Correctness of Abstract Interpretation

Deepak D’Souza and K. V. Raghavan

IISc
Recollection of Abstract Interpretation

It is a tuple \((D, F_D, \gamma)\), such that

- \((D, \leq)\) is a complete join semi-lattice (aka the abstract lattice), with a least element \(\perp\).
- Concretization function \(\gamma : D \rightarrow 2^{\text{State}}\)
- 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 $C : (2^{\text{State}}, \subseteq), \bot = \emptyset, \top = \text{State}, \sqcup = \cup$.
- 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 $C : (2^{\text{State}}, \subseteq), \bot = \emptyset, \top = \text{State}, \sqcup = \cup$.
- Transfer function $f_{LM} = nstate'_{LM}$ for each node $n$ and incoming edge $L$ into $n$ and outgoing edge $M$ from $n$.
- $\gamma : C \rightarrow C$ is identity
- Therefore, collecting states at any point $N = \text{JOP}$ at this point using this interpretation
- This particular abstract interpretation is also known as the concrete interpretation.
Definition: consistent abstractions

An A.I. \((D, F_D, \gamma_D : D \rightarrow 2^{State})\) is said to be a consistent abstraction of (or, be correct wrt) another A.I. \((C, F_C, \gamma_C : C \rightarrow 2^{State})\) under a pair of monotone functions \(\gamma_{DC} : D \rightarrow C\) and \(\alpha_{CD} : C \rightarrow D\) 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 \(\gamma_{DC}(d_0) \geq c_0:\)

\[\text{JOP}_C \leq \gamma_{DC}(\text{JOP}_D)\]
where

- \( \text{JOP}_C \) is obtained by using \((C, f_C)\), with \( c_0 \) as the initial state,
- \( \text{JOP}_D \) 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 \( \gamma \) to mean \( \gamma_{DC} \) and \( \alpha \) to mean \( \alpha_{CD} \).
Definition: \((\alpha, \gamma)\) form Galois Connection

- \(\alpha\) and \(\gamma\) are monotonic
- \(\gamma(\alpha(e)) \geq e\)
- \(\alpha(\gamma(d)) = d\)

\[\begin{align*}
\text{C} &\quad \text{D} \\
e &\quad \text{d}
\end{align*}\]
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 $(\alpha, \gamma)$:

$$\alpha(S \in L_2) = \bot, \text{ if } S = \emptyset$$
$$= (\text{coll}(\{x \mid (x, y) \in S\}), \text{coll}(\{y \mid (x, y) \in S\})) \text{, otherwise}$$

$$\gamma((c, d) \in L_1) = \{(x, y) \mid \text{if } c \text{ is oe then } x = o \lor x = e \text{ else } x = c, \text{ if } d \text{ is oe then } y = o \lor y = e \text{ else } y = d\}$$

where

$$\text{coll}(W) = o, \text{ if } W = \{o\}$$
$$= e, \text{ if } W = \{e\}$$
$$= oe, \text{ if } W = \{o, e\}$$
Another illustration of consistent abstraction

Constant propagation (CP) is a consistent interpretation of the concrete interpretation, under the following \((\alpha, \gamma)\):

\[
\alpha(S \in 2^{\text{State}}) = \begin{cases} 
\bot, & \text{if } S \text{ is empty} \\
\{(x, c) \mid \forall e \in S : e(x) = c\}, & \text{otherwise}
\end{cases}
\]

\[
\gamma(p) = \begin{cases} 
\emptyset, & \text{if } p = \bot \\
\{e \in \text{State} \mid \text{for each } (x, c) \in p : e(x) = c\}, & \text{if } p \text{ is any other element of the lattice}
\end{cases}
\]
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: An abstract interpretation \((D, F_D, \gamma_D)\) is a consistent abstraction of another abstract interpretation \((C, F_C, \gamma_C)\) under a pair \((\alpha, \gamma)\) if

- Each transfer function \(f_{LM,D} \in F_D\) is an abstraction of the corresponding function \(f_{LM,C} \in F_C\).
If \((\alpha, \gamma)\) form a Galois connection then the concrete and abstract join operators satisfy the following properties.

