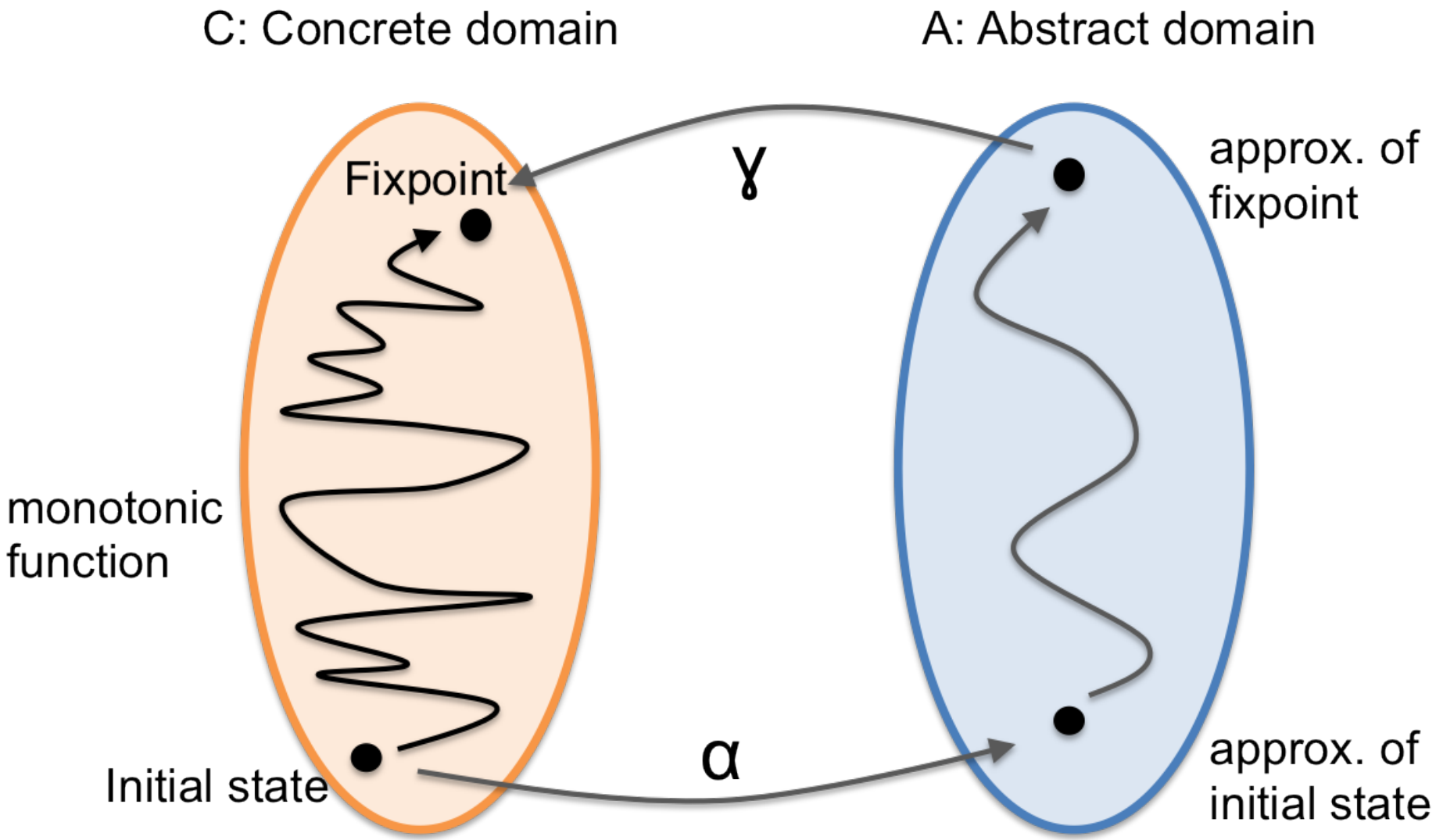
Lecturecise 14
From Lattices to Abstract Interpretation

2013

Abstract Interpretation Big Picture

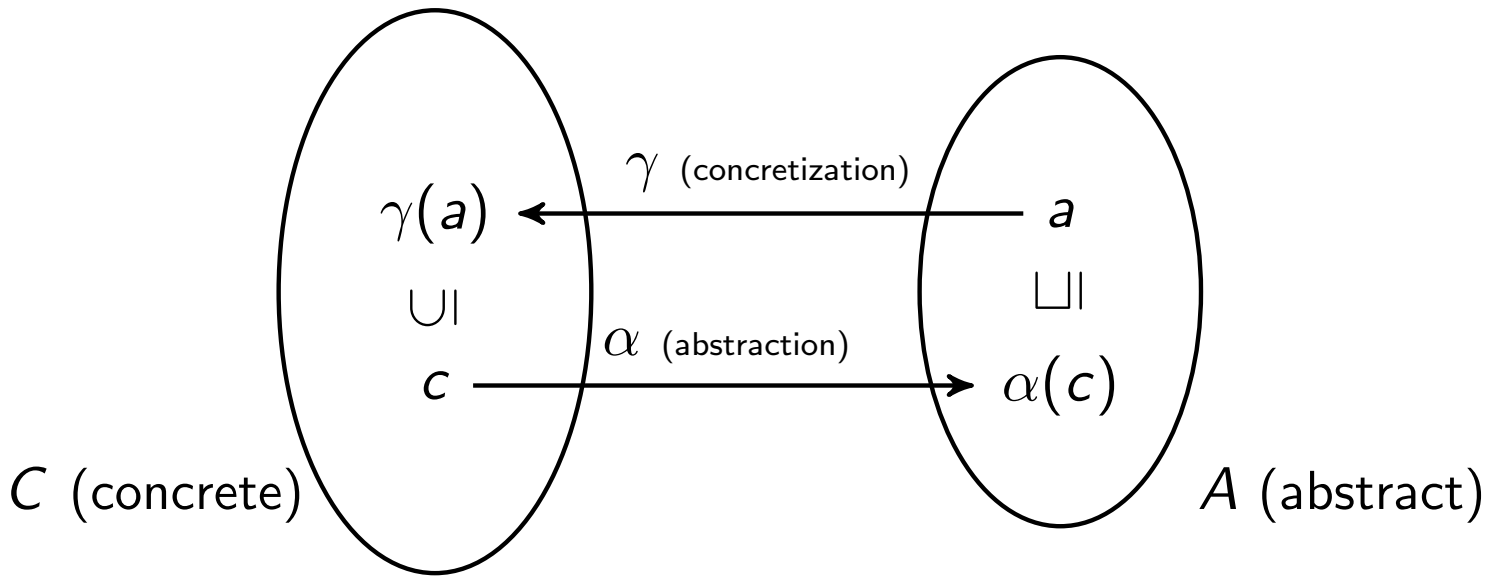


Galois Connection

Galois connection (named after Évariste Galois) is defined by two monotonic functions $\alpha : C \to A$ and $\gamma : A \to C$ between partial orders $(C, \subseteq)$ and $(A, \sqsubseteq)$ such that

