

Lecture 02 – Structural Operational Semantics (SOS)

PROGRAM ANALYSIS & SYNTHESIS

Eran Yahav

Previously...

- static analysis
 - over-approximation of program behavior
- abstract interpretation
 - abstraction, transformers, fixed-point computation
- examples of program analysis
 - parity abstraction
 - heap abstraction (shape analysis)
 - numerical abstractions
- examples of program synthesis
 - optimized search in program space
 - program repair as a game
 - SKETCH
 - SMARTEdit

Today

- What is the meaning of a program?
- semantics
- structural operational semantics
- intro to abstract interpretation

What is the “meaning” of a program?

```
int foo(int a ) {  
    if( 0 < a < 5) c = 42 else c = 73;  
    return c;  
}
```

```
int a() { printf("a"); return 1; }  
int b() { printf("b"); return 2; }  
int c() { printf("c"); return 3; }  
int sum(int x, int y, int z) { return x+y+z; }  
void bar() {  
    printf("%d, sum(a(),b(),c());  
}
```

Semantics

"mathematical models of and methods for describing and reasoning about the behavior of programs"

Why Formal Semantics?

- implementation-independent definition of a programming language
- automatically generating interpreters (and some day maybe full fledged compilers)
- verification and debugging
 - if you don't know what it does, how do you know its incorrect?

Different Approaches

- Denotational Semantics
 - define an input/output relation that assigns meaning to each construct (denotation)
- Structural Operational Semantics
 - define a transition system, transition relation describes evaluation steps of a program
- Axiomatic Semantics
 - define the effect of each construct on logical statements about program state (assertions)

Denotational Semantics

```
int double1(int x) {  
  int t = 0;  
  t = t + x;  
  t = t + x;  
  return t;  
}
```

$$\lambda x. 2 * x$$

```
int double2(int x) {  
  int t = 2 * x;  
  return t;  
}
```

$$\lambda x. 2 * x$$

Operational Semantics

```
int double1(int x) {  
    int t = 0;            $x \mapsto 2$   
    t = t + x;           $[t \mapsto 0, x \mapsto 2]$   
    t = t + x;           $[t \mapsto 2, x \mapsto 2]$   
    return t;            $[t \mapsto 4, x \mapsto 2]$   
    }                    $[t \mapsto 4, x \mapsto 2]$ 
```

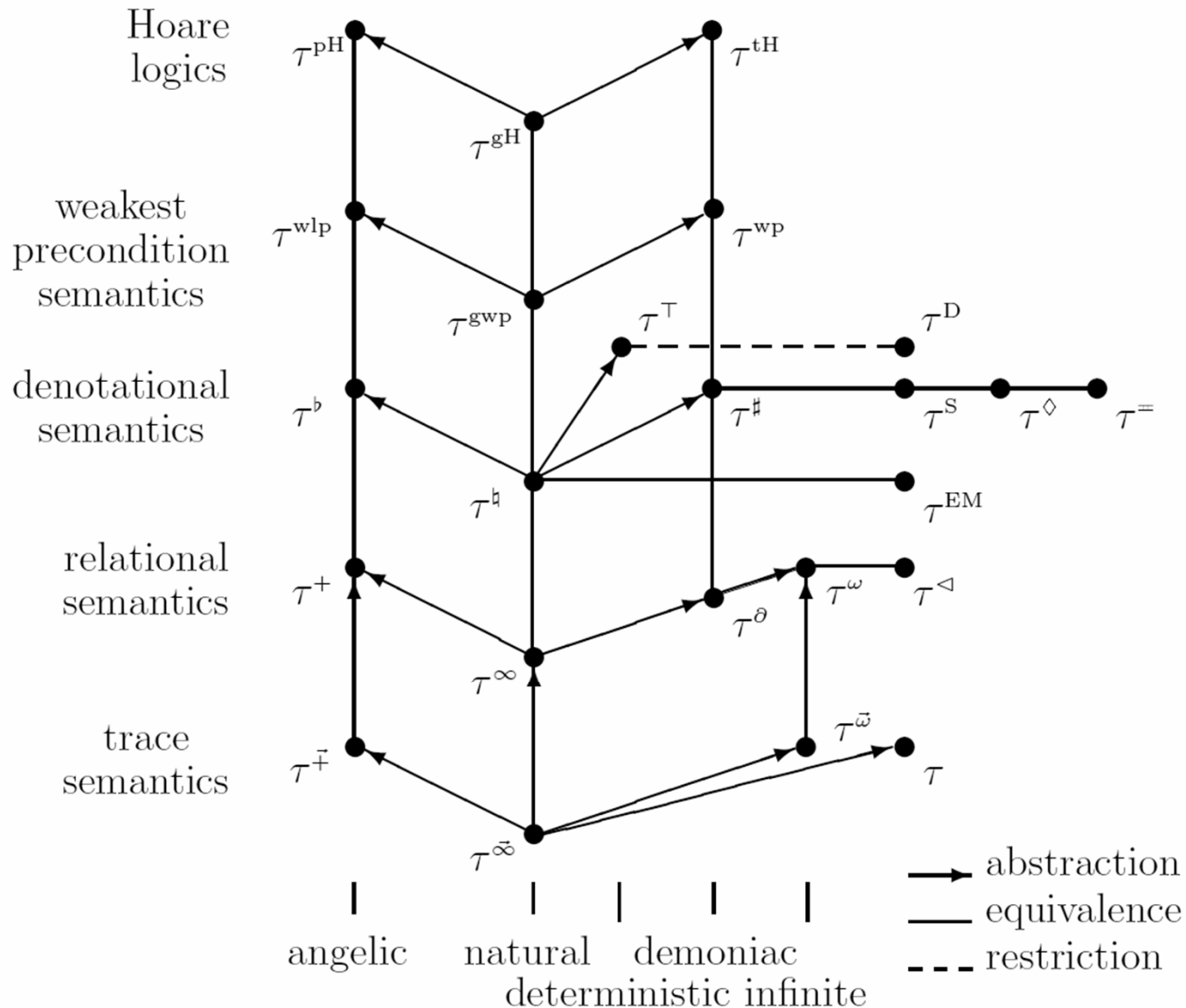
```
int double2(int x) {  
    int t = 2*x;         $[t \mapsto 4, x \mapsto 2]$   
    return t;  
    }
```

