

# A Gentle Introduction to Static Analysis (2)

Woosuk Lee

CSE 6049 Program Analysis



Hanyang University, Korea

# Static Analysis

A **general** method for **automatic** and **sound approximation** of sw run-time behaviors **before** the execution

- “**before**”: statically, without running sw
- “**automatic**”: sw analyzes sw
- “**sound**”: all possibilities into account
- “**approximation**”: cannot be exact
- “**general**”: for any source language and property
  - ▶ C, C++, C#, F#, Java, JavaScript, ML, Scala, Python, JVM, Dalvik, x86, Excel, etc
  - ▶ “buffer-overflow?”, “memory leak?”, “type errors?”, “x = y at line 2?”, “memory use  $\leq 2K$ ?”, etc

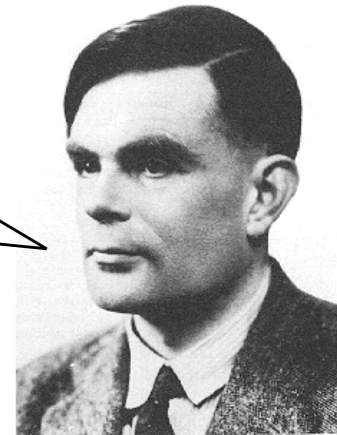
# Abstract Interpretation

- A powerful framework for designing correct static analysis
  - “**framework**” : correct static analysis comes out, reusable
  - “**powerful**” : all static analyses are understood in this framework
  - “**simple**” : prescription is simple
  - “**eye-opening**” : any static analysis is an abstract interpretation

# Why Abstraction?

- Without abstraction,
  - can't capture all possible executions
  - can't terminate

Impossible



Alan Turing

- **Abstraction  $\neq$  omission**

- reality: {2, 4, 6, 8, ... }
- “even number” (abstraction) vs “multiple of 4” (omission)



# Example

---

- Q: What are the possible output values?

```
x = 3;
while (*) {
    x += 2;
}
x -= 1;
print(x);
```

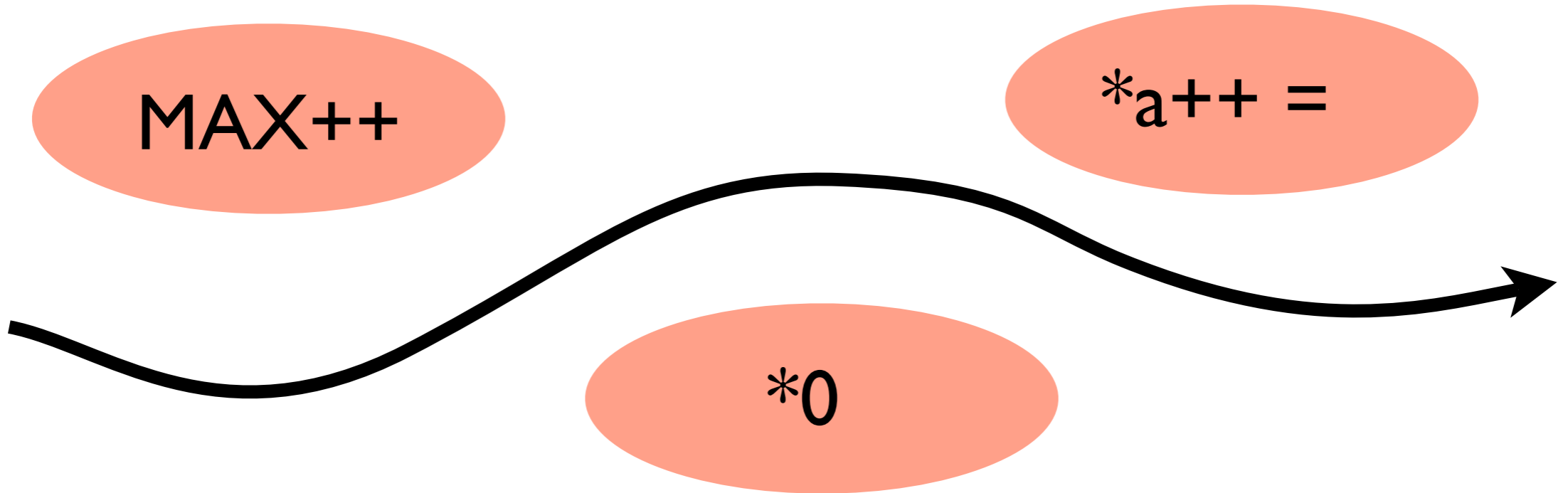
- Concrete interpretation: 2, 4, ... infinitely many possible values
- Abstract interpretation 1: “integers” (coarse)
- Abstract interpretation 2: “positive integers” (precise)
- Abstract interpretation 3: “positive even integers” (more precise)

# Abstraction

MAX++

\*a++ =

\*0

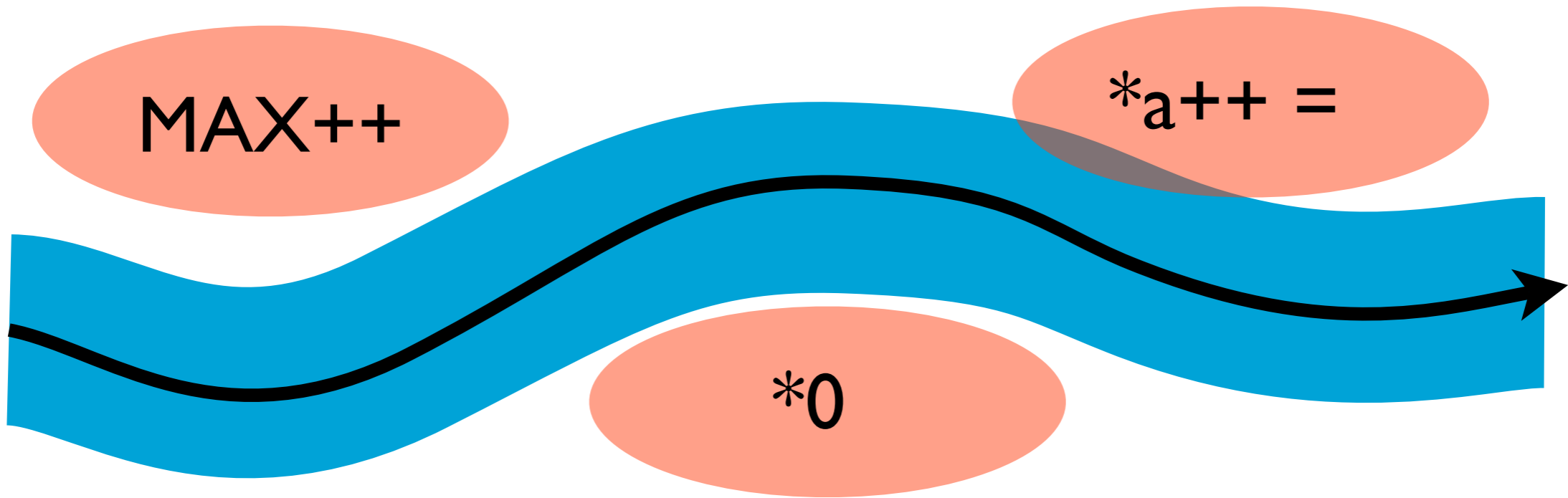


# Abstraction

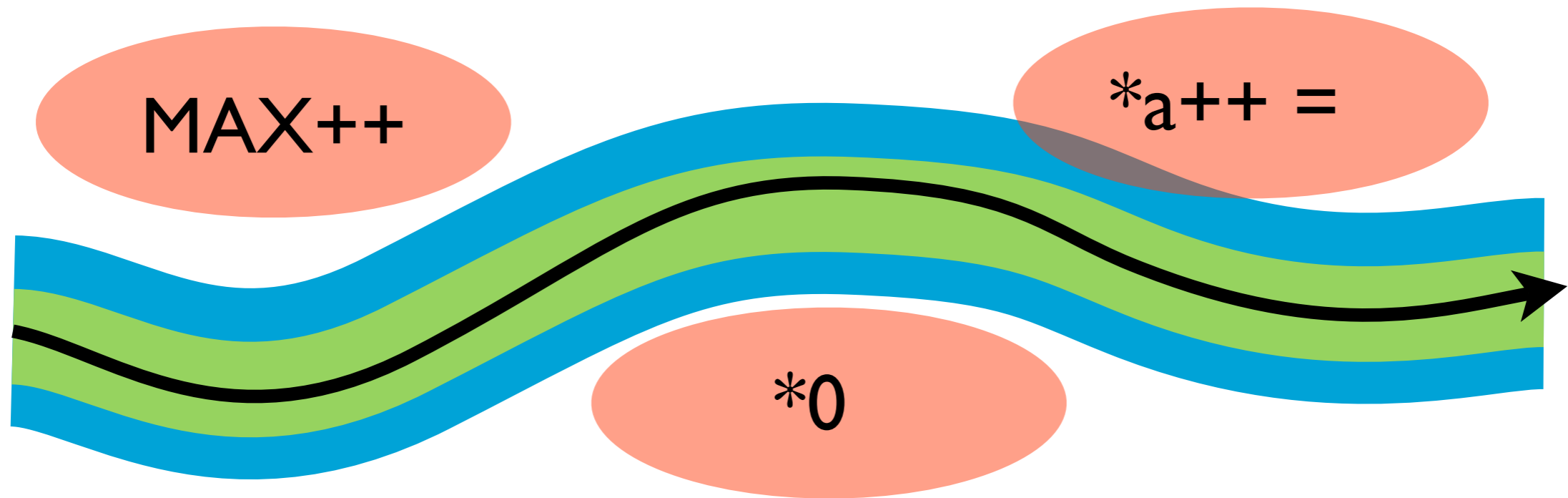
MAX++

\*a++ =

\*0



# Abstraction



# **An Intuitive Explanation of Abstract Interpretation**

# Example Language

---

Initialization with a point that is non-deterministically chosen in a fixed region (e.g.,  $[0,1] \times [0,1]$  square)

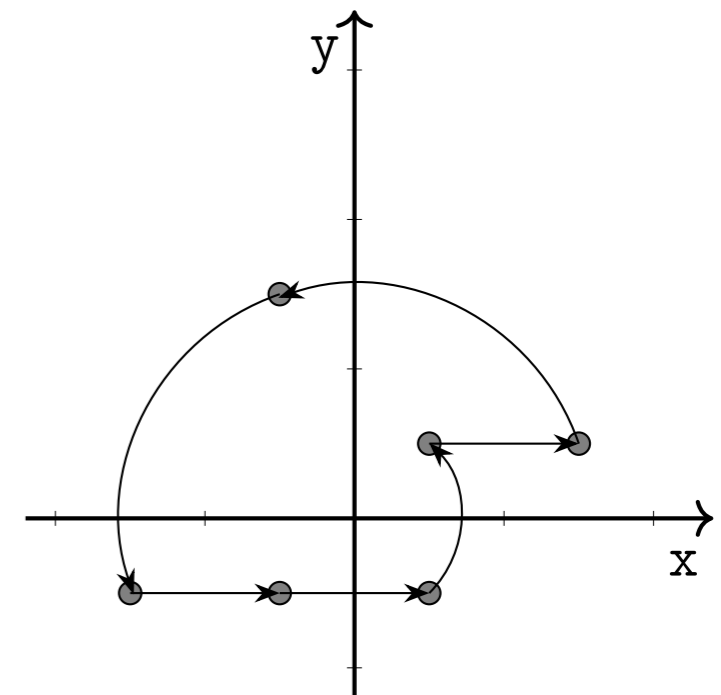
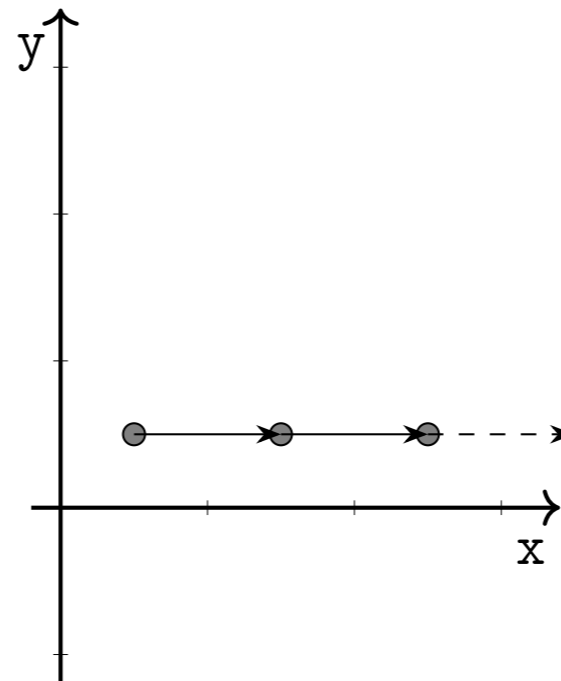
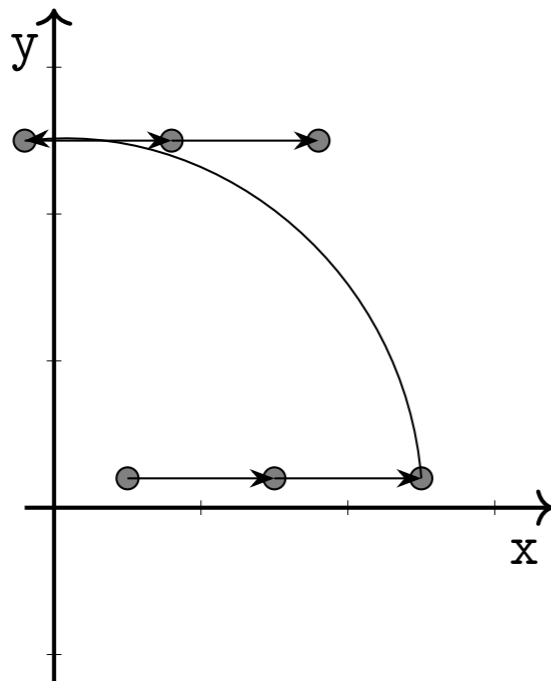
$p ::=$	$\text{init}(\mathcal{R})$	initialization, with a state in $\mathcal{R}$
	$\text{translation}(u, v)$	translation by vector $(u, v)$
	$\text{rotation}(u, v, \theta)$	rotation by center $(u, v)$ and angle $\theta$
	$p ; p$	sequence of operations
	$\{p\} \text{or} \{p\}$	non-deterministic choice
	$\text{iter}\{p\}$	non-deterministic iterations

All programs start with an initialization statement.