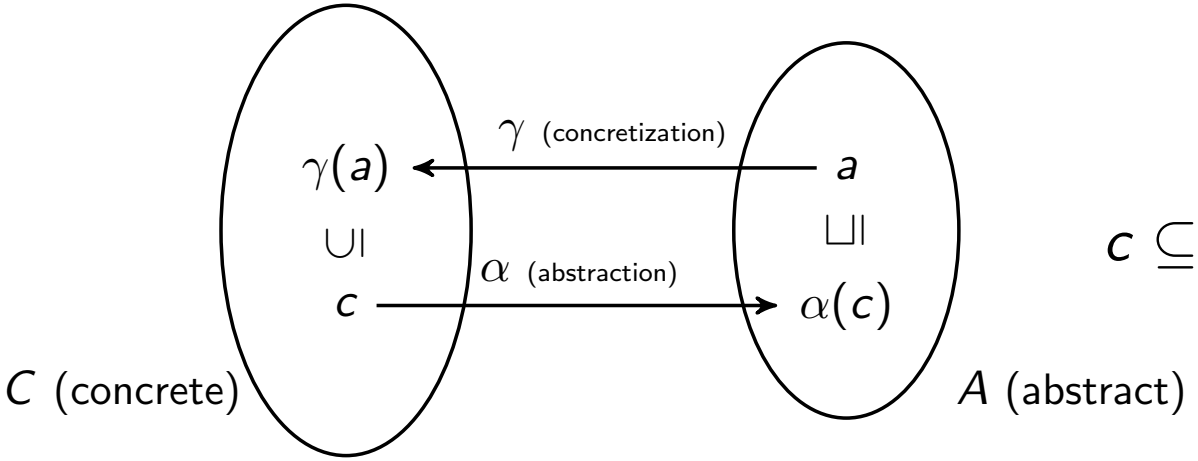
$$\forall c \in C. \forall a \in A. \qquad c \subseteq \gamma(a) \iff \alpha(c) \sqsubseteq a \qquad (*)$$

Intuitively, each side means that c is **over-approximated** by a.
"γ travels to the other side to become α and changes the ordering"



$\subseteq$ could be any partial order, though for us it will typically be actual subset

Example of Galois Connection



$$c \subseteq \gamma(a) \iff \alpha(c) \sqsubseteq a$$

$(C, \subseteq)$, $C = 2^{\mathbb{Z}}$ - set of sets of integers $\quad (A, \sqsubseteq)$, A - set of intervals

$$\begin{aligned} A &= \{\bot\} \cup \{[p,q] \mid p \in \{-\infty\} \cup \mathbb{Z}, q \in \mathbb{Z} \cup \{\infty\}, p \le q\} \\ \gamma([p,q]) &= \{x \in \mathbb{Z} \mid p \le x \le q\}, \quad \gamma(\bot) = \emptyset \\ a_1 \sqsubseteq a_2 &\overset{def}{\iff} \gamma(a_1) \subseteq \gamma(a_2) \\ \alpha(c) &= [\inf(c), \sup(c)], \text{ if } c \neq \emptyset \quad \alpha(\emptyset) = \bot \end{aligned}$$

Lemma:

1. $\sqsubseteq$ defined above is a partial order
2. α, γ are monotonic $c_1 \subseteq c_2 \to \alpha(c_1) \sqsubseteq \alpha(c_2)$
3. (α, γ) is a Galois connection

Galois Connection for Intervals

$$\begin{aligned} \gamma([p,q]) &= \{x \in \mathbb{Z} \mid p \leq x \leq q\}, \qquad \gamma(\bot) = \emptyset \\ \alpha(c) &= [\mathsf{inf}(c), \mathsf{sup}(c)], \text{ if } c \neq \emptyset \qquad \alpha(\emptyset) = \bot \end{aligned}$$

Goal: show $c \subseteq \gamma(a) \iff \alpha(c) \sqsubseteq a$

Consider the case where $c \in C$ is non-empty set and $a = [p, q]$, $p \leq q$

Galois Connection for Intervals

$$\begin{aligned} \gamma([p,q]) &= \{x \in \mathbb{Z} \mid p \leq x \leq q\}, \quad \gamma(\bot) = \emptyset \\ \alpha(c) &= [\inf(c), \sup(c)], \text{ if } c \neq \emptyset \quad \alpha(\emptyset) = \bot \end{aligned}$$

Goal: show $c \subseteq \gamma(a) \iff \alpha(c) \sqsubseteq a$

Consider the case where $c \in C$ is non-empty set and $a = [p, q]$, $p \leq q$

$$\begin{aligned} & c \subseteq \gamma(a) \\ \iff & \end{aligned}$$

Galois Connection for Intervals

$$\begin{aligned} \gamma([p,q]) &= \{x \in \mathbb{Z} \mid p \leq x \leq q\}, \quad \gamma(\bot) = \emptyset \\ \alpha(c) &= [\inf(c), \sup(c)], \text{ if } c \neq \emptyset \quad \alpha(\emptyset) = \bot \end{aligned}$$

Goal: show $c \subseteq \gamma(a) \iff \alpha(c) \sqsubseteq a$

Consider the case where $c \in C$ is non-empty set and $a = [p, q]$, $p \leq q$

$$\begin{aligned} & c \subseteq \gamma(a) \\ \iff & c \subseteq \gamma([p,q]) \\ \iff & \end{aligned}$$

Galois Connection for Intervals

$$\begin{aligned} \gamma([p,q]) &= \{x \in \mathbb{Z} \mid p \leq x \leq q\}, \quad \gamma(\bot) = \emptyset \\ \alpha(c) &= [\inf(c), \sup(c)], \text{ if } c \neq \emptyset \quad \alpha(\emptyset) = \bot \end{aligned}$$

Goal: show $c \subseteq \gamma(a) \iff \alpha(c) \sqsubseteq a$

Consider the case where $c \in C$ is non-empty set and $a = [p, q]$, $p \leq q$

$$\begin{aligned} & c \subseteq \gamma(a) \\ \iff & c \subseteq \gamma([p,q]) \\ \iff & \forall x \in c. p \leq x \wedge x \leq q \\ \iff & \end{aligned}$$

Galois Connection for Intervals

$$\begin{aligned} \gamma([p,q]) &= \{x \in \mathbb{Z} \mid p \leq x \leq q\}, \quad \gamma(\bot) = \emptyset \\ \alpha(c) &= [\inf(c), \sup(c)], \text{ if } c \neq \emptyset \quad \alpha(\emptyset) = \bot \end{aligned}$$

Goal: show $c \subseteq \gamma(a) \iff \alpha(c) \sqsubseteq a$

Consider the case where $c \in C$ is non-empty set and $a = [p, q]$, $p \leq q$

$$\begin{aligned} & c \subseteq \gamma(a) \\ \iff & c \subseteq \gamma([p,q]) \\ \iff & \forall x \in c. p \leq x \wedge x \leq q \\ \iff & (\forall x \in c. p \leq x) \wedge (\forall x \in c. x \leq q) \\ \iff & \end{aligned}$$

Galois Connection for Intervals

$$\begin{aligned} \gamma([p,q]) &= \{x \in \mathbb{Z} \mid p \leq x \leq q\}, \quad \gamma(\bot) = \emptyset \\ \alpha(c) &= [\inf(c), \sup(c)], \text{ if } c \neq \emptyset \quad \alpha(\emptyset) = \bot \end{aligned}$$

Goal: show $c \subseteq \gamma(a) \iff \alpha(c) \sqsubseteq a$

Consider the case where $c \in C$ is non-empty set and $a = [p,q]$, $p \leq q$

$$\begin{aligned} & c \subseteq \gamma(a) \\ \iff & c \subseteq \gamma([p,q]) \\ \iff & \forall x \in c. p \leq x \wedge x \leq q \\ \iff & (\forall x \in c. p \leq x) \wedge (\forall x \in c. x \leq q) \\ \iff & p \leq \inf(c) \wedge q \leq \sup(c) \qquad \sup(c) \leq q \\ \iff & \end{aligned}$$

Galois Connection for Intervals

$$\begin{aligned} \gamma([p,q]) &= \{x \in \mathbb{Z} \mid p \leq x \leq q\}, \quad \gamma(\bot) = \emptyset \\ \alpha(c) &= [\inf(c), \sup(c)], \text{ if } c \neq \emptyset \quad \alpha(\emptyset) = \bot \end{aligned}$$

