Static Program Analysis

Part 10 – abstract interpretation

http://cs.au.dk/~amoeller/spa/

Anders Møller & Michael I. Schwartzbach

Computer Science, Aarhus University

Agenda

- **Collecting semantics**
- Abstraction and concretization
- Soundness
- Optimality

Program semantics as constraint systems

$$ConcreteStates = Vars \rightarrow \mathbb{Z}$$

$$\{[\![v]\!]\} \subseteq ConcreteStates$$

The semantics of expressions

$$ceval : ConcreteStates \times E \to 2^{\mathbb{Z}}$$

$$ceval(\rho, X) = \{\rho(X)\}$$

$$ceval(\rho, I) = \{I\}$$

$$ceval(\rho, \texttt{input}) = \mathbb{Z}$$

$$ceval(\rho, E_1\ \mathtt{op}\ E_2) = \{v_1\ \mathtt{op}\ v_2 \mid v_1 \in ceval(\rho, E_1)\ \wedge\ v_2 \in ceval(\rho, E_2)\}$$

$$ceval(R, E) = \bigcup_{\rho \in R} ceval(\rho, E)$$

Successors and joins

$$csucc : ConcreteStates \times Nodes \rightarrow 2^{Nodes}$$

$$csucc(R, v) = \bigcup_{\rho \in R} csucc(\rho, v)$$

$$CJOIN(v) =$$

$$\{\rho \in ConcreteStates \mid \exists w \in Nodes : \rho \in [\![w]\!] \wedge v \in csucc(\rho, w)\}$$

Semantics of statements

$$[\![X{=}E]\!] = \{\rho[X \mapsto ceval(\rho, E)] \mid \rho \in CJOIN(v)\}$$

$$[\![\mathbf{var}\ X_1, \ldots, X_n]\!] = \{\rho[X_1 \mapsto z_1, \ldots, X_n \mapsto z_n] \mid \rho \in CJOIN(v) \wedge z_1 \in \mathbb{Z} \wedge \cdots \wedge z_n \in \mathbb{Z}\}$$

$$[\![v]\!] = CJOIN(v)$$

The resulting constraint system

$$\{\![v_1]\!\} = cf_1(\{\![v_1]\!\}, \ldots, \{\![v_n]\!\})$$

$$\{\![v_2]\!\} = cf_2(\{\![v_1]\!\}, \ldots, \{\![v_n]\!\})$$

$$\vdots$$

$$\{\![v_n]\!\} = cf_n(\{\![v_1]\!\}, \ldots, \{\![v_n]\!\})$$

$$cf(x_1, \ldots, x_n) = \big(cf_1(x_1, \ldots, x_n), \ldots, cf_n(x_1, \ldots, x_n)\big)$$

$$x = cf(x)$$

Example

```
var x;
x = 0;
while (input) {
  x = x + 2;
}
```


	solution 1	solution 2
$[\![entry]\!]$	$\{[\,]\}$	$\{[\,]\}$
$[\![\texttt{var x}]\!]$	$\{[\texttt{x} \mapsto z] \mid z \in \mathbb{Z}\}$	$\{[\texttt{x} \mapsto z] \mid z \in \mathbb{Z}\}$
$[\![\texttt{x = 0}]\!]$	$\{[\texttt{x} \mapsto 0]\}$	$\{[\texttt{x} \mapsto 0]\}$
$[\![\texttt{input}]\!]$	$\{[\texttt{x} \mapsto z] \mid z \in \{0, 2, 4, \ldots\}\}$	$\{[\texttt{x} \mapsto z] \mid z \in \mathbb{Z}\}$
$[\![\texttt{x = x + 2}]\!]$	$\{[\texttt{x} \mapsto z] \mid z \in \{2, 4, \ldots\}\}$	$\{[\texttt{x} \mapsto z] \mid z \in \mathbb{Z}\}$
$[\![exit]\!]$	$\{[\texttt{x} \mapsto z] \mid z \in \{0, 2, 4, \ldots\}\}$	$\{[\texttt{x} \mapsto z] \mid z \in \mathbb{Z}\}$

the least solution

A fixed point theorem for continuous functions

$f : L \to L$ is continuous if $\quad f(\bigsqcup A) = \bigsqcup_{a \in A} f(a) \quad$ for every $A \subseteq L$