# Semantics

## Example (Semantics)

```
init([0, 1] × [0, 1]);  
translation(1, 0);  
iter{  
  {  
    translation(1, 0)  
  }or{  
    rotation(0, 0, 90°)  
  }  
}
```

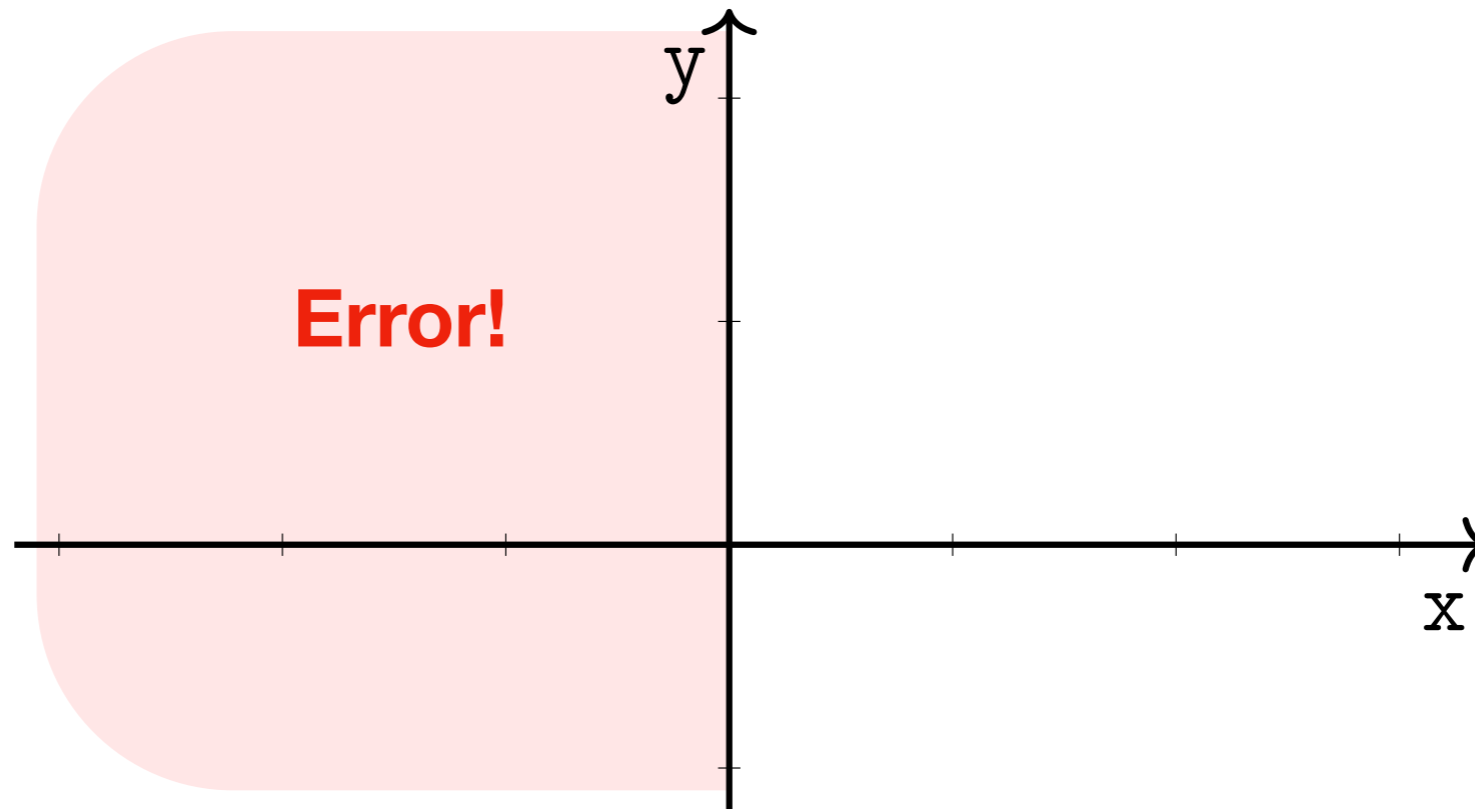


# Analysis Goal Is Safety Property: Reachability

---

Analyze the set of reachable points, to check if the set intersects with a hypothetical error zone:

$$\mathcal{D} = \{(x, y) \mid x < 0\}$$

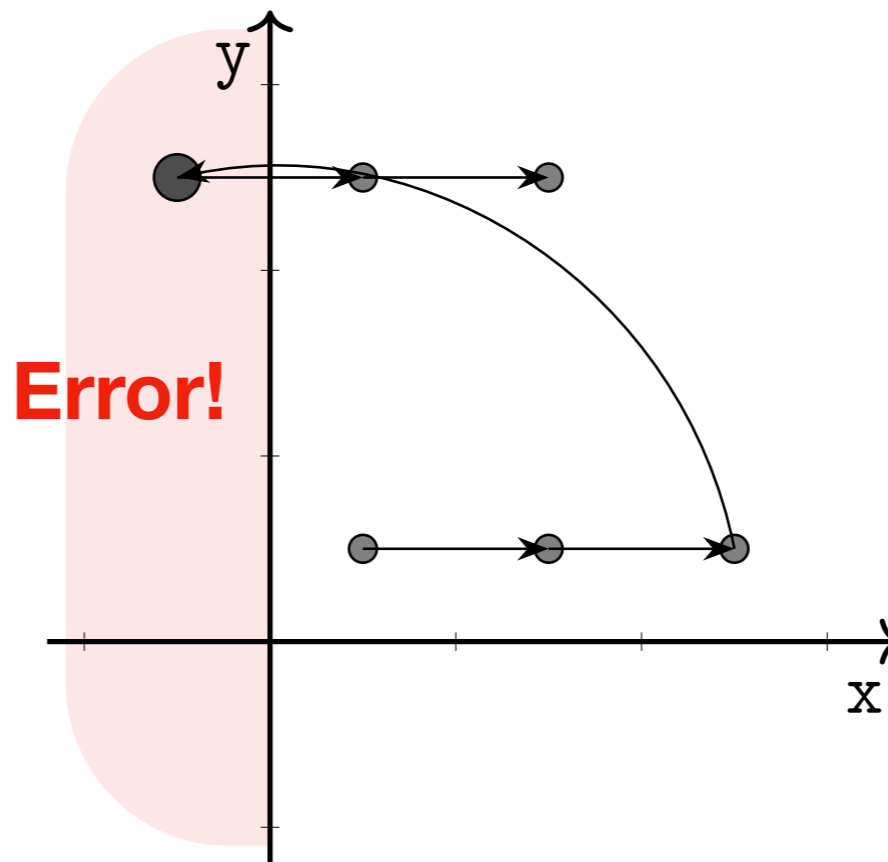




# Correct / Incorrect Executions

---

- Our goal: prove  $\neg \mathcal{D}$

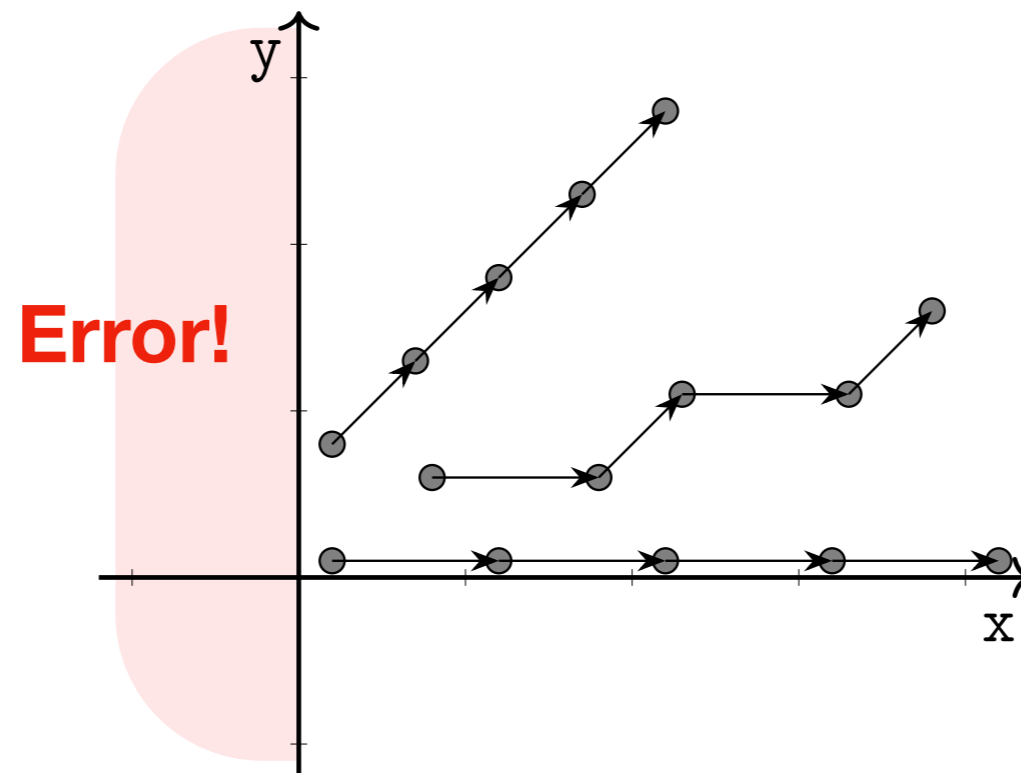


(a) An incorrect execution

# An Example Safe Program

## Example

```
init([0, 1] × [0, 1]);  
iter{  
  {  
    translation(1, 0)  
  }or{  
    translation(0.5, 0.5)  
  }  
}
```



# Need for Static Analysis for Proving $\neg\mathcal{D}$

---

- How can we check  $\neg\mathcal{D}$  for any given program?
- Enumeration of all executions does not work!
  - The set of possible initial states is infinite.
  - The length of executions may be infinite.
  - The set of possible series of non-deterministic choices is infinite.

# How to Finitely Over-Approximate the Set of Reachable Points?

---

## Definition (Abstraction)

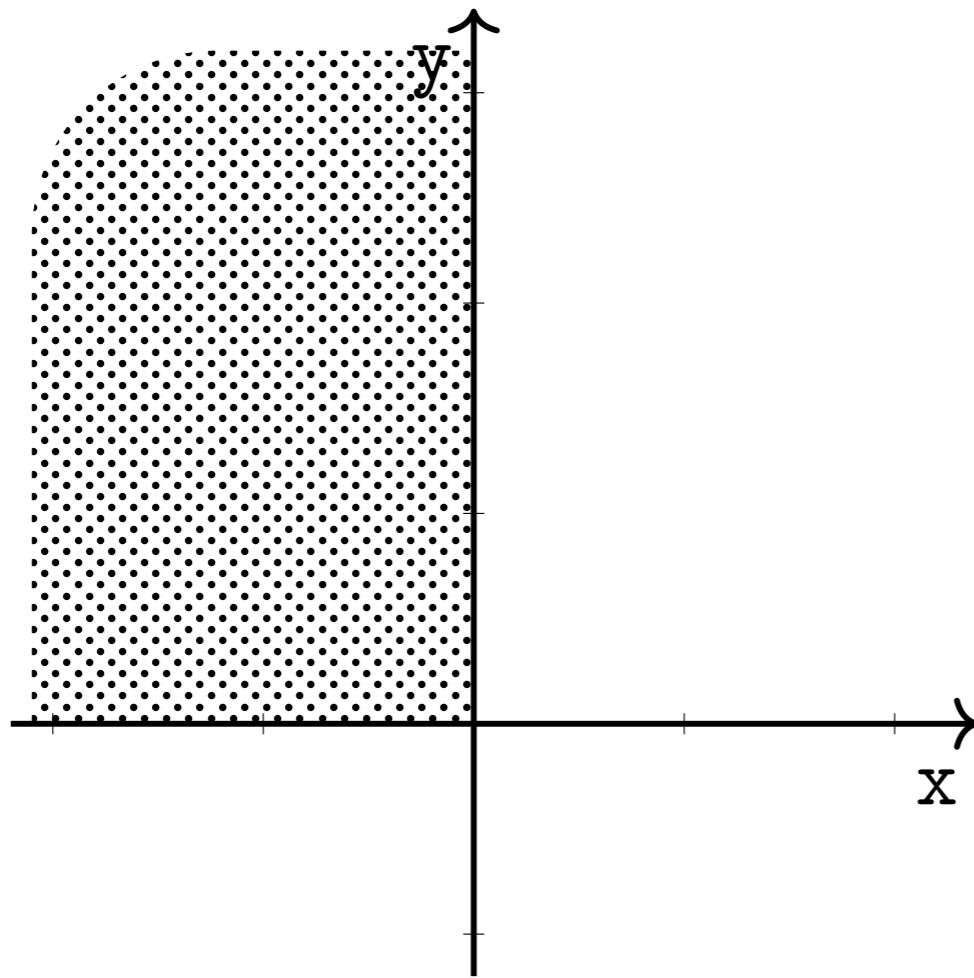
We call *abstraction* a set  $\mathcal{A}$  of logical properties of program states, which are called *abstract properties* or *abstract elements*. A set of abstract properties is called an *abstract domain*.