Goal: show $c \subseteq \gamma(a) \iff \alpha(c) \sqsubseteq a$

Consider the case where $c \in C$ is non-empty set and $a = [p, q]$, $p \leq q$

$$\begin{aligned} & c \subseteq \gamma(a) \\ \iff & c \subseteq \gamma([p,q]) \\ \iff & \forall x \in c. p \leq x \wedge x \leq q \\ \iff & (\forall x \in c. p \leq x) \wedge (\forall x \in c. x \leq q) \\ \iff & p \leq \inf(c) \wedge q \leq \sup(c) \\ \iff & \gamma([\inf(c), \sup(c)]) \subseteq \gamma([p,q]) \\ \iff & \end{aligned}$$

Galois Connection for Intervals

$$\begin{aligned} \gamma([p,q]) &= \{x \in \mathbb{Z} \mid p \leq x \leq q\}, \quad \gamma(\bot) = \emptyset \\ \alpha(c) &= [\inf(c), \sup(c)], \text{ if } c \neq \emptyset \quad \alpha(\emptyset) = \bot \end{aligned}$$

Goal: show $c \subseteq \gamma(a) \iff \alpha(c) \sqsubseteq a$

Consider the case where $c \in C$ is non-empty set and $a = [p,q]$, $p \leq q$

$$\begin{aligned} & c \subseteq \gamma(a) \\ \iff & c \subseteq \gamma([p,q]) \\ \iff & \forall x \in c. p \leq x \wedge x \leq q \\ \iff & (\forall x \in c. p \leq x) \wedge (\forall x \in c. x \leq q) \\ \iff & p \leq \inf(c) \wedge q \leq \sup(c) \\ \iff & \gamma([\inf(c), \sup(c)]) \subseteq \gamma([p,q]) \\ \iff & [\inf(c), \sup(c)] \sqsubseteq [p,q] \\ \iff & \end{aligned}$$

Galois Connection for Intervals

$$\begin{aligned} \gamma([p,q]) &= \{x \in \mathbb{Z} \mid p \leq x \leq q\}, \quad \gamma(\bot) = \emptyset \\ \alpha(c) &= [\inf(c), \sup(c)], \text{ if } c \neq \emptyset \quad \alpha(\emptyset) = \bot \end{aligned}$$

Goal: show $c \subseteq \gamma(a) \iff \alpha(c) \sqsubseteq a$

Consider the case where $c \in C$ is non-empty set and $a = [p, q]$, $p \leq q$

$$\begin{aligned} & c \subseteq \gamma(a) \\ \iff & c \subseteq \gamma([p,q]) \\ \iff & \forall x \in c. p \leq x \wedge x \leq q \\ \iff & (\forall x \in c. p \leq x) \wedge (\forall x \in c. x \leq q) \\ \iff & p \leq \inf(c) \wedge q \leq \sup(c) \\ \iff & \gamma([\inf(c), \sup(c)]) \subseteq \gamma([p,q]) \\ \iff & [\inf(c), \sup(c)] \sqsubseteq [p,q] \\ \iff & \alpha(c) \sqsubseteq a \end{aligned}$$

Galois Connection for Intervals

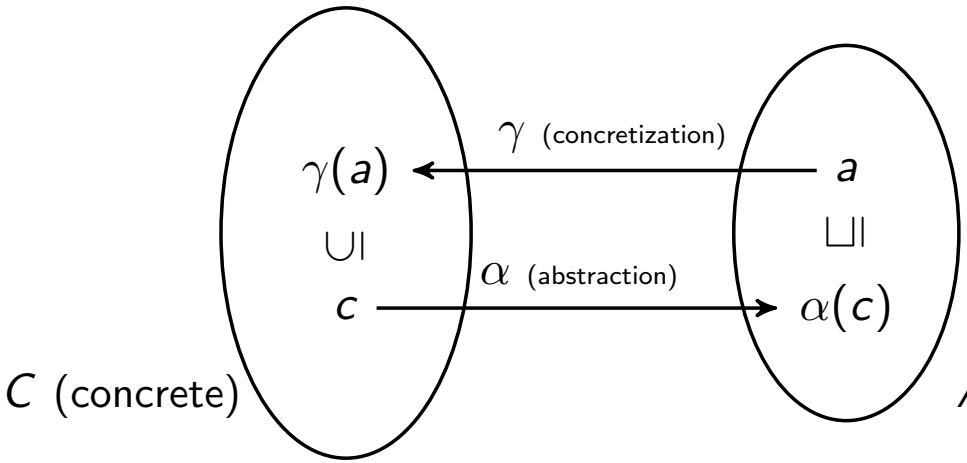
$$\begin{aligned} \gamma([p,q]) &= \{x \in \mathbb{Z} \mid p \leq x \leq q\}, \quad \gamma(\bot) = \emptyset \\ \alpha(c) &= [\inf(c), \sup(c)], \text{ if } c \neq \emptyset \quad \alpha(\emptyset) = \bot \end{aligned}$$

Goal: show $c \subseteq \gamma(a) \iff \alpha(c) \sqsubseteq a$

Consider the case where $c \in C$ is non-empty set and $a = [p, q]$, $p \leq q$

$$\begin{aligned} & c \subseteq \gamma(a) \\ \iff & c \subseteq \gamma([p,q]) \\ \iff & \forall x \in c. p \leq x \wedge x \leq q \\ \iff & (\forall x \in c. p \leq x) \wedge (\forall x \in c. x \leq q) \\ \iff & p \leq \inf(c) \wedge q \leq \sup(c) \\ \iff & \gamma([\inf(c), \sup(c)]) \subseteq \gamma([p,q]) \\ \iff & [\inf(c), \sup(c)] \sqsubseteq [p,q] \\ \iff & \alpha(c) \sqsubseteq a \end{aligned}$$

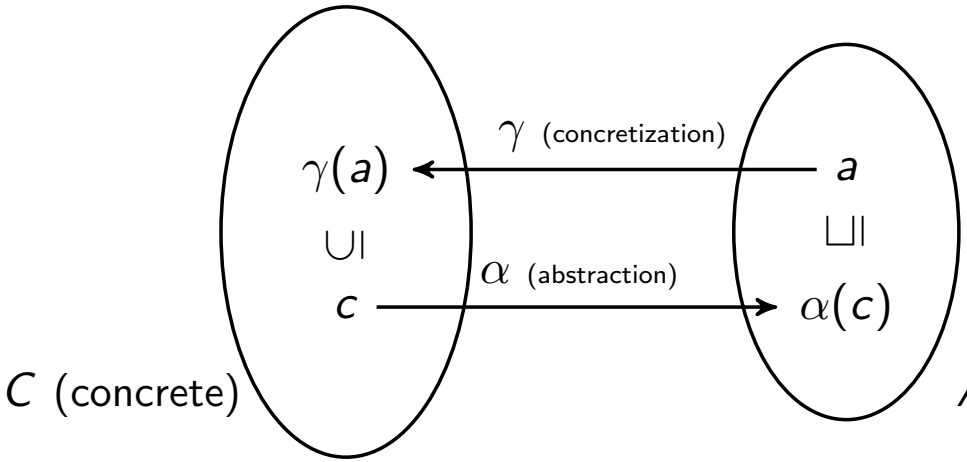
This Galois Connection Connects Larger and Smaller Set



$$c \subseteq \gamma(a) \iff \alpha(c) \sqsubseteq a$$

α is not injective, e.g.

