs. Application: source-level program optimization.
Chapter 11

Denotational Semantics
A Relational Denotational Semantics of IMP

Continuity
A denotational semantics maps syntax to semantics:

\[ D : \text{syntax} \Rightarrow \text{meaning} \]

Examples:

\[ \text{aval} : \text{aexp} \Rightarrow (\text{state} \Rightarrow \text{val}) \]

\[ \text{Big\_step} : \text{com} \Rightarrow (\text{state} \times \text{state}) \text{ set} \]

\( D \) must be defined by primitive recursion over the syntax

\[ D (C \ t_1 \ldots \ t_n) = \ldots (D \ t_1) \ldots (D \ t_n) \ldots \]

Fake:

\[ \text{Big\_step} \ c = \{ (s,t). \ (c,s) \Rightarrow t \} \]
Why?

More abstract:
- operational: How to execute it
- denotational: What does it mean

Simpler proof principles:
- operational: relational, rule induction
- denotational: equational, structural induction
A Relational Denotational Semantics of IMP

Continuity
Relations

\[ Id :: (\texttt{'}a \times \texttt{'}a) \text{ set} \]
\[ Id = \{ p. \exists x. p = (x, x) \} \]

\[ op O :: (\texttt{'}a \times \texttt{'}b) \text{ set} \Rightarrow (\texttt{'}b \times \texttt{'}c) \text{ set} \Rightarrow (\texttt{'}a \times \texttt{'}c) \text{ set} \]
\[ r \ O \ s = \{(x, z). \exists y. (x, y) \in r \land (y, z) \in s\} \]
**D :: com ⇒ com_den**

**type_synonym** \(\text{com}_\text{den} = (\text{state} \times \text{state})\text{ set} \)

\[ D \text{ SKIP} = \text{Id} \]

\[ D \ (x ::= a) = \{(s, t). \ t = s(x ::= \text{aval} \ a \ s)\} \]

\[ D \ (c_1;; c_2) = D \ c_1 \ O \ D \ c_2 \]

\[ D (IF \ b \ THEN \ c_1 \ ELSE \ c_2) = \{(s, t). \ \text{if} \ \text{bval} \ b \ s \ \text{then} \ (s, t) \in D \ c_1 \ \text{else} \ (s, t) \in D \ c_2\}\]
Example

Let
\[ \begin{align*}
  c_1 &= "x" ::= N \ 0 \\
  c_2 &= "y" ::= V "x":
\end{align*} \]

\[
D \ c_1 = \{(s_1, s_2). \ s_2 = s_1("x" := 0)\}
\]

\[
D \ c_2 = \{(s_2, s_3). \ s_3 = s_2("y" := s_2 "x")\}
\]

\[
D \ (c_1;;c_2) = \{(s_1, s_3). \ s_3 = s_1("x" := 0, "y" := 0)\}
\]
$D \ (WHILE \ b \ DO \ c) = ?$

Wanted:

$D \ w = \\
\{(s, \ t). \ if \ bval \ b \ s \ then \ (s, \ t) \in \ D \ c \ O \ D \ w \ else \ s = t\} \$

Problem: not a denotational definition
not allowed by Isabelle

But $D \ w$ should be a solution of the equation.

General principle:

$x$ is a solution of $x = f(x) \iff x$ is a fixpoint of $f$

Define $D \ w$ as the least fixpoint of a suitable $f$
$D \, w = \\
\{(s, t). \, \text{if} \, bval \, b \, s \, \text{then} \, (s, t) \in D \, c \, O \, D \, w \, \text{else} \, s = t\}$

$W :: (state \Rightarrow bool) \Rightarrow com\_den \Rightarrow (com\_den \Rightarrow com\_den)$

$W\, \text{db}\, \text{dc} = \\
(\lambda dw. \, \{(s, t). \, \text{if} \, \text{db} \, s \, \text{then} \, (s, t) \in \text{dc} \, O \, dw \, \text{else} \, s = t\})$

**Lemma**  $W\, \text{db}\, \text{dc}$ is monotone.
We define

\[ D \ (\text{WHILE} \ b \ \text{DO} \ c) = \text{lfp} \ (W \ (bval \ b) \ (D \ c)) \]

By definition (where \( f = W \ (bval \ b) \ (D \ c) \)):

\[ D \ w = \text{lfp} \ f = f \ (\text{lfp} \ f) = W \ (bval \ v) \ (D \ c) \ (D \ w) \]

\[ = \{(s, t). \ \text{if} \ bval \ b \ s \ \text{then} \ (s, t) \in D \ c \ \text{O} \ D \ w \ \text{else} \ s = t\} \]
Why least?

Formally: needed for equivalence proof with big-step. An intuitive example:

\[ w = \text{WHILE } Bc \ True \text{ DO SKIP} \]

Then

\[ W (bval (Bc True)) (D \ SKIP) \]
\[ = W (\lambda s. True) \ Id \]
\[ = \lambda dw. \{(s, t). (s, t) \in Id \ O \ dw\} \]
\[ = \lambda dw. \ dw \]

Every relation is a fixpoint!
Only the least relation \{\} makes computational sense.
A denotational equivalence proof

Example

\[ D \, w = D \,(IF \, b \, THEN \, c;; \, w \, ELSE \, SKIP) \]

where \( w = WHILE \, b \, DO \, c. \)

Let \( f = W \,(bval \, b) \,(D \, c): \)

\[ D \, w \]

\[ = \{ (s, \, t). \, if \, bval \, b \, s \, then \, (s, \, t) \, \in \, D \, c \, O \, D \, w \, else \, s = t \} \]

\[ = D \,(IF \, b \, THEN \, c;; \, w \, ELSE \, SKIP) \]
Equivalence of denotational and big-step semantics

Lemma \((c, s) \Rightarrow t \iff (s, t) \in D_c\)

Proof by rule induction

Lemma \((s, t) \in D_c \iff (s, t) \in Big\_step\_c\)

Proof by induction on \(c\)

Corollary \((s, t) \in D_c \iff (c, s) \Rightarrow t\)
A Relational Denotational Semantics of IMP

Continuity
Chains and continuity

Definition
chain :: (nat ⇒ 'a set) ⇒ bool
chain S = (∀ i. S i ⊆ S (Suc i))

Definition (Continuous)
cont :: ('a set ⇒ 'b set) ⇒ bool
cont f = (∀ S. chain S → f (∪ n S n) = (∪ n f (S n)))

Lemma cont f → mono f
Kleene fixpoint theorem

**Theorem** \( \text{cont } f \iff \text{lfp } f = \left( \bigcup_n f^n \{\} \right) \)
Application to semantics

**Lemma**  $W \, db \, dc$ is continuous.

**Example**

WHILE $x \neq 0$ DO $x := x - 1$

Semantics: $\{(s,t). \ 0 \leq s \ "x" \land t = s("x" := 0)\}$

Let $f = W \, db \, dc$

where $db = bval \ b = (\lambda s. \ s \ "x" \neq 0)$

$dc = D \ c = \{(s, t). \ t = s("x" := s \ "x" - 1)\}$
A proof of determinism

\[
\text{single\_valued } r = \\
(\forall x y z. (x, y) \in r \land (x, z) \in r \rightarrow y = z)
\]

**Lemma** If \( f :: \text{com\_den} \Rightarrow \text{com\_den} \) is continuous and preserves single-valuedness then \( \text{lfp } f \) is single-valued.

**Lemma** \text{single\_valued } (D c)
Chapter 12
Hoare Logic
Partial Correctness

Verification Conditions

Total Correctness
12 Partial Correctness

13 Verification Conditions

14 Total Correctness
Partial Correctness

Introduction

The Syntactic Approach
The Semantic Approach
Soundness and Completeness
We have proved functional programs correct (e.g. a compiler).

We have proved properties of imperative languages (e.g. type safety).

But how do we prove properties of imperative programs?
An example program:

"y" ::= N 0;; wsum

where

\[ wsum \equiv \text{WHILE Less (N 0) (V "x") DO ("y" ::= Plus (V "y") (V "x"));; "x" ::= Plus (V "x") (N (- 1))) } \]

At the end of the execution of "y" ::= N 0;; wsum variable "y" should contain the sum 1 + ... + i where i is the initial value of "x".

\[ \text{sum } i = (if i \leq 0 \text{ then } 0 \text{ else } \text{sum } (i - 1) + i) \]
A proof via operational semantics

Theorem:

"y" ::= N 0;; wsum, s) ⇒ t → t "y" = sum (s "x")

Required Lemma:

(wsum, s) ⇒ t → t "y" = s "y" + sum (s "x")

Proved by rule induction.
Hoare Logic provides a structured approach for reasoning about properties of states during program execution:

- Rules of Hoare Logic (almost) syntax directed
- Automates reasoning about program execution
- No explicit induction

But no free lunch:

- Must prove implications between predicates on states
- Needs invariants.
Partial Correctness

Introduction
The Syntactic Approach
The Semantic Approach
Soundness and Completeness
This is the standard approach. Formulas are syntactic objects. Everything is very concrete and simple. But complex to formalize. Hence we soon move to a semantic view of formulas. Reason for introduction of syntactic approach: didactic

For now, we work with a (syntactically) simplified version of IMP.
Hoare Logic reasons about *Hoare triples* \( \{ P \} \ c \ \{ Q \} \) where

- \( P \) and \( Q \) are *syntactic formulas* involving program variables
- \( P \) is the *precondition*, \( Q \) is the *postcondition*
- \( \{ P \} \ c \ \{ Q \} \) means that if \( P \) is true at the start of the execution, \( Q \) is true at the end of the execution — if the execution terminates! (*partial correctness*)

Informal example:

\[
\{ x = 41 \} \ x := x + 1 \ \{ x = 42 \}
\]

Terminology: \( P \) and \( Q \) are called *assertions*. 
Examples

\{x = 5\} ? \{x = 10\}
\{True\} ? \{x = 10\}
\{x = y\} ? \{x \neq y\}

Boundary cases:
\{True\} ? \{True\}
\{True\} ? \{False\}
\{False\} ? \{Q\}
The rules of Hoare Logic

\{ P \} \text{ SKIP } \{ P \}

\{ Q[a/x] \} \ x := a \ { Q \}

Notation: \( Q[a/x] \) means “\( Q \) with \( a \) substituted for \( x \)”.

Examples:

\{ \} \ x := 5 \ { x = 5 }  
\{ \} \ x := x+5 \ { x = 5 }  
\{ \} \ x := 2*(x+5) \ { x > 20 }  

Alternative explanation of assignment rule:

\{ Q[a] \} \ x := a \ { Q[x] \}
The assignment axiom allows us to compute the precondition from the postcondition.