## Definition (Concretization)

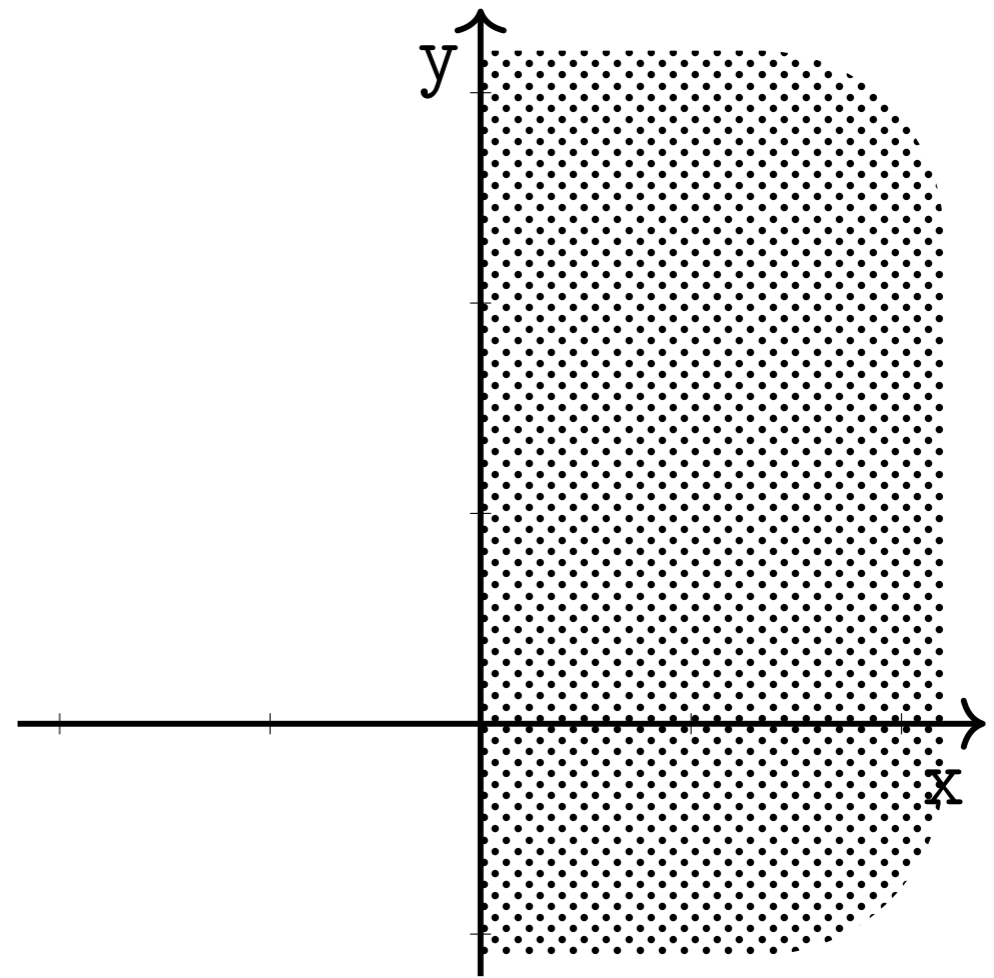
Given an abstract element  $a$  of  $\mathcal{A}$ , we call *concretization* the set of program states that satisfy it. We denote it by  $\gamma(a)$ .

# Abstraction Example I: Sign Abstraction

---



(c) Concretization of  $[x \leq 0, y \geq 0]$



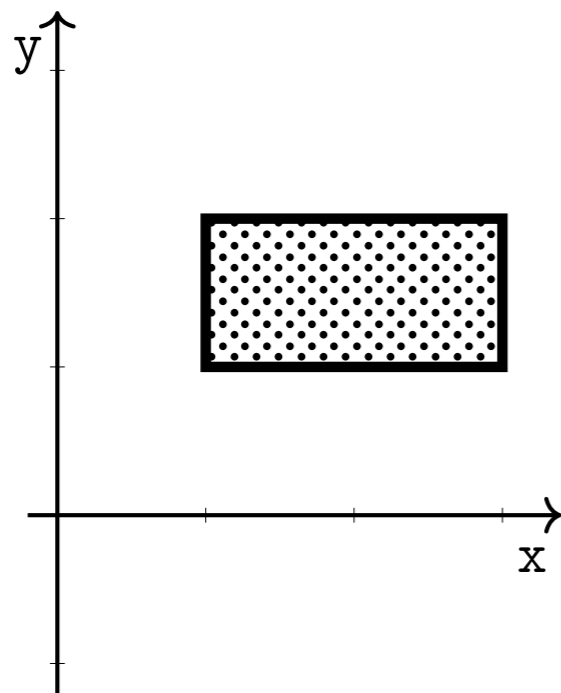
(d) Concretization of  $[x \geq 0, y \geq 0]$

Figure: Signs abstraction

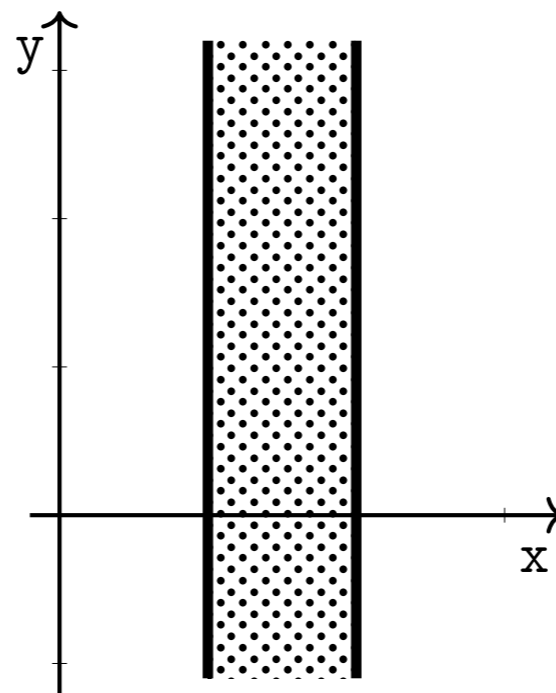
# Abstraction Example 2: Interval Abstraction

---

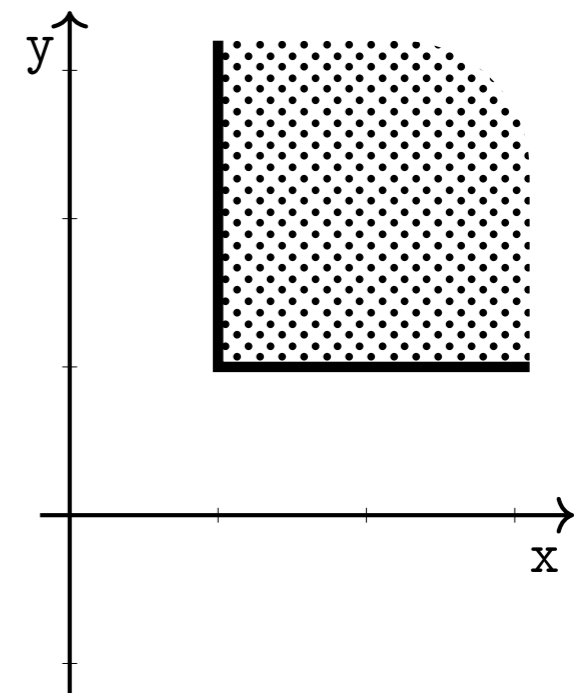
The abstract elements: conjunctions of non-relational inequality constraints:  $c_1 \leq x \leq c_2, c'_1 \leq y \leq c'_2$



(a) Concretization of  $[1 \leq x \leq 3, 1 \leq y \leq 2]$



(b) Concretization of  $[1 \leq x \leq 2]$

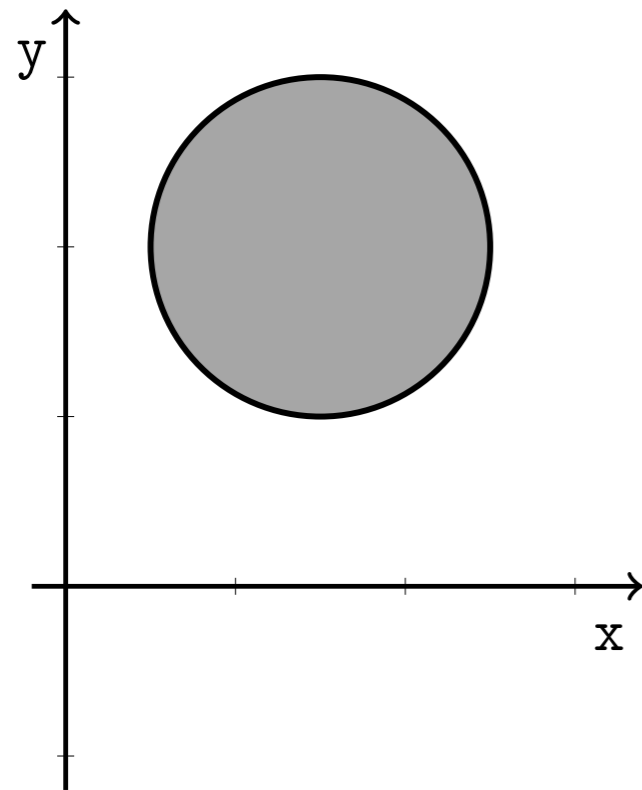


(c) Concretization of  $[1 \leq x, 1 \leq y]$

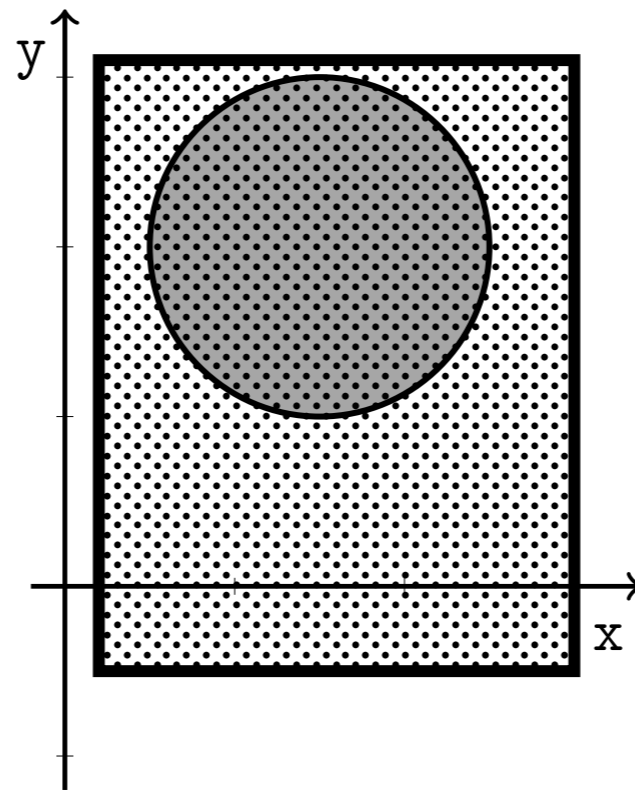
Figure: Intervals abstraction

# Best Abstraction

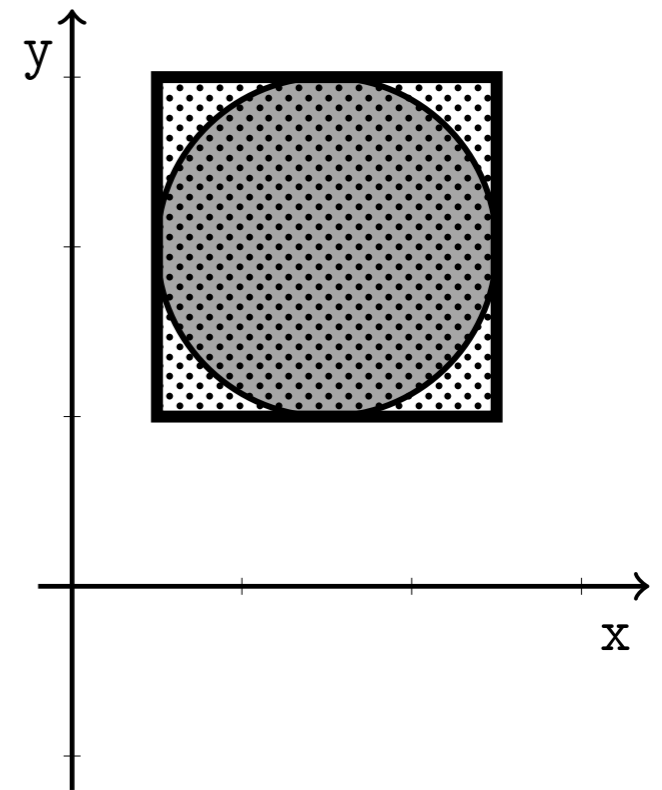
---



(a) A concrete set



(b) Abstractions



(c) Best abstraction

**Figure 2.7**  
Best abstraction

# Best Abstraction

---

- We say  $a$  is the best abstraction of the concrete set  $S$  iff
  - $S \subseteq \gamma(a)$ , and
  - for any  $a'$  such that  $S \subseteq \gamma(a')$ ,  $a'$  is a coarser abstraction than  $a$ .