This Galois Connection Connects Larger and Smaller Set



$$c \subseteq \gamma(a) \iff \alpha(c) \sqsubseteq a$$

α is not injective, e.g. $\alpha(\{0,2\}) = [0,2] = \alpha(\{0,1,2\})$
α and γ are not inverses of each other.
A is smaller than C according to set theory:

$$A \sim \mathbb{Z}^2 \sim \mathbb{Z} \prec 2^{\mathbb{Z}} \sim C$$

There is no bijection between A and C, but there is Galois connection

Galois Connection: Equivalent Formulation

A Galois connection is defined by two monotonic functions $\alpha : C \to A$ and $\gamma : A \to C$ between partial orders $\leq$ on C and $\sqsubseteq$ on A, such that

$$\forall c, a. \quad c \subseteq \gamma(a) \iff \alpha(c) \sqsubseteq a \qquad (*)$$

Show that the condition $(*)$ is equivalent to the conjunction of these two conditions:

$$\forall c \in C. \quad c \subseteq \gamma(\alpha(c))$$
$$\forall a \in A. \quad \alpha(\gamma(a)) \sqsubseteq a$$

$\Rightarrow$: $c \subseteq \gamma(a) \,/\, \alpha$

$\alpha(c) \sqsubseteq \alpha(\gamma(a)) \sqsubseteq a$

$\Leftarrow$: $\alpha(c) \sqsubseteq a \,/\, \gamma$

$c \subseteq \gamma(\alpha(c)) \subseteq \gamma(a)$

$\alpha(c) \sqsubseteq \alpha(c)$

$\gamma(a) \subseteq \gamma(a)$

$\alpha(\gamma(a)) \sqsubseteq a$

Lemma: In every Galois connection (α, γ) the following holds:

- $\forall c \in C. \ \alpha(\gamma(\alpha(c))) = \alpha(c)$
- $\forall a \in A. \ \gamma(\alpha(\gamma(a))) = \gamma(a)$

$\alpha(c) \sqsubseteq \alpha(\gamma(\alpha(c)))$ $\quad a = \alpha(c)$

$\alpha(\gamma(\alpha(c))) \sqsubseteq \alpha(c)$

Galois Insertion

Lemma: If f, g are functions and $f \circ g$ (defined by $(f \circ g)(x) = f(g(x))$ is identity function, then f is surjective and g is injective.

Lemma: Let α and γ satisfy the condition of a Galois connection. Show that the following three conditions are equivalent:

1. $\alpha(\gamma(a)) = a$, for all $a \in A$
2. α is a surjective function
3. γ is an injective function

If these conditions hold, we say (α, γ) is a Galois insertion of $(A, \sqsubseteq)$ into $(C, \subseteq)$.

Galois insertion gives an isomorphism between $(A, \sqsubseteq)$ and its image under γ. Thus, Galois insertion is a renaming of a substructure of $(C, \subseteq)$, together with an abstraction operator α that replaces any element of C with an element of this substructure (its approximation). E.g. replaces a set with an enclosing interval.

Dual Notion to Galois Insertion?

State the condition for $c = \gamma(\alpha(c))$ to hold for all c. When C is the set of sets of concrete states and A is a domain of static analysis, is it more reasonable to expect that $c = \gamma(\alpha(c))$ or $\alpha(\gamma(a)) = a$ to hold?

Least Upper Bounds and Monotonic Functions

Lemma: Let $(A, \sqsubseteq, \sqcup)$ and $(C, \subseteq, \cup)$ be semi-lattices (so $\sqcup$ is lub with respect to $\sqsubseteq$ and $\cup$ with respect to $\subseteq$). Show that, if $\gamma : A \to C$ is a monotonic function then

$$\gamma(a_1) \cup \gamma(a_2) \subseteq \gamma(a_1 \sqcup a_2)$$

i.e., $\sqcup$ over-approximates union.

$a_1 \sqsubseteq a_1 \sqcup a_2$ / γ

$\gamma(a_1) \subseteq \gamma(a_1 \sqcup a_2)$

$\gamma(a_2) \subseteq \gamma(a_1 \sqcup a_2)$

Lemma: Suppose that $\alpha : C \to A$ is monotonic. Which one of the following necessarily holds:

- ~~$\alpha(c_1 \cup c_2) \sqsubseteq \alpha(c_1) \sqcup \alpha(c_2)$~~
- $\alpha(c_1) \sqcup \alpha(c_2) \sqsubseteq \alpha(c_1 \cup c_2)$ ✓

Constructing Least Upper Bounds

Lemma: Let (α, γ) be a Galois **insertion** of a partial order $(A, \sqsubseteq)$ into a semi-lattice $(C, \subseteq, \cup)$. Define operation $*$ on A by

$$a_1 * a_2 = \alpha(\gamma(a_1) \cup \gamma(a_2))$$

Then for every $a_1, a_2 \in A$ the value $a_1 * a_2$ the least upper bound on a_1 and a_2, and thus $(A, \sqsubseteq, \sqcup)$ is also a semi-lattice.

Approximating Fixpoint

Let $C = 2^{\mathbb{Z}}$. Consider $F : C \to C$

$$F(c) = \{0\} \cup \{x + 2 \mid x \in c\}$$

Approximating Fixpoint

Let $C = 2^{\mathbb{Z}}$. Consider $F : C \to C$

$$F(c) = \{0\} \cup \{x + 2 \mid x \in c\}$$

Let $(A, \sqsubseteq)$ be the set of integer intervals. Define $F_\# : A \to A$

$$F_\#(a) = [0, 0] \sqcup G(a)$$

where $G(\bot) = \bot$ and $G([p, q]) = [p + 2, q + 2]$. Thus

$$F_\#(a) = \begin{cases} [0, 0], & \text{if } a = \bot \\ [\min(p + 2, 0), \max(q + 2, 0)], & \text{if } a = [p, q] \end{cases}$$

$F_\#$ satisfies the following crucial soundness property:

$\alpha(F(\gamma(a)) \sqsubseteq F_\#(a)$

$$F(\gamma(a)) \subseteq \gamma(F_\#(a))$$

We have $lfp(F_\#) = \bigsqcup_{k \geq 0}[0, 2k] = [0, \infty]$, so $F_\#([0, \infty]) \sqsubseteq [0, \infty]$, and

$$F(\gamma([0, \infty])) \subseteq \gamma(F_\#([0, \infty])) \subseteq \gamma([0, \infty])$$

Found desired $c = \gamma([0, \infty])$ by searching in A.

Sound Approximation Functions

Consider $(C, \subseteq)$, $(A, \sqsubseteq)$, $F : C \to C$ and $F_\# : A \to A$. The key soundness property that we would like is

$$F(\gamma(a)) \subseteq \gamma(F_\#(a))$$

Suppose we have α as well such that (α, γ) is Galois connection. Then we can define

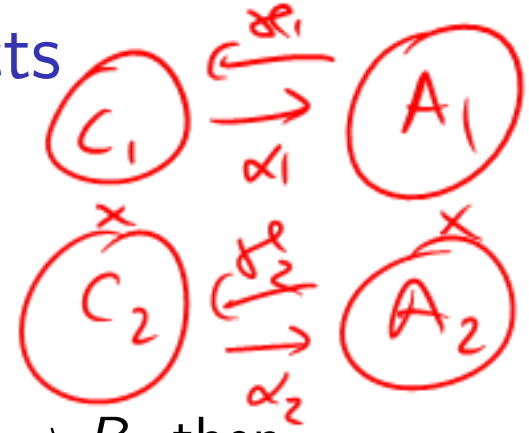
$$F_\#(a) = \alpha(F(\gamma(a)))$$

This implies

$$\alpha(F(\gamma(a))) \sqsubseteq F_\#(a)$$

so by Galois connection condition this is equivalent to $F(\gamma(a)) \subseteq \gamma(F_\#(a))$.

Galois Connection on Products



Notation: if $f_1 : A_1 \to B_1$ and $f_2 : A_2 \to B_2$ then $(f_1 \times f_2) : A_1 \times A_2 \to B_1 \times B_2$ is given pointwise: $(f_1 \times f_2)(x_1, x_2) = (f_1(x_1), f_2(x_2))$.

