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

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

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

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

$$\begin{aligned}
&ceval(\rho, X) = \{\rho(X)\}\\
&ceval(\rho, I) = \{I\}\\
&ceval(\rho, \texttt{input}) = \mathbb{Z}\\
&ceval(\rho, E_1 \,\texttt{op}\, E_2) = \{v_1 \,\texttt{op}\, v_2 \mid v_1 \in ceval(\rho, E_1) \ \wedge \ v_2 \in ceval(\rho, E_2)\}
\end{aligned}$$

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

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

$$CJOIN(v) = \bigcup_{\substack{w \in Nodes \text{ where} \\ v \in csucc(\{\!|w|\!\}, w)}} \{\!|w|\!\}$$

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

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

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

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

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

```
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

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

$$\alpha_{\mathrm{a}}(D) = \begin{cases} \bot & \text{if } D \text{ is empty} \\ + & \text{if } D \text{ is nonempty and contains only positive integers} \\ - & \text{if } D \text{ is nonempty and contains only negative integers} \\ \mathbf{0} & \text{if } D \text{ is nonempty and contains only the integer } 0 \\ \top & \text{otherwise} \end{cases}$$
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 $R_1, \ldots, R_n \subseteq \mathit{ConcreteStates}$

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

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

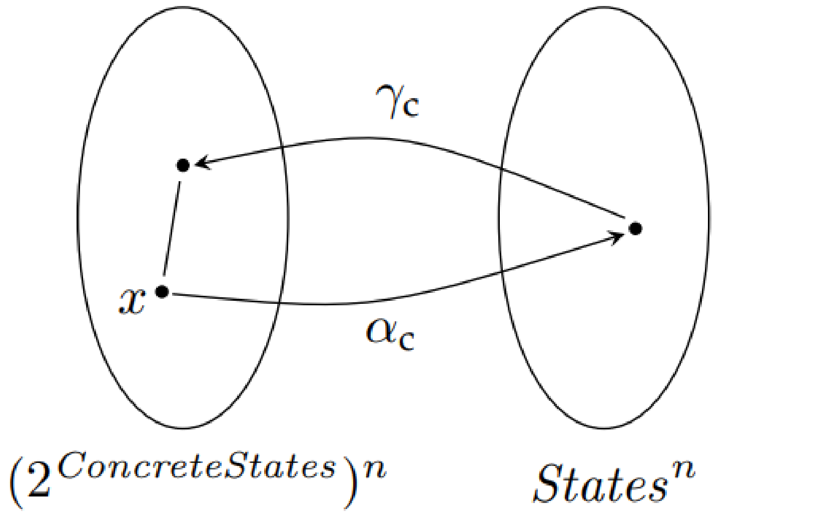
$$\gamma_{\mathrm{b}}(\sigma) = \{\rho \in \mathit{ConcreteStates} \mid \rho(X) \in \gamma_{\mathrm{a}}(\sigma(X)) \text{ for all } X \in \mathit{Vars}\}$$