There is a version to compute the postcondition from the precondition, but it is more complicated. (Exercise!)
More rules of Hoare Logic

\[
\begin{align*}
\{P_1\} & \ c_1 & \{P_2\} & \ {\{P_2\} \ c_2 \ {\{P_3\}}} \\
\{P_1\} & \ c_1;c_2 & \{P_3\}
\end{align*}
\]

\[
\begin{align*}
\{P \land b\} & \ c_1 & \{Q\} & \ {\{P \land \neg b\} \ c_2 \ {\{Q\}}} \\
\{P\} & \ IF \ b \ THEN & c_1 & \ ELSE \ c_2 & \{Q\}
\end{align*}
\]

\[
\begin{align*}
\{P \land b\} & \ c & \{P\} \\
\{P\} & \ WHILE \ b \ DO & c & \{P \land \neg b\}
\end{align*}
\]

In the While-rule, \(P\) is called an \textit{invariant} because it is preserved across executions of the loop body.
The consequence rule

So far, the rules were syntax-directed. Now we add

\[ P' \rightarrow P \quad \{ P \} \ c \ \{ Q \} \quad Q \rightarrow Q' \]

\[ \{ P' \} \ c \ \{ Q' \} \]

*Preconditions can be strengthened, postconditions can be weakened.*
Two derived rules

Problem with assignment and While-rule: special form of pre and postcondition. Better: combine with consequence rule.

\[
\begin{align*}
P & \rightarrow Q[a/x] \\
\{P\} & x := a \{Q\} \\
\{P \land b\} & c \{P\} \quad P \land \neg b \rightarrow Q \\
\{P\} & WHILE \ b \ DO \ c \ \{Q\}
\end{align*}
\]
\{ x = i \}

\textit{y := 0;}

\textit{WHILE 0 < x DO (y := y+x; x := x-1)}

\{ y = \textit{sum i} \}
Example proof exhibits key properties of Hoare logic:

- Choice of rules is syntax-directed and hence automatic.
- Proof of “;” proceeds from right to left.
- Proofs require only invariants and arithmetic reasoning.
Partial Correctness

Introduction
The Syntactic Approach
The Semantic Approach
Soundness and Completeness
Assertions are predicates on states

\[ \text{assn} = \text{state} \Rightarrow \text{bool} \]

Alternative view: \textit{sets of states}

Semantic approach simplifies meta-theory, our main objective.
Validity

\[ \models \{ P \} \ c \ \{ Q \} \quad \longleftrightarrow \quad \forall s \ t. \ P \ s \land (c, \ s) \Rightarrow t \rightarrow Q \ t \]

"\{ P \} \ c \ \{ Q \} \ is \ valid"

In contrast:

\[ \vdash \{ P \} \ c \ \{ Q \} \]

"\{ P \} \ c \ \{ Q \} \ is \ provable/derivable"
Provability

\[ \vdash \{ P \} \ \text{SKIP} \ \{ P \} \]

\[ \vdash \{ \lambda s. \ Q \ (s[a/x]) \} \ x ::= a \ \{ Q \} \]

where \( s[a/x] \equiv s(x := \text{aval} \ a \ s) \)

Example: \( \{ x+5 = 5 \} \ x ::= x+5 \ \{ x = 5 \} \) in semantic terms:

\[ \vdash \{ P \} \ x ::= \text{Plus} \ (V \ x) \ (N \ 5) \ \{ \lambda t. \ t \ x = 5 \} \]

where \( P = (\lambda s. \ (\lambda t. \ t \ x = 5)(s[\text{Plus} \ (V \ x) \ (N \ 5)/x])) \)
\[ = (\lambda s. \ (\lambda t. \ t \ x = 5)(s(x := s \ x + 5))) \]
\[ = (\lambda s. \ s \ x + 5 = 5) \]
\[ \vdash \{ P \} \ c_1 \ \{ Q \} \quad \vdash \{ Q \} \ c_2 \ \{ R \} \]
\[ \vdash \{ P \} \ c_1 {} ; ; \ c_2 \ \{ R \} \]

\[ \vdash \{ \lambda s. \ P \ s \land bval \ b \ s \} \ c_1 \ \{ Q \} \]
\[ \vdash \{ \lambda s. \ P \ s \land \neg \ bval \ b \ s \} \ c_2 \ \{ Q \} \]

\[ \vdash \{ P \} \ IF \ b \ THEN \ c_1 \ ELSE \ c_2 \ \{ Q \} \]

\[ \vdash \{ \lambda s. \ P \ s \land bval \ b \ s \} \ c \ \{ P \} \]
\[ \vdash \{ P \} \ WHILE \ b \ DO \ c \ \{ \lambda s. \ P \ s \land \neg \ bval \ b \ s \} \]
\[ \forall s. \ P' \ s \rightarrow \ P \ s \]
\[ \vdash \{ P \} \ c \ \{ Q \} \]
\[ \forall s. \ Q \ s \rightarrow \ Q' \ s \]
\[ \vdash \{ P' \} \ c \ \{ Q' \} \]
Hoare_Examples.thy
Partial Correctness

Introduction
The Syntactic Approach
The Semantic Approach

Soundness and Completeness
Soundness

Everything that is provable is valid:

\[ \vdash \{ P \} \ c \ \{ Q \} \quad \Rightarrow \quad \models \ {\{ P \} \ c \ {\{ Q \}}} \]

Proof by induction, with a nested induction in the While-case.
Towards completeness: \[ \vdash \implies \top \]
Weakest preconditions

The weakest precondition of command $c$ w.r.t. postcondition $Q$: 

$$wp\ c\ Q = (\lambda s. \forall t. (c, s) \Rightarrow t \rightarrow Q t)$$

The set of states that lead (via $c$) into $Q$.

A foundational semantic notion, not merely for the completeness proof.
Nice and easy properties of $wp$

$\text{wp \ SKIP } Q = Q$

$\text{wp \ (x ::= a)} \ Q = (\lambda s. \ Q \ (s[a/x]))$

$\text{wp \ (c_1;; \ c_2)} \ Q = \text{wp \ c_1 \ (wp \ c_2 \ Q)}$

$\text{wp \ (IF \ b \ THEN \ c_1 \ ELSE \ c_2)} \ Q =$

$(\lambda s. \ \text{if \ bval \ b \ s \ then \ wp \ c_1 \ Q \ s \ else \ wp \ c_2 \ Q \ s)}$

$\neg \ bval \ b \ s \ \implies \ wp \ (WHILE \ b \ DO \ c) \ Q \ s = Q \ s$

$bval \ b \ s \ \implies$

$wp \ (WHILE \ b \ DO \ c) \ Q \ s =$

$wp \ (\ c_;; \ \text{WHILE} \ b \ DO \ c) \ Q \ s$
Completeness

$$\models \{P\} \ c \ \{Q\} \implies \vdash \{P\} \ c \ \{Q\}$$

Proof idea: do not prove $\vdash \{P\} \ c \ \{Q\}$ directly, prove something stronger:

**Lemma** $\vdash \{wp \ c \ Q\} \ c \ \{Q\}$

**Proof** by induction on $c$, for arbitrary $Q$.

Now prove $\vdash \{P\} \ c \ \{Q\}$ from $\vdash \{wp \ c \ Q\} \ c \ \{Q\}$ by the consequence rule rule because

**Fact** $\models \{P\} \ c \ \{Q\} \iff (\forall s. \ P \ s \implies \ wp \ c \ Q \ s)$

Follows directly from defs of $\models$ and $wp$. 
\[ \vdash \{ P \} \ c \ \{ Q \} \iff \models \{ P \} \ c \ \{ Q \} \]

Proving program properties by Hoare logic (\(\vdash\)) is just as powerful as by operational semantics (\(\models\)).
WARNING

Most texts that discuss completeness of Hoare logic state or prove that Hoare logic is only “relatively complete” but not complete.

Reason: the standard notion of completeness assumes some abstract mathematical notion of $\models$.

Our notion of $\models$ is defined within the same (limited) proof system (for HOL) as $\vdash$. 


Partial Correctness

Verification Conditions

Total Correctness
Idea:

Reduce provability in Hoare logic to provability in the assertion language: automate the Hoare logic part of the problem.

More precisely:

From $\{ P \} \ c \ \{ Q \}$ generate an assertion $A$, the verification condition, such that $\vdash \{ P \} \ c \ \{ Q \}$ iff $A$ is provable.

Method:

Simulate syntax-directed application of Hoare logic rules. Collect all assertion language side conditions.
A problem: loop invariants

Where do they come from?

A trivial solution:

Let the user provide them!

How?

Each loop must be annotated with its invariant!
How to synthesize loop invariants automatically is an important research problem. Which we ignore for the moment. But come back to later.
Terminology:

VCG = Verification Condition Generator

All successful verification technology for imperative programs relies on

- VCGs (of one kind or another)
- and powerful (semi-)automatic theorem provers.
The (approx.) plan of attack

1. Introduce annotated commands with loop invariants

2. Define functions for computing
   - weakest preconditions: $\text{pre} :: \text{com} \Rightarrow \text{assn} \Rightarrow \text{assn}$
   - verification conditions: $\text{vc} :: \text{com} \Rightarrow \text{assn} \Rightarrow \text{bool}$

3. Soundness: $\text{vc} \ c \ Q \implies \vdash \{ ? \} \ c \ \{ Q \}$

4. Completeness: if $\vdash \{ P \} \ c \ \{ Q \}$ then $c$ can be annotated (becoming $C$) such that $\text{vc} \ C \ Q$.

The details are a bit different . . .
Annotated commands

Like commands, except for While:

**datatype** \( acom \) = \( \text{Skip} \)
| \( \text{Assign } vname \ aexp \)
| \( \text{Seq } acom \ acom \)
| \( \text{If } bexp \ acom \ acom \)
| \( \text{While } assn \ bexp \ acom \)

Concrete syntax: like commands, except for \( \text{WHILE} \):

\( \{ I \} \ \text{WHILE } b \ \text{DO } c \)
Weakest precondition

\[ \text{pre} :: \text{acom} \Rightarrow \text{assn} \Rightarrow \text{assn} \]

\[ \text{pre SKIP } Q = Q \]

\[ \text{pre } (x ::= a) \ Q = (\lambda s. \ Q \ (s[a/x])) \]

\[ \text{pre } (C_1;;; C_2) \ Q = \text{pre } C_1 \ (\text{pre } C_2 \ Q) \]

\[ \text{pre } (\text{IF } b \ \text{THEN } C_1 \ \text{ELSE } C_2) \ Q = (\lambda s. \text{if } \text{bval } b \ s \ \text{then } \text{pre } C_1 \ Q \ s \ \text{else } \text{pre } C_2 \ Q \ s) \]

\[ \text{pre } (\{I\} \ \text{WHILE } b \ \text{DO } C) \ Q = I \]
Warning

In the presence of loops,

$pre \ C$ may not be the weakest precondition
but may be anything!
Verification condition

\( vc :: a\text{com} \Rightarrow a\text{ssn} \Rightarrow \text{bool} \)

\( vc \ SKIP \ Q = \text{True} \)

\( vc \ (x ::= a) \ Q = \text{True} \)

\( vc \ (C_1 ;; C_2) \ Q = (vc \ C_1 \ (pre \ C_2 \ Q) \land vc \ C_2 \ Q) \)

\( vc \ (IF \ b \ THEN \ C_1 \ ELSE \ C_2) \ Q = \)

\( (vc \ C_1 \ Q \land vc \ C_2 \ Q) \)

\( vc \ (\{I\} \ WHILE \ b \ DO \ C) \ Q = \)

\( ((\forall \ s. \ (I \ s \land bval \ b \ s \rightarrow pre \ C \ I \ s) \land \)

\( (I \ s \land \neg \ bval \ b \ s \rightarrow Q \ s)) \land \)

\( vc \ C \ I) \)
Verification conditions only arise from loops:

• the invariant must be invariant
• and it must imply the postcondition.