# Abstraction Example 3: Convex Polyhedra Abstraction

The abstract elements: conjunctions of linear inequality constraints:

$$c_1x + c_2y \leq c_3$$

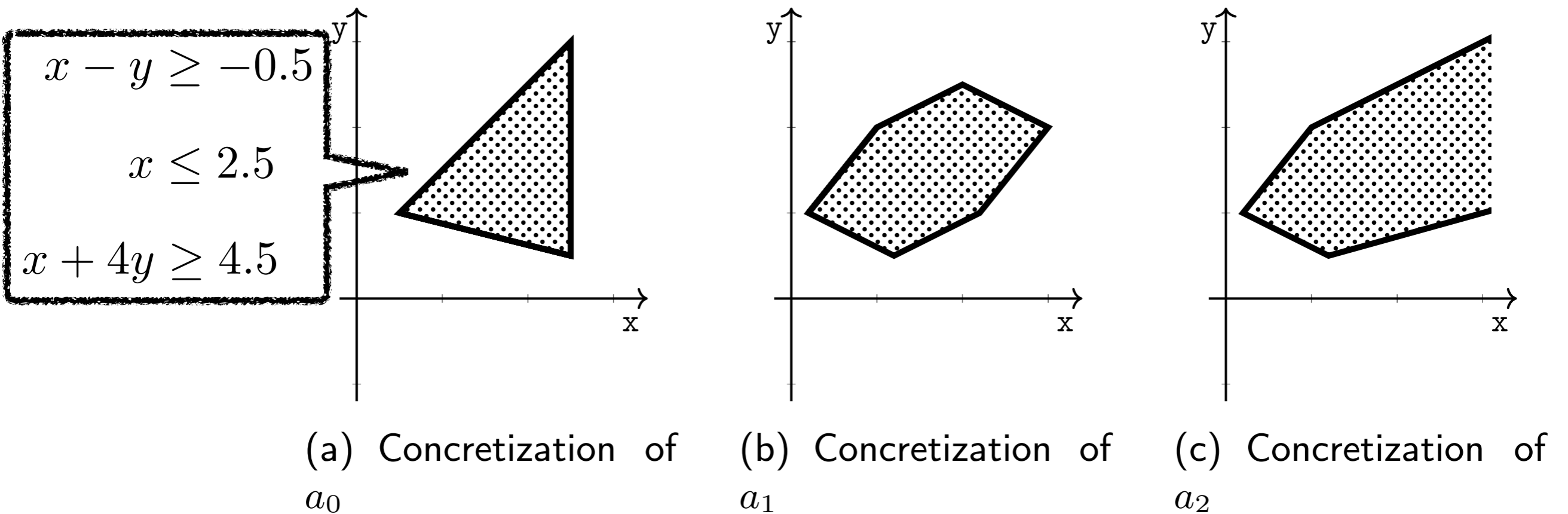


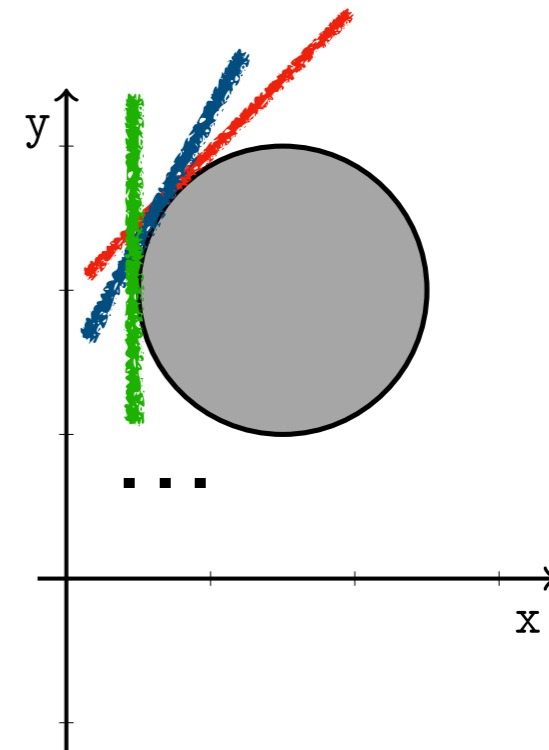
Figure: Convex polyhedra abstraction

# Best Abstraction is Not Always Obtainable

---

- Computing the best abstraction is expensive in general, or sometimes even impossible.

- In case of the diameter, there is no best abstraction since it requires infinitely many linear inequalities.



(a) A concrete set

- Thus in practice, we often use abstractions as precise as possible but may not be the best.

# Reachable States of the Example Program

## Example

```
init([0, 1] × [0, 1]);  
iter{  
  {  
    translation(1, 0)  
  }or{  
    translation(0.5, 0.5)  
  }  
}
```

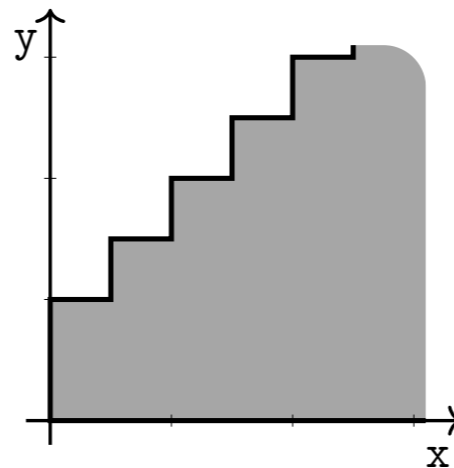
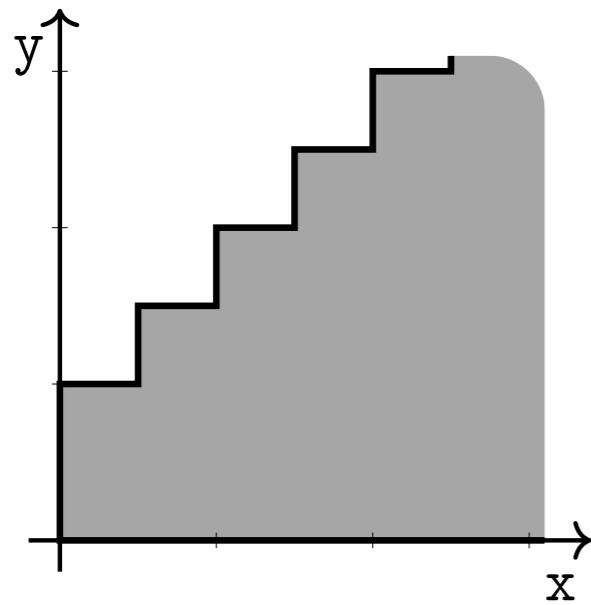


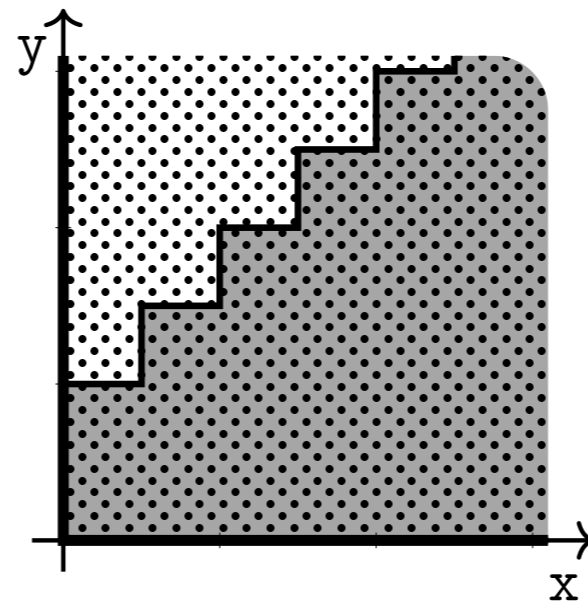
Figure: Reachable states

# Abstractions of the Semantics of the Example Program

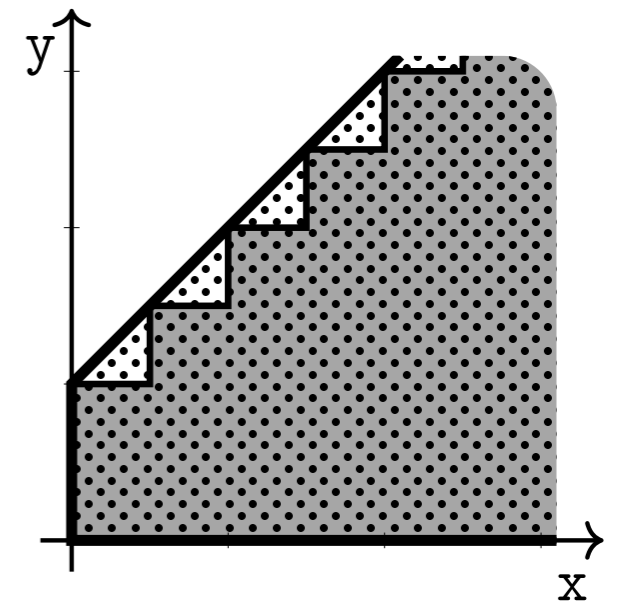
---



(a) Reachable states



(b) Intervals abstraction



(c) Convex polyhedra abstraction

Figure: Program's reachable states and abstraction

# Abstract Semantics Computation

---

Recall the example language

$p ::=$	$\text{init}(\mathcal{R})$	initialization, with a state in $\mathcal{R}$
	$\text{translation}(u, v)$	translation by vector $(u, v)$
	$\text{rotation}(u, v, \theta)$	rotation defined by center $(u, v)$ and angle $\theta$
	$p ; p$	sequence of operations
	$\{p\} \text{or} \{p\}$	non-deterministic choice
	$\text{iter}\{p\}$	non-deterministic iterations

## Approach

A sound analysis for a program is constructed by computing sound abstract semantics of the program's components.

# Sound Analysis Function for the Example Language

---

- Input: a program  $p$  and an abstract area  $a$  (pre-state)
- Output: an abstract area  $a'$  (post-state)

## Definition (sound analysis)

An analysis is sound if and only if **it captures the real executions of the input program.**

If an execution of  $p$  moves a point  $(x, y)$  to point  $(x', y')$ ,  
then for all abstract element  $a$  such that  $(x, y) \in \gamma(a)$ ,  
$$(x', y') \in \gamma(\text{analysis}(p, a))$$

# Sound Analysis Function as a Diagram

---

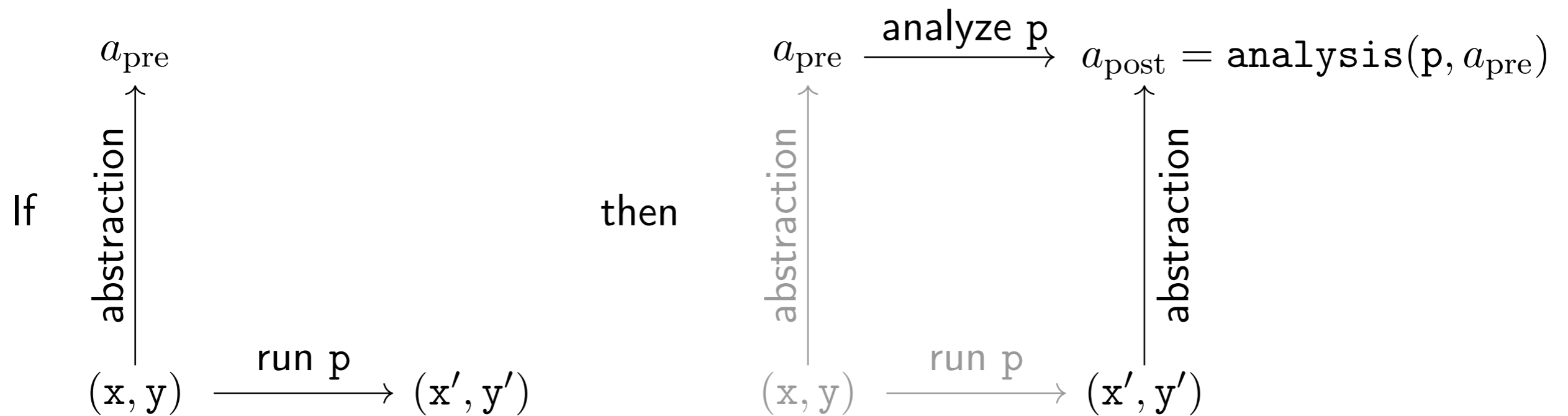
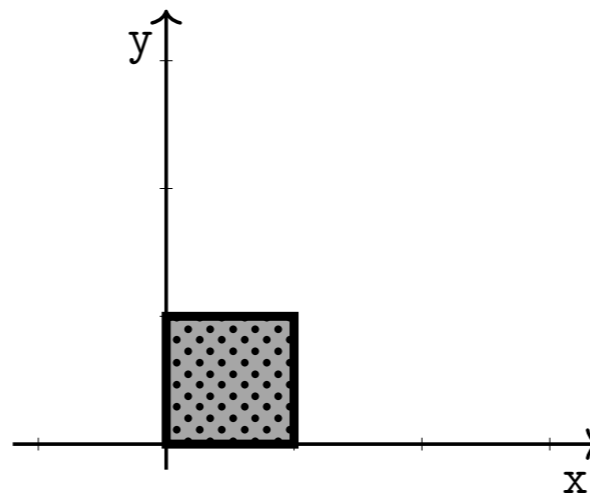


Figure: Sound analysis of a program  $p$

# Abstract Semantics Computation: $\text{init}(\mathcal{R})$

---

- Select, if any, the best abstraction of the region  $\mathcal{R}$ .
- For the example program with the intervals or convex polyhedra abstract domains, analysis of  $\text{init}([0, 1] \times [0, 1])$  is

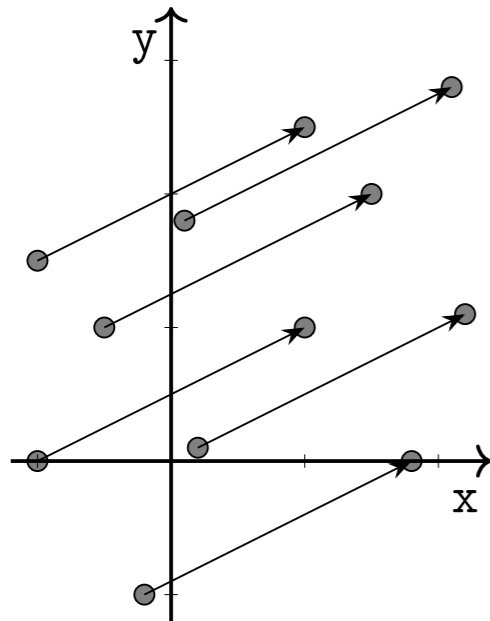


$\text{analysis}(\text{init}(\mathcal{R}), a) = \text{best abstraction of the region } \mathcal{R}$