for any $\sigma \in \mathit{States}$

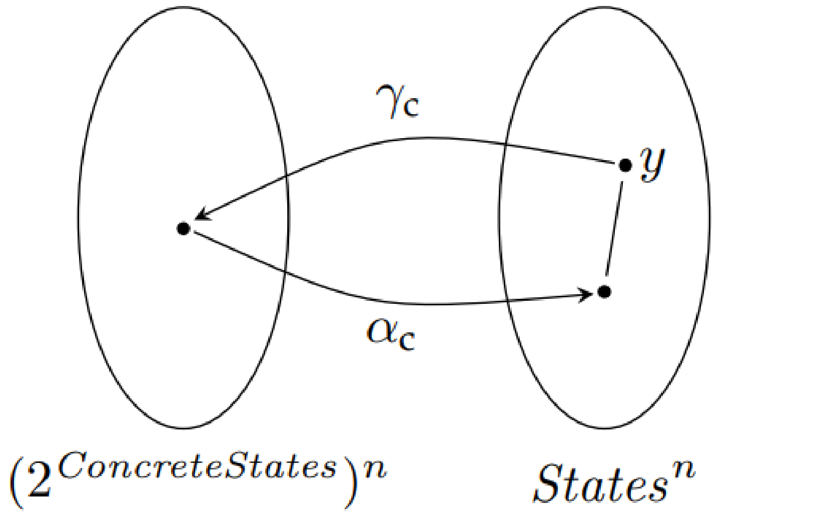
$$\gamma_{\mathrm{c}}(\sigma_1, \dots, \sigma_n) = (\gamma_{\mathrm{b}}(\sigma_1), \dots, \gamma_{\mathrm{b}}(\sigma_n))$$
for any $(\sigma_1, \dots, \sigma_n) \in \mathit{States}^n$

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

$\gamma \circ \alpha$ is extensive

$\alpha \circ \gamma$ is reductive





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

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

$$\begin{tikzcd}
 & \top \arrow[dl, no head] \arrow[dr, no head] & \\
\mathtt{0+} \arrow[dr, no head] & & \mathtt{0-} \arrow[dl, no head] \\
 & \bot &
\end{tikzcd}$$

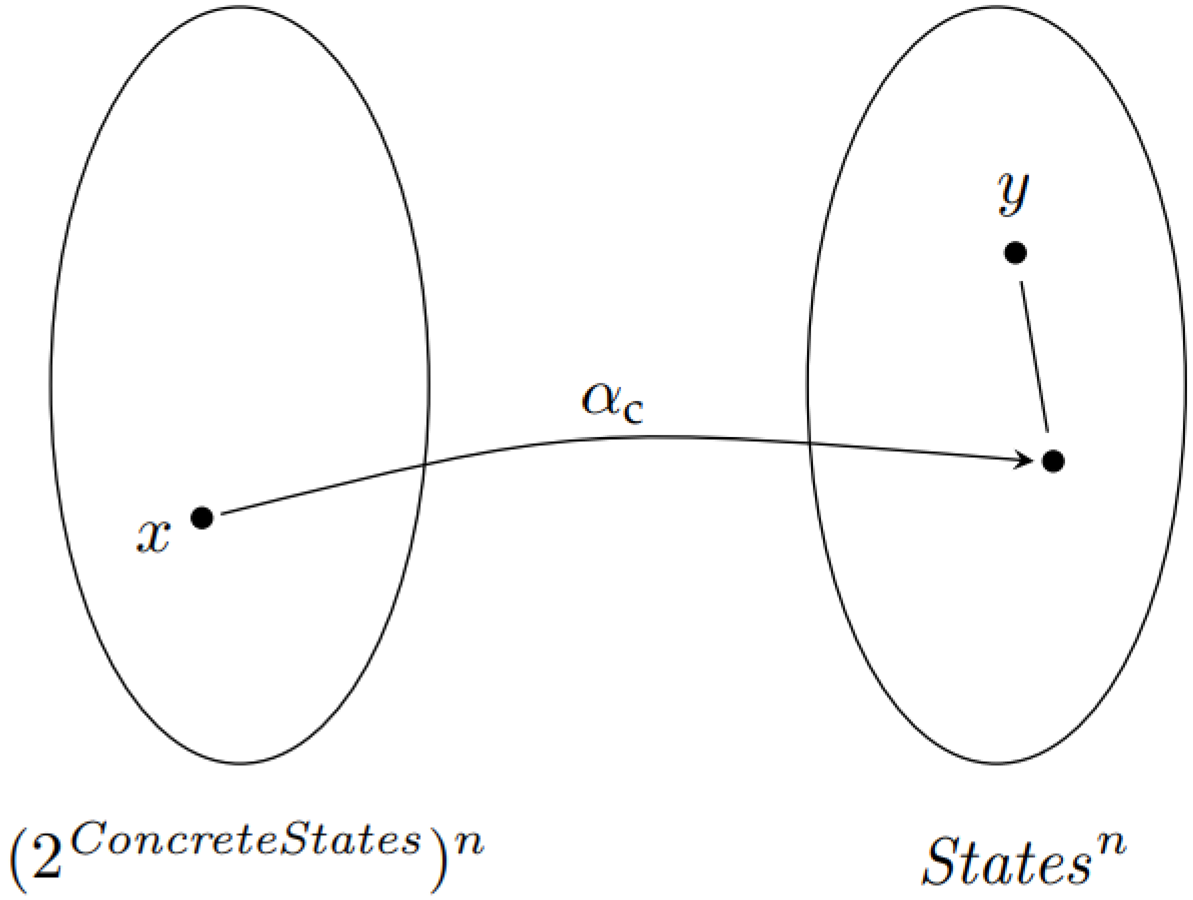
$$\begin{tikzcd}
 & \text{bigint} \arrow[d, no head] & \\
 & \text{int} \arrow[dl, no head] \arrow[dr, no head] & \\
\text{byte} \arrow[dr, no head] & & \text{char} \arrow[dl, no head] \\
 & \text{bool} \arrow[d, no head] & \\
 & \bot &