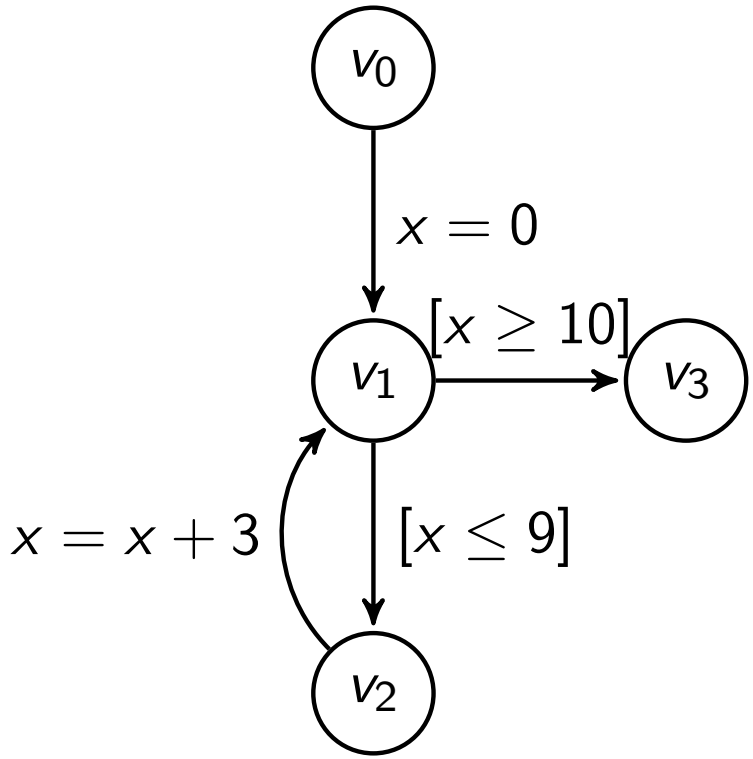
Let (α_1, γ_1) be a Galois connection between $(A_1, \sqsubseteq_1)$ and $(C_1, \subseteq_1)$. Let (α_2, γ_2) be a Galois connection between $(A_2, \sqsubseteq_2)$ and $(C_2, \subseteq_2)$. Then $(\alpha_1 \times \alpha_2, \gamma_1 \times \gamma_2)$ is a Galois connection between product partial orders $(A_1, \sqsubseteq_1) \times (A_2, \sqsubseteq_2)$ and $(C_1, \subseteq_1) \times (C_2, \subseteq_2)$.

$(a_1, a_2) \sqsubseteq (a_1', a_2')$

$a_1 \sqsubseteq_1 a_1' \quad a_2 \sqsubseteq a_2'$

Example Program

```
// v0
x := 0;
// v1
while (x < 10) {
   // v2
   x := x + 3;
}
// v3
```



Concrete Domain: Sets of States

Because there is only one variable:

- state is an element of $\mathbb{Z}$ (value of x)
- sets of states are sets of integers, $C = 2^{\mathbb{Z}}$ (concrete domain)
- for each command K, strongest postcondition function $sp(\cdot, K) : C \to C$

Strongest Postondition

Compute *sp* on example statements:

$$sp(P, x := 0) = \{0\}$$

$$sp(P, x := x + 3) = \{x + 3 \mid x \in P\}$$

$$sp(P, assume(x < 10)) = \{x \mid x \in P \wedge x < 10\}$$

$$sp(P, assume(\neg(x < 10)) = \{x \mid x \in P \wedge x \geq 10\}$$

Sets of States at Each Program Point

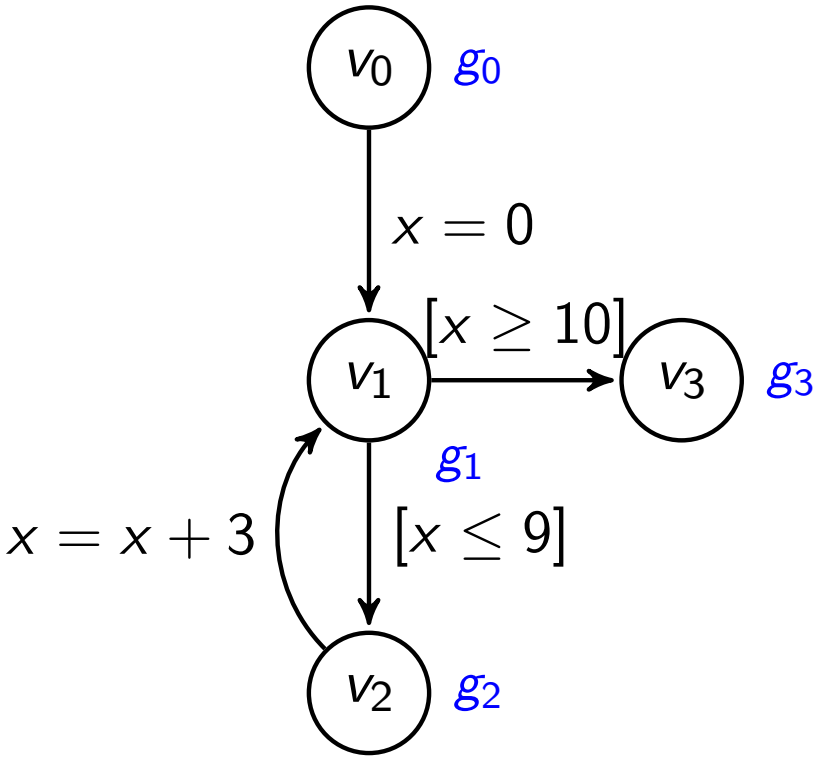
Collecting semantics computes with sets of states at each program point

$$g : \{v_0, v_1, v_2, v_3\} \to C$$

We sometimes write g_i as a shorthand for $g(v_i)$, for $i \in \{0, 1, 2, 3\}$.
In the initial state the value of variable is arbitrary: $I = \mathbb{Z}$

post Function for the Collecting Semantics

From here we can derive F that maps g to new value of g:



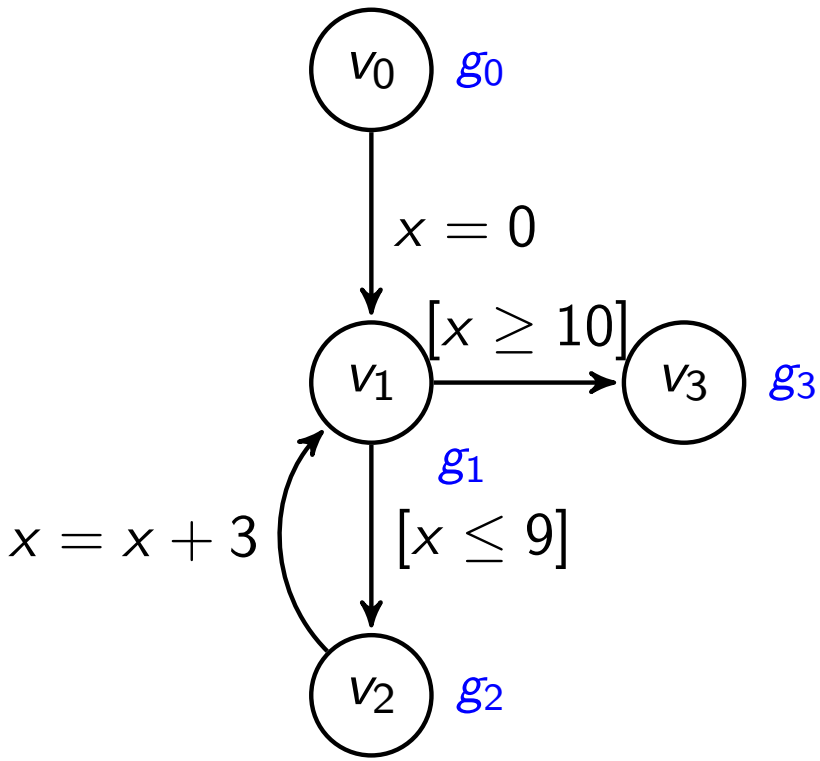
$$F(g_0, g_1, g_2, g_3) =$$
$$(\mathbb{Z},$$
$$sp(g_0, x := 0) \cup sp(g_2, x := x + 3),$$
$$sp(g_1, assume(x \leq 9)),$$
$$sp(g_1, assume(x \geq 10)))$$

