An Abstract Domain to Infer Octagonal Constraints with Absolute Value

Liqian Chen[1] Jiangchao Liu[2] Antoine Miné[2,3]
Deepak Kapur[4] Ji Wang[1]

[1]National University of Defense Technology, China
[2]École Normale Supérieure, Paris, France
[3]CNRS, France
[4]University of New Mexico, NM, USA

11/09/2014 – SAS 2014

Overview

- Motivation
- The octagon abstract domain
- A domain of octagonal constraints with absolute value
- Experiments
- Conclusion

Motivation

Motivation

Goal: numerical static analysis

discover numerical properties of a program statically and automatically

Applications:

- check for runtime errors (e.g., arithmetic overflows, division by zero, array out-of-bounds, etc.)
- optimize programs
- . . .

Theoretical framework: **abstract interpretation**

to design static analyses that are

- sound by construction (no behavior is omitted)
- approximate (trade-off between precision and efficiency)

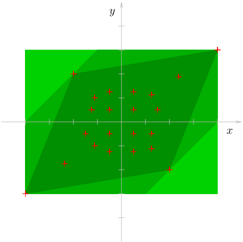
Motivation

Abstract domain: key ingredient of abstract interpretation

- a specific kind of computer-representable properties
  - e.g., a family of constraints
- sound (but maybe incomplete) algorithms for semantic actions
  - e.g., join, meet, widening,. . .

Numerical abstract domains

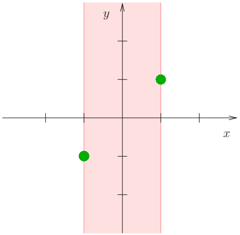
- infer relationships among numerical variables
- examples
  - non-relational: **intervals** ($a \leq x \leq b$)
  - weakly relational: **octagons** ($\pm x \pm y \leq c$)
  - strongly relational: **polyhedra** ($\Sigma_k a_k x_k \leq b$)
  - . . .



Motivation

Convexity limitations: a motivating example

```
real x, y;
x ← 1;
y ← 1;
while (true) {
        x ← −x;
        y ← 1/x;   ①
}
```



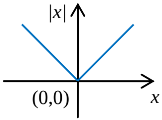
Loc	Most abstract domains	Concrete semantics
①	$x \in [-1, 1]$ $y \in [-\infty, +\infty]$	$(x = -1 \wedge y = -1)$ $\vee (x = 1 \wedge y = 1)$
	Division-by-zero?	Safe !

Motivation

Absolute Value (AV): $y = |x|$

- piecewise linear expressiveness

$$|x| = \begin{cases} x & \text{if } x \geq 0 \\ -x & \text{if } x < 0 \end{cases}$$



Possible applications

- to encode disjunctions of linear constraints in the program
  - $(x \leq -1 \vee x \geq 1) \iff |x| \geq 1$
  - $(x \neq 1 \vee y \neq 2) \iff |x - 1| + |y - 2| > 0$
- AV functions in C: $abs(), fabs(), \ldots$
- MiniMax functions in C: $fmax(), fmin(), \ldots$
  - e.g., $\max(x, y) = \frac{1}{2}(|x - y| + x + y)$
- abstractions for floating-point rounding errors
  - $|R_{\mathbf{f},r}(x) - x| \leq \varepsilon_{\text{rel}} \cdot |x| + \varepsilon_{\text{abs}}$ (float: $\varepsilon_{\text{rel}} = 2^{-23}, \varepsilon_{\text{abs}} = 2^{-149}$)

Motivation

The domain of linear absolute value inequalities: ($\Sigma_k a_k x_k + \Sigma_k b_k |x_k| \leq b$)
[Chen et al. ESOP'11]

- idea: extending polyhedra domain ($\Sigma_k a_k x_k \leq b$) with absolute value
- pros: piecewise linear expressiveness
- cons: exponential complexity

