AAA616: Program Analysis

Introduction to Program Analysis

Hakjoo Oh
2019 Fall

Static Program 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-overrun?", "memory leak?", "type errors?", "x = y at line 2?", "memory use $\leq \boldsymbol{2K}$?", etc

Program Analysis is Undecidable

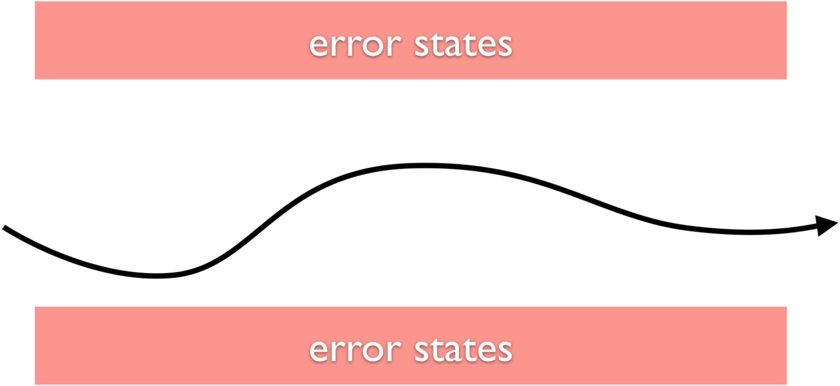
Reasoning about program behavior involves the Halting Problem: e.g.,

$$\texttt{if} \ \cdots \ \texttt{then} \ x := 1 \ \texttt{else} \ (S; x := 2); y := x$$

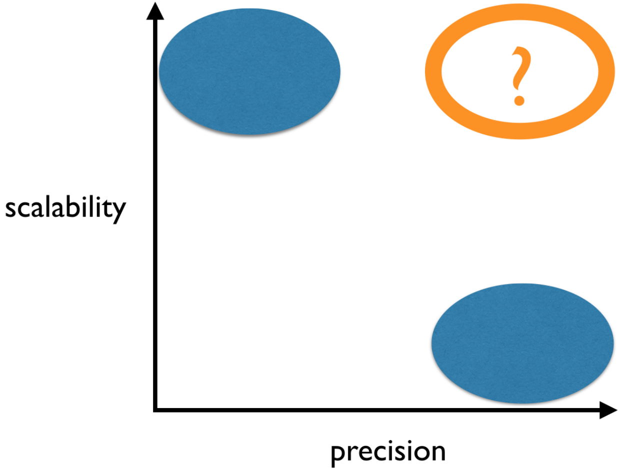
(S does not define x.) What are the possible values of x at the last statement?

Alan Turing (1912–1954)

Side-Stepping Undecidability



Tradeoff between Precision and Scalability



The While Language

$$\begin{array}{rcl} a & \rightarrow & n \mid x \mid a_1 + a_2 \mid a_1 \star a_2 \mid a_1 - a_2 \\ b & \rightarrow & \texttt{true} \mid \texttt{false} \mid a_1 = a_2 \mid a_1 \leq a_2 \mid \neg b \mid b_1 \wedge b_2 \\ c & \rightarrow & x := a \mid \texttt{skip} \mid c_1; c_2 \mid \texttt{if}\ b\ c_1\ c_2 \mid \texttt{while}\ b\ c \end{array}$$

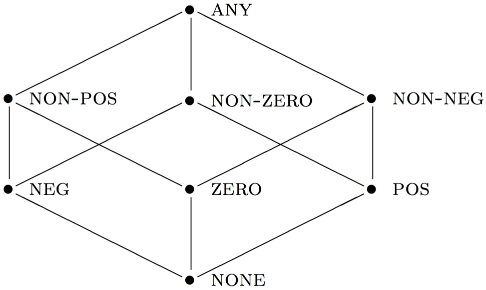
Example 1: Sign Analysis

Execute the program with abstract values ($\mathbf{POS}, \mathbf{NEG}, \mathbf{0}, \top, \bot$):

```
// a >= 0, b >= 0
int mod (int a, int b) {
  int q = 0;
  int r = a;
  while (r >= b) {
    r = r - b;
    q = q + 1;
  }
  return r;
}
```