Sets of States at Each Program Point

The fixpoint condition $F(g) = g$ becomes a system of inequations

$$
\begin{aligned}
g_0 &= \mathbb{Z} \\
g_1 &= sp(g_0, x := 0) \cup sp(g_2, x := x + 3) \\
g_2 &= sp(g_1, assume(x \leq 0)) \\
g_3 &= sp(g_1, assume(x \geq 10)))
\end{aligned}
$$

whereas the postfix point (see Tarski's fixpoint theorem) becomes



$F(g_0, g_1, g_2, g_3) \subseteq (g_0, g_1, g_2, g_3)$

$$
\begin{aligned}
&\mathbb{Z} \subseteq g_0 \\
&sp(g_0, x := 0) \cup sp(g_2, x := x + 3) \subseteq g_1 \\
&sp(g_1, assume(x \leq 9)) \subseteq g_2 \\
&sp(g_1, assume(x \geq 10)) \subseteq g_3
\end{aligned}
$$

Computing Fixpoint

To find the fixpoint, we compute the sequence $F^n(\emptyset, \emptyset, \emptyset, \emptyset)$ for $n \geq 0$:

$$
\begin{aligned}
&(\emptyset, \emptyset, \emptyset, \emptyset)\\
&(\mathbb{Z}, \emptyset, \emptyset, \emptyset)\\
&(\mathbb{Z}, \{0\}, \emptyset, \emptyset)\\
&(\mathbb{Z}, \{0\}, \{0\}, \emptyset)\\
&(\mathbb{Z}, \{0, 3\}, \{0\}, \emptyset)\\
&(\mathbb{Z}, \{0, 3\}, \{0, 3\}, \emptyset)\\
&(\mathbb{Z}, \{0, 3, 6\}, \{0, 3\}, \emptyset)\\
&(\mathbb{Z}, \{0, 3, 6\}, \{0, 3, 6\}, \emptyset)\\
&(\mathbb{Z}, \{0, 3, 6, 9\}, \{0, 3, 6, 9\}, \emptyset)\\
&(\mathbb{Z}, \{0, 3, 6, 9, 12\}, \{0, 3, 6, 9\}, \emptyset)\\
&(\mathbb{Z}, \{0, 3, 6, 9, 12\}, \{0, 3, 6, 9\}, \{12\})\\
&(\mathbb{Z}, \{0, 3, 6, 9, 12\}, \{0, 3, 6, 9\}, \{12\})
\end{aligned}
$$

Thus, all subsequent values remain the same and $(\mathbb{Z}, \{0, 3, 6, 9, 12\}, \{0, 3, 6, 9\}, \{12\})$ is the fixpoint of collecting semantics equations. In general we may need infinitely many iterations to converge.

Now formulate analogous constraints in abstract domain

Abstract Postcondition of Statements: Core of Analysis

We had: $sp(\cdot, c) : C \to C$
Now we have: $sp^{\#}(\cdot, c) : A \to A$
For correctness, we need that for each $a \in A$ and each command r:

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

We would like $sp^{\#}$ to be *as small as possible so that this condition holds.*
By property of Galois Connection, the condition $sp(\gamma(a), r) \subseteq \gamma(sp^{\#}(a, r))$ is equivalent to

$$\alpha(sp(\gamma(a), r)) \sqsubseteq sp^{\#}(a, r)$$

Because we want $sp^{\#}$ to be as small as possible (to obtain correct result), we let equality hold:

$$sp^{\#}(a, r) = \alpha(sp(\gamma(a), r))$$

Because we know α, γ, sp, we can compute the value of $sp^{\#}(a, r)$ by simplifying certain expressions involving sets of states.

Example

For $p \leq q$ we have:

$$\begin{aligned} sp^{\#}([p,q], x := x+3) &= \alpha(sp(\gamma([p,q]), x := x+3)) \\ &= \alpha(sp(\{x \mid\ p \leq x \wedge x \leq q\}, x := x+3)) \\ &= \alpha(\{x+3 \mid p \leq x \wedge x \leq q\}) \\ &= \alpha(\{y \mid p+3 \leq y \wedge y \leq q+3\}) \\ &= [p+3, q+3] \end{aligned}$$

For K an integer constant and $a \neq \bot$, we have

$$sp^{\#}(a, x := K) = [K, K]$$

Note that for every command given by relation r, we have

$$\begin{aligned} sp^{\#}(\bot, r) &= \alpha(sp(\gamma(\bot), r)) \\ &= \alpha(sp(\emptyset, r)) \\ &= \alpha(\emptyset) \\ &= \bot \end{aligned}$$

Variable Range Analysis for Example Program

The general form of abstract interpretation of the collecting semantics is analogous to collecting semantics, but replaces operations on sets with operations on the lattice:

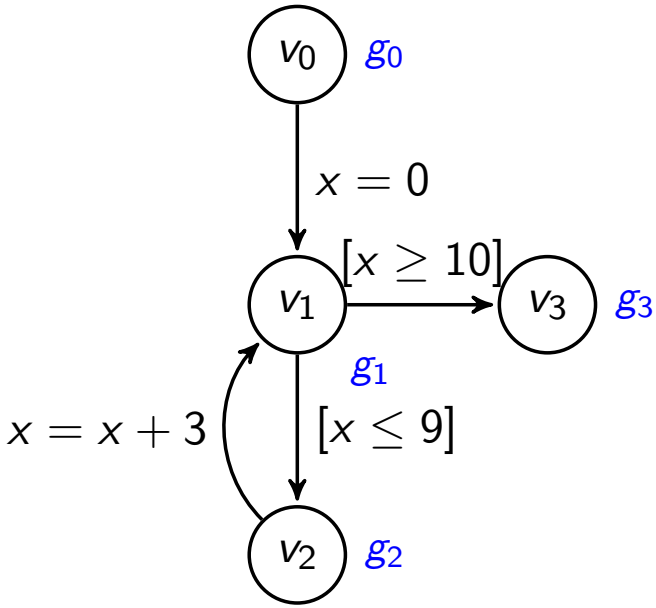
$$F^{\#} : (V \to A) \to (V \to A)$$

$$F(g^{\#})(v') = g^{\#}_{init}(v') \sqcup \bigsqcup_{(v,v') \in E} sp^{\#}(g^{\#}(v), r(v, v'))$$

Here $g^{\#}_{init}(v')$ will be $\bot$ except at the entry into our control-flow graph, where it approximates the set of initial states at the entry point.

Abstract Semantic Function for the Program

In Collecting Semantics for Example Program we had



$$
\begin{aligned}
&F(g_0, g_1, g_2, g_3) = \\
&(\mathbb{Z}, \\
&sp(g_0, x := 0) \cup sp(g_2, x := x + 3), \\
&sp(g_1, assume(x \leq 9)), \\
&sp(g_1, assume(x \geq 10)))
\end{aligned}
$$

Here we have:

$$
\begin{aligned}
&F^{\#}(g_0^{\#}, g_1^{\#}, g_2^{\#}, g_3^{\#}) = \\
&(\top, \\
&sp^{\#}(g_0^{\#}, x := 0) \sqcup sp^{\#}(g_2^{\#}, x := x + 3), \\
&sp^{\#}(g_1^{\#}, assume(x \leq 9)), \\
&sp^{\#}(g_1^{\#}, assume(x \geq 10)))
\end{aligned}
$$

Solving Abstract Function

Doing the analysis means computing $(F^{\#})^n(\bot, \bot, \bot, \bot)$ for $n \geq 0$:

$$
\begin{array}{l}
(\bot, \bot, \bot, \bot) \\
(\top, \bot, \bot, \bot) \\
(\top, [0,0], \bot, \bot) \\
(\top, [0,0], [0,0], \bot) \\
(\top, [0,3], [0,0], \bot) \\
(\top, [0,3], [0,3], \bot) \\
(\top, [0,6], [0,3], \bot) \\
(\top, [0,6], [0,6], \bot) \\
(\top, [0,9], [0,6], \bot) \\
(\top, [0,9], [0,9], \bot) \\
(\top, [0,12], [0,9], \bot) \\
(\top, [0,12], [0,9], [10,12]) \\
(\top, [0,12], [0,9], [10,12])
\end{array}
$$

Note the approximation (especially in the last step) compared to the collecting semantics we have computed before for our example program.