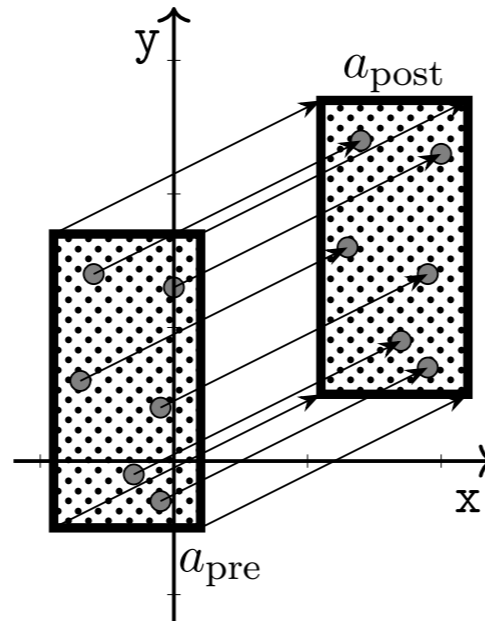


# Abstract Semantics Computation: translation( $u, v$ )

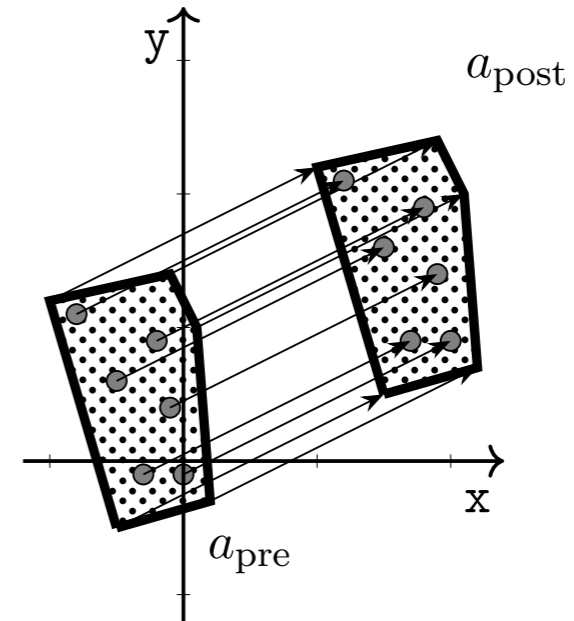
---



(a) Concrete semantics



(b) Intervals

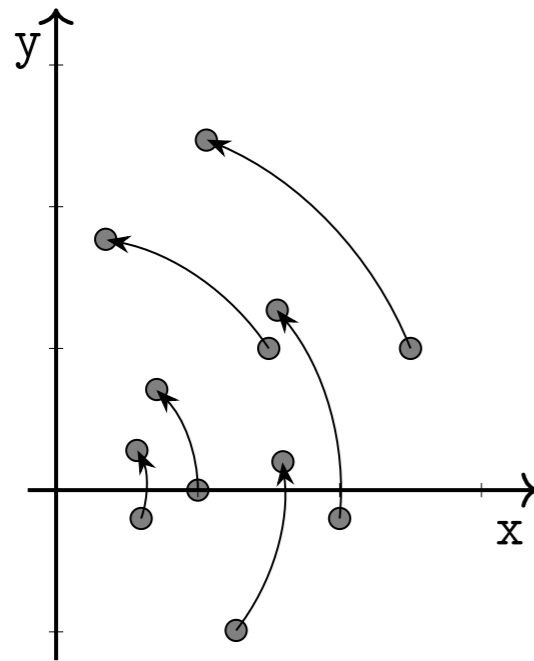


(c) Convex polyhedra

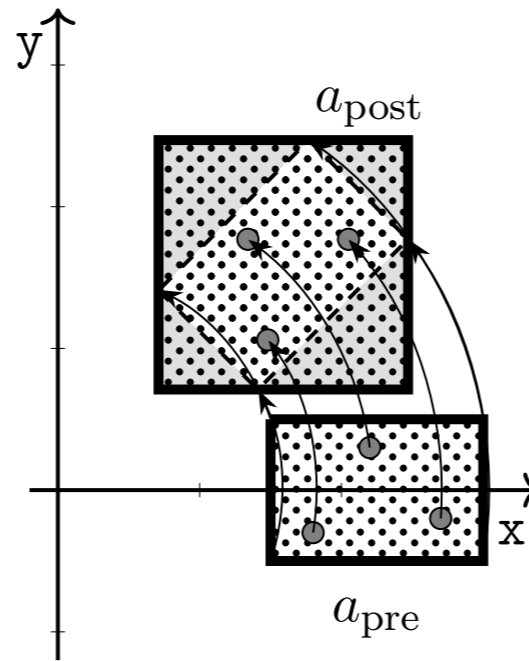
$\text{analysis}(\text{translation}(u, v), a) = \begin{cases} \text{return an abstract state that contains} \\ \text{the translation of } a \end{cases}$

# Abstract Semantics Computation: $\text{rotation}(u, v, \theta)$

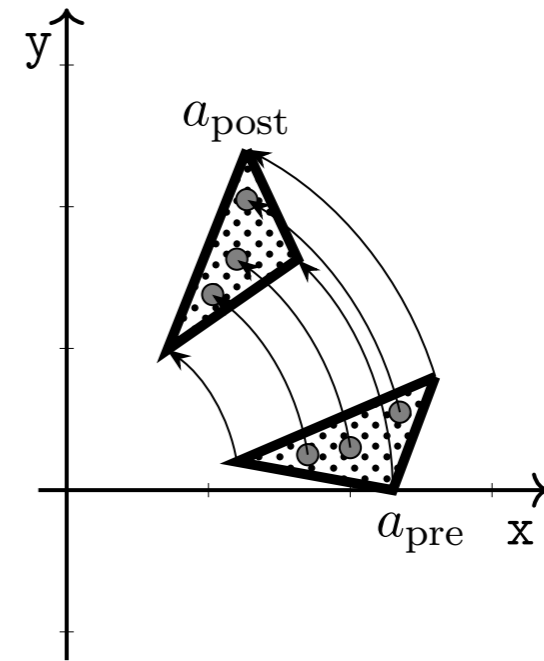
---



(d) Concrete semantics



(e) Intervals



(f) Convex polyhedra

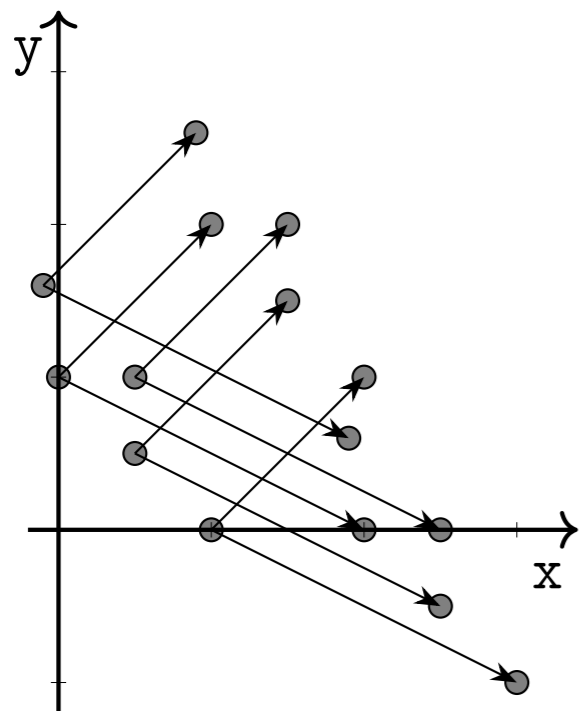
$\text{analysis}(\text{rotation}(u, v, \theta), a) = \begin{cases} \text{return an abstract state that contains} \\ \text{the rotation of } a \end{cases}$

# Abstract Semantics Computation: $p_0; p_1$

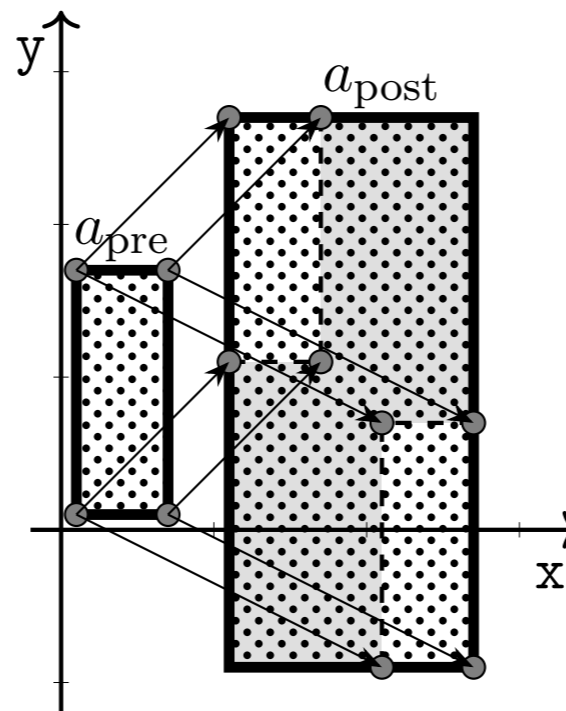
---

$$\text{analysis}(p_0; p_1, a) = \text{analysis}(p_1, \text{analysis}(p_0, a))$$

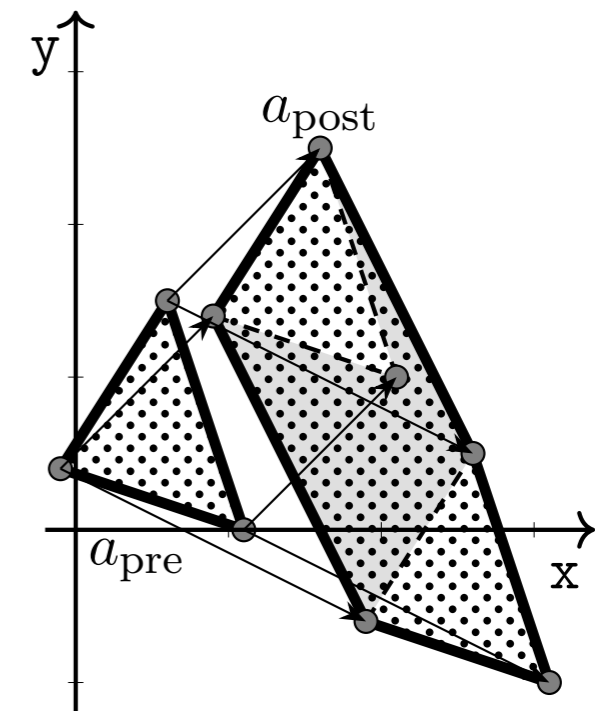
# Abstract Semantics Computation: $\{p\}$ or $\{p\}$



(g) Concrete semantics



(h) Intervals



(i) Convex polyhedra

$$\text{analysis}(\{p_0\} \text{ or } \{p_1\}, a) = \text{union}(\text{analysis}(p_1, a), \text{analysis}(p_0, a))$$

# Abstract Semantics Computation: iter{b}

---

$$p ::= \left\{ \begin{array}{l} \mathbf{iter}\{ \\ \quad b \\ \} \end{array} \right. \equiv \begin{array}{l} \{\} \\ \mathbf{or}\{b\} \\ \mathbf{or}\{b;b\} \\ \mathbf{or}\{b;b;b\} \\ \mathbf{or}\{b;b;b;b\} \\ \vdots \end{array}$$

# Abstract Semantics Computation: iter{p}

---

program  $p_0$  is  $\{\}$   
program  $p_1$  is  $\{\}$  **or**  $\{b\}$   
program  $p_2$  is  $\{\}$  **or**  $\{b\}$  **or**  $\{b; b\}$   
program  $p_3$  is  $\{\}$  **or**  $\{b\}$  **or**  $\{b; b\}$  **or**  $\{b; b; b\}$   
 $\vdots$

$p_{k+1}$  is equivalent to  $p_k$  **or**  $\{p_k; b\}$

**Therefore,**

$\text{analysis}(p_{k+1}, a) = \text{union}(\text{analysis}(p_k, a), \text{analysis}(b, \text{analysis}(p_k, a)))$

# Abstract Semantics Computation: iter{p}

---

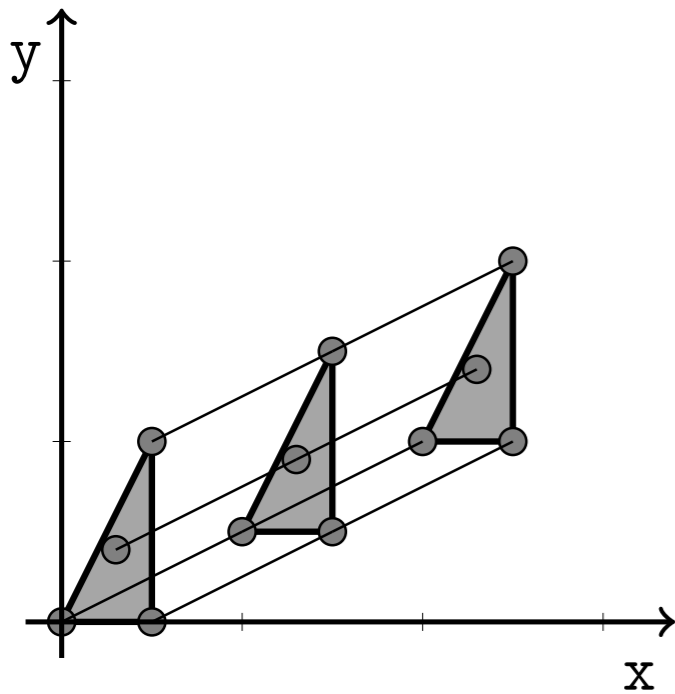
$$\text{analysis}(\mathbf{iter}\{p\}, a) = \left\{ \begin{array}{l} R \leftarrow a; \\ \text{repeat} \\ \quad T \leftarrow R; \\ \quad R \leftarrow \text{union}(R, \text{analysis}(p, R)) \\ \text{until } \text{inclusion}(R, T) \\ \text{return } T; \end{array} \right.$$

operator `inclusion` returns **true** only when it succeeds checking inclusion

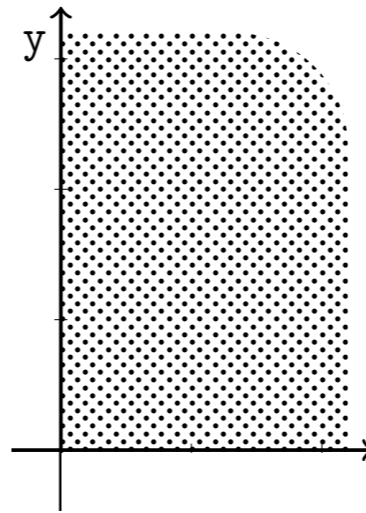
# Abstract Semantics Computation: $\text{iter}\{p\}$