Sign Domain

The complete lattice $(\mathbf{Sign}, \sqsubseteq)$:



The lattice is an abstraction of integers:

$$\alpha_{\mathbb{Z}} : \wp(\mathbb{Z}) \to \mathbf{Sign}, \qquad \gamma_{\mathbb{Z}} : \mathbf{Sign} \to \wp(\mathbb{Z})$$

Abstract States

The complete lattice of abstract states $(\widehat{\textbf{State}}, \sqsubseteq)$:

$$\widehat{\textbf{State}} = Var \to \textbf{Sign}$$

with the pointwise ordering:

$$\hat{s}_1 \sqsubseteq \hat{s}_2 \iff \forall x \in Var.\ \hat{s}_1(x) \sqsubseteq \hat{s}_2(x).$$

The least upper bound of $Y \subseteq \widehat{\textbf{State}}$,

$$\bigsqcup Y = \lambda x.\ \bigsqcup_{\hat{s} \in Y} \hat{s}(x).$$

Lemma

Let S be a non-empty set and $(D, \sqsubseteq)$ be a poset. Then, the poset $(S \to D, \sqsubseteq)$ with the ordering

$$f_1 \sqsubseteq f_2 \iff \forall s \in S.\ f_1(s) \sqsubseteq f_2(s)$$

is a complete lattice (resp., CPO) if D is a complete lattice (resp., CPO).

Abstract States

The complete lattice of abstract states $(\widehat{\textbf{State}}, \sqsubseteq)$:

$$\widehat{\textbf{State}} = Var \to \textbf{Sign}$$

with the pointwise ordering:

$$\hat{s}_1 \sqsubseteq \hat{s}_2 \iff \forall x \in Var.\ \hat{s}_1(x) \sqsubseteq \hat{s}_2(x).$$

The least upper bound of $Y \subseteq \widehat{\textbf{State}}$,

$$\bigsqcup Y = \lambda x.\ \bigsqcup_{\hat{s} \in Y} \hat{s}(x).$$

$$\alpha : \wp(\textbf{State}) \to \widehat{\textbf{State}}$$

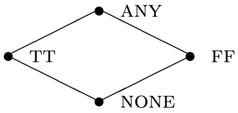
$$\alpha(S) = \lambda x.\ \bigsqcup_{s \in S} \alpha_{\mathbb{Z}}(\{s(x)\})$$

$$\gamma : \widehat{\textbf{State}} \to \wp(\textbf{State})$$

$$\gamma(\hat{s}) = \{s \in \textbf{State} \mid \forall x \in Var.\ s(x) \in \gamma_{\mathbb{Z}}(\hat{s}(x))\}$$

Abstract Booleans

The truth values $\mathbf{T} = \{true, false\}$ are abstracted by the complete lattice $(\widehat{\mathbf{T}}, \sqsubseteq)$:



The abstraction and concretization functions for the lattice:

$$\alpha_{\mathbf{T}} : \wp(\mathbf{T}) \to \widehat{\mathbf{T}}, \qquad \gamma_{\mathbf{T}} : \widehat{\mathbf{T}} \to \wp(\mathbf{T})$$

Abstract Semantics

$$
\begin{aligned}
\widehat{\mathcal{A}}[\![a]\!] &: \widehat{\textsf{State}} \to \textsf{Sign} \\
\widehat{\mathcal{A}}[\![n]\!](\hat{s}) &= \alpha_{\mathbb{Z}}(\{n\}) \\
\widehat{\mathcal{A}}[\![x]\!](\hat{s}) &= \hat{s}(x) \\
\widehat{\mathcal{A}}[\![a_1 + a_2]\!](\hat{s}) &= \widehat{\mathcal{A}}[\![a_1]\!](\hat{s}) +_S \widehat{\mathcal{A}}[\![a_2]\!](\hat{s}) \\
\widehat{\mathcal{A}}[\![a_1 \star a_2]\!](\hat{s}) &= \widehat{\mathcal{A}}[\![a_1]\!](\hat{s}) \star_S \widehat{\mathcal{A}}[\![a_2]\!](\hat{s}) \\
\widehat{\mathcal{A}}[\![a_1 - a_2]\!](\hat{s}) &= \widehat{\mathcal{A}}[\![a_1]\!](\hat{s}) -_S \widehat{\mathcal{A}}[\![a_2]\!](\hat{s})
\end{aligned}
$$