Axiomatic Semantics

```
int double1(int x) {  
  {  $x = x_0$  }  
  int t = 0;  
  {  $x = x_0 \wedge t = 0$  }  
  t = t + x;  
  {  $x = x_0 \wedge t = x_0$  }  
  t = t + x;  
  {  $x = x_0 \wedge t = 2 * x_0$  }  
  return t;  
}
```

```
int double2(int x) {  
  {  $x = x_0$  }  
  int t = 2 * x;  
  {  $x = x_0 \wedge t = 2 * x_0$  }  
  return t;  
}
```

Relating Semantics



What is the “meaning” of this program?

```
[y := x]1;  
[z := 1]2;  
while [y > 0]3 (  
  [z := z * y]4;  
  [y := y - 1]5;  
)  
[y := 0]6
```

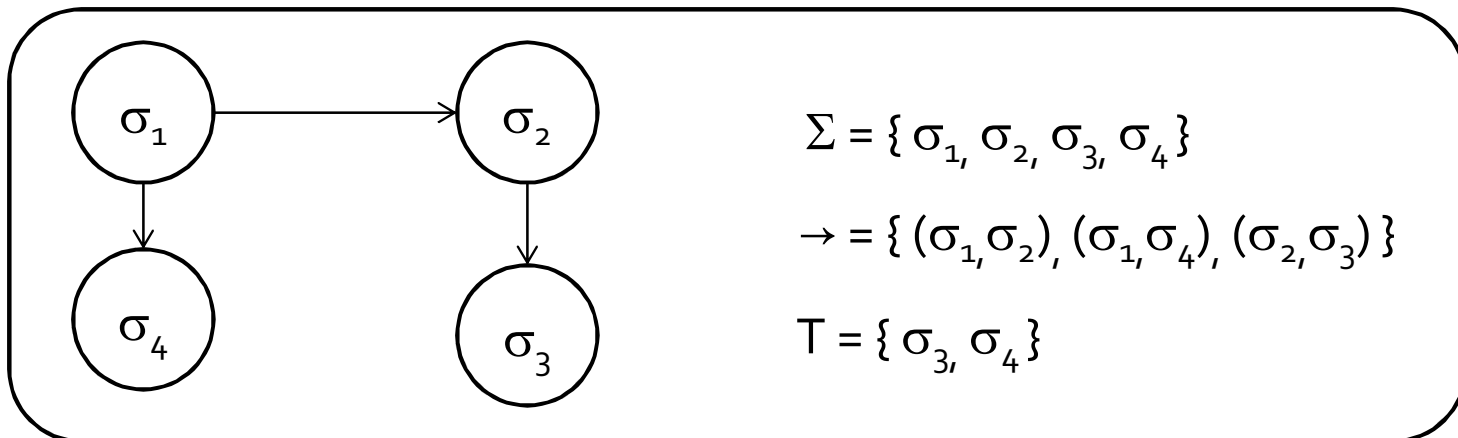
what is the “meaning” of an arithmetic expression?

- $z * y$
- $y - 1$

- First: syntax of simple arithmetic expressions
- For now, assume no variables
- $a ::= n$
 - | $a_1 + a_2$
 - | $a_1 - a_2$
 - | $a_1 * a_2$
 - | (a_1)

Structural Operational Semantics

- Defines a transition system $(\Sigma, \rightarrow, \mathsf{T})$
 - configurations Σ : snapshots of current state of the program
 - transitions $\rightarrow \subseteq \Sigma \times \Sigma$: steps between configurations
 - final configurations $\mathsf{T} \subseteq \Sigma$



Structural Operational Semantics

Useful Notations

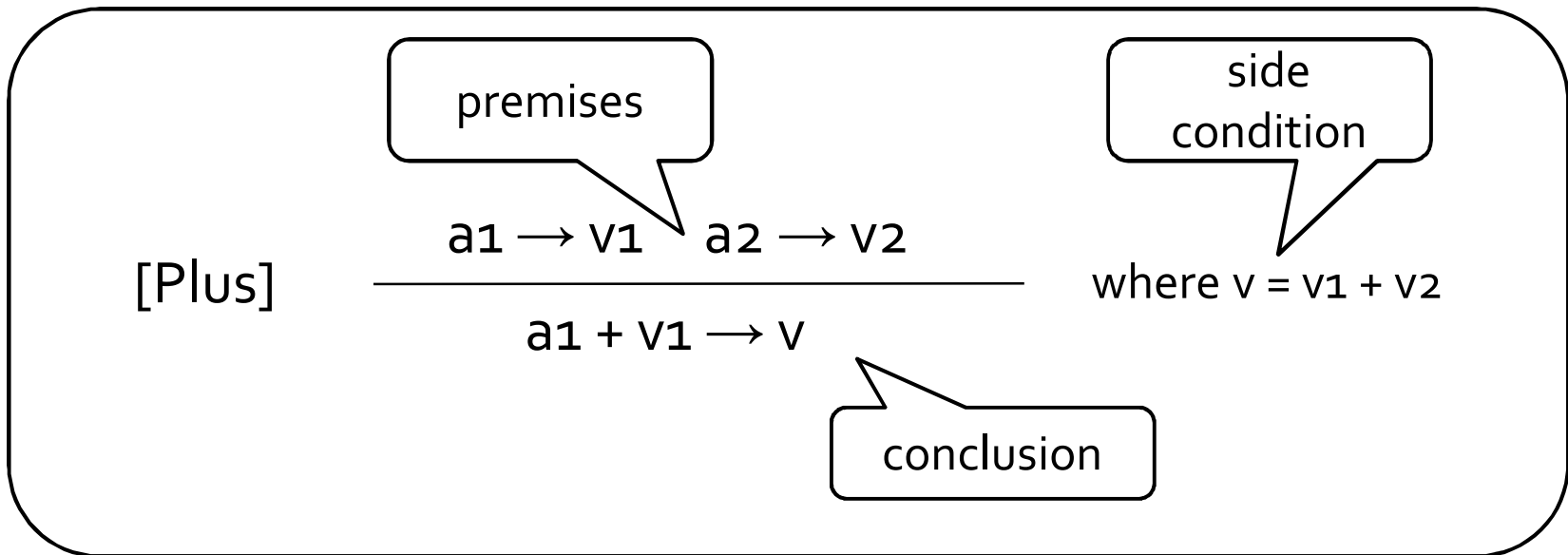
- We write $\sigma \rightarrow \sigma'$ when $(\sigma, \sigma') \in \rightarrow$
- \rightarrow^* denotes the reflexive transitive closure of the relation \rightarrow
 - $\sigma \rightarrow^* \sigma'$ when there is a sequence $\sigma = \sigma_0 \rightarrow \sigma_1 \rightarrow \dots \rightarrow \sigma_n = \sigma'$ for some $n \geq 0$