Everything else in the definition of $vc$ is just bureaucracy: collecting assertions and passing them around.
Hoare triples operate on \(com\), functions \(pre\) and \(vc\) operate on \(acom\). Therefore we define

\[
\text{strip} :: \ acom \Rightarrow com
\]

\[
\text{strip} \ SKIP = SKIP
\]

\[
\text{strip} \ (x ::= a) = x ::= a
\]

\[
\text{strip} \ (C_1 ;; C_2) = \text{strip} \ C_1 ;; \text{strip} \ C_2
\]

\[
\text{strip} \ (\text{IF} \ b \ \text{THEN} \ C_1 \ \text{ELSE} \ C_2) = \text{IF} \ b \ \text{THEN} \ \text{strip} \ C_1 \ \text{ELSE} \ \text{strip} \ C_2
\]

\[
\text{strip} \ (\{I\} \ \text{WHILE} \ b \ \text{DO} \ C) = \text{WHILE} \ b \ \text{DO} \ \text{strip} \ C
\]
Soundness of $vc \ & \ pre$ w.r.t. $\vdash$

$$vc \ C \ Q \iff \vdash \{pre \ C \ Q\} \ \text{strip} \ C \ \{Q\}$$

Proof by induction on $C$, for arbitrary $Q$.

Corollary:

$$[vc \ C \ Q; \ \forall s. \ P \ s \ \longrightarrow \ pre \ C \ Q \ s] \\implies \vdash \{P\} \ \text{strip} \ C \ \{Q\}$$

How to prove some $\vdash \{P\} \ c \ \{Q\}$:

- Annotate $c$ yielding $C$, i.e. $\text{strip} \ C = c$.
- Prove Hoare-free premise of corollary.

But is premise provable if $\vdash \{P\} \ c \ \{Q\}$ is?
\[ \text{Why could premise not be provable although conclusion is?} \]

- Some annotation in $C$ is not invariant.
- $vc$ or $pre$ are wrong (e.g. accidentally always produce $False$).

Therefore we prove completeness: suitable annotations exist such that premise is provable.
Completeness of $vc$ & $pre$ w.r.t. $\vdash$

$\vdash \{ P \} \ c \ \{ Q \} \implies \exists \ C. \ \text{strip} \ C = c \land \ vc \ C \ Q \land (\forall \ s. \ P \ s \implies \ pre \ C \ Q \ s)$

Proof by rule induction. Needs two monotonicity lemmas:

$[\forall \ s. \ P \ s \implies P' \ s; \ pre \ C \ P \ s] \implies pre \ C \ P' \ s$

$[\forall \ s. \ P \ s \implies P' \ s; \ vc \ C \ P] \implies vc \ C \ P'$
12 Partial Correctness

13 Verification Conditions

14 Total Correctness
• Partial Correctness:
  *if* command terminates, postcondition holds

• Total Correctness:
  command terminates *and* postcondition holds

**Total Correctness = Partial Correctness + Termination**

Formally:

\[
(\models_t \{ P \} \ c \ \{ Q \}) = \\
(\forall s. \ P \ s \rightarrow (\exists t. \ (c, \ s) \Rightarrow t \land Q \ t))
\]

Assumes that semantics is deterministic!

Exercise: Reformulate for nondeterministic language
⊢ _t: A proof system for total correctness

Only need to change the \textit{WHILE} rule.

Some measure function \textit{state} \rightarrow \textit{nat} must decrease with every loop iteration

\[
\bigwedge n. \vdash_t \{ \lambda s. \ P s \land bval \ b \ s \land n = f \ s \} \ c \ \{ \lambda s. \ P s \land f s < n \} \\
\vdash_t \{ P \} \ \text{WHILE} \ b \ \text{DO} \ c \ \{ \lambda s. \ P s \land \neg \ bval \ b \ s \}
\]
**WHILE** rule can be generalized from a function to a relation:

\[ \forall n. \Gamma_t \{ \lambda s. P s \land bval b s \land T s n \} \xRightarrow{c} \{ \lambda s. P s \land (\exists n' < n. T s n') \} \]

\[ \Gamma_t \{ \lambda s. P s \land (\exists n. T s n) \} \xRightarrow{WHILE b DO c} \{ \lambda s. P s \land \neg bval b s \} \]
Hoare_Total.thy

Example
Soundness

\[ \vdash_t \{ P \} \ c \ \{ Q \} \implies \models_t \{ P \} \ c \ \{ Q \} \]

Proof by induction, with a nested induction on \( n \) in the While-case.
Completeness

\[ \models t \{ P \} c \{ Q \} \implies \vdash t \{ P \} c \{ Q \} \]

Follows easily from

\[ \vdash t \{ \text{wp}_t c \ Q \} c \{ Q \} \]

where

\[ \text{wp}_t c \ Q = (\lambda s. \exists t. (c, s) \Rightarrow t \land Q t). \]
Proof of $\vdash_t \wp_t \ c \ Q \ c \ \{Q\}$ is by induction on $c$. In the `WHILE b DO c` case, use the `WHILE` rule with

\[
\begin{array}{c}
\neg \ bval \ b \ s \\
\hline
T \ s \ 0
\end{array} \quad \quad \quad \quad
\begin{array}{c}
bval \ b \ s \\
\hline
(c, s) \Rightarrow s' \\
\hline
T \ s' \ n
\end{array}
\]

\[
T \ s \ (n + 1)
\]

$T \ s \ n$ means that `WHILE b DO c` started in state $s$ needs $n$ iterations to terminate.
Chapter 13

Abstract Interpretation
Introduction
Annotated Commands
Collecting Semantics
Abstract Interpretation: Orderings
A Generic Abstract Interpreter
Executable Abstract State
Termination
Analysis of Boolean Expressions
Widening and Narrowing
• Abstract interpretation is a generic approach to static program analysis.
• It subsumes and improves our earlier approaches.

Aim:
For each program point, compute the possible values of all variables

Method:
Execute/interpret program with abstract instead of concrete values, eg intervals instead of numbers.
Applications: Optimization

- Constant folding
- Unreachable and dead code elimination
- Array access optimization:
  
  \[
  a[i] := 1; \ a[j] := 2; \ x := a[i] \leadsto
  a[i] := 1; \ a[j] := 2; \ x := 1
  \]
  
  if \( i \neq j \)

- …
Applications: Debugging/Verification

Detect presence or absence of certain runtime exceptions/errors:

• Interval analysis: $i \in [m, n]$:
  - No division by 0 in $e/i$ if $0 \notin [m, n]$
  - No ArrayIndexOutOfBoundsException in $a[i]$ if $0 \leq m \land n < a.length$
  - ...

• Null pointer analysis

• ...

• Null pointer analysis

• ...
A consequence of Rice’s theorem: 

*In general, the possible values of a variable cannot be computed precisely.*

Program analyses overapproximate: they compute a *superset* of the possible values of a variable.

If an analysis says that some value

- cannot arise, this is definitely the case.
- can arise, this is only potentially the case.

Beware of *false alarms* because of overapproximation.
Error

Program Analysis

No Alarm  False Alarm  True Alarm
Annotated commands

Like in Hoare logic, we annotate

\[
\{ \ldots \} 
\]

program text with semantic information. Not just loops but also all intermediate program points, for example:

\[
x := 0 \{ \ldots \}; y := 0 \{ \ldots \}
\]
Annotated WHILE

View

\[ \{ \text{Inv} \} \]

WHILE \( b \) DO  \( \{ P \} \) \( c \)

\( \{ Q \} \)

as a *control flow graph* with annotated nodes:
The starting point: Collecting Semantics

Collects all possible states for each program point:

\[
\begin{align*}
&\text{x := 0 } \{ \langle x := 0 \rangle \} ; \\
&\{ \langle x := 0 \rangle, \langle x := 2 \rangle, \langle x := 4 \rangle \} \\
&\text{WHILE } x < 3 \\
&\text{DO } \{ \langle x := 0 \rangle, \langle x := 2 \rangle \} \\
&\text{x := x+2 } \{ \langle x := 2 \rangle, \langle x := 4 \rangle \} \\
&\{ \langle x := 4 \rangle \}
\end{align*}
\]
Infinite sets of states:

{\ldots, \langle x := -1 \rangle, \langle x := 0 \rangle, \langle x := 1 \rangle, \ldots } \\
\textbf{WHILE} \ x < 3 \\
\textbf{DO} \{ \ldots, \langle x := 1 \rangle, \langle x := 2 \rangle \} \\
\quad \ x := x+2 \ \{ \ldots, \langle x := 3 \rangle, \langle x := 4 \rangle \} \\
\{ \langle x := 3 \rangle, \langle x := 4 \rangle, \ldots \}
Multiple variables:

\[
x := 0; \quad y := 0 \begin{cases} <x:=0, y:=0> \\ <x:=0, y:=0>, <x:=2, y:=1>, <x:=4, y:=2> \end{cases}
\]

WHILE \( x < 3 \)
DO \( \begin{cases} <x:=0, y:=0>, <x:=2, y:=1> \end{cases} \)
\[
x := x+2; \quad y := y+1
\begin{cases} <x:=2, y:=1>, <x:=4, y:=2> \end{cases}
\begin{cases} <x:=4, y:=2> \end{cases}
\]
A first approximation

\[(\text{vname} \Rightarrow \text{val}) \text{ set} \quad \sim \quad \text{vname} \Rightarrow \text{val set}\]

\[
x := 0 \{ <x := \{0\}> \} ;
\{ <x := \{0,2,4\}> \}
\]

\[\text{WHILE } x < 3 \]

\[\text{DO} \{ <x := \{0,2\}> \}
\]

\[x := x+2 \{ <x := \{2,4\}> \}
\]

\[\{ <x := \{4\}> \}\]
Loses relationships between variables but simplifies matters a lot.

Example:

\[
\{ <x:=0,\ y:=0>, \ <x:=1,\ y:=1> \} 
\]

is approximated by

\[
<x:=\{0,1\},\ y:=\{0,1\}> 
\]

which also subsumes

\[
<x:=0,\ y:=1> \text{ and } <x:=1,\ y:=0>. 
\]
Abstract Interpretation

Approximate sets of concrete values by *abstract values*

Example: approximate sets of numbers by intervals

Execute/interpret program with abstract values
Example

Consistently annotated program:

\begin{verbatim}
x := 0 \{ \langle x := [0,0] \rangle \} ;
\{ \langle x := [0,4] \rangle \}
WHILE x < 3
DO \{ \langle x := [0,2] \rangle \}
  x := x+2 \{ \langle x := [2,4] \rangle \}
\{ \langle x := [3,4] \rangle \}
\end{verbatim}

The annotations are computed by

- starting from an un-annotated program and
- iterating abstract execution
- until the annotations stabilize.
x := 0

WHILE x < 3
DO
  x := x + 2
15 Introduction
16 Annotated Commands
17 Collecting Semantics
18 Abstract Interpretation: Orderings
19 A Generic Abstract Interpreter
20 Executable Abstract State
21 Termination
22 Analysis of Boolean Expressions
23 Widening and Narrowing
Concrete syntax

\[ \text{'a acom} ::= \text{SKIP \{} \text{'a} \} \mid \text{string ::= aexp \{} \text{'a} \} \]
\[ \mid \text{'a acom ;; 'a acom} \]
\[ \mid \text{IF bexp THEN \{} \text{'a} \} \text{'a acom ELSE \{} \text{'a} \} \text{'a acom} \]
\[ \{ \text{'a} \} \]
\[ \mid \{ \text{'a} \} \text{WHILE bexp DO \{} \text{'a} \} \text{'a acom} \]
\[ \{ \text{'a} \} \]

\text{'a}: \text{type of annotations}

Example: \[ "x" ::= N 1 \{9}\};; \text{SKIP \{}6\} :: \text{nat acom} \]
Abstract syntax

datatype

'a acom =

  SKIP 'a

| Assign string aexp 'a
| Seq ('a acom) ('a acom)
| If bexp 'a ('a acom) 'a ('a acom) 'a
| While 'a bexp 'a ('a acom) 'a
Auxiliary functions: \emph{strip}

