Short Lecture 13
Predicate Abstraction and Intervals

2015

Predicate Abstraction

Abstract interpretation domain (lattice) is determined by a set of formulas (predicates) $\mathcal{P}$ on program variables.
Example: $\mathcal{P} = \{P_0, P_1, P_2, P_3\}$ where

$$\begin{array}{lcl} P_0 & \equiv & \text{false} \\ P_1 & \equiv & 0 < x \\ P_2 & \equiv & 0 < y \\ P_3 & \equiv & x < y \end{array}$$

Analysis tries to construct invariants from these predicates using

- conjunctions, e.g. $P_1 \wedge P_3$ (our focus here, for simplicity)
- conjunctions and disjunctions, e.g. $P_3 \wedge (P_1 \vee P_2)$

Predicate Abstraction

Abstract interpretation domain (lattice) is determined by a set of formulas (predicates) $\mathcal{P}$ on program variables.
Example: $\mathcal{P} = \{P_0, P_1, P_2, P_3\}$ where

$$\begin{array}{lcl} P_0 & \equiv & \text{false} \\ P_1 & \equiv & 0 < x \\ P_2 & \equiv & 0 < y \\ P_3 & \equiv & x < y \end{array}$$

Analysis tries to construct invariants from these predicates using

- conjunctions, e.g. $P_1 \wedge P_3$ (our focus here, for simplicity)
- conjunctions and disjunctions, e.g. $P_3 \wedge (P_1 \vee P_2)$

We assume $P_0 \equiv$ false, other predicates in $\mathcal{P}$ - arbitrary

- expressed in logic of some theorem prover (e.g. SMT solver)

Example of Analysis Result

$\mathcal{P} = \{\mathit{false}, 0 < x, 0 <= x, 0 < y, x < y, x = 0, y = 1, x < 1000, 1000 \leq x\}$

```
x = 0;
y = 1;
// 0<y, x<y,x=0,y=1, x<1000
while // 0<y, 0≤x, x<y
(x < 1000) {
   // 0<y, 0≤x, x<y, x<1000
   x = x + 1;
   // 0<y, 0≤x, 0<x
   y = 2∗x;
   // 0<y, 0≤x, 0<x, x<y
   y = y + 1;
   // 0<y, 0≤x, 0<x, x<y
   print(y);
}
// 0<y, 0≤x, x<y, 1000 ≤ x
```

Example of Analysis Result

$\mathcal{P} = \{false, 0 < x, 0 <= x, 0 < y, x < y, x = 0, y = 1, x < 1000, 1000 \leq x\}$

```
x = 0;
y = 1;
// 0<y, x<y,x=0,y=1, x<1000
while // 0<y, 0≤x, x<y
(x < 1000) {
   // 0<y, 0≤x, x<y, x<1000
   x = x + 1;
   // 0<y, 0≤x, 0<x
   y = 2*x;
   // 0<y, 0≤x, 0<x, x<y
   y = y + 1;
   // 0<y, 0≤x, 0<x, x<y
   print(y);
}
// 0<y, 0≤x, x<y, 1000 ≤ x
```

Start by assuming all predicates hold in all non-entry points.

Example of Analysis Result

$\mathcal{P} = \{\mathit{false}, 0 < x, 0 <= x, 0 < y, x < y, x = 0, y = 1, x < 1000, 1000 \leq x\}$

```
x = 0;
y = 1;
// 0<y, x<y,x=0,y=1, x<1000
while // 0<y, 0≤x, x<y
(x < 1000) {
   // 0<y, 0≤x, x<y, x<1000
   x = x + 1;
   // 0<y, 0≤x, 0<x
   y = 2∗x;
   // 0<y, 0≤x, 0<x, x<y
   y = y + 1;
   // 0<y, 0≤x, 0<x, x<y
   print(y);
}
// 0<y, 0≤x, x<y, 1000 ≤ x
```

Start by assuming all predicates hold in all non-entry points.

Check Hoare triples, remove predicates from postcondition that do not hold

Lattice of Conjunctions of Predicates and Concretization

$\mathcal{P} = \{P_0, P_1, \ldots, P_n\}$ - predicates

- formulas whose free variables denote program variables

$L = A = 2^{\mathcal{P}}$, so for $a \in A$ we have $a \subseteq \mathcal{P}$
Example: $a_0 = \{0 < x, x < y\}$.
$s \models F$ means: formula F is true for variables given by the program state s

$$\gamma(a) = \{s \mid s \models \bigwedge_{P \in a} P\}$$

Shorthand: $\bigwedge a$ means $\bigwedge_{P \in a} P$
Example: $\gamma(a_0) =$

Lattice of Conjunctions of Predicates and Concretization

$\mathcal{P} = \{P_0, P_1, \ldots, P_n\}$ - predicates

- formulas whose free variables denote program variables

$L = A = 2^{\mathcal{P}}$, so for $a \in A$ we have $a \subseteq \mathcal{P}$
Example: $a_0 = \{0 < x, x < y\}$.
$s \models F$ means: formula F is true for variables given by the program state s

$$\gamma(a) = \{s \mid s \models \bigwedge_{P \in a} P\}$$

Shorthand: $\bigwedge a$ means $\bigwedge_{P \in a} P$
Example: $\gamma(a_0) = \{s \mid s \models 0 < x \wedge x < y\}$.

