

Basic Concepts of Abstract Interpretation*

Soonho Kong
soonhok@cs.cmu.edu

15-817A Model Checking and Abstract Interpretation
Carnegie Mellon University

Mar 23, 2011

*Work of P. Cousot and R. Cousot



P. Cousot and R. Cousot.

Basic Concepts of Abstract Interpretation.

In *Building the Information Society*, R. Jacquard (Ed.), pages
359–366. Kluwer Academic Publishers 2004.

Goal

To Understand basic concepts of abstract interpretation.

Contents

- 1 Overview
- 2 Introduction
- 3 Transition Systems
- 4 Partial Trace Semantics
- 5 The Reflexive Transitive Closure Semantics
- 6 The Reachability Semantics
- 7 The Interval Semantics
- 8 Convergence Acceleration
- 9 Conclusion

Introduction

Abstract Interpretation:

a theory of approximation of mathematical structures, in particular those involved in the semantic models of computer systems.

Transition Systems

Programs are formalized as transition systems τ :

$$\tau = \langle \Sigma, \Sigma_i, t \rangle$$

- Σ : a set of states
- $\Sigma_i \subseteq \Sigma$: the set of initial states
- $t \subseteq \Sigma \times \Sigma$: a transition relation between a state and its possible successors.

Example, the transition system

$$\langle \mathbb{Z}, \{0\}, \{ \langle x, x' \rangle \mid x' = x + 1 \} \rangle$$

of program `x := 0; while true do x := x + 1.`

Partial Trace Semantics

A finite partial execution trace : $\sigma = s_0 s_1 \dots s_n$

- $s_0 \in \Sigma$
- For all $i < n$, $\langle s_i, s_{i+1} \rangle \in t$

Partial traces of length 0 : ϕ

Partial traces of length 1 : $\Sigma_\tau^1 = \{s \mid s \in \Sigma\}$

Partial traces of length $n + 1$:

$$\Sigma_\tau^{n+1} = \{\sigma s s' \mid \sigma s \in \Sigma_\tau^n \wedge \langle s, s' \rangle \in t\}$$

Collecting semantics of τ : all partial traces of all finite lengths

$$\Sigma_\tau^{\vec{*}} = \bigcup_{n \geq 0} \Sigma_\tau^n$$

Partial Trace Semantics in Fixpoint Form

For the function $\mathcal{F}_\tau^{\vec{*}}$

$$\mathcal{F}_\tau^{\vec{*}}(X) = \{s \mid s \in \Sigma\} \cup \{\sigma s s' \mid \sigma s \in X \wedge \langle s, s' \rangle \in t\}$$

$\Sigma_\tau^{\vec{*}}$ is the least fixpoint of $\mathcal{F}_\tau^{\vec{*}}$, that is

- $\mathcal{F}_\tau^{\vec{*}}(\Sigma_\tau^{\vec{*}}) = \Sigma_\tau^{\vec{*}}$
- For all X such that $\mathcal{F}_\tau^{\vec{*}}(X) = X$, $\Sigma_\tau^{\vec{*}} \subseteq X$

Therefore,

$$\Sigma_\tau^{\vec{*}} = \text{lfp } \mathcal{F}_\tau^{\vec{*}} = \bigcup_{n \geq 0} \mathcal{F}_\tau^{\vec{*}n}(\phi)$$

Partial Trace Semantics in Fixpoint Form - Proof I

$$\mathcal{F}_\tau^{\vec{*}}(\Sigma_\tau^{\vec{*}}) = \Sigma_\tau^{\vec{*}}$$

The proof is as follows:

$$\mathcal{F}_\tau^{\vec{*}}(\Sigma_\tau^{\vec{*}}) = \mathcal{F}_\tau^{\vec{*}}\left(\bigcup_{n \geq 0} \Sigma_\tau^n\right) \quad \text{def. } \Sigma_\tau^{\vec{*}}$$

$$= \{s \mid s \in \Sigma\} \cup \{\sigma s s' \mid \sigma s \in \left(\bigcup_{n \geq 0} \Sigma_\tau^n\right) \wedge \langle s, s' \rangle \in \mathbf{t}\} \quad \text{def. } \mathcal{F}_\tau^{\vec{*}}$$