Strips all annotations from an annotated command

\begin{align*}
\text{strip} &:: \ 'a \ acom \Rightarrow \ com \\
\text{strip} \ (\text{SKIP} \ \{P\}) &\ = \ \text{SKIP} \\
\text{strip} \ (x ::= e \ \{P\}) &\ = \ x ::= e \\
\text{strip} \ (C_1 ;; C_2) &\ = \ \text{strip} \ C_1 ;; \ \text{strip} \ C_2 \\
\text{strip} \ (\text{IF} \ b \ \text{THEN} \ \{P_1\} \ C_1 \ \text{ELSE} \ \{P_2\} \ C_2 \ \{P\}) &\ = \ \text{IF} \ b \ \text{THEN} \ \text{strip} \ C_1 \ \text{ELSE} \ \text{strip} \ C_2 \\
\text{strip} \ (\{I\} \ \text{WHILE} \ b \ \text{DO} \ \{P\} \ C \ \{Q\}) &\ = \ \text{WHILE} \ b \ \text{DO} \ \text{strip} \ C
\end{align*}

We call \(C\) and \(C'\) \emph{strip-equal} iff \(\text{strip} \ C = \text{strip} \ C'\).
Auxiliary functions: \textit{annos}

The list of annotations in an annotated command
(from left to right)

\begin{align*}
\text{annos} :: \ 'a \ acom & \Rightarrow \ 'a \ \text{list} \\
\text{annos} (\text{SKIP} \ \{ P \}) & = [P] \\
\text{annos} (x ::= e \ \{ P \}) & = [P] \\
\text{annos} (C_1 ;; C_2) & = \text{annos} C_1 @ \text{annos} C_2 \\
\text{annos} (\text{IF} \ b \ \text{THEN} \ \{ P_1 \} \ C_1 \ \text{ELSE} \ \{ P_2 \} \ C_2 \ \{ Q \}) & = P_1 \# \text{annos} C_1 @ P_2 \# \text{annos} C_2 @ [Q] \\
\text{annos} (\{I\} \ \text{WHILE} \ b \ \text{DO} \ \{ P \} \ C \ \{ Q \}) & = I \# P \# \text{annos} C @ [Q]
\end{align*}
Auxiliary functions: \textit{anno}

\begin{align*}
\text{anno} &:: 'a \text{ acom} \Rightarrow \text{ nat} \Rightarrow 'a \\
\text{anno } C \ p & = \text{ annos } C ! \ p
\end{align*}

The \textit{p}-th annotation (starting from 0)
Auxiliary functions: \textit{post}

\begin{align*}
\textit{post} &:: 'a \text{ acom} \Rightarrow 'a \\
\textit{post }C & = \text{ last } (\text{annos } C)
\end{align*}

The rightmost\/last\/post annotation
Auxiliary functions: \( \text{map\_acom} \)

\[
\text{map\_acom} :: (\text{a} \Rightarrow \text{b}) \Rightarrow \text{a acom} \Rightarrow \text{b acom}
\]

\(\text{map\_acom}\ f\ C\) applies \(f\) to all annotations in \(C\)
Collecting Semantics
Annotate commands with the set of states that can occur at each annotation point.

The annotations are generated iteratively:

\[
\text{step} :: \text{state set} \Rightarrow \text{state set acom} \Rightarrow \text{state set acom}
\]

Each step executes all atomic commands simultaneously, propagating the annotations one step further.

start states flowing into the command
\[ \text{step} \]

\[
\text{step } S \ (\text{SKIP } \{\_\}) = \text{SKIP } \{S\}
\]

\[
\text{step } S \ (x ::= e \ \{\_\}) =
\]
\[
x ::= e \ \{\{s(x ::= \text{aval } e \ s) \mid s. \ s \in S\}\}
\]

\[
\text{step } S \ (C_1 ;; C_2) = \text{step } S \ C_1 ;; \text{step } (\text{post } C_1) \ C_2
\]

\[
\text{step } S \ (\text{IF } b \ \text{THEN } \{P_1\} \ C_1 \ \text{ELSE } \{P_2\} \ C_2 \ \{\_\}) =
\]
\[
\text{IF } b \ \text{THEN } \{\{s \in S. \ bval \ b \ s\}\} \ \text{step } P_1 \ C_1
\]
\[
\text{ELSE } \{\{s \in S. \ \neg \ bval \ b \ s\}\} \ \text{step } P_2 \ C_2
\]
\[
\{\text{post } C_1 \cup \text{post } C_2\}
\]
\textit{step}

\begin{align*}
\text{step } S \left( \{ I \} \quad \text{WHILE } b \quad \text{DO } \{ P \} \quad C \{ \_ \} \right) &= \\
\{ S \cup \text{post } C \} \\
\text{WHILE } b \\
\text{DO } \{ \{ s \in I. \ bval b s \} \} \\
\text{step } P \ C \\
\{ \{ s \in I. \neg bval b s \} \}
\end{align*}
Collecting semantics

View command as a control flow graph
• where you constantly feed in some fixed input set $S$ (typically all possible states)
• and pump/propagate it around the graph
• until the annotations stabilize — this may happen in the limit only!

Stabilization means fixpoint:

$$\text{step } S \ C = \ C$$
Collecting_Examples.thy
Abstract example

Let \( C = \{ I \} \)

\[
\text{WHILE } b \\
\text{DO } \{ P \} C_0 \\
\{ Q \}
\]

\( \text{step } S C = C \) means

\[
I = S \cup \text{post } C_0 \\
P = \{ s \in I. \ bval b s \} \\
C_0 = \text{step } P C_0 \\
Q = \{ s \in I. \neg bval b s \}
\]

Fixpoint = solution of equation system
Iteration is just one way of solving equations
Why *least* fixpoint?

\[
\{ I \} \\
\text{WHILE true} \\
\text{DO \{ } I \text{ \} SKIP \{ } I \text{ \} } \\
\{ \{ \} \} \\
\]

Is fixpoint of \textit{step} \{ \} for every \( I \)

But the “reachable” fixpoint is \( I = \{ \} \)
Does $step$ always have a least fixpoint?
A type 'a is a **partial order** if

- there is a predicate \( \leq :: 'a \Rightarrow 'a \Rightarrow bool \)
- that is **reflexive** \( (x \leq x) \),
- **transitive** \( ([x \leq y; y \leq z] \implies x \leq z) \) and
- **antisymmetric** \( ([x \leq y; y \leq x] \implies x = y) \)
Complete lattice

Definition
A partial order ‘a is a complete lattice if every set S :: 'a set has a greatest lower bound l :: 'a:

- ∀ s ∈ S. l ≤ s
- If ∀ s ∈ S. l' ≤ s then l' ≤ l

The greatest lower bound (infimum) of S is often denoted by \( \bigsqcap S \).

Fact Type 'a set is a complete lattice where \( \leq = \subseteq \) and \( \bigsqcap = \cap \)
**Lemma** In a complete lattice, every set $S$ of elements also has a *least upper bound* (*supremum*) $\bigsqcup S$:

- $\forall s \in S. \ s \leq \bigsqcup S$
- If $\forall s \in S. \ s \leq u$ then $\bigsqcup S \leq u$

The least upper bound is the greatest lower bound of all upper bounds: $\bigsqcup S = \bigsqcap \{u. \ \forall s \in S. \ s \leq u\}$.

Thus complete lattices can be defined via the existence of all infima or all suprema or both.
Existence of least fixpoints

**Definition** A function \( f \) on a partial order \( \leq \) is **monotone** if \( x \leq y \implies f \, x \leq f \, y \).

**Theorem** (Knaster-Tarski) Every monotone function on a complete lattice has the least (pre-)fixpoint

\[ \cap \{ p. \, f \, p \leq p \} \]

**Proof** just like the version for sets.
Ordering ′a acom

An ordering on ′a can be lifted to ′a acom by comparing the annotations of strip-equal commands:

\[ C_1 \leq C_2 \iff \text{strip } C_1 = \text{strip } C_2 \land \left( \forall p < \text{length } (\text{annos } C_1). \text{ anno } C_1 p \leq \text{ anno } C_2 p \right) \]

**Lemma** If ′a is a partial order, so is ′a acom.
Ordering 'a acom

Example:

\[
\begin{align*}
x &::= N 0 \{\{a\}\} \leq x ::= N 0 \{\{a, b\}\} &\iff True \\
x &::= N 0 \{\{a\}\} \leq x ::= N 0 \{\{\}\\} &\iff False \\
x &::= N 0 \{S\} \leq x ::= N 1 \{S\} &\iff False
\end{align*}
\]
The collecting semantics needs to order \textit{state set} \textit{acom}.

Annotations are (state) sets ordered by \( \subseteq \), which form a complete lattice.

Does \textit{state set} \textit{acom} also form a complete lattice?

Almost \ldots
What is the infimum of $\text{SKIP } \{ S \}$ and $\text{SKIP } \{ T \}$?

$\text{SKIP } \{ S \cap T \}$

What is the infimum of $\text{SKIP } \{ S \}$ and $x ::= N \ 0 \ \{ T \}$?

Only strip-equal commands have an infimum
It turns out:

- if `'a` is a complete lattice,
- then for each \( c :: com \)
- the set \( \{ C :: 'a acom. strip C = c \} \)
  is also a complete lattice
- but the whole type `'a acom` is not.

Therefore we make the carrier set explicit.
Complete lattice as a set

**Definition** Let 'a be a partially ordered type. A set $L :: 'a set$ is a complete lattice if every $M \subseteq L$ has a greatest lower bound $\bigwedge M \in L$.

Given sets $A$ and $B$ and a function $f$, $f \in A \rightarrow B$ means $\forall a \in A. \ f \ a \in B$.

**Theorem** (Knaster-Tarski) Let $L :: 'a set$ be a complete lattice and $f \in L \rightarrow L$ a monotone function. Then $f$ (restricted to $L$) has the least fixpoint

$$lfp \ f = \bigwedge \{ p \in L. \ f \ p \leq p \}.$$
Application to $acom$

Let $'a$ be a complete lattice and $c :: com$.
Then $L = \{ C :: 'a acom. strip C = c \}$ is a complete lattice.

The infimum of a set $M \subseteq L$ is computed “pointwise”:

*Annotate $c$ at annotation point $p$ with the infimum of the annotations of all $C \in M$ at $p$."

Example

\[ \bigcap \{ SKIP \{ A \}, SKIP \{ B \}, \ldots \} \]
\[ = SKIP \{ \bigcap \{ A, B, \ldots \} \} \]

Formally . . .
Auxiliary function: \textit{annotate}

\begin{align*}
\text{annotate} &:: (\text{nat} \Rightarrow \, 'a) \Rightarrow \text{com} \Rightarrow \, 'a \ acom \\
\text{Set annotation number } p \text{ (as counted by } \text{anno} \text{) to } f \, p. \text{ Definition is technical. The characteristic lemma:} \\
\text{anno} \ (\text{annotate} \ f \ c) \ p & = f \ p
\end{align*}
**Lemma** Let 'a be a complete lattice and $c :: com$. Then $L = \{C :: 'a acom. strip C = c\}$ is a complete lattice where the infimum of $M \subseteq L$ is

$$\text{annotate} \left( \lambda p. \bigcap \{\text{anno } C \; p \mid C. \; C \in M\} \right) c$$

**Proof** straightforward (pointwise).
The Collecting Semantics

The underlying complete lattice is now *state set*. Therefore $L = \{ C :: \text{state set acom. strip } C = c \}$ is a complete lattice for any $c$.