Big-step vs. Small-step

- Big-step
 - $\sigma \rightarrow \sigma'$ describes the entire computation
 - σ' is always a terminal configuration
- Small-step
 - $\sigma \Rightarrow \sigma'$ describes a single step of a larger computation
 - σ' need not be a terminal configuration
- pros/cons to each
- big-step hard in the presence of concurrency

Simple Arithmetic Expressions

(big step semantics)



$a \rightarrow v$ means “expression a evaluates to the value v ”

$a \in AExp, v \in Z$

Simple Arithmetic Expressions

(big step semantics)

$$\text{[Plus]} \quad \frac{a1 \rightarrow v1 \quad a2 \rightarrow v2}{a1 + v1 \rightarrow v} \quad \text{where } v = v1 + v2$$

$$\text{[Minus]} \quad \frac{a1 \rightarrow v1 \quad a2 \rightarrow v2}{a1 - v1 \rightarrow v} \quad \text{where } v = v1 - v2$$

$$\text{[Mult]} \quad \frac{a1 \rightarrow v1 \quad a2 \rightarrow v2}{a1 * v1 \rightarrow v} \quad \text{where } v = v1 * v2$$

$$\text{[Paren]} \quad \frac{a1 \rightarrow v1}{(a1) \rightarrow v}$$

$$\text{[Num]} \quad n \rightarrow v \text{ if } N[[n]] = v$$

Simple Arithmetic Expressions

(big step semantics)

- Transition system (Σ, \rightarrow, T)
 - configurations $\Sigma = AExp \cup Z$
 - transitions $\rightarrow \subseteq \Sigma \times \Sigma$: defined by the rules on the previous slide
 - final configurations $T = Z$
- Transitions are syntax directed

Derivation Tree

- show that $(2+4)*(4+3) \rightarrow 42$

$$2 \rightarrow 2 \quad 4 \rightarrow 4$$

$$4 \rightarrow 4 \quad 3 \rightarrow 3$$

$$\frac{2 \rightarrow 2 \quad 4 \rightarrow 4}{2 + 4 \rightarrow 6}$$

$$\frac{4 \rightarrow 4 \quad 3 \rightarrow 3}{4 + 3 \rightarrow 7}$$

$$\frac{2 + 4 \rightarrow 6}{(2 + 4) \rightarrow 6}$$

$$\frac{4 + 3 \rightarrow 7}{(4 + 3) \rightarrow 7}$$

$$\frac{(2+4) \rightarrow 6 \quad (4+3) \rightarrow 7}{(2+4)*(4+3) \rightarrow 42}$$

Derivation Tree

$$\begin{array}{r} \frac{2 \rightarrow 2 \quad 4 \rightarrow 4}{2 + 4 \rightarrow 6} \quad \frac{4 \rightarrow 4 \quad 3 \rightarrow 3}{4 + 3 \rightarrow 7} \\ \frac{(2 + 4) \rightarrow 6 \quad (4 + 3) \rightarrow 7}{(2+4)*(4 + 3) \rightarrow 42} \end{array}$$

Simple Arithmetic Expressions

(small step semantics)

$$\text{[Plus-1]} \quad \frac{a1 \Rightarrow a1'}{a1 + a2 \Rightarrow a1' + a2}$$

$$\text{[Plus-2]} \quad \frac{a2 \Rightarrow a2'}{a1 + a2 \Rightarrow a1 + a2'}$$

$$\text{[Plus-3]} \quad v1 + v2 \Rightarrow v \text{ where } v = v1 + v2$$

- intermediate values
- intermediate configurations

Determinacy

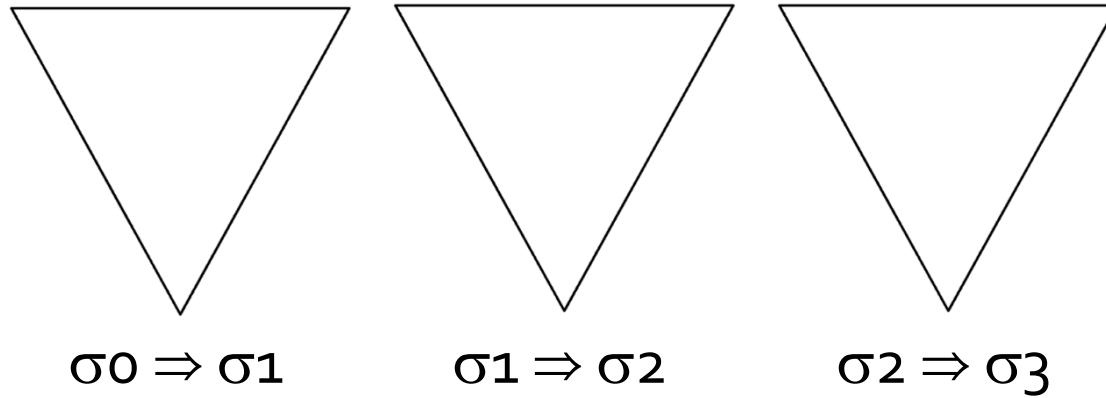
- We would like the big-step semantics of arithmetic expressions to be deterministic
 - $a \rightarrow v_1$ and $a \rightarrow v_2$ then $v_1 = v_2$
- induction on the height of the derivation tree (“transition induction”)
 - show that rules for roots are deterministic
 - show that transition rules are deterministic

Determinacy

- Is the small-step semantics of arithmetic expressions deterministic?
- we want
 - if $a \Rightarrow v_1$ and $a \Rightarrow v_2$ then $v_1 = v_2$
- but we have, for example
 - $2 + 3 \Rightarrow 2 + 3$
 - $2 + 3 \Rightarrow \mathbf{2} + 3$

Small Step and Big Step

small step



big step

Diagram illustrating a big step, represented by a large downward-pointing triangle. The step is labeled below the triangle:

$$\sigma_0 \rightarrow \sigma_3$$

The WHILE Language: Syntax

Var	set of variables	$A \in AExp$	arithmetic expressions
Lab	set of labels	$B \in BExp$	boolean expressions
Op_a	arithmetic operators	$S \in Stmt$	statements
Op_b	boolean operators		
Op_r	relational operators		

