Correctness of Abstract Interpretation

Deepak D'Souza and K. V. Raghavan

IISc

Recollection of Abstract Interpretation

It is a tuple (D, F_D, γ), such that

- $(D, \leq)$ is a complete join semi-lattice (aka the abstract lattice), with a least element $\bot$.
- Concretization function $\gamma_D : D \rightarrow 2^{State}$
- Monotone transfer function $(f_{LM} : D \rightarrow D) \in F_D$ for each node n and incoming edge L into n and outgoing edge M from n.
  - Junction nodes have identity transfer function.

An aside: Collecting semantics stated as an abstract interpretation

- Concrete lattice $C : (2^{State}, \subseteq)$, $\bot = \emptyset$, $\top = State$, $\sqcup = \cup$.
- Transfer function $f_{LM} = nstate'_{LM}$ for each node n and incoming edge L into n and outgoing edge M from n.
- $\gamma : C \to C$ is identity

An aside: Collecting semantics stated as an abstract interpretation

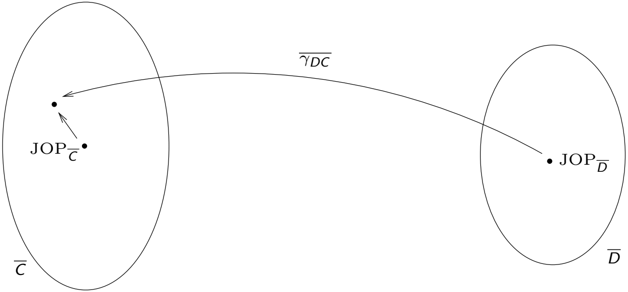
- Concrete lattice $C : (2^{State}, \subseteq)$, $\bot = \emptyset$, $\top = State$, $\sqcup = \cup$.
- Transfer function $f_{LM} = nstate'_{LM}$ for each node n and incoming edge L into n and outgoing edge M from n.
- $\gamma : C \to C$ is identity
- As seen earlier, collecting states at any point $N =$ JOP at this point using this interpretation
- This particular abstract interpretation is also known as the concrete interpretation.

Definition: consistent abstractions

An A.I. $(D, F_D, \gamma_D : D \to 2^{State})$ is said to be a consistent abstraction of (or, be correct wrt) another A.I. $(C, F_C, \gamma_C : C \to 2^{State})$ *under* a pair of monotone functions $\gamma_{DC} : D \to C$ and $\alpha_{CD} : C \to D$ iff:
(a) $(\alpha_{CD}, \gamma_{DC})$ form a Galois connection, and

(b) for all programs, and for all $d_0 \in D$ and $c_0 \in C$ such that $\gamma_{DC}(d_0) \geq c_0$:

$$\mathrm{JOP}_{\overline{C}} \leq \overline{\gamma_{DC}}(\mathrm{JOP}_{\overline{D}})$$



Definition – contd.

where

- $\mathrm{JOP}_{\overline{C}}$ is obtained by using (C, f_C), with c_0 as the initial state,
- $\mathrm{JOP}_{\overline{D}}$ is by obtained using (D, f_D), with d_0 as the initial state, and
- $\overline{x}$ is the "vectorized" form of x, i.e., x for all points in a program.

Note: Throughout remaining slides we use γ to mean γ_{DC} and α to mean α_{CD}.

Definition: (α, γ) form Galois Connection

- α and γ are monotonic
- $\gamma(\alpha(e)) \geq e$, for all $e \in C$
- $\alpha(\gamma(d)) = d$, for all $d \in D$