Lattice of Conjunctions of Predicates and Concretization

$\mathcal{P} = \{P_0, P_1, \ldots, P_n\}$ - predicates

- formulas whose free variables denote program variables

$L = A = 2^{\mathcal{P}}$, so for $a \in A$ we have $a \subseteq \mathcal{P}$
Example: $a_0 = \{0 < x, x < y\}$.
$s \models F$ means: formula F is true for variables given by the program state s

$$\gamma(a) = \{s \mid s \models \bigwedge_{P \in a} P\}$$

Shorthand: $\bigwedge a$ means $\bigwedge_{P \in a} P$
Example: $\gamma(a_0) = \{s \mid s \models 0 < x \wedge x < y\}$. We often assume states are pairs (x, y). Then $\gamma(a_0) =$

Lattice of Conjunctions of Predicates and Concretization

$\mathcal{P} = \{P_0, P_1, \ldots, P_n\}$ - predicates

- formulas whose free variables denote program variables

$L = A = 2^{\mathcal{P}}$, so for $a \in A$ we have $a \subseteq \mathcal{P}$
Example: $a_0 = \{0 < x, x < y\}$.
$s \models F$ means: formula F is true for variables given by the program state s

$$\gamma(a) = \{s \mid s \models \bigwedge_{P \in a} P\}$$

Shorthand: $\bigwedge a$ means $\bigwedge_{P \in a} P$
Example: $\gamma(a_0) = \{s \mid s \models 0 < x \wedge x < y\}$. We often assume states are pairs (x, y). Then $\gamma(a_0) = \{(x, y) \mid 0 < x \wedge x < y\}$. $\gamma(a) = \{(x, y) \mid \bigwedge a\}$

Lattice of Conjunctions of Predicates and Concretization

$\mathcal{P} = \{P_0, P_1, \ldots, P_n\}$ - predicates

- formulas whose free variables denote program variables

$L = A = 2^{\mathcal{P}}$, so for $a \in A$ we have $a \subseteq \mathcal{P}$
Example: $a_0 = \{0 < x, x < y\}$.
$s \models F$ means: formula F is true for variables given by the program state s

$$\gamma(a) = \{s \mid s \models \bigwedge_{P \in a} P\}$$

Shorthand: $\bigwedge a$ means $\bigwedge_{P \in a} P$
Example: $\gamma(a_0) = \{s \mid s \models 0 < x \land x < y\}$. We often assume states are pairs (x, y). Then $\gamma(a_0) = \{(x, y) \mid 0 < x \land x < y\}$. $\gamma(a) = \{(x, y) \mid \bigwedge a\}$
If $a_1 \subseteq a_2$ then $\bigwedge a_2$ implies $\bigwedge a_1$, so $\gamma(a_2) \subseteq \gamma(a_1)$.

Lattice of Conjunctions of Predicates and Concretization

$\mathcal{P} = \{P_0, P_1, \ldots, P_n\}$ - predicates

- formulas whose free variables denote program variables

$L = A = 2^{\mathcal{P}}$, so for $a \in A$ we have $a \subseteq \mathcal{P}$
Example: $a_0 = \{0 < x, x < y\}$.
$s \models F$ means: formula F is true for variables given by the program state s

$$\gamma(a) = \{s \mid s \models \bigwedge_{P \in a} P\}$$

Shorthand: $\bigwedge a$ means $\bigwedge_{P \in a} P$
Example: $\gamma(a_0) = \{s \mid s \models 0 < x \land x < y\}$. We often assume states are pairs (x, y). Then $\gamma(a_0) = \{(x, y) \mid 0 < x \land x < y\}$. $\gamma(a) = \{(x, y) \mid \bigwedge a\}$
If $a_1 \subseteq a_2$ then $\bigwedge a_2$ implies $\bigwedge a_1$, so $\gamma(a_2) \subseteq \gamma(a_1)$.
Define:

$$a_1 \sqsubseteq a_2 \quad \iff \quad a_2 \subseteq a_1$$

Lemma: $a_1 \sqsubseteq a_2 \rightarrow \gamma(a_1) \subseteq \gamma(a_2)$

Lattice of Conjunctions of Predicates and Concretization

$\mathcal{P} = \{P_0, P_1, \ldots, P_n\}$ - predicates

- formulas whose free variables denote program variables

$L = A = 2^{\mathcal{P}}$, so for $a \in A$ we have $a \subseteq \mathcal{P}$
Example: $a_0 = \{0 < x, x < y\}$.
$s \models F$ means: formula F is true for variables given by the program state s

$$\gamma(a) = \{s \mid s \models \bigwedge_{P \in a} P\}$$

Shorthand: $\bigwedge a$ means $\bigwedge_{P \in a} P$
Example: $\gamma(a_0) = \{s \mid s \models 0 < x \wedge x < y\}$. We often assume states are pairs (x, y). Then $\gamma(a_0) = \{(x, y) \mid 0 < x \wedge x < y\}$. $\gamma(a) = \{(x, y) \mid \bigwedge a\}$
If $a_1 \subseteq a_2$ then $\bigwedge a_2$ implies $\bigwedge a_1$, so $\gamma(a_2) \subseteq \gamma(a_1)$.
Define:

$$a_1 \sqsubseteq a_2 \quad \iff \quad a_2 \subseteq a_1$$

Lemma: $a_1 \sqsubseteq a_2 \rightarrow \gamma(a_1) \subseteq \gamma(a_2)$
Does the converse hold?