Lemma 1

Lemma 2
Proof of lemmas

Proof of Lemma 2:

- \( d_1 \sqcup d_2 \) is \( \geq \) both \( d_1 \) and \( d_2 \) (property of join).
- Therefore, due to monotonicity of \( \gamma \), \( \gamma(d_1 \sqcup d_2) \) is \( \geq \) both \( \gamma(d_1) \) and \( \gamma(d_2) \).
- Therefore, by property of join, \( \gamma(d_1 \sqcup d_2) \geq \gamma(d_1) \sqcup \gamma(d_2) \). \( \square \).

Proof of Lemma 1:

- Using an argument similar to above it can be shown that \( \alpha(c_1 \sqcup c_2) \geq \alpha(c_1) \sqcup \alpha(c_2) \).
- Let \( c_3 \equiv c_1 \sqcup c_2 \), \( d_1 \equiv \alpha(c_1) \), \( d_2 \equiv \alpha(c_2) \), \( d_3 \equiv d_1 \sqcup d_2 \), and \( d_4 \equiv \alpha(c_3) \).
- We now prove that \( \alpha(c_1 \sqcup c_2) \sqsubseteq \alpha(c_1) \sqcup \alpha(c_2) \) is not possible.
  Assume, for contradiction, that \( d_4 \sqsubseteq d_3 \).
- Due to Galois connection property, \( \gamma(d_1) \sqsupseteq c_1 \) and \( \gamma(d_2) \sqsupseteq c_2 \). Now, since \( d_3 \) dominates \( d_1 \) and \( d_2 \), due to monotonicity of \( \gamma \), it follows that \( \gamma(d_3) \) dominates \( c_1 \) and \( c_2 \). Therefore, \( \gamma(d_3) \) dominates \( c_3 \).
Proof of Lemma 1 – continued

Case 1

Case 2

Now, one of the following two cases has to hold.

- **Case 1**: \( \gamma(d_3) = c_3 \). Following the property of Galois connections, we get \( \alpha(c_3) = d_3 \). But this contradicts \( \alpha(c_3) = d_4 \).
- **Case 2**: \( \gamma(d_3) = c_4 \uplus c_3 \). Therefore, due to the Galois connection property, \( \alpha(c_4) = d_3 \). This, in conjunction with \( \alpha(c_3) = d_4 \), violates monotonicity of \( \alpha \).

Therefore, \( \alpha(c_1 \uplus c_2) = \alpha(c_1) \uplus \alpha(c_2) \).
Definition: $f_{n,D}$ is an abstraction of $f_{n,C}$

$f_{n,C}$ and $f_{n,D}$ satisfy one of the following (each of them implies the other):

\[ C \xrightarrow{\alpha} f_{n,C} \xleftarrow{\alpha} D \]

\[ C \xrightarrow{\gamma} f_{n,C} \xleftarrow{\gamma} D \]
Definition: $f_{n,D}$ is an abstraction of $f_{n,C}$

$f_{n,C}$ and $f_{n,D}$ satisfy one of the following (each of them implies the other):

Exercise: Given a statement $n = "x = x + y"$, and treating the constant propagation lattice as $D$, identify an element $c$ of the $2^{State}$ domain such that $\alpha(f_{n,C}(c)) \sqsubseteq f_{n,D}(\alpha(c))$. 
Lemma 3

Statement: Consider any edge $M \to N$. Let $d$ be an abstract value and $c$ be a concrete value at $M$ such that $\alpha(c) \leq d$. $\alpha(f_{MN,c}(c)) \leq f_{MN,D}(d)$.

