

Computer aided verification

lecture 9

Abstract interpretation

Sławomir Lasota
University of Warsaw

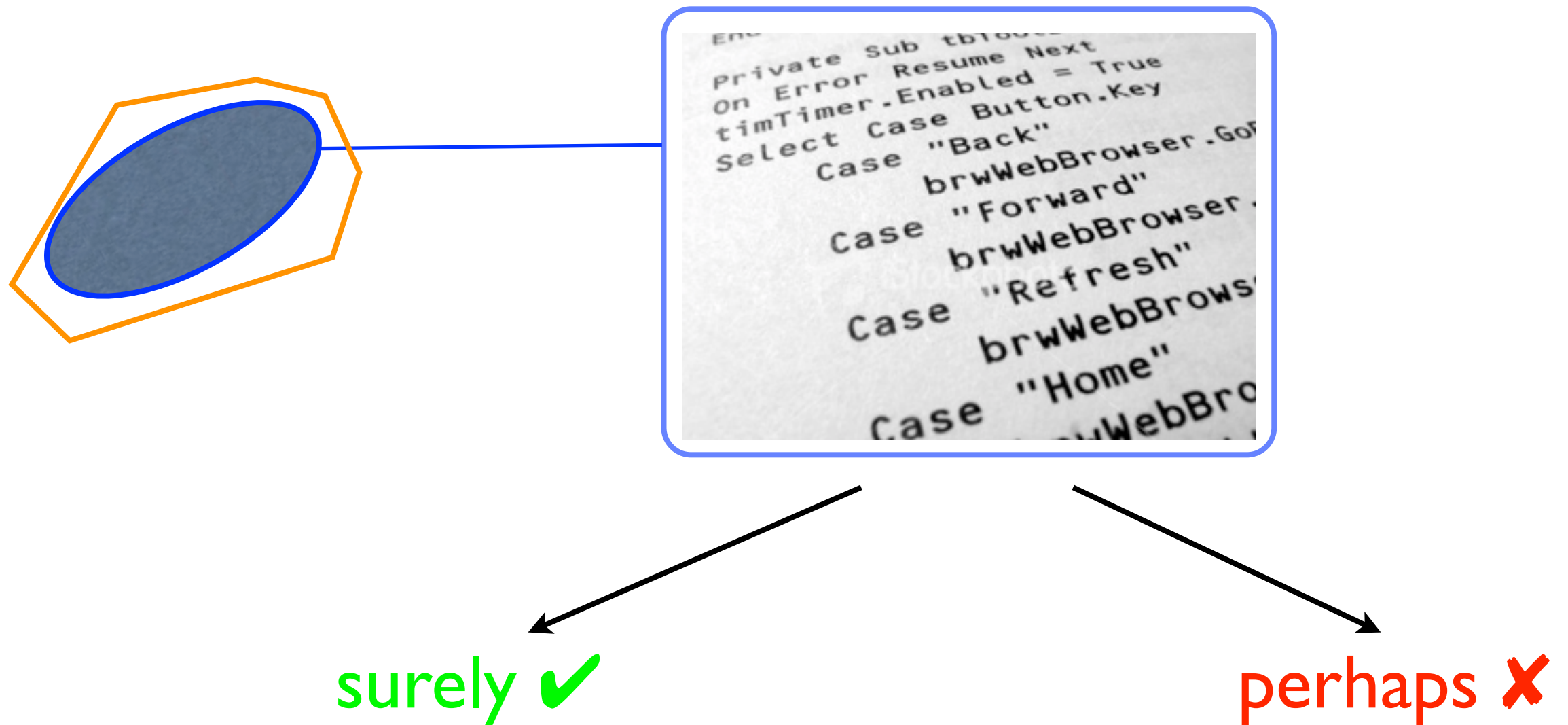
Literature

- F. Nielson, H.R. Nielson, C. Hankin, [Principles of Program Analysis](#), Springer, 2005.
- <http://www.imm.dtu.dk/~riis/PPA/slides2.pdf>
- V. D'Silva, D. Kroening, G. Weissenbacher, [A Survey of Automated Techniques for Formal Software Verification](#). IEEE Trans. on CAD of Integrated Circuits and Systems 27(7):1165-1178, 2008.

Pioniers

- P. Naur 1965
- P. Cousot, R. Cousot 1977

Approximate analysis



$$123 \cdot 457 + 76543 \stackrel{?}{=} 132654$$

$$123 \cdot 457 + 76543 \stackrel{?}{=} 132654$$

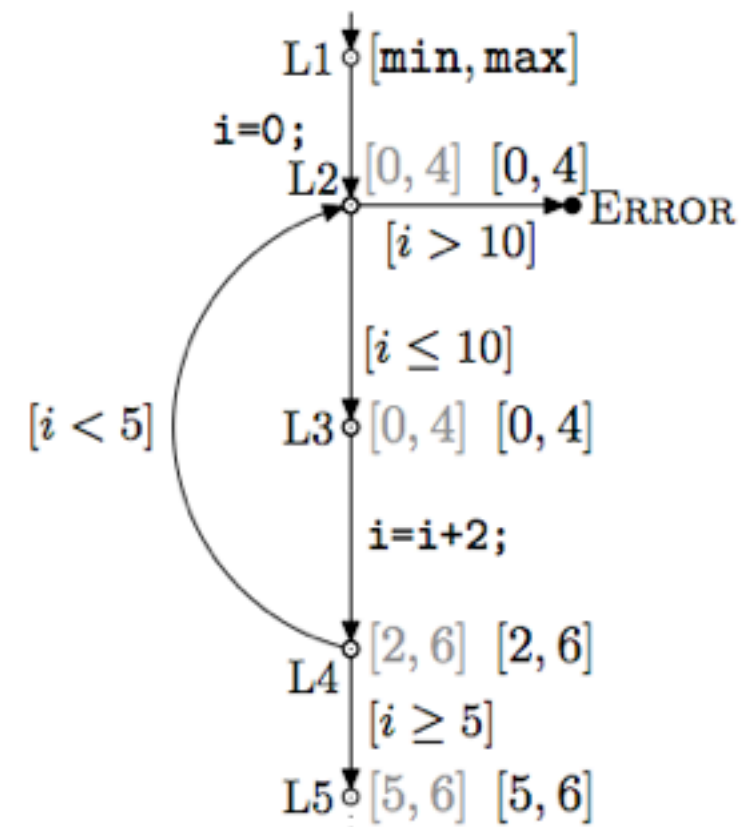
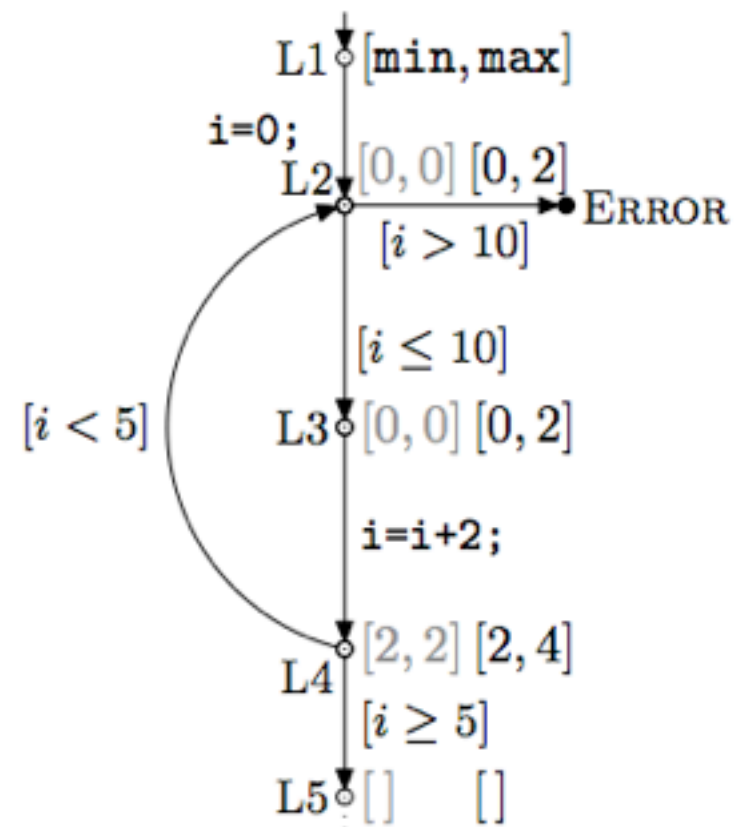
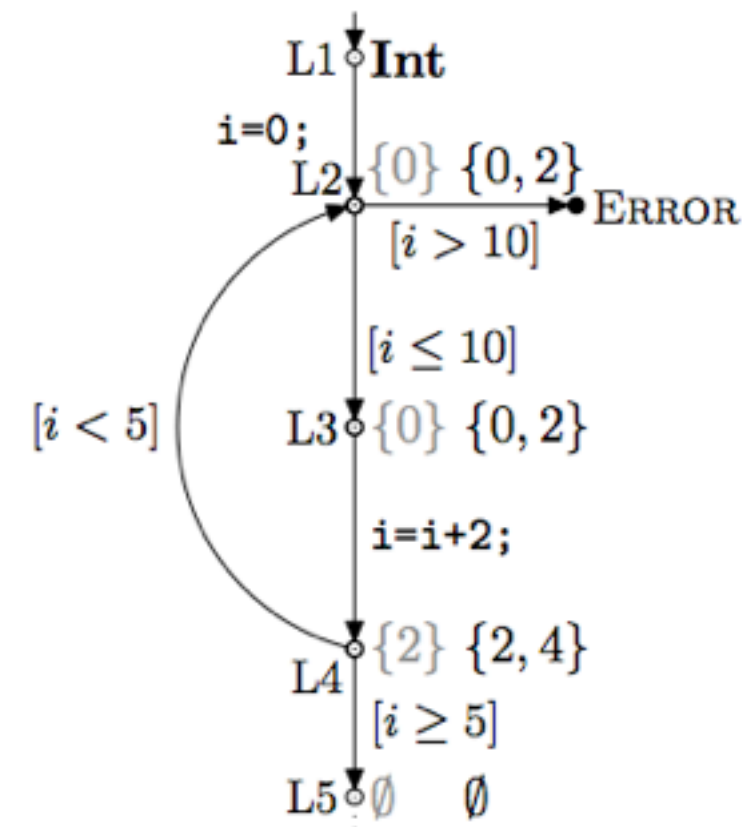
$$6 \cdot 7 + 7 \stackrel{?}{=} 3 \pmod{9}$$

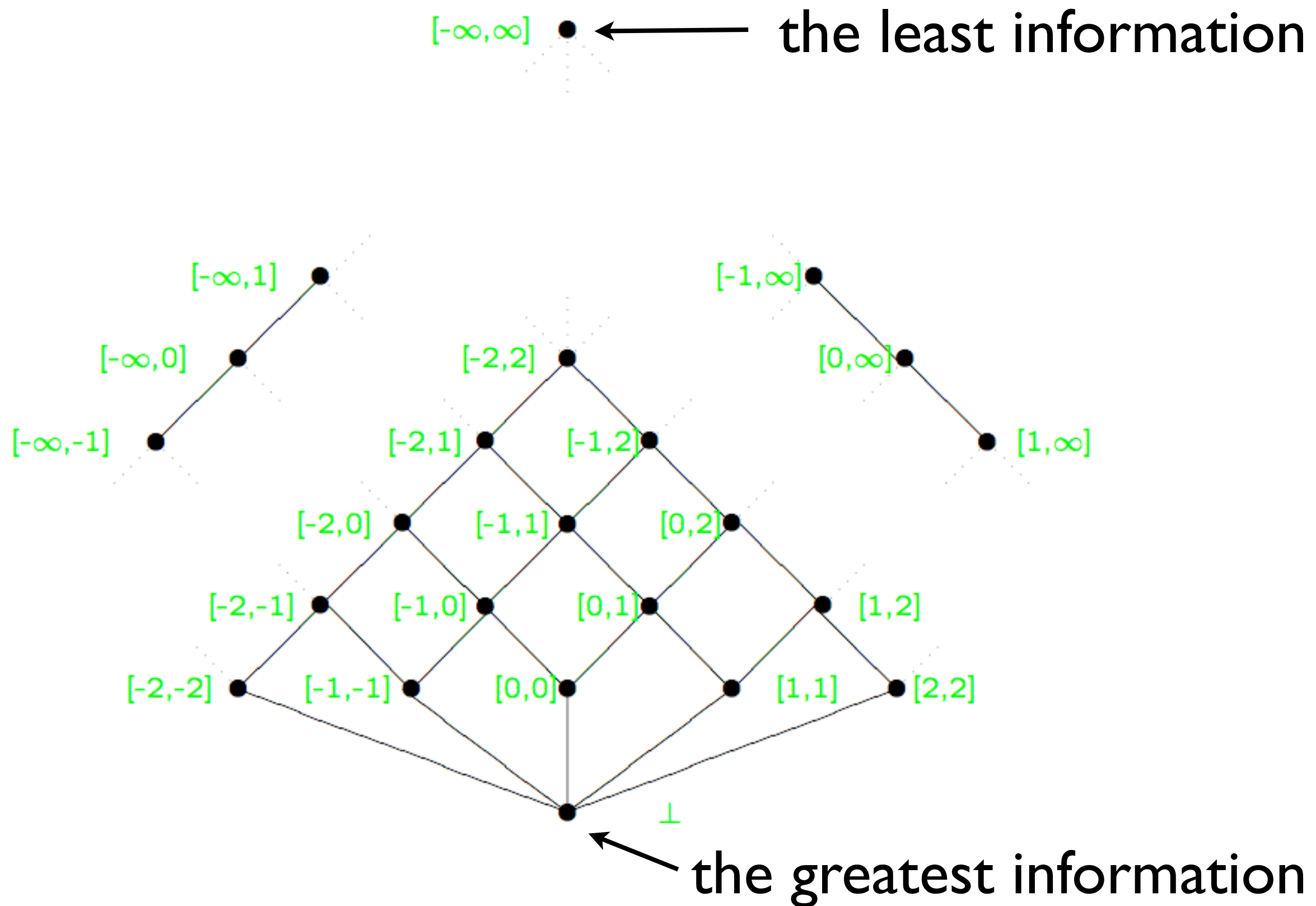
$$6 + 7 \stackrel{?}{=} 3 \pmod{9}$$

```

int i = 0;
do {
    assert(i <= 10);
    i = i+2;
} while (i < 5);

```





Approximate analysis

- static analysis
- source code is analyzed (**control-flow diagram**)
- false alarms (false positives)
- typically oriented towards specific properties
- fully automatic
- scalable
- a diagnostic information if error

Approximate analysis - methods

- data-flow analysis
- control-flow analysis
- type analysis
- WCET analysis
- ...
- abstract interpretation

Approximate analysis - applications

- code optimization, program transformations
- program verification (static analysis)
- estimation of quality of code
- abstract interpretation - systematization

Code optimization

- constants propagation (at compile time)
- copies propagation
- available expressions analysis (elimination of comp.)
- live variables analysis (elimination of dead code)
- definition-use and use-definition analysis
- strictness analysis
- array bounds analysis
- ...

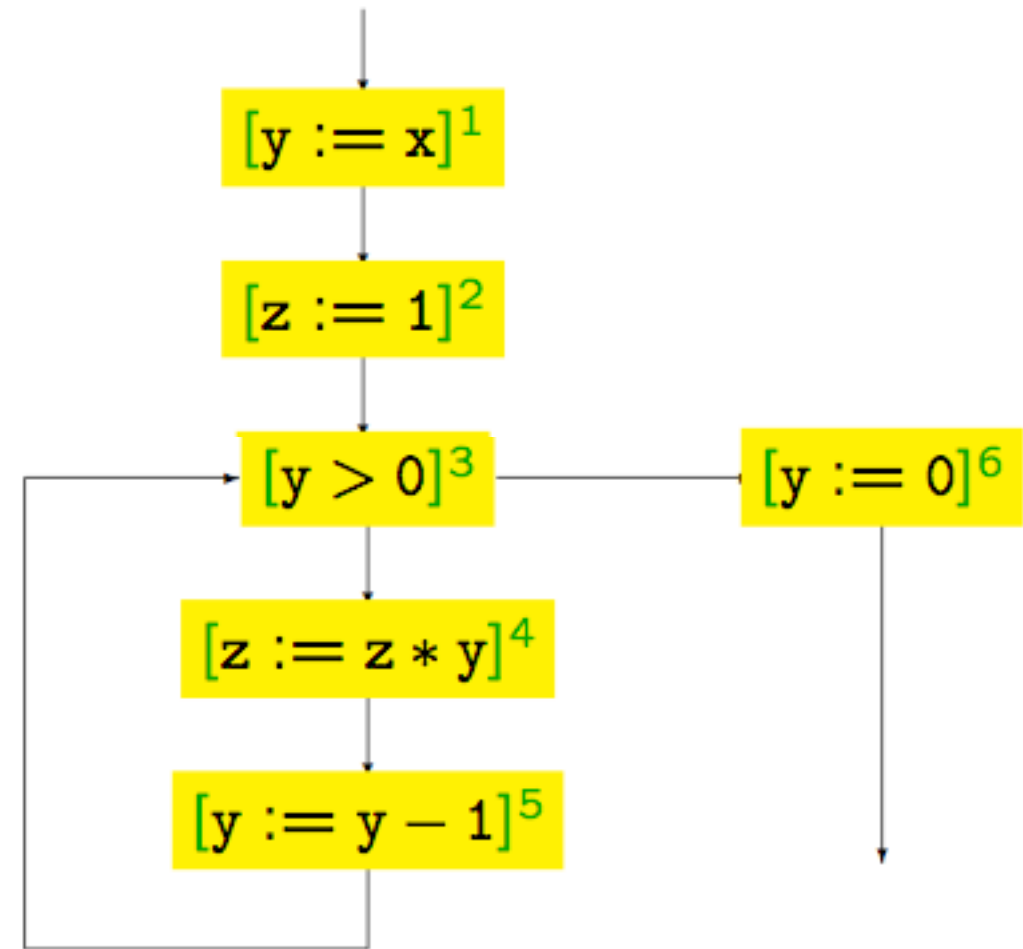
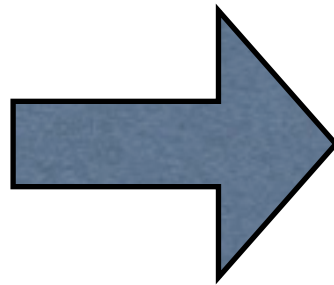
Program verification

- division by 0, exceptions
- pointers
 - NULL dereferences
 - static and dynamic data (stack and heap)
 - shape analysis
- aliasing analysis
- array bounds
- detection of invariants
- ...