Abstract Interpretation

Given control-flow graph: (V, E, r) where

- $V = \{v_1, \ldots, v_n\}$ is set of program points
- $E \subseteq V \times V$ are control-flow graph edges
- $r : E \to 2^{S \times S}$, so each $r(v, v') \subseteq S \times S$ is relation describing the meaning of command between v and v'

Key steps:

- design abstract domain A that represents sets of program states
- define $\gamma : A \to C$ giving meaning to elements of A
- define lattice ordering $\sqsubseteq$ on A such that $a_1 \sqsubseteq a_2 \to \gamma(a_1) \subseteq \gamma(a_2)$
- define $sp^{\#} : A \times 2^{S \times S} \to A$ that maps an abstract element and a CFG statement to new abstract element, such that
  $sp(\gamma(a), r) \subseteq \gamma(sp^{\#}(a, r))$
  For example, by defining function α so that (α, γ) becomes a *Galois Connection* and defining $sp^{\#}(a) = \alpha(sp(\gamma(a), r))$.

Running Abstract Interpretation

- Extend $sp^{\#}$ to work on control-flow graphs, by defining $F^{\#} : (V \to A) \to (V \to A)$ as follows (below, $g^{\#} : V \to A$)

$$F^{\#}(g^{\#})(v') = Init(v') \sqcup \bigsqcup_{(v,v') \in E} sp^{\#}(g^{\#}(v), r(v, v'))$$

- Compute $g_*^{\#} = lfp(F^{\#})$ (this is easier than computing semantics because lattice A^n is simpler than C^n):

$$g_*^{\#} = \bigsqcup_{n \geq 0} (F^{\#})^n(\bot^{\#})$$

  where $\bot^{\#}(v) = \bot_A$ for all $v \in V$.

The resulting fixpoint describes an inductive program invariant.

Termination and Efficiency of Abstract Interpretation

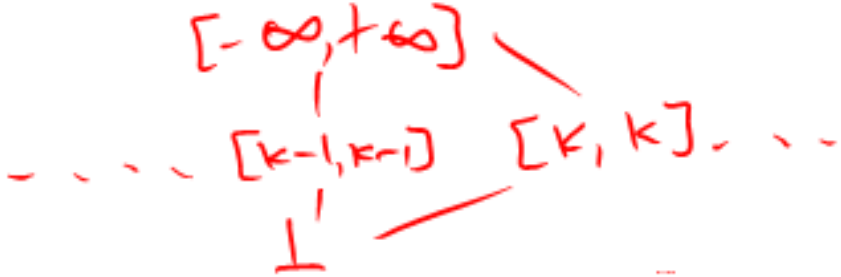
Definition: A **chain** of length n is a sequence $s_0, s_1, \ldots, s_n$ such that

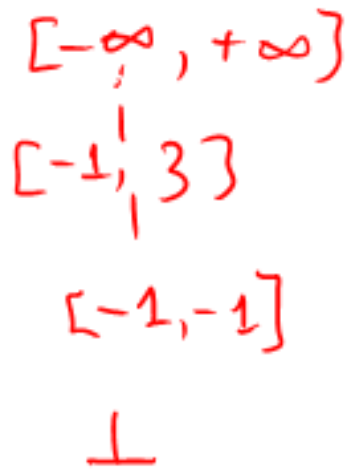
$$s_0 \sqsubset s_1 \sqsubset s_2 \sqsubset \ldots \sqsubset s_n$$

where $x \sqsubset y$ means, as usual, $x \sqsubseteq y \wedge x \neq y$

Definition: A partial order has a **finite height** n if it has a chain of length n and every chain is of length at most n.

A finite lattice is of finite height.





Example

The constant propagation lattice $\mathbb{Z} \cup \{\bot, \top\}$ is an infinite lattice of height 2. One example chain of length 2 is

$$\bot \sqsubset 42 \sqsubset \top$$

Here the γ function is given by

- $\gamma(k) = \{k\}$ when $k \in \mathbb{Z}$
- $\gamma(\top) = \mathbb{Z}$
- $\gamma(\bot) = \emptyset$

The ordering is given by $a_1 \subseteq a_2$ iff $\gamma(a_1) \subseteq \gamma(a_2)$

Example

If a state of a (one-variable) program is given by an integer, then a concrete lattice element is a set of integers. This lattice has infinite height. There is a chain

$$\{0\} \subset \{0,1\} \subset \{0,1,2\} \subset \ldots \subset \{0,1,2,\ldots,n\}$$

for every n.

Convergence in Lattices of Finite Height

Consider a finite-height lattice $(L, \sqsubseteq)$ of height n and function

$$F : L \to L$$

What is the maximum length of sequence $\bot, F(\bot), F^2(\bot), \ldots$?
Give an effectively computable expression for $lfp(F)$.

Computing the Height when Combining Lattices

Let $H(L, \leq)$ denote the height of the lattice $(L, \leq)$.

Product

Given lattices $(L_1, \sqsubseteq_1)$ and $(L_2, \sqsubseteq_2)$, consider product lattice with set $L_1 \times L_2$ and potwise order

$$(x_1, x_2) \sqsubseteq (x_1', x_2')$$

iff . .

What is the height of the product lattice?

Exponent

Given lattice $(L, \sqsubseteq)$ and set V, consider the lattice $(L^V, \sqsubseteq')$ defined by

$$g \sqsubseteq' h$$

iff $\forall v \in V . g(v) \sqsubseteq h(v)$.

What is the height of the exponent lattice?

Computing the Height when Combining Lattices

Let $H(L, \leq)$ denote the height of the lattice $(L, \leq)$.

Product

Given lattices $(L_1, \sqsubseteq_1)$ and $(L_2, \sqsubseteq_2)$, consider product lattice with set $L_1 \times L_2$ and potwise order

$$(x_1, x_2) \sqsubseteq (x_1', x_2')$$

iff ...

What is the height of the product lattice?

Exponent

Given lattice $(L, \sqsubseteq)$ and set V, consider the lattice $(L^V, \sqsubseteq')$ defined by

$$g \sqsubset' h$$

iff $\forall v \in V . g(v) \sqsubseteq h(v)$.

What is the height of the exponent lattice?

Answer: height of L times the cardinality of V.

Widening and Narrowing in Variable Range Analysis

Interval analysis domain, for each program point, maps each program variable to an interval.
Analysis domain has elements $g^{\#} : V \to I$ where I denotes the set of such intervals.

Height of lattice for unbounded integers: infinite.

Height of lattice of one interval for 64-bit integers: around 2^{64}

Moreover, if we have q variables in program and p program points, height of lattice for the analysis domain is pq times larger.

How to guarantee (reasonably fast) termination?

Widening technique

If the iteration does not seem to be converging, take a "jump" and make the interval much **wider** (larger).
Finite set of *jump points* J (e.g. set of all integer constants in the program)
In fixpoint computation, compose H_i with function

$$w([a,b]) = [\max\{x \in J \mid x \leq a\}, \min\{x \in J \mid b \leq x\}]$$

We require the condition:

$$x \sqsubseteq W(x)$$

for all x.
The condition holds for the example above.