$$= \{s \mid s \in \Sigma\} \cup \bigcup_{n \geq 0} \{\sigma s s' \mid \sigma s \in (\Sigma_\tau^n) \wedge \langle s, s' \rangle \in \mathbf{t}\} \quad \text{set theory}$$

$$= \Sigma_\tau^1 \cup \bigcup_{n \geq 0} \Sigma_\tau^{n+1} \quad \text{def. } \Sigma_\tau^1 \text{ and } \Sigma_\tau^{n+1}$$

$$= \bigcup_{n' \geq 1} \Sigma_\tau^{n'} = \bigcup_{n \geq 0} \Sigma_\tau^n$$

by letting $n' = n + 1$ and since $\Sigma_\tau^n = \phi$

Partial Trace Semantics in Fixpoint Form - Proof II

For all X such that $\mathcal{F}_\tau^{\vec{*}}(X) = X$, $\Sigma_\tau^{\vec{*}} \subseteq X$

We prove by induction that $\forall n \geq 0 : \Sigma_\tau^n \subseteq X$

- 1 Base Case : $\Sigma_\tau^0 = \phi \subseteq X$
- 2 Inductive Hypothesis : $\Sigma_\tau^n \subseteq X$

Since $\sigma s \in \Sigma_\tau^n \rightarrow \sigma s \in X$,

$\{\sigma s s' \mid \sigma s \in \Sigma_\tau^n \wedge \langle s, s' \rangle \in t\} \subseteq \{\sigma s s' \mid \sigma s \in X \wedge \langle s, s' \rangle \in t\}$

Therefore,

$$\Sigma_\tau^{n+1} \subseteq \mathcal{F}_\tau^{\vec{*}}(\Sigma_\tau^n) \subseteq \mathcal{F}_\tau^{\vec{*}}(X) = X$$

The Reflexive Transitive Closure Semantics as an Abstraction

- Abstraction of the partial trace semantics

$$\alpha^*(X) = \{\vec{\alpha}(\sigma) \mid \sigma \in X\} \quad \text{where } \vec{\alpha}(s_0s_1 \dots s_n) = \langle s_0, s_n \rangle$$

$\alpha^*(\Sigma_{\tau}^{\vec{\alpha}})$ is the reflexive transitive closure t^* of the transition relation t .

- Concretization

$$\gamma^*(Y) = \{\sigma \mid \vec{\alpha}(\sigma) \in Y\} = \{s_0s_1 \dots s_n \mid \langle s_0, s_n \rangle \in Y\}$$

- $X \subseteq \gamma^*(\alpha^*(X))$

Answering Concrete Questions in the Abstract

Answering concrete question about X using a simpler abstract question on $\alpha^*(X)$.

Example : $s \dots s' \dots s'' \in X? \rightarrow \langle s, s'' \rangle \in \alpha^*(X)?$

Galois Connections

Given any set X of partial traces and Y of pair of states,

$$\alpha^*(X) \subseteq Y \iff X \subseteq \gamma^*(Y)$$

which is a characteristic property of Galois connections.

Proof.

$$\begin{aligned} \alpha^*(X) \subseteq Y &\iff \{\vec{\alpha}^*(\sigma) \mid \sigma \in X\} \subseteq Y && \text{def. } \alpha^* \\ &\iff \forall \sigma \in X : \vec{\alpha}(\sigma) \in Y \\ &\iff X \subseteq \{\sigma \mid \vec{\alpha}(\sigma) \in Y\} && \text{def. } \subseteq \\ &\iff X \subseteq \gamma^*(Y) && \text{def. } \gamma^* \end{aligned}$$

Galois Connections

Galois connections preserve joins.

$$\alpha^*\left(\bigcup_{i \in I} X_i\right) = \bigcup_{i \in I} \alpha^*(X_i)$$

Proof.

$$\begin{aligned}\alpha^*\left(\bigcup_{i \in I} X_i\right) &= \{\vec{\alpha}^*(\sigma) \mid \sigma \in \bigcup_{i \in I} X_i\} \\ &= \bigcup_{i \in I} \{\vec{\alpha}^*(\sigma) \mid \sigma \in X_i\} \\ &= \bigcup_{i \in I} \alpha^*(X_i)\end{aligned}$$

The Reflexive Transitive Closure Semantics in Fixpoint Form

* General Principle in Abstract Interpretation.