Data-flow analyses

Data-flow analysis

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



(while-programs)

finite set of control locations

$$S = \{1, \dots, n\}, \quad \rightsquigarrow \subseteq S \times S$$

$$\text{init}(S) \subseteq S$$

$$\text{State} = S \times \text{Store}$$

$$\text{Store} = \text{Var} \rightarrow \text{Val}$$

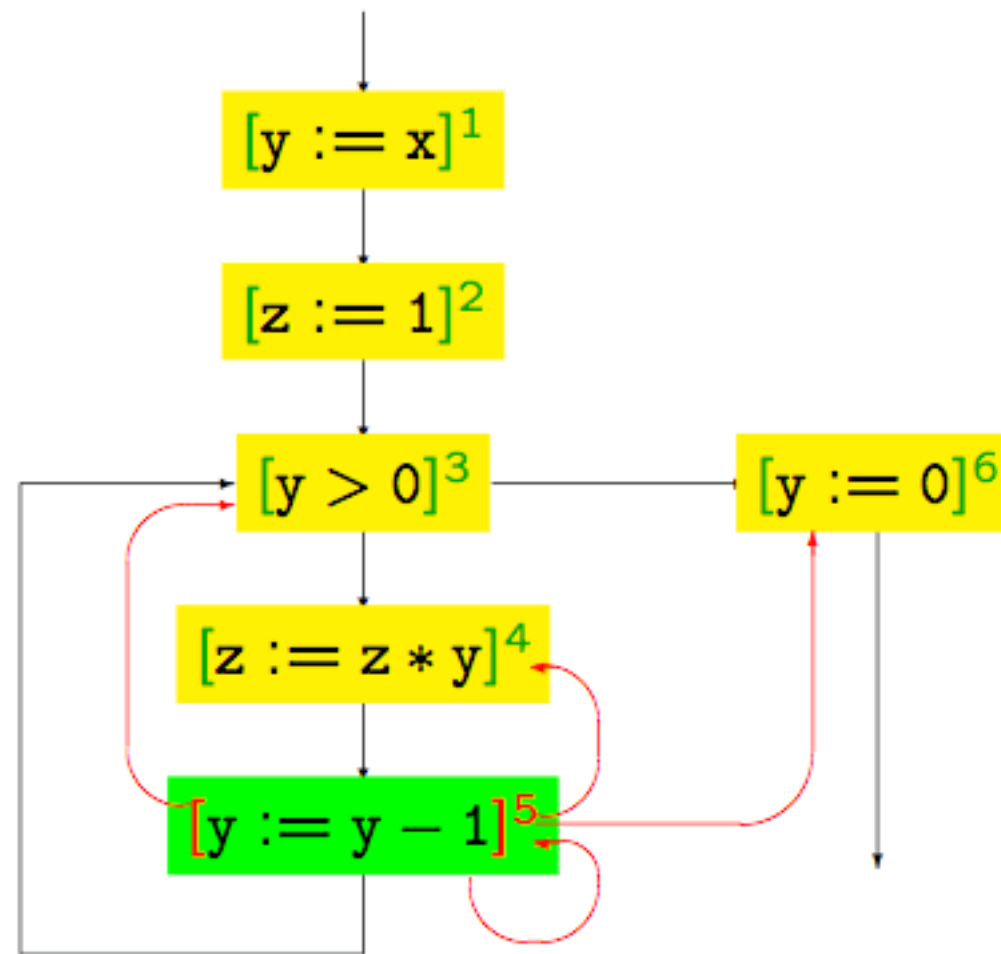
Data-flow analyses

- reaching definitions analysis
- available expressions analysis
- live variables analysis
- very busy expressions analysis
- constants propagation
- ...

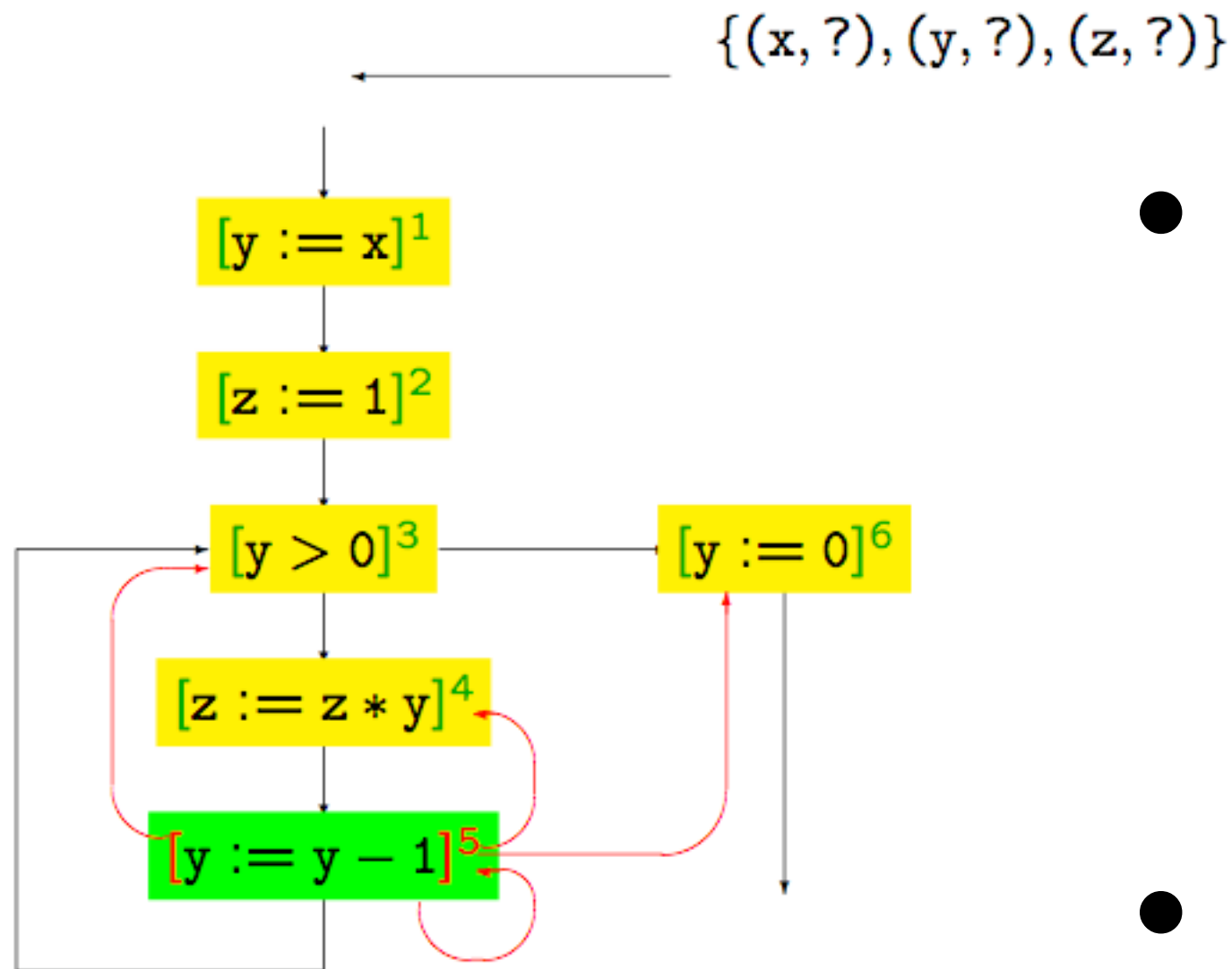
Reaching definitions analysis

In each control location compute the set of assignments that possibly have been executed (and not „overwritten”) prior to reaching this location.

Reaching definitions analysis



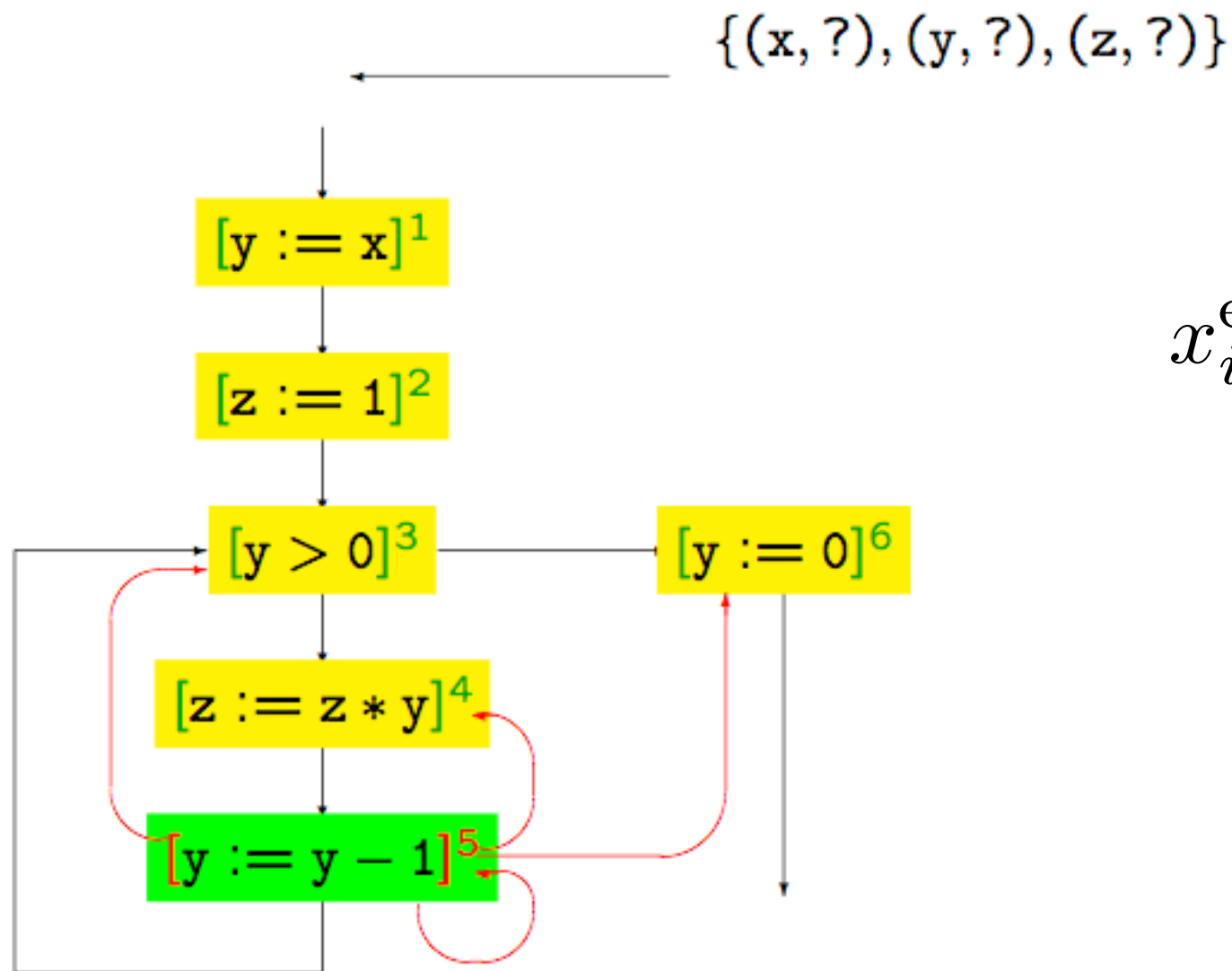
Reaching definitions analysis



- formalize the problem as a set of equations
 - variables represent information before and after each instruction
- **the lest** solution
- iterative algorithm

kill(i)	gen(i)	U
---------	--------	---

Reaching definitions analysis



$$x_i^{\text{exit}} = x_i^{\text{entry}} \setminus \text{kill}(i) \cup \text{gen}(i)$$

$$x_i^{\text{entry}} = \bigcup_{j \rightsquigarrow i} x_j^{\text{exit}}$$

$$x_1^{\text{entry}} = \{(x, ?), (y, ?), (z, ?)\}$$

we are interested in **the least** solution

Reaching definitions analysis

$$\text{kill}(z := e) = \{(z, ?)\} \cup \{(z, j) \mid j \in S\}$$

$$\text{kill}(b) = \emptyset$$

$$\text{kill}(\text{skip}) = \emptyset$$

$$\text{gen}([z := e]^i) = \{(z, i)\}$$

$$\text{gen}(b) = \emptyset$$

$$\text{gen}(\text{skip}) = \emptyset$$

$$x_i^{\text{exit}} = x_i^{\text{entry}} \setminus \text{kill}(i) \cup \text{gen}(i)$$

$$x_i^{\text{entry}} = \bigcup_{j \rightsquigarrow i} x_j^{\text{exit}}$$

$$x_1^{\text{entry}} = \{(x, ?), (y, ?), (z, ?)\}$$

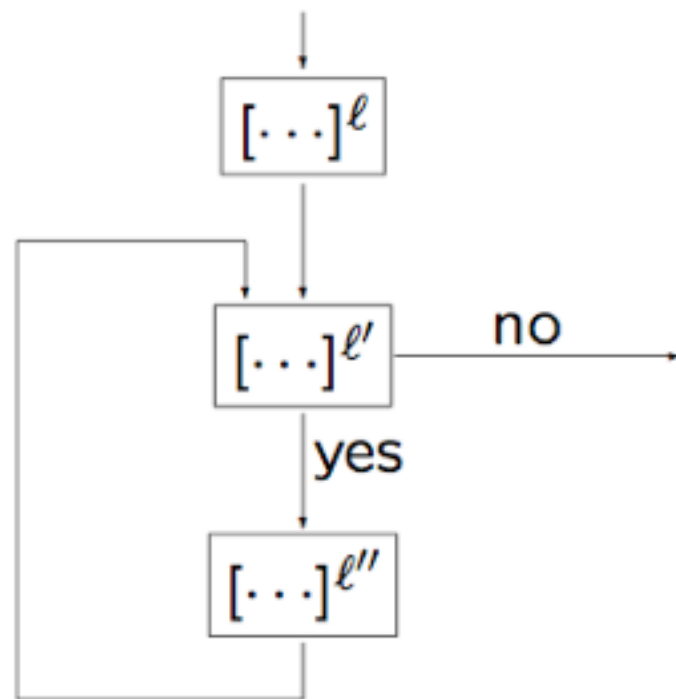
we are interested in **the least** solution

Reaching definitions analysis

	←	$\{(x, ?), (y, ?), (z, ?)\}$
$[y := x]^1;$	←	$\{(x, ?), (y, 1), (z, ?)\}$
$[z := 1]^2;$	←	$\{(x, ?), (y, 1), (y, 5), (z, 2), (z, 4)\}$
while $[y > 0]^3$ do	←	$\{(x, ?), (y, 1), (y, 5), (z, 2), (z, 4)\}$
$[z := z * y]^4;$	←	$\{(x, ?), (y, 1), (y, 5), \boxed{(z, 2)}, (z, 4)\}$
$[y := y - 1]^5$	←	$\{(x, ?), \boxed{(y, 1)}, (y, 5), \boxed{(z, 2)}, (z, 4)\}$
od;	←	$\{(x, ?), (y, 1), (y, 5), (z, 2), (z, 4)\}$
$[y := 0]^6$	←	$\{(x, ?), (y, 6), (z, 2), (z, 4)\}$