If f is continuous:

$$fix(f) = \bigsqcup_{i \geq 0} f^i(\perp)$$

(even when L has infinite height!)

cf is continuous

Semantics vs. analysis

```
var a,b,c;
a = 42;
b = 87;
if (input) {
  c = a + b;
} else {
  c = a - b;
}
```


$$\{\![\texttt{b = 87}]\!\} = \{[\mathsf{a} \mapsto 42, \mathsf{b} \mapsto 87, \mathsf{c} \mapsto z] \mid z \in \mathbb{Z}\}$$
$$\{\![\texttt{c = a - b}]\!\} = \{[\mathsf{a} \mapsto 42, \mathsf{b} \mapsto 87, \mathsf{c} \mapsto -45]\}$$
$$\{\![\mathit{exit}]\!\} = \{[\mathsf{a} \mapsto 42, \mathsf{b} \mapsto 87, \mathsf{c} \mapsto 129], [\mathsf{a} \mapsto 42, \mathsf{b} \mapsto 87, \mathsf{c} \mapsto -45]\}$$

$$[\![\texttt{b = 87}]\!] = [\mathsf{a} \mapsto +, \mathsf{b} \mapsto +, \mathsf{c} \mapsto \top]$$
$$[\![\texttt{c = a - b}]\!] = [\mathsf{a} \mapsto +, \mathsf{b} \mapsto +, \mathsf{c} \mapsto \top]$$
$$[\![\mathit{exit}]\!] = [\mathsf{a} \mapsto +, \mathsf{b} \mapsto +, \mathsf{c} \mapsto \top]$$

Agenda

- Collecting semantics
- **Abstraction and concretization**
- Soundness
- Optimality

Abstraction functions for sign analysis

$$\alpha_{\mathrm{a}} : 2^{\mathbb{Z}} \to \mathit{Sign}$$
$$\alpha_{\mathrm{b}} : 2^{\mathit{ConcreteStates}} \to \mathit{States}$$
$$\alpha_{\mathrm{c}} : (2^{\mathit{ConcreteStates}})^n \to \mathit{States}^n$$

$$\alpha_{\mathrm{a}}(D) = \begin{cases} \bot & \text{if } D \text{ is empty} \\ + & \text{if } D \text{ is nonempty and contains only positive integers} \\ - & \text{if } D \text{ is nonempty and contains only negative integers} \\ \mathbf{0} & \text{if } D \text{ is nonempty and contains only the integer } 0 \\ \top & \text{otherwise} \end{cases}$$

for any $D \in 2^{\mathbb{Z}}$

$\alpha_{\mathrm{b}}(R) = \sigma$ where $\sigma(X) = \alpha_{\mathrm{a}}(\{\rho(X) \mid \rho \in R\})$
for any $R \subseteq \mathit{ConcreteStates}$ and $X \in \mathit{Vars}$

$\alpha_{\mathrm{c}}(R_1, \ldots, R_n) = (\alpha_b(R_1), \ldots, \alpha_b(R_n))$
for any $R_1, \ldots, R_n \subseteq \mathit{ConcreteStates}$

Concretization functions for sign analysis

$$\gamma_{\mathrm{a}} : \mathit{Sign} \to 2^{\mathbb{Z}}$$
$$\gamma_{\mathrm{b}} : \mathit{States} \to 2^{\mathit{ConcreteStates}}$$
$$\gamma_{\mathrm{c}} : \mathit{States}^n \to (2^{\mathit{ConcreteStates}})^n$$

$$\gamma_{\mathrm{a}}(s) = \begin{cases} \emptyset & \text{if } s = \bot \\ \{1, 2, 3, \dots\} & \text{if } s = + \\ \{-1, -2, -3, \dots\} & \text{if } s = - \\ \{0\} & \text{if } s = \mathbf{0} \\ \mathbb{Z} & \text{if } s = \top \end{cases}$$

for any $s \in \mathit{Sign}$

$$\gamma_{\mathrm{b}}(\sigma) = \{\rho \in \mathit{ConcreteStates} \mid \rho(X) \in \gamma_{\mathrm{a}}(\sigma(X)) \text{ for all } X \in \mathit{Vars}\}$$

for any $\sigma \in \mathit{States}$

$$\gamma_{\mathrm{c}}(\sigma_1, \dots, \sigma_n) = (\gamma_{\mathrm{b}}(\sigma_1), \dots, \gamma_{\mathrm{b}}(\sigma_n))$$

