Principles of Program Analysis:

Control Flow Analysis

Transparencies based on Chapter 3 of the book: Flemming Nielson, Hanne Riis Nielson and Chris Hankin: Principles of Program Analysis. Springer Verlag 2005.

The Dynamic Dispatch Problem

$\vdots$

$[\texttt{call p(p1,1,v)}]^{\ell_c^1}_{\ell_r^1}$

$\vdots$

$[\texttt{call p(p2,2,v)}]^{\ell_c^2}_{\ell_r^2}$

$\vdots$

proc $\texttt{p(}$**procval** $\texttt{q,}$ **val** $\texttt{x,}$ **res** $\texttt{y)}$ **is**$^{\ell_n}$

$\vdots$

$[\texttt{call q (x,y)}]^{\ell_c^p}_{\ell_r^p}$

$\vdots$

end$^{\ell_x}$

which procedure is called?

These problems arise for:

- imperative languages with procedures as parameters
- object oriented languages
- functional languages

Example:

```
let f = fn x => x 1;
    g = fn y => y+2;
    h = fn z => z+3
in (f g) + (f h)
```

The aim of Control Flow Analysis:

For each function application, which functions may be applied?

Control Flow Analysis computes the interprocedural flow relation used when formulating interprocedural Data Flow Analysis.

Syntax of the Fun Language

Syntactic categories:

e	$\in$	**Exp**	expressions (or labelled terms)
t	$\in$	**Term**	terms (or unlabelled expressions)
f, x	$\in$	**Var**	variables
c	$\in$	**Const**	constants
op	$\in$	**Op**	binary operators
ℓ	$\in$	**Lab**	labels

Syntax:

$$e ::= t^{\ell}$$

$$\begin{array}{rcl} t & ::= & c \mid x \mid \texttt{fn}\ x\ \texttt{=>}\ e_0 \mid \texttt{fun}\ f\ x\ \texttt{=>}\ e_0 \mid e_1\ e_2 \\ & \mid & \texttt{if}\ e_0\ \texttt{then}\ e_1\ \texttt{else}\ e_2 \mid \texttt{let}\ x\ \texttt{=}\ e_1\ \texttt{in}\ e_2 \mid e_1\ op\ e_2 \end{array}$$

(Labels correspond to program points or nodes in the parse tree.)

Examples:

- $((\texttt{fn x => x}^1)^2\ (\texttt{fn y => y}^3)^4)^5$

- $(\texttt{let f = }(\texttt{fn x => }(\texttt{x}^1\ \texttt{1}^2)^3)^4\texttt{;}$
  $\texttt{in }(\texttt{let g = }(\texttt{fn y => y}^5)^6\texttt{;}$
  $\quad\texttt{in }(\texttt{let h = }(\texttt{fn z => z}^7)^8$
  $\quad\quad\texttt{in }((\texttt{f}^9\ \texttt{g}^{10})^{11}\texttt{ + }(\texttt{f}^{12}\ \texttt{h}^{13})^{14})^{15})^{16})^{17})^{18}$

- $(\texttt{let g = }(\texttt{fun f x => }(\texttt{f}^1\ (\texttt{fn y => y}^2)^3)^4)^5$
  $\texttt{in }(\texttt{g}^6\ (\texttt{fn z => z}^7)^8)^9)^{10}$

Abstract 0-CFA Analysis

- Abstract domains
- Specification of the analysis
- Well-definedness of the analysis

Towards defining the Abstract Domains

The *result* of a 0-CFA analysis is a pair $(\widehat{\mathsf{C}}, \widehat{\rho})$:

- $\widehat{\mathsf{C}}$ is the *abstract cache* associating abstract values with each labelled program point

- $\widehat{\rho}$ is the *abstract environment* associating abstract values with each variable

Example:

$$((\texttt{fn x => x}^1)^2\ (\texttt{fn y => y}^3)^4)^5$$

Three guesses of a 0-CFA analysis result:

	$(\widehat{C}_e, \widehat{\rho}_e)$	$(\widehat{C}'_e, \widehat{\rho}'_e)$	$(\widehat{C}''_e, \widehat{\rho}''_e)$
1	$\{\texttt{fn y => y}^3\}$	$\{\texttt{fn y => y}^3\}$	$\{\texttt{fn x => x}^1, \texttt{fn y => y}^3\}$
2	$\{\texttt{fn x => x}^1\}$	$\{\texttt{fn x => x}^1\}$	$\{\texttt{fn x => x}^1, \texttt{fn y => y}^3\}$
3	$\emptyset$	$\emptyset$	$\{\texttt{fn x => x}^1, \texttt{fn y => y}^3\}$
4	$\{\texttt{fn y => y}^3\}$	$\{\texttt{fn y => y}^3\}$	$\{\texttt{fn x => x}^1, \texttt{fn y => y}^3\}$
5	$\{\texttt{fn y => y}^3\}$	$\{\texttt{fn y => y}^3\}$	$\{\texttt{fn x => x}^1, \texttt{fn y => y}^3\}$
x	$\{\texttt{fn y => y}^3\}$	$\emptyset$	$\{\texttt{fn x => x}^1, \texttt{fn y => y}^3\}$
y	$\emptyset$	$\emptyset$	$\{\texttt{fn x => x}^1, \texttt{fn y => y}^3\}$

Example:

```
(let g = (fun f x => (f^1 (fn y => y^2)^3)^4)^5
 in (g^6 (fn z => z^7)^8)^9)^10
```

Abbreviations:

$$
\begin{array}{rcl}
\mathtt{f} & = & \texttt{fun f x => (f}^1\ \texttt{(fn y => y}^2\texttt{)}^3\texttt{)}^4 \\
\mathsf{id}_y & = & \texttt{fn y => y}^2 \\
\mathsf{id}_z & = & \texttt{fn z => z}^7
\end{array}
$$

One guess of a 0-CFA analysis result:

$$
\begin{array}{rclrclrcl}
\widehat{\mathsf{C}}_{\mathsf{lp}}(1) & = & \{\mathsf{f}\} & \widehat{\mathsf{C}}_{\mathsf{lp}}(6) & = & \{\mathsf{f}\} & \widehat{\rho}_{\mathsf{lp}}(\mathtt{f}) & = & \{\mathsf{f}\} \\
\widehat{\mathsf{C}}_{\mathsf{lp}}(2) & = & \emptyset & \widehat{\mathsf{C}}_{\mathsf{lp}}(7) & = & \emptyset & \widehat{\rho}_{\mathsf{lp}}(\mathtt{g}) & = & \{\mathsf{f}\} \\
\widehat{\mathsf{C}}_{\mathsf{lp}}(3) & = & \{\mathsf{id}_y\} & \widehat{\mathsf{C}}_{\mathsf{lp}}(8) & = & \{\mathsf{id}_z\} & \widehat{\rho}_{\mathsf{lp}}(\mathtt{x}) & = & \{\mathsf{id}_y, \mathsf{id}_z\} \\
\widehat{\mathsf{C}}_{\mathsf{lp}}(4) & = & \emptyset & \widehat{\mathsf{C}}_{\mathsf{lp}}(9) & = & \emptyset & \widehat{\rho}_{\mathsf{lp}}(\mathtt{y}) & = & \emptyset \\
\widehat{\mathsf{C}}_{\mathsf{lp}}(5) & = & \{\mathsf{f}\} & \widehat{\mathsf{C}}_{\mathsf{lp}}(10) & = & \emptyset & \widehat{\rho}_{\mathsf{lp}}(\mathtt{z}) & = & \emptyset
\end{array}
$$

Abstract Domains

Formally:

$$
\begin{array}{llllr}
\widehat{v} & \in & \widehat{\mathbf{Val}} & = & \mathcal{P}(\mathbf{Term}) & \textit{abstract values} \\
\widehat{\rho} & \in & \widehat{\mathbf{Env}} & = & \mathbf{Var} \to \widehat{\mathbf{Val}} & \textit{abstract environments} \\
\widehat{\mathsf{C}} & \in & \widehat{\mathbf{Cache}} & = & \mathbf{Lab} \to \widehat{\mathbf{Val}} & \textit{abstract caches}
\end{array}
$$

An abstract value $\widehat{v}$ is a set of terms of the forms

- **fn** x **=>** e_0
- **fun** f x **=>** e_0

Control Flow Analysis versus Use-Definition chains

The aim: to trace how definition points reach use points

- Control Flow Analysis
  - definition points: where function abstractions are created
  - use points: where functions are applied

- Use-Definition chains
  - definition points: where variables are assigned a value
  - use points: where values of variables are accessed

Specification of the 0-CFA

When is a proposed guess $(\widehat{\mathsf{C}}, \widehat{\rho})$ of an analysis results an *acceptable 0-CFA analysis* for the program?

Different approaches:

- abstract specification
- syntax-directed and constraint-based specifications
- algorithms for computing the *best* result

Specification of the Abstract 0-CFA

$(\widehat{\mathsf{C}}, \widehat{\rho}) \models e$ means that $(\widehat{\mathsf{C}}, \widehat{\rho})$ is an *acceptable Control Flow Analysis* of the expression e

The relation $\models$ has functionality:

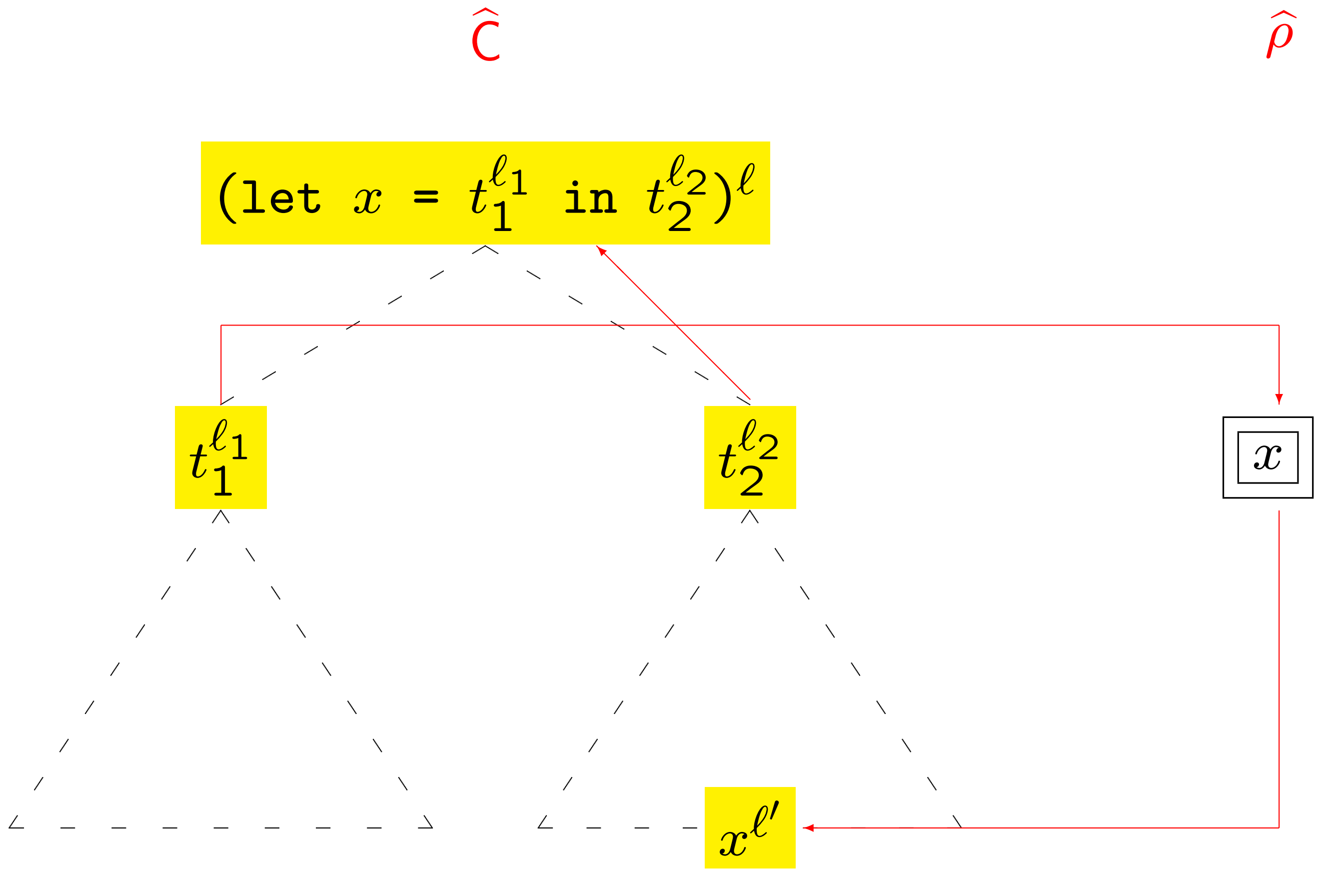
$$\models \; : (\widehat{\mathbf{Cache}} \times \widehat{\mathbf{Env}} \times \mathbf{Exp}) \to \{\mathit{true}, \mathit{false}\}$$

Clauses for Abstract 0-CFA (1)

$(\widehat{\mathsf{C}}, \widehat{\rho}) \models c^{\ell}$ always

$(\widehat{\mathsf{C}}, \widehat{\rho}) \models x^{\ell}$ $\quad \underline{\text{iff}} \quad$ $\widehat{\rho}(x) \subseteq \widehat{\mathsf{C}}(\ell)$

$$
\begin{array}{ll}
(\widehat{\mathsf{C}}, \widehat{\rho}) \models (\texttt{let } x \texttt{ = } t_1^{\ell_1} \texttt{ in } t_2^{\ell_2})^{\ell} & \\
\quad \underline{\text{iff}} & (\widehat{\mathsf{C}}, \widehat{\rho}) \models t_1^{\ell_1} \ \wedge \ (\widehat{\mathsf{C}}, \widehat{\rho}) \models t_2^{\ell_2} \ \wedge \\
& \widehat{\mathsf{C}}(\ell_1) \subseteq \widehat{\rho}(x) \ \wedge \ \widehat{\mathsf{C}}(\ell_2) \subseteq \widehat{\mathsf{C}}(\ell)
\end{array}
$$



$\widehat{\mathsf{C}}$
$\widehat{\rho}$
(let x = $t_1^{\ell_1}$ in $t_2^{\ell_2}$)$^\ell$
$t_1^{\ell_1}$
$t_2^{\ell_2}$
x
$x^{\ell'}$

Clauses for Abstract 0-CFA (2)

$$(\widehat{\mathsf{C}}, \widehat{\rho}) \models (\texttt{if } t_0^{\ell_0} \texttt{ then } t_1^{\ell_1} \texttt{ else } t_2^{\ell_2})^{\ell}$$

$$\underline{\text{iff}} \quad (\widehat{\mathsf{C}}, \widehat{\rho}) \models t_0^{\ell_0} \wedge$$

$$(\widehat{\mathsf{C}}, \widehat{\rho}) \models t_1^{\ell_1} \wedge (\widehat{\mathsf{C}}, \widehat{\rho}) \models t_2^{\ell_2} \wedge$$

$$\widehat{\mathsf{C}}(\ell_1) \subseteq \widehat{\mathsf{C}}(\ell) \wedge \widehat{\mathsf{C}}(\ell_2) \subseteq \widehat{\mathsf{C}}(\ell)$$

$$(\widehat{\mathsf{C}}, \widehat{\rho}) \models (t_1^{\ell_1} \; op \; t_2^{\ell_2})^{\ell}$$

$$\underline{\text{iff}} \quad (\widehat{\mathsf{C}}, \widehat{\rho}) \models t_1^{\ell_1} \wedge (\widehat{\mathsf{C}}, \widehat{\rho}) \models t_2^{\ell_2}$$

Clauses for Abstract 0-CFA (3)