Reaching definitions analysis

$[z:=x+y]^\ell; \text{while } [\text{true}]^{\ell'} \text{ do } [\text{skip}]^{\ell''}$



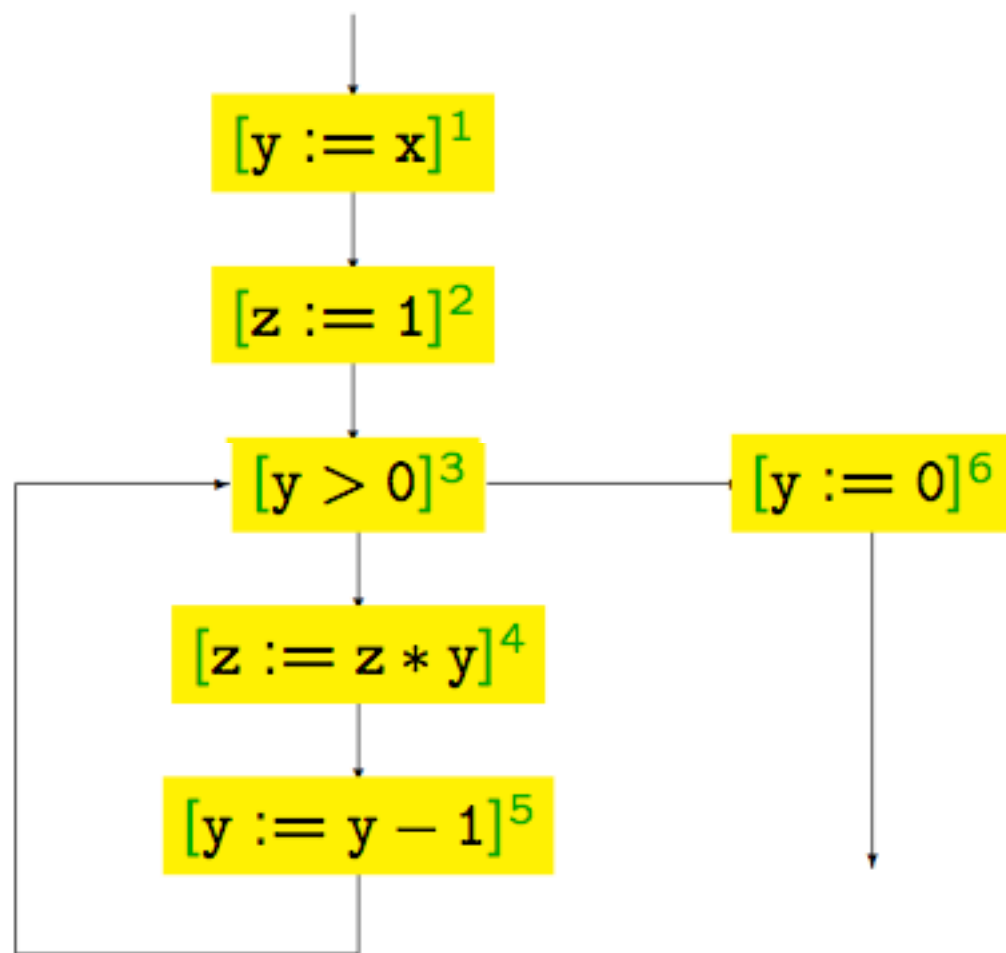
solutions: $x_{l'}^{\text{entry}} \supseteq \{(x, ?), (y, ?), (z, 1)\}$

we are interested in **the least** solution

Available expressions analysis

In each control location compute the set of expressions whose value is surely computed whenever this location is entered.

Available expressions analysis



$$x_i^{\text{exit}} = x_i^{\text{entry}} \setminus \text{kill}(i) \cup \text{gen}(i)$$

$$x_i^{\text{entry}} = \bigcap_{j \rightsquigarrow i} x_j^{\text{exit}}$$

$$x_1^{\text{entry}} = \emptyset$$

we are interested in **the greatest** solution

Available expressions analysis

$[x:=a+b]^1; [y:=a*b]^2; \text{while } [y>a+b]^3 \text{ do } ([a:=a+1]^4; [x:=a+b]^5)$

i	x_i^{entry}	x_i^{exit}
1	\emptyset	$\{a+b\}$
2	$\{a+b\}$	$\{a+b, a*b\}$
3	$\{a+b\}$	$\{a+b\}$
4	$\{a+b\}$	\emptyset
5	\emptyset	$\{a+b\}$

$$x_i^{\text{exit}} = x_i^{\text{entry}} \setminus \text{kill}(i) \cup \text{gen}(i)$$

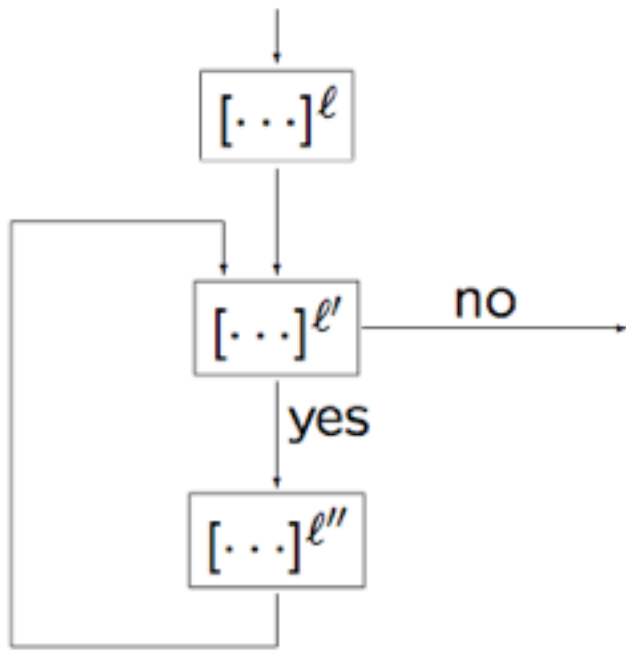
$$x_i^{\text{entry}} = \bigcap_{j \rightsquigarrow i} x_j^{\text{exit}}$$

$$x_1^{\text{entry}} = \emptyset$$

we are interested in **the greatest** solution

Available expressions analysis

$[z:=x+y]^\ell; \text{while } [\text{true}]^{\ell'} \text{ do } [\text{skip}]^{\ell''}$



$$x_i^{\text{exit}} = x_i^{\text{entry}} \setminus \text{kill}(i) \cup \text{gen}(i)$$

$$x_i^{\text{entry}} = \bigcap_{j \rightsquigarrow i} x_j^{\text{exit}}$$

two solutions:

$$x_{l'}^{\text{entry}} = \{x + y\}, \emptyset$$

$$x_1^{\text{entry}} = \emptyset$$

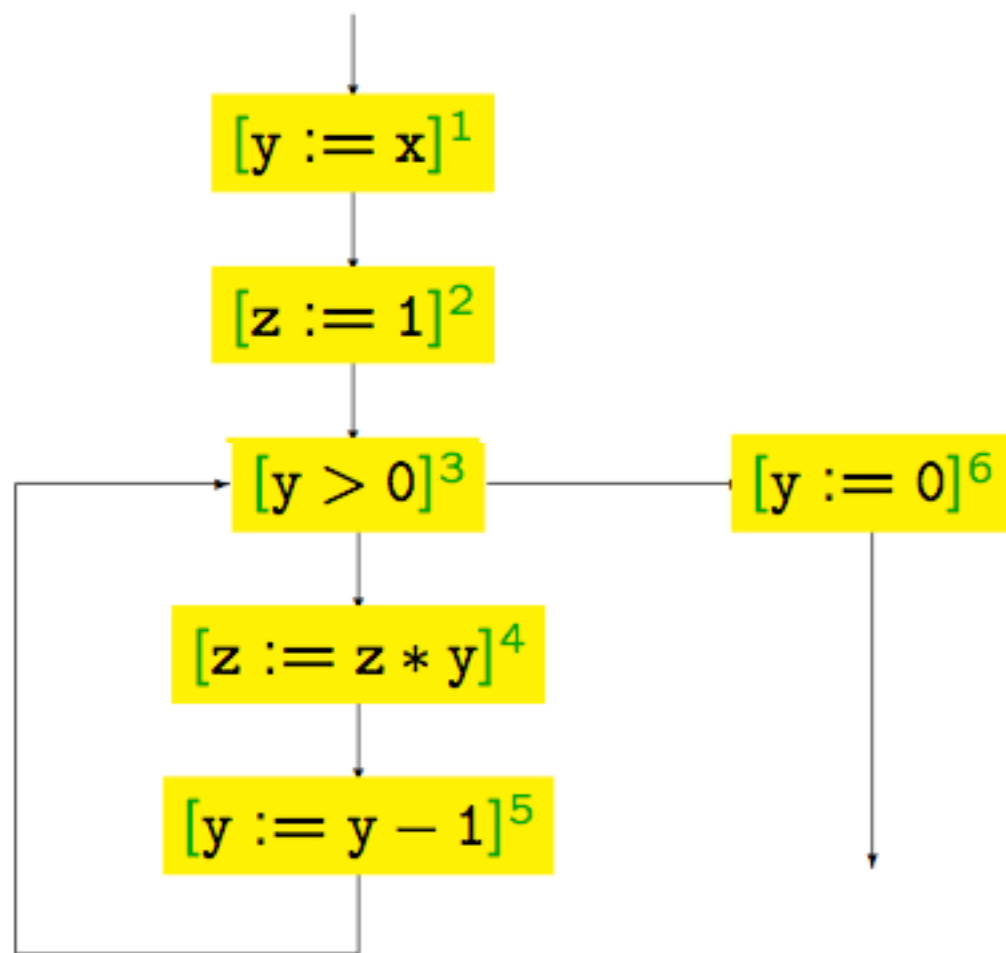
we are interested in **the greatest** solution

Live variables analysis

In each program location compute the set of variables that are live (possibly used in future before being redefined) when exiting this location.

backwards analysis

Live variables analysis



$$x_i^{\text{entry}} = x_i^{\text{exit}} \setminus \text{kill}(i) \cup \text{gen}(i)$$

$$x_i^{\text{exit}} = \bigcup_{i \rightsquigarrow j} x_j^{\text{entry}}$$

$$x_6^{\text{exit}} = \emptyset$$

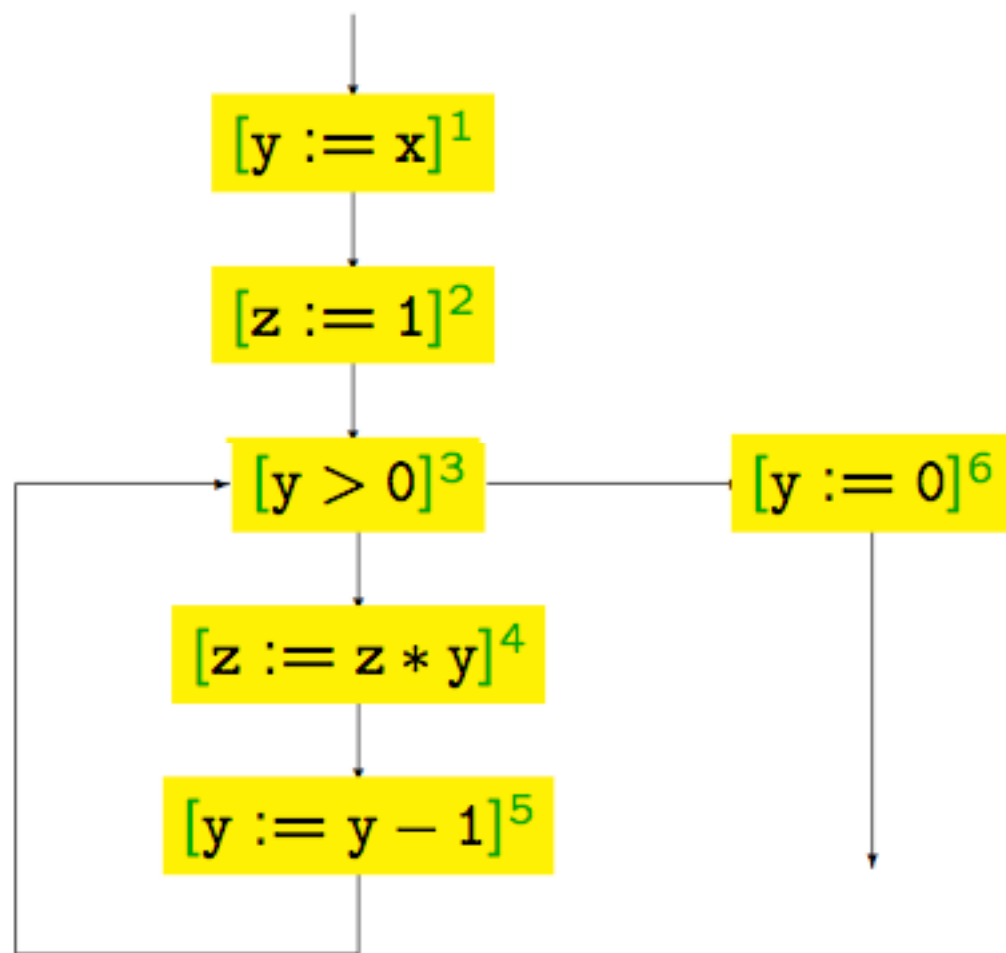
we are interested in **the least** solution

Very busy expressions analysis

In each control location compute the set of expressions that will surely be computed in future before redefinition of any of variables appearing in the expression.

backwards analysis

Very busy expressions analysis



$$x_i^{\text{entry}} = x_i^{\text{exit}} \setminus \text{kill}(i) \cup \text{gen}(i)$$

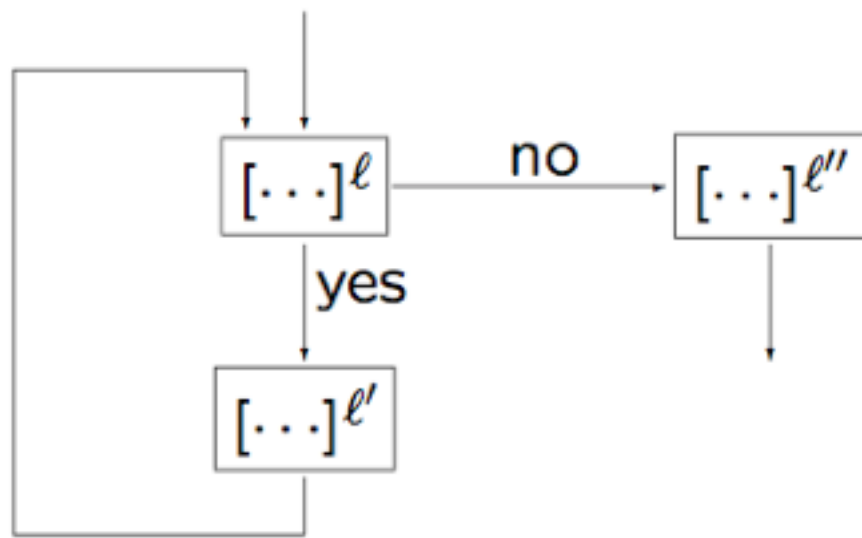
$$x_i^{\text{entry}} = \bigcap_{i \rightsquigarrow j} x_j^{\text{entry}}$$

$$x_6^{\text{exit}} = \emptyset$$

we are interested in **the greatest** solution

Very busy expressions analysis

`(while [x>1]ℓ do [skip]ℓ'); [x:=x+1]ℓ''`



$$x_i^{\text{entry}} = x_i^{\text{exit}} \setminus \text{kill}(i) \cup \text{gen}(i)$$

$$x_i^{\text{entry}} = \bigcap_{i \rightsquigarrow j} x_j^{\text{entry}}$$

two solutions:

$$x_l^{\text{exit}} = \{x + 1\}, \emptyset$$

$$x_6^{\text{exit}} = \emptyset$$

we are interested in **the greatest** solution

Abstract interpretation

Generalization

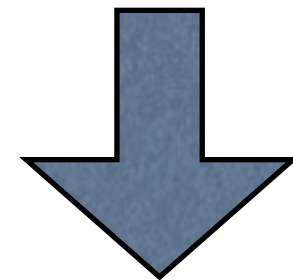
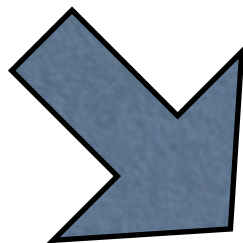
L - abstract space

(L, \sqsubseteq) - complete lattice

\sqcup, \sqcap - bounds

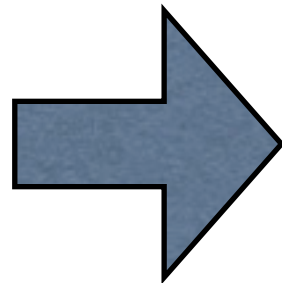
$$f : S \rightarrow \text{Mon}(L \rightarrow L)$$

$$f_{\text{init}} : \text{init}(S) \rightarrow L$$



$$S = \{1, \dots, n\}, \quad \rightsquigarrow \subseteq S \times S$$

$$\text{init}(S) \subseteq S$$



$$x_i^{\text{entry}} = \bigsqcup_{j \rightsquigarrow i} x_j^{\text{exit}} \sqcup f_{\text{init}}(x)$$

$$x_i^{\text{exit}} = f(x)(x_i^{\text{entry}})$$

Generalization

abstract
interpretation

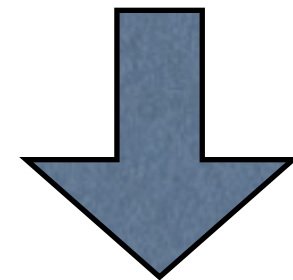
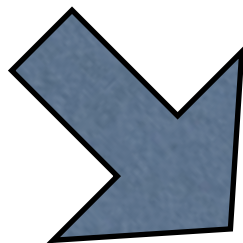
L - abstract space

(L, \sqsubseteq) - complete lattice

\sqcup, \sqcap - bounds

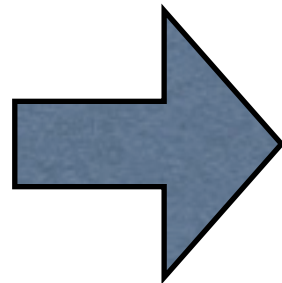
$$f : S \rightarrow \text{Mon}(L \rightarrow L)$$

$$f_{\text{init}} : \text{init}(S) \rightarrow L$$



$$S = \{1, \dots, n\}, \quad \rightsquigarrow \subseteq S \times S$$

$$\text{init}(S) \subseteq S$$



$$x_i^{\text{entry}} = \bigsqcup_{j \rightsquigarrow i} x_j^{\text{exit}} \sqcup f_{\text{init}}(x)$$

$$x_i^{\text{exit}} = f(x)(x_i^{\text{entry}})$$

Data-flow analyses

- reaching definitions analysis
- available expressions analysis
- live variables analysis
- very busy expressions analysis
- constants propagation
- ...