Lattice of Conjunctions of Predicates and Concretization

$\mathcal{P} = \{P_0, P_1, \ldots, P_n\}$ - predicates

- formulas whose free variables denote program variables

$L = A = 2^{\mathcal{P}}$, so for $a \in A$ we have $a \subseteq \mathcal{P}$
Example: $a_0 = \{0 < x, x < y\}$.
$s \models F$ means: formula F is true for variables given by the program state s

$$\gamma(a) = \{s \mid s \models \bigwedge_{P \in a} P\}$$

Shorthand: $\bigwedge a$ means $\bigwedge_{P \in a} P$
Example: $\gamma(a_0) = \{s \mid s \models 0 < x \wedge x < y\}$. We often assume states are pairs (x, y). Then $\gamma(a_0) = \{(x, y) \mid 0 < x \wedge x < y\}$. $\gamma(a) = \{(x, y) \mid \bigwedge a\}$
If $a_1 \subseteq a_2$ then $\bigwedge a_2$ implies $\bigwedge a_1$, so $\gamma(a_2) \subseteq \gamma(a_1)$.
Define:

$$a_1 \sqsubseteq a_2 \iff a_2 \subseteq a_1$$

Lemma: $a_1 \sqsubseteq a_2 \rightarrow \gamma(a_1) \subseteq \gamma(a_2)$
Does the converse hold? no (will see examples later)

Lattice Operations: Example

$$\{x < 2, 0 < x, x < y\} \sqsubseteq \{0 < x, 0 < y\} \sqsubseteq \{0 < x\} \sqsubseteq \emptyset$$

Draw the Hasse diagram for any lattice $(A, \sqsubseteq)$ i.e. $(2^{\mathcal{P}}, \supseteq)$ for $\mathcal{P} = \{P_0, P_1, P_2\}$ a three-element set.

Lattice Operations: Example

$$\{x < 2, 0 < x, x < y\} \sqsubseteq \{0 < x, 0 < y\} \sqsubseteq \{0 < x\} \sqsubseteq \emptyset$$

Draw the Hasse diagram for any lattice $(A, \sqsubseteq)$ i.e. $(2^{\mathcal{P}}, \supseteq)$ for $\mathcal{P} = \{P_0, P_1, P_2\}$ a three-element set.

What is the top and what is the bottom element of this lattice?

Lattice Operations: Example

$$\{x < 2, 0 < x, x < y\} \sqsubseteq \{0 < x, 0 < y\} \sqsubseteq \{0 < x\} \sqsubseteq \emptyset$$

Draw the Hasse diagram for any lattice $(A, \sqsubseteq)$ i.e. $(2^{\mathcal{P}}, \supseteq)$ for $\mathcal{P} = \{P_0, P_1, P_2\}$ a three-element set.

What is the top and what is the bottom element of this lattice?
What is $\sqcup$? Compute $\{0 < x, x < 2\} \sqcup \{0 < x, x < y\} =$

Lattice Operations: Example

$$\{x < 2, 0 < x, x < y\} \sqsubseteq \{0 < x, 0 < y\} \sqsubseteq \{0 < x\} \sqsubseteq \emptyset$$

Draw the Hasse diagram for any lattice $(A, \sqsubseteq)$ i.e. $(2^{\mathcal{P}}, \supseteq)$ for $\mathcal{P} = \{P_0, P_1, P_2\}$ a three-element set.

What is the top and what is the bottom element of this lattice?
What is $\sqcup$? Compute $\{0 < x, x < 2\} \sqcup \{0 < x, x < y\} = \{0 < x\}$ (draw)

Lattice Operations: Example

$$\{x < 2, 0 < x, x < y\} \sqsubseteq \{0 < x, 0 < y\} \sqsubseteq \{0 < x\} \sqsubseteq \emptyset$$

Draw the Hasse diagram for any lattice $(A, \sqsubseteq)$ i.e. $(2^{\mathcal{P}}, \supseteq)$ for $\mathcal{P} = \{P_0, P_1, P_2\}$ a three-element set.

What is the top and what is the bottom element of this lattice?
What is $\sqcup$? Compute $\{0 < x, x < 2\} \sqcup \{0 < x, x < y\} = \{0 < x\}$ (draw)
What is the size and the height of the lattice?

Lattice Height

A finite chain inside a partial order is a strictly ordered sequence of elements: $x_0 \sqsubset x_1 \ldots \sqsubset x_n$
Here $x \sqsubset y$ means that both $x \sqsubseteq y$ and $x \neq y$
Length of such chain is n (number of $\sqsubset$ signs)
Note that $x_i \sqsubset x_j$ for $i < j$ by transitivity
Thus all elements in a chain are distinct.
An infinite chain is infinite sequence of elements where $x_i \sqsubset x_{i+1}$ for all i
A lattice is finite-height if all chains are finite. Then the maximum length of chains is called the height of the lattice.