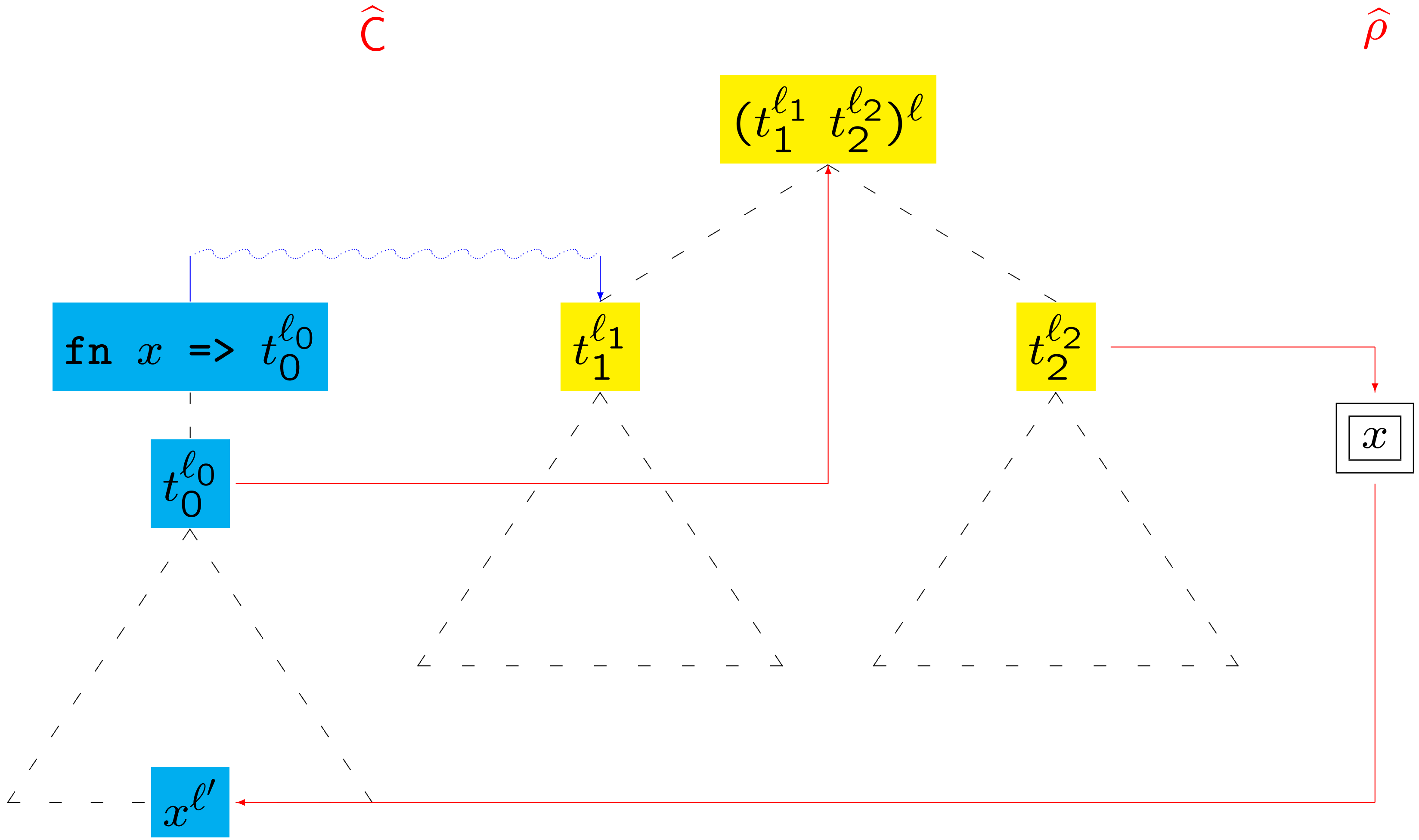
$(\widehat{\mathsf{C}}, \widehat{\rho}) \models (\mathtt{fn}\ x\ \mathtt{=>}\ t_0^{\ell_0})^\ell$ iff $\{\mathtt{fn}\ x\ \mathtt{=>}\ t_0^{\ell_0}\} \subseteq \widehat{\mathsf{C}}(\ell)$

$(\widehat{\mathsf{C}}, \widehat{\rho}) \models (t_1^{\ell_1}\ t_2^{\ell_2})^\ell$

<u>iff</u> $(\widehat{\mathsf{C}}, \widehat{\rho}) \models t_1^{\ell_1} \ \wedge\ (\widehat{\mathsf{C}}, \widehat{\rho}) \models t_2^{\ell_2}\ \wedge$

$(\forall (\mathtt{fn}\ x\ \mathtt{=>}\ t_0^{\ell_0}) \in \widehat{\mathsf{C}}(\ell_1) :\quad (\widehat{\mathsf{C}}, \widehat{\rho}) \models t_0^{\ell_0}\ \wedge$

$\widehat{\mathsf{C}}(\ell_2) \subseteq \widehat{\rho}(x)\ \wedge\ \widehat{\mathsf{C}}(\ell_0) \subseteq \widehat{\mathsf{C}}(\ell))$


$\widehat{\mathsf{C}}$
$\widehat{\rho}$
$(t_1^{\ell_1}\ t_2^{\ell_2})^{\ell}$
fn x => $t_0^{\ell_0}$
$t_1^{\ell_1}$
$t_2^{\ell_2}$
$t_0^{\ell_0}$
x
$x^{\ell'}$

Clauses for Abstract 0-CFA (4)

$(\widehat{C}, \widehat{\rho}) \models (\texttt{fun}\ f\ x\ \texttt{=>}\ e_0)^\ell$ iff $\{\texttt{fun}\ f\ x\ \texttt{=>}\ e_0\} \subseteq \widehat{C}(\ell)$

$$
\begin{array}{ll}
(\widehat{C}, \widehat{\rho}) \models (t_1^{\ell_1}\ t_2^{\ell_2})^\ell & \\
\underline{\text{iff}} & (\widehat{C}, \widehat{\rho}) \models t_1^{\ell_1} \ \wedge\ (\widehat{C}, \widehat{\rho}) \models t_2^{\ell_2} \ \wedge \\
& (\forall (\texttt{fn}\ x\ \texttt{=>}\ t_0^{\ell_0}) \in \widehat{C}(\ell_1) : \quad (\widehat{C}, \widehat{\rho}) \models t_0^{\ell_0} \ \wedge \\
& \qquad \widehat{C}(\ell_2) \subseteq \widehat{\rho}(x) \ \wedge\ \widehat{C}(\ell_0) \subseteq \widehat{C}(\ell)) \ \wedge \\
& (\forall (\texttt{fun}\ f\ x\ \texttt{=>}\ t_0^{\ell_0}) \in \widehat{C}(\ell_1) : \quad (\widehat{C}, \widehat{\rho}) \models t_0^{\ell_0} \ \wedge \\
& \qquad \widehat{C}(\ell_2) \subseteq \widehat{\rho}(x) \ \wedge\ \widehat{C}(\ell_0) \subseteq \widehat{C}(\ell) \ \wedge \\
& \qquad \{\texttt{fun}\ f\ x\ \texttt{=>}\ t_0^{\ell_0}\} \subseteq \widehat{\rho}(f))
\end{array}
$$

Example:

Two guesses for $((\texttt{fn x => x}^1)^2\ (\texttt{fn y => y}^3)^4)^5$

	$(\widehat{C}_e, \widehat{\rho}_e)$	$(\widehat{C}'_e, \widehat{\rho}'_e)$
1	$\{\texttt{fn y => y}^3\}$	$\{\texttt{fn y => y}^3\}$
2	$\{\texttt{fn x => x}^1\}$	$\{\texttt{fn x => x}^1\}$
3	$\emptyset$	$\emptyset$
4	$\{\texttt{fn y => y}^3\}$	$\{\texttt{fn y => y}^3\}$
5	$\{\texttt{fn y => y}^3\}$	$\{\texttt{fn y => y}^3\}$
x	$\{\texttt{fn y => y}^3\}$	$\emptyset$
y	$\emptyset$	$\emptyset$

Checking the guesses:

$$(\widehat{C}_e, \widehat{\rho}_e) \models ((\texttt{fn x => x}^1)^2\ (\texttt{fn y => y}^3)^4)^5$$

$$(\widehat{C}'_e, \widehat{\rho}'_e) \not\models ((\texttt{fn x => x}^1)^2\ (\texttt{fn y => y}^3)^4)^5$$

Well-definedness of the Abstract 0-CFA

Difficulty: The clause for function application is *not* of a form that allows us to define $(\widehat{\mathsf{C}}, \widehat{\rho}) \models e$ by Structural Induction in the expression e

$$
\begin{array}{ll}
(\widehat{\mathsf{C}}, \widehat{\rho}) \models (t_1^{\ell_1}\ t_2^{\ell_2})^{\ell} & \\
\quad \underline{\text{iff}} & (\widehat{\mathsf{C}}, \widehat{\rho}) \models t_1^{\ell_1} \ \wedge\ (\widehat{\mathsf{C}}, \widehat{\rho}) \models t_2^{\ell_2} \ \wedge \\
& (\forall (\texttt{fn}\ x\ \texttt{=>}\ t_0^{\ell_0}) \in \widehat{\mathsf{C}}(\ell_1) : \quad (\widehat{\mathsf{C}}, \widehat{\rho}) \models t_0^{\ell_0} \ \wedge \\
& \qquad \widehat{\mathsf{C}}(\ell_2) \subseteq \widehat{\rho}(x) \ \wedge\ \widehat{\mathsf{C}}(\ell_0) \subseteq \widehat{\mathsf{C}}(\ell))
\end{array}
$$

Solution: The relation $\models$ is defined by coinduction, that is, as the greatest fixed point of a functional.

The functional $\mathcal{Q}$

The clauses for $\models$ define a function:

$$\mathcal{Q} : ((\widehat{\mathbf{Cache}} \times \widehat{\mathbf{Env}} \times \mathbf{Exp}) \to \{\mathit{true}, \mathit{false}\}) \\ \to ((\widehat{\mathbf{Cache}} \times \widehat{\mathbf{Env}} \times \mathbf{Exp}) \to \{\mathit{true}, \mathit{false}\})$$

Example:

$$(\widehat{\mathsf{C}}, \widehat{\rho}) \models (\mathtt{let}\ x\ \mathtt{=}\ t_1^{\ell_1}\ \mathtt{in}\ t_2^{\ell_2})^{\ell}$$
$$\underline{\text{iff}} \quad (\widehat{\mathsf{C}}, \widehat{\rho}) \models t_1^{\ell_1} \ \wedge\ (\widehat{\mathsf{C}}, \widehat{\rho}) \models t_2^{\ell_2} \ \wedge\ \widehat{\mathsf{C}}(\ell_1) \subseteq \widehat{\rho}(x) \ \wedge\ \widehat{\mathsf{C}}(\ell_2) \subseteq \widehat{\mathsf{C}}(\ell)$$

becomes

$$\mathcal{Q}(\ Q\ )(\widehat{\mathsf{C}}, \widehat{\rho}, (\mathtt{let}\ x\ \mathtt{=}\ t_1^{\ell_1}\ \mathtt{in}\ t_2^{\ell_2})^{\ell})$$
$$= Q(\widehat{\mathsf{C}}, \widehat{\rho}, t_1^{\ell_1}) \ \wedge\ Q(\widehat{\mathsf{C}}, \widehat{\rho}, t_2^{\ell_2}) \ \wedge\ \widehat{\mathsf{C}}(\ell_1) \subseteq \widehat{\rho}(x) \ \wedge\ \widehat{\mathsf{C}}(\ell_2) \subseteq \widehat{\mathsf{C}}(\ell)$$

Properties of $\mathcal{Q}$

$\mathcal{Q}$ is a monotone function on the complete lattice

$$((\widehat{\mathbf{Cache}} \times \widehat{\mathbf{Env}} \times \mathbf{Exp}) \to \{\mathit{true}, \mathit{false}\}, \sqsubseteq)$$

where the ordering $\sqsubseteq$ is defined by:

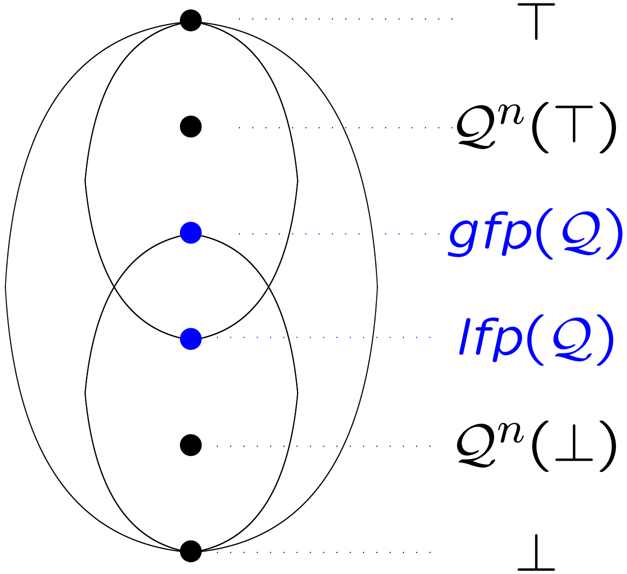
$$Q_1 \sqsubseteq Q_2 \quad \text{iff} \quad \forall(\widehat{\mathsf{C}}, \widehat{\rho}, e) : (Q_1(\widehat{\mathsf{C}}, \widehat{\rho}, e) = \mathit{true}) \;\Rightarrow\; (Q_2(\widehat{\mathsf{C}}, \widehat{\rho}, e) = \mathit{true})$$

Hence $\mathcal{Q}$ has fixed points and we shall define $\models$ coinductively:

$\models$ is the *greatest fixed point* of $\mathcal{Q}$

Tarski's Theorem:

A monotone function on a complete lattice has a complete lattice of fixed points and in particular a least and a greatest fixed point.



$$\mathcal{Q} : ((\widehat{\mathbf{Cache}} \times \widehat{\mathbf{Env}} \times \mathbf{Exp}) \to \{\mathit{true}, \mathit{false}\}) \to ((\widehat{\mathbf{Cache}} \times \widehat{\mathbf{Env}} \times \mathbf{Exp}) \to \{\mathit{true}, \mathit{false}\})$$

Coinductive definition:

$$\mathit{gfp}(\mathcal{Q}) = \bigsqcup\{P \mid \mathcal{Q}(P) \sqsupseteq P\}$$

Inductive definition:

$$\mathit{lfp}(\mathcal{Q}) = \bigsqcap\{P \mid \mathcal{Q}(P) \sqsubseteq P\} = \bigsqcup_n \mathcal{Q}^n(\bot)$$

assuming that $\mathcal{Q}(P)(\widehat{\mathsf{C}}, \widehat{\rho}, e)$ only depends on finitely many values of P

Inductive Definition

$$P = \mathit{lfp}(\mathcal{Q}) = \bigsqcup_n \mathcal{Q}^n(\bot) \quad \text{assuming } \cdots$$

P can be expressed as

$P(\widehat{\mathsf{C}}, \widehat{\rho}, x^\ell)$ iff $\widehat{\rho}(x) \subseteq \widehat{\mathsf{C}}(\ell)$

$P(\widehat{\mathsf{C}}, \widehat{\rho}, (\texttt{let } x \texttt{ = } t_1^{\ell_1} \texttt{ in } t_2^{\ell_2})^\ell)$ iff
$P(\widehat{\mathsf{C}}, \widehat{\rho}, t_1^{\ell_1}) \wedge P(\widehat{\mathsf{C}}, \widehat{\rho}, t_2^{\ell_2})$
$\widehat{\mathsf{C}}(\ell_1) \subseteq \widehat{\rho}(x) \wedge \widehat{\mathsf{C}}(\ell_2) \subseteq \widehat{\mathsf{C}}(\ell)$

$\vdots$

simply because $P = \mathcal{Q}(P)$

Example:

0 is a number

$n+1$ is a number iff n is a number

(Peano's Axioms)

to check $P(\widehat{\mathsf{C}}, \widehat{\rho}, e)$

simply unfold using the clauses:

- if it terminates
  - and yields true: then it holds
  - and yields false: then it does not
- if it loops
  - because it repeats itself:
    - then it does not hold
  - but we cannot detect it $\cdots$

Example:

$2 = 0+1+1$ is a number

because $0+1$ is because 0 is

Inductive Definition

to prove: $\forall(\widehat{\mathsf{C}}, \widehat{\rho}, e) : P(\widehat{\mathsf{C}}, \widehat{\rho}, e) \Rightarrow R(\widehat{\mathsf{C}}, \widehat{\rho}, e)$

show: $R(\widehat{\mathsf{C}}, \widehat{\rho}, x^\ell)$ if $\widehat{\rho}(x) \subseteq \widehat{\mathsf{C}}(\ell)$ axiom

$$\frac{R(\widehat{\mathsf{C}}, \widehat{\rho}, t_1^{\ell_1}) \quad R(\widehat{\mathsf{C}}, \widehat{\rho}, t_2^{\ell_2})}{R(\widehat{\mathsf{C}}, \widehat{\rho}, (\texttt{let } x \texttt{ = } t_1^{\ell_1} \texttt{ in } t_2^{\ell_2})^\ell)}$$ inference rule

if $\widehat{\mathsf{C}}(\ell_1) \subseteq \widehat{\rho}(x) \wedge \widehat{\mathsf{C}}(\ell_2) \subseteq \widehat{\mathsf{C}}(\ell)$

$\vdots$

Examples:

- mathematical induction: $R(0)$, $\dfrac{R(n)}{R(n+1)}$
- structural induction
- induction on the shape of inference tree

Coinductive Definition

$$P = gfp(\mathcal{Q}) = \bigsqcup\{R \mid R \sqsubseteq \mathcal{Q}(R)\}$$

P can be expressed as

$P(\widehat{\mathsf{C}}, \widehat{\rho}, x^\ell)$ <u>iff</u> $\widehat{\rho}(x) \subseteq \widehat{\mathsf{C}}(\ell)$

$P(\widehat{\mathsf{C}}, \widehat{\rho}, (\texttt{let } x \texttt{ = } t_1^{\ell_1} \texttt{ in } t_2^{\ell_2})^\ell)$ <u>iff</u>
$P(\widehat{\mathsf{C}}, \widehat{\rho}, t_1^{\ell_1}) \wedge P(\widehat{\mathsf{C}}, \widehat{\rho}, t_2^{\ell_2})$
$\widehat{\mathsf{C}}(\ell_1) \subseteq \widehat{\rho}(x) \wedge \widehat{\mathsf{C}}(\ell_2) \subseteq \widehat{\mathsf{C}}(\ell)$

⋮

simply because $P = \mathcal{Q}(P)$

to check $P(\widehat{\mathsf{C}}, \widehat{\rho}, e)$

find some R such that
$R(\widehat{\mathsf{C}}, \widehat{\rho}, e)$ can be shown to hold

that is prove:

$R(\widehat{\mathsf{C}}, \widehat{\rho}, x^\ell)$ if $\widehat{\rho}(x) \subseteq \widehat{\mathsf{C}}(\ell)$

$$\frac{R(\widehat{\mathsf{C}}, \widehat{\rho}, t_1^{\ell_1}) \quad R(\widehat{\mathsf{C}}, \widehat{\rho}, t_2^{\ell_2})}{R(\widehat{\mathsf{C}}, \widehat{\rho}, (\texttt{let } x \texttt{ = } t_1^{\ell_1} \texttt{ in } t_2^{\ell_2})^\ell)}$$

if $\widehat{\mathsf{C}}(\ell_1) \subseteq \widehat{\rho}(x) \wedge \widehat{\mathsf{C}}(\ell_2) \subseteq \widehat{\mathsf{C}}(\ell)$

⋮

and use $P = \bigsqcup\{R \mid R \sqsubseteq \mathcal{Q}(R)\}$

Coinductive Definition

to prove: $\forall(\widehat{\mathsf{C}}, \widehat{\rho}, e) : P(\widehat{\mathsf{C}}, \widehat{\rho}, e) \Rightarrow R(\widehat{\mathsf{C}}, \widehat{\rho}, e)$

- try to prove it using $P = \mathcal{Q}(P)$
  i.e. by using the way P is expressed

- if it fails try to do induction (on the structure or size) of e

- if it fails $\cdots$ you will need an extra insight

Example: loop

```
(let g = (fun f x => (f¹ (fn y => y²)³)⁴)⁵
 in (g⁶ (fn z => z⁷)⁸)⁹)¹⁰
```

$(\texttt{let g = } (\texttt{fun f x => } (\texttt{f}^1\ (\texttt{fn y => y}^2)^3)^4)^5$
$\texttt{ in } (\texttt{g}^6\ (\texttt{fn z => z}^7)^8)^9)^{10}$

Abbreviations:

$$\begin{aligned} \mathsf{f} &= \texttt{fun f x => } (\texttt{f}^1\ (\texttt{fn y => y}^2)^3)^4 \\ \mathsf{id}_y &= \texttt{fn y => y}^2 \\ \mathsf{id}_z &= \texttt{fn z => z}^7 \end{aligned}$$

One guess of a 0-CFA analysis result:

$$\begin{aligned} \widehat{\mathsf{C}}_{\mathsf{lp}}(1) &= \{\mathsf{f}\} & \widehat{\mathsf{C}}_{\mathsf{lp}}(6) &= \{\mathsf{f}\} & \widehat{\rho}_{\mathsf{lp}}(\mathtt{f}) &= \{\mathsf{f}\} \\ \widehat{\mathsf{C}}_{\mathsf{lp}}(2) &= \emptyset & \widehat{\mathsf{C}}_{\mathsf{lp}}(7) &= \emptyset & \widehat{\rho}_{\mathsf{lp}}(\mathtt{g}) &= \{\mathsf{f}\} \\ \widehat{\mathsf{C}}_{\mathsf{lp}}(3) &= \{\mathsf{id}_y\} & \widehat{\mathsf{C}}_{\mathsf{lp}}(8) &= \{\mathsf{id}_z\} & \widehat{\rho}_{\mathsf{lp}}(\mathtt{x}) &= \{\mathsf{id}_y, \mathsf{id}_z\} \\ \widehat{\mathsf{C}}_{\mathsf{lp}}(4) &= \emptyset & \widehat{\mathsf{C}}_{\mathsf{lp}}(9) &= \emptyset & \widehat{\rho}_{\mathsf{lp}}(\mathtt{y}) &= \emptyset \\ \widehat{\mathsf{C}}_{\mathsf{lp}}(5) &= \{\mathsf{f}\} & \widehat{\mathsf{C}}_{\mathsf{lp}}(10) &= \emptyset & \widehat{\rho}_{\mathsf{lp}}(\mathtt{z}) &= \emptyset \end{aligned}$$

Naively checking the solution gives rise to circularity:

To show

$$(\widehat{\mathsf{C}}_{\mathsf{lp}}, \widehat{\rho}_{\mathsf{lp}}) \models \mathsf{loop}$$

we have (among others) to show

$$(\widehat{\mathsf{C}}_{\mathsf{lp}}, \widehat{\rho}_{\mathsf{lp}}) \models (\mathtt{g}^6\ (\mathtt{fn\ z\ =>\ z}^7)^8)^9$$

and to prove this we have (among others) to show

$$(\widehat{\mathsf{C}}_{\mathsf{lp}}, \widehat{\rho}_{\mathsf{lp}}) \models (\mathtt{f}^1\ (\mathtt{fn\ y\ =>\ y}^2)^3)^4$$

and to show this we have (among others) to show

$$(\widehat{\mathsf{C}}_{\mathsf{lp}}, \widehat{\rho}_{\mathsf{lp}}) \models (\mathtt{f}^1\ (\mathtt{fn\ y\ =>\ y}^2)^3)^4$$

because $\widehat{\mathsf{C}}_{\mathsf{lp}}(3) \subseteq \widehat{\rho}_{\mathsf{lp}}(\mathtt{x})$, $\widehat{\mathsf{C}}_{\mathsf{lp}}(4) \subseteq \widehat{\mathsf{C}}_{\mathsf{lp}}(4)$ and $\mathsf{f} \in \widehat{\rho}_{\mathsf{lp}}(\mathtt{f})$.

The Lesson

The co-inductive definition solves the circularity:

> It allows us to assume that $(\widehat{\mathsf{C}}_{\mathsf{lp}}, \widehat{\rho}_{\mathsf{lp}}) \models (\mathtt{f}^1\ (\mathtt{fn\ y\ =>\ y}^2)^3)^4$ holds at the "inner level" and proving that it also holds at the "outer level"

An inductive definition does not give us this possibility!

Theoretical Properties:

- structural operational semantics
- semantic correctness
- the existence of least solutions

Choice of Semantics

- operational or denotational semantics?
  - an operational semantics more easily models intensional properties

- small-step or big-step operational semantics?
  - a small-step semantics allows us to reason about looping programs

- operational semantics based on environments or substitutions?
  - an environment based semantics preserves the identity of functions

Configurations and Transitions

Semantic categories:

$$v \in \mathbf{Val} \quad \textit{values}$$

$$\rho \in \mathbf{Env} \quad \textit{environments}$$

defined by:

$$v ::= c \mid \texttt{close}\ t\ \texttt{in}\ \rho \qquad \textit{closures}$$

$$\rho ::= [\ ] \mid \rho[x \mapsto v]$$

Transitions have the form

$$\rho \vdash e_1 \longrightarrow e_2$$

meaning that *one step* of computation of the expression e_1 in the environment ρ will transform it into e_2.

Transitions

$$\rho \vdash x^\ell \rightarrow v^\ell \text{ if } x \in \mathit{dom}(\rho) \text{ and } v = \rho(x)$$

$$\rho \vdash (\texttt{fn}\ x\ \texttt{=>}\ e_0)^\ell \rightarrow (\texttt{close}\ (\texttt{fn}\ x\ \texttt{=>}\ e_0)\ \texttt{in}\ \rho_0)^\ell$$

$$\text{where } \rho_0 = \rho \mid \mathit{FV}(\texttt{fn}\ x\ \texttt{=>}\ e_0)$$

$$\rho \vdash (\texttt{fun}\ f\ x\ \texttt{=>}\ e_0)^\ell \rightarrow (\texttt{close}\ (\texttt{fun}\ f\ x\ \texttt{=>}\ e_0)\ \texttt{in}\ \rho_0)^\ell$$

$$\text{where } \rho_0 = \rho \mid \mathit{FV}(\texttt{fun}\ f\ x\ \texttt{=>}\ e_0)$$

static scope!

Intermediate Expressions and Terms

$$ie \in \mathbf{IExp} \quad \textit{intermediate expressions}$$

$$it \in \mathbf{ITerm} \quad \textit{intermediate terms}$$

extending the syntax:

$$ie ::= it^{\ell}$$

$$\begin{array}{rcl} it & ::= & c \mid x \mid \texttt{fn}\ x\ \texttt{=>}\ e_0 \mid \texttt{fun}\ f\ x\ \texttt{=>}\ e_0 \mid ie_1\ ie_2 \\ & \mid & \texttt{if}\ ie_0\ \texttt{then}\ e_1\ \texttt{else}\ e_2 \mid \texttt{let}\ x\ \texttt{=}\ ie_1\ \texttt{in}\ e_2 \mid ie_1\ op\ ie_2 \\ & \mid & \texttt{close}\ t\ \texttt{in}\ \rho \mid \texttt{bind}\ \rho\ \texttt{in}\ ie \end{array}$$

The correct form of transitions

$$\rho \vdash ie_1 \rightarrow ie_2$$

Transitions

$$\frac{\rho \vdash ie_1 \longrightarrow ie_1'}{\rho \vdash (ie_1\ ie_2)^\ell \longrightarrow (ie_1'\ ie_2)^\ell} \qquad \frac{\rho \vdash ie_2 \longrightarrow ie_2'}{\rho \vdash (v_1^{\ell_1}\ ie_2)^\ell \longrightarrow (v_1^{\ell_1}\ ie_2')^\ell}$$

$$\rho \vdash ((\texttt{close}\ (\texttt{fn}\ x\ \texttt{=>}\ e_1)\ \texttt{in}\ \rho_1)^{\ell_1}\ v_2^{\ell_2})^\ell \longrightarrow (\texttt{bind}\ \rho_1[x \mapsto v_2]\ \texttt{in}\ e_1)^\ell$$

$$\rho \vdash ((\texttt{close}\ (\texttt{fun}\ f\ x\ \texttt{=>}\ e_1)\ \texttt{in}\ \rho_1)^{\ell_1}\ v_2^{\ell_2})^\ell \longrightarrow (\texttt{bind}\ \rho_2[x \mapsto v_2]\ \texttt{in}\ e_1)^\ell$$

where $\rho_2 = \rho_1[f \mapsto \texttt{close}\ (\texttt{fun}\ f\ x\ \texttt{=>}\ e_1)\ \texttt{in}\ \rho_1]$

$$\frac{\rho_1 \vdash ie_1 \longrightarrow ie_1'}{\rho \vdash (\texttt{bind}\ \rho_1\ \texttt{in}\ ie_1)^\ell \longrightarrow (\texttt{bind}\ \rho_1\ \texttt{in}\ ie_1')^\ell}$$

$$\rho \vdash (\texttt{bind}\ \rho_1\ \texttt{in}\ v_1^{\ell_1})^\ell \longrightarrow v_1^{\ell}$$

the outermost label remains the same

Example:

$$[\,] \;\vdash\; ((\texttt{fn x => x}^1)^2\ (\texttt{fn y => y}^3)^4)^5$$

$$\longrightarrow\quad ((\texttt{close (fn x => x}^1\texttt{) in [ ]})^2\ (\texttt{fn y => y}^3)^4)^5$$

$$\longrightarrow\quad ((\texttt{close (fn x => x}^1\texttt{) in [ ]})^2\ (\texttt{close (fn y => y}^3\texttt{) in [ ]})^4)^5$$

$$\longrightarrow\quad (\texttt{bind [x} \mapsto \texttt{(close (fn y => y}^3\texttt{) in [ ])] in x}^1)^5$$

$$\longrightarrow\quad (\texttt{bind [x} \mapsto \texttt{(close (fn y => y}^3\texttt{) in [ ])] in}$$
$$(\texttt{close (fn y => y}^3\texttt{) in [ ]})^1)^5$$

$$\longrightarrow\quad (\texttt{close (fn y => y}^3\texttt{) in [ ]})^5$$

Transitions

$$\frac{\rho \vdash ie_0 \longrightarrow ie'_0}{\rho \vdash (\texttt{if } ie_0 \texttt{ then } e_1 \texttt{ else } e_2)^\ell \longrightarrow (\texttt{if } ie'_0 \texttt{ then } e_1 \texttt{ else } e_2)^\ell}$$

$$\rho \vdash (\texttt{if true}^{\ell_0} \texttt{ then } t_1^{\ell_1} \texttt{ else } t_2^{\ell_2})^\ell \longrightarrow t_1^\ell$$

$$\rho \vdash (\texttt{if false}^{\ell_0} \texttt{ then } t_1^{\ell_1} \texttt{ else } t_2^{\ell_2})^\ell \longrightarrow t_2^\ell$$

$$\frac{\rho \vdash ie_1 \longrightarrow ie'_1}{\rho \vdash (\texttt{let } x \texttt{ = } ie_1 \texttt{ in } e_2)^\ell \longrightarrow (\texttt{let } x \texttt{ = } ie'_1 \texttt{ in } e_2)^\ell}$$

$$\rho \vdash (\texttt{let } x \texttt{ = } v^{\ell_1} \texttt{ in } e_2)^\ell \longrightarrow (\texttt{bind } [x \mapsto v] \texttt{ in } e_2)^\ell$$

$$\frac{\rho \vdash ie_1 \longrightarrow ie'_1}{\rho \vdash (ie_1 \; op \; ie_2)^\ell \longrightarrow (ie'_1 \; op \; ie_2)^\ell} \qquad \frac{\rho \vdash ie_2 \longrightarrow ie'_2}{\rho \vdash (v_1^{\ell_1} \; op \; ie_2)^\ell \longrightarrow (v_1^{\ell_1} \; op \; ie'_2)^\ell}$$

$$\rho \vdash (v_1^{\ell_1} \; op \; v_2^{\ell_2})^\ell \longrightarrow v^\ell \qquad \text{if } v = v_1 \; \textbf{op} \; v_2$$

