

Compositional Recurrence Analysis Revisited

Zachary Kincaid¹ Jason Breck² Ashkan Forouhi Boroujeni² Thomas Reps^{2,3}

¹Princeton University ²University of Wisconsin-Madison ³GammaTech, Inc.

June 19, 2017

How can we apply loop analyses to recursive procedures?

Over-approximating the behavior of loops

- Iterative program analysis [Cousot & Cousot POPL 1977]
 - Repeatedly evaluate the program under an abstract semantics until convergence upon a property that over-approximates all reachable states.

Over-approximating the behavior of loops

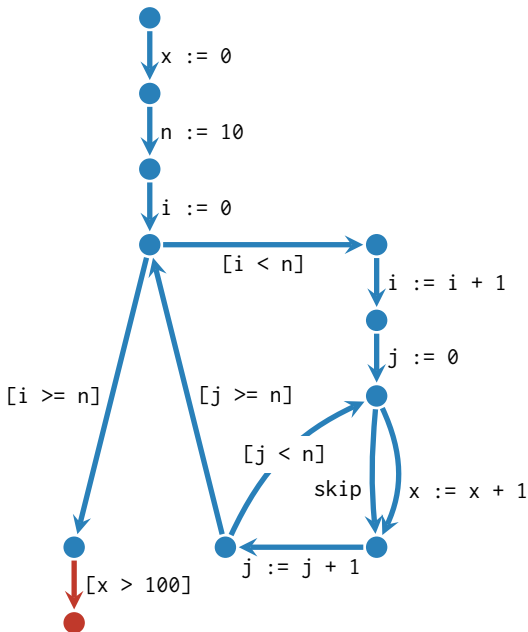
- Iterative program analysis [Cousot & Cousot POPL 1977]
 - Repeatedly evaluate the program under an abstract semantics until convergence upon a property that over-approximates all reachable states.
- Algebraic program analysis [Tarjan JACM 1981]
 - 1 Compute a *path expression* to a point of interest (e.g., an assertion)
 - 2 Evaluate the path expression in the *semantic algebra* defining the analysis to yield a property that over-approximates all paths.

```
x := 0
n := 10
i := 0
outer: if(i >= n):
    goto end
    i := i + 1
inner: j := 0
    if(*):
        x := x + 1