Abstract Semantics

$+_S$	NONE	NEG	ZERO	POS	NON-POS	NON-ZERO	NON-NEG	ANY
NONE	NONE	NONE	NONE	NONE	NONE	NONE	NONE	NONE
NEG	NONE	NEG	NEG	ANY	NEG	ANY	ANY	ANY
ZERO	NONE	POS	ZERO	POS	NON-POS	NON-ZERO	NON-NEG	ANY
POS	NONE	ANY	POS	POS	ANY	ANY	POS	ANY
NON-POS	NONE	NEG	NON-POS	ANY	NON-POS	ANY	ANY	ANY
NON-ZERO	NONE	ANY	NON-ZERO	ANY	ANY	ANY	ANY	ANY
NON-NEG	NONE	ANY	NON-NEG	POS	ANY	ANY	NON-NEG	ANY
ANY	NONE	ANY	ANY	ANY	ANY	ANY	ANY	ANY

$\star_S$	NEG	ZERO	POS
NEG	POS	ZERO	NEG
ZERO	ZERO	ZERO	ZERO
POS	NEG	ZERO	POS

$-_S$	NEG	ZERO	POS
NEG	ANY	NEG	NEG
ZERO	POS	ZERO	NEG
POS	POS	POS	ANY

Abstract Semantics

$$
\begin{array}{rcl}
\widehat{\mathcal{B}}[\![b]\!] & : & \widehat{\textsf{State}} \to \widehat{\textsf{T}} \\
\widehat{\mathcal{B}}[\![\texttt{true}]\!](\hat{s}) & = & \mathrm{TT} \\
\widehat{\mathcal{B}}[\![\texttt{false}]\!](\hat{s}) & = & \mathrm{FF} \\
\widehat{\mathcal{B}}[\![a_1 = a_2]\!](\hat{s}) & = & \widehat{\mathcal{B}}[\![a_1]\!](\hat{s}) =_S \widehat{\mathcal{B}}[\![a_2]\!](\hat{s}) \\
\widehat{\mathcal{B}}[\![a_1 \leq a_2]\!](\hat{s}) & = & \widehat{\mathcal{B}}[\![a_1]\!](\hat{s}) \leq_S \widehat{\mathcal{B}}[\![a_2]\!](\hat{s}) \\
\widehat{\mathcal{B}}[\![\neg b]\!](\hat{s}) & = & \neg_S \widehat{\mathcal{B}}[\![b]\!](\hat{s}) \\
\widehat{\mathcal{B}}[\![b_1 \wedge b_2]\!](\hat{s}) & = & \widehat{\mathcal{B}}[\![b_1]\!](\hat{s}) \wedge_S \widehat{\mathcal{B}}[\![b_2]\!](\hat{s})
\end{array}
$$