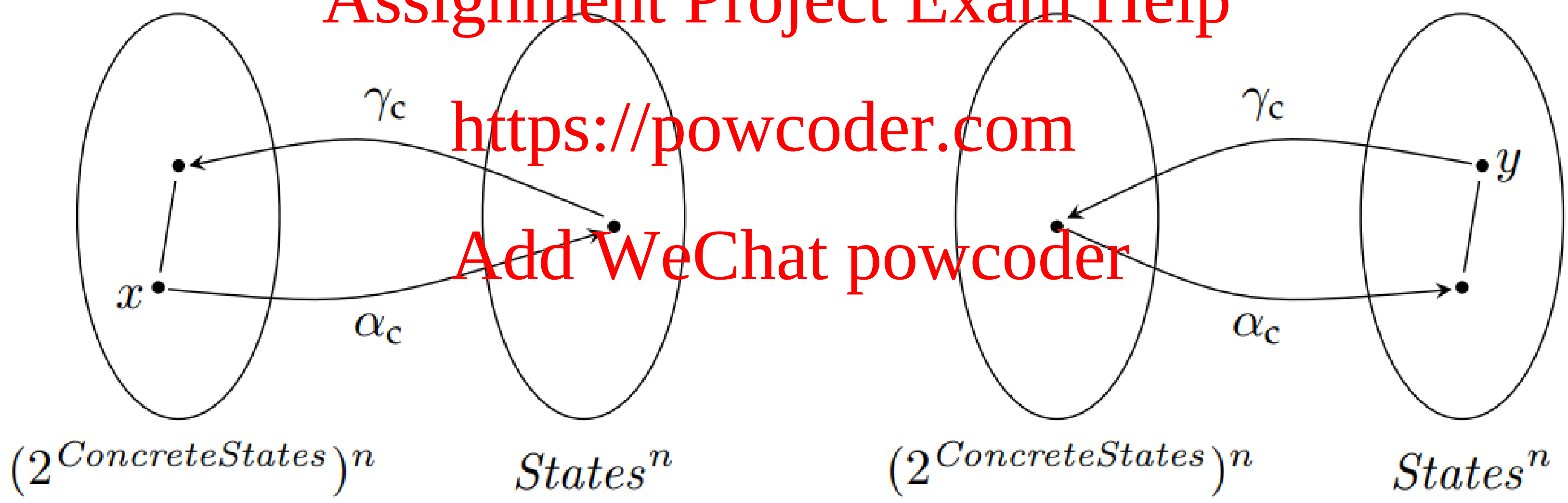
for any $(\sigma_1, \dots, \sigma_n) \in \mathit{States}^n$

Galois connections

The pair of monotone functions, α and γ, is called a *Galois connection* if

$$\gamma \circ \alpha \text{ is extensive}$$

$$\alpha \circ \gamma \text{ is reductive}$$



Galois connections

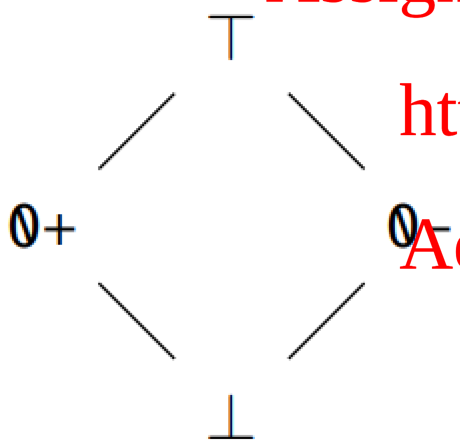
For Galois connections, the concretization function uniquely determines the abstraction function and vice versa:

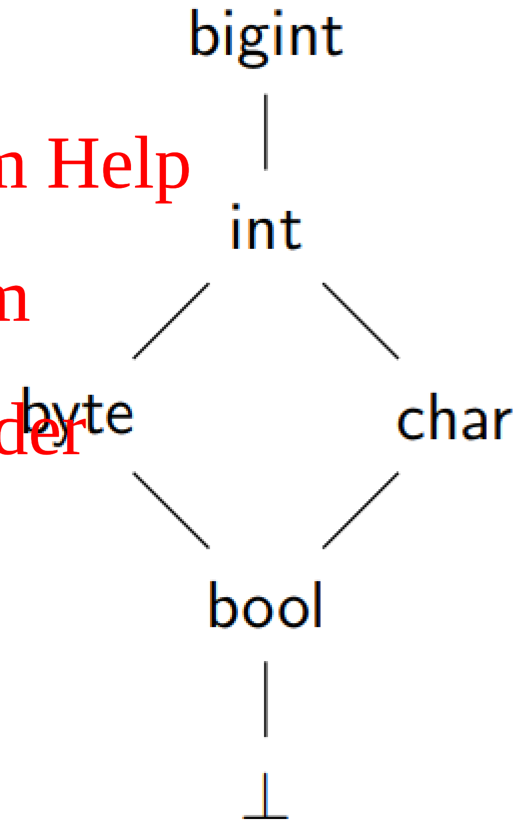
$$\gamma(y) = \bigsqcup_{x \in L_1 \text{ where } \alpha(x) \sqsubseteq y} x$$

$$\alpha(x) = \bigsqcap_{y \in L_2 \text{ where } x \sqsubseteq \gamma(y)} y$$

Galois connections

For each of these two lattices, given the “obvious” concretization function, is there an abstraction function such that the concretization function and the abstraction function form a Galois connection?