        j := j + 1
    if(j < n):
        goto inner
    goto outer
end: assert(x <= 100)
```

```

x := 0
n := 10
i := 0
outer: if(i >= n):
    goto end
    i := i + 1
inner: j := 0
    if(*):
        x := x + 1
        j := j + 1
        if(j < n):
            goto inner
        goto outer
end: assert(x <= 100)

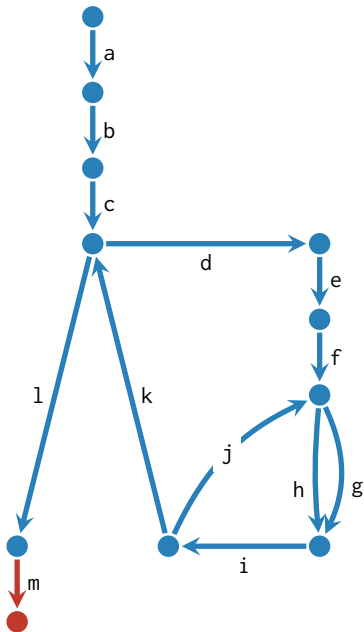
```



```

x := 0
n := 10
i := 0
outer: if(i >= n):
    goto end
    i := i + 1
inner: j := 0
    if(*):
        x := x + 1
        j := j + 1
        if(j < n):
            goto inner
        goto outer
end: assert(x <= 100)

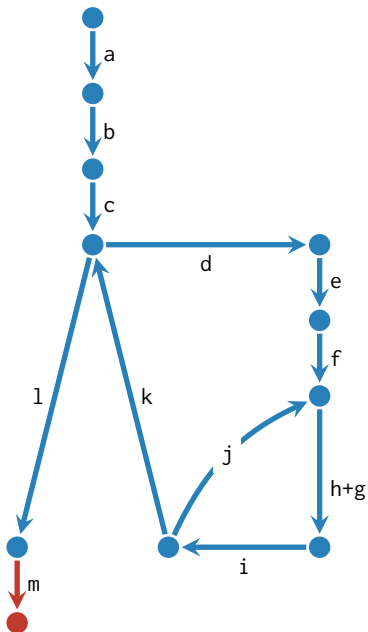
```



```

x := 0
n := 10
i := 0
outer: if(i >= n):
    goto end
    i := i + 1
inner: j := 0
    if(*):
        x := x + 1
        j := j + 1
        if(j < n):
            goto inner
        goto outer
end: assert(x <= 100)

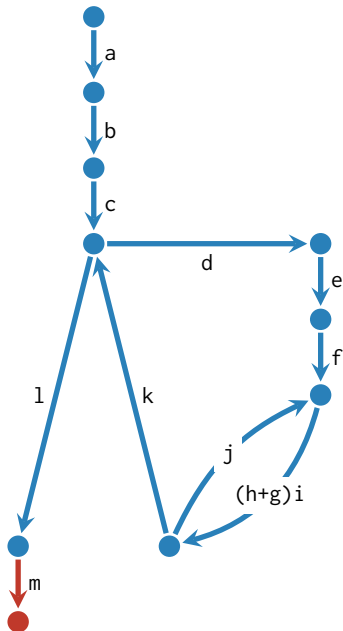
```




```

x := 0
n := 10
i := 0
outer: if(i >= n):
    goto end
    i := i + 1
inner: j := 0
    if(*):
        x := x + 1
        j := j + 1
        if(j < n):
            goto inner
        goto outer
end: assert(x <= 100)

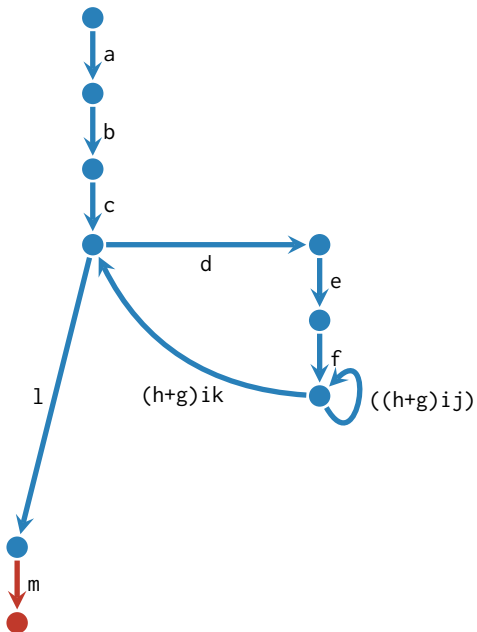
```