$a ::= x \mid n \mid a_1 op_a a_2$

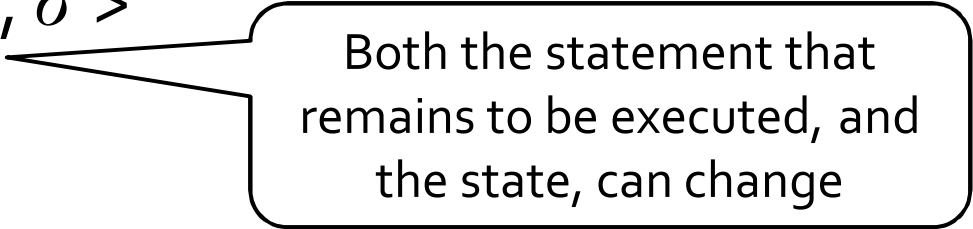
$b ::= true \mid false \mid not\ b \mid b_1 op_b b_2 \mid a_1 op_r a_2$

$S ::= [x := a]^{lab}$
| $[skip]^{lab}$
| $S_1; S_2$
| $if\ [b]^{lab}\ then\ S_1\ else\ S_2$
| $while\ [b]^{lab}\ do\ S$

(We are going to abuse syntax later for readability)

The WHILE Language: Structural Operational Semantics

- $\sigma \in \text{State} = \text{Var} \rightarrow Z$
- Configuration:
 - $\langle S, \sigma \rangle$
 - σ for terminal configuration
- Transitions:
 - $\langle S, \sigma \rangle \Rightarrow \langle S', \sigma' \rangle$
 - $\langle S, \sigma \rangle \Rightarrow \sigma'$



Both the statement that remains to be executed, and the state, can change

The WHILE Language: Structural Operational Semantics

- Transition system (Σ, \rightarrow, T)
 - configurations
 $\Sigma = (\text{Stmt} \times \text{State}) \cup \text{State}$
 - transitions $\rightarrow \subseteq \Sigma \times \Sigma$
 - final configurations $T = \text{State}$

Arithmetic Expressions

$A: AExp \rightarrow (State \rightarrow Z)$

$$A[x]\sigma = \sigma(x)$$

$$A[n]\sigma = N[n]$$

$$A[a1 \text{ op } a2]\sigma = A[a1]\sigma \text{ op } A[a2]\sigma$$

Boolean Expressions

$B: \text{BExp} \rightarrow (\text{State} \rightarrow \{\text{true}, \text{false}\})$

$$B[\text{not } b]\sigma = \neg B[b]\sigma$$

$$B[b_1 \text{ op}_b b_2]\sigma = B[b_1]\sigma \text{ op}_b B[b_2]\sigma$$

$$B[a_1 \text{ op}_r a_2]\sigma = A[a_1]\sigma \text{ op}_r A[a_2]\sigma$$

The WHILE Language: Structural Operational Semantics

$$[\text{ass}] \quad \langle [x := a]^{\text{lab}}, \sigma \rangle \Rightarrow \sigma[x \mapsto A[[a]]\sigma]$$

$$[\text{skip}] \quad \langle [\text{skip}]^{\text{lab}}, \sigma \rangle \Rightarrow \sigma$$

$$[\text{seq}_1] \quad \frac{\langle S_1, \sigma \rangle \Rightarrow \langle S'_1, \sigma' \rangle}{\langle S_1; S_2, \sigma \rangle \Rightarrow \langle S'_1; S_2, \sigma' \rangle}$$

$$[\text{seq}_2] \quad \frac{\langle S_1, \sigma \rangle \Rightarrow \sigma'}{\langle S_1; S_2, \sigma \rangle \Rightarrow \langle S_2, \sigma' \rangle}$$

The WHILE Language: Structural Operational Semantics

[if ₁]	$\langle \text{if } [b]^{\text{lab}} \text{ then } S_1 \text{ else } S_2, \sigma \rangle \Rightarrow \langle S_1, \sigma \rangle$	if $B[[b]] \sigma = \text{true}$
[if ₂]	$\langle \text{if } [b]^{\text{lab}} \text{ then } S_1 \text{ else } S_2, \sigma \rangle \Rightarrow \langle S_2, \sigma \rangle$	if $B[[b]] \sigma = \text{false}$
[wh ₁]	$\langle \text{while } [b]^{\text{lab}} \text{ do } S, \sigma \rangle \Rightarrow \langle (S; \text{while } [b]^{\text{lab}} \text{ do } S), \sigma \rangle$	if $B[[b]] \sigma = \text{true}$
[wh ₂]	$\langle \text{while } [b]^{\text{lab}} \text{ do } S, \sigma \rangle \Rightarrow \sigma$	if $B[[b]] \sigma = \text{false}$

(Table 2.6 from PPA)

Derivation Sequences

- Finite derivation sequence
 - A sequence $\langle S_0, \sigma_0 \rangle \dots \sigma_n$
 - $\langle S_i, \sigma_i \rangle \Rightarrow \langle S_{i+1}, \sigma_{i+1} \rangle$
 - σ_n terminal configuration
- Infinite derivation sequence
 - A sequence $\langle S_0, \sigma_0 \rangle \dots$
 - $\langle S_i, \sigma_i \rangle \Rightarrow \langle S_{i+1}, \sigma_{i+1} \rangle$

Termination in small-step semantics

```
while [o = o]1 (  
  [skip]2;  
)
```

$\langle \text{while } [o = o]^1 ([\text{skip}]^2), \sigma \rangle \Rightarrow$

$\langle [\text{skip}]^2; \text{while } [o = o]^1 ([\text{skip}]^2), \sigma \rangle \Rightarrow$

$\langle \text{while } [o = o]^1 ([\text{skip}]^2), \sigma \rangle \Rightarrow$

$\langle [\text{skip}]^2; \text{while } [o = o]^1 ([\text{skip}]^2), \sigma \rangle \Rightarrow \dots$

Termination in small-step semantics

- We say that S terminates from a start state σ when there exists a state σ' such that $\langle S, \sigma \rangle \Rightarrow^* \sigma'$

Termination in big-step semantics

```
while [o = o]1 (  
  [skip]2;  
)
```

- what would be the transition in the big-step semantics for this example?