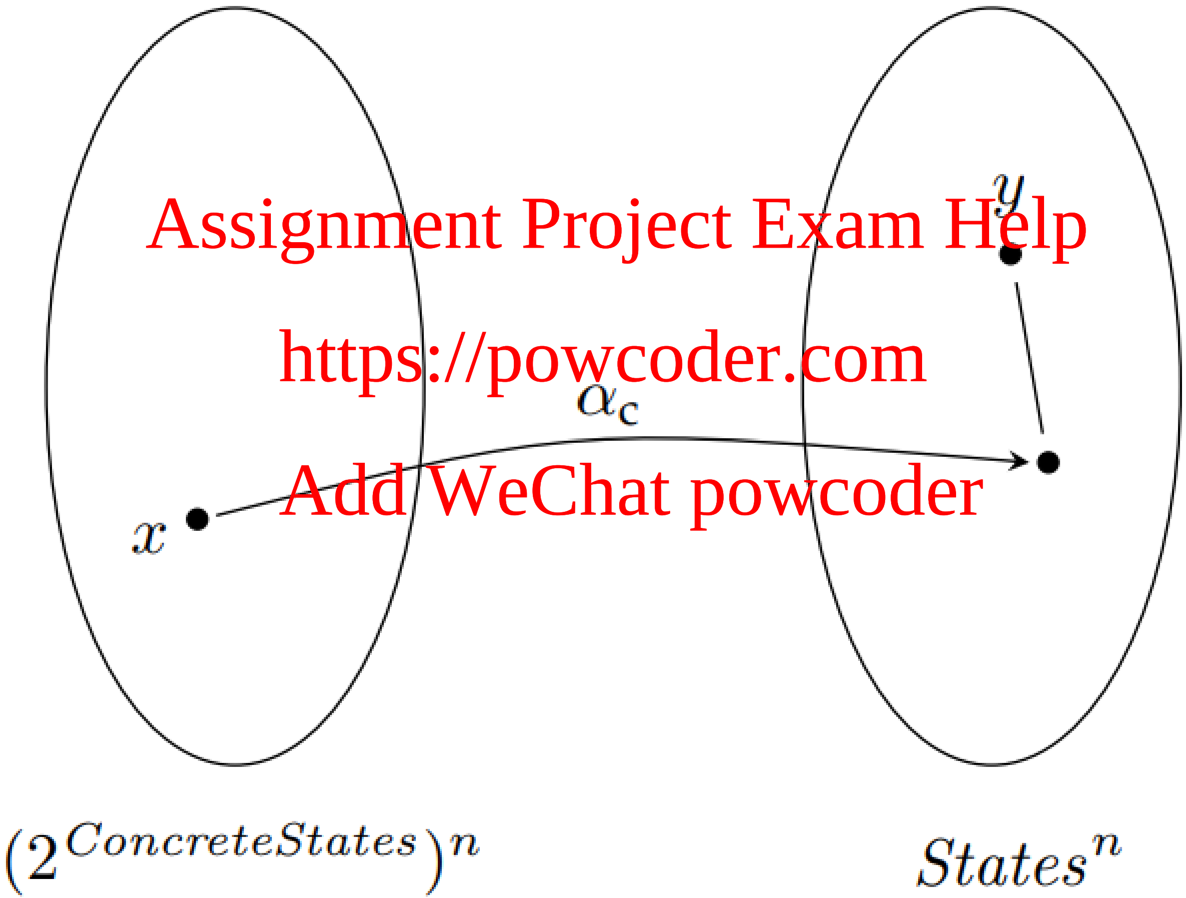


Agenda

- Collecting semantics
- Abstraction and concretization
- **Soundness**
- Optimality