Data-flow analyses

- reaching definitions analysis $\mathcal{P}(\text{Var} \times (\mathcal{S} \cup \{?\}))$
- available expressions analysis
- live variables analysis
- very busy expressions analysis
- constants propagation
- ...

Data-flow analyses

- reaching definitions analysis $\mathcal{P}(\text{Var} \times (\mathcal{S} \cup \{?\}))$
- available expressions analysis $\mathcal{P}(\text{Expr})$
- live variables analysis
- very busy expressions analysis
- constants propagation
- ...

Data-flow analyses

- reaching definitions analysis $\mathcal{P}(\text{Var} \times (\mathcal{S} \cup \{?\}))$
- available expressions analysis $\mathcal{P}(\text{Expr})$
- live variables analysis $\mathcal{P}(\text{Var})$
- very busy expressions analysis
- constants propagation
- ...

Data-flow analyses

- reaching definitions analysis $\mathcal{P}(\text{Var} \times (\mathcal{S} \cup \{?\}))$
- available expressions analysis $\mathcal{P}(\text{Expr})$
- live variables analysis $\mathcal{P}(\text{Var})$
- very busy expressions analysis $\mathcal{P}(\text{Expr})$
- constants propagation
- ...

Data-flow analyses

- reaching definitions analysis $\mathcal{P}(\text{Var} \times (\mathcal{S} \cup \{?\}))$
- available expressions analysis $\mathcal{P}(\text{Expr})$
- live variables analysis $\mathcal{P}(\text{Var})$
- very busy expressions analysis $\mathcal{P}(\text{Expr})$
- constants propagation $\text{Var} \rightarrow \mathbb{Z}^\top$
- ...

Distributivity

$$f(s)(l_1 \sqcup l_2) = f(s)(l_1) \sqcup f(s)(l_2)$$

holds whenever:

$$L = \mathcal{P}(\mathcal{D}) \quad \mathcal{D} - \text{finite}$$

$$f(s)(l) = l \setminus l_1 \cup l_2$$

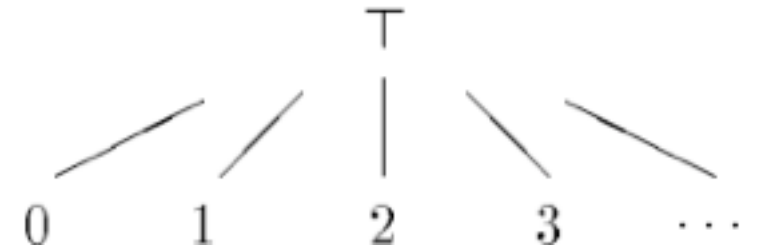
may not hold

Constants propagation

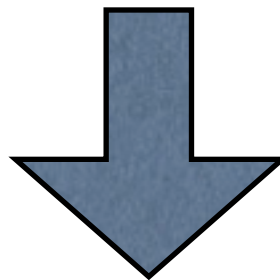
In each control location compute the set of variables that have a constant value independent from the history.

Constants propagation

$$L = \text{Var} \rightarrow \mathbb{Z}^\top$$



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



`[x:=6]1; [y:=3]2; while [x > 3]3 do ([x:=x - 1]4; [z:=9]6)`

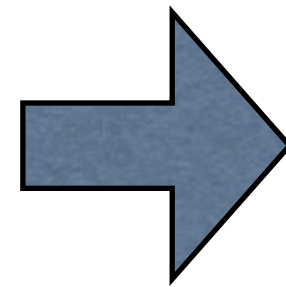
$$l_1(y) = 5 \quad l_2(y) = -5$$

$$f(6)(l_1 \sqcup l_2)(z) = \top$$

$$f(6)(l_1)(z) = f(6)(l_2)(z) = 25$$

Algorithm

$$x_i^{\text{entry}} = \bigsqcup_{j \rightsquigarrow i} x_j^{\text{exit}} \sqcup f_{\text{init}}(x)$$
$$x_i^{\text{exit}} = f(x)(x_i^{\text{entry}})$$



$$\vec{x} = \vec{f}(\vec{x})$$

complete lattice $L^{S \times \{\text{entry}, \text{exit}\}}$ with the coordinate-wise order

the lest fix-point of the monotonic function \vec{f}

iterative algorithm

we assume that L has only **finite chains**

LFP

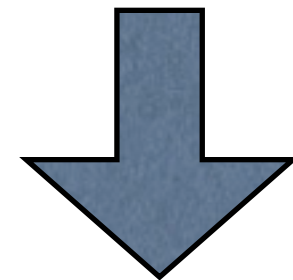
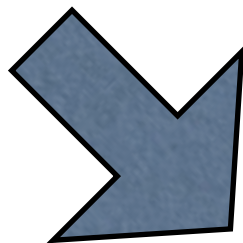
L - abstract space

(L, \sqsubseteq) - complete lattice

\sqcup, \sqcap - bounds

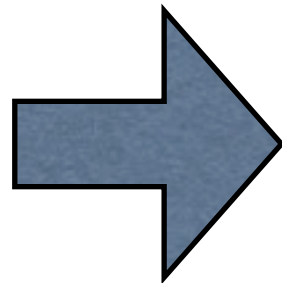
$$f : S \rightarrow \text{Mon}(L \rightarrow L)$$

$$f_{\text{init}} : \text{init}(S) \rightarrow L$$



$$S = \{1, \dots, n\}, \quad \rightsquigarrow \subseteq S \times S$$

$$\text{init}(S) \subseteq S$$



$$x_i^{\text{entry}} = \bigsqcup_{j \rightsquigarrow i} x_j^{\text{exit}} \sqcup f_{\text{init}}(x)$$

$$x_i^{\text{exit}} = f(x)(x_i^{\text{entry}})$$

MOP

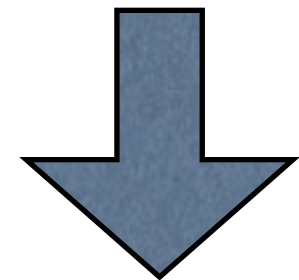
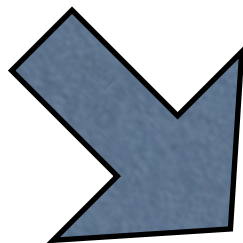
L - abstract space

(L, \sqsubseteq) - complete lattice

\sqcup, \sqcap - bounds

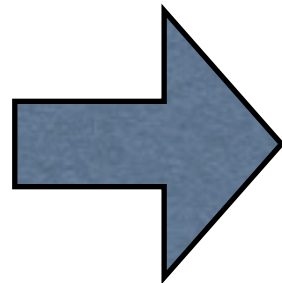
$$f : S \rightarrow \text{Mon}(L \rightarrow L)$$

$$f_{\text{init}} : \text{init}(S) \rightarrow L$$



$$S = \{1, \dots, n\}, \quad \rightsquigarrow \subseteq S \times S$$

$$\text{init}(S) \subseteq S$$



$$y_i^{\text{entry}} = \sqcup \{f(p) \mid p \in \text{paths}^{\text{entry}}(i)\}$$

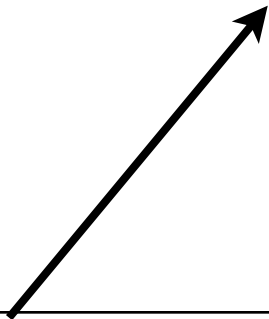
$$y_i^{\text{exit}} = \sqcup \{f(p) \mid p \in \text{paths}^{\text{exit}}(i)\}$$

$$\text{MOP} \sqsubseteq \text{LFP}$$

$$\vec{y} \sqsubseteq \vec{x}$$

MOP \sqsubseteq LFP

not always computable



$\vec{y} \sqsubseteq \vec{x}$

MOP = LFP

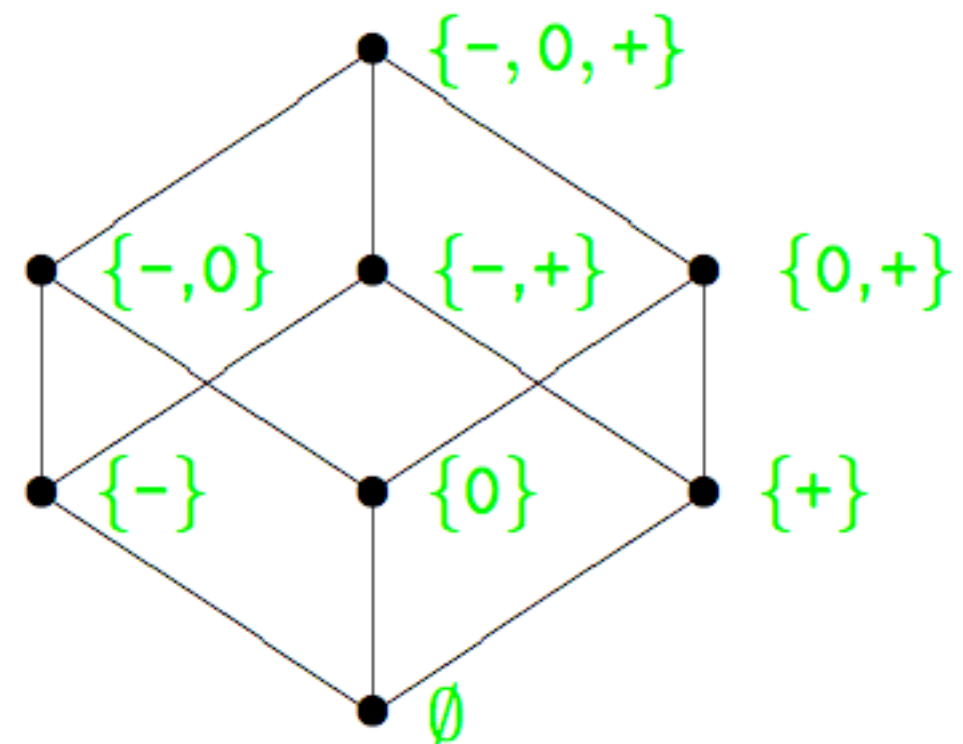
$$\vec{y} = \vec{x}$$

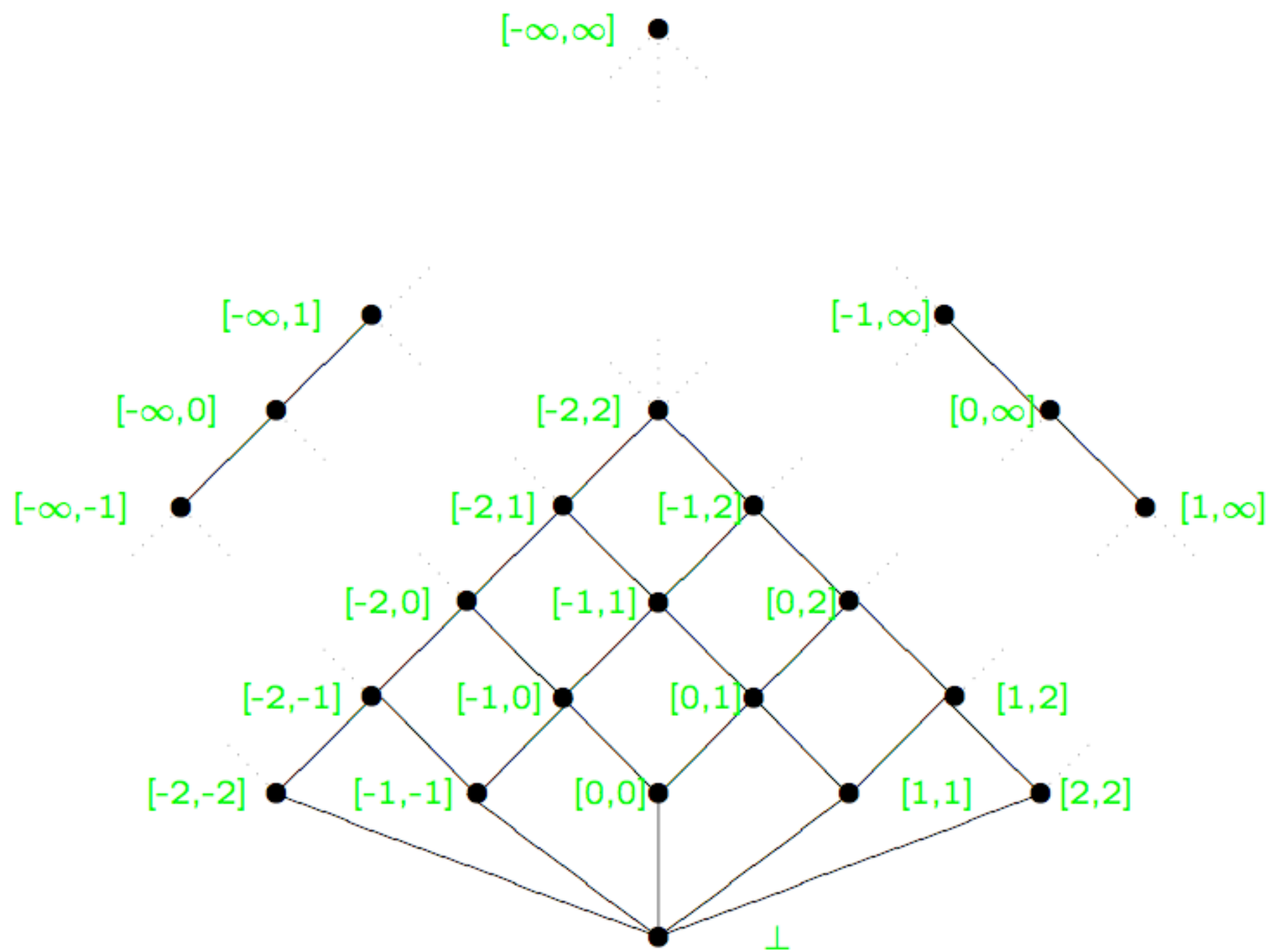
when distributivity holds

Abstract domains

Non-relational domains

- signs $\mathcal{P}(-, 0, +)$
- intervals $[n, m]$
- parity
- congruence modulo k