Example:

$$
\begin{array}{ll}
[\,] \vdash & (\texttt{let g} = (\texttt{fun f x => } (\texttt{f}^1\ (\texttt{fn y => y}^2)^3)^4)^5 \\
& \quad \texttt{in } (\texttt{g}^6\ (\texttt{fn z => z}^7)^8)^9)^{10} \\
\longrightarrow & (\texttt{let g} = \mathbf{f}^5 \texttt{ in } (\texttt{g}^6\ (\texttt{fn z => z}^7)^8)^9)^{10} \\
\longrightarrow & (\texttt{bind } [\texttt{g} \mapsto \mathbf{f}] \texttt{ in } (\texttt{g}^6\ (\texttt{fn z => z}^7)^8)^9)^{10} \\
\longrightarrow & (\texttt{bind } [\texttt{g} \mapsto \mathbf{f}] \texttt{ in } (\mathbf{f}^6\ (\texttt{fn z => z}^7)^8)^9)^{10} \\
\longrightarrow & (\texttt{bind } [\texttt{g} \mapsto \mathbf{f}] \texttt{ in } (\mathbf{f}^6\ \mathbf{id}_z^8)^9)^{10} \\
\longrightarrow & (\texttt{bind } [\texttt{g} \mapsto \mathbf{f}] \texttt{ in } (\texttt{bind } [\texttt{f} \mapsto \mathbf{f}][\texttt{x} \mapsto \mathbf{id}_z] \texttt{ in } (\texttt{f}^1\ (\texttt{fn y => y}^2)^3)^4)^9)^{10} \\
\longrightarrow^* & (\texttt{bind } [\texttt{g} \mapsto \mathbf{f}] \texttt{ in } (\texttt{bind } [\texttt{f} \mapsto \mathbf{f}][\texttt{x} \mapsto \mathbf{id}_z] \texttt{ in} \\
& \quad (\texttt{bind } [\texttt{f} \mapsto \mathbf{f}][\texttt{x} \mapsto \mathbf{id}_y] \texttt{ in } (\texttt{f}^1\ (\texttt{fn y => y}^2)^3)^4)^4)^9)^{10} \\
\longrightarrow^* & \ldots
\end{array}
$$

Abbreviations:

$$
\begin{array}{rcl}
\mathbf{f} & = & \texttt{close } (\texttt{fun f x => } (\texttt{f}^1\ (\texttt{fn y => y}^2)^3)^4) \texttt{ in } [\,] \\
\mathbf{id}_y & = & \texttt{close } (\texttt{fn y => y}^2) \texttt{ in } [\,] \\
\mathbf{id}_z & = & \texttt{close } (\texttt{fn z => z}^7) \texttt{ in } [\,]
\end{array}
$$

Semantic Correctness

A *subject reduction result:* an acceptable result of the analysis remains acceptable under evaluation

Analysis of intermediate expressions

$$(\widehat{\mathsf{C}}, \widehat{\rho}) \models (\texttt{bind}\ \rho\ \texttt{in}\ it_0^{\ell_0})^{\ell}$$
$$\underline{\text{iff}} \quad (\widehat{\mathsf{C}}, \widehat{\rho}) \models it_0^{\ell_0} \ \wedge\ \widehat{\mathsf{C}}(\ell_0) \subseteq \widehat{\mathsf{C}}(\ell) \ \wedge\ \rho\ \mathcal{R}\ \widehat{\rho}$$

$$(\widehat{\mathsf{C}}, \widehat{\rho}) \models (\texttt{close}\ t_0\ \texttt{in}\ \rho)^{\ell}$$
$$\underline{\text{iff}} \quad \{t_0\} \subseteq \widehat{\mathsf{C}}(\ell) \ \wedge\ \rho\ \mathcal{R}\ \widehat{\rho}$$

Correctness Relation

The global abstract environment, $\widehat{\rho}$ models *all* the local environments of the semantics

Correctness relation

$$\mathcal{R} : (\mathbf{Env} \times \widehat{\mathbf{Env}}) \rightarrow \{\mathit{true}, \mathit{false}\}$$

We demand that $\rho \; \mathcal{R} \; \widehat{\rho}$ for all local environments, ρ, occurring in the intermediate expressions

Define

$$\rho \; \mathcal{R} \; \widehat{\rho} \quad \underline{\text{iff}} \quad \forall x \in \mathit{dom}(\rho) \subseteq \mathit{dom}(\widehat{\rho}) \; \forall t_x \; \forall \rho_x : \\ (\rho(x) = \texttt{close } t_x \texttt{ in } \rho_x) \; \Rightarrow \; (t_x \in \widehat{\rho}(x) \land \rho_x \; \mathcal{R} \; \widehat{\rho})$$

(Well-defined by induction in the size of ρ.)

Example:

Suppose that:

$$\begin{array}{rcl} \rho & = & [x \mapsto \texttt{close}\ t_1\ \texttt{in}\ \rho_1][y \mapsto \texttt{close}\ t_2\ \texttt{in}\ \rho_2] \\ \rho_1 & = & [\,] \\ \rho_2 & = & [x \mapsto \texttt{close}\ t_3\ \texttt{in}\ \rho_3] \\ \rho_3 & = & [\,] \end{array}$$

Then $\rho\ \mathcal{R}\ \widehat{\rho}$ amounts to $\{t_1, t_3\} \subseteq \widehat{\rho}(x) \wedge \{t_2\} \subseteq \widehat{\rho}(y)$.

Alternative definition of Correctness Relation

Split the definition of $\mathcal{R}$ into two components:

$$\begin{aligned} \mathcal{V} : \quad & (\mathbf{Val} \times (\widehat{\mathbf{Env}} \times \widehat{\mathbf{Val}})) \to \{\mathit{true}, \mathit{false}\} \\ \mathcal{R} : \quad & (\mathbf{Env} \times \widehat{\mathbf{Env}}) \to \{\mathit{true}, \mathit{false}\} \end{aligned}$$

and define

$$\begin{array}{lll} v \; \mathcal{V} \; (\widehat{\rho}, \widehat{v}) & \underline{\text{iff}} & \forall t \; \forall \rho : (v = \texttt{close } t \texttt{ in } \rho) \;\Rightarrow\; (t \in \widehat{v} \;\wedge\; \rho \; \mathcal{R} \; \widehat{\rho}) \\ \rho \; \mathcal{R} \; \widehat{\rho} & \underline{\text{iff}} & \forall x \in \mathit{dom}(\rho) \subseteq \mathit{dom}(\widehat{\rho}) : \rho(x) \; \mathcal{V} \; (\widehat{\rho}, \widehat{\rho}(x)) \end{array}$$

Correctness Result

$$
\begin{array}{ccccccccccc}
\rho & \vdash & ie & \longrightarrow & ie' & \longrightarrow & ie'' & \longrightarrow & \cdots & \longrightarrow & v^{\ell} \\
\updownarrow \; \mathcal{R} & & \updownarrow \; \models & & \updownarrow \; \models & & \updownarrow \; \models & & & & \updownarrow \; \models \\
\widehat{\rho} & & (\widehat{\mathsf{C}}, \widehat{\rho}) & & (\widehat{\mathsf{C}}, \widehat{\rho}) & & (\widehat{\mathsf{C}}, \widehat{\rho}) & & \cdots & & (\widehat{\mathsf{C}}, \widehat{\rho})
\end{array}
$$

Formal details of Correctness Result

Theorem:

If $\rho \; \mathcal{R} \; \widehat{\rho}$ and $\rho \vdash ie \rightarrow ie'$ then $(\widehat{\mathsf{C}}, \widehat{\rho}) \models ie$ implies $(\widehat{\mathsf{C}}, \widehat{\rho}) \models ie'$.

Intuitively:

If there is a possible evaluation of the program such that the function at a call point evaluates to some abstraction, then this abstraction has to be in the set of possible abstractions computed by the analysis.

Observe: the theorem expresses that *all* acceptable analysis results remain acceptable under evaluation!

Thus we do *not* rely on the existence of a least or “best” solution.

Proof of Correctness Result

We assume that $\rho \; \mathcal{R} \; \widehat{\rho}$ and $(\widehat{\mathsf{C}}, \widehat{\rho}) \models ie$ and prove $(\widehat{\mathsf{C}}, \widehat{\rho}) \models ie'$ by induction on the structure of the inference tree for $\rho \vdash ie \rightarrow ie'$.

Most cases amount to inspecting the defining clause for $(\widehat{\mathsf{C}}, \widehat{\rho}) \models ie$.

This method of proof applies to *all* fixed points of a recursive definition and in particular also to the (more familiar least and) greatest fixed point(s).

Crucial fact: If $(\widehat{\mathsf{C}}, \widehat{\rho}) \models it^{\ell_1}$ and $\widehat{\mathsf{C}}(\ell_1) \subseteq \widehat{\mathsf{C}}(\ell_2)$ then $(\widehat{\mathsf{C}}, \widehat{\rho}) \models it^{\ell_2}$.

Example:

Semantics:

$$[\,] \vdash ((\texttt{fn x => x}^1)^2\ (\texttt{fn y => y}^3)^4)^5 \longrightarrow^* (\texttt{close}\ (\texttt{fn y => y}^3)\ \texttt{in}\ [\,])^5$$

	$(\widehat{C}_e, \widehat{\rho}_e)$
1	$\{\texttt{fn y => y}^3\}$
2	$\{\texttt{fn x => x}^1\}$
3	$\emptyset$
4	$\{\texttt{fn y => y}^3\}$
5	$\{\texttt{fn y => y}^3\}$
x	$\{\texttt{fn y => y}^3\}$
y	$\emptyset$

Analysis:

$$(\widehat{C}_e, \widehat{\rho}_e) \models ((\texttt{fn x => x}^1)^2\ (\texttt{fn y => y}^3)^4)^5$$

Correctness relation:

$$[\,]\ \mathcal{R}\ \widehat{\rho}_e$$

Correctness theorem: $(\widehat{C}_e, \widehat{\rho}_e) \models (\texttt{close}\ (\texttt{fn y => y}^3)\ \texttt{in}\ [\,])^5$

Existence of Solutions

- Does each expression e admit a Control Flow Analysis?

  i.e. does there exist $(\widehat{\mathsf{C}}, \widehat{\rho})$ such that $(\widehat{\mathsf{C}}, \widehat{\rho}) \models e$?

- Does each expression e have a "least" Control Flow Analysis?

  i.e. does there exists $(\widehat{\mathsf{C}}_0, \widehat{\rho}_0)$ such that $(\widehat{\mathsf{C}}_0, \widehat{\rho}_0) \models e$ and such that whenever $(\widehat{\mathsf{C}}, \widehat{\rho}) \models e$ then $(\widehat{\mathsf{C}}_0, \widehat{\rho}_0)$ is "less than" $(\widehat{\mathsf{C}}, \widehat{\rho})$?

Here "least" is with respect to the partial ordering

$$(\widehat{\mathsf{C}}_1, \widehat{\rho}_1) \sqsubseteq (\widehat{\mathsf{C}}_2, \widehat{\rho}_2) \quad \underline{\text{iff}} \quad \begin{array}{l}(\forall \ell \in \mathbf{Lab} : \widehat{\mathsf{C}}_1(\ell) \subseteq \widehat{\mathsf{C}}_2(\ell)) \ \wedge \\ (\forall x \in \mathbf{Var} : \widehat{\rho}_1(x) \subseteq \widehat{\rho}_2(x))\end{array}$$

Existence of Solutions (cont.)

Two answers:

- there exists algorithms for the efficient computation of least solutions for all expressions

- all intermediate expressions enjoy a Moore family property

A subset Y of a complete lattice $L = (L, \sqsubseteq)$ is a *Moore family* if and only if $(\bigsqcap Y') \in Y$ for all subsets Y' of L

Proposition: The set $\{(\widehat{\mathsf{C}}, \widehat{\rho}) \mid (\widehat{\mathsf{C}}, \widehat{\rho}) \models ie\}$ is a Moore family for all intermediate expressions ie

Existence of Solutions (cont.)

All intermediate expressions admit a Control Flow Analysis

Let Y' be the empty set; then $\bigsqcap Y'$ is an element of $\{(\widehat{\mathsf{C}}, \widehat{\rho}) \mid (\widehat{\mathsf{C}}, \widehat{\rho}) \models ie\}$ showing that there exists at least one analysis of ie.

All intermediate expressions have a least Control Flow Analysis

Let Y' be the set $\{(\widehat{\mathsf{C}}, \widehat{\rho}) \mid (\widehat{\mathsf{C}}, \widehat{\rho}) \models ie\}$; then $\bigsqcap Y'$ is an element of $\{(\widehat{\mathsf{C}}, \widehat{\rho}) \mid (\widehat{\mathsf{C}}, \widehat{\rho}) \models ie\}$ so it will also be an analysis of ie. Clearly $\bigsqcap Y' \sqsubseteq (\widehat{\mathsf{C}}, \widehat{\rho})$ for all other analyses $(\widehat{\mathsf{C}}, \widehat{\rho})$ of ie so it is the least analysis result.

Example:

$$(\widehat{\mathsf{C}}_{\mathsf{e}}', \widehat{\rho}_{\mathsf{e}}') \models ((\texttt{fn x => x}^1)^2\ (\texttt{fn y => y}^3)^4)^5$$

$$(\widehat{\mathsf{C}}_{\mathsf{e}}'', \widehat{\rho}_{\mathsf{e}}'') \models ((\texttt{fn x => x}^1)^2\ (\texttt{fn y => y}^3)^4)^5$$

The Moore family result ensures that

$$(\widehat{\mathsf{C}}_{\mathsf{e}}' \sqcap \widehat{\mathsf{C}}_{\mathsf{e}}'', \widehat{\rho}_{\mathsf{e}}' \sqcap \widehat{\rho}_{\mathsf{e}}'') \models ((\texttt{fn x => x}^1)^2\ (\texttt{fn y => y}^3)^4)^5$$

	$(\widehat{\mathsf{C}}_{\mathsf{e}}, \widehat{\rho}_{\mathsf{e}})$	$(\widehat{\mathsf{C}}_{\mathsf{e}}', \widehat{\rho}_{\mathsf{e}}')$	$(\widehat{\mathsf{C}}_{\mathsf{e}}'', \widehat{\rho}_{\mathsf{e}}'')$
1	$\{\texttt{fn y => y}^3\}$	$\{\texttt{fn y => y}^3\}$	$\{\texttt{fn y => y}^3\}$
2	$\{\texttt{fn x => x}^1\}$	$\{\texttt{fn x => x}^1\}$	$\{\texttt{fn x => x}^1\}$
3	$\emptyset$	$\{\texttt{fn x => x}^1\}$	$\{\texttt{fn y => y}^3\}$
4	$\{\texttt{fn y => y}^3\}$	$\{\texttt{fn y => y}^3\}$	$\{\texttt{fn y => y}^3\}$
5	$\{\texttt{fn y => y}^3\}$	$\{\texttt{fn y => y}^3\}$	$\{\texttt{fn y => y}^3\}$
x	$\{\texttt{fn y => y}^3\}$	$\{\texttt{fn y => y}^3\}$	$\{\texttt{fn y => y}^3\}$
y	$\emptyset$	$\{\texttt{fn x => x}^1\}$	$\{\texttt{fn y => y}^3\}$

Coinduction versus Induction

The abstract Control Flow Analysis is defined *coinductively*

$\models$ is the *greatest* fixed point of a function $\mathcal{Q}$

An alternative might be an *inductive* definition

$\models'$ is the *least* fixed point of the function $\mathcal{Q}$.

Proposition: There exists $e_\star \in \mathbf{Exp}$ such that $\{(\widehat{\mathsf{C}}, \widehat{\rho}) \mid (\widehat{\mathsf{C}}, \widehat{\rho}) \models' e_\star\}$ is *not* a Moore family.

Syntax Directed 0-CFA Analysis

Reformulate the abstract specification:

(i) Syntax directed specification

(ii) Constructing a finite set of constraints

(iii) Compute the least solution of the set of constraints

Common Phenomenon

A specification $\models_A$ is reformulated into a specification $\models_B$ ensuring that

$$(\widehat{\mathsf{C}}, \widehat{\rho}) \models_A e_\star \Leftarrow (\widehat{\mathsf{C}}, \widehat{\rho}) \models_B e_\star$$

so that "$\models_B$" is a *safe approximation* to "$\models_A$" and hence the best (i.e. least) solution to "$\models_B e_\star$" will also be a solution to "$\models_A e_\star$".