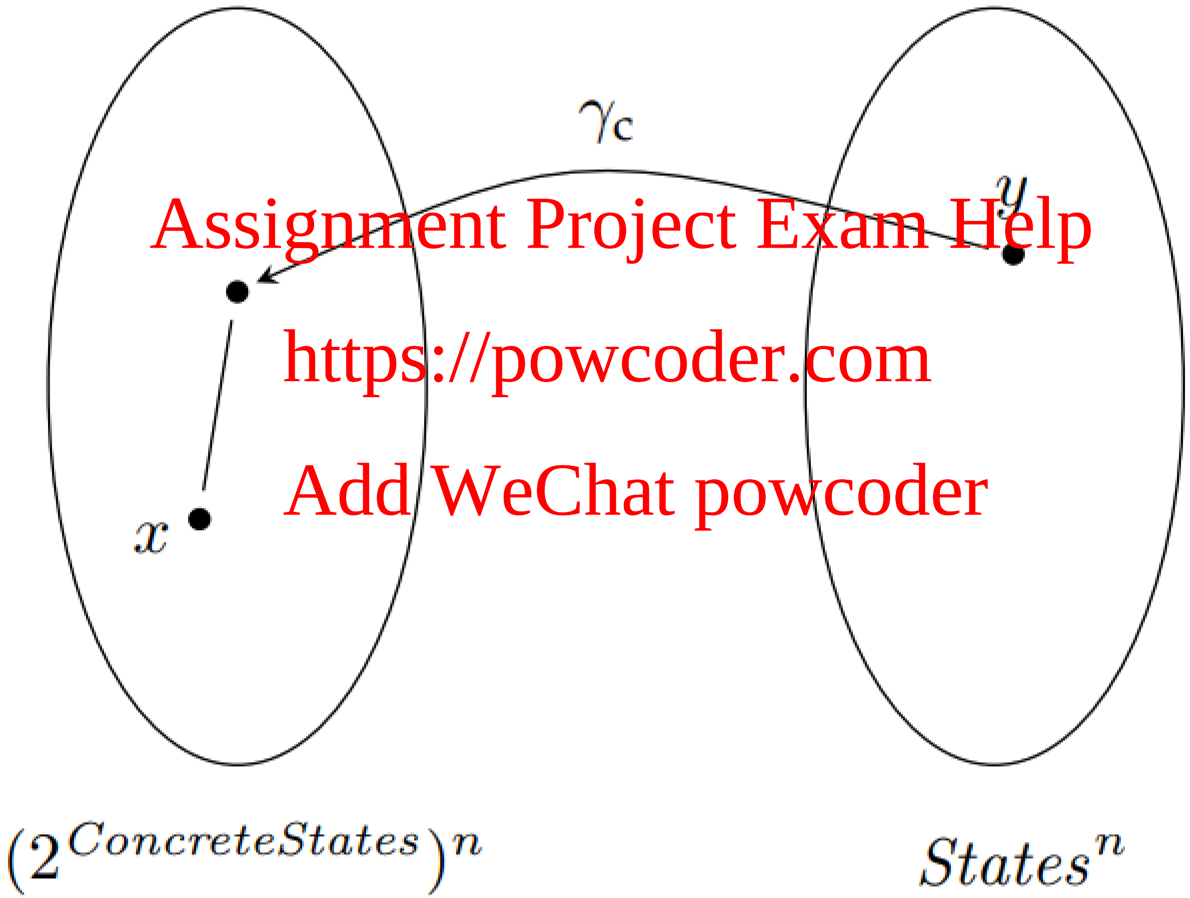
Soundness

$$\alpha(x) \sqsubseteq y$$



Soundness

$$x \sqsubseteq \gamma(y)$$



Safe approximations

$$\alpha_{\mathsf{a}}(ceval(R, E)) \sqsubseteq eval(\alpha_{\mathsf{b}}(R), E)$$

$$csucc(R, v) \subseteq succ(v) \text{ for any } R \subseteq ConcreteStates$$

$$\alpha_{\mathsf{b}}(CJOIN(v)) \sqsubseteq JOIN(v)$$

if $\alpha_{\mathsf{b}}(\{[\![w]\!]\}) \sqsubseteq [\![w]\!]$ for all $w \in Nodes$.

Safe approximations

if v represents an assignment statement $X = E$:

$$cf_v(\{\![v_1]\!\}, \ldots, \{\![v_n]\!\}) = \{\rho[X \mapsto z] \mid \rho \in CJOIN(v) \ \wedge \ z \in ceval(\rho, E)\}$$

$$af_v([\![v_1]\!], \ldots, [\![v_n]\!]) = \sigma[X \mapsto eval(\sigma, E)] \text{ where } \sigma = JOIN(v)$$

$$\alpha_{\mathsf{b}}(cf_v(R_1, \ldots, R_n)) \sqsubseteq af_v(\alpha_{\mathsf{b}}(R_1), \ldots, \alpha_{\mathsf{b}}(R_n))$$