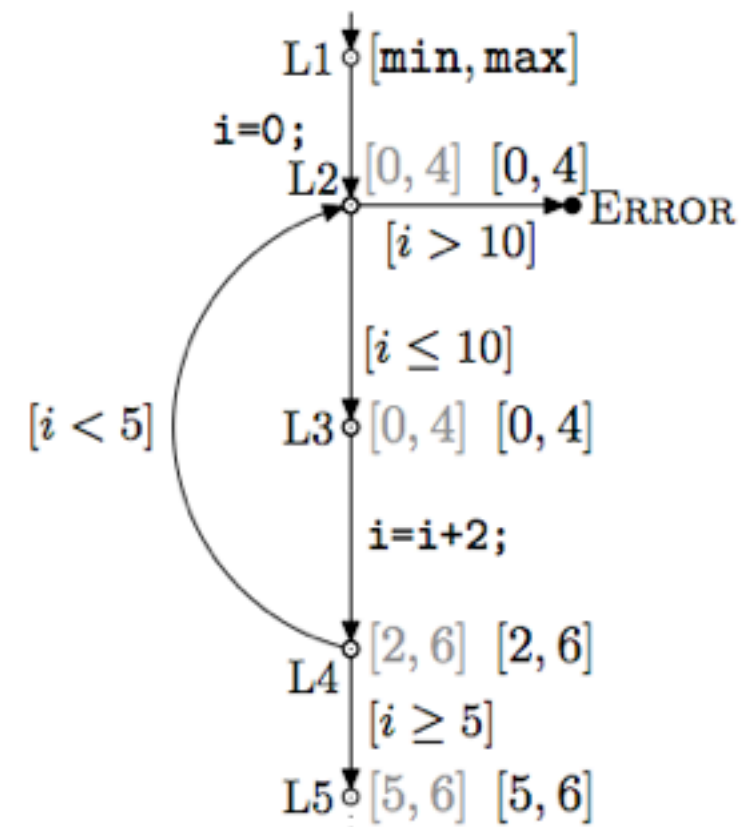
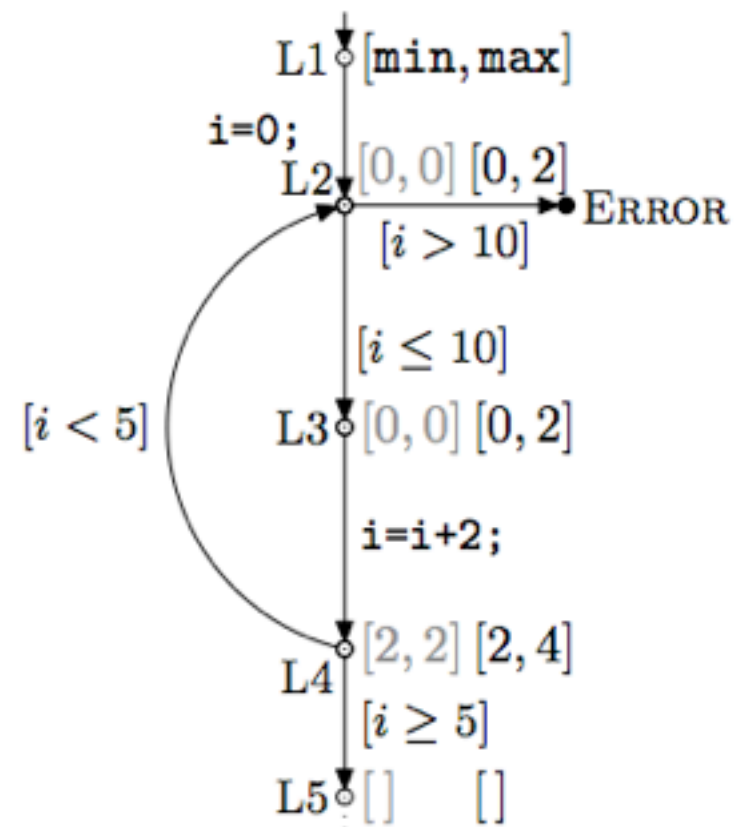
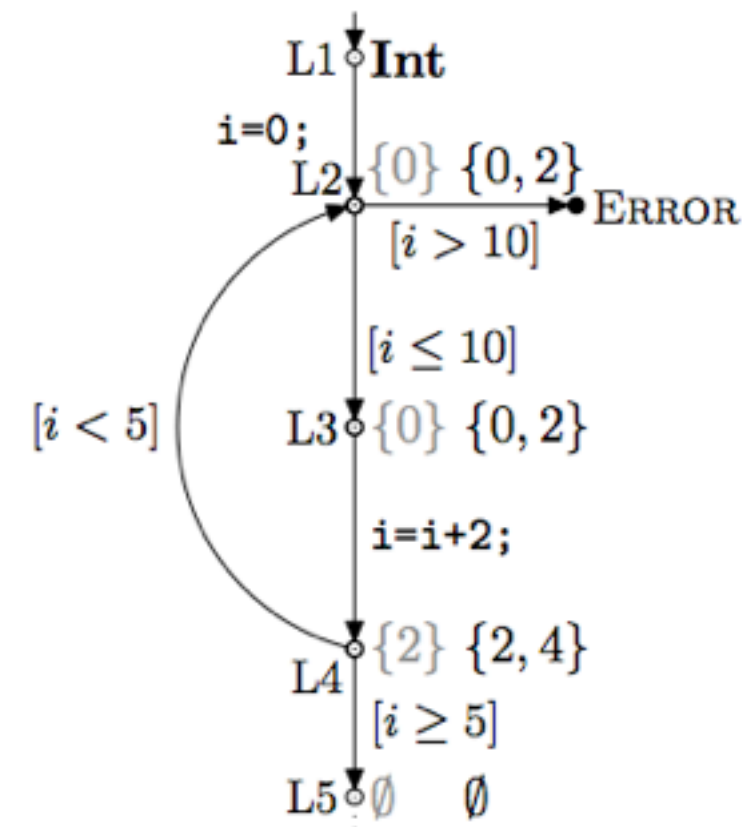




```

int i = 0;
do {
    assert(i <= 10);
    i = i+2;
} while (i < 5);

```



Expressive power

- signs

- intervals

- DBM (difference bounds matrices) $x - y \leq c$

- octagon $\begin{matrix} + & + \\ - & - \end{matrix} x \quad y \leq c$

- octahedra $\begin{matrix} + & + \\ - & - \end{matrix} x_1 \dots x_n \leq c$

- polyhedra $a_1 x_1 + \dots + a_n x_n \leq c$

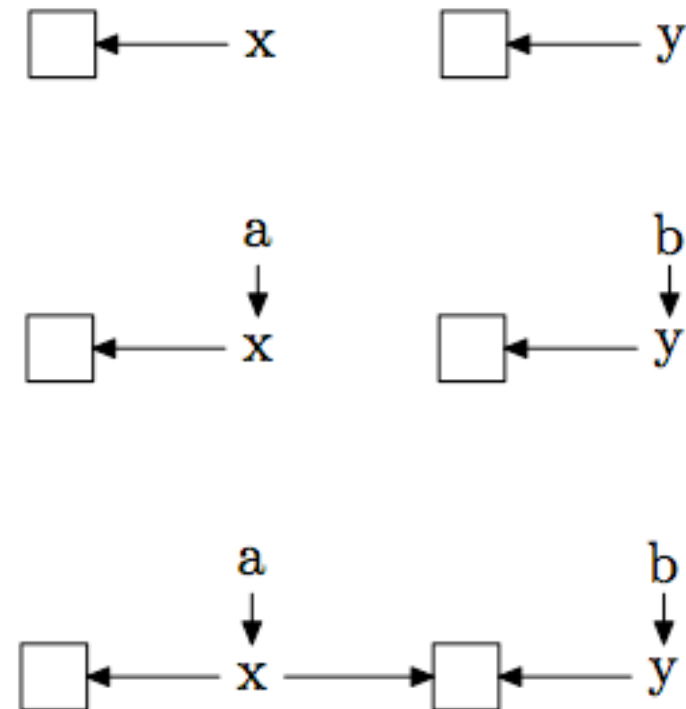
precision
↓

Pointer analyses domains

- points-to graphs

Example: alias analysis

```
int **a, **b, *x, *y;  
x = (int*) malloc(sizeof(int));  
y = (int*) malloc(sizeof(int));  
a = &x;  
b = &y;  
*a = y;
```



`a` and `b` **do not** point to the same location

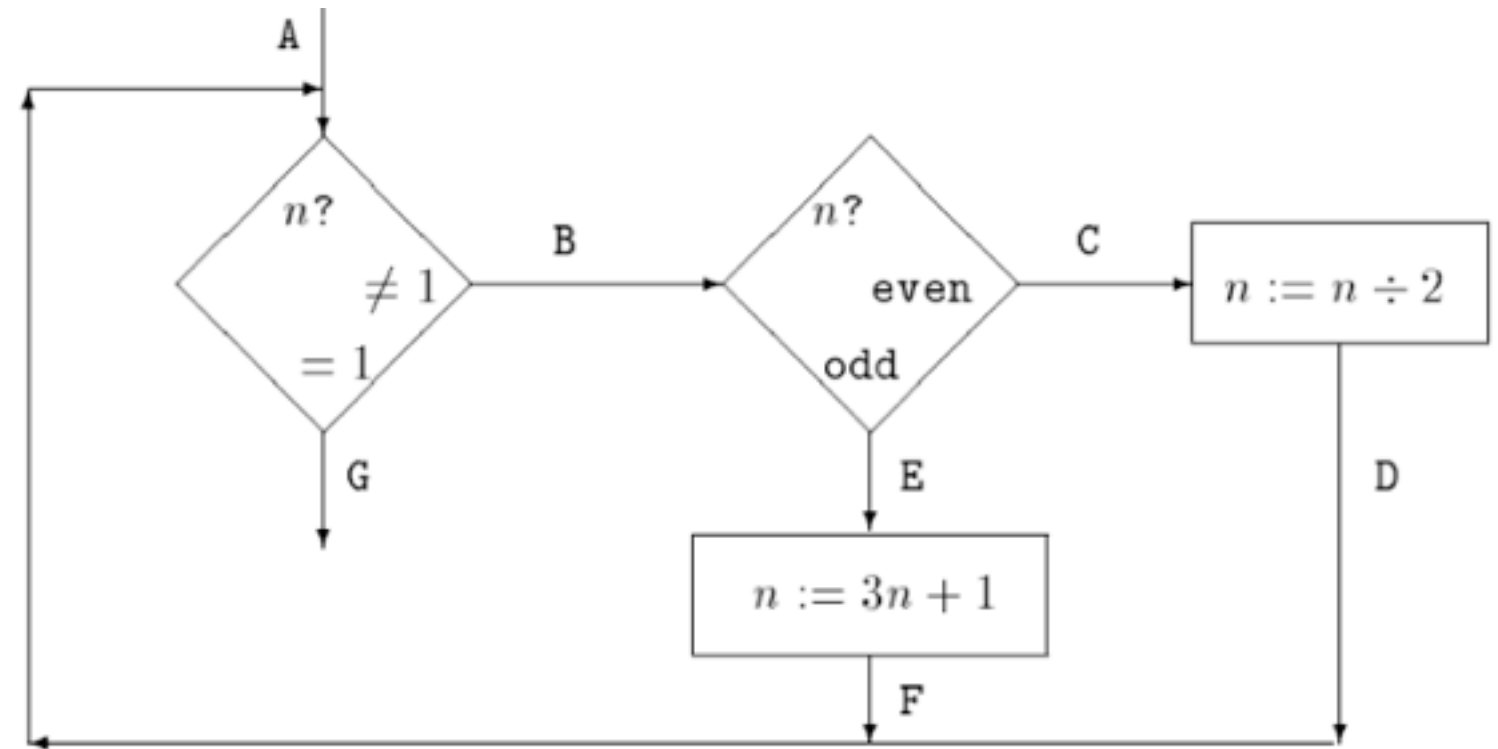
`x` and `y` **may** point to the same location

Abstract semantics

```

A: while  $n \neq 1$  do
  B: if  $n$  even
    then (C:  $n := n \div 2$ ; D: )
    else (E:  $n := 3 * n + 1$ ; F: )
  fi
od
G:

```



$$S = \{A, B, C, D, E, F, G\}$$

$$\text{State} = S \times \text{Store}$$

$$\text{Store} = \text{Var} \rightarrow \text{Val}$$

```
A: while  $n \neq 1$  do  
  B: if  $n$  even  
    then (C:  $n := n \div 2$ ; D: )  
    else (E:  $n := 3 * n + 1$ ; F: )  
  fi  
od  
G:
```

**concrete
semantics**



**abstract
semantics**

Domains

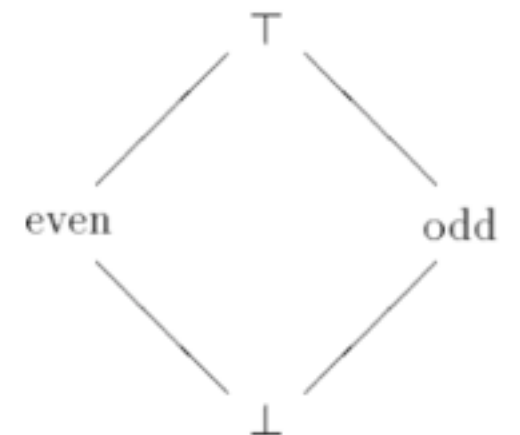
**concrete
semantics**



**abstract
semantics**

$$V = \text{Store} = \text{Var} \rightarrow \mathbb{Z}$$

$$L = \text{Var} \rightarrow \{\perp, \text{even}, \text{odd}, \top\}$$



Abstract semantics

$n := n \div 2;$

$\perp \mapsto \perp$

odd, even, $\top \mapsto \top$

$n := 3 * n + 1;$

$\perp \mapsto \perp$

odd \mapsto even

even \mapsto odd

$\top \mapsto \top$

**concrete
semantics**

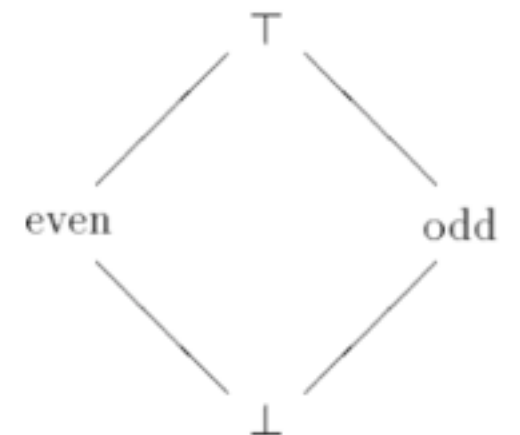


**abstract
semantics**

$$V = \text{Store} = \text{Var} \rightarrow \mathbb{Z}$$

do these two semantics agree?

$$L = \text{Var} \rightarrow \{\perp, \text{even}, \text{odd}, \top\}$$



Representation function

concrete
semantics

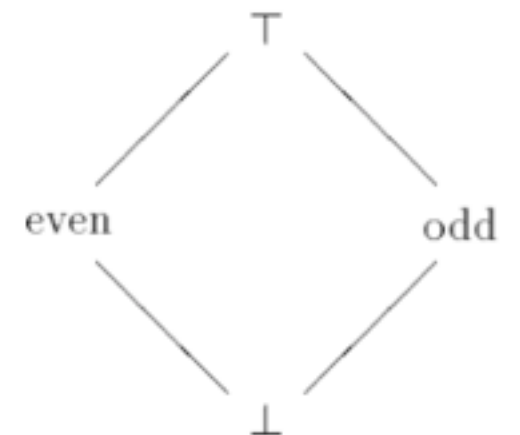
$$V = \text{Store} = \text{Var} \rightarrow \mathbb{Z}$$

$\beta : V \rightarrow L$
monotonic

$$\beta(v) = \begin{cases} \text{even} & \text{if } v \text{ even} \\ \text{odd} & \text{if } v \text{ odd} \end{cases}$$

abstract
semantics

$$L = \text{Var} \rightarrow \{\perp, \text{even}, \text{odd}, \top\}$$



Representation function

concrete
semantics

the best approximation

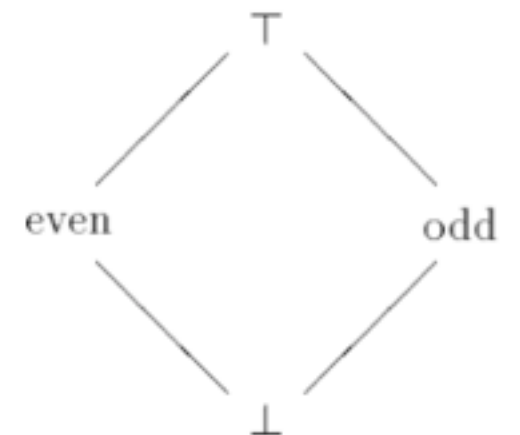
$$V = \text{Store} = \text{Var} \rightarrow \mathbb{Z}$$