- 1 The concrete(partial trace) semantics is expressed in fixpoint form.

$$\Sigma_{\tau}^{\vec{*}} = \text{lfp } \mathcal{F}_{\tau}^{\vec{*}}$$

- 2 The abstract(reflexive transitive closure) semantics is an abstraction of the concrete semantics by a Galois connections and it can be expressed in fixpoint form, too.

$$\alpha^*(\Sigma_{\tau}^{\vec{*}}) = \text{lfp } \mathcal{F}_{\tau}^*$$

- 3 2 can be generalized to order theory, and is known as the fixpoint transfer theorem.

The Reflexive Transitive Closure Semantics in Fixpoint Form - Propositions & Definitions

- ① Proposition 1. $\alpha^*(\phi) = \phi$
 $\phi \subseteq \gamma^*(\phi) \iff \alpha^*(\phi) \subseteq \phi$. Therefore $\alpha^*(\phi) = \phi$.

- ② Propostion 2.

Commutation Property: $\alpha^*(\mathcal{F}_\tau^{\rightarrow*}(X)) = \mathcal{F}_\tau^*(\alpha^*(X))$

① Definition 1. $\mathbb{I}_\Sigma = \{\langle s, s \rangle \mid s \in \Sigma\}$

② Definition 2. $\mathcal{F}_\tau^*(Y) = \mathbb{I}_\Sigma \cup Y \circ t$

$$\begin{aligned} & \alpha^*(\mathcal{F}_\tau^{\rightarrow*}(X)) && \\ &= \alpha^* (\{s \mid s \in \Sigma\} \cup \{\sigma s s' \mid \sigma s \in X \wedge \langle s, s' \rangle \in t\}) && \text{def. } \mathcal{F}_\tau^{\rightarrow*} \\ &= \{\vec{\alpha}(s) \mid s \in \Sigma\} \cup \{\vec{\alpha}(\sigma s s') \mid \sigma s \in X \wedge \langle s, s' \rangle \in t\} && \text{def. } \alpha^* \\ &= \{\langle s, s \rangle \mid s \in \Sigma\} \cup \{\langle \sigma_0, s' \rangle \mid \exists s : \sigma s \in X \wedge \langle s, s' \rangle \in t\} && \text{def. } \vec{\alpha} \\ &= \mathbb{I}_\Sigma \cup \{\langle \sigma_0, s' \rangle \mid \exists s : \langle \sigma_0, s \rangle \in \alpha^*(X) \wedge \langle s, s' \rangle \in t\} && \text{def. } \mathbb{I}_\Sigma, \alpha^* \\ &= \mathbb{I}_\Sigma \cup \alpha^*(X) \circ t \\ &= \mathcal{F}_\tau^*(\alpha^*(X)) \end{aligned}$$

The Reflexive Transitive Closure Semantics in Fixpoint Form - Proof

Showing

$$\alpha^*(\Sigma_{\tau}^{\rightarrow*}) = \text{lfp } \mathcal{F}_{\tau}^*$$

is equivalent to prove that

$$\alpha^*\left(\bigcup_{n \geq 0} \mathcal{F}_{\tau}^{\rightarrow*n}(\phi)\right) = \bigcup_{n \geq 0} \mathcal{F}_{\tau}^{*n}(\phi)$$

Using induction on

$$\forall n : \alpha^*(\mathcal{F}_{\tau}^{\rightarrow*n}(\phi)) = \mathcal{F}_{\tau}^{*n}(\phi)$$

The Reflexive Transitive Closure Semantics in Fixpoint Form - Proof

$$\forall n : \alpha^*(\mathcal{F}_\tau^{\rightarrow n}(\phi)) = \mathcal{F}_\tau^{*n}(\phi)$$

① Base Case:

$$\alpha^*(\mathcal{F}_\tau^{\rightarrow 0}(\phi)) = \phi = \mathcal{F}_\tau^{*0}(\phi)$$