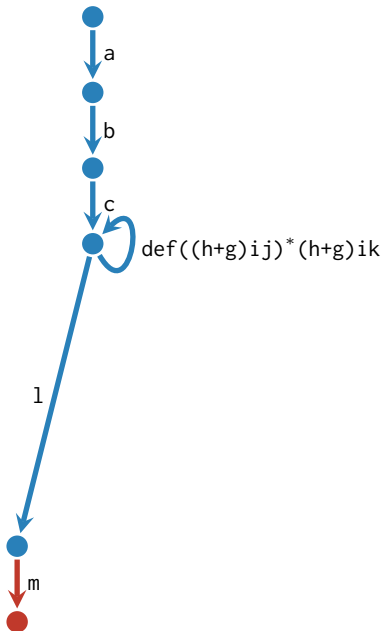
```

x := 0
n := 10
i := 0
outer: if(i >= n):
    goto end
    i := i + 1
inner: j := 0
    if(*):
        x := x + 1
        j := j + 1
        if(j < n):
            goto inner
        goto outer
end: assert(x <= 100)

```



```
x := 0
n := 10
i := 0
outer: if(i >= n):
    goto end
    i := i + 1
inner: j := 0
    if(*):
        x := x + 1
        j := j + 1
        if(j < n):
            goto inner
        goto outer
end: assert(x <= 100)
```



Evaluation of a path expression:

- $\mathcal{D} = \langle D, \otimes, \oplus, *, 0, 1 \rangle$ is a *semantic algebra*

Evaluation of a path expression:

- $\mathcal{D} = \langle D, \otimes, \oplus, *, 0, 1 \rangle$ is a *semantic algebra*



Space of program properties

Composition operators

Evaluation of a path expression:

- $\mathcal{D} = \langle D, \underbrace{\otimes, \oplus, *}, 0, 1 \rangle$ is a *semantic algebra*

Space of program properties

Evaluation of a path expression:

- $\mathcal{D} = \langle D, \otimes, \oplus, *, 0, 1 \rangle$ is a *semantic algebra*
- $\llbracket \cdot \rrbracket$: Control flow edges $\rightarrow D$ is a *semantic function*

Evaluation of a path expression:

- $\mathcal{D} = \langle D, \otimes, \oplus, *, 0, 1 \rangle$ is a *semantic algebra*
- $\llbracket \cdot \rrbracket$: Control flow edges $\rightarrow D$ is a *semantic function*

$$\begin{aligned} \llbracket abc(def((h+g)ij)^*(h+g)ik)^*lm \rrbracket &= \llbracket a \rrbracket \otimes \llbracket b \rrbracket \otimes \llbracket c \rrbracket \\ &\quad \otimes \left(\llbracket d \rrbracket \otimes \llbracket e \rrbracket \otimes \llbracket f \rrbracket \right. \\ &\quad \quad \otimes \left((\llbracket h \rrbracket \oplus \llbracket g \rrbracket) \otimes \llbracket i \rrbracket \otimes \llbracket j \rrbracket \right)^* \\ &\quad \quad \left. \otimes (\llbracket h \rrbracket \oplus \llbracket g \rrbracket) \otimes \llbracket i \rrbracket \otimes \llbracket k \rrbracket \right)^* \\ &\quad \otimes \llbracket l \rrbracket \otimes \llbracket m \rrbracket \end{aligned}$$

Compositional recurrence analysis [Farzan & Kincaid FMCAD 2015]

- D is the set of *transition formulas* in non-linear integer arithmetic

$$\llbracket x := x + 1 \rrbracket \triangleq x' = x + 1 \wedge y' = y$$

Compositional recurrence analysis [Farzan & Kincaid FMCAD 2015]

- D is the set of *transition formulas* in non-linear integer arithmetic

$$\llbracket x := x + 1 \rrbracket \triangleq x' = x + 1 \wedge y' = y$$

- $\varphi \otimes \psi \triangleq \exists \mathbf{x}'' . \varphi[\mathbf{x}' \mapsto \mathbf{x}''] \wedge \psi[\mathbf{x} \mapsto \mathbf{x}'']$

Compositional recurrence analysis [Farzan & Kincaid FMCAD 2015]

- D is the set of *transition formulas* in non-linear integer arithmetic

$$\llbracket x := x + 1 \rrbracket \triangleq x' = x + 1 \wedge y' = y$$

- $\varphi \otimes \psi \triangleq \exists \mathbf{x}'' . \varphi[\mathbf{x}' \mapsto \mathbf{x}''] \wedge \psi[\mathbf{x} \mapsto \mathbf{x}'']$
- $\varphi \oplus \psi \triangleq \varphi \vee \psi$

Compositional recurrence analysis [Farzan & Kincaid FMCAD 2015]

- D is the set of *transition formulas* in non-linear integer arithmetic

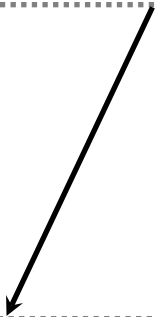
$$\llbracket x := x + 1 \rrbracket \triangleq x' = x + 1 \wedge y' = y$$