Approaches

- always apply widening (we will assume this)
- iterate a few times with H_i only (without using w), if we are not at a fixpoint at this program point, then widen.
- this is not monotonic: if you start at fixpoint, it converges, if start below, can jump over fixpoint!

Standard iteration: $\bot, \ldots, (F^\#)^n(\bot), \ldots$
Widening: $\bot, \ldots, ((W \circ F^\#)^n(\bot), \ldots$

Example where widening works nicely

Consider program:

```
x = 0;
while (x < 1000) {
  x = x + 1;
}
```

Interval analysis without widening will need around 1000 iterations to converge to interval $[1000, 1000]$ for x at the end of the program.
This may be too slow.

Let us derive the set J by taking all constants that appear in the program, as well as $-\infty$ and $+\infty$:

$$J = \{-\infty, 0, 1, 1000, +\infty\}$$

After a few iterations, widening maps interval $[0, 2]$ into $[0, 1000]$. This gives $[0, 999]$ for x at loop entry and again $[1000, 1000]$ for x at the end of the program, but in many fewer iterations.

Example showing problems with widening

Consider program:

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

Interval analysis without widening will need around 1000 iterations to converge to

$$x \mapsto [1000, 1000]; \qquad y \mapsto [1, 2001]$$

This may be too slow.
Now apply widening with the same J as before. When within loop we obtain $x \mapsto [0, 1000]$, applying widening function to the interval $[0, 2000]$ for y results in $[0, +\infty)$. We obtain $y \mapsto [1, +\infty)$ at the end of the program:

$$x \mapsto [1000, 1000]; \qquad y \mapsto [1, +\infty)$$

Narrowing

Observation

Consider a monotonic function, such as $f(x) = 0.5x + 1$ on the set of real numbers.

If we consider a sequence $x_0, f(x_0), \ldots$, this sequence is

- monotonically increasing iff $x_0 < x_1$ (e.g. for $x_0 = 0$)
- monotonically decreasing iff $x_1 < x_0$ (e.g. for $x_0 = 3$)

Informally, the sequence continues of the direction in which it starts in the first step.

This is because $x_0 < x_1$ implies by monotonicity of f that $x_1 < x_2$ etc., whereas $x_1 < x_0$ implies $x_2 < x_1$.

The Idea

Let $W : A \to A$ such that $x \sqsubseteq W(x)$.

After finding fixpoint of $(W \circ F)^{\#}$, apply $F^{\#}$ to improve precision.

Widen and Narrow

Lemma: Let $F^{\#}$ and W be monotonic functions on a partial order $\sqsubseteq$ such that $x \sqsubseteq W(x)$ for all x. Define the following:

- $x_* = \sqcup_{n \geq 0} (F^{\#})^n(\bot)$
- $y_* = \sqcup_{n \geq 0} (W \circ F^{\#})^n(\bot)$
- $z_* = \sqcap_{n \geq 0} (F^{\#})^n(y_*)$

where we also assume that the two $\sqcup$ and one $\sqcap$ exist. Then

- x_* is the least fixpoint of $F^{\#}$ and z_*, is the least fixpoint of $W \circ F^{\#}$ (by Tarski's Fixpoint Theorem), and
- $x_* \sqsubseteq z_* \sqsubseteq y_*$.

Proof

By induction, for each n we have

$$(F^{\#})^n(\bot) \sqsubseteq (W \circ F^{\#})^n(\bot)$$

Thus by Comparing Fixpoints of Sequences, we have $x_* \sqsubseteq y_*$.
Next, we have that

$$x_* = F^{\#}(x_*) \sqsubseteq F^{\#}(y_*) \sqsubseteq (W \circ F^{\#})(y_*) \sqsubseteq y_*$$

Thus, $F^{\#}(y_*) \sqsubseteq y_*$. From there by induction and monotonicity of $F^{\#}$ we obtain

$$(F^{\#})^{n+1}(y_*) \sqsubseteq (F^{\#})^n(y_*)$$

i.e. the sequence $(F^{\#})^n(y_*)$ is **decreasing**. Therefore, y_* is its upper bound and therefore $z_* \sqsubseteq y_*$.
On the other hand, we have by monotonicity of $F^{\#}$, the fact that x_* is fixpoint, and $x_* \sqsubseteq y_*$ that:

$$x_* = (F^{\#})^n(x_*) \sqsubseteq (F^{\#})^n(y_*)$$

Thus, x_* is the lower bound on $(F^{\#})^n(y_*)$, so $x_* \sqsubseteq z_*$.

Note

Even if z_* does not exist, we can simply compute $(F^\#)^n(y_*)$ for any chosen value of n, it is still a sound over-approximation, because it approximates x_*, which approximates the concrete value:

$$x_* \sqsubseteq z_n$$

so

$$s_* \subseteq \gamma(x_*) \subseteq \gamma(z_n)$$

Being able to stop at any point gives us an **anytime algorithm**.

Example showing how narrowing may improve result after widening

In the above example for the program, the results obtained using widening are:

```
x = 0;
y = 1;
// x -> [0,0], y -> [1,1]
// (merge point)
// x -> [0,1000], y -> [1,+infty)
while (x < 1000) {
  // x -> [0,999], y -> [1,+infty)
  x = x + 1;
  // x -> [0,1000], y -> [1,+infty)
  y = 2*x;
  // x -> [0,1000], y -> [0,+infty)
  y = y + 1;
  // x -> [0,1000], y -> [1,+infty)
  print(y);
}
// x -> [1000,1000], y -> [1,+infty)
```

Example cont.

Let us now apply one ordinary iteration, without widening. We obtain:

```
x = 0;
y = 1;
// x -> [0,0], y -> [1,1]
// (merge point)
// x -> [0,1000], y -> [1,2001]
while (x < 1000) {
  // x -> [0,999], y -> [1,+infty)
  x = x + 1;
  // x -> [0,1000], y -> [1,+infty)
  y = 2*x;
  // x -> [0,1000], y -> [0,2000]
  y = y + 1;
  // x -> [0,1000], y -> [1,2001]
  print(y);
}
// x -> [1000,1000], y -> [1,2001]
```

Thus, we obtained a good first approximation by a few iterations with widening and then improved it with a single iteration without widening.

Exercises

Exercise 1:
Consider an analysis that has two integer variables, for which we track intervals, and one boolean variable, whose value we track exactly.
Give the type of $F^{\#}$ for such program.

Exercise 2:
Consider the program that manipulates two integer variables x, y.
Consider any assignment $x = e$, where e is a linear combination of integer variables, for example,

$$x = 2 * x - 5 * y$$

Consider an interval analysis that maps each variable to its value.
Describe an algorithm that will, given a syntax tree of $x = e$ and intervals for x (denoted $[a_x, b_x]$) and y (denoted $[a_y, b_y]$) find the new interval $[a, b]$ for x after the assignment statement.

Exercise 3

a)

For a program whose state is one integer variable and whose abstraction is an interval, derive general transfer functions $sp^{\#}(a, c)$ for the following statements c, where K is an arbitrary compile-time constant known in the program:

- $x = K$
- $x = x + K$
- $assume(x \leq K)$
- $assume(x \geq K)$

b)

Consider a program with two integer variables, x,y. Consider analysis that stores one interval for each variable.

- Define the domain of lattice elements a that are computed for each program point.
- Give the definition for statement $sp^{\#}(a, y = x + y + K)$

Exercise 3

c)

Draw the control-flow graph for the following program.

Run abstract interpretation that maintains an interval for x at each program point, until you reach a fixpoint.

What are the fixpoint values at program points v_4 and v_5?

```
// v0
x := 0;
// v1
while (x < 10) {
  // v2
  x := x + 3;
}
// v3
if (x >= 0) {
  if (x <= 15) {
    a[x]=7; // index in range
  } else {
    // v4
    error;
  }
} else {
  // v5
  error;
}
```