Abstract Semantics

$=_S$	NEG	ZERO	POS
NEG	ANY	FF	FF
ZERO	FF	TT	FF
POS	FF	FF	ANY

$\leq_S$	NEG	ZERO	POS
NEG	ANY	TT	TT
ZERO	FF	TT	TT
POS	FF	FF	ANY

$\neg_T$	
NONE	NONE
TT	FF
FF	TT
ANY	ANY

$\wedge_T$	NONE	TT	FF	ANY
NONE	NONE	NONE	NONE	NONE
TT	NONE	TT	FF	ANY
FF	NONE	FF	FF	FF
ANY	NONE	ANY	FF	ANY

Abstract Semantics

$$
\begin{aligned}
\widehat{\mathcal{C}}[\![c]\!] &: \widehat{\textsf{State}} \to \widehat{\textsf{State}} \\
\widehat{\mathcal{C}}[\![x := a]\!] &= \lambda \hat{s}.\hat{s}[x \mapsto \widehat{\mathcal{A}}[\![a]\!](\hat{s})] \\
\widehat{\mathcal{C}}[\![\texttt{skip}]\!] &= \textsf{id} \\
\widehat{\mathcal{C}}[\![c_1; c_2]\!] &= \widehat{\mathcal{C}}[\![c_2]\!] \circ \widehat{\mathcal{C}}[\![c_1]\!] \\
\widehat{\mathcal{C}}[\![\texttt{if}\ b\ c_1\ c_2]\!] &= \widehat{\textsf{cond}}(\widehat{\mathcal{B}}[\![b]\!], \widehat{\mathcal{C}}[\![c_1]\!], \widehat{\mathcal{C}}[\![c_2]\!]) \\
\widehat{\mathcal{C}}[\![\texttt{while}\ b\ c]\!] &= \mathit{fix}\,\widehat{F} \\
& \quad \text{where } \widehat{F}(g) = \widehat{\textsf{cond}}(\widehat{\mathcal{B}}[\![b]\!], g \circ \widehat{\mathcal{C}}[\![c]\!], \textsf{id})
\end{aligned}
$$

$$
\widehat{\textsf{cond}}(f, g, h)(\hat{s}) = \begin{cases} \bot & \cdots f(\hat{s}) = \mathrm{NONE} \\ f(\hat{s}) & \cdots f(\hat{s}) = \mathrm{TT} \\ g(\hat{s}) & \cdots f(\hat{s}) = \mathrm{FF} \\ f(\hat{s}) \sqcup g(\hat{s}) & \cdots f(\hat{s}) = \mathrm{ANY} \end{cases}
$$

Example 2: Taint Analysis (Information Flow Analysis)

Can the information from the untrustworthy source be transferred to the sink?

```
x:=source(); ...; sink(y)
```

Applications to sw security:

- privacy leak
- SQL injection
- buffer overflow
- integer overflow
- XSS
- ...

Abstract Domain

- The complete lattice of the abstract values $(\widehat{\mathbf{T}}, \sqsubseteq)$:

$$\widehat{\mathbf{T}} = \{\mathrm{LOW}, \mathrm{HIGH}\}$$

  with the ordering $\mathrm{LOW} \sqsubseteq \mathrm{HIGH}$, $\mathrm{LOW} \sqsubseteq \mathrm{LOW}$, and $\mathrm{HIGH} \sqsubseteq \mathrm{HIGH}$.

- The lattice of states:

$$\widehat{\mathbf{State}} = \mathit{Var} \to \widehat{\mathbf{T}}$$

Abstract Semantics

$$
\begin{array}{rcl}
\widehat{\mathcal{A}}[\![a]\!] & : & \widehat{\textsf{State}} \to \widehat{\textsf{T}} \\
\widehat{\mathcal{A}}[\![n]\!](\hat{s}) & = & \begin{cases} \text{LOW} & \cdots n \text{ is public} \\ \text{HIGH} & \cdots n \text{ is private} \end{cases} \\
\widehat{\mathcal{A}}[\![x]\!](\hat{s}) & = & \hat{s}(x) \\
\widehat{\mathcal{A}}[\![a_1 + a_2]\!](\hat{s}) & = & \widehat{\mathcal{A}}[\![a_1]\!](\hat{s}) \sqcup \widehat{\mathcal{A}}[\![a_2]\!](\hat{s}) \\
\widehat{\mathcal{A}}[\![a_1 \star a_2]\!](\hat{s}) & = & \widehat{\mathcal{A}}[\![a_1]\!](\hat{s}) \sqcup \widehat{\mathcal{A}}[\![a_2]\!](\hat{s}) \\
\widehat{\mathcal{A}}[\![a_1 - a_2]\!](\hat{s}) & = & \widehat{\mathcal{A}}[\![a_1]\!](\hat{s}) \sqcup \widehat{\mathcal{A}}[\![a_2]\!](\hat{s})
\end{array}
$$