- $\varphi \otimes \psi \triangleq \exists \mathbf{x}'' . \varphi[\mathbf{x}' \mapsto \mathbf{x}''] \wedge \psi[\mathbf{x} \mapsto \mathbf{x}'']$
- $\varphi \oplus \psi \triangleq \varphi \vee \psi$
- $\varphi^* \triangleq \dots$

CRA's iteration operator

```
while(i < n):  
  if (*):  
    x := x + i  
  else  
    y := y + i  
  i := i + 1
```

..... loop body

$$i < n$$
$$\wedge \left(\begin{array}{l} (x' = x + i \wedge y' = y) \\ \vee (y' = y + i \wedge x' = x) \end{array} \right)$$
$$\wedge i' = i + 1$$
$$\wedge n' = n$$


..... loop abstraction

$$\exists k. k \geq 0 \wedge i' = i + k \wedge x' + y' = x + y + k(k+1)/2 + ki_0 \wedge x' \geq x \wedge y' \geq y$$

CRA's iteration operator

```
while(i < n):  
  if (*):  
    x := x + i  
  else  
    y := y + i  
  i := i + 1
```

..... loop body

$$\begin{aligned} & i < n \\ & \wedge \left(\begin{array}{l} (x' = x + i \wedge y' = y) \\ \vee (y' = y + i \wedge x' = x) \end{array} \right) \\ & \wedge i' = i + 1 \\ & \wedge n' = n \end{aligned}$$

..... recurrences

$$\begin{aligned} i^{(k)} &= i^{(k-1)} + 1 \\ x^{(k)} + y^{(k)} &= x^{(k-1)} + y^{(k-1)} + i \\ x^{(k)} &\geq x^{(k-1)} \\ y^{(k)} &\geq y^{(k-1)} \end{aligned}$$

..... loop abstraction

$$\exists k. k \geq 0 \wedge i' = i + k \wedge x' + y' = x + y + k(k+1)/2 + ki_0 \wedge x' \geq x \wedge y' \geq y$$

CRA's iteration operator

```
while(i < n):  
  if (*):  
    x := x + i  
  else  
    y := y + i  
  i := i + 1
```

$$\begin{array}{l} \text{loop body} \\ i < n \\ \wedge \left(\begin{array}{l} (x' = x + i \wedge y' = y) \\ \vee (y' = y + i \wedge x' = x) \end{array} \right) \\ \wedge i' = i + 1 \\ \wedge n' = n \end{array}$$

$$\begin{array}{l} \text{recurrences} \\ i^{(k)} = i^{(k-1)} + 1 \\ x^{(k)} + y^{(k)} = x^{(k-1)} + y^{(k-1)} + i \\ x^{(k)} \geq x^{(k-1)} \\ y^{(k)} \geq y^{(k-1)} \end{array}$$

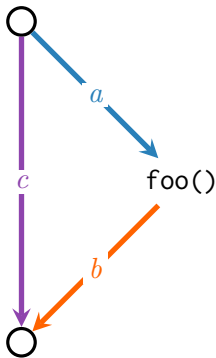
$$\begin{array}{l} \text{closed forms} \\ i^{(k)} = i^{(0)} + k \\ x^{(k)} + y^{(k)} = x^{(0)} + y^{(0)} + \frac{k(k+1)}{2} + ki_0 \\ x^{(k)} \geq x^{(0)} \\ y^{(k)} \geq y^{(0)} \end{array}$$

$$\text{loop abstraction} \\ \exists k. k \geq 0 \wedge i' = i + k \wedge x' + y' = x + y + k(k+1)/2 + ki_0 \wedge x' \geq x \wedge y' \geq y$$

How can we apply CRA to recursive procedures?

Recursive procedures have non-regular path languages

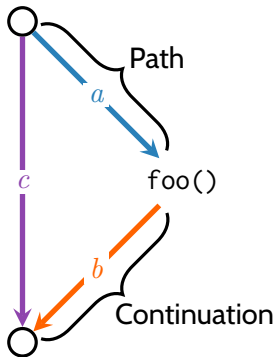
foo():



$paths(foo) = \{a^i c b^i : i \geq 0\}$ is not regular!