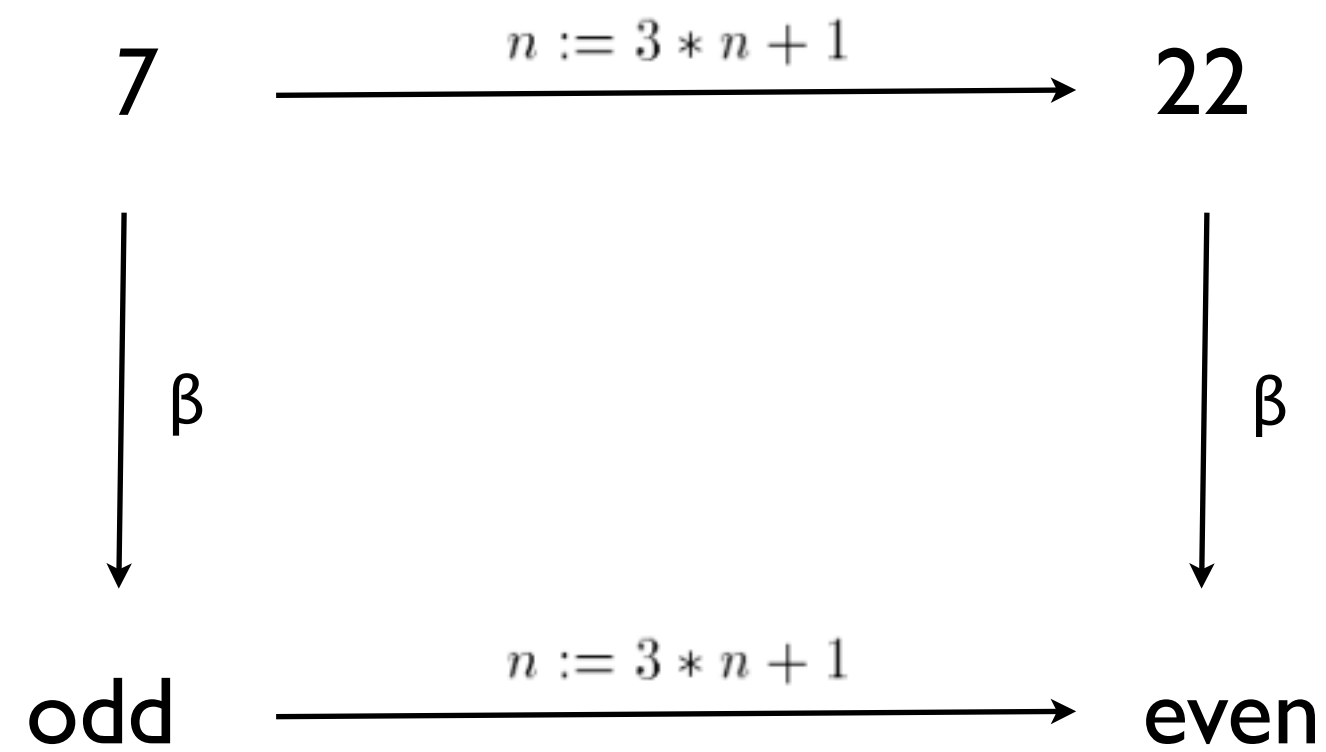
$\beta : V \rightarrow L$
monotonic

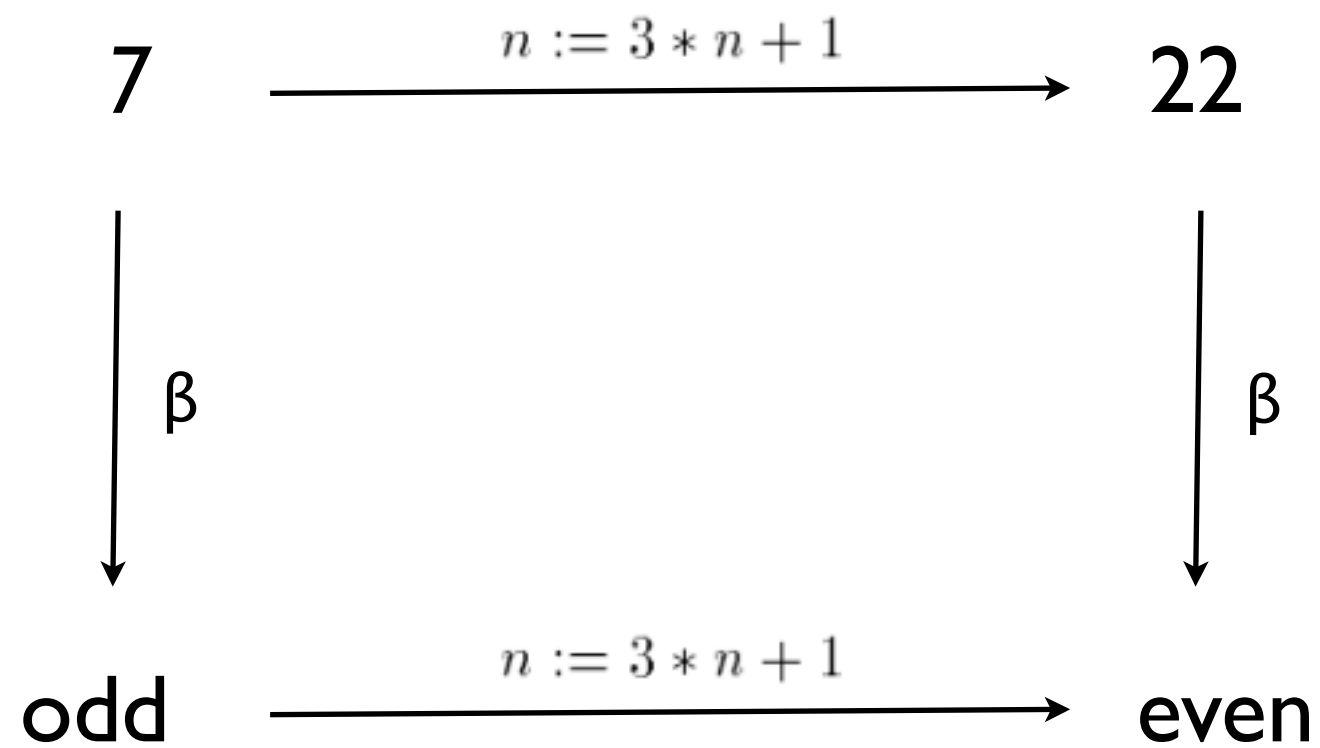
$$\beta(v) = \begin{cases} \text{even} & \text{if } v \text{ even} \\ \text{odd} & \text{if } v \text{ odd} \end{cases}$$

abstract
semantics

$$L = \text{Var} \rightarrow \{\perp, \text{even}, \text{odd}, \top\}$$

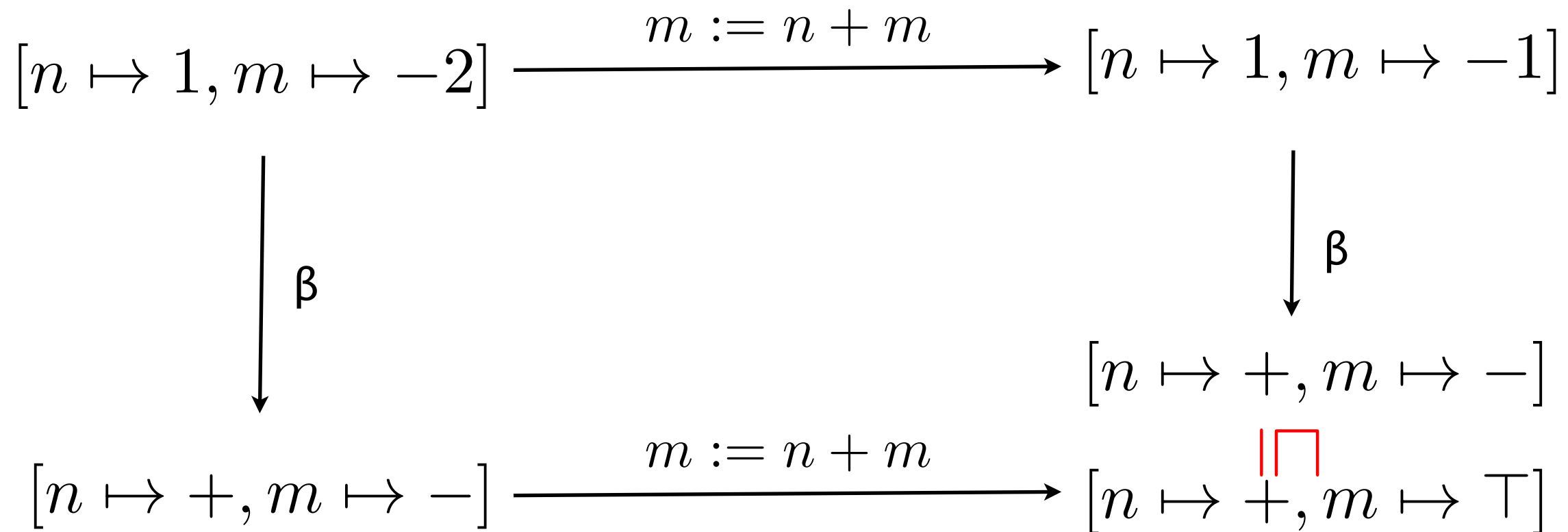






β is not always a homomorphism!

β is not always a homomorphism!



```
A: while  $n \neq 1$  do  
  B: if  $n$  even  
    then (C:  $n := n \div 2$ ; D: )  
    else (E:  $n := 3 * n + 1$ ; F: )  
  fi  
od  
G:
```

standard
semantics



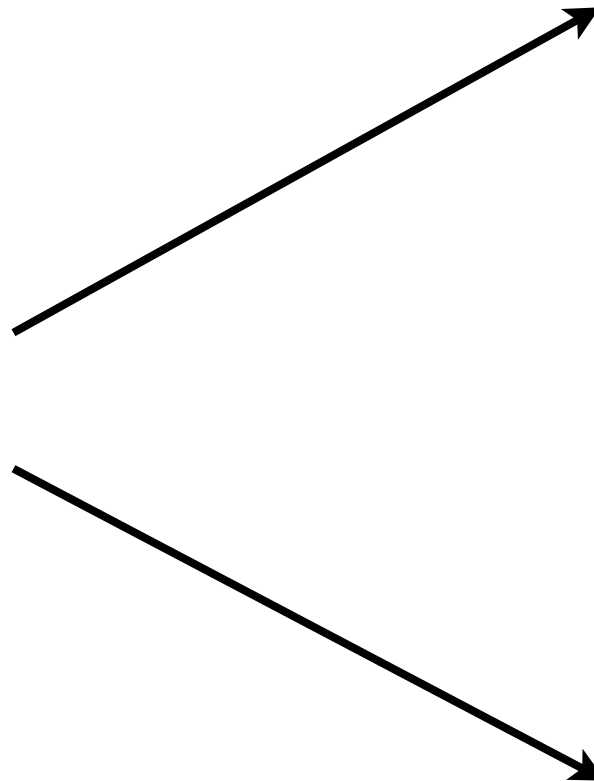
abstract
semantics

cumulative
semantics

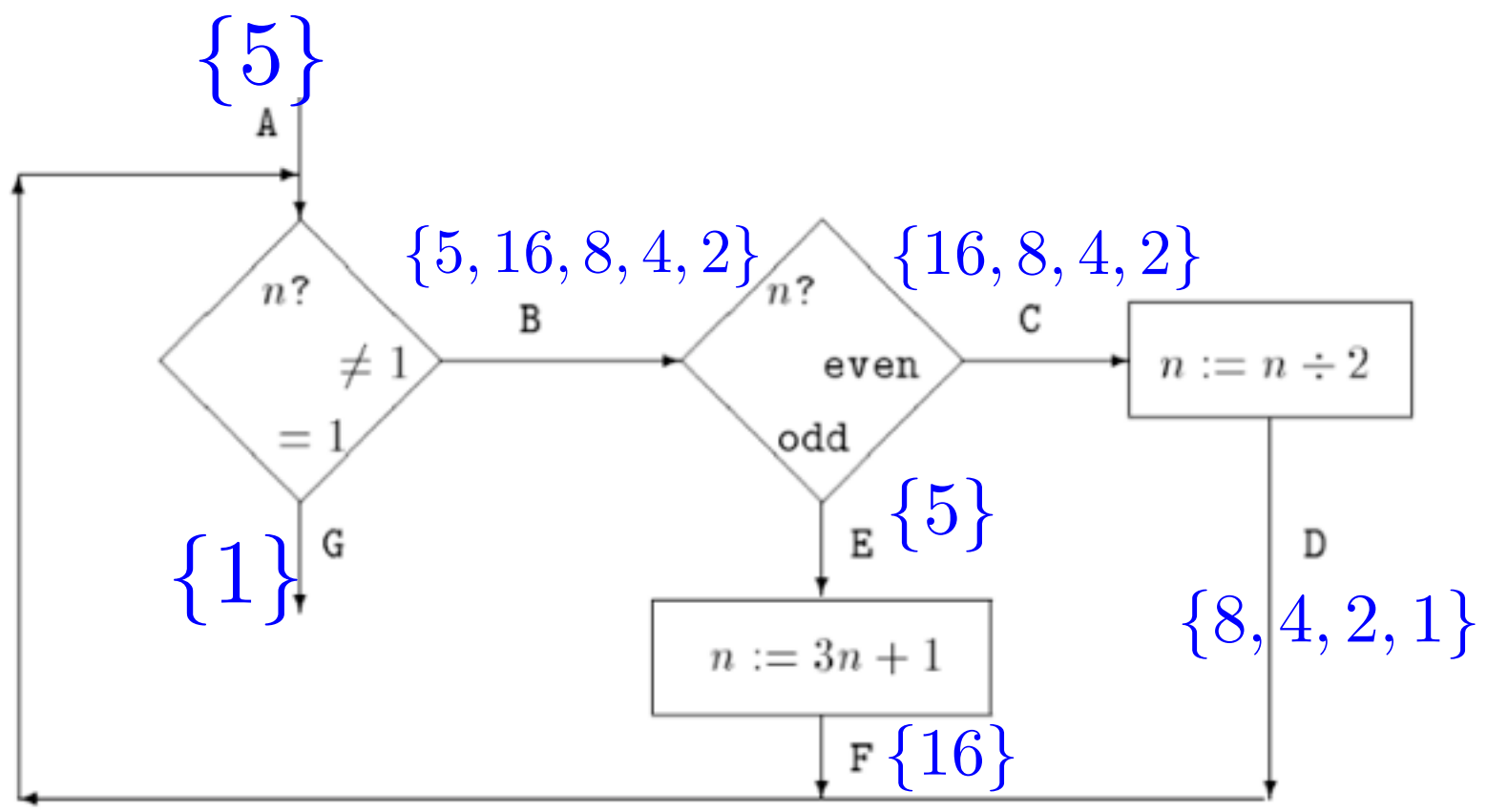
standard
semantics

```
A: while  $n \neq 1$  do
  B: if  $n$  even
    then (C:  $n := n \div 2$ ; D: )
    else (E:  $n := 3 * n + 1$ ; F: )
  fi
od
G:
```

abstract
semantics



```
A: while  $n \neq 1$  do
  B: if  $n$  even
    then (C:  $n := n \div 2$ ; D: )
    else (E:  $n := 3 * n + 1$ ; F: )
  fi
od
G:
```



Abstraction function

concrete
semantics

$\mathcal{P}(V)$

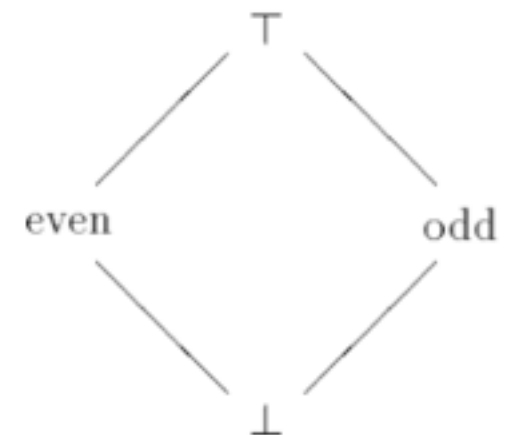
abstraction

$\alpha : \mathcal{P}(V) \rightarrow L$

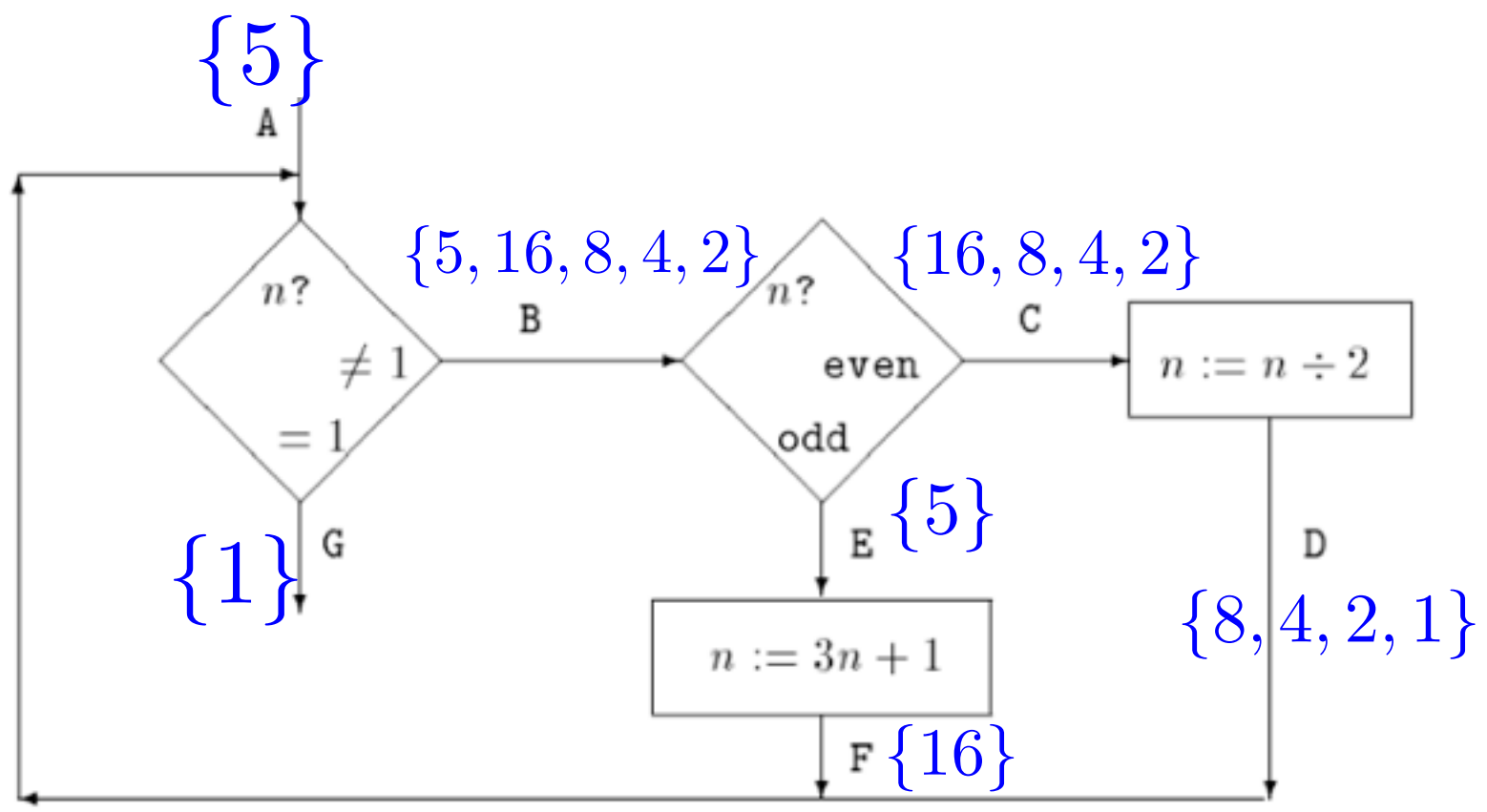
$\alpha(X) = \sqcup \{ \beta(v) \mid v \in X \}$