② Inductive Hypothesis: $\alpha^*(\mathcal{F}_\tau^{\rightarrow n}(\phi)) = \mathcal{F}_\tau^{*n}(\phi)$

$$\begin{aligned}\alpha^*(\mathcal{F}_\tau^{\rightarrow n+1}(\phi)) &= \alpha^*(\mathcal{F}_\tau^{\rightarrow}(\mathcal{F}_\tau^{\rightarrow n}(\phi))) \\ &= \mathcal{F}_\tau^*(\alpha^*(\mathcal{F}_\tau^{\rightarrow n}(\phi))) && \text{commutative} \\ &= \mathcal{F}_\tau^*\mathcal{F}_\tau^{*n}(\phi) && \text{inductive hypothesis} \\ &= \mathcal{F}_\tau^{*n+1}(\phi)\end{aligned}$$

The Reachability Semantics as an Abstraction

The reachability semantics of the transition system $\tau = \langle \Sigma, \Sigma_i, t \rangle$

$$\{s' \mid \exists s \in \Sigma_i : \langle s, s' \rangle \in t^*\}$$

is the set of states that are reachable from the initial states Σ_i .

The Reachability Semantics as an Abstraction

Definition $\text{post}[r]Z$: The right-image of the set Z by relation r

$$\text{post}[r]Z = \{s' \mid \exists s \in Z : \langle s, s' \rangle \in r\}$$

The Reachability Semantics as an Abstraction

Abstraction of the reflexive transitive closure semantics Y is defined as

$$\begin{aligned}\alpha^\bullet(Y) &= \{s' \mid \exists s \in \Sigma_i : \langle s, s' \rangle \in Y\} \\ &= \text{post}[Y]\Sigma_i\end{aligned}$$

Concretization of the reachability semantics Z is defined as

$$\gamma^\bullet(Z) = \{\langle s, s' \rangle \mid s \in \Sigma_i \implies s' \in Z\}$$

Galois Connection

We have the Galois Connection:

$$\alpha^\bullet(Y) \subseteq Z \iff Y \subseteq \gamma^\bullet(Z)$$

Proof.

$$\begin{aligned} \alpha^\bullet(Y) \subseteq Z &\iff \{s' \mid \exists s \in \Sigma_i : \langle s, s' \rangle \in Y\} \subseteq Z && \text{def. } \alpha^\bullet \\ &\iff \forall s' : \forall s \in \Sigma_i : \langle s, s' \rangle \in Y \implies s' \in Z && \text{def. } \subseteq \\ &\iff \forall \langle s, s' \rangle \in Y : s \in \Sigma_i \implies s' \in Z && \text{def. } \implies \\ &\iff Y \subseteq \{\langle s, s' \rangle \mid s \in \Sigma_i \implies s' \in Z\} && \text{def. } \subseteq \\ &\iff Y \subseteq \gamma^\bullet(Z) && \text{def. } \gamma^\bullet \end{aligned}$$

The Reachability Semantics in fixpoint form

- 1 Define $\mathcal{F}_\tau^\bullet(Z) = \Sigma_i \cup \text{post}[t]Z$.
- 2 Establish commutation property $\alpha^\bullet(\mathcal{F}_\tau^*(Y)) = \alpha^\bullet(\mathcal{F}_\tau^\bullet(Y))$

$$\begin{aligned} & \alpha^\bullet(\mathcal{F}_\tau^*(Y)) \\ &= \{s' \mid \exists s \in \Sigma_i : \langle s, s' \rangle \in (\mathbb{I}_\Sigma \cup Y \circ t)\} && \text{def. } \alpha^\bullet \& \mathcal{F}_\tau^* \\ &= \{s' \mid \exists s \in \Sigma_i : s' = s\} \cup \\ & \quad \{s' \mid \exists s \in \Sigma_i : \exists s'' : \langle s, s'' \rangle \in Y \wedge \langle s'', s' \rangle \in t\} && \text{def. } \mathbb{I}_\Sigma \& \circ \\ &= \Sigma_i \cup \{s' \mid \exists s'' \in \alpha^\bullet(Y) \wedge \langle s'', s' \rangle \in t\} && \text{def. } \alpha^\bullet \\ &= \alpha^\bullet(\mathcal{F}_\tau^\bullet(Y)) && \text{def } \mathcal{F}_\tau^\bullet(Z) \end{aligned}$$

- 3 By the fixpoint transfer theorem,

$$\alpha^\bullet(t^*) = \alpha^\bullet(\text{lfp } \mathcal{F}_\tau^*) = \text{lfp } \mathcal{F}_\tau^\bullet$$