**Lemma** step $S \in L \to L$ and is monotone.

Therefore Knaster-Tarski is applicable and we define

\[
CS :: \text{com } \Rightarrow \text{state set acom}
\]

\[
CS c = \text{lfp } c \ (\text{step } \text{UNIV})
\]

[lfp is defined in the context of some lattice $L$. Our concrete $L$ depends on $c$. Therefore lfp depends on $c$, too.]
Relationship to big-step semantics

For simplicity: compare only pre and post-states

**Theorem** \((c, s) \Rightarrow t \iff t \in \text{post}(CS\ c)\)

Follows directly from

\[
\left[ (c, s) \Rightarrow t; \ s \in S \right] \implies t \in \text{post}(\text{lfp} \ c \ (\text{step} \ S))
\]
Proof of

\[ \left[ (c, s) \Rightarrow t; \ s \in S \right] \Rightarrow t \in \text{post}(\text{lfp } c \ (\text{step } S)) \]

uses

\[ \text{post}(\text{lfp } c \ f) = \bigcap \{ \text{post } C \mid C. \ \text{strip } C = c \wedge f \ C \leq C \} \]

and

\[ \left[ (c, s) \Rightarrow t; \ \text{strip } C = c; \ s \in S; \ \text{step } S \ C \leq C \right] \Rightarrow t \in \text{post } C \]

which is proved by induction on the big step.
In a nutshell:
collecting semantics overapproximates big-step semantics

Later:
program analysis overapproximates collecting semantics
Together:
program analysis overapproximates big-step semantics

The other direction

\[ t \in post(lfp_c(step\ S)) \implies \exists s \in S. (c, s) \Rightarrow t \]

is also true but is not proved in this course.
Abstract Interpretation: Orderings

A Generic Abstract Interpreter

Executable Abstract State

Termination

Analysis of Boolean Expressions

Widening and Narrowing
Approximating the Collecting semantics

A conceptual step:

\[(vname \Rightarrow val) \text{ set } \sim vname \Rightarrow val \text{ set}\]

A domain-specific step:

\[val \text{ set } \sim \ 'av\]

where \ 'av \ is some ordered type of abstract values that we can compute on.
Example: parity analysis

Abstract values:

\[ \textbf{datatype} \quad \text{parity} = \text{Even} \mid \text{Odd} \mid \text{Either} \]

\[ \begin{array}{ll}
\text{Either} & \rightarrow Z \\
\text{Even} & \rightarrow 2\mathbb{Z} \\
\text{Odd} & \rightarrow 2\mathbb{Z} + 1
\end{array} \]

concretization function \( \gamma_{\text{parity}} \)
A concretisation function $\gamma$
maps an abstract value to a set of concrete values

Bigger abstract values represent more concrete values
Example: parity

Fact Type \( \text{parity} \) is a partial order.
A partial order 'a has a top element ⊤ :: 'a if

\[ a \leq \top \]
A type ′a is a *semilattice* if
- it is a partial order and
- there is a least upper bound operation
\[\sqcup :: \text{′a} \Rightarrow \text{′a} \Rightarrow \text{′a}\]
\[x \leq x \sqcup y \quad y \leq x \sqcup y\]
\[\left[ x \leq z; \; y \leq z \right] \implies x \sqcup y \leq z\]

Application: abstract ∪, join two computation paths
We often call △ the *join* operation.
Fact If \( \mathcal{A} \) is a semilattice, then the least upper bound of two elements is uniquely determined.

If \( u_1 \) and \( u_2 \) are least upper bounds of \( x \) and \( y \), then \( u_1 \leq u_2 \) and \( u_2 \leq u_1 \).
Example: parity

Fact Type \textit{parity} is a semilattice with top element.
Isabelle’s type classes

A type class is defined by

- a set of required functions (the interface)
- and a set of axioms about those functions

Examples

class order: partial orders
class semilattice_sup: semilattices
class semilattice_sup_top: semilattices with top element

A type belongs to some class if

- the interface functions are defined on that type
- and satisfy the axioms of the class (proof needed!)

Notation: \( \tau :: C \) means type \( \tau \) belongs to class \( C \)

Example: \( \text{parity} :: \text{semilattice\_sup} \)
HOL/Orderings.thy
Abs_Int1_parity.thy

Orderings and instances
From abstract values to abstract states

Need to abstract collecting semantics:

\[
\text{state set}
\]

- First attempt:
  \[
  \text{'av st } = \text{ vname } \Rightarrow \text{'av}
  \]
  where \text{'av} is the type of abstract values

- Problem: cannot abstract empty set of states (unreachable program points!)

- Solution: type \text{'av st option}
Lifting semilattice and $\gamma$ to $\text{'av st option}$

**Lemma**  
If $\text{'a :: semilattice\_sup\_top}$  
then $\text{'b \Rightarrow \text{'a :: semilattice\_sup\_top}}$

**Proof**  
\[(f \leq g) = (\forall x. f \; x \leq g \; x)\]  
\[f \sqcup g = (\lambda x. f \; x \sqcup g \; x)\]  
\[\top = (\lambda x. \top)\]

**definition**  
\[\gamma\_fun :: (\text{'a \Rightarrow \text{'c set}) \Rightarrow (\text{'b \Rightarrow \text{'a}) \Rightarrow (\text{'b \Rightarrow \text{'c})set}\]  
where $\gamma\_fun \; \gamma \; F = \{f. \forall x. f \; x \in \gamma \; (F \; x)\}$

**Lemma**  
If $\gamma$ is monotone then $\gamma\_fun \; \gamma$ is monotone.
Lemma  If 'a :: semilattice_sup_top
        then 'a option :: semilattice_sup_top

Proof
(Some x ≤ Some y) = (x ≤ y)
(None ≤ _) = True
(Some _ ≤ None) = False

Some x ▷ Some y = Some (x ▷ y)
None ▷ y = y
x ▷ None = x
⊤ = Some ⊤

Corollary  If 'a :: semilattice_sup_top
        then 'a st option :: semilattice_sup_top
fun \( \gamma_{\text{option}} :: (\text{'a} \Rightarrow \text{'c set}) \Rightarrow \text{'a option} \Rightarrow \text{'c set} \)

where

\( \gamma_{\text{option}} \gamma \ None = \{} \)
\( \gamma_{\text{option}} \gamma \ (\text{Some } a) = \gamma \ a \)

Lemma If \( \gamma \) is monotone then \( \gamma_{\text{option}} \gamma \) is monotone.
Remember:

**Lemma** If \( \mathcal{A} \) :: \( \text{order} \) then \( \mathcal{A} \text{ acom} \) :: \( \text{order} \).

Partial order is enough, semilattice not needed.

Lifting \( \gamma : \mathcal{A} \Rightarrow \mathcal{C} \) to \( \mathcal{A} \text{ acom} \Rightarrow \mathcal{C} \text{ acom} \) is easy: \( \text{map}_\text{acom} \)

**Lemma**

If \( \gamma \) is monotone then \( \text{map}_\text{acom} \gamma \) is monotone.
• Stepwise development of a generic abstract interpreter as a parameterized module

• Parameters/Input: abstract type of values together with abstractions of the operations on concrete type \( \text{val} = \text{int} \).

• Result/Output: abstract interpreter that approximates the collecting semantics by computing on abstract values.

• Realization in Isabelle as a locale
Parameters (I)

Abstract values: type \( 	exttt{av} :: \text{semilattice}\_\text{sup}\_\text{top} \)

Concretization function: \( \gamma :: \texttt{av} \Rightarrow \text{val set} \)

Assumptions:
\[
\begin{align*}
a \leq b & \implies \gamma a \subseteq \gamma b \\
\gamma \top &= \text{UNIV}
\end{align*}
\]
Parameters (II)

Abstract arithmetic:  
\[ \text{num'} :: \text{val} \Rightarrow \text{'av} \]
\[ \text{plus'} :: \text{'av} \Rightarrow \text{'av} \Rightarrow \text{'av} \]

Intention:  
\text{num'} abstracts the meaning of \( N \)
\text{plus'} abstracts the meaning of \( \text{Plus} \)

Required for each constructor of \text{aexp} (except \( V \))

Assumptions:
\[ i \in \gamma (\text{num'} \ i) \]
\[ [i_1 \in \gamma \ a_1; \ i_2 \in \gamma \ a_2] \implies i_1 + i_2 \in \gamma (\text{plus'} \ a_1 \ a_2) \]

The \( n \in \gamma \ a \) relationship is maintained
Lifted concretization functions

\[ \gamma_s :: \mathcal{av} \mathcal{st} \rightarrow \text{state set} \]
\[ \gamma_s = \gamma_{\text{fun}} \gamma \]

\[ \gamma_o :: \mathcal{av} \mathcal{st} \mathcal{option} \rightarrow \text{state set} \]
\[ \gamma_o = \gamma_{\text{option}} \gamma_s \]

\[ \gamma_c :: \mathcal{a} \mathcal{st} \mathcal{option} \mathcal{acom} \rightarrow \text{state set acom} \]
\[ \gamma_c = \text{map}_{\text{acom}} \gamma_o \]

All of them are monotone.
Abstract interpretation of $aexp$

fun $aval' :: aexp \Rightarrow 'av \; st \Rightarrow 'av$

$aval' \ (N \ n) \ S = num' \ n$

$aval' \ (V \ x) \ S = S \ x$

$aval' \ (Plus \ a_1 \ a_2) \ S = plus' \ (aval' \ a_1 \ S) \ (aval' \ a_2 \ S)$

Correctness of $aval'$ wrt $aval$:

Lemma $s \in \gamma_s \ S \Longrightarrow aval \ a \ s \in \gamma \ (aval' \ a \ S)$

Proof by induction on $a$
  using the assumptions about the parameters.
Example instantiation with \textit{parity}

\[ \leq/\sqcup \text{ and } \gamma_{\text{parity}} : \text{ see earlier} \]

\[
\text{num}_{\text{parity}} i = (\text{if } i \mod 2 = 0 \text{ then Even else Odd})
\]

\[
\text{plus}_{\text{parity}} \text{ Even Even } = \text{ Even} \\
\text{plus}_{\text{parity}} \text{ Odd Odd } = \text{ Even} \\
\text{plus}_{\text{parity}} \text{ Even Odd } = \text{ Odd} \\
\text{plus}_{\text{parity}} \text{ Odd Even } = \text{ Odd} \\
\text{plus}_{\text{parity}} \text{ Either } y = \text{ Either} \\
\text{plus}_{\text{parity}} x \text{ Either } = \text{ Either}
\]
Example instantiation with \textit{parity}

\textbf{Input:} \[\gamma \mapsto \gamma_{\text{parity}}\]
\[\text{num}' \mapsto \text{num\_parity}\]
\[\text{plus}' \mapsto \text{plus\_parity}\]

\textbf{Must prove parameter assumptions}

\textbf{Output:} \[\text{aval}' \mapsto \text{aval\_parity}\]

\textbf{Example} The value of

\[\text{aval\_parity} (\text{Plus} (V "x") (V "x"))\]
\[ ((\lambda_. \text{Either})("x" := \text{Odd}))\]

is \textit{Even}. 
Abs_Int1_parity.thy

Locale interpretation
Abstract interpretation of \textit{bexp}

For now, boolean expressions are not analysed.
Abstract interpretation of \textit{com}