Lattice of Predicates: Basic Facts

$\mathcal{P} = \{P_0, P_1, \ldots, P_n\}$. Lattice elements $a \in 2^{\mathcal{P}}$ (subsets of $\mathcal{P}$)
More predicates in conjunction $\Rightarrow$ stronger condition $\Rightarrow$ smaller set
Therefore we have:

- $\perp$ - bottom (smallest set of states) is:

Lattice of Predicates: Basic Facts

$\mathcal{P} = \{P_0, P_1, \ldots, P_n\}$. Lattice elements $a \in 2^{\mathcal{P}}$ (subsets of $\mathcal{P}$)
More predicates in conjunction $\Rightarrow$ stronger condition $\Rightarrow$ smaller set
Therefore we have:

- $\perp$ - bottom (smallest set of states) is: $\mathcal{P}$

Lattice of Predicates: Basic Facts

$\mathcal{P} = \{P_0, P_1, \ldots, P_n\}$. Lattice elements $a \in 2^{\mathcal{P}}$ (subsets of $\mathcal{P}$)
More predicates in conjunction $\Rightarrow$ stronger condition $\Rightarrow$ smaller set
Therefore we have:

- $\perp$ - bottom (smallest set of states) is: $\mathcal{P}$
- $\top$ - top (largest set of states) is:

Lattice of Predicates: Basic Facts

$\mathcal{P} = \{P_0, P_1, \ldots, P_n\}$. Lattice elements $a \in 2^{\mathcal{P}}$ (subsets of $\mathcal{P}$)
More predicates in conjunction $\Rightarrow$ stronger condition $\Rightarrow$ smaller set
Therefore we have:

- $\bot$ - bottom (smallest set of states) is: $\mathcal{P}$
- $\top$ - top (largest set of states) is: $\emptyset$ (predicate 'true')

Lattice of Predicates: Basic Facts

$\mathcal{P} = \{P_0, P_1, \ldots, P_n\}$. Lattice elements $a \in 2^{\mathcal{P}}$ (subsets of $\mathcal{P}$)
More predicates in conjunction $\Rightarrow$ stronger condition $\Rightarrow$ smaller set
Therefore we have:

- $\bot$ - bottom (smallest set of states) is: $\mathcal{P}$
- $\top$ - top (largest set of states) is: $\emptyset$ (predicate 'true')
- $\sqcup$ (approximates union) is:

Lattice of Predicates: Basic Facts

$\mathcal{P} = \{P_0, P_1, \ldots, P_n\}$. Lattice elements $a \in 2^{\mathcal{P}}$ (subsets of $\mathcal{P}$)
More predicates in conjunction $\Rightarrow$ stronger condition $\Rightarrow$ smaller set
Therefore we have:

- $\perp$ - bottom (smallest set of states) is: $\mathcal{P}$
- $\top$ - top (largest set of states) is: $\emptyset$ (predicate 'true')
- $\sqcup$ (approximates union) is: $\cap$

Lattice of Predicates: Basic Facts

$\mathcal{P} = \{P_0, P_1, \ldots, P_n\}$. Lattice elements $a \in 2^{\mathcal{P}}$ (subsets of $\mathcal{P}$)
More predicates in conjunction $\Rightarrow$ stronger condition $\Rightarrow$ smaller set
Therefore we have:

- $\perp$ - bottom (smallest set of states) is: $\mathcal{P}$
- $\top$ - top (largest set of states) is: $\emptyset$ (predicate 'true')
- $\sqcup$ (approximates union) is: $\cap$
- Size of the lattice with $n+1$ element:

Lattice of Predicates: Basic Facts

$\mathcal{P} = \{P_0, P_1, \ldots, P_n\}$. Lattice elements $a \in 2^{\mathcal{P}}$ (subsets of $\mathcal{P}$)
More predicates in conjunction $\Rightarrow$ stronger condition $\Rightarrow$ smaller set
Therefore we have:

- $\perp$ - bottom (smallest set of states) is: $\mathcal{P}$
- $\top$ - top (largest set of states) is: $\emptyset$ (predicate 'true')
- $\sqcup$ (approximates union) is: $\cap$
- Size of the lattice with $n+1$ element: 2^{n+1}

Lattice of Predicates: Basic Facts

$\mathcal{P} = \{P_0, P_1, \ldots, P_n\}$. Lattice elements $a \in 2^{\mathcal{P}}$ (subsets of $\mathcal{P}$)
More predicates in conjunction $\Rightarrow$ stronger condition $\Rightarrow$ smaller set
Therefore we have:

- $\perp$ - bottom (smallest set of states) is: $\mathcal{P}$
- $\top$ - top (largest set of states) is: $\emptyset$ (predicate 'true')
- $\sqcup$ (approximates union) is: $\cap$
- Size of the lattice with $n+1$ element: 2^{n+1}
- Height of the lattice with $n+1$ element:

Lattice of Predicates: Basic Facts

$\mathcal{P} = \{P_0, P_1, \ldots, P_n\}$. Lattice elements $a \in 2^{\mathcal{P}}$ (subsets of $\mathcal{P}$)
More predicates in conjunction $\Rightarrow$ stronger condition $\Rightarrow$ smaller set
Therefore we have:

- $\bot$ - bottom (smallest set of states) is: $\mathcal{P}$
- $\top$ - top (largest set of states) is: $\emptyset$ (predicate 'true')
- $\sqcup$ (approximates union) is: $\cap$
- Size of the lattice with $n+1$ element: 2^{n+1}
- Height of the lattice with $n+1$ element: $n+1$

Lattice of Predicates: Basic Facts

$\mathcal{P} = \{P_0, P_1, \ldots, P_n\}$. Lattice elements $a \in 2^{\mathcal{P}}$ (subsets of $\mathcal{P}$)
More predicates in conjunction $\Rightarrow$ stronger condition $\Rightarrow$ smaller set
Therefore we have:

- $\perp$ - bottom (smallest set of states) is: $\mathcal{P}$
- $\top$ - top (largest set of states) is: $\emptyset$ (predicate 'true')
- $\sqcup$ (approximates union) is: $\cap$
- Size of the lattice with $n+1$ element: 2^{n+1}
- Height of the lattice with $n+1$ element: $n+1$

Given $a \in 2^{\mathcal{P}}$ we abbreviate $\bigwedge_{P \in a} P$ as $\bigwedge a$

Abstract Strongest Postcondition $sp^{\#}(a, c)$

Abstract strongest postcondition (= transfer function in data-flow analysis)
Consider a command c and a set of predicates $a \subseteq \mathcal{P}$, we define *abstract* strongest postcondition of a as the conjunction

$$sp^{\#}(a, c) = \{ P' \in \mathcal{P} \mid \ \{\bigwedge a\} c \{P'\} \ \}$$

all predicates from $\mathcal{P}$ that hold after c. Note that $\{\ldots\} c \{\ldots\}$ after "|" denotes a Hoare triple. $sp^{\#} : A \times \mathit{Commands} \to A$.
By conjunctivity of Hoare triple, the result denotes a valid postcondition:

$$\{\bigwedge a\} c \{\bigwedge sp^{\#}(a, c)\}$$

Thus $sp_F(\bigwedge a, c) \implies sp^{\#}(a, c)$ holds as sp_F is strongest. However, converse implication need not - abstract postcondition is only an over-approximation
This definition of $sp^{\#}(a, c)$ gives the strongest condition *that we can write* as a conjunction of the allowed predicates $\mathcal{P}$, whereas sp_F need not be expressible using $\mathcal{P}$

Example of Computing Abstract Strongest Postcondition

$\mathcal{P} = \{\mathit{false}, 0 < x, 0 < y, x < y\}$
Compute $sp^{\#}(\{0 < x\}, y := x + 1)$

Example of Computing Abstract Strongest Postcondition

$\mathcal{P} = \{false, 0 < x, 0 < y, x < y\}$
Compute $sp^{\#}(\{0 < x\}, y := x + 1)$
We can test for each predicate $P' \in \mathcal{P}$ whether

$$x > 0 \wedge (y' = x + 1 \wedge x' = x) \implies P'(x', y')$$

Example of Computing Abstract Strongest Postcondition

$\mathcal{P} = \{\mathit{false}, 0 < x, 0 < y, x < y\}$
Compute $sp^{\#}(\{0 < x\}, y := x + 1)$
We can test for each predicate $P' \in \mathcal{P}$ whether

$$x > 0 \land (y' = x + 1 \land x' = x) \implies P'(x', y')$$

We obtain that the condition holds for $0 < x$, $0 < y$, and for $x < y$, but not for *false*. Thus,

$$sp^{\#}(\{0 < x\}, y := x + 1) = \{0 < x, 0 < y, x < y\}$$

Example of Computing Abstract Strongest Postcondition

$\mathcal{P} = \{false, 0 < x, 0 < y, x < y\}$
Compute $sp^{\#}(\{0 < x\}, y := x + 1)$
We can test for each predicate $P' \in \mathcal{P}$ whether

$$x > 0 \wedge (y' = x + 1 \wedge x' = x) \implies P'(x', y')$$

We obtain that the condition holds for $0 < x$, $0 < y$, and for $x < y$, but not for *false*. Thus,

$$sp^{\#}(\{0 < x\}, y := x + 1) = \{0 < x, 0 < y, x < y\}$$

Compute

$$sp^{\#}(\{0 < x\}, y := x - 1) =$$

Example of Computing Abstract Strongest Postcondition

$\mathcal{P} = \{\mathit{false}, 0 < x, 0 < y, x < y\}$
Compute $sp^{\#}(\{0 < x\}, y := x + 1)$
We can test for each predicate $P' \in \mathcal{P}$ whether

$$x > 0 \land (y' = x + 1 \land x' = x) \implies P'(x', y')$$

We obtain that the condition holds for $0 < x$, $0 < y$, and for $x < y$, but not for *false*. Thus,

$$sp^{\#}(\{0 < x\}, y := x + 1) = \{0 < x, 0 < y, x < y\}$$

Compute

$$sp^{\#}(\{0 < x\}, y := x - 1) = \{0 < x\}$$

Example of Computing Abstract Strongest Postcondition

$\mathcal{P} = \{false, 0 < x, 0 < y, x < y\}$
Compute $sp^{\#}(\{0 < x\}, y := x + 1)$
We can test for each predicate $P' \in \mathcal{P}$ whether

$$x > 0 \wedge (y' = x + 1 \wedge x' = x) \implies P'(x', y')$$

We obtain that the condition holds for $0 < x$, $0 < y$, and for $x < y$, but not for *false*. Thus,

$$sp^{\#}(\{0 < x\}, y := x + 1) = \{0 < x, 0 < y, x < y\}$$

Compute

$$sp^{\#}(\{0 < x\}, y := x - 1) = \{0 < x\}$$
$$sp^{\#}(\{0 < x, x < y\}, x := x - 1) =$$