## Example: Sign Abstraction

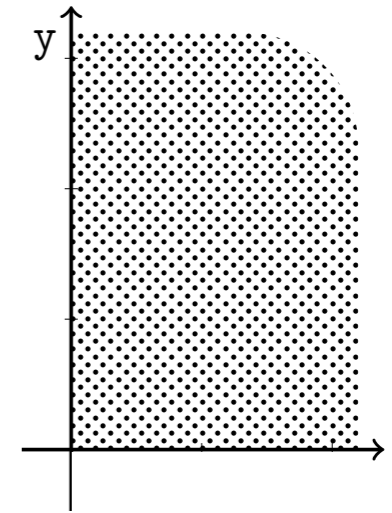
```
init({(x, y) | 0 ≤ y ≤ 2x and x ≤ 0.5});  
iter{  
  translation(1, 0.5)  
}
```



(a) Concrete semantics



(b) Analysis of  $p_0$  (0 iteration)



(c) Analysis of  $p_1$  (up to 1 iteration)

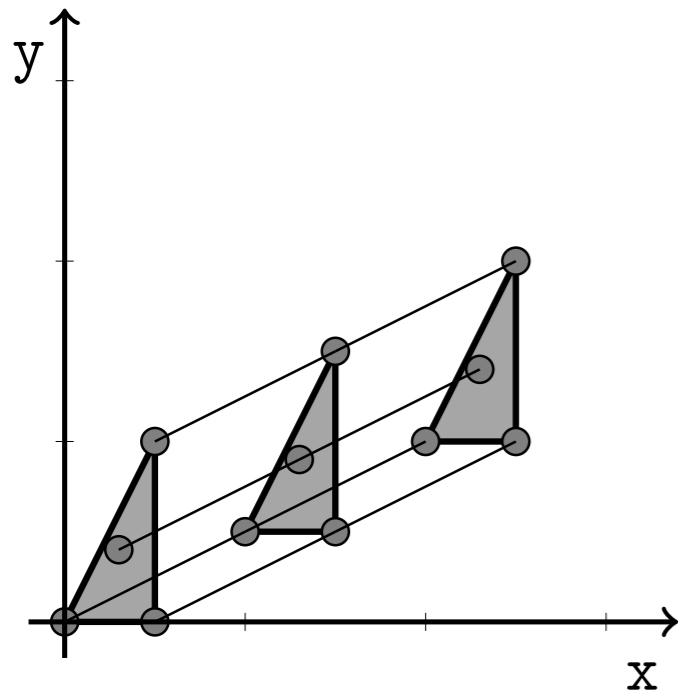
**No change. Done!**



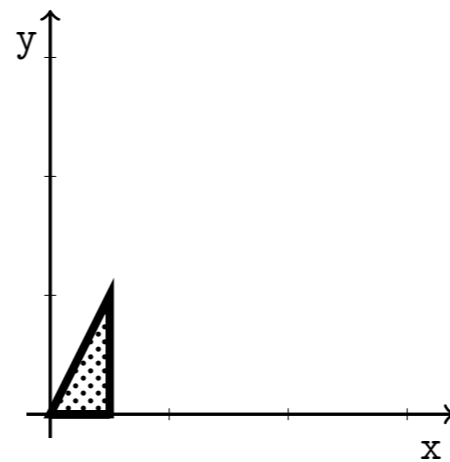
# Abstract Semantics Computation: iter{p}

## Example: Convex Polyhedra

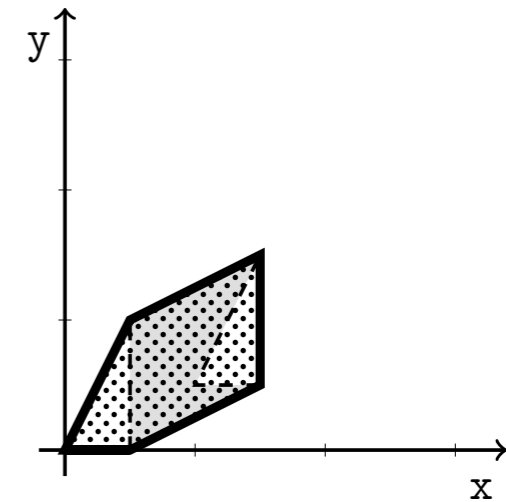
```
init({(x, y) | 0 ≤ y ≤ 2x and x ≤ 0.5});  
iter{  
  translation(1, 0.5)  
}
```



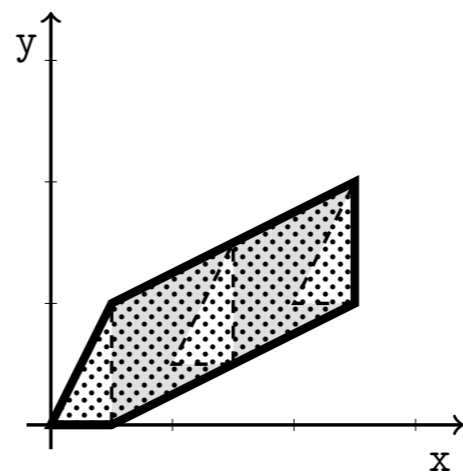
(a) Concrete semantics



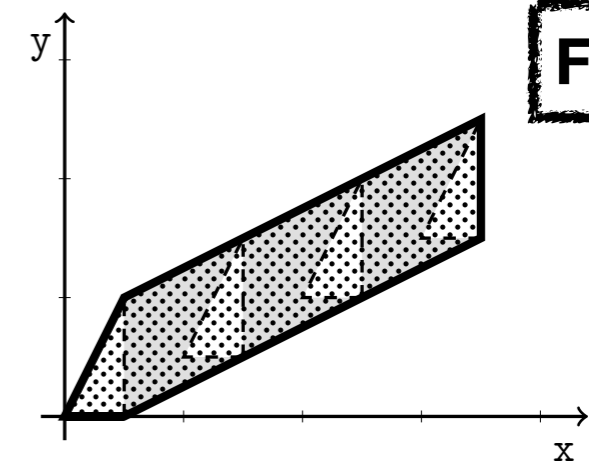
(b) Analysis of  $p_0$  (0 iteration)



(c) Analysis of  $p_1$  (up to 1 iteration)



(d) Analysis of  $p_2$  (up to 2 iterations)



(e) Analysis of  $p_3$  (up to 3 iterations)

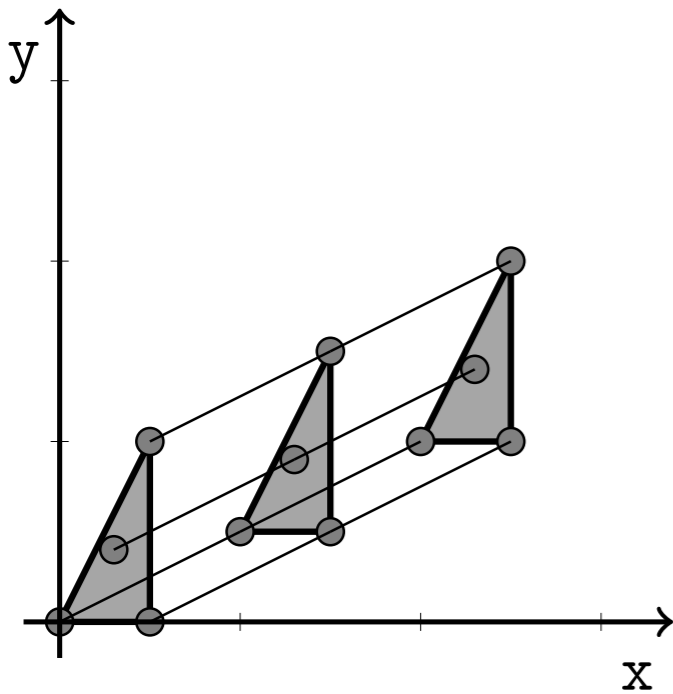
**Forever!**

...

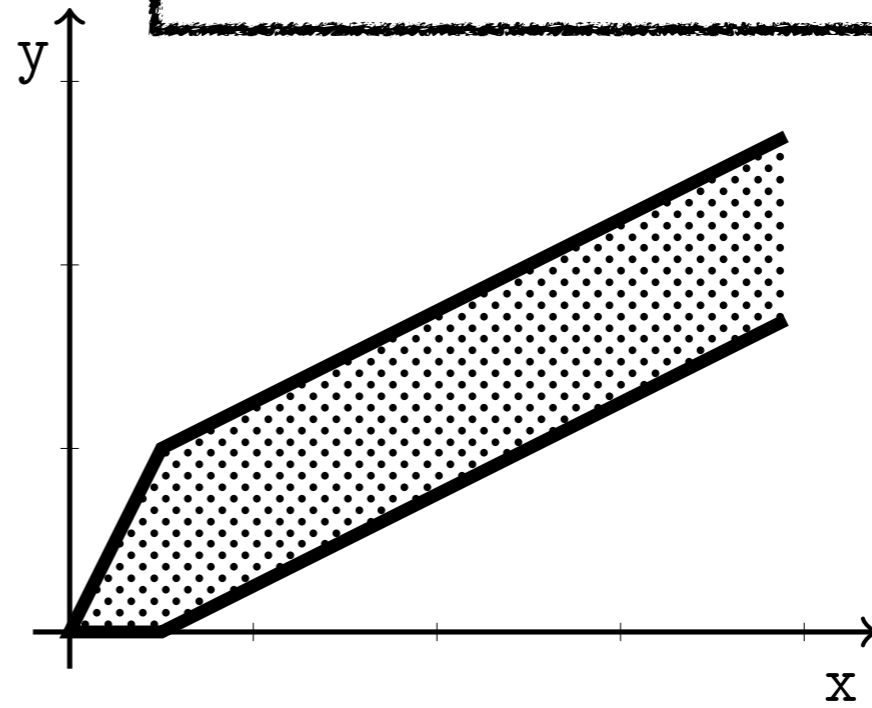
# Abstract Semantics Computation: $\text{iter}\{p\}$

## Example: Convex Polyhedra

```
init({(x, y) |  $0 \leq y \leq 2x$  and  $x \leq 0.5$ });  
iter{  
  translation(1, 0.5)  
}
```



(a) Concrete semantics



(f) Expected result

# Widening

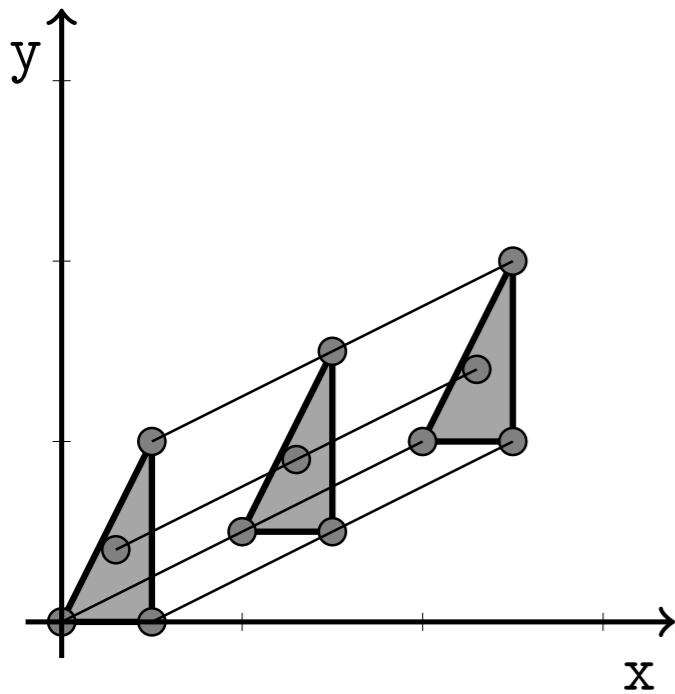
---

- To ensure termination of the analysis, we need to *enforce* the convergence of the iterations.
- In case of convex polyhedra
  - An abstract element = (finitely many) inequalities
  - If we decrease the number of inequalities at each iteration, it will eventually terminate.