Semantic Equivalence

- formal semantics enables us to reason about programs and their equivalence
- S_1 and S_2 are semantically equivalent when
 - for all σ and σ' $\langle S_1, \sigma \rangle \Rightarrow^* \sigma'$ iff $\langle S_2, \sigma \rangle \Rightarrow^* \sigma'$
- We write $S_1 \sim S_2$ when S_1 and S_2 are semantically equivalent

Abnormal Termination

- add a statement **abort** for aborting execution
- in the big-step semantics
 - while (o=o) skip; ~ abort
- big-step semantics does not distinguish between abnormal termination and infinite-loops
- in the small-step semantics
 - while (o=o) skip; ~ abort
- but we can distinguish the cases if we look at the transitions
 - $\langle \text{abort}, \sigma \rangle \Rightarrow \text{abort}, \sigma$
 - infinite trace of skips

Nondeterminism

big-step semantics

- new language construct $s_1 \text{ OR } s_2$

$$\text{[OR}_1\text{-BSS]} \quad \frac{\langle S_1, \sigma \rangle \rightarrow \sigma'}{\langle S_1 \text{ OR } S_2, \sigma \rangle \rightarrow \sigma'}$$

$$\text{[OR}_2\text{-BSS]} \quad \frac{\langle S_2, \sigma \rangle \rightarrow \sigma'}{\langle S_1 \text{ OR } S_2, \sigma \rangle \rightarrow \sigma'}$$

Nondeterminism

small-step semantics

$$[\text{OR}_1\text{-SSS}] \quad \langle S_1 \text{ OR } S_2, \sigma \rangle \Rightarrow \langle S_1, \sigma \rangle$$

$$[\text{OR}_1\text{-SSS}] \quad \langle S_1 \text{ OR } S_2, \sigma \rangle \Rightarrow \langle S_2, \sigma \rangle$$

Nondeterminism

- $(x = 1)$ OR $\text{while}(o=o)$ skip;
- big-step semantics suppresses infinite loops
- small step semantics has the infinite sequence created by picking the while

$\langle (x = 1) \text{ OR } \text{while}(o=o) \text{ skip};, \sigma \rangle \Rightarrow \langle \text{while}(o=o) \text{ skip};, \sigma \rangle \Rightarrow \dots$

What is the “meaning” of this program?

```
[y := x]1;  
[z := 1]2;  
while [y > 0]3 (  
  [z := z * y]4;  
  [y := y - 1]5;  
)  
[y := 0]6
```

now we can answer this question using derivation sequences

Example of Derivation Sequence

```
[y := x]1;  
[z := 1]2;  
while [y > 0]3 (  
  [z := z * y]4;  
  [y := y - 1]5;  
  )  
[y := 0]6
```

$\langle [y := x]^1; [z := 1]^2; \text{while } [y > 0]^3 ([z := z * y]^4; [y := y - 1]^5); [y := 0]^6, \{ x \mapsto 42, y \mapsto 0, z \mapsto 0 \} \rangle \Rightarrow$

$\langle [z := 1]^2; \text{while } [y > 0]^3 ([z := z * y]^4; [y := y - 1]^5); [y := 0]^6, \{ x \mapsto 42, y \mapsto 42, z \mapsto 0 \} \rangle \Rightarrow$

$\langle \text{while } [y > 0]^3 ([z := z * y]^4; [y := y - 1]^5); [y := 0]^6, \{ x \mapsto 42, y \mapsto 42, z \mapsto 1 \} \rangle \Rightarrow$

$\langle ([z := z * y]^4; [y := y - 1]^5); \text{while } [y > 0]^3 ([z := z * y]^4; [y := y - 1]^5); [y := 0]^6, \{ x \mapsto 42, y \mapsto 42, z \mapsto 1 \} \rangle \dots$

Traces

$\langle [y := x]^1; [z := 1]^2; \text{while } [y > 0]^3 ([z := z * y]^4; [y := y - 1]^5); [y := 0]^6, \{ x \mapsto 42, y \mapsto 0, z \mapsto 0 \} \rangle \Rightarrow$

$\langle [z := 1]^2; \text{while } [y > 0]^3 ([z := z * y]^4; [y := y - 1]^5); [y := 0]^6, \{ x \mapsto 42, y \mapsto 42, z \mapsto 0 \} \rangle \Rightarrow$

$\langle \text{while } [y > 0]^3 ([z := z * y]^4; [y := y - 1]^5); [y := 0]^6, \{ x \mapsto 42, y \mapsto 42, z \mapsto 1 \} \rangle \Rightarrow$

$\langle ([z := z * y]^4; [y := y - 1]^5); \text{while } [y > 0]^3 ([z := z * y]^4; [y := y - 1]^5); [y := 0]^6, \{ x \mapsto 42, y \mapsto 42, z \mapsto 1 \} \rangle \dots$

$\langle [y := x]^1; [z := 1]^2; \text{while } [y > 0]^3 ([z := z * y]^4; [y := y - 1]^5); [y := 0]^6, \{ x \mapsto 42, y \mapsto 0, z \mapsto 0 \} \rangle \Rightarrow$ $[y := x]^1$

$\langle [z := 1]^2; \text{while } [y > 0]^3 ([z := z * y]^4; [y := y - 1]^5); [y := 0]^6, \{ x \mapsto 42, y \mapsto 42, z \mapsto 0 \} \rangle \Rightarrow$ $[z := 1]^2$

$\langle \text{while } [y > 0]^3 ([z := z * y]^4; [y := y - 1]^5); [y := 0]^6, \{ x \mapsto 42, y \mapsto 42, z \mapsto 1 \} \rangle \Rightarrow$ $[y > 0]^3$

$\langle ([z := z * y]^4; [y := y - 1]^5); \text{while } [y > 0]^3 ([z := z * y]^4; [y := y - 1]^5); [y := 0]^6, \{ x \mapsto 42, y \mapsto 42, z \mapsto 1 \} \rangle \dots$