Tensor domains [Reps, Turetsky, Prabhu POPL 2016]

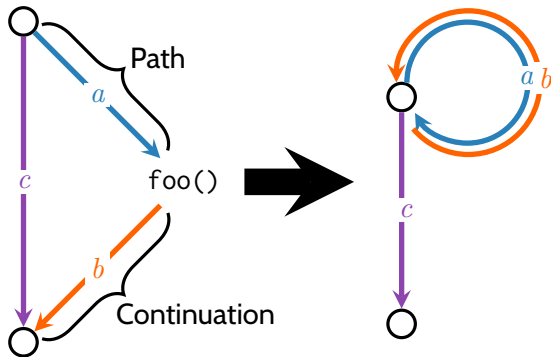
foo():



$paths(foo) = \{a^i c b^i : i \geq 0\}$ is not regular!

Tensor domains [Reps, Turetsky, Prabhu POPL 2016]

foo():



$paths(foo) = \{a^i c b^i : i \geq 0\}$ is not regular!

Tensorred paths

A tensorred path (p, k) is a pair consisting of a path p and a continuation k .

Tensorred paths

A tensorred path (p, k) is a pair consisting of a path p and a continuation k .

- $T_1 \otimes T_2 \triangleq \{(p_2 p_1, k_1 k_2) : (p_1, k_1) \in T_1, (p_2, k_2) \in T_2\}$
 - $(cd, dc) \otimes (ab, ba) = (abcd, dcba)$

Tensorred paths

A tensorred path (p, k) is a pair consisting of a path p and a continuation k .

- $T_1 \otimes T_2 \triangleq \{(p_2 p_1, k_1 k_2) : (p_1, k_1) \in T_1, (p_2, k_2) \in T_2\}$
 - $(cd, dc) \otimes (ab, ba) = (abcd, dcba)$
- $T_1 \oplus T_2 \triangleq T_1 \cup T_2$

Tensorred paths

A tensorred path (p, k) is a pair consisting of a path p and a continuation k .

- $T_1 \otimes T_2 \triangleq \{(p_2 p_1, k_1 k_2) : (p_1, k_1) \in T_1, (p_2, k_2) \in T_2\}$
 - $(cd, dc) \otimes (ab, ba) = (abcd, dcba)$
- $T_1 \oplus T_2 \triangleq T_1 \cup T_2$
- $T^* \triangleq \bigcup_n T^n$
 - $(a, b)^* = \{(a^i, b^i) : i \geq 0\}$

Tensorized paths

A tensorized path (p, k) is a pair consisting of a path p and a continuation k .

- $T_1 \otimes T_2 \triangleq \{(p_2 p_1, k_1 k_2) : (p_1, k_1) \in T_1, (p_2, k_2) \in T_2\}$
 - $(cd, dc) \otimes (ab, ba) = (abcd, dcba)$
- $T_1 \oplus T_2 \triangleq T_1 \cup T_2$
- $T^* \triangleq \bigcup_n T^n$
 - $(a, b)^* = \{(a^i, b^i) : i \geq 0\}$

Tensor product pairs a paths and continuations

$$P \odot K \triangleq \{(p, k) : p \in P, k \in K\}$$

Detensor product places a path between a path & continuation

$$Q \times T \triangleq \{pqk : q \in Q, (p, k) \in T\}$$

For example, $c \times (a \odot b)^* = \{a^i c b^i : i \geq 0\}$.