If additionally

$$(\widehat{\mathsf{C}}, \widehat{\rho}) \models_A e_\star \Rightarrow (\widehat{\mathsf{C}}, \widehat{\rho}) \models_B e_\star$$

then we can be assured that *no solutions are lost* and hence the best (i.e. least) solution to "$\models_B e_\star$" will also be the best (i.e. least) solution to "$\models_A e_\star$".

Syntax Directed Specification (1)

$$(\widehat{\mathsf{C}}, \widehat{\rho}) \models_s (\texttt{fn}\ x\ \texttt{=>}\ e_0)^\ell$$
$$\underline{\text{iff}} \quad \{\texttt{fn}\ x\ \texttt{=>}\ e_0\} \subseteq \widehat{\mathsf{C}}(\ell) \ \land$$
$$(\widehat{\mathsf{C}}, \widehat{\rho}) \models_s e_0$$

$$(\widehat{\mathsf{C}}, \widehat{\rho}) \models_s (\texttt{fun}\ f\ x\ \texttt{=>}\ e_0)^\ell$$
$$\underline{\text{iff}} \quad \{\texttt{fun}\ f\ x\ \texttt{=>}\ e_0\} \subseteq \widehat{\mathsf{C}}(\ell) \ \land$$
$$(\widehat{\mathsf{C}}, \widehat{\rho}) \models_s e_0 \ \land \ \{\texttt{fun}\ f\ x\ \texttt{=>}\ e_0\} \subseteq \widehat{\rho}(f)$$

$$(\widehat{\mathsf{C}}, \widehat{\rho}) \models_s (t_1^{\ell_1}\ t_2^{\ell_2})^\ell$$
$$\underline{\text{iff}} \quad (\widehat{\mathsf{C}}, \widehat{\rho}) \models_s t_1^{\ell_1} \ \land \ (\widehat{\mathsf{C}}, \widehat{\rho}) \models_s t_2^{\ell_2} \ \land$$
$$(\forall (\texttt{fn}\ x\ \texttt{=>}\ t_0^{\ell_0}) \in \widehat{\mathsf{C}}(\ell_1) :$$
$$\widehat{\mathsf{C}}(\ell_2) \subseteq \widehat{\rho}(x) \ \land \ \widehat{\mathsf{C}}(\ell_0) \subseteq \widehat{\mathsf{C}}(\ell) \quad \boxed{}\ ) \ \land$$
$$(\forall (\texttt{fun}\ f\ x\ \texttt{=>}\ t_0^{\ell_0}) \in \widehat{\mathsf{C}}(\ell_1) :$$
$$\widehat{\mathsf{C}}(\ell_2) \subseteq \widehat{\rho}(x) \ \land \ \widehat{\mathsf{C}}(\ell_0) \subseteq \widehat{\mathsf{C}}(\ell) \quad \boxed{}\ )$$

Syntax Directed Specification (2)

$(\widehat{\mathsf{C}}, \widehat{\rho}) \models_s c^\ell$ always

$(\widehat{\mathsf{C}}, \widehat{\rho}) \models_s x^\ell$ <u>iff</u> $\widehat{\rho}(x) \subseteq \widehat{\mathsf{C}}(\ell)$

$(\widehat{\mathsf{C}}, \widehat{\rho}) \models_s (\texttt{if } t_0^{\ell_0} \texttt{ then } t_1^{\ell_1} \texttt{ else } t_2^{\ell_2})^\ell$
<u>iff</u> $(\widehat{\mathsf{C}}, \widehat{\rho}) \models_s t_0^{\ell_0} \wedge$
$(\widehat{\mathsf{C}}, \widehat{\rho}) \models_s t_1^{\ell_1} \wedge (\widehat{\mathsf{C}}, \widehat{\rho}) \models_s t_2^{\ell_2} \wedge$
$\widehat{\mathsf{C}}(\ell_1) \subseteq \widehat{\mathsf{C}}(\ell) \wedge \widehat{\mathsf{C}}(\ell_2) \subseteq \widehat{\mathsf{C}}(\ell)$

$(\widehat{\mathsf{C}}, \widehat{\rho}) \models_s (\texttt{let } x \texttt{ = } t_1^{\ell_1} \texttt{ in } t_2^{\ell_2})^\ell$
<u>iff</u> $(\widehat{\mathsf{C}}, \widehat{\rho}) \models_s t_1^{\ell_1} \wedge (\widehat{\mathsf{C}}, \widehat{\rho}) \models_s t_2^{\ell_2} \wedge$
$\widehat{\mathsf{C}}(\ell_1) \subseteq \widehat{\rho}(x) \wedge \widehat{\mathsf{C}}(\ell_2) \subseteq \widehat{\mathsf{C}}(\ell)$

$(\widehat{\mathsf{C}}, \widehat{\rho}) \models_s (t_1^{\ell_1} \ op \ t_2^{\ell_2})^\ell$ <u>iff</u> $(\widehat{\mathsf{C}}, \widehat{\rho}) \models_s t_1^{\ell_1} \wedge (\widehat{\mathsf{C}}, \widehat{\rho}) \models_s t_2^{\ell_2}$

Example: loop

```
(let g = (fun f x => (f¹ (fn y => y²)³)⁴)⁵
 in (g⁶ (fn z => z⁷)⁸)⁹)¹⁰
```

Abbreviations:

$$\begin{array}{rcl} \mathsf{f} & = & \texttt{fun f x => (f}^1\ \texttt{(fn y => y}^2\texttt{)}^3\texttt{)}^4 \\ \mathsf{id}_y & = & \texttt{fn y => y}^2 \\ \mathsf{id}_z & = & \texttt{fn z => z}^7 \end{array}$$

One guess of a 0-CFA analysis result:

$$\begin{array}{rclrclrcl} \widehat{\mathsf{C}}_{\mathsf{lp}}(1) & = & \{\mathsf{f}\} & \widehat{\mathsf{C}}_{\mathsf{lp}}(6) & = & \{\mathsf{f}\} & \widehat{\rho}_{\mathsf{lp}}(\mathtt{f}) & = & \{\mathsf{f}\} \\ \widehat{\mathsf{C}}_{\mathsf{lp}}(2) & = & \emptyset & \widehat{\mathsf{C}}_{\mathsf{lp}}(7) & = & \emptyset & \widehat{\rho}_{\mathsf{lp}}(\mathtt{g}) & = & \{\mathsf{f}\} \\ \widehat{\mathsf{C}}_{\mathsf{lp}}(3) & = & \{\mathsf{id}_y\} & \widehat{\mathsf{C}}_{\mathsf{lp}}(8) & = & \{\mathsf{id}_z\} & \widehat{\rho}_{\mathsf{lp}}(\mathtt{x}) & = & \{\mathsf{id}_y, \mathsf{id}_z\} \\ \widehat{\mathsf{C}}_{\mathsf{lp}}(4) & = & \emptyset & \widehat{\mathsf{C}}_{\mathsf{lp}}(9) & = & \emptyset & \widehat{\rho}_{\mathsf{lp}}(\mathtt{y}) & = & \emptyset \\ \widehat{\mathsf{C}}_{\mathsf{lp}}(5) & = & \{\mathsf{f}\} & \widehat{\mathsf{C}}_{\mathsf{lp}}(10) & = & \emptyset & \widehat{\rho}_{\mathsf{lp}}(\mathtt{z}) & = & \emptyset \end{array}$$

Example: Checking the solution

To show

$$(\widehat{\mathsf{C}}_{\mathsf{lp}}, \widehat{\rho}_{\mathsf{lp}}) \models_s \mathsf{loop}$$

we have (among others) to show

$$(\widehat{\mathsf{C}}_{\mathsf{lp}}, \widehat{\rho}_{\mathsf{lp}}) \models_s (\mathtt{g}^6\ (\mathtt{fn\ z\ =>\ z}^7)^8)^9$$

and

$$(\widehat{\mathsf{C}}_{\mathsf{lp}}, \widehat{\rho}_{\mathsf{lp}}) \models_s (\mathtt{f}^1\ (\mathtt{fn\ y\ =>\ y}^2)^3)^4$$

and this is straightforward.

The Lesson

No need for co-induction because the definition is syntax-directed

Preservation of Solutions

Define $(\widehat{\mathsf{C}}_\star^\top, \widehat{\rho}_\star^\top)$ by:

$$\widehat{\mathsf{C}}_\star^\top(\ell) = \begin{cases} \emptyset & \text{if } \ell \notin \mathbf{Lab}_\star \\ \mathbf{Term}_\star & \text{if } \ell \in \mathbf{Lab}_\star \end{cases}$$

$$\widehat{\rho}_\star^\top(x) = \begin{cases} \emptyset & \text{if } x \notin \mathbf{Var}_\star \\ \mathbf{Term}_\star & \text{if } x \in \mathbf{Var}_\star \end{cases}$$

Then all the solutions to "$\models_s e_\star$" that are "less than" $(\widehat{\mathsf{C}}_\star^\top, \widehat{\rho}_\star^\top)$ are solutions to "$\models e_\star$" as well:

Proposition: If $(\widehat{\mathsf{C}}, \widehat{\rho}) \models_s e_\star$ and $(\widehat{\mathsf{C}}, \widehat{\rho}) \sqsubseteq (\widehat{\mathsf{C}}_\star^\top, \widehat{\rho}_\star^\top)$ then $(\widehat{\mathsf{C}}, \widehat{\rho}) \models e_\star$.

(That $(\widehat{\mathsf{C}}, \widehat{\rho}) \sqsubseteq (\widehat{\mathsf{C}}_\star^\top, \widehat{\rho}_\star^\top)$ means that $(\widehat{\mathsf{C}}, \widehat{\rho})$ lives in a "closed universe".)

Proposition:

$\{(\widehat{\mathsf{C}},\widehat{\rho}) \sqsubseteq (\widehat{\mathsf{C}}_\star^\top,\widehat{\rho}_\star^\top) \mid (\widehat{\mathsf{C}},\widehat{\rho}) \models_s e_\star\}$ is a Moore family.

Corollaries:

- each expression $e_\star$ has a Control Flow Analysis that is "less than" $(\widehat{\mathsf{C}}_\star^\top,\widehat{\rho}_\star^\top)$, and

- each expression $e_\star$ has a "least" Control Flow Analysis that is "less than" $(\widehat{\mathsf{C}}_\star^\top,\widehat{\rho}_\star^\top)$.

Constraint Based 0-CFA Analysis

$\mathcal{C}_\star[\![e_\star]\!]$ is a set of constraints of the form

$$\mathit{lhs} \subseteq \mathit{rhs}$$

$$\{t\} \subseteq \mathit{rhs}' \Rightarrow \mathit{lhs} \subseteq \mathit{rhs}$$

where

$$\mathit{rhs} \quad ::= \quad \mathsf{C}(\ell) \;\mid\; \mathsf{r}(x)$$

$$\mathit{lhs} \quad ::= \quad \mathsf{C}(\ell) \;\mid\; \mathsf{r}(x) \;\mid\; \{t\}$$

and all occurrences of t are of the form **fn** x **=>** e_0 or **fun** f x **=>** e_0

Constraint Based Control Flow Analysis (1)

$$\mathcal{C}_\star[\![(\texttt{fn}\ x\ \texttt{=>}\ e_0)^\ell]\!] = \{\ \{\texttt{fn}\ x\ \texttt{=>}\ e_0\} \subseteq \mathsf{C}(\ell)\ \} \cup \mathcal{C}_\star[\![e_0]\!]$$

$$\begin{aligned}\mathcal{C}_\star[\![(\texttt{fun}\ f\ x\ \texttt{=>}\ e_0)^\ell]\!] &= \{\ \{\texttt{fun}\ f\ x\ \texttt{=>}\ e_0\} \subseteq \mathsf{C}(\ell)\ \} \cup \mathcal{C}_\star[\![e_0]\!] \\ &\quad \cup \{\ \{\texttt{fun}\ f\ x\ \texttt{=>}\ e_0\} \subseteq \mathsf{r}(f)\ \}\end{aligned}$$

$$\begin{aligned}\mathcal{C}_\star[\![(t_1^{\ell_1}\ t_2^{\ell_2})^\ell]\!] = \mathcal{C}_\star[\![t_1^{\ell_1}]\!] \cup \mathcal{C}_\star[\![t_2^{\ell_2}]\!] \\ \cup \{\ \{t\} \subseteq \mathsf{C}(\ell_1) \Rightarrow \mathsf{C}(\ell_2) \subseteq \mathsf{r}(x)\ &\mid t = (\texttt{fn}\ x\ \texttt{=>}\ t_0^{\ell_0}) \in \mathbf{Term}_\star\} \\ \cup \{\ \{t\} \subseteq \mathsf{C}(\ell_1) \Rightarrow \mathsf{C}(\ell_0) \subseteq \mathsf{C}(\ell)\ &\mid t = (\texttt{fn}\ x\ \texttt{=>}\ t_0^{\ell_0}) \in \mathbf{Term}_\star\} \\ \cup \{\ \{t\} \subseteq \mathsf{C}(\ell_1) \Rightarrow \mathsf{C}(\ell_2) \subseteq \mathsf{r}(x)\ &\mid t = (\texttt{fun}\ f\ x\ \texttt{=>}\ t_0^{\ell_0}) \in \mathbf{Term}_\star\} \\ \cup \{\ \{t\} \subseteq \mathsf{C}(\ell_1) \Rightarrow \mathsf{C}(\ell_0) \subseteq \mathsf{C}(\ell)\ &\mid t = (\texttt{fun}\ f\ x\ \texttt{=>}\ t_0^{\ell_0}) \in \mathbf{Term}_\star\}\end{aligned}$$

(Eager rather than lazy unfolding – easy but costly.)

Constraint Based Control Flow Analysis (2)

$$\mathcal{C}_\star[\![c^\ell]\!] = \emptyset$$

$$\mathcal{C}_\star[\![x^\ell]\!] = \{\, \mathsf{r}(x) \subseteq \mathsf{C}(\ell) \,\}$$

$$\begin{aligned}\mathcal{C}_\star[\![(\texttt{if } t_0^{\ell_0} \texttt{ then } t_1^{\ell_1} \texttt{ else } t_2^{\ell_2})^\ell]\!] &= \mathcal{C}_\star[\![t_0^{\ell_0}]\!] \cup \mathcal{C}_\star[\![t_1^{\ell_1}]\!] \cup \mathcal{C}_\star[\![t_2^{\ell_2}]\!] \\ &\cup \{\, \mathsf{C}(\ell_1) \subseteq \mathsf{C}(\ell) \,\} \\ &\cup \{\, \mathsf{C}(\ell_2) \subseteq \mathsf{C}(\ell) \,\}\end{aligned}$$

$$\begin{aligned}\mathcal{C}_\star[\![(\texttt{let } x \texttt{ = } t_1^{\ell_1} \texttt{ in } t_2^{\ell_2})^\ell]\!] &= \mathcal{C}_\star[\![t_1^{\ell_1}]\!] \cup \mathcal{C}_\star[\![t_2^{\ell_2}]\!] \\ &\cup \{\, \mathsf{C}(\ell_1) \subseteq \mathsf{r}(x) \,\} \cup \{\, \mathsf{C}(\ell_2) \subseteq \mathsf{C}(\ell) \,\}\end{aligned}$$

$$\mathcal{C}_\star[\![(t_1^{\ell_1} \ \mathit{op}\ t_2^{\ell_2})^\ell]\!] = \mathcal{C}_\star[\![t_1^{\ell_1}]\!] \cup \mathcal{C}_\star[\![t_2^{\ell_2}]\!]$$