abstract
semantics

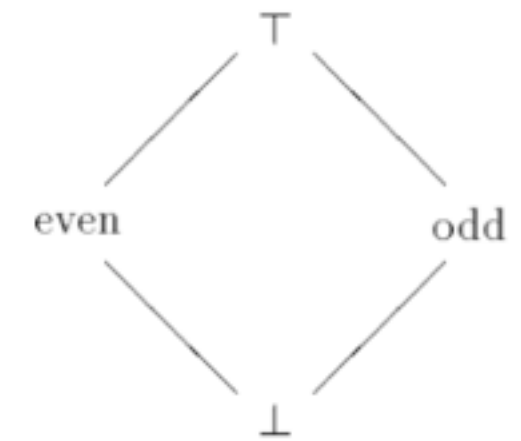
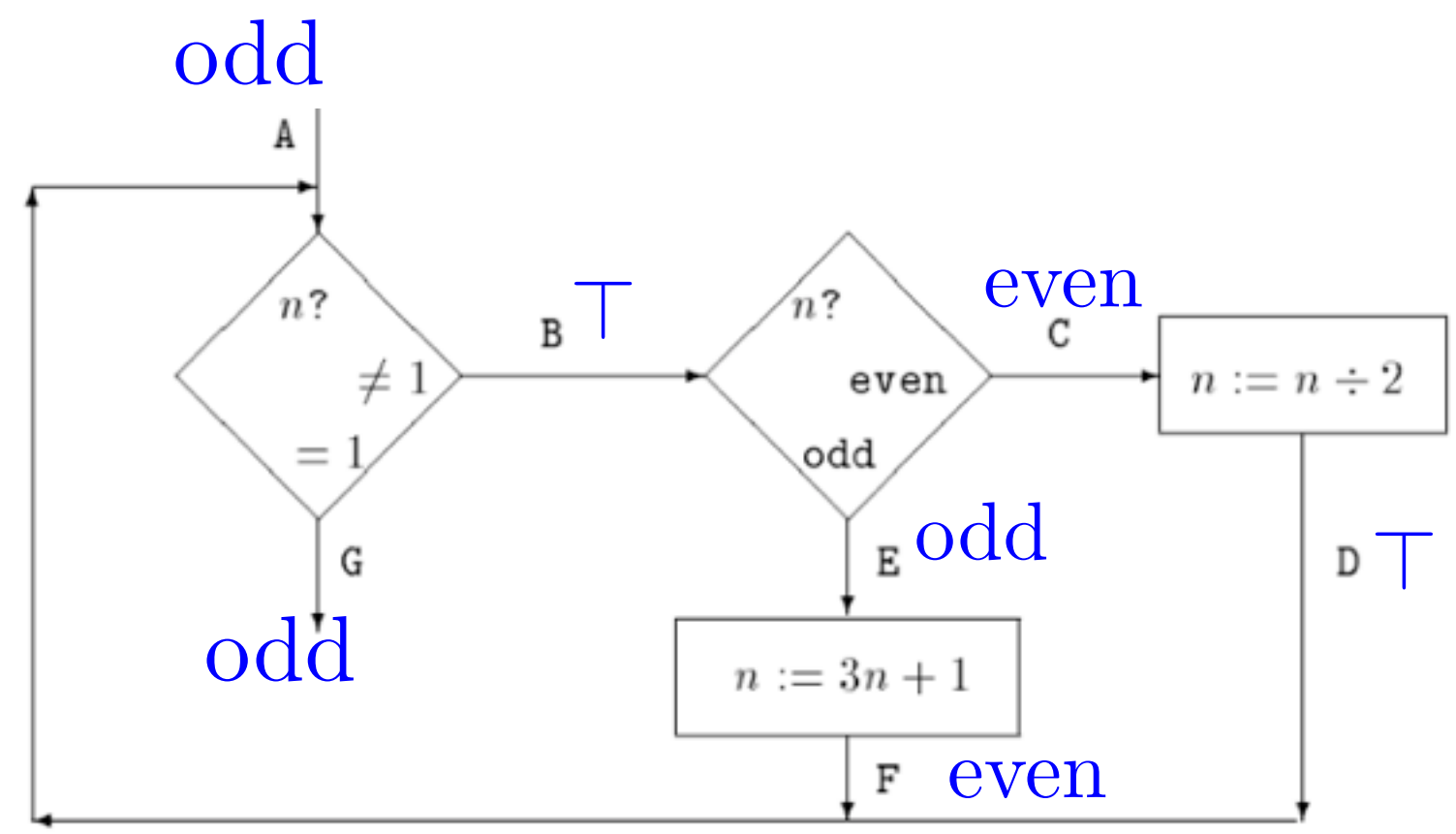
$L = \text{Var} \rightarrow \{ \perp, \text{even}, \text{odd}, \top \}$



```
A: while  $n \neq 1$  do
  B: if  $n$  even
    then (C:  $n := n \div 2$ ; D: )
    else (E:  $n := 3 * n + 1$ ; F: )
  fi
od
G:
```

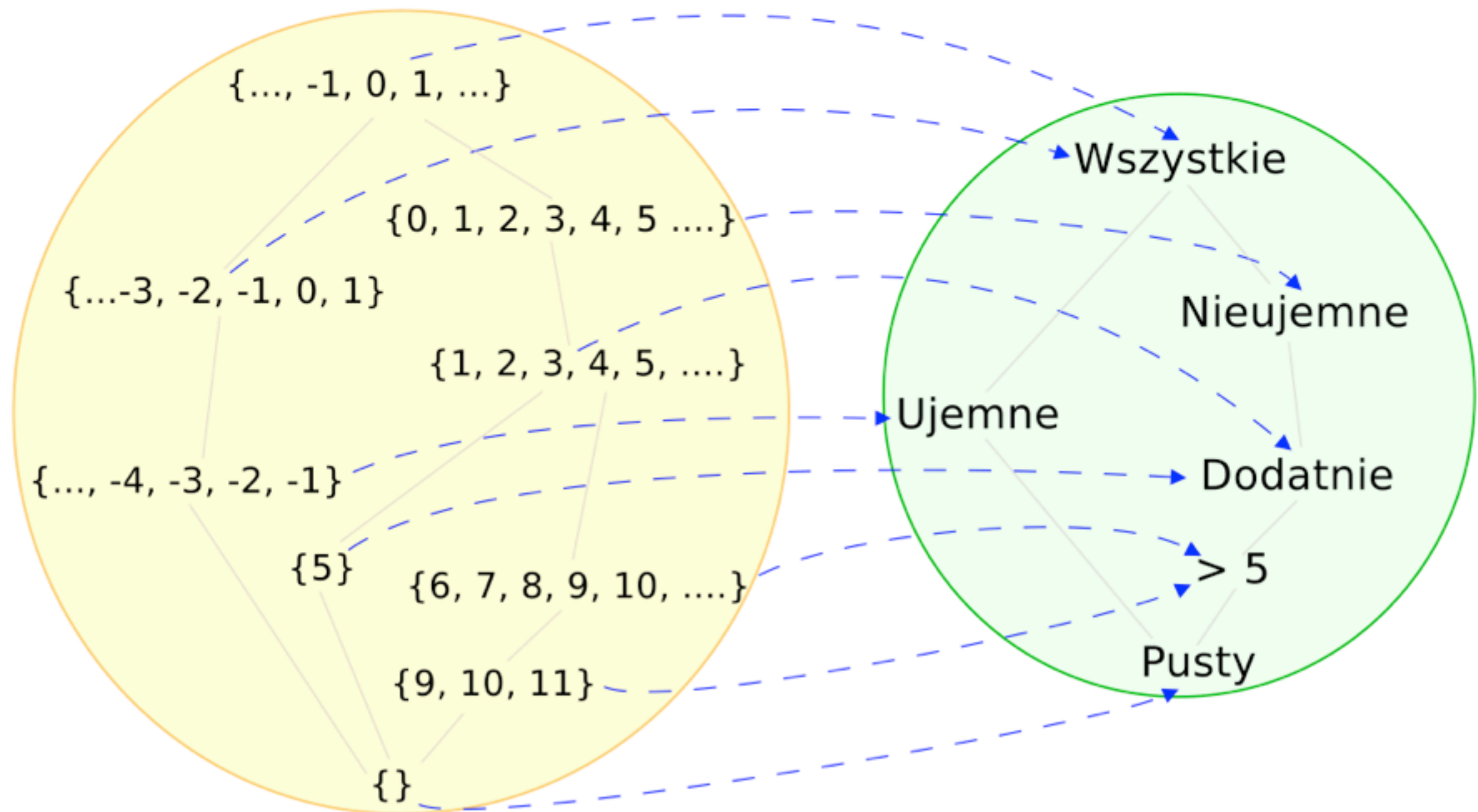


```
A: while  $n \neq 1$  do
  B: if  $n$  even
    then (C:  $n := n \div 2$ ; D: )
    else (E:  $n := 3 * n + 1$ ; F: )
  fi
od
G:
```



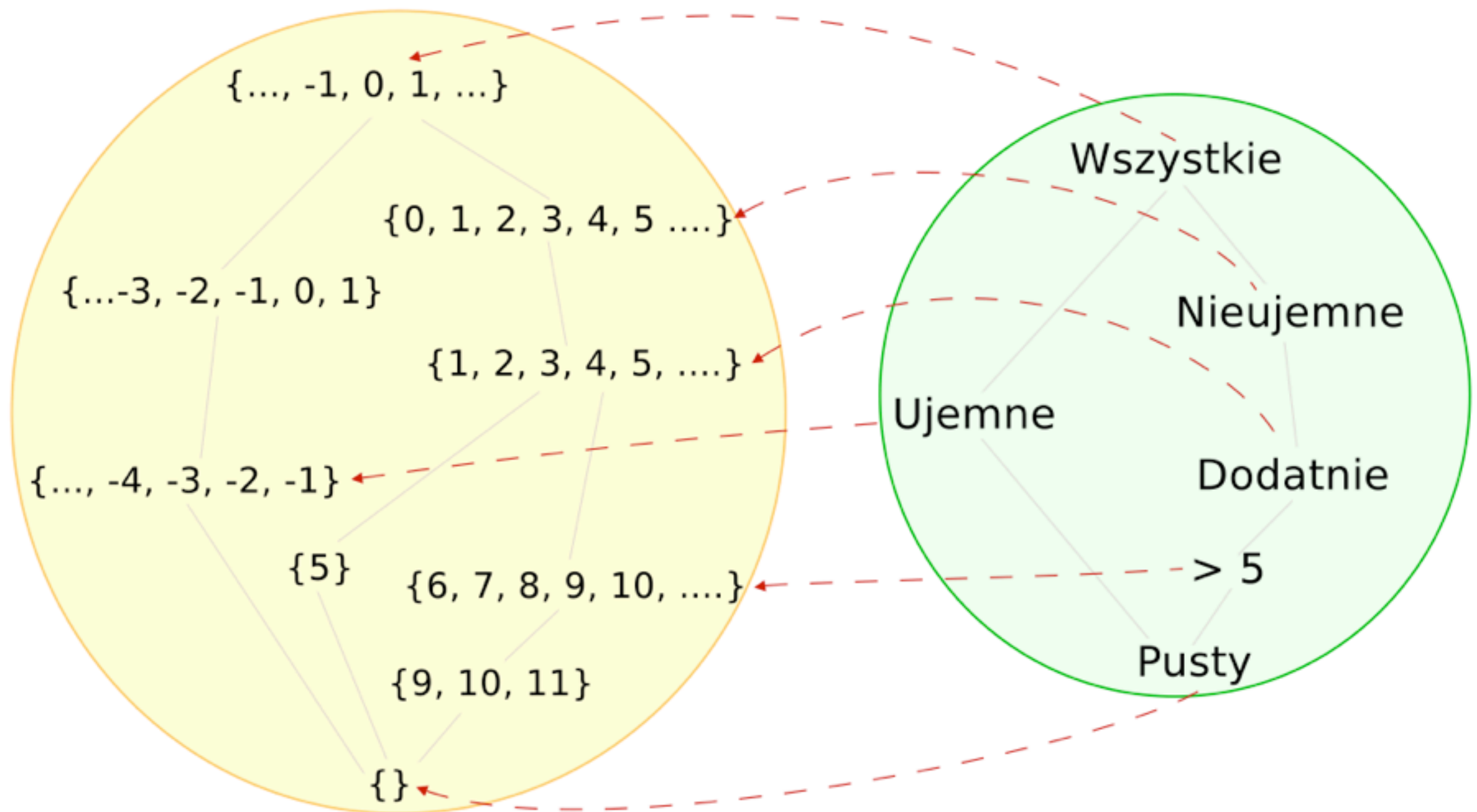
Abstraction function (example)

α maps a set of concrete values to the most exact abstract value

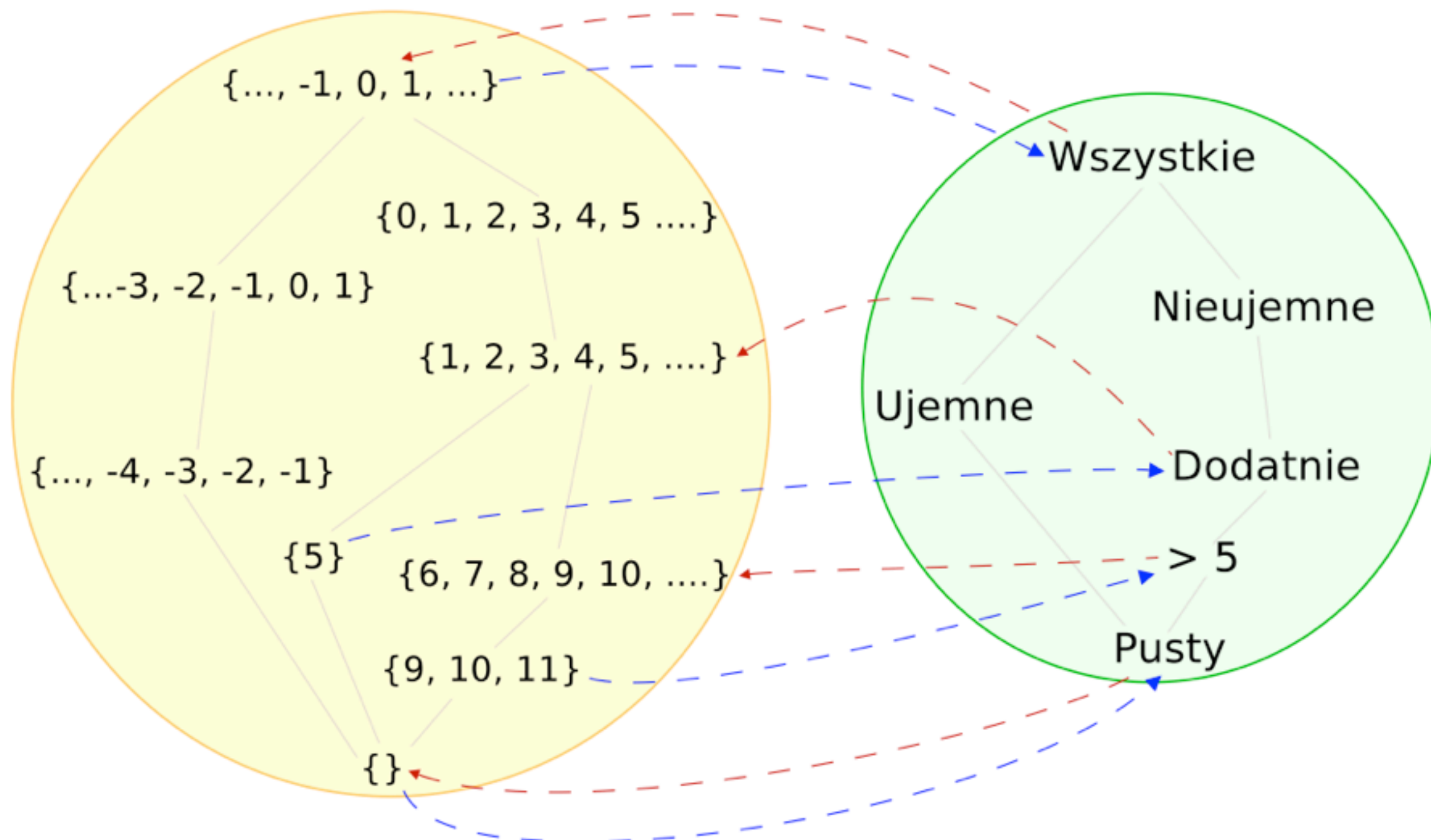


Concretization function (example)