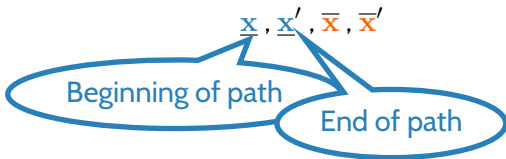
Tensor domain of CRA

Tensor transition formula \sim formula over *four* copies of the program variables

$$\underline{x}, \underline{x}', \bar{x}, \bar{x}'$$

Tensor domain of CRA

Tensor transition formula \sim formula over *four* copies of the program variables



Tensor domain of CRA

Tensor domain
variables

copies of the program

Beginning of continuation

End of continuation

$\underline{x}, \underline{x}', \bar{x}, \bar{x}'$

Beginning of path

End of path

Tensor domain of CRA

Tensored transition formula \sim formula over *four* copies of the program variables

$$\underline{x}, \underline{x}', \bar{x}, \bar{x}'$$

- $\Phi \otimes \Psi \triangleq \exists \underline{x}'', \bar{x}''. \Phi[\underline{x} \mapsto \underline{x}'', \bar{x}' \mapsto \bar{x}''] \wedge \Psi[\underline{x}' \mapsto \underline{x}'', \bar{x} \mapsto \bar{x}'']$

Tensor domain of CRA

Tensored transition formula \sim formula over *four* copies of the program variables

$$\underline{x}, \underline{x}', \bar{x}, \bar{x}'$$

- $\Phi \otimes \Psi \triangleq \exists \underline{x}'', \bar{x}''. \Phi[\underline{x} \mapsto \underline{x}'', \bar{x}' \mapsto \bar{x}''] \wedge \Psi[\underline{x}' \mapsto \underline{x}'', \bar{x} \mapsto \bar{x}'']$
- $\Phi \oplus \Psi, \Phi^*$ as for the untensored case

Tensor domain of CRA

Tensored transition formula \sim formula over *four* copies of the program variables

$$\underline{x}, \underline{x}', \bar{x}, \bar{x}'$$

- $\Phi \otimes \Psi \triangleq \exists \underline{x}'', \bar{x}''. \Phi[\underline{x} \mapsto \underline{x}'', \bar{x}' \mapsto \bar{x}''] \wedge \Psi[\underline{x}' \mapsto \underline{x}'', \bar{x} \mapsto \bar{x}'']$
- $\Phi \oplus \Psi, \Phi^*$ as for the untensored case
- $\varphi \odot \psi \triangleq \varphi[\underline{x} \mapsto \underline{x}, \underline{x}' \mapsto \underline{x}'] \wedge \psi[\underline{x} \mapsto \bar{x}, \underline{x}' \mapsto \bar{x}']$
 - E.g., $(x' = x + 1) \odot (y' = y + 2) = (\underline{x}' = \underline{x} + 1 \wedge \bar{y}' = \bar{y} + 2)$
- $\varphi \times \Psi \triangleq \exists \underline{x}, \underline{x}', \bar{x}, \bar{x}'. \varphi[\underline{x} \mapsto \underline{x}', \underline{x}' \mapsto \bar{x}] \wedge \Psi[\underline{x} \mapsto \underline{x}, \bar{x}' \mapsto \underline{x}']$
 - E.g., $(y' = x) \times (\underline{x}' = \underline{x} + 1 \wedge \bar{y}' = \bar{y} + 2) = (y' = x + 3)$

Solving non-linear recursive systems

```
fib(n):  
  if(i > 1):  
    f1 := fib(n-1)  
    f2 := fib(n-2)  
    return f1 + f2  
  else  
    return 1
```


Solving non-linear recursive systems

$paths(\text{fib}) \sim$ least fixed point of $X = aXbXc + d$

Solving non-linear recursive systems

$paths(\text{fib}) \sim$ least fixed point of $X = aXbXc + d$

[Reps, Turetsky, Prabhu POPL 2016] solves this via *Newton iteration*:

Solve a sequence of *linearized* systems until convergence on a property that over-approximates all paths.