Example:

$$
\begin{aligned}
&\mathcal{C}_\star[\![((\texttt{fn x => x}^1)^2\ (\texttt{fn y => y}^3)^4)^5]\!] = \\
&\quad \{\ \{\texttt{fn x => x}^1\} \subseteq \mathsf{C}(2), \\
&\quad\ \ \mathsf{r}(\texttt{x}) \subseteq \mathsf{C}(1), \\
&\quad\ \ \{\texttt{fn y => y}^3\} \subseteq \mathsf{C}(4), \\
&\quad\ \ \mathsf{r}(\texttt{y}) \subseteq \mathsf{C}(3), \\
&\quad\ \ \{\texttt{fn x => x}^1\} \subseteq \mathsf{C}(2) \Rightarrow \mathsf{C}(4) \subseteq \mathsf{r}(\texttt{x}), \\
&\quad\ \ \{\texttt{fn x => x}^1\} \subseteq \mathsf{C}(2) \Rightarrow \mathsf{C}(1) \subseteq \mathsf{C}(5), \\
&\quad\ \ \{\texttt{fn y => y}^3\} \subseteq \mathsf{C}(2) \Rightarrow \mathsf{C}(4) \subseteq \mathsf{r}(\texttt{y}), \\
&\quad\ \ \{\texttt{fn y => y}^3\} \subseteq \mathsf{C}(2) \Rightarrow \mathsf{C}(3) \subseteq \mathsf{C}(5)\ \}
\end{aligned}
$$

Preservation of Solutions

Translating syntactic entities to sets of terms:

$$\begin{array}{rcl} (\widehat{\mathsf{C}},\widehat{\rho})[\![\mathsf{C}(\ell)]\!] & = & \widehat{\mathsf{C}}(\ell) \\ (\widehat{\mathsf{C}},\widehat{\rho})[\![\mathsf{r}(x)]\!] & = & \widehat{\rho}(x) \\ (\widehat{\mathsf{C}},\widehat{\rho})[\![\{t\}]\!] & = & \{t\} \end{array}$$

Satisfaction relation for constraints: $(\widehat{\mathsf{C}},\widehat{\rho}) \models_c (\mathit{lhs} \subseteq \mathit{rhs})$

$$\begin{array}{l} (\widehat{\mathsf{C}},\widehat{\rho}) \models_c (\mathit{lhs} \subseteq \mathit{rhs}) \\ \quad \underline{\text{iff}} \quad (\widehat{\mathsf{C}},\widehat{\rho})[\![\mathit{lhs}]\!] \subseteq (\widehat{\mathsf{C}},\widehat{\rho})[\![\mathit{rhs}]\!] \end{array}$$

$$\begin{array}{l} (\widehat{\mathsf{C}},\widehat{\rho}) \models_c (\{t\} \subseteq \mathit{rhs}' \Rightarrow \mathit{lhs} \subseteq \mathit{rhs}) \\ \quad \underline{\text{iff}} \quad (\{t\} \subseteq (\widehat{\mathsf{C}},\widehat{\rho})[\![\mathit{rhs}']\!] \wedge (\widehat{\mathsf{C}},\widehat{\rho})[\![\mathit{lhs}]\!] \subseteq (\widehat{\mathsf{C}},\widehat{\rho})[\![\mathit{rhs}]\!]) \\ \quad \;\vee \quad\;\; (\{t\} \not\subseteq (\widehat{\mathsf{C}},\widehat{\rho})[\![\mathit{rhs}']\!]) \end{array}$$

Proposition: $(\widehat{\mathsf{C}},\widehat{\rho}) \models_s e_\star$ if and only if $(\widehat{\mathsf{C}},\widehat{\rho}) \models_c \mathcal{C}_\star[\![e_\star]\!]$.

Solving the Constraints (1)

Input: a set of constraints $\mathcal{C}_\star[\![e_\star]\!]$

Output: the least solution $(\widehat{\mathsf{C}}, \widehat{\rho})$ to the constraints

Data structures: a graph with one node for each $\mathsf{C}(\ell)$ and $\mathsf{r}(x)$ (where $\ell \in \mathbf{Lab}_\star$ and $x \in \mathbf{Var}_\star$) and zero, one or two edges for each constraint in $\mathcal{C}_\star[\![e_\star]\!]$

- W: the worklist of the nodes whose outgoing edges should be traversed
- D: an array that for each node gives an element of $\widehat{\mathbf{Val}}_\star$
- E: an array that for each node gives a list of constraints influenced (and outgoing edges)

Auxiliary procedure:

procedure add(q,d) is if $\neg$ ($d \subseteq$ D[q]) then D[q] := D[q] $\cup$ d;
W := cons(q,W);

Solving the Constraints (2)

Step 1 Initialisation

W := nil;

for q in Nodes do D[q] := $\emptyset$; E[q] := nil;

Step 2 Building the graph

for cc in $\mathcal{C}_\star[\![e_\star]\!]$ do

case cc of $\{t\} \subseteq p$: add(p,$\{t\}$);

$p_1 \subseteq p_2$: E[p_1] := cons(cc,E[p_1]);

$\{t\} \subseteq p \Rightarrow p_1 \subseteq p_2$: E[$p_1$] := cons($cc$,E[$p_1$]);

E[p] := cons(cc,E[p]);

Step 3 Iteration

while W $\neq$ nil do

q := head(W); W := tail(W);

for cc in E[q] do

case cc of $p_1 \subseteq p_2$: add(p_2, D[p_1]);

$\{t\} \subseteq p \Rightarrow p_1 \subseteq p_2$: if $t \in$ D[p] then add(p_2, D[p_1]);

Step 4 Recording the solution

for ℓ in $\mathbf{Lab}_\star$ do $\widehat{\mathsf{C}}(\ell)$:= D[C(ℓ)]; for x in $\mathbf{Var}_\star$ do $\widehat{\rho}(x)$:= D[r(x)];

Example:

Initialisation of data structures

p	D[p]	E[p]
C(1)	$\emptyset$	[id$_x$ $\subseteq$ C(2) $\Rightarrow$ C(1) $\subseteq$ C(5)]
C(2)	id$_x$	[id$_y$ $\subseteq$ C(2) $\Rightarrow$ C(3) $\subseteq$ C(5), id$_y$ $\subseteq$ C(2) $\Rightarrow$ C(4) $\subseteq$ r(y), id$_x$ $\subseteq$ C(2) $\Rightarrow$ C(1) $\subseteq$ C(5), id$_x$ $\subseteq$ C(2) $\Rightarrow$ C(4) $\subseteq$ r(x)]
C(3)	$\emptyset$	[id$_y$ $\subseteq$ C(2) $\Rightarrow$ C(3) $\subseteq$ C(5)]
C(4)	id$_y$	[id$_y$ $\subseteq$ C(2) $\Rightarrow$ C(4) $\subseteq$ r(y), id$_x$ $\subseteq$ C(2) $\Rightarrow$ C(4) $\subseteq$ r(x)]
C(5)	$\emptyset$	[ ]
r(x)	$\emptyset$	[r(x) $\subseteq$ C(1)]
r(y)	$\emptyset$	[r(y) $\subseteq$ C(3)]

Example:

Iteration steps

W	[C(4),C(2)]	[r(x),C(2)]	[C(1),C(2)]	[C(5),C(2)]	[C(2)]	[ ]
p	D[p]	D[p]	D[p]	D[p]	D[p]	D[p]
C(1)	$\emptyset$	$\emptyset$	id$_y$	id$_y$	id$_y$	id$_y$
C(2)	id$_x$	id$_x$	id$_x$	id$_x$	id$_x$	id$_x$
C(3)	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$
C(4)	id$_y$	id$_y$	id$_y$	id$_y$	id$_y$	id$_y$
C(5)	$\emptyset$	$\emptyset$	$\emptyset$	id$_y$	id$_y$	id$_y$
r(x)	$\emptyset$	id$_y$	id$_y$	id$_y$	id$_y$	id$_y$
r(y)	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$

Correctness:

Given input $\mathcal{C}_\star[\![e_\star]\!]$ the worklist algorithm terminates and the result $(\widehat{\mathsf{C}}, \widehat{\rho})$ produced by the algorithm satisfies

$$(\widehat{\mathsf{C}}, \widehat{\rho}) = \bigsqcap \{(\widehat{\mathsf{C}}', \widehat{\rho}') \mid (\widehat{\mathsf{C}}', \widehat{\rho}') \models_c \mathcal{C}_\star[\![e_\star]\!]\}$$

and hence it is the least solution to $\mathcal{C}_\star[\![e_\star]\!]$.

Complexity:

The algorithm takes at most $O(n^3)$ steps if the original expression $e_\star$ has size n.

Adding Data Flow Analysis

Idea: extend the set $\widehat{\mathbf{Val}}$ to contain other abstract values than just abstractions

- powerset (possibly finite)
- complete lattice (possibly satisfying Ascending Chain Condition)

Abstract Values as Powersets

Let **Data** be a set of *abstract data values* (i.e. abstract properties of booleans and integers)

$$\widehat{v} \in \widehat{\mathbf{Val}}_d = \mathcal{P}(\mathbf{Term} \cup \mathsf{\mathbf{Data}}) \quad \text{abstract values}$$

For each constant $c \in \mathbf{Const}$ we need an element $d_c \in$ **Data**

For each operator $\mathit{op} \in \mathbf{Op}$ we need a total function

$$\widehat{\mathsf{op}} : \widehat{\mathbf{Val}}_d \times \widehat{\mathbf{Val}}_d \to \widehat{\mathbf{Val}}_d$$

typically

$$\widehat{v}_1 \ \widehat{\mathsf{op}} \ \widehat{v}_2 = \bigcup \{ d_{op}(d_1, d_2) \mid d_1 \in \widehat{v}_1 \cap \mathsf{\mathbf{Data}}, d_2 \in \widehat{v}_2 \cap \mathsf{\mathbf{Data}} \}$$

for some d_{op} : **Data** $\times$ **Data** $\to$ $\mathcal{P}$(**Data**)

Example: *Detection of Signs Analysis*

$\mathbf{Data}_{\mathsf{sign}} = \{$tt, ff, -, 0, +$\}$

$d_{\mathtt{true}} =$ tt

$d_7 =$ +

$\widehat{+}$ is defined from

d_{+}	tt	ff	-	0	+
tt	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$
ff	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$
-	$\emptyset$	$\emptyset$	{-}	{-}	{-, 0, +}
0	$\emptyset$	$\emptyset$	{-}	{0}	{+}
+	$\emptyset$	$\emptyset$	{-, 0, +}	{+}	{+}

Abstract Values as Powersets (1)

$$(\widehat{\mathsf{C}}, \widehat{\rho}) \models_d (\texttt{fn}\ x\ \texttt{=>}\ e_0)^\ell \quad \underline{\text{iff}} \quad \{\texttt{fn}\ x\ \texttt{=>}\ e_0\} \subseteq \widehat{\mathsf{C}}(\ell)$$

$$(\widehat{\mathsf{C}}, \widehat{\rho}) \models_d (\texttt{fun}\ f\ x\ \texttt{=>}\ e_0)^\ell \quad \underline{\text{iff}} \quad \{\texttt{fun}\ f\ x\ \texttt{=>}\ e_0\} \subseteq \widehat{\mathsf{C}}(\ell)$$

$$\begin{array}{ll}
\multicolumn{2}{l}{(\widehat{\mathsf{C}}, \widehat{\rho}) \models_d (t_1^{\ell_1}\ t_2^{\ell_2})^\ell} \\
\underline{\text{iff}} & (\widehat{\mathsf{C}}, \widehat{\rho}) \models_d t_1^{\ell_1} \ \wedge\ (\widehat{\mathsf{C}}, \widehat{\rho}) \models_d t_2^{\ell_2} \ \wedge \\
& (\forall (\texttt{fn}\ x\ \texttt{=>}\ t_0^{\ell_0}) \in \widehat{\mathsf{C}}(\ell_1) : \\
& \quad (\widehat{\mathsf{C}}, \widehat{\rho}) \models_d t_0^{\ell_0} \ \wedge \\
& \quad \widehat{\mathsf{C}}(\ell_2) \subseteq \widehat{\rho}(x) \ \wedge\ \widehat{\mathsf{C}}(\ell_0) \subseteq \widehat{\mathsf{C}}(\ell)) \ \wedge \\
& (\forall (\texttt{fun}\ f\ x\ \texttt{=>}\ t_0^{\ell_0}) \in \widehat{\mathsf{C}}(\ell_1) : \\
& \quad (\widehat{\mathsf{C}}, \widehat{\rho}) \models_d t_0^{\ell_0} \ \wedge \\
& \quad \widehat{\mathsf{C}}(\ell_2) \subseteq \widehat{\rho}(x) \ \wedge\ \widehat{\mathsf{C}}(\ell_0) \subseteq \widehat{\mathsf{C}}(\ell) \ \wedge \\
& \quad \{\texttt{fun}\ f\ x\ \texttt{=>}\ t_0^{\ell_0}\} \subseteq \widehat{\rho}(f))
\end{array}$$

Abstract Values as Powersets (2)

$(\widehat{\mathsf{C}}, \widehat{\rho}) \models_d c^\ell \quad \underline{\text{iff}} \quad \{d_c\} \subseteq \widehat{\mathsf{C}}(\ell)$

$(\widehat{\mathsf{C}}, \widehat{\rho}) \models_d x^\ell \quad \underline{\text{iff}} \quad \widehat{\rho}(x) \subseteq \widehat{\mathsf{C}}(\ell)$

$$
\begin{array}{l}
(\widehat{\mathsf{C}}, \widehat{\rho}) \models_d (\texttt{if } t_0^{\ell_0} \texttt{ then } t_1^{\ell_1} \texttt{ else } t_2^{\ell_2})^\ell \\
\quad \underline{\text{iff}} \quad (\widehat{\mathsf{C}}, \widehat{\rho}) \models_d t_0^{\ell_0} \ \wedge \\
\qquad (d_{\texttt{true}} \in \widehat{\mathsf{C}}(\ell_0) \Rightarrow (\boxed{(\widehat{\mathsf{C}}, \widehat{\rho}) \models_d t_1^{\ell_1}} \ \wedge \ \widehat{\mathsf{C}}(\ell_1) \subseteq \widehat{\mathsf{C}}(\ell))) \ \wedge \\
\qquad (d_{\texttt{false}} \in \widehat{\mathsf{C}}(\ell_0) \Rightarrow (\boxed{(\widehat{\mathsf{C}}, \widehat{\rho}) \models_d t_2^{\ell_2}} \ \wedge \ \widehat{\mathsf{C}}(\ell_2) \subseteq \widehat{\mathsf{C}}(\ell)))
\end{array}
$$

$$
\begin{array}{l}
(\widehat{\mathsf{C}}, \widehat{\rho}) \models_d (\texttt{let } x \texttt{ = } t_1^{\ell_1} \texttt{ in } t_2^{\ell_2})^\ell \\
\quad \underline{\text{iff}} \quad (\widehat{\mathsf{C}}, \widehat{\rho}) \models_d t_1^{\ell_1} \ \wedge \ (\widehat{\mathsf{C}}, \widehat{\rho}) \models_d t_2^{\ell_2} \ \wedge \ \widehat{\mathsf{C}}(\ell_1) \subseteq \widehat{\rho}(x) \ \wedge \ \widehat{\mathsf{C}}(\ell_2) \subseteq \widehat{\mathsf{C}}(\ell)
\end{array}
$$

$$
\begin{array}{l}
(\widehat{\mathsf{C}}, \widehat{\rho}) \models_d (t_1^{\ell_1} \ op \ t_2^{\ell_2})^\ell \\
\quad \underline{\text{iff}} \quad (\widehat{\mathsf{C}}, \widehat{\rho}) \models_d t_1^{\ell_1} \ \wedge \ (\widehat{\mathsf{C}}, \widehat{\rho}) \models_d t_2^{\ell_2} \ \wedge \ \widehat{\mathsf{C}}(\ell_1) \ \widehat{\mathsf{op}} \ \widehat{\mathsf{C}}(\ell_2) \subseteq \widehat{\mathsf{C}}(\ell)
\end{array}
$$

Example:

```
(let f = (fn x => (if (x^1 > 0^2)^3 then (fn y => y^4)^5
              else (fn z => 25^6)^7)^8)^9
 in ((f^10 3^11)^12 0^13)^14)^15
```

A pure 0-CFA analysis will not be able to discover that the `else`-branch of the conditional will never be executed.

When we combine the analysis with a Detection of Signs Analysis then the analysis can determine that only `fn y =>` $\mathtt{y}^4$ is a possible abstraction at label 12.