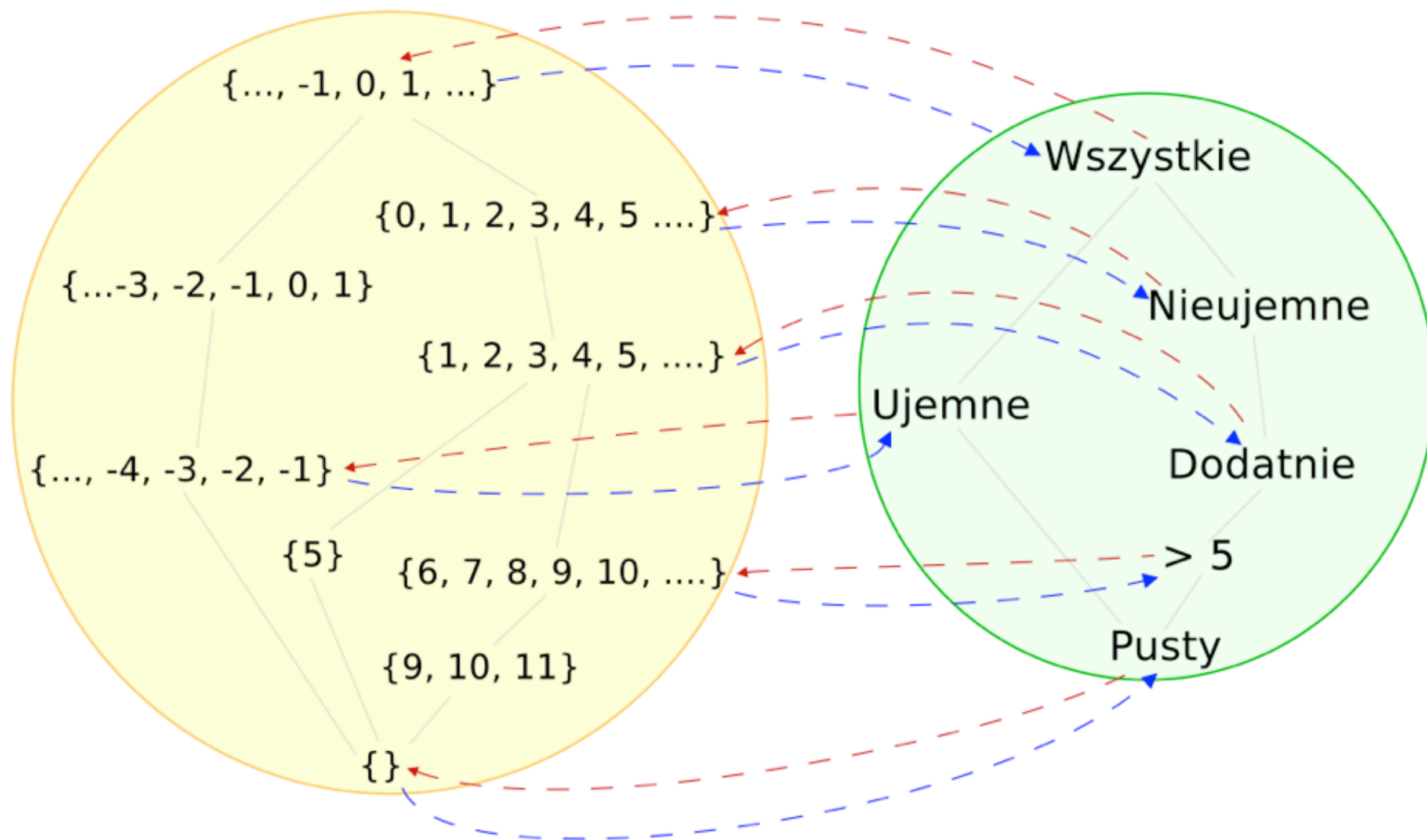
γ maps an abstract value to the set of represented concrete values



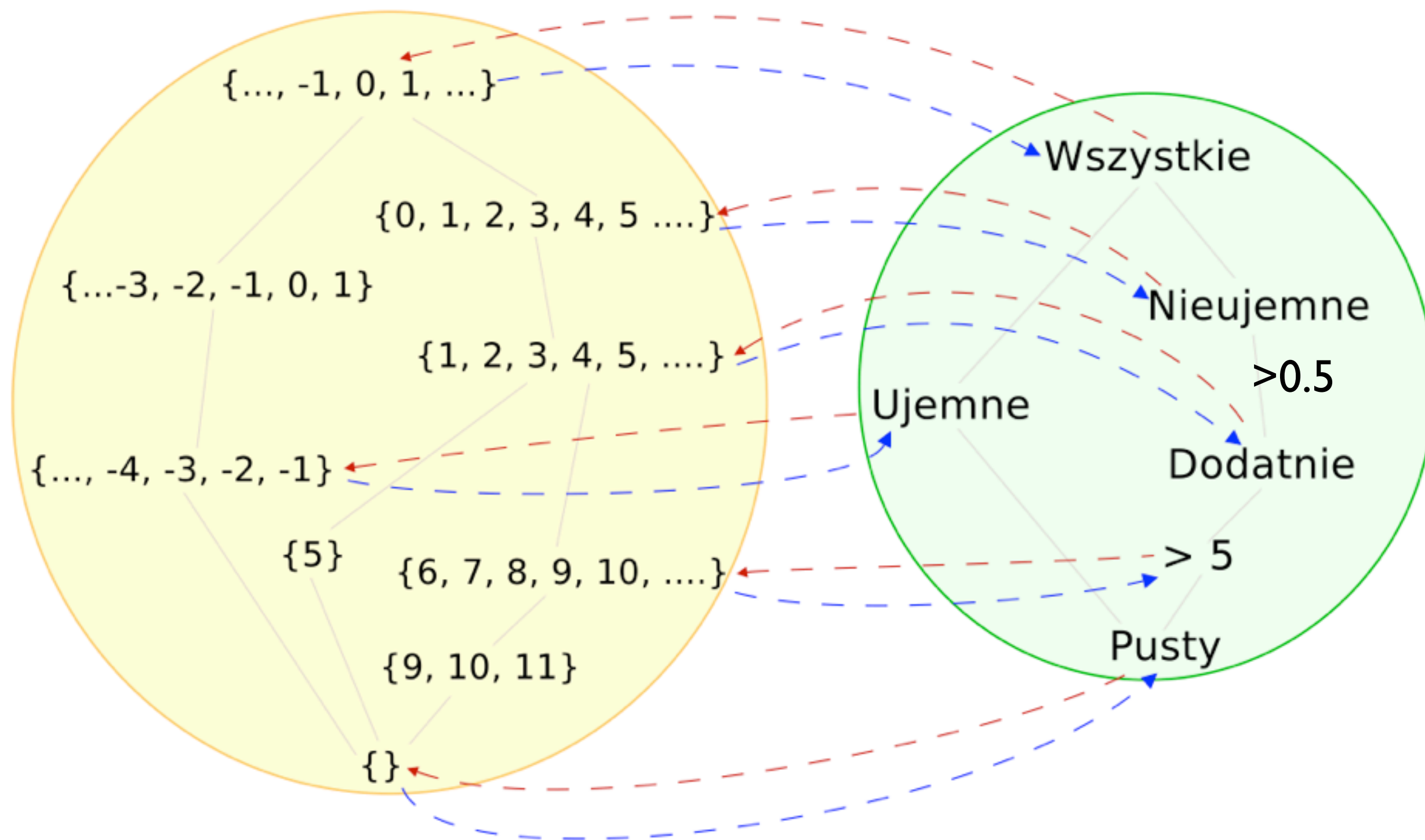
$$x \subseteq \gamma \cdot \alpha(x)$$



$$\alpha \cdot \gamma(a) \leq a$$

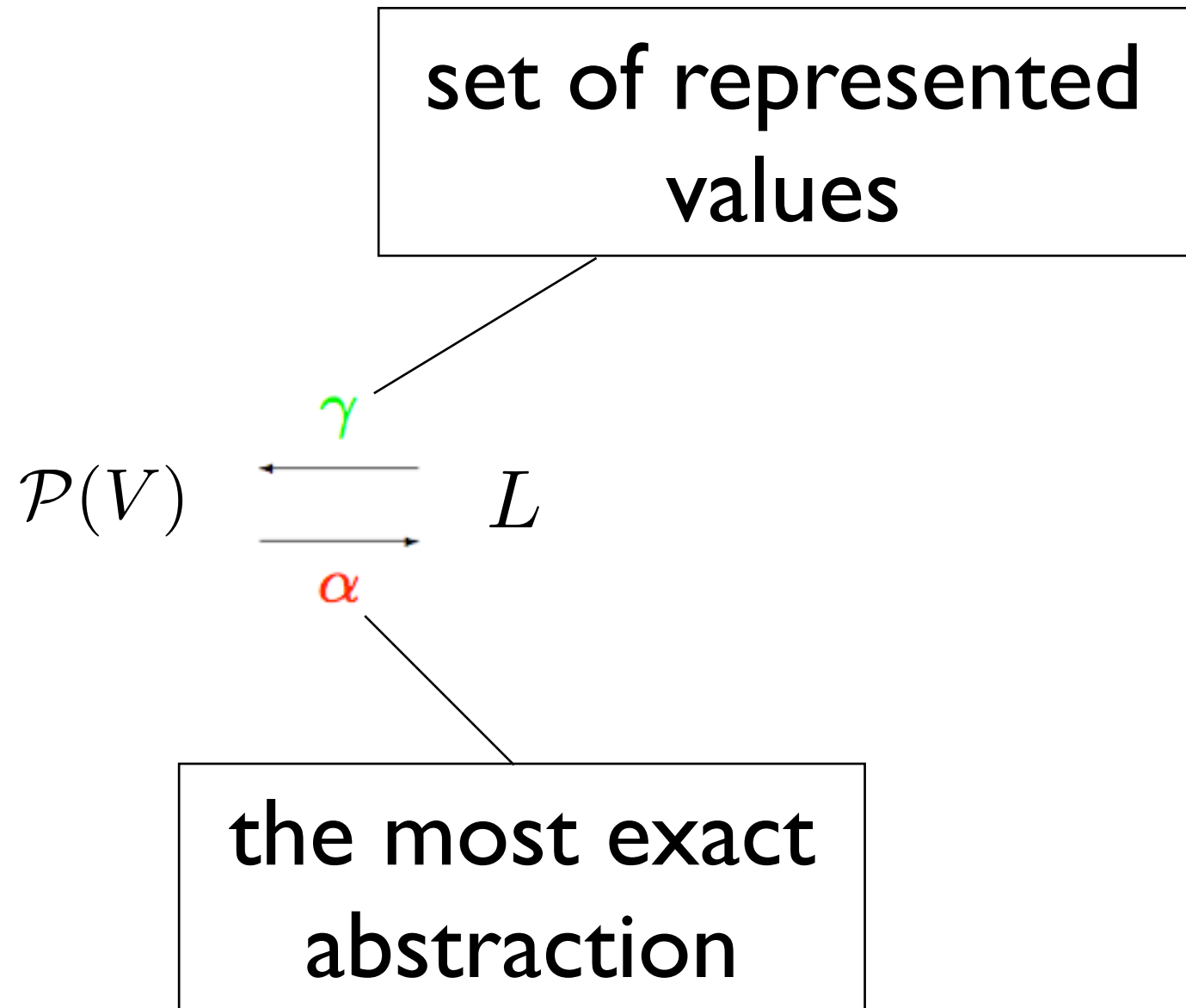


$$\alpha \cdot \gamma(a) \leq a$$



Galois connection

Concrete and abstract domain



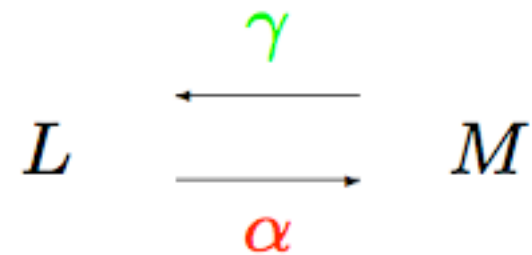
Example

$$\mathcal{P}(\mathbb{Z}) \begin{array}{c} \xleftarrow{\gamma} \\ \xrightarrow{\alpha} \end{array} \text{Intervals}$$

$\alpha(X) =$ the smallest interval containing X

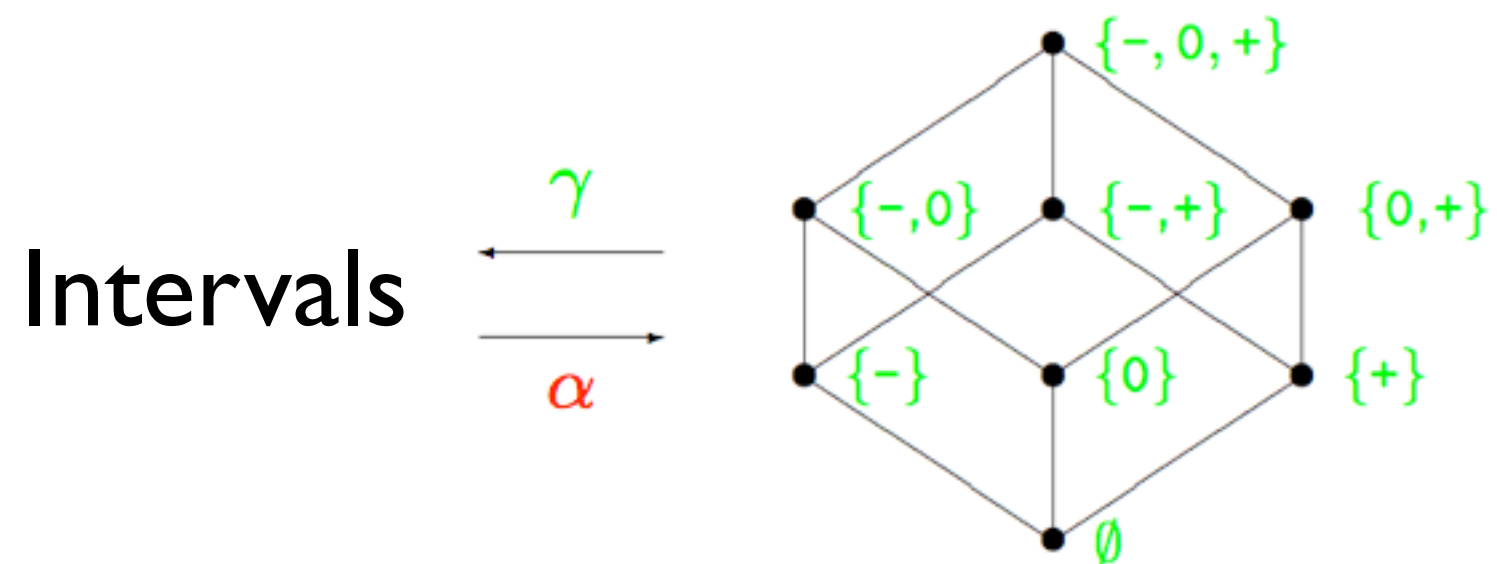
$$\gamma(I) = I$$

Two abstract domains



M is more abstract (less exact) than L

Example



Example