Abstract Semantics

$$
\begin{array}{rcl}
\widehat{\mathcal{B}}[\![b]\!] & : & \widehat{\textsf{State}} \to \widehat{\textsf{T}} \\
\widehat{\mathcal{B}}[\![\texttt{true}]\!](\hat{s}) & = & \mathrm{LOW} \\
\widehat{\mathcal{B}}[\![\texttt{false}]\!](\hat{s}) & = & \mathrm{LOW} \\
\widehat{\mathcal{B}}[\![a_1 = a_2]\!](\hat{s}) & = & \widehat{\mathcal{B}}[\![a_1]\!](\hat{s}) \sqcup \widehat{\mathcal{B}}[\![a_2]\!](\hat{s}) \\
\widehat{\mathcal{B}}[\![a_1 \leq a_2]\!](\hat{s}) & = & \widehat{\mathcal{B}}[\![a_1]\!](\hat{s}) \sqcup \widehat{\mathcal{B}}[\![a_2]\!](\hat{s}) \\
\widehat{\mathcal{B}}[\![\neg b]\!](\hat{s}) & = & \widehat{\mathcal{B}}[\![b]\!](\hat{s}) \\
\widehat{\mathcal{B}}[\![b_1 \wedge b_2]\!](\hat{s}) & = & \widehat{\mathcal{B}}[\![b_1]\!](\hat{s}) \sqcup \widehat{\mathcal{B}}[\![b_2]\!](\hat{s})
\end{array}
$$

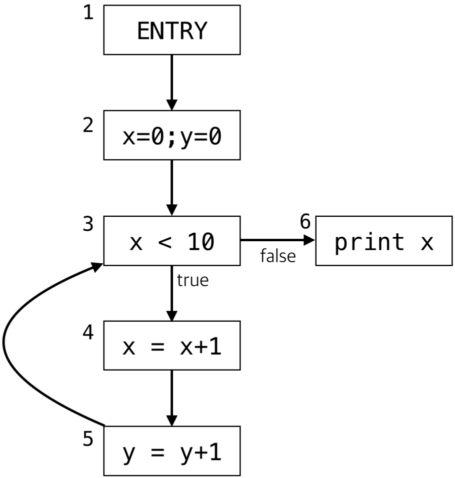
Abstract Semantics

$$
\begin{array}{rcl}
\widehat{\mathcal{C}}[\![c]\!] & : & \widehat{\textsf{State}} \to \widehat{\textsf{State}} \\
\widehat{\mathcal{C}}[\![x := a]\!] & = & \lambda \hat{s}.\hat{s}[x \mapsto \widehat{\mathcal{A}}[\![a]\!](\hat{s})] \\
\widehat{\mathcal{C}}[\![\texttt{skip}]\!] & = & \textbf{id} \\
\widehat{\mathcal{C}}[\![c_1; c_2]\!] & = & \widehat{\mathcal{C}}[\![c_2]\!] \circ \widehat{\mathcal{C}}[\![c_1]\!] \\
\widehat{\mathcal{C}}[\![\texttt{if}\ b\ c_1\ c_2]\!] & = & \lambda \hat{s}.\widehat{\mathcal{C}}[\![c_1]\!](\hat{s}) \sqcup \widehat{\mathcal{C}}[\![c_2]\!](\hat{s}) \\
\widehat{\mathcal{C}}[\![\texttt{while}\ b\ c]\!] & = & \mathit{fix} \widehat{F} \\
 & & \text{where } \widehat{F}(g) = \lambda \hat{s}.\hat{s} \sqcup (g \circ \widehat{\mathcal{C}}[\![c]\!])(\hat{s})
\end{array}
$$

Example 3: Interval Analysis

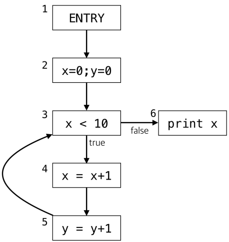
- ```
  x = 0;
  while (x < 10) {
    assert (x < 10);
    x++;
  }
  assert (x == 10);
  ```
- ```
  x = 0;
  y = 0;
  while (x < 10) {
    assert (y < 10);
    x++; y++;
  }
  assert (y == 10);
  ```

Example 3: Interval Analysis



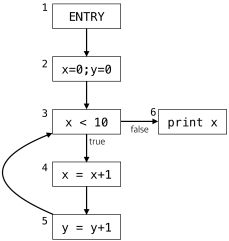
Node	Result
1	$x \mapsto \bot$ $y \mapsto \bot$
2	$x \mapsto [0, 0]$ $y \mapsto [0, 0]$
3	$x \mapsto [0, 9]$ $y \mapsto [0, +\infty]$
4	$x \mapsto [1, 10]$ $y \mapsto [0, +\infty]$
5	$x \mapsto [1, 10]$ $y \mapsto [1, +\infty]$
6	$x \mapsto [10, 10]$ $y \mapsto [0, +\infty]$