Traces

$$\begin{aligned}
 & \langle [y := x]^1; [z := 1]^2; \text{while } [y > 0]^3 ([z := z * y]^4; [y := y - 1]^5); [y := 0]^6, \{x \mapsto 42, y \mapsto 0, z \mapsto 0\} \rangle \Rightarrow [y := x]^1 \\
 & \langle [z := 1]^2; \text{while } [y > 0]^3 ([z := z * y]^4; [y := y - 1]^5); [y := 0]^6, \{x \mapsto 42, y \mapsto 42, z \mapsto 0\} \rangle \Rightarrow [z := 1]^2 \\
 & \langle \text{while } [y > 0]^3 ([z := z * y]^4; [y := y - 1]^5); [y := 0]^6, \{x \mapsto 42, y \mapsto 42, z \mapsto 1\} \rangle \Rightarrow [y > 0]^3 \\
 & \langle ([z := z * y]^4; [y := y - 1]^5); \text{while } [y > 0]^3 ([z := z * y]^4; [y := y - 1]^5); [y := 0]^6, \{x \mapsto 42, y \mapsto 42, z \mapsto 1\} \rangle \dots \\
 \\
 & \langle 1, \{x \mapsto 42, y \mapsto 0, z \mapsto 0\} \rangle \Rightarrow [y := x]^1 \\
 & \langle 2, \{x \mapsto 42, y \mapsto 42, z \mapsto 0\} \rangle \Rightarrow [z := 1]^2 \\
 & \langle 3, \{x \mapsto 42, y \mapsto 42, z \mapsto 1\} \rangle \Rightarrow [y > 0]^3 \\
 & \langle 4, \{x \mapsto 42, y \mapsto 42, z \mapsto 1\} \rangle \dots
 \end{aligned}$$

Traces

$\langle 1, \{x \mapsto 42, y \mapsto 0, z \mapsto 0\} \rangle \xRightarrow{[y := x]^1}$

$\langle 2, \{x \mapsto 42, y \mapsto 42, z \mapsto 0\} \rangle \xRightarrow{[z := 1]^2}$

$\langle 3, \{x \mapsto 42, y \mapsto 42, z \mapsto 1\} \rangle \xRightarrow{[y > 0]^3}$

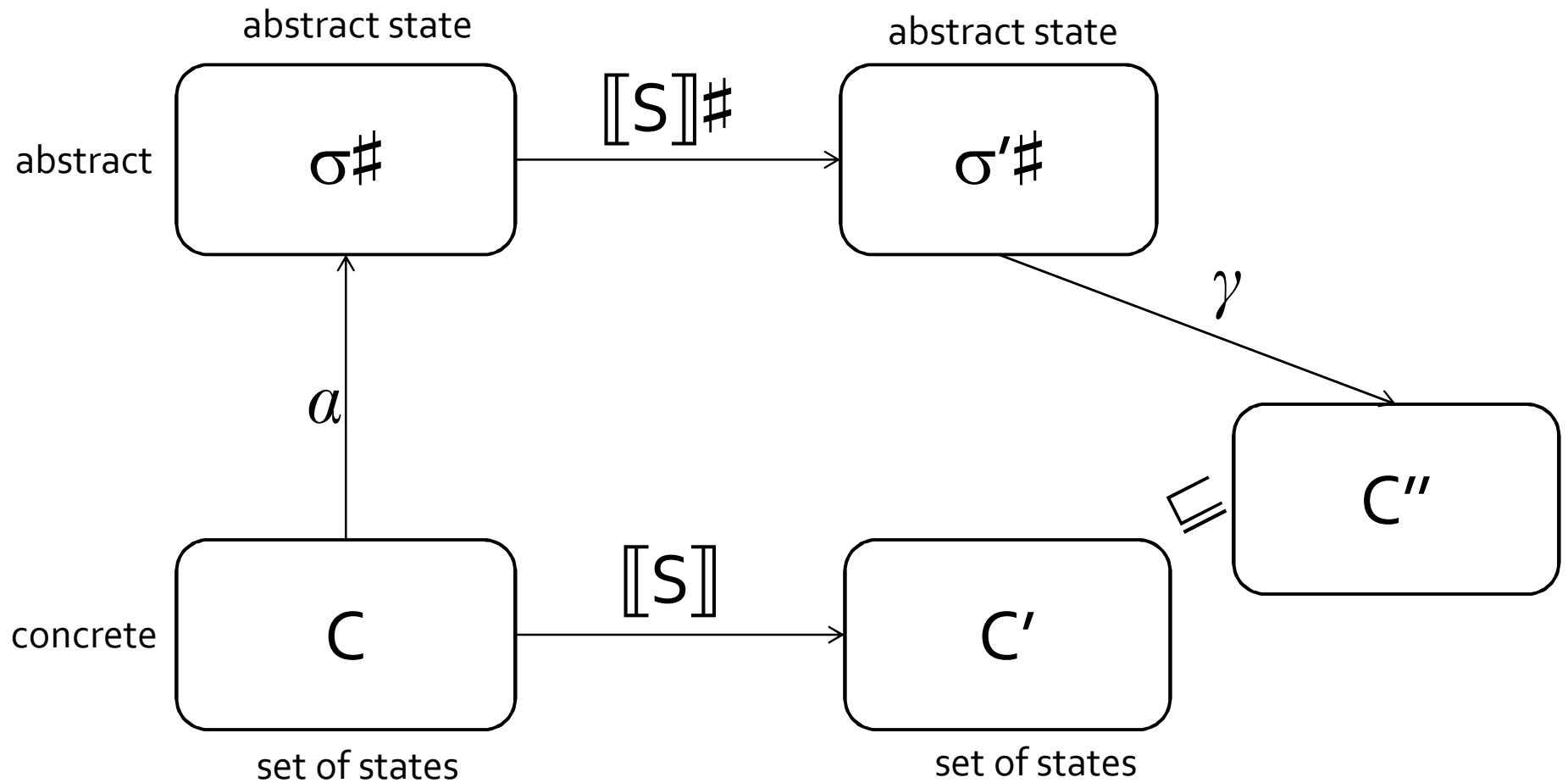
$\langle 4, \{x \mapsto 42, y \mapsto 42, z \mapsto 1\} \rangle \dots$

Trace Semantics

$$\begin{array}{l}
 [y := x]^1; \\
 [z := 1]^2; \\
 \text{while } [y > 0]^3 (\\
 [z := z * y]^4; \\
 [y := y - 1]^5; \\
) \\
 [y := 0]^6
 \end{array}
 \quad
 \left|
 \begin{array}{l}
 \begin{array}{l}
 \langle 1, \{x \mapsto 42, y \mapsto 0, z \mapsto 0\} \rangle \Rightarrow \langle 2, \{x \mapsto 42, y \mapsto 42, z \mapsto 0\} \rangle \Rightarrow \\
 \langle 3, \{x \mapsto 42, y \mapsto 42, z \mapsto 1\} \rangle \Rightarrow \langle 4, \{x \mapsto 42, y \mapsto 42, z \mapsto 1\} \rangle \dots
 \end{array} \\
 \hline
 \begin{array}{l}
 \langle 1, \{x \mapsto 73, y \mapsto 0, z \mapsto 0\} \rangle \Rightarrow \langle 2, \{x \mapsto 73, y \mapsto 73, z \mapsto 0\} \rangle \Rightarrow \\
 \langle 3, \{x \mapsto 73, y \mapsto 73, z \mapsto 1\} \rangle \Rightarrow \langle 4, \{x \mapsto 73, y \mapsto 73, z \mapsto 1\} \rangle \dots
 \end{array} \\
 \vdots
 \end{array}
 \right.$$