\end{tikzcd}$$

Agenda

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

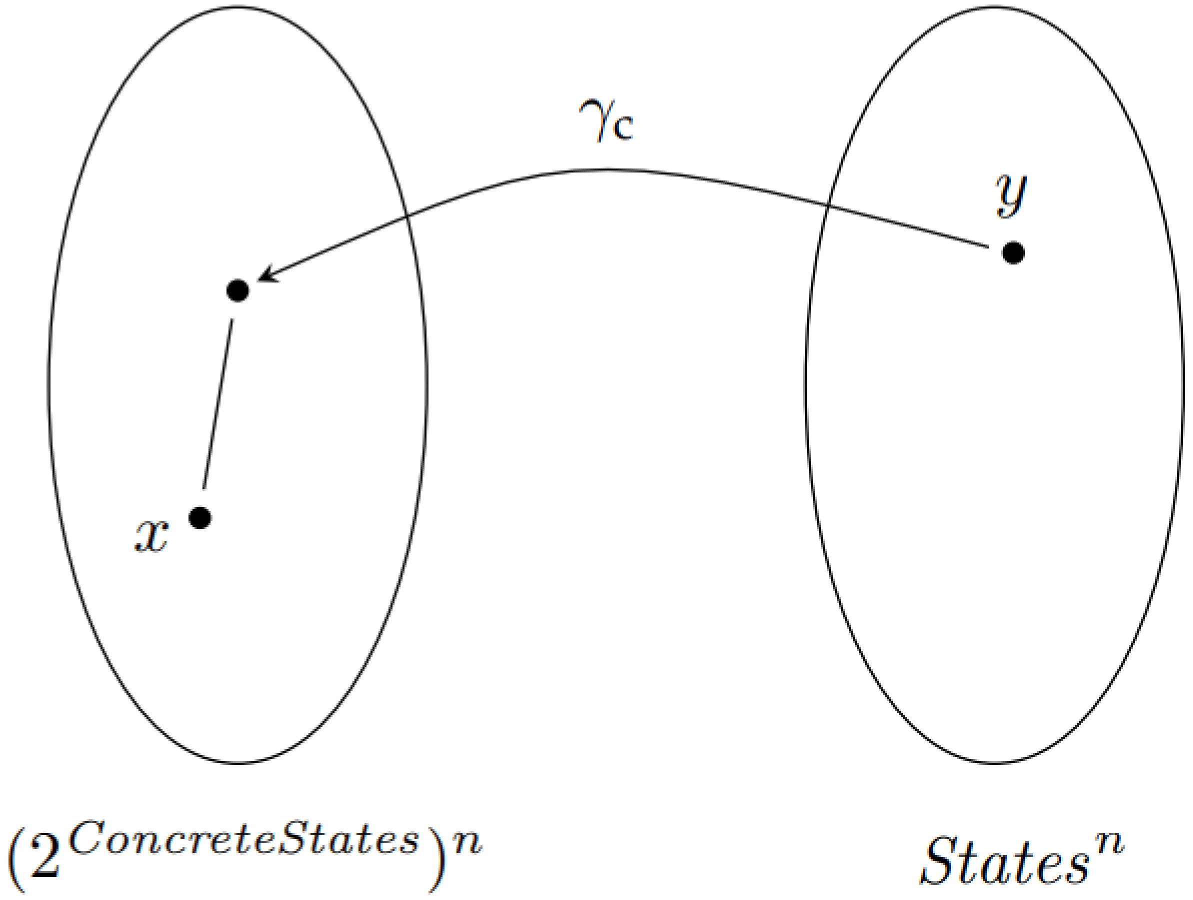
Soundness

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



Soundness

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



safe approximation

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

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

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

if $\alpha_{\mathsf{b}}(\{\llbracket w \rrbracket\}) \sqsubseteq \llbracket w \rrbracket$ for all $w \in Nodes$.

safe approximation

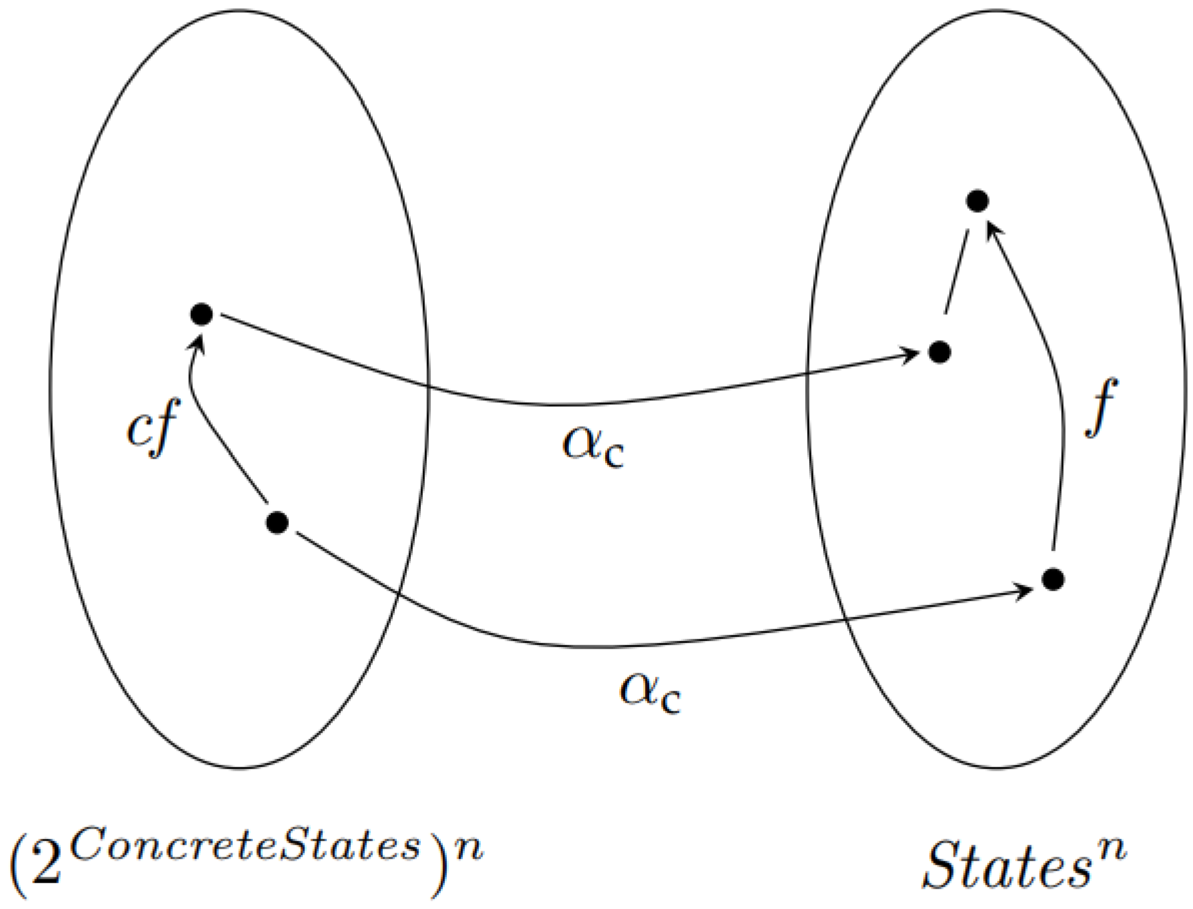
if v represents an assignment statement $X = E$:

$$cf_v(\{\![v_1]\!\}, \ldots, \{\![v_n]\!\}) = \{\rho[X \mapsto ceval(\rho, E)] \mid \rho \in CJOIN(v)\}$$
$$f_v([\![v_1]\!], \ldots, [\![v_n]\!]) = \sigma[X \mapsto eval(\sigma, E)] \text{ where } \sigma = JOIN(v)$$

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