Fixed Point Computation Does Not Terminate



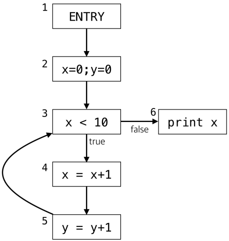
Node	initial	1	2	3	10	11	k	∞
1	$x \mapsto \bot$ $y \mapsto \bot$	$x \mapsto \bot$ $y \mapsto \bot$	$x \mapsto \bot$ $y \mapsto \bot$	$x \mapsto \bot$ $y \mapsto \bot$	$x \mapsto \bot$ $y \mapsto \bot$	$x \mapsto \bot$ $y \mapsto \bot$	$x \mapsto \bot$ $y \mapsto \bot$	$x \mapsto \bot$ $y \mapsto \bot$
2	$x \mapsto \bot$ $y \mapsto \bot$	$x \mapsto [0,0]$ $y \mapsto [0,0]$	$x \mapsto [0,0]$ $y \mapsto [0,0]$	$x \mapsto [0,0]$ $y \mapsto [0,0]$	$x \mapsto [0,0]$ $y \mapsto [0,0]$	$x \mapsto [0,0]$ $y \mapsto [0,0]$	$x \mapsto [0,0]$ $y \mapsto [0,0]$	$x \mapsto [0,0]$ $y \mapsto [0,0]$
3	$x \mapsto \bot$ $y \mapsto \bot$	$x \mapsto [0,0]$ $y \mapsto [0,0]$	$x \mapsto [0,1]$ $y \mapsto [0,1]$	$x \mapsto [0,2]$ $y \mapsto [0,2]$	$x \mapsto [0,9]$ $y \mapsto [0,9]$	$x \mapsto [0,9]$ $y \mapsto [0,10]$	$x \mapsto [0,9]$ $y \mapsto [0,k-1]$	$x \mapsto [0,9]$ $y \mapsto [0,+\infty]$
4	$x \mapsto \bot$ $y \mapsto \bot$	$x \mapsto [1,1]$ $y \mapsto [0,0]$	$x \mapsto [1,2]$ $y \mapsto [0,1]$	$x \mapsto [1,3]$ $y \mapsto [0,2]$	$x \mapsto [1,10]$ $y \mapsto [0,9]$	$x \mapsto [1,10]$ $y \mapsto [0,10]$	$x \mapsto [1,10]$ $y \mapsto [0,k-1]$	$x \mapsto [1,10]$ $y \mapsto [0,+\infty]$
5	$x \mapsto \bot$ $y \mapsto \bot$	$x \mapsto [1,1]$ $y \mapsto [1,1]$	$x \mapsto [1,2]$ $y \mapsto [1,2]$	$x \mapsto [1,3]$ $y \mapsto [1,3]$	$x \mapsto [1,10]$ $y \mapsto [1,10]$	$x \mapsto [1,10]$ $y \mapsto [1,11]$	$x \mapsto [1,10]$ $y \mapsto [1,k]$	$x \mapsto [1,10]$ $y \mapsto [1,+\infty]$
6	$x \mapsto \bot$ $y \mapsto \bot$	$x \mapsto \bot$ $y \mapsto [0,0]$	$x \mapsto \bot$ $y \mapsto [0,1]$	$x \mapsto \bot$ $y \mapsto [0,2]$	$x \mapsto [10,10]$ $y \mapsto [0,9]$	$x \mapsto [10,10]$ $y \mapsto [0,10]$	$x \mapsto [10,10]$ $y \mapsto [0,k-1]$	$x \mapsto [10,10]$ $y \mapsto [0,+\infty]$