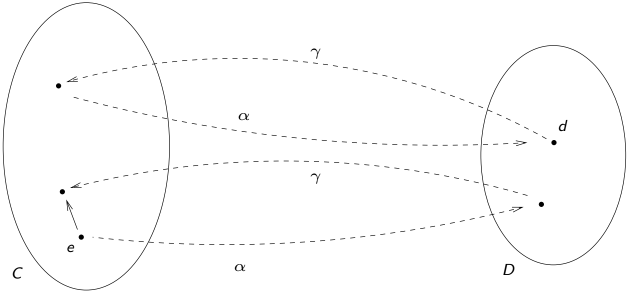


Illustration of consistent abstraction

- Consider the lattices L_1 and L_2 from the introduction slides.
- L_1 is a consistent abstraction of L_2 under the following (α, γ):

$$\begin{array}{rcl} \alpha(S \in L_2) & = & \bot, \text{ if } S = \emptyset \\ & = & (coll(\{x \mid (x,y) \in S\}), coll(\{y \mid (x,y) \in S\})), \\ & & \text{otherwise} \\ \gamma((c,d) \in L_1) & = & \{(x,y) \mid \text{if } c \text{ is } oe \text{ then } x = o \vee x = e \text{ else } x = c, \\ & & \text{if } d \text{ is } oe \text{ then } y = o \vee y = e \text{ else } y = d\} \end{array}$$

where

$$\begin{array}{rcl} coll(W) & = & o, \text{ if } W = \{o\} \\ & = & e, \text{ if } W = \{e\} \\ & = & oe, \text{ if } W = \{o, e\} \end{array}$$

Another illustration of consistent abstraction

Constant propagation (CP) is a consistent abstraction of the concrete interpretation, under the following (α, γ):

$$
\begin{array}{lcl}
\alpha(S \in 2^{State}) & = & \bot, \\
 & & \text{if } S \text{ is empty} \\
 & = & \{(x, c) \mid \forall e \in S : \ e(x) = c\}, \\
 & & \text{otherwise} \\
\gamma(p) & = & \emptyset, \\
 & & \text{if } p = \bot \\
 & = & \{e \in State \mid \text{for each } (x, c) \in p : e(x) = c\}, \\
 & & \text{if } p \text{ is any other element of the lattice}
\end{array}
$$

Properties of consistent abstractions

- Note: If an abstract interpretation $(D, F_D, \gamma : D \rightarrow 2^{State})$ is a consistent abstraction of $(2^{State}, nstate', identity)$, then we say that $(D, F_D, \gamma : D \rightarrow 2^{State})$ is correct.
- Consistent-abstraction-of is a transitive property. That is, if $(D, F_D, \gamma_D : D \rightarrow 2^{State})$ is a consistent abstraction of $(C, F_C, \gamma_C : C \rightarrow 2^{State})$ under $\gamma_{DC} : D \rightarrow C$, and $(C, F_C, \gamma_C : C \rightarrow 2^{State})$ is a consistent abstraction of $(B, F_B, \gamma_B : B \rightarrow 2^{State})$ under $\gamma_{CB} : C \rightarrow B$, then $(D, F_D, \gamma_D : D \rightarrow 2^{State})$ is a consistent abstraction of $(B, F_B, \gamma_B : B \rightarrow 2^{State})$ under $\gamma_{CB} \circ \gamma_{DC}$.

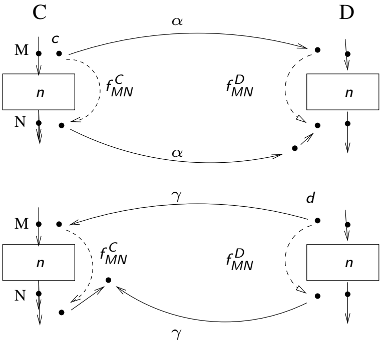
A sufficient condition for correctness

Theorem 1: An abstract interpretation (D, F_D, γ_D) is a consistent abstraction of another abstract interpretation (C, F_C, γ_C) under a pair (α, γ) if

- (α, γ) form a Galois connection, and
- Each transfer function $f^D_{MN} \in F_D$ is an abstraction of the corresponding function $f^D_{MN} \in F_C$.