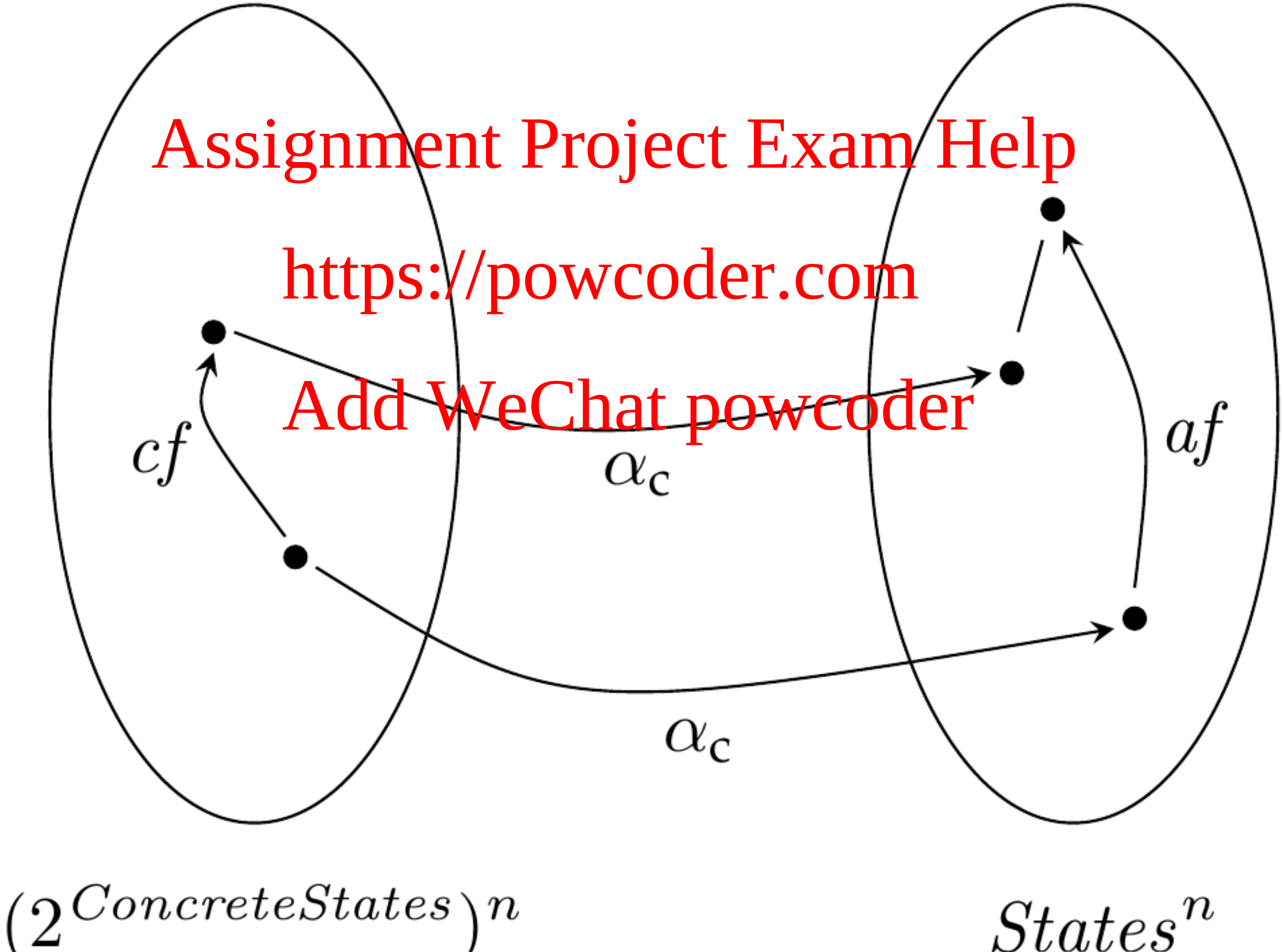
The two constraint systems

$$cf(\{\![v_1]\!\}, \ldots, \{\![v_n]\!\}) = ((cf_{v_1}(\{\![v_1]\!\}, \ldots, \{\![v_n]\!\}), \ldots, cf_{v_n}(\{\![v_1]\!\}, \ldots, \{\![v_n]\!\}))$$

$$af([\![v_1]\!], \ldots, [\![v_n]\!]) = ((af_{v_1}([\![v_1]\!], \ldots, [\![v_n]\!]), \ldots, af_{v_n}([\![v_1]\!], \ldots, [\![v_n]\!]))$$

Safe approximations

$$\alpha_{\mathrm{c}}(cf(R_1, \ldots, R_n)) \sqsubseteq af(\alpha_{\mathrm{c}}(R_1, \ldots, R_n))$$



The soundness theorem

Let L_1 and L_2 be lattices where L_2 has finite height, assume $\alpha : L_1 \to L_2$ and $\gamma : L_2 \to L_1$ form a Galois connection, $cf : L_1 \to L_1$ is continuous, and $af : L_2 \to L_2$ is monotone.

If af is a sound abstraction of cf, then $\alpha(fix(cf)) \sqsubseteq fix(af)$.