The Interval Semantics as an Abstraction

The set of states of a transition system $\tau = \langle \Sigma, \Sigma_i, t \rangle$ is totally ordered $\langle \Sigma, < \rangle$ with extrema $-\infty$ and $+\infty$, the interval semantics $\alpha^{\perp-1}(\alpha^\bullet(t^*))$ of τ provides bounds on its reachable states $\alpha^\bullet(t^*)$:

$$\alpha^{\perp-1}(Z) = [\min Z, \max Z]$$

$$\min(\phi) = \infty \quad \max(\phi) = -\infty$$

Concretization:

$$\gamma^H([l, h]) = \{s \in \Sigma \mid l \leq s \leq h\}$$

Abstract implication:

$$[l, h] \sqsubseteq [l', h'] \iff (l' \leq l \wedge h \leq h')$$

Galois Connection

- We have the Galois Connection:

$$\alpha^H(Z) \sqsubseteq [l, h] \iff Z \subseteq \gamma^H([l, h])$$

Proof.

$$\begin{aligned} \alpha^H(Z) \sqsubseteq [l, h] &\iff [\min Z, \max Z] \sqsubseteq [l, h] && \text{def. } \alpha^H \\ &\iff l \leq \min Z \wedge \max Z \leq h && \text{def. } \sqsubseteq \\ &\iff Z \subseteq \{s \in \Sigma \mid l \leq s \leq h\} && \text{def. min\&max} \\ &\iff Z \subseteq \gamma^H([l, h]) && \text{def. } \gamma^H \end{aligned}$$

- By defining

$$\bigsqcup_{i \in I} [l_i, h_i] = [\min_{i \in I} l_i, \max_{i \in I} h_i]$$

, Galois connection preserves least upper bounds

$$\alpha^H\left(\bigcup_{i \in I} Z_i\right) = \bigsqcup_{i \in I} \alpha^H(Z_i)$$

The Interval Semantics in Fixpoint Form

- 1 Define $[\min \Sigma_i, \max \Sigma_i] \cup \alpha^{\perp} \circ \text{post}[t] \circ \gamma^H(I) \sqsubseteq \mathcal{F}_{\tau}^H(I)$
- 2 Establish semi-commutation property

$$\alpha^{\perp}(\mathcal{F}_{\tau}^{\bullet}(Z)) \sqsubseteq \mathcal{F}_{\tau}^H(\alpha^{\perp}(Z))$$

$$\begin{aligned} \alpha^{\perp}(\mathcal{F}_{\tau}^{\bullet}(Z)) &= \alpha^{\perp}(\Sigma_i \cup \text{post}[t]Z) && \text{def } \mathcal{F}_{\tau}^{\bullet} \\ &= \alpha^{\perp}(\Sigma_i) \cup \alpha^{\perp}(\text{post}[t][Z]) && \text{Galois Connection} \\ &\sqsubseteq [\min \Sigma_i, \max \Sigma_i] \cup \alpha^{\perp}(\text{post}[t](\gamma^H(\alpha^{\perp}(Z)))) \\ &\sqsubseteq \mathcal{F}_{\tau}^H(\alpha^{\perp}(Z)) \end{aligned}$$

- 3 By the fixpoint approximation:

$$\alpha^{\perp}(\mathcal{F}_{\tau}^{\bullet}(t^*)) = \alpha^{\perp}(\text{lfp } \mathcal{F}_{\tau}^{\bullet}) \sqsubseteq \text{lfp } \mathcal{F}_{\tau}^H$$

Convergence Acceleration

In general, $\text{lfp } \mathcal{F}_\tau^H = \bigsqcup_{n \geq 0} \mathcal{F}_\tau^H(\phi = [+ \infty, - \infty])$ diverge. Example, the transition system

$$\langle \mathbb{Z}, \{0\}, \{\langle x, x' \rangle \mid x' = x + 1\} \rangle$$

of program `x := 0; while true do x := x + 1.`

$$\mathcal{F}_\tau^H([l, h]) = [0, 0] \cup [l + 1, h + 1]$$

It diverges: $[+\infty, -\infty], [0, 0], [0, 1], [0, 2], \dots$

Widening

To accelerate convergence, introduce a widening ∇ such that,

$$(X \sqsubseteq X \nabla Y) \wedge (Y \sqsubseteq X \nabla Y)$$

$$\begin{aligned} I^0 &= \phi = [+∞, -∞] \\ I^{n+1} &= I^n && \text{if } \mathcal{F}_\tau^H(I^n) \sqsubseteq I^n \\ &= I^n \nabla \mathcal{F}_\tau^H(I^n) && \text{otherwise.} \end{aligned}$$