Abstracting the collecting semantics

\[
\text{step} :: \quad \tau \Rightarrow \tau \text{ acom} \Rightarrow \tau \text{ acom} \\
\text{where} \quad \tau = \text{state set}
\]

to

\[
\text{step'} :: \quad \tau \Rightarrow \tau \text{ acom} \Rightarrow \tau \text{ acom} \\
\text{where} \quad \tau = 'av \text{ st option}
\]

Idea: define both as instances of a generic step function:

\[
\text{Step} :: 'a \Rightarrow 'a \text{ acom} \Rightarrow 'a \text{ acom}
\]
Parameterized wrt

- type 'a with □
- the interpretation of assignments and tests:
  \[ asem :: vname \Rightarrow aexp \Rightarrow 'a \Rightarrow 'a \]
  \[ bsem :: bexp \Rightarrow 'a \Rightarrow 'a \]
Step a \((SKIP \ {\_})\) = SKIP \{a\}

Step a \((x ::= e \ {\_})\) = x ::= e \{asem x e a\}

Step a \((C_1;; C_2)\) = Step a \(C_1;; \) Step \((post \ C_1) \ C_2\)

Step a \((\text{IF } b \ \text{THEN } \{P_1\} \ C_1 \ \text{ELSE } \{P_2\} \ C_2 \ {\_})\) =
IF b THEN \{bsem b a\} Step P_1 C_1
ELSE \{bsem (Not b) a\} Step P_2 C_2
{post C_1 \sqcup post C_2}

Step a \((\{I\} \ \text{WHILE } b \ \text{DO } \{P\} \ C \ {\_})\) =
\{a \sqcup post C\} \ \text{WHILE } b \ \text{DO } \{bsem b I\} \ Step P \ C
\{bsem (Not b) I\}
Instantiating \textit{Step}\textsuperscript{\textcopyright}

The truth: \textit{asem} and \textit{bsem} are (hidden) parameters of \textit{Step}: \texttt{Step asem bsem} ...

\[
\text{step} = \\
\text{Step} (\lambda x \; e \; S. \; \{s(x := \text{aval} \; e \; s) \mid s. \; s \in S\}) \\
(\lambda b \; S. \; \{s \in S. \; \text{bval} \; b \; s\})
\]

\[
\text{step}' = \text{Step} \; \text{asem} \; (\lambda b \; S. \; S)
\]

\text{where}

\[
\text{asem} \; x \; e \; S = \\
(\text{case} \; S \; \text{of} \; \text{None} \Rightarrow \text{None} \\
| \; \text{Some} \; S \Rightarrow \text{Some} \; (S(x := \text{aval}' \; e \; S)))
\]
Example: iterating \textit{step\_parity}

\[(step\_parity\ S)^k \ C\]

where

\[
\begin{align*}
C &= x ::= N\ 3 \ \{None\} ; \\
    & \{None\} \\
    & \text{WHILE} \ b \ \text{DO} \ \{None\} \\
    & \quad x ::= \text{Plus} \ (V\ x) \ (N\ 5) \ \{None\} \\
    & \quad \{None\} \\
S &= Some\ (\lambda_.\ \text{Either}) \\
S_p &= Some\ ((\lambda_.\ \text{Either})(x ::= p))
\end{align*}
\]
Correctness of $step'$ wrt $step$

The conretization of $step'$ overapproximates $step$:

**Corollary** $step \ (\gamma_o \ S) \ (\gamma_c \ C) \leq \gamma_c \ (step' \ S \ C)$

where $S :: 'av \ st \ option$

$C :: 'av \ st \ option \ acom$

**Lemma** $Step \ f \ g \ (\gamma_o \ S) \ (\gamma_c \ C) \leq \gamma_c \ (Step \ f' \ g' \ S \ C)$

if for all $x, e, b$: $f \ x \ e \ (\gamma_o \ S) \subseteq \gamma_o \ (f' \ x \ e \ S)$

$g \ b \ (\gamma_o \ S) \subseteq \gamma_o \ (g' \ b \ S)$

**Proof** by an easy induction on $C$
The abstract interpreter

- Ideally: iterate $step'$ until a fixpoint is reached
- May take too long
- Sufficient: any pre-fixpoint: $step' S C \leq C$

Means iteration does not increase annotations, i.e. annotations are consistent but maybe too big
Unbounded search

From the HOL library:

\[
\text{while\_option} ::
\]

\[
('a \Rightarrow \text{bool}) \Rightarrow ('a \Rightarrow 'a) \Rightarrow 'a \Rightarrow 'a \text{ option}
\]

such that

\[
\text{while\_option } b \ f \ x =
\]

\[
(\text{if } b \ x \ \text{then} \ \text{while\_option} \ b \ f \ (f \ x) \ \text{else} \ \text{Some} \ x)
\]

and \hspace{1em} \text{while\_option } b \ f \ x = \text{None}

if the recursion does not terminate.
Pre-fixpoint:

\[
pfp :: (\:\'a \Rightarrow \:\'a) \Rightarrow \:\'a \Rightarrow \:\'a \text{ option}
\]

\[
pfp \ f = \text{while\_option} (\\lambda x. \neg f \ x \leq x) \ f
\]

Start iteration with least annotated command:

\[
\text{bot \ } c = \text{annotate} (\\lambda p. \ None) \ c
\]
The generic abstract interpreter

definition $AI :: \text{com} \Rightarrow 'av \text{st} \text{option acom option}$

where $AI c = \text{pfp} \ (\text{step}' \top) \ (\text{bot} c)$

Theorem $AI c = \text{Some} \ C \implies CS c \leq \gamma_c C$

Proof From the assumption: $\text{step}' \top C \leq C$

By monotonicity: $\gamma_c (\text{step}' \top C) \leq \gamma_c C$

By step/step': $\text{step} \ (\gamma_o \top) \ (\gamma_c C) \leq \gamma_c (\text{step}' \top C)$

Hence $\gamma_c C$ is a pfp of $\text{step} \ (\gamma_o \top) = \text{step} \ \text{UNIV}$

Because $CS$ is the least pfp of $\text{step} \ \text{UNIV}$: $CS c \leq \gamma_c C$
Problem

AI is not directly executable

because \( pfp \) compares \( f \ C \leq C \)
where \( C :: 'av \ st \ option \ acom \)
which compares functions \( vname \Rightarrow 'av \)
which is not computable: \( vname \) is infinite.
15 Introduction
16 Annotated Commands
17 Collecting Semantics
18 Abstract Interpretation: Orderings
19 A Generic Abstract Interpreter
20 Executable Abstract State
21 Termination
22 Analysis of Boolean Expressions
23 Widening and Narrowing
Solution

Record only the finite set of variables actually present in a program.

An association list representation:

**type_synonym** \( 'a \text{ st}_\text{rep} = (vname \times 'a) \text{ list} \)

From \( 'a \text{ st}_\text{rep} \) back to \( vname \Rightarrow 'a \):

**fun** fun_rep :: ('a::top) st_rep ⇒ (vname ⇒ 'a)

fun_rep \([(x, a) \# ps]\) = (fun_rep ps)(x := a)

fun_rep [] = (λx. ⊤)

Missing variables are mapped to \( ⊤ \)

Example: fun_rep [("x", a), ("x", b)]

= ((λx. ⊤)("x" := b))("x" := a) = (λx. ⊤)("x" := a)
Comparing association lists

Compare them only on their finite “domains”:

\[
\text{less_eq_st_rep } ps_1 \ ps_2 = \\
(\forall x \in \text{set } (\text{map } \text{fst } ps_1) \cup \text{set } (\text{map } \text{fst } ps_2). \\
\text{fun_rep } ps_1 \ x \leq \text{fun_rep } ps_2 \ x)
\]

Not a partial order because not antisymmetric!
Example: \[(["x", a], ["y", b]) \text{ and } (["y", b], ["x", a])\]
Quotient type \( \texttt{'}a \texttt{ st} \)

Define \( \texttt{eq\_st } \texttt{ps}_1 \texttt{ ps}_2 = (\texttt{fun\_rep} \texttt{ ps}_1 = \texttt{fun\_rep} \texttt{ ps}_2) \)

Overwrite \( \texttt{'}a \texttt{ st} = \texttt{vname} \Rightarrow \texttt{'}a \) by

\texttt{quotient\_type} \( \texttt{'}a \texttt{ st} = (\texttt{'}a::\texttt{top}) \texttt{ st\_rep} / \texttt{eq\_st} \)

Elements of \( \texttt{'}a \texttt{ st} \):

\texttt{equivalence classes} \ \ [\texttt{ps}]_{\texttt{eq\_st}} = \{ \texttt{ps}'. \ \texttt{eq\_st} \texttt{ ps} \texttt{ ps}' \} 

Abbreviate \( [\texttt{ps}]_{\texttt{eq\_st}} \) by \( \texttt{St} \texttt{ ps} \)
Alternative to quotient: canonical representatives

For example, the subtype of sorted association lists:

- \([\langle "x", a \rangle, \langle "y", b \rangle]\)
- \([\langle "y", b \rangle, \langle "x", a \rangle]\)

More concrete, and probably a bit more complicated
Auxiliary functions on 'a st

Turning an abstract state into a function:
\[ \text{fun } (\text{St } ps) = \text{fun_rep } ps \]

Updating an abstract state:
\[ \text{update } (\text{St } ps) \ x \ a = \text{St } ((x, a) \# ps) \]
Turning 'a st into a semilattice

\[(St \, ps_1 \leq St \, ps_2) = less\_eq\_st\_rep \, ps_1 \, ps_2\]

\[St \, ps_1 \sqcup St \, ps_2 = St(map\_2\_st\_rep \, (op \sqcup) \, ps_1 \, ps_2)\]

**fun** map2_st_rep ::

\[('a::top \Rightarrow 'a \Rightarrow 'a) \Rightarrow 'a \, st\_rep \Rightarrow 'a \, st\_rep \Rightarrow 'a \, st\_rep\]

map2_st_rep f ((x, y) # ps1) ps2 =

let y2 = fun_rep ps2 x

in (x, f y y2) # map2_st_rep f ps1 ps2

map2_st_rep f [] ps2 = map (λ(x, y). (x, f ⊤ y)) ps2
Modified abstract interpreter

Everything as before, except for $S :: 'a v s t$:

\[
\begin{align*}
S \; x & \quad \leadsto \quad \text{fun} \; S \; x \\
S(x := a) & \quad \leadsto \quad \text{update} \; S \; x \; a
\end{align*}
\]

Now $AI$ is executable!
Abs_Int1_parity.thy
Abs_Int1_const.thy

Examples
Beyond partial correctness

- AI may compute any pfp
- AI may not terminate

The solution: Monotonicity

\[ \implies \]

Precision \( AI \) computes least pre-fixpoints

Termination \( AI \) terminates if \( 'av \) is of bounded height
Monotonicity

The *monotone framework* also demands monotonicity of abstract arithmetic:

\[
[a_1 \leq b_1; \ a_2 \leq b_2] \implies \text{plus}' \ a_1 \ a_2 \leq \text{plus}' \ b_1 \ b_2
\]

**Theorem** In the monotone framework, \( \text{aval}' \) is also monotone

\[
S_1 \leq S_2 \implies \text{aval}' \ e \ S_1 \leq \text{aval}' \ e \ S_2
\]

and therefore \( \text{step}' \) is also monotone:

\[
[S_1 \leq S_2; \ C_1 \leq C_2] \implies \text{step}' \ S_1 \ C_1 \leq \text{step}' \ S_2 \ C_2
\]
Precision: smaller is better

If \( f \) is monotone and \( \perp \) is a least element, then \( pfp \ f \ \perp \) is a least pre-fixpoint of \( f \)
**Lemma** Let $\leq$ be a partial order on a set $L$ with least element $\bot \in L$: $x \in L \implies \bot \leq x$.