Fixed Point Computation with Widening and Narrowing



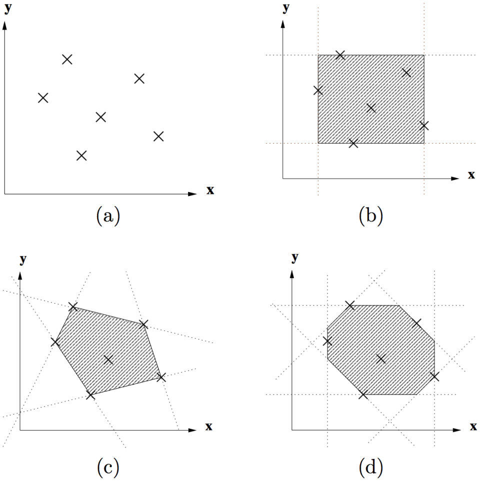
Node	initial	**1**	**2**	**3**
1	$x \mapsto \bot$ $y \mapsto \bot$	$x \mapsto \bot$ $y \mapsto \bot$	$x \mapsto \bot$ $y \mapsto \bot$	$x \mapsto \bot$ $y \mapsto \bot$
2	$x \mapsto \bot$ $y \mapsto \bot$	$x \mapsto [0,0]$ $y \mapsto [0,0]$	$x \mapsto [0,0]$ $y \mapsto [0,0]$	$x \mapsto [0,0]$ $y \mapsto [0,0]$
3	$x \mapsto \bot$ $y \mapsto \bot$	$x \mapsto [0,0]$ $y \mapsto [0,0]$	$x \mapsto [0,9]$ $y \mapsto [0,+\infty]$	$x \mapsto [0,9]$ $y \mapsto [0,+\infty]$
4	$x \mapsto \bot$ $y \mapsto \bot$	$x \mapsto [1,1]$ $y \mapsto [0,0]$	$x \mapsto [1,10]$ $y \mapsto [0,+\infty]$	$x \mapsto [1,10]$ $y \mapsto [0,+\infty]$
5	$x \mapsto \bot$ $y \mapsto \bot$	$x \mapsto [1,1]$ $y \mapsto [1,1]$	$x \mapsto [1,10]$ $y \mapsto [1,+\infty]$	$x \mapsto [1,10]$ $y \mapsto [1,+\infty]$
6	$x \mapsto \bot$ $y \mapsto \bot$	$x \mapsto \bot$ $y \mapsto [0,0]$	$x \mapsto [10,+\infty]$ $y \mapsto [0,+\infty]$	$x \mapsto [10,+\infty]$ $y \mapsto [0,+\infty]$

Fixed Point Computation with Widening and Narrowing



Node	initial	1	2
1	$x \mapsto \bot$ $y \mapsto \bot$	$x \mapsto \bot$ $y \mapsto \bot$	$x \mapsto \bot$ $y \mapsto \bot$
2	$x \mapsto [0,0]$ $y \mapsto [0,0]$	$x \mapsto [0,0]$ $y \mapsto [0,0]$	$x \mapsto [0,0]$ $y \mapsto [0,0]$
3	$x \mapsto [0,9]$ $y \mapsto [0,+\infty]$	$x \mapsto [0,9]$ $y \mapsto [0,+\infty]$	$x \mapsto [0,9]$ $y \mapsto [0,+\infty]$
4	$x \mapsto [1,10]$ $y \mapsto [0,+\infty]$	$x \mapsto [1,10]$ $y \mapsto [0,+\infty]$	$x \mapsto [1,10]$ $y \mapsto [0,+\infty]$
5	$x \mapsto [1,10]$ $y \mapsto [1,+\infty]$	$x \mapsto [1,10]$ $y \mapsto [1,+\infty]$	$x \mapsto [1,10]$ $y \mapsto [1,+\infty]$
6	$x \mapsto [10,+\infty]$ $y \mapsto [0,+\infty]$	$x \mapsto [10,10]$ $y \mapsto [0,+\infty]$	$x \mapsto [10,10]$ $y \mapsto [0,+\infty]$

cf) Numerical Abstractions



(image from The Octagon Abstract Domain by Antonine Mine)

Interval vs. Octagon

```
i = 0;
p = 0;

while (i < 12) {
    i = i + 1;
    p = p + 1;
}
assert(i==p)
```

Interval analysis

i	[12,12]
p	[0,+oo]

Octagon analysis

i	[12,12]
p	[12,12]
p-i	[0,0]
p+i	[24,24]