Newton iteration

$$\begin{aligned}\nu_0 &= 0 \\ \nu_1 &= d \times ((a \odot b\nu_0c) \oplus (a\nu_0b \odot c))^* \\ \nu_2 &= d \times ((a \odot b\nu_1c) \oplus (a\nu_1b \odot c))^* \\ &\vdots \\ &(\text{repeat until } \nu_{n+1} = \nu_n)\end{aligned}$$

The problem with Newton iteration

- 1 Transition formulas have infinite ascending chains (convergence is not guaranteed)
- 2 Transition formula equivalence is undecidable (convergence can't be detected)

CRA's iteration operator

```
while(i < n):  
  if (*):  
    x := x + i  
  else  
    y := y + i  
  i := i + 1
```

..... loop body

$$\begin{aligned} & i < n \\ & \wedge \left(\begin{array}{l} (x' = x + i \wedge y' = y) \\ \vee (y' = y + i \wedge x' = x) \end{array} \right) \\ & \wedge i' = i + 1 \\ & \wedge n' = n \end{aligned}$$

α

..... recurrences

$$\begin{aligned} i^{(k)} &= i^{(k-1)} + 1 \\ x^{(k)} + y^{(k)} &= x^{(k-1)} + y^{(k-1)} + i \\ x^{(k)} &\geq x^{(k-1)} \\ y^{(k)} &\geq y^{(k-1)} \end{aligned}$$

β

..... loop abstraction

$$\exists k. k \geq 0 \wedge i' = i + k \wedge x' + y' = x + y + k(k+1)/2 + ki_0 \wedge x' \geq x \wedge y' \geq y$$

CRA's iteration operator

```
while(i < n):  
  if (*):  
    x := x + i  
  else  
    y := y + i  
  i := i + 1
```

..... loop body

$$\begin{aligned} & i < n \\ & \wedge \left(\begin{array}{l} (x' = x + i \wedge y' = y) \\ \vee (y' = y + i \wedge x' = x) \end{array} \right) \\ & \wedge i' = i + 1 \\ & \wedge n' = n \end{aligned}$$

..... recurrences

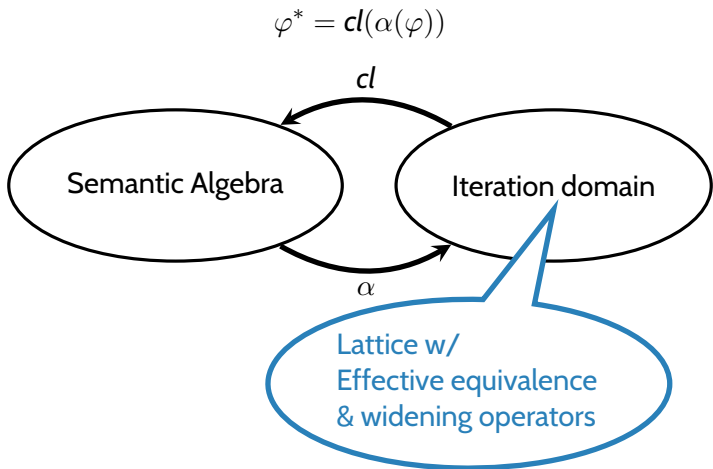
$$\begin{aligned} i^{(k)} &= i^{(k-1)} + 1 \\ x^{(k)} + y^{(k)} &= x^{(k-1)} + y^{(k-1)} + i \\ x^{(k)} &\geq x^{(k-1)} \\ y^{(k)} &\geq y^{(k-1)} \end{aligned}$$

Polyhedron

..... loop abstraction

$$\exists k. k \geq 0 \wedge i' = i + k \wedge x' + y' = x + y + k(k+1)/2 + ki_0 \wedge x' \geq x \wedge y' \geq y$$

Iteration domains



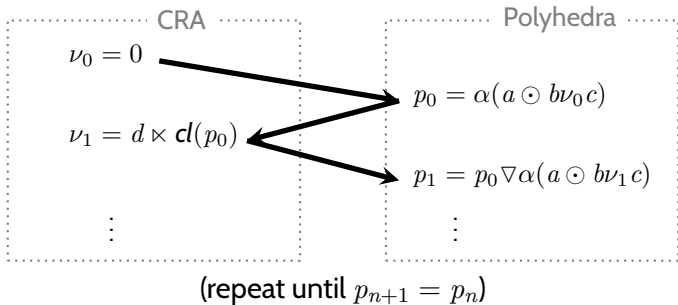
*Key idea: we have an opportunity to detect / enforce convergence at every place we apply the * operator.*