Let $f \in L \to L$ be a monotone function.

If $\text{while\_option} \ (\lambda x. \neg f x \leq x) \ f \bot = \text{Some} \ p$ then $p$ is the least pre-fixpoint of $f$ on $L$.

That is, if $f q \leq q$ for some $q \in L$, then $p \leq q$.

**Proof** Clearly $f p \leq p$.

Given any pre-fixpoint $q \in L$, property

$$P x = (x \in L \land x \leq q)$$

is an invariant of the while loop:

$$P \bot \text{ holds and } P x \implies f x \leq f q \leq q$$

Hence upon termination $P p$ must hold and thus $p \leq q$. 
Application to

\[ AI \ c = pfp \ (step' \top) \ (bot \ c) \]
\[ pfp \ f = while\_option \ (\lambda x. \neg f \ x \leq x) \ f \]

Because \( bot \ c \) is a least element and \( step' \) is monotone, \( AI \) returns least pre-fixpoints
Termination

Because \textit{step}' is monotone, starting from \textit{bot c} generates an ascending $<$ chain of annotated commands. We exhibit a measure function $m_c$ that decreases with every loop iteration:

$$C_1 < C_2 \implies m_c C_2 < m_c C_1$$

Modulo some details ...
The measure function $m_c$ is constructed from a measure function $m$ on 'av in several steps.

Parameters: $m :: 'av \Rightarrow \text{nat}$

$h :: \text{nat}$

Assumptions: $m x \leq h$

$x < y \implies m y < m x$

Parameter $h$ is the height of $<$:
every chain $x_0 < x_1 < \ldots$ has length at most $h$.

Application to parity and const: $h = 1$
Measure functions

\[ m_c : \text{'av st option acom} \Rightarrow \text{nat} \]
\[ m_c \ C = (\sum a \leftarrow \text{annos} \ C. \ m_o \ a \ (\text{vars} \ C)) \]

\[ m_o : \text{'av st option} \Rightarrow \text{vname set} \Rightarrow \text{nat} \]
\[ m_o \ (\text{Some} \ S) \ X = m_s \ S \ X \]
\[ m_o \ None \ X = h * \text{card} \ X + 1 \]

\[ m_s : \text{'av st} \Rightarrow \text{vname set} \Rightarrow \text{nat} \]
\[ m_s \ S \ X = (\sum x \in X. \ m \ (S \ x)) \]
All measure functions are bounded:

\[
\text{finite } X \implies m_s S X \leq h \ast \text{card } X
\]
\[
\text{finite } X \implies m_o \text{ opt } X \leq h \ast \text{card } X + 1
\]
\[
m_c C \leq \text{length (annos } C) \ast (h \ast \text{card (vars } C) + 1)
\]

Therefore \( AI c \) requires at most \( p \ast (h \ast n + 1) \) steps
where \( p = \) the number of annotation points of \( c \)
and \( n = \) the number of variables in \( c \).
Anti-monotonicity does not hold!

Example:

\[
\text{finite } X \implies S_1 < S_2 \implies m_s S_2 X < m_s S_1 X
\]

because \( S_1 < S_2 \iff S_1 \leq S_2 \land (\exists x. S_1 x < S_2 x) \)

Need to know that \( S_1 \) and \( S_2 \) are the same outside \( X \). Follows if both are \( \top \) outside \( X \).
\textbf{top\_on}

\begin{align*}
top\_on_s &:: \quad \text{'av st} \Rightarrow \text{vname set} \Rightarrow \text{bool} \\
top\_on_s \ S \ X &= (\forall \ x \in X. \ S \ x = \top) \\
top\_on_o &:: \quad \text{'av st option} \Rightarrow \text{vname set} \Rightarrow \text{bool} \\
top\_on_o \ (\text{Some} \ S) \ X &= top\_on_s \ S \ X \\
top\_on_o \ \text{None} \ X &= \text{True} \\
top\_on_c &:: \quad \text{'av st option acom} \Rightarrow \text{bool} \\
top\_on_c \ C \ X &= (\forall \ a \in \text{set} \ (\text{annos} \ C). \ top\_on_o \ a \ X)
\end{align*}
Now we can formulate and prove anti-monotonicity:

\[
\begin{align*}
&\text{finite } X; \ S_1 = S_2 \ on \ -X; \ S_1 < S_2 \\
&\implies m_s \ S_2 \ X < m_s \ S_1 \ X
\end{align*}
\]

\[
\begin{align*}
&\text{finite } X; \ top_{\text{on}_o} \ o_1 (-X); \ top_{\text{on}_o} \ o_2 (-X); \\
&\quad o_1 < o_2 \\
&\implies m_o \ o_2 \ X < m_o \ o_1 \ X
\end{align*}
\]

\[
\begin{align*}
&\text{top}_{\text{on}_c} \ C_1 (-\text{vars } C_1); \ top_{\text{on}_c} \ C_2 (-\text{vars } C_2); \\
&\quad C_1 < C_2 \\
&\implies m_c \ C_2 < m_c \ C_1
\end{align*}
\]
Now we can prove termination

$$\exists C. \ AI c = Some \ C$$

because $step'$ leaves $top_on_s$ invariant:

$$top_on_c \ C (\neg \ vars \ C) \implies$$
$$top_on_c (step' \top \ C) (\neg \ vars \ C)$$
Warning: $step'$ is very inefficient.

It is applied to every subcommand in every step. Thus the actual complexity of $AI$ is $O(p^2 \times n \times h)$

Better iteration policy:
Ignore subcommands where nothing has changed.

Practical algorithms often use a control flow graph and a worklist recording the nodes where annotations have changed.

As usual: efficiency complicates proofs.
Abs_Int1_parity.thy
Abs_Int1_const.thy

Termination
Introduction

Annotated Commands

Collecting Semantics

Abstract Interpretation: Orderings

A Generic Abstract Interpreter

Executable Abstract State

Termination

Analysis of Boolean Expressions

Widening and Narrowing
Need to simulate collecting semantics ($S :: state set$):

$$\{s \in S. \ bval \ b \ s\}$$

Given $S :: 'av \ st$, reduce it to some $S' \leq S$ such that

if $s \in \gamma_s S$ and $bval \ b \ s$ then $s \in \gamma_s S'$

- No state satisfying $b$ is lost
- but $\gamma_s S'$ may still contain states not satisfying $b$.
- Trivial solution: $S' = S$

Computing $S'$ from $S$ requires $\square$
A type `'a` is a **lattice** if

- it is a semilattice
- there is a greatest lower bound operation
  \[ \sqcap :: 'a \Rightarrow 'a \Rightarrow 'a \]

\[ x \sqcap y \leq x \quad x \sqcap y \leq y \]

\[ [z \leq x; \ z \leq y] \implies z \leq x \sqcap y \]

Note: \( \sqcap \) is also called **infimum** or **meet**.

Type class: **lattice**
Bounded lattice

A type `'a` is a *bounded lattice* if

- it is a lattice
- there is a top element \( \top \) :: `'a`
- and a *bottom* element \( \bot \) :: `'a`
  \[ \bot \leq a \]

Type class: *bounded_lattice*

**Fact** Any complete lattice is a bounded lattice.
Concretization

We strengthen the abstract interpretation framework by assuming

- ′av :: bounded_lattice
- \( \gamma a_1 \cap \gamma a_2 \subseteq \gamma (a_1 \cap a_2) \)

\[ \implies \gamma (a_1 \cap a_2) = \gamma a_1 \cap \gamma a_2 \]
\[ \implies \cap \text{ is precise!} \]

How about \( \gamma a_1 \cup \gamma a_2 \) and \( \gamma (a_1 \sqcup a_2) \)?

- \( \gamma \bot = \{\} \)
Backward analysis of $aexp$

Given

$e :: aexp$

$a :: 'av$ (the intended value of $e$)

$S :: 'av st$

restrict $S$ to some $S' \leq S$ such that

$$\{ s \in \gamma_s S. \ aval e s \in \gamma a \} \subseteq \gamma_s S'$$

$\gamma_s S'$ overapproximates the subset of $\gamma_s S$ that makes $e$ evaluate to $a$.

What if $\{ s \in \gamma_s S. \ aval e s \in \gamma a \}$ is empty?

Work with $'av st option$ instead of $'av st$
inv_aval' \ N

\[
\text{inv_aval'} :: \\
\text{aexp} \Rightarrow \text{'av} \Rightarrow \text{'av st option} \Rightarrow \text{'av st option}
\]

\[
\text{inv_aval'} (N \ n) \ a \ S = \\
(\text{if test_num'} n \ a \ \text{then} \ S \ \text{else} \ \text{None})
\]

An extension of the interface of our framework:

\[
\text{test_num'} :: \text{int} \Rightarrow \text{'av} \Rightarrow \text{bool}
\]

Assumption:

\[
\text{test_num'} i \ a = (i \in \gamma a)
\]

Note: \( i \in \gamma a \) not necessarily executable
\[ inv_{\text{aval}'} \ (V \ x) \ a \ S = \]

\[
\begin{cases} 
\text{case } S \text{ of } \text{None} \Rightarrow \text{None} \\
\mid \text{Some } S \Rightarrow \\
\quad \text{let } a' = \text{fun } S \ x \sqcap a \\
\quad \text{in if } a' = \bot \text{ then None} \\
\quad \quad \text{else Some (update } S \ x \ a')
\end{cases}
\]

Avoid \( \bot \) component in abstract state, turn abstract state into \textit{None} instead.
**inv_aval' Plus**

A further extension of the interface of our framework:

\[ \text{inv_plus'} :: 'av \Rightarrow 'av \Rightarrow 'av \Rightarrow 'av \times 'av \]

**Assumption:**

\[ \text{inv_plus'} \ a \ a_1 \ a_2 = (a_1', a_2') \Rightarrow \\
\gamma a_1' \supseteq \{ i_1 \in \gamma a_1. \exists i_2 \in \gamma a_2. i_1+i_2 \in \gamma a \} \land \\
\gamma a_2' \supseteq \{ i_2 \in \gamma a_2. \exists i_1 \in \gamma a_1. i_1+i_2 \in \gamma a \} \]

**Definition:**

\[ \text{inv_aval'} (\text{Plus } e_1 \ e_2) \ a \ S = \\
\text{(let } (a_1, a_2) = \text{inv_plus'} \ a (\text{aval'' } e_1 \ S) \ (\text{aval'' } e_2 \ S) \\
\text{in } \text{inv_aval'} e_1 \ a_1 (\text{inv_aval'} e_2 \ a_2 \ S)) \]

(Analogously for all other arithmetic operations)
Backward analysis of $bexp$

Given

$b :: bexp$

$\text{res} :: bool$ (the intended value of $b$)