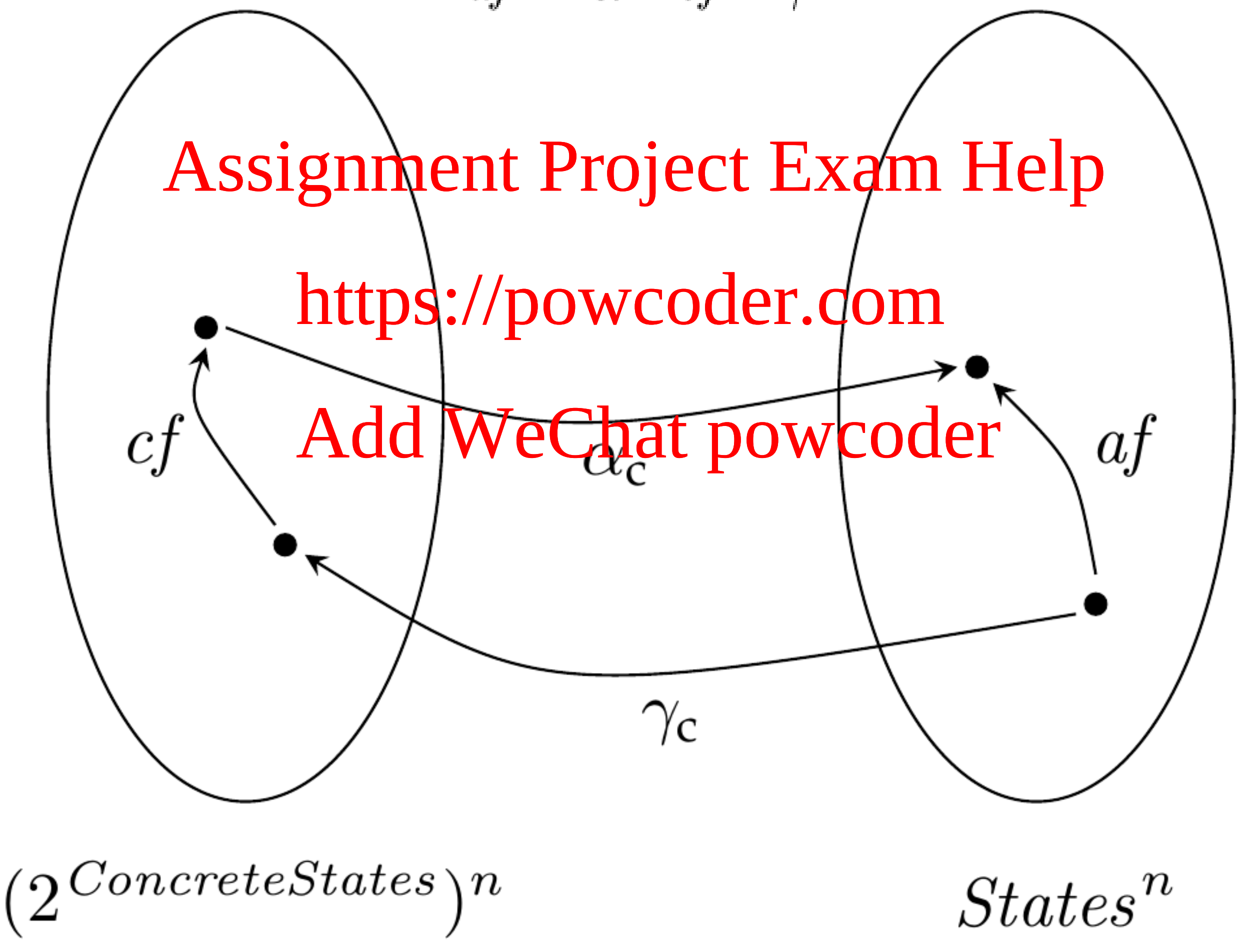
Agenda

- Collecting semantics
- Abstraction and concretization
- Soundness
- **Optimality**

Optimal approximations

af is an *optimal* approximation of cf if

$$af = \alpha \circ cf \circ \gamma$$



Optimal approximations in sign analysis?

$\widehat{*}$ is optimal:

$$s_1 \mathbin{\widehat{*}} s_2 = \alpha_{\mathrm{a}}(\gamma_{\mathrm{a}}(s_1) \cdot \gamma_{\mathrm{a}}(s_2))$$

$eval$ is *not* optimal:

$$\sigma(\mathbf{x}) = \top$$

$$eval(\sigma, \mathtt{x}\text{-}\mathtt{x}) = \top$$

$$\alpha_{\mathrm{b}}(ceval(\gamma_{\mathrm{b}}(\sigma), \mathtt{x}\text{-}\mathtt{x})) = \mathbf{0}$$

Even if we could make $eval$ optimal, the analysis result is not always optimal:

```
x = input;
y = x;
z = x - y;
```