limit I^λ is finite ($\lambda \in \mathbb{N}$) and is a fixpoint overapproximation

$$\text{lfp } \mathcal{F}_\tau^H \sqsubseteq I^\lambda$$

Example of Widening

An example of interval widening

- 1 choosing finite sequence

$$-\infty = r_0 < r_1 < \dots < r_k = +\infty$$

- 2

$$[+\infty, -\infty] \nabla [l, h] = [l, h]$$

$$[l, h] \nabla [l', h'] = [\text{if } l > l' \text{ then } \max\{r_i \mid r_i \leq l'\} \text{ else } l, \\ \text{if } h < h' \text{ then } \min\{r_i \mid h' \leq r_i\} \text{ else } h]$$

Example of Widening

Example, the transition system

$$\langle \mathbb{Z}, \{0\}, \{\langle x, x' \rangle \mid x' = x + 1\} \rangle$$

of program `x := 0; while x < 100 do x := x + 1.`

$$\mathcal{F}_\tau^H([l, h]) = [0, 0] \cup [l + 1, \min(99, h) + 1]$$

1 Sequence $r = -\infty < -1 < 0 < 1 < \infty$

2

$$I^0 = [+\infty, -\infty]$$

$$I^1 = [0, 0] \sqcup [1, 1] = [0, 1]$$

$$I^2 = [0, 1] \sqcup [0, 2] = [0, +\infty]$$

$$I^3 = [0, +\infty]$$

Narrowing

The limit of an iteration with widening can be improved by a narrowing Δ , such that

$$Y \sqsubseteq X \implies Y \sqsubseteq (X \Delta Y) \sqsubseteq X$$

All terms in the iterates with narrowing

$$\begin{aligned} J^0 &= I^\lambda \\ J^{n+1} &= J^n \Delta \mathcal{F}_\tau^H(J^0) \end{aligned}$$

improve the result obtained by widening.

$$\text{lfp } \mathcal{F}_\tau^H \sqsubseteq J^n \sqsubseteq I^\lambda$$

Example of Narrowing

$$[l, h] \triangle [l', h'] = [\text{if } \exists i : l = r_i \text{ then } l' \text{ else } l, \text{ if } \exists j : h = r_j \text{ then } h' \text{ else } h]$$

Example, the transition system

$$\langle \mathbb{Z}, \{0\}, \{\langle x, x' \rangle \mid x' = x + 1\} \rangle$$

of program `x := 0; while x < 100 do x := x + 1.`

$$J^0 = [0, +\infty]$$

$$J^1 = [0, +\infty] \triangle [0, 100] = [0, 100]$$

$$J^2 = [0, 100] \triangle [0, 100] = [0, 100]$$

Composition of Abstractions

The design of three abstractions of the partial trace semantics $\Sigma_{\tau}^{\rightarrow*}$ of a transition system τ was compositional. Composition of Galois connections is a Galois connection so the successive arguments on sound approximation do compose nicely.

$$\alpha^{\text{H}} \circ \alpha^{\bullet} \circ \alpha^*, \gamma^* \circ \gamma^{\bullet} \circ \gamma^{\text{H}}$$

Hierarchy of Semantics

The four semantics of a transition system $\tau = \langle \Sigma, \Sigma_i, t \rangle$ form a hierarchy

- 1 Partial traces $\Sigma_{\tau}^{\rightarrow*}$
- 2 Reflexive transitive closure $\alpha^*(\Sigma_{\tau}^{\rightarrow*})$
- 3 Reachability $\alpha^{\bullet} \circ \alpha^*(\Sigma_{\tau}^{\rightarrow*})$
- 4 Interval semantics $\alpha^{\perp} \circ \alpha^{\bullet} \circ \alpha^*(\Sigma_{\tau}^{\rightarrow*})$

Thanks

Thank you for listening.