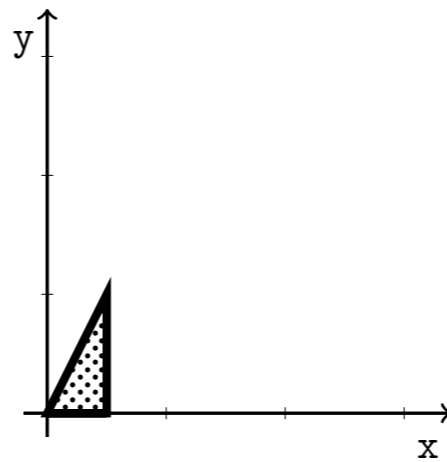
# Abstract Semantics Computation: iter{p}

## Example: Convex Polyhedra

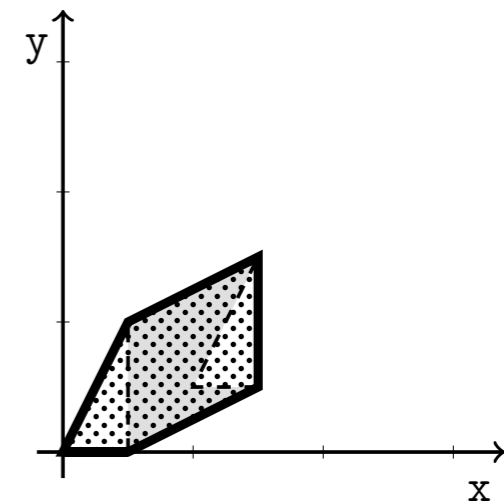
```
init({(x, y) | 0 ≤ y ≤ 2x and x ≤ 0.5});  
iter{  
  translation(1, 0.5)  
}
```



(a) Concrete semantics



(b) Analysis of  $p_0$  (0 iteration)



(c) Analysis of  $p_1$  (up to 1 iteration)

$$\begin{array}{l} 0 \leq y \\ y \leq 2x \\ x \leq 0.5 \end{array}$$

$$\begin{array}{l} 0 \leq y \\ y \leq 2x \\ x \leq 1.5 \\ y \leq 0.5x + 1 \\ y \geq 0.5x - 1 \end{array}$$

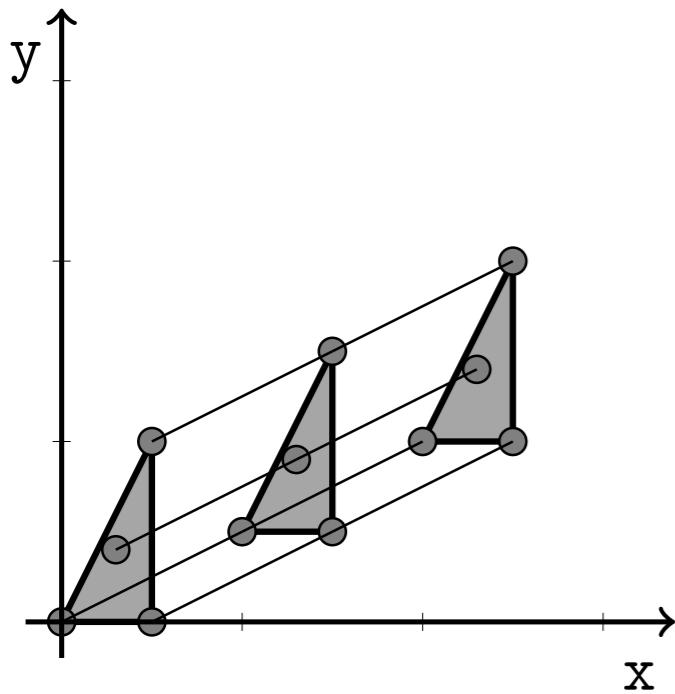
# Abstract Semantics Computation: iter{p}

## Example: Convex Polyhedra

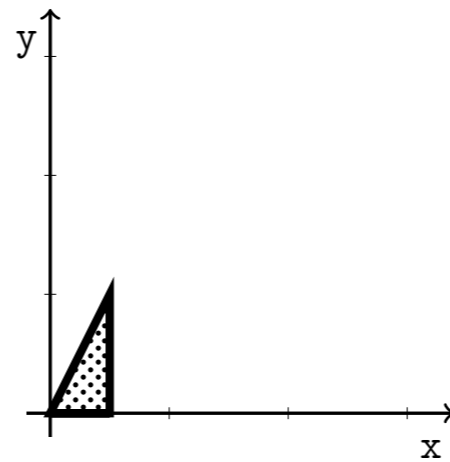
```

init({(x, y) | 0 ≤ y ≤ 2x and x ≤ 0.5});
iter{
  translation(1, 0.5)
}

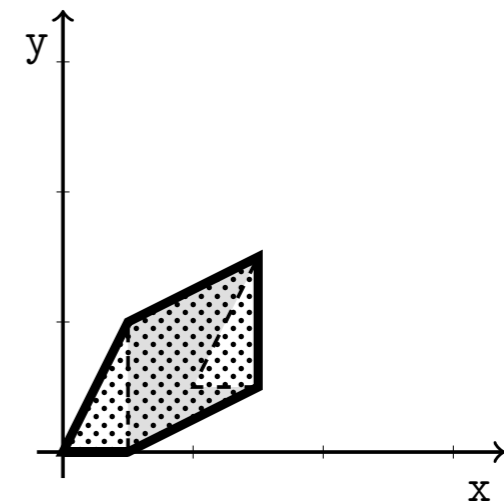
```



(a) Concrete semantics



(b) Analysis of  $p_0$  (0 iteration)



(c) Analysis of  $p_1$  (up to 1 iteration)

Remained

Modified

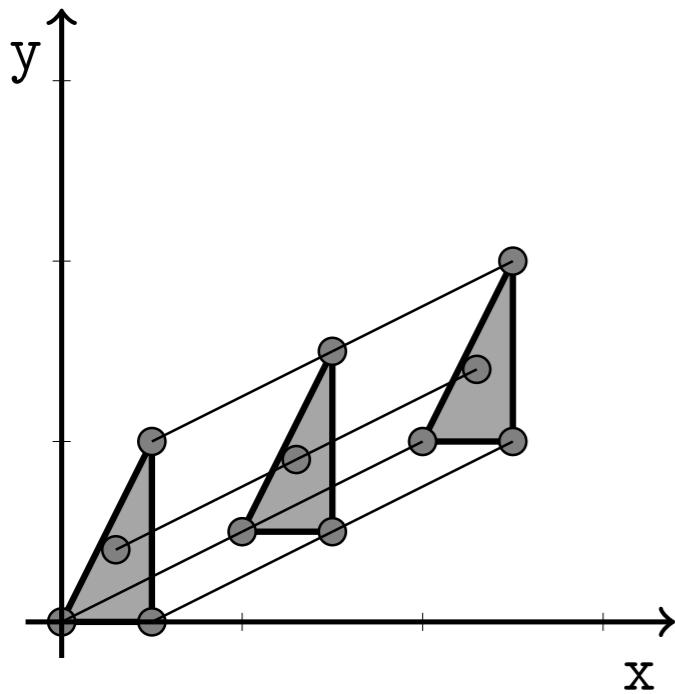
$$\begin{array}{l}
 0 \leq y \\
 y \leq 2x \\
 x \leq 0.5
 \end{array}$$

$$\begin{array}{l}
 0 \leq y \\
 y \leq 2x \\
 x \leq 1.5 \\
 y \leq 0.5x + 1 \\
 y \geq 0.5x - 1
 \end{array}$$

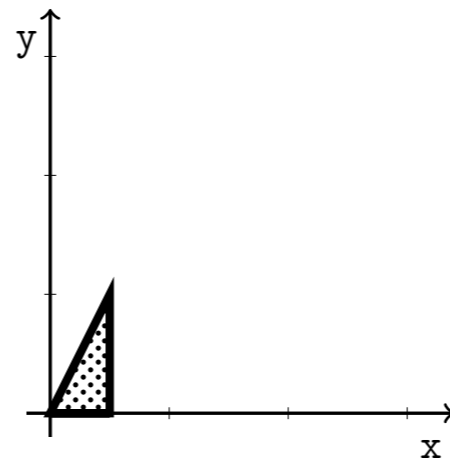
# Abstract Semantics Computation: iter{p}

## Example: Convex Polyhedra

```
init({(x, y) | 0 ≤ y ≤ 2x and x ≤ 0.5});  
iter{  
  translation(1, 0.5)  
}
```



(a) Concrete semantics



(b) Analysis of p<sub>0</sub> (0 iteration)

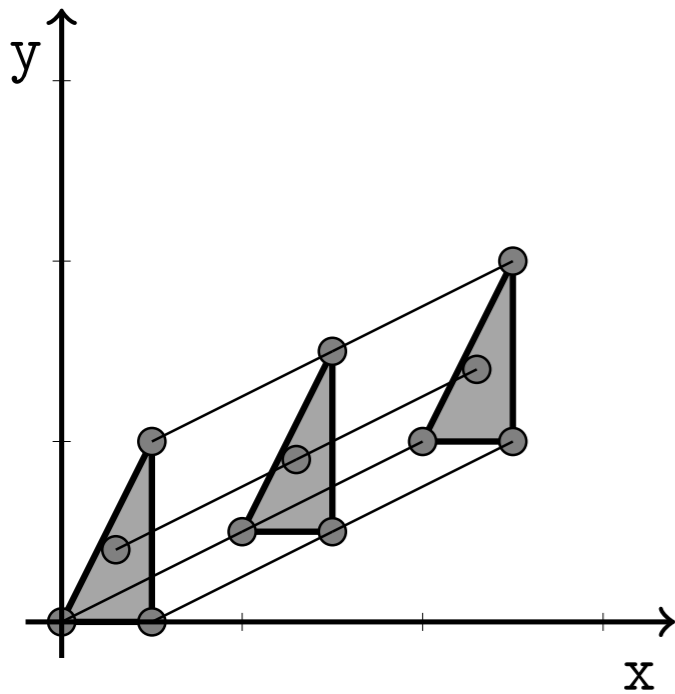
**Discard this**

$0 \leq y$   
 $y \leq 2x$   
 ~~$x \leq 0.5$~~

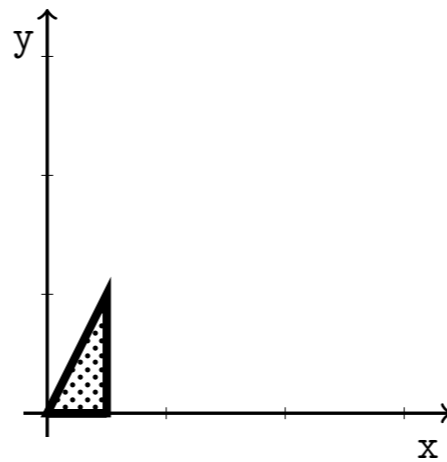
# Abstract Semantics Computation: iter{p}

## Example: Convex Polyhedra

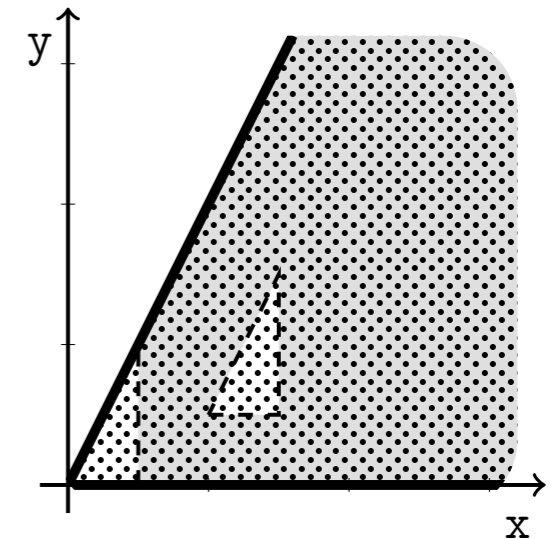
```
init({(x, y) | 0 ≤ y ≤ 2x and x ≤ 0.5});  
iter{  
  translation(1, 0.5)  
}
```



(a) Concrete semantics



(b) Analysis of p<sub>0</sub> (0 iteration)



(b) Iteration 1

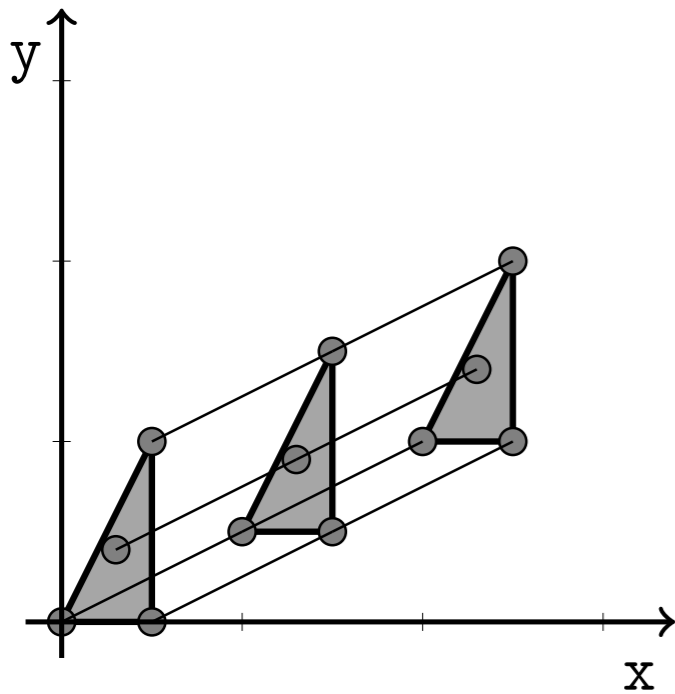
$$\begin{array}{l} 0 \leq y \\ y \leq 2x \\ x \leq 0.5 \end{array}$$

$$\begin{array}{l} 0 \leq y \\ y \leq 2x \end{array}$$

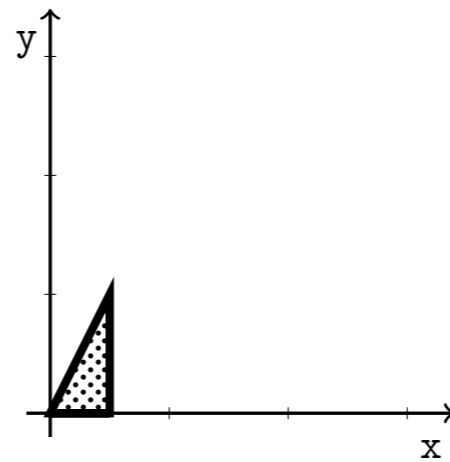
# Abstract Semantics Computation: iter{p}