Proof: The first condition on transfer functions tells us that $\alpha(f_{MN,c}(c)) \leq f_{MN,D}(\alpha(c))$. Using the lemma’s prerequisite $\alpha(c) \leq d$, and by monotonicity of $f_{MN,D}$, we get $f_{MN,D}(\alpha(c)) \leq f_{MN,D}(d)$. Therefore $\alpha(f_{MN,c}(c)) \leq f_{MN,D}(d)$ □
Lemma 3 proof illustration
Lemma 4: If $\alpha(c_0) \leq d_0$, then $\overline{\alpha}(\text{JOP}_C) \leq \text{JOP}_D$.

Proof:
Consider any path $p$ in the CFG starting from the entry point $E$. We will first prove using induction that for any $i \geq 0$, where $p^i$ is the prefix of $p$ containing $i$ edges, $\alpha(f_{p^i,C}(c_0)) \leq f_{p^i,D}(d_0)$, where $f_{p^i,C}$ ($f_{p^i,D}$) is the composition of the concrete (abstract) transfer functions of the edges in $p^i$.

- Base case ($i = 0$): The property to prove reduces to $\alpha(c_0) \leq d_0$. Recall that this is a pre-requisite of this lemma.

- Inductive case: The inductive hypothesis is that $\alpha(f_{p^{i-1},C}(c_0)) \leq f_{p^{i-1},D}(d_0)$. Let the $i^{th}$ edge of $p$ be $L \to M$. Applying Lemma 3 on this edge we get $\alpha(f_{LM,C}(f_{p^{i-1},C}(c_0))) \leq f_{LM,D}(f_{p^{i-1},D}(d_0))$. This reduces to $\alpha(f_{p^i,C}(c_0)) \leq f_{p^i,D}(d_0)$. The inductive case is done.
Illustration of inductive case of Lemma 4
Lemma 4 – continued

- From the result proved above we derive
  
  \[ \alpha(c_p) \leq d_p \]  
  
  where \( p \) is any path, \( c_p = f_{p,C}(c_0) \) and \( d_p = f_{p,D}(d_0) \).

- Let \( N \) be any program point, and let
  \[ P_N = \{ p \mid p \text{ is a path from } E \text{ to } N \}. \]
Lemma 4 – continued

- Property (1), plus the property of joins, gives us

\[ \bigvee_{p \in P_N} (\alpha(c_p)) \leq \bigvee_{p \in P_N} (d_p) \]  \hspace{1cm} (2)

\[ = \text{JOP}_D[N] \]  \hspace{1cm} (3)

- By Lemma 1 we have

\[ \bigvee_{p \in P_N} (\alpha(c_p)) = \alpha( \bigvee_{p \in P_N} (c_p)) \]  \hspace{1cm} (4)

\[ = \alpha(\text{JOP}_C[N]) \]  \hspace{1cm} (5)

- Using Properties 3 and 5, and extending over all program points \( N \) we get

\[ \overline{\alpha}(\text{JOP}_C) \leq \text{JOP}_D \]

We are done.
Proof of main theorem

Pick any \( c_0 \in C \) and \( d_0 \in D \) such that \( \gamma(d_0) \geq c_0 \).

\[
\begin{align*}
\alpha(\gamma(d_0)) & \geq \alpha(c_0) \quad \text{(monotonicity of } \alpha) \\
d_0 & \geq \alpha(c_0) \quad \text{(Galois connection property)} \\
\overline{\alpha}(JOP_{\overline{C}}) & \leq JOP_{\overline{D}} \quad \text{(Lemma 4)} \\
\overline{\gamma}(\overline{\alpha}(JOP_{\overline{C}})) & \leq \overline{\gamma}(JOP_{\overline{D}}) \quad \text{(monotonicity of } \gamma) \\
JOP_{\overline{C}} & \leq \overline{\gamma}(JOP_{\overline{D}}) \quad \text{(property of Galois connection)}
\end{align*}
\]