$S :: \text{'av st option}$

restrict $S$ to some $S' \leq S$ such that

$$\{s \in \gamma_o S. \ bval b s = \text{res}\} \subseteq \gamma_o S'$$

$\gamma_s S'$ overapproximates the subset of $\gamma_s S$ that makes $b$ evaluate to $\text{res}$. 
\[\text{inv\_bval}' :: \]
\[bexp \to \text{bool} \to \text{'av st option} \to \text{'av st option}\]
\[\text{inv\_bval}' (Bc v) \text{ res } S = (\text{if } v = \text{ res then } S \text{ else None })\]
\[\text{inv\_bval}' (\text{Not } b) \text{ res } S = \text{inv\_bval}' b (\neg \text{ res}) S\]
\[\text{inv\_bval}' (\text{And } b_1 b_2) \text{ res } S = \]
\[\text{if res then } \text{inv\_bval}' b_1 \text{ True } (\text{inv\_bval}' b_2 \text{ True } S) \]
\[\text{else } \text{inv\_bval}' b_1 \text{ False } S \sqcup \text{inv\_bval}' b_2 \text{ False } S\]
\[\text{inv\_bval}' (\text{Less } e_1 e_2) \text{ res } S = \]
\[\text{let } (a_1, a_2) = \text{inv\_less'} \text{ res } (\text{aval}'' e_1 S) (\text{aval}'' e_2 S) \]
\[\text{in } \text{inv\_aval'} e_1 a_1 (\text{inv\_aval'} e_2 a_2 S)\]
A further extension of the interface of our framework:

\[ \text{inv\_less'} :: \text{bool} \Rightarrow \text{'av} \Rightarrow \text{'av} \Rightarrow \text{'av} \times \text{'av} \]

Assumption:

\[
\text{inv\_less'} \ res \ a_1 \ a_2 = (a_1', a_2') \implies \\
\forall a_1' \exists a_2' \in \gamma a_1. (i_1 < i_2) = \text{res} \\
\forall a_2' \exists a_1' \in \gamma a_2. (i_1 < i_2) = \text{res}
\]
Example: intervals, informally

\[\text{inv\_plus'} \ [0, 4] [10, 20] [-10, 0] = ([10, 14], [-10, -6])\]

\[\text{inv\_less'} \ True \ [0, 20] [-5, 5] = ([0, 4], [1, 5])\]

\[\text{inv\_bval'} (x + y < z) \ True\]

\[\{x \mapsto [10, 20], y \mapsto [-10, 0], z \mapsto [-5, 5]\}:\]

\[\text{inv\_aval'} z \ [1, 5] \ \{\bullet\} = \{\bullet, z \mapsto [1, 5]\}\]

\[\text{inv\_aval'} (x + y) \ [0, 4] \ \{\bullet\}:\]

\[\text{inv\_aval'} y \ [-10, -6] \ \{\bullet\} = \{\bullet, y \mapsto [-10, -6], \bullet\}\]

\[\text{inv\_aval'} x \ [10, 14] \ \{\bullet\} = \{x \mapsto [10, 14], y \mapsto [-10, -6], z \mapsto [1, 5]\}\]
Before: \[ \text{step}' = \text{Step} \ asem \ (\lambda b \ S. \ S) \]
Now: \[ \text{step}' = \text{Step} \ asem \ (\lambda b. \ \text{inv\_bval}' \ b \ \text{True}) \]
Correctness proof

Almost as before, but with correctness lemmas for $inv\_aval'$

$$\{ s \in \gamma_o S. \ aval \ e \ s \in \gamma \ a \} \subseteq \gamma_o (inv\_aval' \ e \ a \ S)$$

and $inv\_bval'$:

$$\{ s \in \gamma_o S. \ bv = bval \ b \ s \} \subseteq \gamma_o (inv\_bval' \ b \ bv \ S)$$
Extended interface to abstract interpreter:

- `'av :: bounded_lattice`
  \[ \gamma \bot = \{\} \text{ and } \gamma a_1 \cap \gamma a_2 \subseteq \gamma (a_1 \cap a_2) \]

- `test_num' :: int ⇒ 'av ⇒ bool`
  `test_num' i a = (i ∈ \gamma a)`

- `inv_plus' :: 'av ⇒ 'av ⇒ 'av ⇒ 'av × 'av`
  \[ [inv_plus' a a_1 a_2] = (a_1', a_2'); \]
  \[ i_1 ∈ \gamma a_1; i_2 ∈ \gamma a_2; i_1 + i_2 ∈ \gamma a] \]
  \[ \implies i_1 ∈ \gamma a_1' \land i_2 ∈ \gamma a_2' \]

- `inv_less' :: bool ⇒ 'av ⇒ 'av ⇒ 'av × 'av`
  \[ [inv_less' (i_1 < i_2) a_1 a_2] = (a_1', a_2'); \]
  \[ i_1 ∈ \gamma a_1; i_2 ∈ \gamma a_2] \]
  \[ \implies i_1 ∈ \gamma a_1' \land i_2 ∈ \gamma a_2' \]
Abs_Int2_ivl.thy
The problem

If there are infinite ascending \( \leq \) chains of abstract values then the abstract interpreter may not terminate.

Canonical example: intervals

\[
[0,0] \leq [0,1] \leq [0,2] \leq [0,3] \leq \ldots
\]

Can happen even if the program terminates!
Widening

- $x_0 = \bot$, $x_{i+1} = f(x_i)$ may not terminate while searching for a pfp: $f(x_i) \leq x_i$
- Widen in each step: $x_{i+1} = x_i \nabla f(x_i)$ until a pfp is found.
- We assume
  - $\nabla$ “extrapolates” its arguments: $x, y \leq x \nabla y$
  - $\nabla$ “jumps” far enough to prevent nontermination
Example: Widening on (non-empty) intervals

\[
[l_1, h_1] \triangledown [l_2, h_2] = [l, h]
\]

where

\[
l = (\text{if } l_1 > l_2 \text{ then } -\infty \text{ else } l_1)
\]
\[
h = (\text{if } h_1 < h_2 \text{ then } \infty \text{ else } h_1)
\]
Warning

- $x_{i+1} = f(x_i)$ finds a least pfp if it terminates, $f$ is monotone, and $x_0 = \bot$
- $x_{i+1} = x_i \nabla f(x_i)$ may return any pfp in the worst case $\top$

We win termination, we lose precision
iteration of $f$

iteration with widening

narrowing iteration of $f$
A **widening operator** $\nabla :: 'a \Rightarrow 'a \Rightarrow 'a$ on a preorder must satisfy $x \leq x \nabla y$ and $y \leq x \nabla y$.

Widening operators can be extended from $'a$ to $'a$ st, $'a$ option and $'a$ acom.
Abstract interpretation with widening

New assumption: 

\[ \text{'av has widening operator} \]

\[
\text{iter_widen :: ('a ⇒ 'a) ⇒ 'a ⇒ 'a option}
\]

\[
\text{iter_widen f =}
\]

\[
\text{while_option (λx. ¬ f x ≤ x) (λx. x ▽ f x)}
\]

Correctness (returns pfp): by definition

Abstract interpretation of \( c \):

\[
\text{iter_widen (step' ⊤) (bot c)}
\]
x ::= N 0 \{A_0\};
\{A_1\}
WHILE Less (V x) (N 100)
DO \{A_2\}
\quad x ::= Plus (V x) (N 1) \{A_3\}
\{A_4\}
Narrowing

Widening returns a (potentially) imprecise pfp $p$.

If $f$ is monotone, further iteration improves $p$:

$$p \geq f(p) \geq f^2(p) \geq \ldots$$

and each $f^i(p)$ is still a pfp!

- need not terminate: $[0, \infty] \geq [1, \infty] \geq \ldots$
- but we can stop at any point!
A narrowing operator $\triangle ::= \ 'a \Rightarrow \ 'a \Rightarrow \ 'a$

must satisfy $y \leq x \implies y \leq x \triangle y \leq x$.

**Lemma** Let $f$ be monotone. If $f \ p \leq p$ then $f(p \triangle f \ p) \leq p \triangle f \ p \leq p$

iter_narrow $f \ p =$

while_option ($\lambda x. \ x \triangle f \ x < x$) ($\lambda x. \ x \triangle f \ x$) $p$

If $f$ is monotone and $p$ a pfp of $f$ and the loop terminates, then (by the lemma) we obtain a pfp of $f$ below $p$.

Iteration as long as progress is made: $x \triangle f \ x < x$
Example: Narrowing on (non-empty) intervals

\[ [l_1, h_1] \bigtriangleup [l_2, h_2] = [l, h] \]

where

\[
  l = (\text{if } l_1 = -\infty \text{ then } l_2 \text{ else } l_1)
\]

\[
  h = (\text{if } h_1 = \infty \text{ then } h_2 \text{ else } h_1)
\]
Abstract interpretation with widening & narrowing

New assumption: 'av also has a narrowing operator

\[ pfp\_wn \ f \ x = \]
\[ (\text{case } \text{iter\_widen} \ f \ x \ \text{of } \text{None} \Rightarrow \text{None}
| \text{Some } p \Rightarrow \text{iter\_narrow} \ f \ p) \]

\[ AI\_wn \ c = pfp\_wn \ (\text{step'} \bot) \ (\text{bot } c) \]

**Theorem** \( AI\_wn \ c = \text{Some } C \implies CS \ c \leq \gamma_c \ C \)

**Proof** as before
Termination

of

\( \text{while\_option} \ (\lambda x. \ P \ x) \ (\lambda x. \ g \ x) \)

via measure function \( m \)
such that \( m \) goes down with every iteration:

\[
P \ x \implies m \ x > m(g \ x)
\]

May need some invariant \( \text{Inv} \) as additional premise:

\[
\text{Inv} \ x \implies P \ x \implies m \ x > m(g \ x)
\]
Termination of iter_widen

iter_widen \( f = \)
while_option (λx. ¬ f x ≤ x) (λx. x ▽ f x)

As before (almost): Assume \( m :: 'av ⇒ \text{nat} \) and \( h :: \text{nat} \) such that \( m x ≤ h \) and \( x ≤ y \) \( ⇒ \) \( m y ≤ m x \)
and additionally \( ¬ y ≤ x \) \( ⇒ \) \( m (x ▽ y) < m x \)

Define the same functions \( m_s/m_o/m_c \) as before.

Termination of iter_widen on 'a st option acom:

Lemma \( ¬ C_2 ≤ C_1 ⇒ m_c (C_1 ▽ C_2) < m_c C_1 \)
if top_onc \( C_1 \) (− vars \( C_1 \)), top_onc \( C_2 \) (− vars \( C_2 \))
and strip \( C_1 = \) strip \( C_2 \)
Termination of \textit{iter\_narrow}

\[
\text{iter\_narrow } f = \text{while\_option } (\lambda x. x \triangle f x < x) (\lambda x. x \triangle f x)
\]

Assume \( n :: 'av \Rightarrow \text{nat} \) such that
\[
[y \leq x; x \triangle y < x] \Rightarrow n (x \triangle y) < n x
\]

Define \( n_s/n_o/n_c \) like \( m_s/m_o/m_c \)

Termination of \textit{iter\_narrow on 'a st option acom}:

\textbf{Lemma} \[
[y \leq x; x \triangle y < x] \Rightarrow n_c (C_1 \triangle C_2) < n_c C_1 \quad \text{if} \quad \text{strip } C_1 = \text{strip } C_2,
\]
\[
top_{\text{on}_c} C_1 (\neg \text{vars } C_1) \quad \text{and} \quad top_{\text{on}_c} C_2 (\neg \text{vars } C_2)
\]
Measuring non-empty intervals

\[ m \left[ l, h \right] = (\text{if } l = -\infty \text{ then } 0 \text{ else } 1) + (\text{if } h = \infty \text{ then } 0 \text{ else } 1) \]

\[ h = 2 \]

\[ n \text{ ivl} = 2 - m \text{ ivl} \]