## Example: Convex Polyhedra

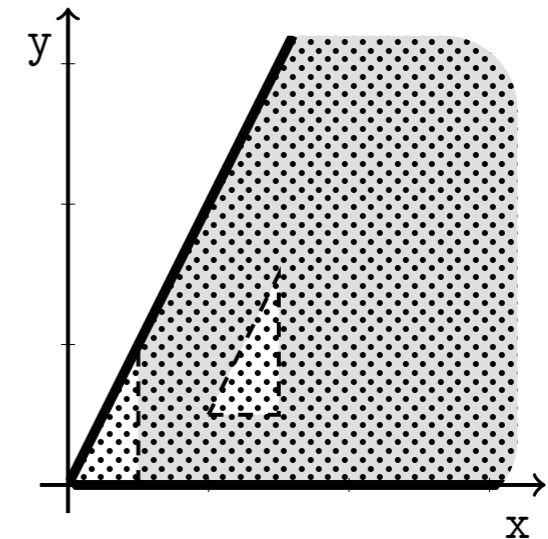
```
init({(x, y) | 0 ≤ y ≤ 2x and x ≤ 0.5});  
iter{  
  translation(1, 0.5)  
}
```



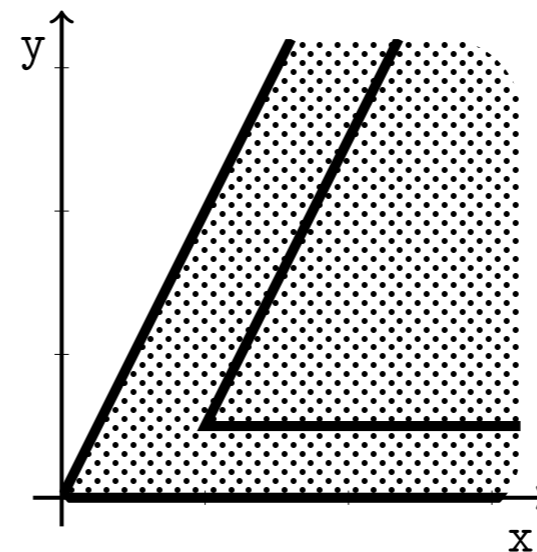
(a) Concrete semantics



(b) Analysis of  $p_0$  (0 iteration)



(b) Iteration 1



(c) Iteration 2 and limit



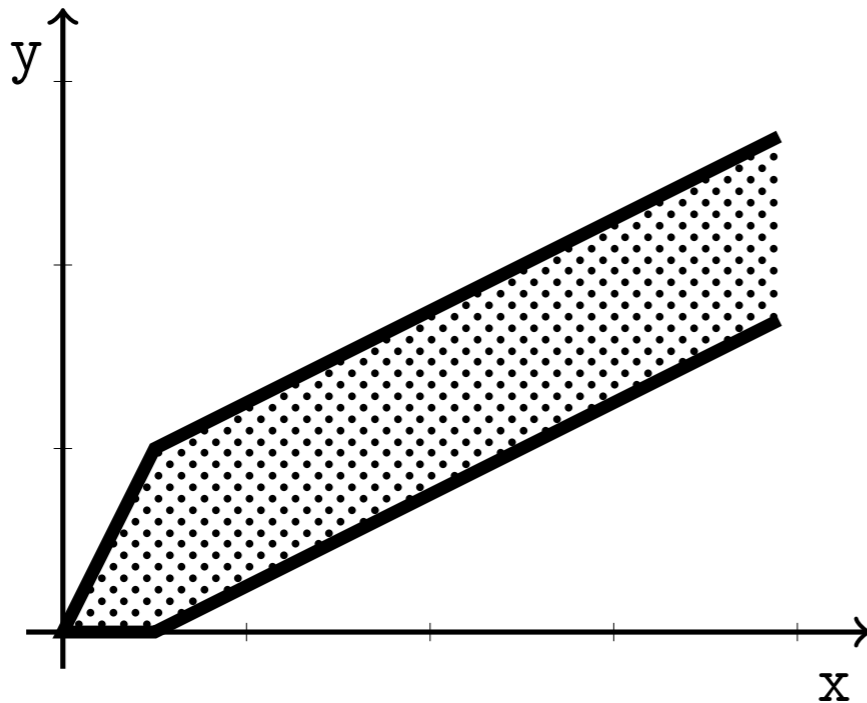
# Imprecision due to Widening

---

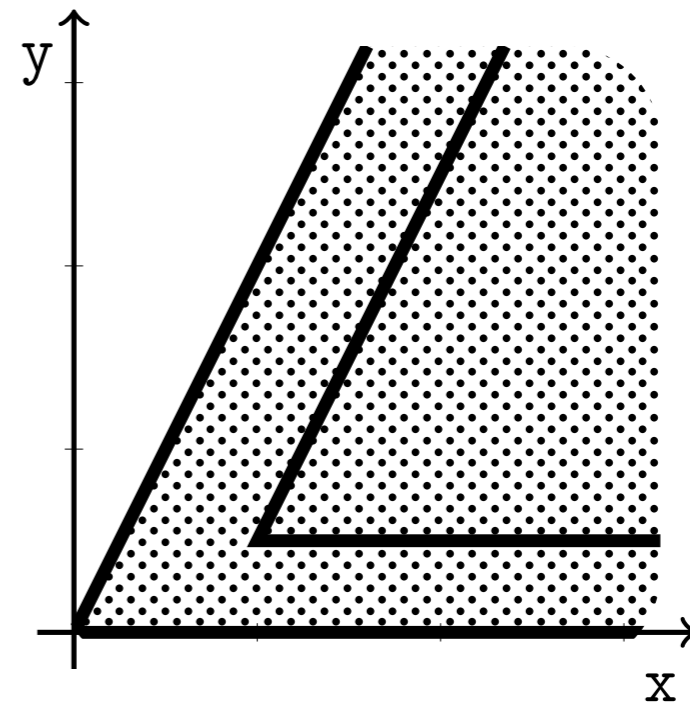
operator `widen`

{ over-approximates unions  
enforces convergence

- Widening guarantees termination of the analysis.
- However, it incurs significant precision loss.



(f) Expected result



(c) Iteration 2 and limit

# Abstract Iteration with Widening

---

Recall

$$\begin{aligned} \text{iter}\{p\} &= \{\} \text{ or } \{p\} \text{ or } \{p; p\} \text{ or } \dots \\ &= \lim_i p_i \end{aligned}$$

where

$$p_0 = \{\} \quad p_{k+1} = p_k \text{ or } \{p_k; p\}$$

Hence,

$$\text{analysis}(\text{iter}\{p\}, a) = \left\{ \begin{array}{l} R \leftarrow a; \\ \text{repeat} \\ \quad T \leftarrow R; \\ \quad R \leftarrow \text{widen}(R, \text{analysis}(p, R)); \\ \text{until } \text{inclusion}(R, T) \\ \text{return } T; \end{array} \right.$$

operator widen  $\left\{ \begin{array}{l} \text{over approximates unions} \\ \text{enforces finite convergence} \end{array} \right.$

# Loop Unrolling for Precision Improvement

---

```
init({(x,y) | 0 ≤ y ≤ 2x and x ≤ 0.5});  
iter{  
    translation(1,0.5)  
}
```

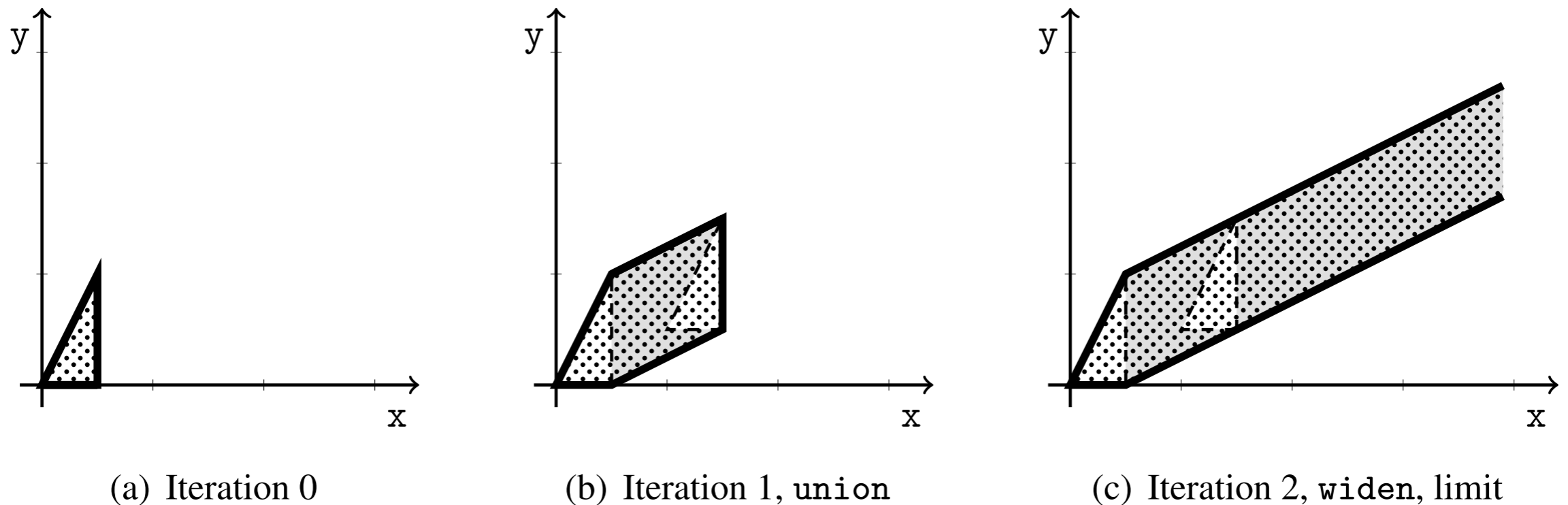


**Loop unrolling once**

```
init({(x,y) | 0 ≤ y ≤ 2x and x ≤ 0.5});  
{ }or{  
    translation(1,0.5)  
}  
iter{  
    translation(1,0.5)  
}
```

# Loop Unrolling for Precision Improvement

---



**Figure 2.17**

Abstract iteration with widening and unrolling

# Abstract Semantics Function analysis At a Glance

---

The `analysis(p, a)` is finitely computable and sound.

$$\begin{aligned} \text{analysis}(\text{init}(\mathfrak{R}), a) &= \text{best abstraction of the region } \mathfrak{R} \\ \text{analysis}(\text{translation}(u, v), a) &= \left\{ \begin{array}{l} \text{return an abstract state that contains} \\ \text{the translation of } a \end{array} \right. \\ \text{analysis}(\text{rotation}(u, v, \theta), a) &= \left\{ \begin{array}{l} \text{return an abstract state that contains} \\ \text{the rotation of } a \end{array} \right. \\ \text{analysis}(\{p_0\} \text{ or } \{p_1\}, a) &= \text{union}(\text{analysis}(p_1, a), \text{analysis}(p_0, a)) \\ \text{analysis}(p_0; p_1, a) &= \text{analysis}(p_1, \text{analysis}(p_0, a)) \\ \text{analysis}(\text{iter}\{p\}, a) &= \left\{ \begin{array}{l} R \leftarrow a; \\ \text{repeat} \\ \quad T \leftarrow R; \\ \quad R \leftarrow \text{widen}(R, \text{analysis}(p, R)); \\ \text{until inclusion}(R, T) \\ \text{return } T; \end{array} \right. \end{aligned}$$

# Soundness of Abstract Semantics Function analysis

---

## Sound analysis

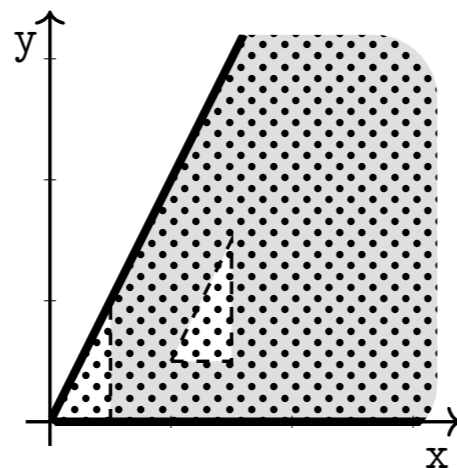
If an execution of  $p$  from a state  $(x, y)$  generates the state  $(x', y')$ ,  
then for all abstract element  $a$  such that  $(x, y) \in \gamma(a)$ ,  
$$(x', y') \in \gamma(\text{analysis}(p, a))$$

**Theorem.** The **analysis** function is sound.

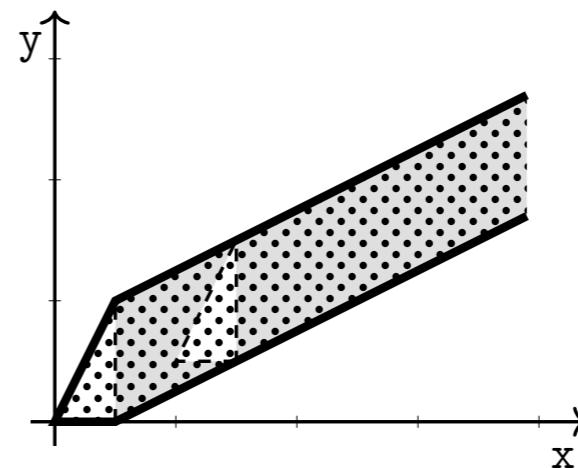
# Verification of the Property of Interest

---

- Does program compute a point inside no-fly zone  $\mathcal{D}$ ?
- Need to collect the set of reachable points.
- Run `analysis(p, -)` and collect all points  $\mathcal{R}$  from every call to `analysis`.
- Since `analysis` is sound, the result is an over approx. of the reachable points.
- **If  $\mathcal{R} \cap \mathcal{D} = \emptyset$ , then  $p$  is verified. Otherwise, we don't know.**



(a) An example  $\mathcal{R}$



(b) A more precise  $\mathcal{R}$