$$X = aXbXc + d \rightsquigarrow X = d \times (a \odot bXc)^*$$



All variables appear below *

$$X = aXbXc + d \rightsquigarrow X = d \times (a \odot bXc)^*$$



Over-approximating recursive procedures

Given a system of recursive equations describing a set of paths,

- 1 Using the tensor domain, rewrite the system so that every variable appears below a star (similar to Gauss-Jordan elimination)
- 2 Compute solution to resulting system iteratively, using iteration domains to detect and enforce convergence.

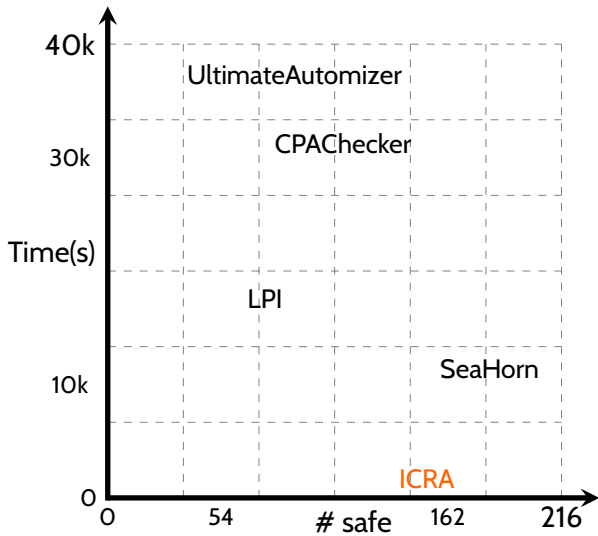
Implementation & Evaluation

ICRA was implemented on top of CRA and WALi

- (uses Cil C frontend, Z3 SMT solver, Apron abstract domain library)

Experimental set-up

- Ran on 216 *safe* benchmarks collected from SV-Comp, C4B (resource bound verification problems), and misc examples
- Compare with SeaHorn, CPAchecker, LPI, Ultimate Automizer



Summary

Algebraic analyses can be extended to recursive procedures using

- 1 *Tensor domains*, to re-arrange recursion into loops
- 2 *Iteration domains*, to detect and enforce convergence

Experimental results

Benchmark Suite	Total	ICRA		UAut.		CPA		LPI		SEA	
	#A	Time	#A	Time	#A	Time	#A	Time	#A	Time	#A
recursive	18/7	40.7	7	1952.1	8	1817.8	10	62.0	0	1334.0	14
rec.-simple	36/38	168.7	21	6979.3	28	2760.4	32	179.5	3	743.8	36
Rec. (tot.)	54/45	209.4	28	8931.4	36	4578.1	42	241.5	3	2077.8	50
loop-accel.	19/16	20.8	13	6696.5	7	4565.7	13	4227.7	13	2713.1	15
loop-invgen	18/1	53.1	16	1876.2	7	4909.6	2	1282.3	15	506.0	16
loop-lit	15/1	316.5	12	2722.9	5	2720.6	7	444.9	13	305.2	13
loops	34/32	209.7	22	3984.1	19	4380.1	28	3356.8	26	1821.5	27
loop-new	8/0	304.8	7	2147.9	1	1866.1	3	929.6	4	302.8	6
Loops (tot.)	94/50	904.8	70	17427.6	39	18442.2	53	10241.3	71	5648.6	77
C4B	35/0	30.3	30	6156.6	1	7817.8	2	6726.7	0	1867.6	29
misc	10/4	76.7	10	492.2	8	334.4	7	332.2	1	5.3	10
rec-loop-lit	15/1	312.7	9	2755.5	3	51.0	6	40.4	0	922.6	12
rec-loop-new	8/0	6.2	5	1546.9	2	25.6	2	19.6	0	905.7	4
Misc.-Rec.	33/5	395.6	24	4794.6	13	410.9	15	392.2	1	1833.7	26