- In the beginning, there was the trace semantics...
- note that input (x) can be anything
- clearly, the trace semantics is not computable

Abstract Interpretation



Dataflow Analysis

	_____	$\{(x,?), (y,?), (z,?)\}$
1: $y := x;$		
	_____	$\{(x,?), (y,1), (z,?)\}$
2: $z := 1;$		
	_____	$\{(x,?), (y,1), (z,2)\}$
3: while $y > 0$ {		
	_____	$\{(x,?), (y,1), (z,2)\}$
4: $z := z * y;$		
	_____	$\{(x,?), (y,?), (z,4)\}$
5: $y := y - 1$		
	_____	$\{(x,?), (y,5), (z,4)\}$
}		
	_____	$\{(x,?), (y,1), (z,2)\}$
6: $y := 0$		
	_____	$\{(x,?), (y,?), (z,?)\}$



- Reaching Definitions

- The assignment **lab: var := exp** reaches **lab'** if there is an execution where **var** was last assigned at **lab**

(adapted from Nielson, Nielson & Hankin)

Dataflow Analysis

		{ (x,?), (y,?), (z,?) }
1: y := x;		
2: z := 1;		{ (x,?), (y,1), (z,?) }
3: while y > 0 {		{ (x,?), (y,1), (z,2), (y,5), (z,4) }
4: z := z * y;		{ (x,?), (y,1), (z,2), (y,5), (z,4) }
5: y := y - 1		{ (x,?), (y,?), (z,4), (y,5) }
}		{ (x,?), (y,5), (z,4) }
6: y := 0		{ (x,?), (y,1), (z,2), (y,5), (z,4) }
		{ (x,?), (y,6), (z,2), (z,4) }

- Reaching Definitions

- The assignment **lab: var := exp** reaches **lab'** if there is an execution where **var** was last assigned at **lab**

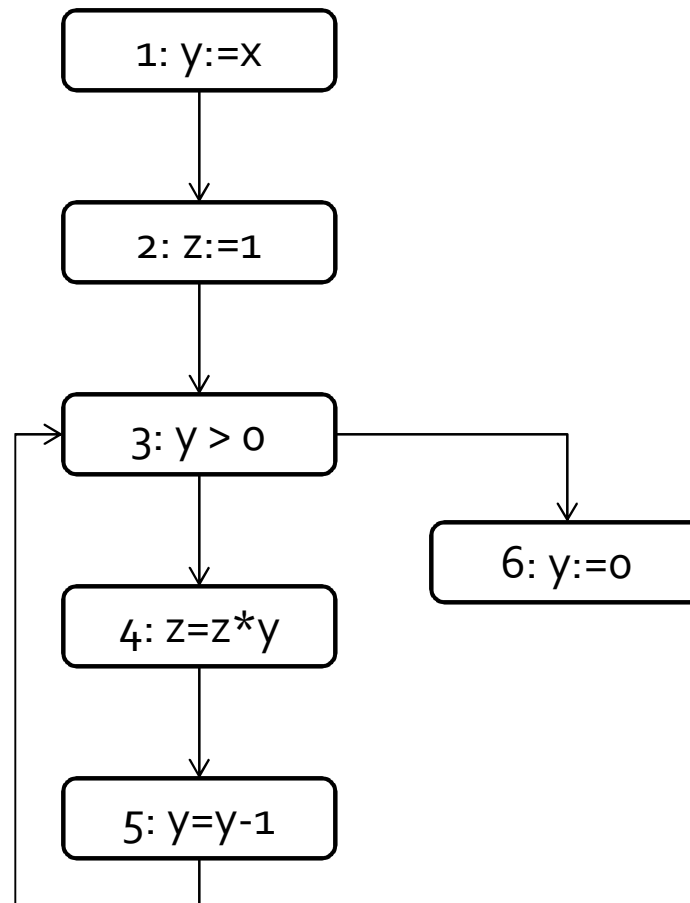
(adapted from Nielson, Nielson & Hankin)

Dataflow Analysis

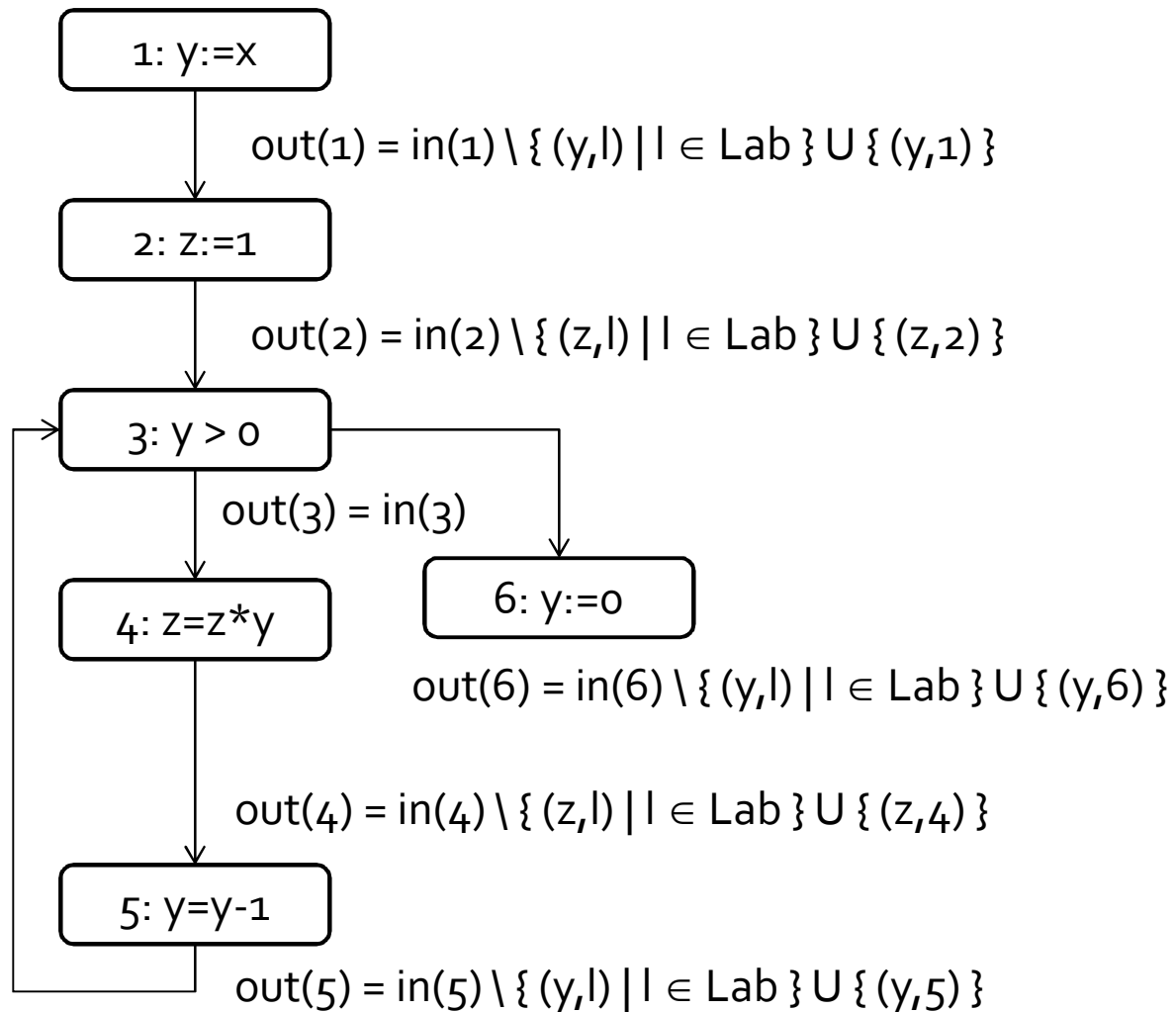
- Build control-flow graph
- Assign transfer functions
- Compute fixed point