Example:

	$(\widehat{C}, \widehat{\rho})$	$(\widehat{C}, \widehat{\rho})$
1	$\emptyset$	$\{+\}$
2	$\emptyset$	$\{0\}$
3	$\emptyset$	$\{tt\}$
4	$\emptyset$	$\{0\}$
5	$\{\texttt{fn y => y}^4\}$	$\{\texttt{fn y => y}^4\}$
6	$\emptyset$	$\emptyset$
7	$\{\texttt{fn z => 25}^6\}$	$\emptyset$
8	$\{\ \texttt{fn y => y}^4\texttt{, fn z => 25}^6\ \}$	$\{\texttt{fn y => y}^4\}$
9	$\{\texttt{fn x => }(\cdots)^8\}$	$\{\texttt{fn x => }(\cdots)^8\}$
10	$\{\texttt{fn x => }(\cdots)^8\}$	$\{\texttt{fn x => }(\cdots)^8\}$
11	$\emptyset$	$\{+\}$
12	$\{\ \texttt{fn y => y}^4\texttt{, fn z => 25}^6\ \}$	$\{\texttt{fn y => y}^4\}$
13	$\emptyset$	$\{0\}$
14	$\emptyset$	$\{0\}$
15	$\emptyset$	$\{0\}$
f	$\{\texttt{fn x => }(\cdots)^8\}$	$\{\texttt{fn x => }(\cdots)^8\}$
x	$\emptyset$	$\{+\}$
y	$\emptyset$	$\{0\}$
z	$\emptyset$	$\emptyset$

Abstract Values as Complete Lattices

A *monotone structure* consists of:

- a complete lattice L, and
- a set $\mathcal{F}$ of monotone functions of $L \times L \to L$.

An *instance* of a monotone structure consists of the structure $(L, \mathcal{F})$ and

- a mapping $\iota.$ from the constants $c \in \mathbf{Const}$ to values in L, and
- a mapping $f.$ from the binary operators *op* $\in \mathbf{Op}$ to functions of $\mathcal{F}$.

Example:

A monotone structure corresponding to the previous development will have L to be $\mathcal{P}(\textbf{Data})$ and $\mathcal{F}$ to be the monotone functions of $\mathcal{P}(\textbf{Data}) \times \mathcal{P}(\textbf{Data}) \to \mathcal{P}(\textbf{Data})$.

(L satisfies the Ascending Chain Property iff **Data** is finite.)

An instance of the monotone structure is then obtained by taking

$$\iota_c = \{d_c\}$$

for all constants c (and with $d_c \in \textbf{Data}$ as above) and

$$f_{op}(l_1, l_2) = \bigcup\{d_{op}(d_1, d_2) \mid d_1 \in l_1, d_2 \in l_2\}$$

for all binary operators *op* (and where $d_{op} : \textbf{Data} \times \textbf{Data} \to \mathcal{P}(\textbf{Data})$ is as above).

Example: A monotone structure for *Constant Propagation Analysis* will have L to be $\mathbf{Z}_{\perp}^{\top} \times \mathcal{P}(\{\mathsf{tt}, \mathsf{ff}\})$ and $\mathcal{F}$ to be the monotone functions of $L \times L \to L$.

An instance of the monotone structure is obtained by taking e.g. $\iota_7 = (7, \emptyset)$ and $\iota_{\mathtt{true}} = (\perp, \{\mathsf{tt}\})$. For a binary operator as + we can take:

$$f_{+}(l_1, l_2) = \begin{cases} (z_1 + z_2, \emptyset) & \text{if } l_1 = (z_1, \cdots), l_2 = (z_2, \cdots), \\ & \text{ and } z_1, z_2 \in \mathbf{Z} \\ (\perp, \emptyset) & \text{if } l_1 = (z_1, \cdots), l_2 = (z_2, \cdots), \\ & \text{ and } z_1 = \perp \text{ or } z_2 = \perp \\ (\top, \emptyset) & \text{otherwise} \end{cases}$$

Abstract Domains

For the Control Flow Analysis:

$$\begin{array}{llllll} \hat{v} & \in & \widehat{\mathbf{Val}} & = & \mathcal{P}(\mathbf{Term}) & \text{abstract values} \\ \hat{\rho} & \in & \widehat{\mathbf{Env}} & = & \mathbf{Var} \to \widehat{\mathbf{Val}} & \text{abstract environments} \\ \widehat{\mathsf{C}} & \in & \widehat{\mathbf{Cache}} & = & \mathbf{Lab} \to \widehat{\mathbf{Val}} & \text{abstract caches} \end{array}$$

For the Data Flow Analysis:

$$\begin{array}{llllll} \hat{d} & \in & \widehat{\mathbf{Data}} & = & L & \text{abstract data values} \\ \hat{\delta} & \in & \mathbf{D}\widehat{\mathbf{Env}} & = & \mathbf{Var} \to \widehat{\mathbf{Data}} & \text{abstract data environments} \\ \widehat{\mathsf{D}} & \in & \mathbf{D}\widehat{\mathbf{Cache}} & = & \mathbf{Lab} \to \widehat{\mathbf{Data}} & \text{abstract data caches} \end{array}$$

Abstract Values as Complete Lattices (1)

$$(\widehat{C}, \widehat{D}, \widehat{\rho}, \widehat{\delta}) \models_D (\texttt{fn}\ x\ \texttt{=>}\ e_0)^\ell \quad \underline{\text{iff}} \quad \{\texttt{fn}\ x\ \texttt{=>}\ e_0\} \subseteq \widehat{C}(\ell)$$

$$(\widehat{C}, \widehat{D}, \widehat{\rho}, \widehat{\delta}) \models_D (\texttt{fun}\ f\ x\ \texttt{=>}\ e_0)^\ell \quad \underline{\text{iff}} \quad \{\texttt{fun}\ f\ x\ \texttt{=>}\ e_0\} \subseteq \widehat{C}(\ell)$$

$$\begin{array}{l} (\widehat{C}, \widehat{D}, \widehat{\rho}, \widehat{\delta}) \models_D (t_1^{\ell_1}\ t_2^{\ell_2})^\ell \\ \quad \underline{\text{iff}} \quad (\widehat{C}, \widehat{D}, \widehat{\rho}, \widehat{\delta}) \models_D t_1^{\ell_1} \ \wedge\ (\widehat{C}, \widehat{D}, \widehat{\rho}, \widehat{\delta}) \models_D t_2^{\ell_2} \ \wedge \\ \qquad (\forall (\texttt{fn}\ x\ \texttt{=>}\ t_0^{\ell_0}) \in \widehat{C}(\ell_1) :\ (\widehat{C}, \widehat{D}, \widehat{\rho}, \widehat{\delta}) \models_D t_0^{\ell_0} \ \wedge \\ \qquad\qquad \widehat{C}(\ell_2) \subseteq \widehat{\rho}(x) \ \wedge\ \widehat{D}(\ell_2) \sqsubseteq \widehat{\delta}(x) \ \wedge \\ \qquad\qquad \widehat{C}(\ell_0) \subseteq \widehat{C}(\ell) \ \wedge\ \widehat{D}(\ell_0) \sqsubseteq \widehat{D}(\ell)) \ \wedge \\ \qquad (\forall (\texttt{fun}\ f\ x\ \texttt{=>}\ t_0^{\ell_0}) \in \widehat{C}(\ell_1) :\ (\widehat{C}, \widehat{D}, \widehat{\rho}, \widehat{\delta}) \models_D t_0^{\ell_0} \ \wedge \\ \qquad\qquad \widehat{C}(\ell_2) \subseteq \widehat{\rho}(x) \ \wedge\ \widehat{D}(\ell_2) \sqsubseteq \widehat{\delta}(x) \ \wedge \\ \qquad\qquad \widehat{C}(\ell_0) \subseteq \widehat{C}(\ell) \ \wedge\ \widehat{D}(\ell_0) \sqsubseteq \widehat{D}(\ell) \ \wedge \\ \qquad\qquad \{\texttt{fun}\ f\ x\ \texttt{=>}\ t_0^{\ell_0}\} \subseteq \widehat{\rho}(f)) \end{array}$$

Abstract Values as Complete Lattices (2)

$(\widehat{\mathsf{C}}, \widehat{\mathsf{D}}, \widehat{\rho}, \widehat{\delta}) \models_D c^\ell$ $\quad$ <u>iff</u> $\quad \iota_C \sqsubseteq \widehat{\mathsf{D}}(\ell)$

$(\widehat{\mathsf{C}}, \widehat{\mathsf{D}}, \widehat{\rho}, \widehat{\delta}) \models_D x^\ell$ $\quad$ <u>iff</u> $\quad \widehat{\rho}(x) \subseteq \widehat{\mathsf{C}}(\ell) \;\wedge\; \widehat{\delta}(x) \sqsubseteq \widehat{\mathsf{D}}(\ell)$

$$
\begin{array}{l}
(\widehat{\mathsf{C}}, \widehat{\mathsf{D}}, \widehat{\rho}, \widehat{\delta}) \models_D (\texttt{if } t_0^{\ell_0} \texttt{ then } t_1^{\ell_1} \texttt{ else } t_2^{\ell_2})^\ell \\
\quad \underline{\text{iff}} \quad (\widehat{\mathsf{C}}, \widehat{\mathsf{D}}, \widehat{\rho}, \widehat{\delta}) \models_D t_0^{\ell_0} \;\wedge \\
\qquad\quad (\iota_{\texttt{true}} \sqsubseteq \widehat{\mathsf{D}}(\ell_0) \Rightarrow (\widehat{\mathsf{C}}, \widehat{\mathsf{D}}, \widehat{\rho}, \widehat{\delta}) \models_D t_1^{\ell_1} \;\wedge \\
\qquad\qquad\qquad \widehat{\mathsf{C}}(\ell_1) \subseteq \widehat{\mathsf{C}}(\ell) \wedge \widehat{\mathsf{D}}(\ell_1) \sqsubseteq \widehat{\mathsf{D}}(\ell)) \;\wedge \\
\qquad\quad (\iota_{\texttt{false}} \sqsubseteq \widehat{\mathsf{D}}(\ell_0) \Rightarrow (\widehat{\mathsf{C}}, \widehat{\mathsf{D}}, \widehat{\rho}, \widehat{\delta}) \models_D t_2^{\ell_2} \;\wedge \\
\qquad\qquad\qquad \widehat{\mathsf{C}}(\ell_2) \subseteq \widehat{\mathsf{C}}(\ell) \wedge \widehat{\mathsf{D}}(\ell_2) \sqsubseteq \widehat{\mathsf{D}}(\ell))
\end{array}
$$

Abstract Values as Complete Lattices (3)

$$(\widehat{\mathsf{C}}, \widehat{\mathsf{D}}, \widehat{\rho}, \widehat{\delta}) \models_D (\texttt{let}\ x\ \texttt{=}\ t_1^{\ell_1}\ \texttt{in}\ t_2^{\ell_2})^{\ell}$$

$$\begin{array}{ll} \underline{\text{iff}} & (\widehat{\mathsf{C}}, \widehat{\mathsf{D}}, \widehat{\rho}, \widehat{\delta}) \models_D t_1^{\ell_1} \ \wedge \\ & (\widehat{\mathsf{C}}, \widehat{\mathsf{D}}, \widehat{\rho}, \widehat{\delta}) \models_D t_2^{\ell_2} \ \wedge \\ & \widehat{\mathsf{C}}(\ell_1) \subseteq \widehat{\rho}(x) \wedge \widehat{\mathsf{D}}(\ell_1) \sqsubseteq \widehat{\delta}(x) \ \wedge \ \widehat{\mathsf{C}}(\ell_2) \subseteq \widehat{\mathsf{C}}(\ell) \wedge \widehat{\mathsf{D}}(\ell_2) \sqsubseteq \widehat{\mathsf{D}}(\ell) \end{array}$$

$$(\widehat{\mathsf{C}}, \widehat{\mathsf{D}}, \widehat{\rho}, \widehat{\delta}) \models_D (t_1^{\ell_1}\ op\ t_2^{\ell_2})^{\ell}$$

$$\begin{array}{ll} \underline{\text{iff}} & (\widehat{\mathsf{C}}, \widehat{\mathsf{D}}, \widehat{\rho}, \widehat{\delta}) \models_D t_1^{\ell_1} \ \wedge \ (\widehat{\mathsf{C}}, \widehat{\mathsf{D}}, \widehat{\rho}, \widehat{\delta}) \models_D t_2^{\ell_2} \ \wedge \\ & f_{op}(\widehat{\mathsf{D}}(\ell_1), \widehat{\mathsf{D}}(\ell_2)) \sqsubseteq \widehat{\mathsf{D}}(\ell) \end{array}$$

Example:

	$(\widehat{C}, \widehat{\rho})$	$(\widehat{C}, \widehat{\rho})$	$(\widehat{C}, \widehat{\rho})$	$(\widehat{D}, \widehat{\delta})$
1	$\emptyset$	{+}	$\emptyset$	{+}
2	$\emptyset$	{0}	$\emptyset$	{0}
3	$\emptyset$	{tt}	$\emptyset$	{tt}
4	$\emptyset$	{0}	$\emptyset$	{0}
5	{fn y => y^4}	{fn y => y^4}	{fn y => y^4}	$\emptyset$
6	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$
7	{fn z => 25^6}	$\emptyset$	$\emptyset$	$\emptyset$
8	{ fn y => y^4, fn z => 25^6 }	{fn y => y^4}	{fn y => y^4}	$\emptyset$
9	{fn x => $(\cdots)^8$}	{fn x => $(\cdots)^8$}	{fn x => $(\cdots)^8$}	$\emptyset$
10	{fn x => $(\cdots)^8$}	{fn x => $(\cdots)^8$}	{fn x => $(\cdots)^8$}	$\emptyset$
11	$\emptyset$	{+}	$\emptyset$	{+}
12	{ fn y => y^4, fn z => 25^6 }	{fn y => y^4}	{fn y => y^4}	$\emptyset$
13	$\emptyset$	{0}	$\emptyset$	{0}
14	$\emptyset$	{0}	$\emptyset$	{0}
15	$\emptyset$	{0}	$\emptyset$	{0}
f	{fn x => $(\cdots)^8$}	{fn x => $(\cdots)^8$}	{fn x => $(\cdots)^8$}	$\emptyset$
x	$\emptyset$	{+}	$\emptyset$	{+}
y	$\emptyset$	{0}	$\emptyset$	{0}
z	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$

Staging the specification

Alternative clause for the conditional where the data flow component *cannot* influence the control flow component:

$$
\begin{array}{ll}
(\widehat{\mathsf{C}}, \widehat{\mathsf{D}}, \widehat{\rho}, \widehat{\delta}) \models_D (\texttt{if}\ t_0^{\ell_0}\ \texttt{then}\ t_1^{\ell_1}\ \texttt{else}\ t_2^{\ell_2})^{\ell} & \\
\quad\quad \underline{\text{iff}} \quad (\widehat{\mathsf{C}}, \widehat{\mathsf{D}}, \widehat{\rho}, \widehat{\delta}) \models_D t_0^{\ell_0}\ \wedge & \\
\quad\quad\quad\quad (\widehat{\mathsf{C}}, \widehat{\mathsf{D}}, \widehat{\rho}, \widehat{\delta}) \models_D t_1^{\ell_1}\ \wedge\ \widehat{\mathsf{C}}(\ell_1) \subseteq \widehat{\mathsf{C}}(\ell) \wedge \widehat{\mathsf{D}}(\ell_1) \sqsubseteq \widehat{\mathsf{D}}(\ell)\ \wedge & \\
\quad\quad\quad\quad (\widehat{\mathsf{C}}, \widehat{\mathsf{D}}, \widehat{\rho}, \widehat{\delta}) \models_D t_2^{\ell_2}\ \wedge\ \widehat{\mathsf{C}}(\ell_2) \subseteq \widehat{\mathsf{C}}(\ell) \wedge \widehat{\mathsf{D}}(\ell_2) \sqsubseteq \widehat{\mathsf{D}}(\ell) &
\end{array}
$$

Compare with flow-insensitive Data Flow Analyses.

Adding Context Information

Mono-variant analysis: does not distinguish the various instances of variables and program points from one another. (Compare with context-insensitive interprocedural analysis.) 0-CFA is a typical example.

Poly-variant analysis: distinguishes between the various instances of variables and program points. (Compare with context-sensitive interprocedural analysis.)