Example of Computing Abstract Strongest Postcondition

$\mathcal{P} = \{\mathit{false}, 0 < x, 0 < y, x < y\}$
Compute $\mathit{sp}^{\#}(\{0 < x\}, y := x + 1)$
We can test for each predicate $P' \in \mathcal{P}$ whether

$$x > 0 \wedge (y' = x + 1 \wedge x' = x) \implies P'(x', y')$$

We obtain that the condition holds for $0 < x$, $0 < y$, and for $x < y$, but not for *false*. Thus,

$$\mathit{sp}^{\#}(\{0 < x\}, y := x + 1) = \{0 < x, 0 < y, x < y\}$$

Compute

$$\mathit{sp}^{\#}(\{0 < x\}, y := x - 1) = \{0 < x\}$$
$$\mathit{sp}^{\#}(\{0 < x, x < y\}, x := x - 1) = \{0 < y, x < y\}$$

Example of Computing Abstract Strongest Postcondition

$\mathcal{P} = \{\mathit{false}, 0 < x, 0 < y, x < y\}$
Compute $sp^{\#}(\{0 < x\}, y := x + 1)$
We can test for each predicate $P' \in \mathcal{P}$ whether

$$x > 0 \wedge (y' = x + 1 \wedge x' = x) \implies P'(x', y')$$

We obtain that the condition holds for $0 < x$, $0 < y$, and for $x < y$, but not for *false*. Thus,

$$sp^{\#}(\{0 < x\}, y := x + 1) = \{0 < x, 0 < y, x < y\}$$

Compute

$$sp^{\#}(\{0 < x\}, y := x - 1) = \{0 < x\}$$
$$sp^{\#}(\{0 < x, x < y\}, x := x - 1) = \{0 < y, x < y\}$$

What is the relation between $\{0 < x, x < y\}$ and $\{0 < x, 0 < y, x < y\}$?

Example of Computing Abstract Strongest Postcondition

$\mathcal{P} = \{\mathit{false}, 0 < x, 0 < y, x < y\}$
Compute $sp^{\#}(\{0 < x\}, y := x + 1)$
We can test for each predicate $P' \in \mathcal{P}$ whether

$$x > 0 \land (y' = x + 1 \land x' = x) \implies P'(x', y')$$

We obtain that the condition holds for $0 < x$, $0 < y$, and for $x < y$, but not for *false*. Thus,

$$sp^{\#}(\{0 < x\}, y := x + 1) = \{0 < x, 0 < y, x < y\}$$

Compute

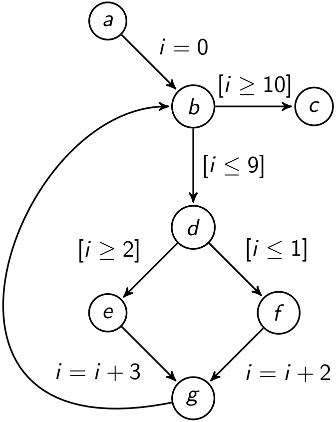
$$sp^{\#}(\{0 < x\}, y := x - 1) = \{0 < x\}$$
$$sp^{\#}(\{0 < x, x < y\}, x := x - 1) = \{0 < y, x < y\}$$

What is the relation between $\{0 < x, x < y\}$ and $\{0 < x, 0 < y, x < y\}$?
Different in lattice, denote same states. This is not a problem.

Control-Flow Graphs with Commands on Edges

Control-flow graphs are (V, E) where E contains triples (u, c, v) where $u, v \in V$ and each c is a command labeling the edge (u, v)

```
//a
i = 0;
    //b
while (i < 10) {
  //d
  if (i > 1)
    //e
    i = i + 3;
  else
    //f
    i = i + 2;
  //g
}
//c
```



Analysis Algorithm

Control-flow graph (V, E) where E contains triples (u, c, v)

Analysis maintains a map $g : V \to A$ from vertices to lattice elements

For program entry point: the set of all predicates that are true in every initial state (over-approximation of the set of initial states).

For other program points, initially put conjunction of all predicates ($\perp$). Then update the value at a point when any predecessor changes:

$$g(v) := \bigsqcup_{(u,c,v) \in E} sp^{\#}(g(u), c)$$

This process terminates since the lattice has finite height. Lattice elements grow (sets of predicates shrink).
Checking if the process terminates is same as checking that we have computed a loop invariant.