Control-Flow Graph

```
1: y := x;  
2: z := 1;  
3: while y > 0 {  
4:  z := z * y;  
5:  y := y - 1  
}  
6: y := 0
```



Transfer Functions



$in(1) = \{(x,?), (y,?), (z,?)\}$

$in(2) = out(1)$

$in(3) = out(2) \cup out(5)$

$in(4) = out(3)$

$in(5) = out(4)$

$in(6) = out(3)$

System of Equations

$$\text{in}(1) = \{ (x,?), (y,?), (z,?) \}$$

$$\text{in}(2) = \text{out}(1)$$

$$\text{in}(3) = \text{out}(2) \cup \text{out}(5)$$

$$\text{in}(4) = \text{out}(3)$$

$$\text{in}(5) = \text{out}(4)$$

$$\text{in}(6) = \text{out}(3)$$

$$\text{out}(1) = \text{in}(1) \setminus \{ (y,l) \mid l \in \text{Lab} \} \cup \{ (y,1) \}$$

$$\text{out}(2) = \text{in}(2) \setminus \{ (z,l) \mid l \in \text{Lab} \} \cup \{ (z,2) \}$$

$$\text{out}(3) = \text{in}(3)$$

$$\text{out}(4) = \text{in}(4) \setminus \{ (z,l) \mid l \in \text{Lab} \} \cup \{ (z,4) \}$$

$$\text{out}(5) = \text{in}(5) \setminus \{ (y,l) \mid l \in \text{Lab} \} \cup \{ (y,5) \}$$

$$\text{out}(6) = \text{in}(6) \setminus \{ (y,l) \mid l \in \text{Lab} \} \cup \{ (y,6) \}$$

$$F: (\wp(\text{Var} \times \text{Lab}))^{12} \rightarrow (\wp(\text{Var} \times \text{Lab}))^{12}$$

Least Fixed Point

- We will see later why it exists
- For now, mostly informally...

$$F: (\wp(\text{Var} \times \text{Lab}))^{12} \rightarrow (\wp(\text{Var} \times \text{Lab}))^{12}$$

$$RD \sqsubseteq RD' \text{ when } \forall i: RD(i) \subseteq RD'(i)$$

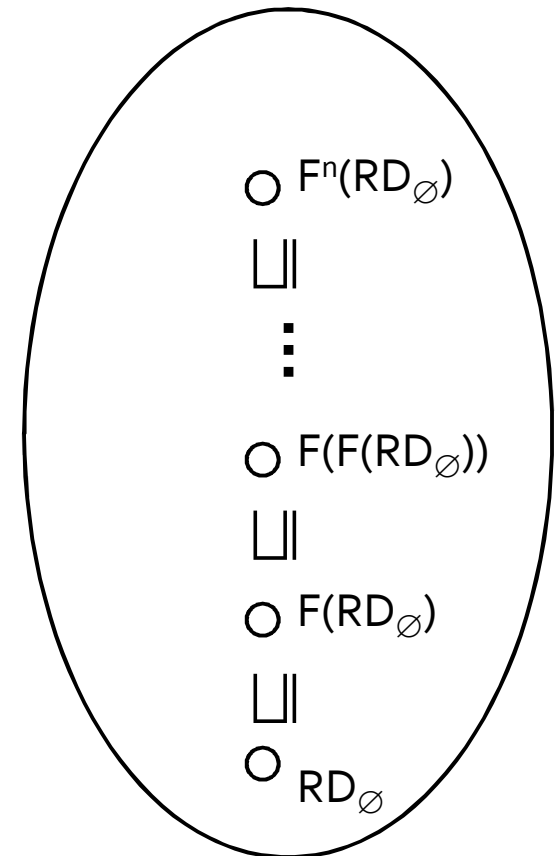
F is monotone:

$$RD \sqsubseteq RD' \text{ implies that } F(RD) \sqsubseteq F(RD')$$

$$RD_{\emptyset} = (\emptyset, \emptyset, \dots, \emptyset)$$

$$F(RD_{\emptyset}), F(F(RD_{\emptyset})), F(F(F(RD_{\emptyset}))), \dots F^n(RD_{\emptyset})$$

$$F^{n+1}(RD_{\emptyset}) = F^n(RD_{\emptyset})$$



Things that Should Trouble You

- How did we get the transfer functions?
- How do we know these transfer functions are safe (conservative)?
- How do we know that these transfer functions are optimal?

References

- “Transitions and Trees” / Huttel
- “Principles of Program Analysis” / Nielson, Nielson, and Hankin