Definition: f^D_{MN} is an abstraction of f^C_{MN}

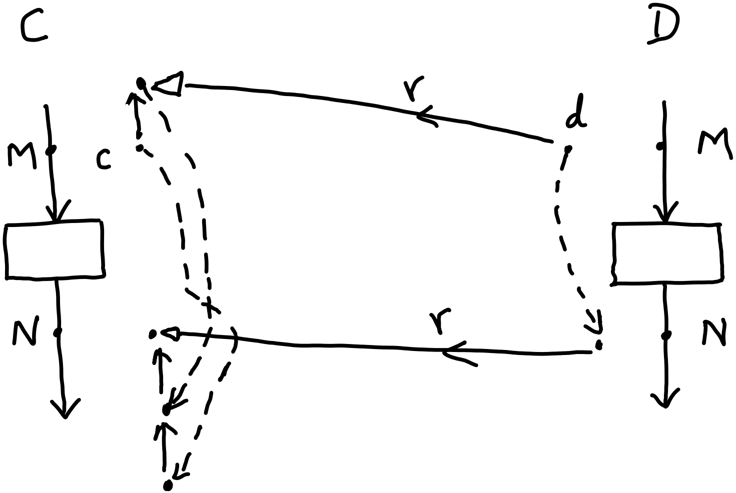
f^C_{MN} and f^D_{MN} satisfy *one* of the following (each of them implies the other):



Lemma 1

Statement: Consider any edge $M \rightarrow N$. If d is any element of D and c is any element of C such that $\gamma(d) \geq c$, then $\gamma(f^D_{MN}(d)) \geq f^C_{MN}(c)$.
Proof: The second condition on transfer functions tells us that $\gamma(f^D_{MN}(d)) \geq f^C_{MN}(\gamma(d))$. Using the lemma's prerequisite $\gamma(d) \geq c$, and by monotonicity of f^C_{MN}, we get $\gamma(f^D_{MN}(d)) \geq f^C_{MN}(c)$.

Lemma 1 proof illustration

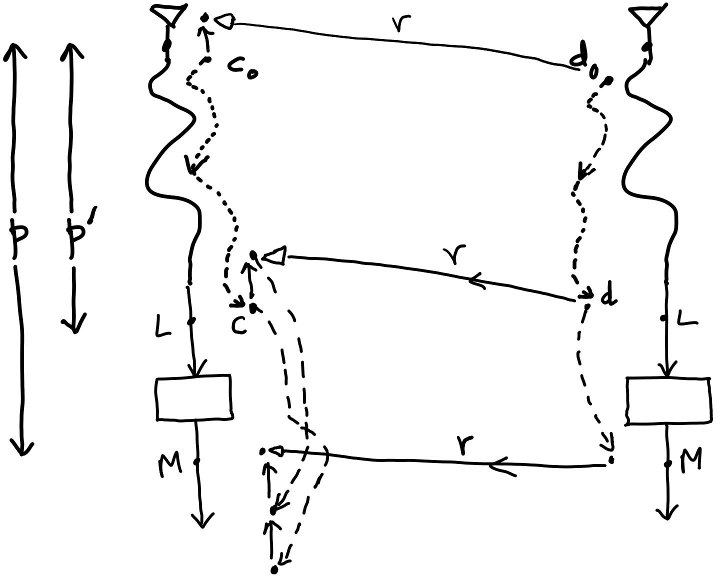


Lemma 2

Lemma 2: If $\gamma(d_0) \geq c_0$, then for any path p, $\gamma(f_p^D(d_0)) \geq f_p^C(c_0)$.
Proof:
The proof is by induction on the length of the path p. Let i be the length of the path p.

- Base case ($i = 0$): The property to prove reduces to $\gamma(d_0) \geq c_0$. Recall that this is a pre-requisite of this lemma.
- Inductive case $i > 0$: Let p' denote the prefix of path p that excludes the last edge of p. The inductive hypothesis is that $\gamma(f_{p'}^D(d_0)) \geq f_{p'}^C(c_0)$. Let the last edge of p be $L \rightarrow M$. Applying Lemma 1 on this edge we get $\gamma(f_{LM}^D(f_{p'}^D(d_0))) \geq f_{LM}^C(f_{p'}^C(c_0))$. This reduces to $\gamma(f_p^D(d_0)) \geq f_p^C(c_0)$. The inductive case is done.