Example:

$$\texttt{(let f = (fn x => x}^{1}\texttt{)}^{2}\texttt{ in ((f}^{3}\texttt{ f}^{4}\texttt{)}^{5}\texttt{ (fn y => y}^{6}\texttt{)}^{7}\texttt{)}^{8}\texttt{)}^{9}$$

The least 0-CFA analysis:

$$\begin{array}{lcllcl}
\widehat{\mathsf{C}}_{\mathsf{id}}(1) & = & \{\texttt{fn x => x}^1, \texttt{fn y => y}^6\} & \widehat{\mathsf{C}}_{\mathsf{id}}(2) & = & \{\texttt{fn x => x}^1\} \\
\widehat{\mathsf{C}}_{\mathsf{id}}(3) & = & \{\texttt{fn x => x}^1\} & \widehat{\mathsf{C}}_{\mathsf{id}}(4) & = & \{\texttt{fn x => x}^1\} \\
\widehat{\mathsf{C}}_{\mathsf{id}}(5) & = & \{\texttt{fn x => x}^1, \texttt{fn y => y}^6\} & \widehat{\mathsf{C}}_{\mathsf{id}}(6) & = & \{\texttt{fn y => y}^6\} \\
\widehat{\mathsf{C}}_{\mathsf{id}}(7) & = & \{\texttt{fn y => y}^6\} & \widehat{\mathsf{C}}_{\mathsf{id}}(8) & = & \{\texttt{fn x => x}^1, \texttt{fn y => y}^6\} \\
\widehat{\mathsf{C}}_{\mathsf{id}}(9) & = & \{\texttt{fn x => x}^1, \texttt{fn y => y}^6\} & & & \\
\widehat{\rho}_{\mathsf{id}}(\texttt{f}) & = & \{\texttt{fn x => x}^1\} & \widehat{\rho}_{\mathsf{id}}(\texttt{x}) & = & \{\texttt{fn x => x}^1, \texttt{fn y => y}^6\} \\
\widehat{\rho}_{\mathsf{id}}(\texttt{y}) & = & \{\texttt{fn y => y}^6\} & & &
\end{array}$$

The analysis says that the expression may evaluate to $\texttt{fn x => x}^1$ or $\texttt{fn y => y}^6$.

However, only $\texttt{fn y => y}^6$ is a possible result.

A purely syntactic solution:

Expand

```
(let f = (fn x => x) in ((f f) (fn y => y)))
```

into

```
let f1 = (fn x1 => x1)
in  let f2 = (fn x2 => x2) in (f1 f2) (fn y => y)
```

and analyse the expanded expression.

The 0-CFA analysis is now able to deduce that the overall expression will evaluate to `fn y => y` only.

A purely semantic solution: Uniform *k*-CFA

Idea: extend the set $\widehat{\mathbf{Val}}$ to include context information

In a (uniform) *k-CFA* a context δ records the last k dynamic call points; hence contexts will be sequences of labels of length at most k and they will be updated whenever a function application is analysed. (Compare call strings of length at most k.)

Abstract Domains

$$
\begin{array}{llllll}
\delta & \in & \boxed{\Delta} & = & \mathbf{Lab}^{\leq k} & \text{context information} \\
ce & \in & \mathbf{CEnv} & = & \mathbf{Var} \to \boxed{\Delta} & \text{context environments} \\
\widehat{v} & \in & \widehat{\mathbf{Val}} & = & \mathcal{P}(\mathbf{Term} \times \mathbf{CEnv}) & \text{abstract values} \\
\widehat{\rho} & \in & \widehat{\mathbf{Env}} & = & (\mathbf{Var} \times \boxed{\Delta}) \to \widehat{\mathbf{Val}} & \text{abstract environments} \\
\widehat{\mathsf{C}} & \in & \widehat{\mathbf{Cache}} & = & (\mathbf{Lab} \times \boxed{\Delta}) \to \widehat{\mathbf{Val}} & \text{abstract caches}
\end{array}
$$

(Uniform because Δ used both for $\widehat{\mathbf{Env}}$ and $\widehat{\mathbf{Cache}}$.)

Acceptability Relation

$$(\widehat{\mathsf{C}}, \widehat{\rho}) \models_{\delta}^{ce} e$$

where

- ce is the current context environment – will be changed when new bindings are made
- δ is the current context – will be changed when functions are called

Idea: The formula expresses that $(\widehat{\mathsf{C}}, \widehat{\rho})$ is an acceptable analysis of e in the *context* specified by ce and δ.

Control Flow Analysis with Context (1)

$(\widehat{\mathsf{C}}, \widehat{\rho}) \models^{ce}_{\delta} (\texttt{fn}\ x\ \texttt{=>}\ e_0)^{\ell}$ iff $\{(\texttt{fn}\ x\ \texttt{=>}\ e_0, ce)\} \subseteq \widehat{\mathsf{C}}(\ell, \delta)$

$(\widehat{\mathsf{C}}, \widehat{\rho}) \models^{ce}_{\delta} (\texttt{fun}\ f\ x\ \texttt{=>}\ e_0)^{\ell}$ iff $\{(\texttt{fun}\ f\ x\ \texttt{=>}\ e_0, ce)\} \subseteq \widehat{\mathsf{C}}(\ell, \delta)$

$(\widehat{\mathsf{C}}, \widehat{\rho}) \models^{ce}_{\delta} (t_1^{\ell_1}\ t_2^{\ell_2})^{\ell}$

iff $(\widehat{\mathsf{C}}, \widehat{\rho}) \models^{ce}_{\delta} t_1^{\ell_1} \ \wedge\ (\widehat{\mathsf{C}}, \widehat{\rho}) \models^{ce}_{\delta} t_2^{\ell_2} \ \wedge$

$(\forall (\texttt{fn}\ x\ \texttt{=>}\ t_0^{\ell_0}, ce_0) \in \widehat{\mathsf{C}}(\ell_1, \delta):$

$(\widehat{\mathsf{C}}, \widehat{\rho}) \models^{ce'_0}_{\delta_0} t_0^{\ell_0} \ \wedge\ \widehat{\mathsf{C}}(\ell_2, \delta) \subseteq \widehat{\rho}(x, \delta_0) \ \wedge\ \widehat{\mathsf{C}}(\ell_0, \delta_0) \subseteq \widehat{\mathsf{C}}(\ell, \delta)$

where $\delta_0 = \lceil \delta, \ell \rceil_k$ and $ce'_0 = ce_0[x \mapsto \delta_0]) \ \wedge$

$(\forall (\texttt{fun}\ f\ x\ \texttt{=>}\ t_0^{\ell_0}, ce_0) \in \widehat{\mathsf{C}}(\ell_1, \delta):$

$(\widehat{\mathsf{C}}, \widehat{\rho}) \models^{ce'_0}_{\delta_0} t_0^{\ell_0} \ \wedge\ \widehat{\mathsf{C}}(\ell_2, \delta) \subseteq \widehat{\rho}(x, \delta_0) \ \wedge\ \widehat{\mathsf{C}}(\ell_0, \delta_0) \subseteq \widehat{\mathsf{C}}(\ell, \delta) \ \wedge$

$\{(\texttt{fun}\ f\ x\ \texttt{=>}\ t_0^{\ell_0}, ce_0)\} \subseteq \widehat{\rho}(f, \delta_0)$

where $\delta_0 = \lceil \delta, \ell \rceil_k$ and $ce'_0 = ce_0[f \mapsto \delta_0, x \mapsto \delta_0])$

Control Flow Analysis with Context (2)

$(\widehat{\mathsf{C}}, \widehat{\rho}) \models_{\delta}^{ce} c^{\ell}$ always

$(\widehat{\mathsf{C}}, \widehat{\rho}) \models_{\delta}^{ce} x^{\ell}$ $\quad$ <u>iff</u> $\quad$ $\widehat{\rho}(x, ce(x)) \subseteq \widehat{\mathsf{C}}(\ell, \delta)$

$(\widehat{\mathsf{C}}, \widehat{\rho}) \models_{\delta}^{ce} (\texttt{if } t_0^{\ell_0} \texttt{ then } t_1^{\ell_1} \texttt{ else } t_2^{\ell_2})^{\ell}$
<u>iff</u> $\quad$ $(\widehat{\mathsf{C}}, \widehat{\rho}) \models_{\delta}^{ce} t_0^{\ell_0} \ \wedge \ (\widehat{\mathsf{C}}, \widehat{\rho}) \models_{\delta}^{ce} t_1^{\ell_1} \ \wedge \ (\widehat{\mathsf{C}}, \widehat{\rho}) \models_{\delta}^{ce} t_2^{\ell_2} \ \wedge$
$\widehat{\mathsf{C}}(\ell_1, \delta) \subseteq \widehat{\mathsf{C}}(\ell, \delta) \ \wedge \ \widehat{\mathsf{C}}(\ell_2, \delta) \subseteq \widehat{\mathsf{C}}(\ell, \delta)$

$(\widehat{\mathsf{C}}, \widehat{\rho}) \models_{\delta}^{ce} (\texttt{let } x \texttt{ = } t_1^{\ell_1} \texttt{ in } t_2^{\ell_2})^{\ell}$
<u>iff</u> $\quad$ $(\widehat{\mathsf{C}}, \widehat{\rho}) \models_{\delta}^{ce} t_1^{\ell_1} \ \wedge \ (\widehat{\mathsf{C}}, \widehat{\rho}) \models_{\delta}^{ce'} t_2^{\ell_2} \ \wedge$
$\widehat{\mathsf{C}}(\ell_1, \delta) \subseteq \widehat{\rho}(x, \delta) \ \wedge \ \widehat{\mathsf{C}}(\ell_2, \delta) \subseteq \widehat{\mathsf{C}}(\ell, \delta)$
where $ce' = ce[x \mapsto \delta]$

$(\widehat{\mathsf{C}}, \widehat{\rho}) \models_{\delta}^{ce} (t_1^{\ell_1} \ op \ t_2^{\ell_2})^{\ell}$ $\quad$ <u>iff</u> $\quad$ $(\widehat{\mathsf{C}}, \widehat{\rho}) \models_{\delta}^{ce} t_1^{\ell_1} \ \wedge \ (\widehat{\mathsf{C}}, \widehat{\rho}) \models_{\delta}^{ce} t_2^{\ell_2}$

Example:

$$\texttt{(let f = (fn x => x}^1)^2\ \texttt{in}\ ((\texttt{f}^3\ \texttt{f}^4)^5\ \texttt{(fn y => y}^6)^7)^8)^9$$

Contexts of interest for uniform 1-CFA:

- Λ: the initial context
- 5: the context when the application point labelled 5 has been passed
- 8: the context when the application point labelled 8 has been passed

Context environments of interest for uniform 1-CFA:

$ce_0 = [\ ]$	the initial (empty) context environment
$ce_1 = ce_0[\mathtt{f} \mapsto \Lambda]$	the context environment for the analysis of the body of the `let`-construct
$ce_2 = ce_0[\mathtt{x} \mapsto 5]$	the context environment used for the analysis of the body of `f` initiated at the application point 5
$ce_3 = ce_0[\mathtt{x} \mapsto 8]$	the context environment used for the analysis of the body of `f` initiated at the application point 8.

Example: Let us take $\widehat{\mathsf{C}}_{\mathsf{id}}{}'$ and $\widehat{\rho}_{\mathsf{id}}{}'$ to be:

$$\widehat{\mathsf{C}}_{\mathsf{id}}{}'(1,5) = \{(\texttt{fn x => x}^1,\mathsf{ce}_0)\} \qquad \widehat{\mathsf{C}}_{\mathsf{id}}{}'(1,8) = \{(\texttt{fn y => y}^6,\mathsf{ce}_0)\}$$
$$\widehat{\mathsf{C}}_{\mathsf{id}}{}'(2,\Lambda) = \{(\texttt{fn x => x}^1,\mathsf{ce}_0)\} \qquad \widehat{\mathsf{C}}_{\mathsf{id}}{}'(3,\Lambda) = \{(\texttt{fn x => x}^1,\mathsf{ce}_0)\}$$
$$\widehat{\mathsf{C}}_{\mathsf{id}}{}'(4,\Lambda) = \{(\texttt{fn x => x}^1,\mathsf{ce}_0)\} \qquad \widehat{\mathsf{C}}_{\mathsf{id}}{}'(5,\Lambda) = \{(\texttt{fn x => x}^1,\mathsf{ce}_0)\}$$
$$\widehat{\mathsf{C}}_{\mathsf{id}}{}'(7,\Lambda) = \{(\texttt{fn y => y}^6,\mathsf{ce}_0)\} \qquad \widehat{\mathsf{C}}_{\mathsf{id}}{}'(8,\Lambda) = \{(\texttt{fn y => y}^6,\mathsf{ce}_0)\}$$
$$\widehat{\mathsf{C}}_{\mathsf{id}}{}'(9,\Lambda) = \{(\texttt{fn y => y}^6,\mathsf{ce}_0)\}$$

$$\widehat{\rho}_{\mathsf{id}}{}'(\texttt{f},\Lambda) = \{(\texttt{fn x => x}^1,\mathsf{ce}_0)\}$$
$$\widehat{\rho}_{\mathsf{id}}{}'(\texttt{x},5) = \{(\texttt{fn x => x}^1,\mathsf{ce}_0)\} \qquad \widehat{\rho}_{\mathsf{id}}{}'(\texttt{x},8) = \{(\texttt{fn y => y}^6,\mathsf{ce}_0)\}$$

This is an acceptable analysis result:

$$(\widehat{\mathsf{C}}_{\mathsf{id}}{}', \widehat{\rho}_{\mathsf{id}}{}') \models^{\mathsf{ce}_0}_{\Lambda} (\texttt{let f = (fn x => x}^1)^2 \texttt{ in } ((\texttt{f}^3\ \texttt{f}^4)^5\ (\texttt{fn y => y}^6)^7)^8)^9$$

Complexity

Uniform k-CFA has exponential worst case complexity even when $k = 1$

Assume that the expression has size n and that it has p different variables. Then Δ has $O(n)$ elements and hence there will be $O(p \cdot n)$ different pairs (x, δ) and $O(n^2)$ different pairs (ℓ, δ). This means that $(\widehat{\mathsf{C}}, \widehat{\rho})$ can be seen as an $O(n^2)$ tuple of values from $\widehat{\mathbf{Val}}$. Since $\widehat{\mathbf{Val}}$ itself is a powerset of pairs of the form (t, ce) and there are $O(n \cdot n^p)$ such pairs it follows that $\widehat{\mathbf{Val}}$ has height $O(n \cdot n^p)$. Since $O(p) = O(n)$ we have the exponential worst case complexity.

0-CFA analysis has polynomial worst case complexity

It corresponds to letting Δ be a singleton. Repeating the above calculations we can see $(\widehat{\mathsf{C}}, \widehat{\rho})$ as an $O(p + n)$ tuple of values from $\widehat{\mathbf{Val}}$, and $\widehat{\mathbf{Val}}$ will be a lattice of height $O(n)$.

Variations (based on call-strings)

Uniform k-CFA

$$\begin{array}{llllll} ce & \in & \mathbf{CEnv} & = & \mathbf{Var} \to \Delta & \text{context environments} \\ \widehat{v} & \in & \widehat{\mathbf{Val}} & = & \mathcal{P}(\mathbf{Term} \times \mathbf{CEnv}) & \text{abstract values} \\ \widehat{\rho} & \in & \widehat{\mathbf{Env}} & = & (\mathbf{Var} \times \Delta) \to \widehat{\mathbf{Val}} & \text{abstract environments} \\ \widehat{\mathsf{C}} & \in & \widehat{\mathbf{Cache}} & = & (\mathbf{Lab} \times \Delta) \to \widehat{\mathbf{Val}} & \text{abstract caches} \end{array}$$

k-CFA

$$\widehat{\mathsf{C}} \in \widehat{\mathbf{Cache}} = (\mathbf{Lab} \times \mathbf{CEnv}) \to \widehat{\mathbf{Val}} \quad \text{abstract caches}$$

Polynomial k-CFA

$$\widehat{v} \in \widehat{\mathbf{Val}} = \mathcal{P}(\mathbf{Term} \times \Delta) \quad \text{abstract values}$$