safe approximation

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



Let L_1 and L_2 be lattices such that $\alpha : L_1 \to L_2$ and $\gamma : L_2 \to L_1$ form a Galois connection, and let $cf : L_1 \to L_1$ and $f : L_2 \to L_2$ be monotone functions.

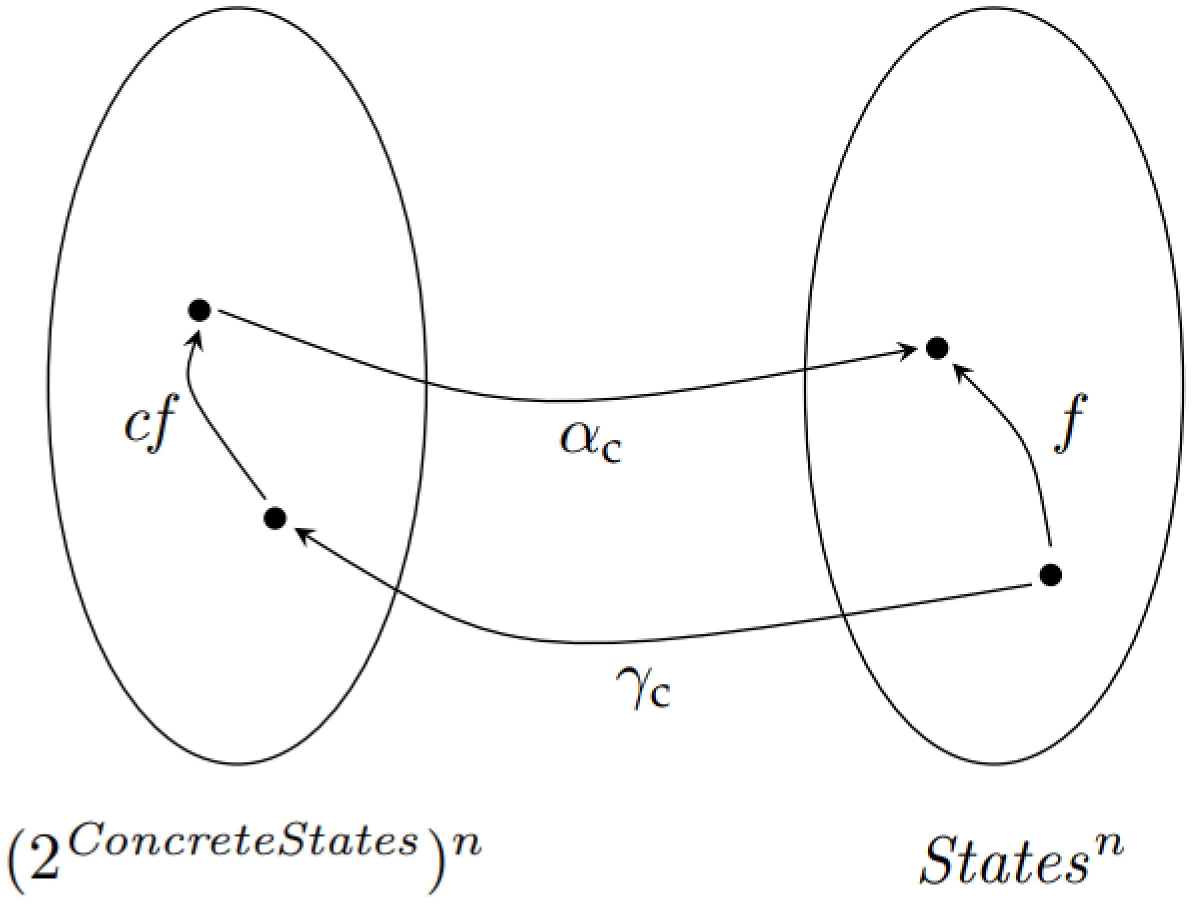
If f is a safe approximation of cf, then $\alpha(\mathit{fix}(cf)) \sqsubseteq \mathit{fix}(f)$.

Agenda

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

f is an *optimal* approximation of cf if

$$f = \alpha \circ cf \circ \gamma$$



$\widehat{*}$ is optimal:

$$s_1 \widehat{*} s_2 = \alpha_a\big(\gamma_a(s_1) * \gamma_a(s_2)\big)$$

eval is *not* optimal:

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

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

$$\alpha_b\big(ceval(\gamma_b(\sigma), \mathtt{x-x})\big) = \mathbf{0}$$

f is *not* optimal:

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