Illustration of inductive case of Lemma 2



Proof of Theorem 1

Given $d_0 \in D$ and $c_0 \in C$ such that $\gamma(d_0) \geq c_0$. Pick any point N in the given program. Let P_N be the set of paths that begin at point I and end at N.

- By Lemma 2, for any path $p \in P_N$, we infer $\gamma(f_p^D(d_0)) \geq f_p^C(c_0)$.
- The result above implies:

$$\bigsqcup_{p \in P_N} (\gamma(f_p^D(d_0))) \geq \bigsqcup_{p \in P_N} (f_p^C(c_0)) \qquad (1)$$

- By monotonicity of γ, we infer:

$$\gamma(\bigsqcup_{p \in P_N} (f_p^D(d_0))) \geq \bigsqcup_{p \in P_N} (\gamma(f_p^D(d_0))) \qquad (2)$$

Proof of Theorem 1 – continued

- Using transitivity, Equations (1) and (2) imply:

$$\gamma\Big(\bigsqcup_{p \in P_N} (f_p^D(d_0))\Big) \geq \bigsqcup_{p \in P_N} (f_p^C(c_0)) \qquad (3)$$

- Using the definition of abstract JOP, we infer:

$$\gamma(\mathrm{JOP}_D^N) \geq \mathrm{JOP}_C^N \qquad (4)$$

- Hence, we get:

$$\overline{\gamma_{DC}}(\mathrm{JOP}_{\overline{D}}) \;\geq\; \mathrm{JOP}_{\overline{C}} \qquad (5)$$

More theorems

1. If α, γ form a Galois connection between $(D, F_D, \gamma_D$ and (C, F_C, γ_C), then for all $d_1, d_2 \in D$, $\gamma(d_1 \sqcap d_2) = \gamma(d_1) \sqcap \gamma(d_2)$.

This has an interesting application:

- If $d_{1,N}$ is the JOP at a point N due to a correct abstract interpretation $(D, F_{1,D}, \gamma_D)$ and if $d_{2,N}$ is the JOP at point N due to another correct abstract interpretation $(D, F_{2,D}, \gamma_D)$ (both JOPs computed using a common entry value $d_0 \in D$), then $d_{1,N} \sqcap d_{2,N}$ is more precise than $d_{1,N}$ or $d_{1,N}$ individually as an abstract JOP, while still over-approximating the collecting semantics.
- Alternatively, for each edge MN, we can use the "meet" transfer function $f_{MN} \equiv f_{1,MN} \sqcap f_{2,MN}$, and compute the abstract JOP using these "meet" transfer functions. The abstract JOP obtained this way will be $\leq d_{1,N} \sqcap d_{2,N}$ mentioned in the preceding bullet, and will also over-approximate the collecting semantics.

More theorems

2. If α, γ is a Galois connection between $(D, F_D, \gamma_D$ and (C, F_C, γ_C), then for any $d \in D$, $\gamma(d)$ is equal to $\sqcup\{c \in C \mid \alpha(c) \sqsubseteq d\}$, and for any $c \in C$, $\alpha(c)$ is equal to $\sqcap\{d \in D \mid \gamma(d) \sqsupseteq c\}$.

- Note, this does *not* imply that for every monotone function γ (resp. α), there exists an α (resp. γ) such that (α, γ) form a Galois connection.

3. If (α, γ) form a Galois connection, and each transfer function $f^D_{LM} \in F_D$ is an abstraction of the corresponding function $f^C_{LM} \in F_C$, then: γ-image of least solution of dataflow equations using (D, F_D, γ_D) dominates least solution of dataflow equations using (C, F_C, γ_C).