Running the Example from the Initial State

$$\mathcal{P} = \{\mathit{false}, 0 < x, 0 <= x, 0 < y, x < y, x = 0, y = 1, x < 1000, 1000 \leq x\}$$

```
// true
x = 0;
// false, 0<x, 0≤x, 0<y, x<y, x=0, y=1, x<1000,1000≤x
y = 1;
while // false, 0<x, 0≤x, 0<y, x<y, x=0, y=1, x<1000,1000≤x
(x < 1000) {
   // false, 0<x, 0≤x, 0<y, x<y, x=0, y=1, x<1000,1000≤x
   x = x + 1;
   // false, 0<x, 0≤x, 0<y, x<y, x=0, y=1, x<1000,1000≤x
   y = 2*x;
   // false, 0<x, 0≤x, 0<y, x<y, x=0, y=1, x<1000,1000≤x
   y = y + 1;
   // false, 0<x, 0≤x, 0<y, x<y, x=0, y=1, x<1000,1000≤x
}
// false, 0<x, 0≤x, 0<y, x<y, x=0, y=1, x<1000,1000≤x
```

Example of Limitations of Conjunctions

$\mathcal{P} = \{false, 0 < x, x \leq 0, 0 < y\}$

if (x $>$ 0) $\{$
  y $=$ x
$\}$
// Q
if (x $>$ 0) $\{$
  if(y $>$ 0) 1/y
  else error
$\}$

Assuming arbitrary initial state, what is the best we can compute as Q using conjunctions from $\mathcal{P}$?

Example of Limitations of Conjunctions

$\mathcal{P} = \{false, 0 < x, x \leq 0, 0 < y\}$

```
if (x > 0) {
  y = x
}
// Q
if (x > 0) {
  if(y > 0) 1/y
  else error
}
```

Assuming arbitrary initial state, what is the best we can compute as Q using conjunctions from $\mathcal{P}$? 'true'

Using disjunctions of conjunctions:

Example of Limitations of Conjunctions

$\mathcal{P} = \{false, 0 < x, x \leq 0, 0 < y\}$

```
if (x > 0) {
  y = x
}
// Q
if (x > 0) {
  if(y > 0) 1/y
  else error
}
```

Assuming arbitrary initial state, what is the best we can compute as Q using conjunctions from $\mathcal{P}$? 'true'

Using disjunctions of conjunctions: $(x > 0 \land y > 0) \lor (x \leq 0)$
Allows us to prove absence of error in the remaining code

Disjunctive Analysis to Overcome Limitations

Lattice with *disjunction* of conjunctions

- Sets of sets of predicates - exponentially larger
- Reduce by using as few predicates as possible, different possible predicates for each program point, limit sizes of conjuncts, ...

Important topic: automatically discover predicates.

- in general as hard as discovering loop invariants
- yet we only need to discover pieces of invariants
- and we can conservatively suggest more candidate predicates (any predicate set gives a sound analysis)

Interval Analysis

For a machine integer x, compute the interval $[a, b]$ such that $x \in [a, b]$
Worst-case interval: $[minI, maxI] = [-2^{31}, 2^{31} - 1]$

- each machine integer is between smallest and largest representable one

In addition, we introduce a special $\perp$ interval to represent an empty set of states
Consider relation c whose semantics is relation r on initial and final integer
Define $sp^{\#}$ as the interval for the values that x can take after c:

$$sp^{\#}([a, b], r) = \alpha(\{x' \mid x \in [a, b] \wedge (x, x') \in r\})$$

Here α computes the interval for a set of values:

- $\alpha(S) = [\min(S), \max(S)]$, if $S \neq \emptyset$, whereas $\alpha(\emptyset) = \perp$

We define $sp^{\#}(\perp) = \perp$, since image of empty set is an empty set

$sp^{\#}([0, 10], x{=}x + 7) = [7, 17]$ $\quad$ $sp^{\#}([-5, -5], x{=}x * x) = [0, 25]$
$sp^{\#}([1000, maxI], x = x + 30) = [minI, maxI]$

Size of the Interval Lattice

$L = \{\bot\} \cup \{[a, b] \mid minI \leq a \leq b \leq maxI\}$
Here $\bot \sqsubseteq [a, b]$ for all proper intervals $[a, b]$. Between intervals,

$$[a, b] \sqsubseteq [a', b'] \text{ if and only if } a' \leq a \leq b \leq b'$$

Number of elements in the lattice:

Size of the Interval Lattice

$L = \{\bot\} \cup \{[a, b] \mid minI \leq a \leq b \leq maxI\}$
Here $\bot \sqsubseteq [a, b]$ for all proper intervals $[a, b]$. Between intervals,

$$[a, b] \sqsubseteq [a', b'] \text{ if and only if } a' \leq a \leq b \leq b'$$

Number of elements in the lattice:$1 + \frac{2^{32}(2^{32}+1)}{2} = 1 + 2^{31} + 2^{63}$

Size of the Interval Lattice

$L = \{\perp\} \cup \{[a, b] \mid minI \leq a \leq b \leq maxI\}$
Here $\perp \sqsubseteq [a, b]$ for all proper intervals $[a, b]$. Between intervals,

$$[a, b] \sqsubseteq [a', b'] \text{ if and only if } a' \leq a \leq b \leq b'$$

Number of elements in the lattice: $1 + \frac{2^{32}(2^{32}+1)}{2} = 1 + 2^{31} + 2^{63}$
Size of the longest chain:

Size of the Interval Lattice

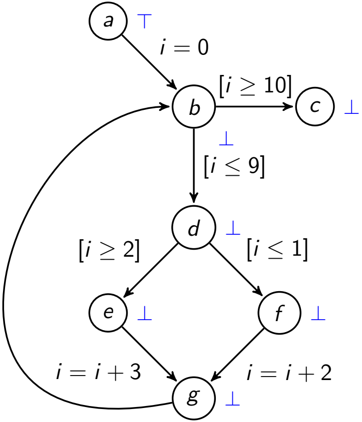
$L = \{\perp\} \cup \{[a, b] \mid minI \leq a \leq b \leq maxI\}$
Here $\perp \sqsubseteq [a, b]$ for all proper intervals $[a, b]$. Between intervals,

$$[a, b] \sqsubseteq [a', b'] \text{ if and only if } a' \leq a \leq b \leq b'$$

Number of elements in the lattice:$1 + \frac{2^{32}(2^{32}+1)}{2} = 1 + 2^{31} + 2^{63}$
Size of the longest chain: 2^{32}

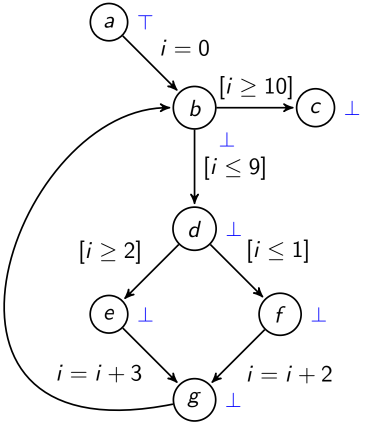
Starting Point for Analysis

```
//a
i = 0;
      //b
while (i < 10) {
   //d
   if (i > 1)
      //e
      i = i + 3;
   else
      //f
      i = i + 2;
   //g
}
//c
```



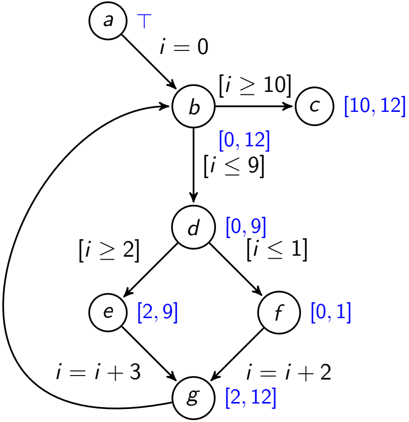
Starting Point for Analysis

```
//a
i = 0;
      //b
while (i < 10) {
   //d
   if (i > 1)
      //e
      i = i + 3;
   else
      //f
      i = i + 2;
   //g
}
//c
```



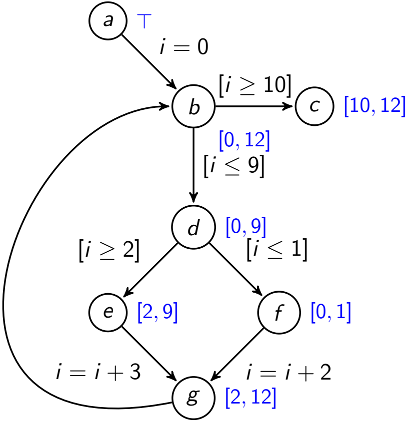
Fixpoint Found

```
//a
i = 0;
    //b
while (i < 10) {
  //d
  if (i > 1)
    //e
    i = i + 3;
  else
    //f
    i = i + 2;
  //g
}
//c
```



Fixpoint Found

```
//a
i = 0;
      //b
while (i < 10) {
  //d
  if (i > 1)
    //e
    i = i + 3;
  else
    //f
    i = i + 2;
  //g
}
//c
```



Note: in general, we maintain interval for each variable

General Remarks

Whatever we choose as our abstract domain A (typically some lattice), it is good to have a function γ that gives meaning to elements of A
Often, elements $a \in A$ represent sets of states:

- predicate abstraction: states that satisfy the conjunction of predicates
- interval analysis: states whose variables belong to the intervals

When we think about correctness conditions intuitively, we can "almost ignore" γ (but it is needed for statements to type check).
Each analysis is given by transfer functions such as $sp^{\#}$ which need to satisfy, for each $a \in A$:

$$sp(c, \gamma(a)) \subseteq \gamma(sp^{\#}(c, a))$$

$sp^{\#}$ gives a larger set of states (it finds only some, not all properties)

computed properties imply assertions of interest $\Rightarrow$ we proved the assertions
otherwise $\Rightarrow$ either assertions do not hold, or analysis was too conservative

Smaller and Larger Lattices

Simple data-flow analysis are often defined by first defining abstraction for one program variable (e.g. an interval). Let this be lattice $(L, \sqsubseteq)$

Then, we have one such value for each variable from set of variable names N. We obtain lattice

$$(L, \sqsubseteq)^N$$

Elements are functions $N \to L$. Ordering is point-wise

Finally, the analysis maintains such value for each program point V, so we have elements $V \to (N \to L)$