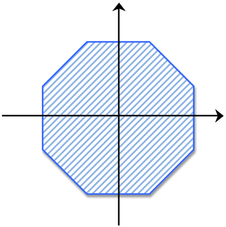
New idea: weakly relational abstract domain with absolute value

- goal: scalable with non-convex expressiveness
- first choice: extending the octagon domain with absolute value
  - octagons: scalable, widely used in practice (e.g., in ASTRÉE)

The octagon abstract domain

The octagon abstract domain

$$\begin{cases} x & & \leq a_1 \\ -x & & \leq a_2 \\ & y & \leq a_3 \\ & -y & \leq a_4 \\ x & +y & \leq a_5 \\ x & -y & \leq a_6 \\ -x & +y & \leq a_7 \\ -x & -y & \leq a_8 \end{cases}$$



The octagon abstract domain : [Miné 01]

- weakly relational: invariants of the form $\pm x \pm y \leq c$
- representation: Difference Bound Matrix (DBM)
- key operation: shortest-path closure via Floyd-Warshall algorithm
- scalable: $\mathcal{O}(n^2)$ in memory and $\mathcal{O}(n^3)$ in time

The octagon abstract domain

Domain representation

- efficient encoding: DBM
- idea: rewrite octagonal constraints on $V = \{V_1, \dots, V_n\}$ as potential constraints on $V' = \{V'_1, \dots, V'_{2n}\}$ where
  - V'_{2k-1} represents $+V_k$
  - V'_{2k} represents $-V_k$

the constraint	is represented by		
$V_i - V_j \leq a$	$V'_{2i-1} - V'_{2j-1} \leq a$	and	$V'_{2j} - V'_{2i} \leq a$
$V_i + V_j \leq b$	$V'_{2i-1} - V'_{2j} \leq b$	and	$V'_{2j-1} - V'_{2i} \leq b$
$-V_i - V_j \leq c$	$V'_{2i} - V'_{2j-1} \leq c$	and	$V'_{2j} - V'_{2i-1} \leq c$
$V_i \leq d$	$V'_{2i-1} - V'_{2i} \leq 2d$		
$-V_i \leq e$	$V'_{2i} - V'_{2i-1} \leq 2e$		

The octagon abstract domain

Key domain operation: closure

x-y octagon + y-z octagon ⇒ x-z octagon

$$\left\{\begin{array}{rrl} x & & \leq a_1 \\ -x & & \leq a_2 \\ & y & \leq a_3 \\ & -y & \leq a_4 \\ x & +y & \leq a_5 \\ x & -y & \leq a_6 \\ -x & +y & \leq a_7 \\ -x & -y & \leq a_8 \end{array}\right. + \left\{\begin{array}{rrl} y & & \leq a_3 \\ -y & & \leq a_4 \\ & z & \leq a'_3 \\ & -z & \leq a'_4 \\ y & +z & \leq a'_5 \\ y & -z & \leq a'_6 \\ -y & +z & \leq a'_7 \\ -y & -z & \leq a'_8 \end{array}\right. \Rightarrow \left\{\begin{array}{rrl} x & & \leq ? \\ -x & & \leq ? \\ & z & \leq ? \\ & -z & \leq ? \\ x & +z & \leq ? \\ x & -z & \leq ? \\ -x & +z & \leq ? \\ -x & -z & \leq ? \end{array}\right.$$

- Floyd-Warshall algorithm

```
for k ← 0 to |V| − 1
    for i ← 0 to |V| − 1
        for j ← 0 to |V| − 1
            d[i, j] ← min(d[i, j], d[i, k]+d[k, j])    /* i ⇝(d_ik) k ⇝(d_kj) j */
```

Complexity: $\mathcal{O}(|V|^3)$

An abstract domain of octagonal constraints with absolute value

Domain representation

Octagonal constraints with absolute value

- octagonal constraints: $\pm x \pm y \leq a$
- absolute value on one variable: $\pm x \pm |y| \leq b$
- absolute value on two variables: $\pm |x| \pm |y| \leq c$

Note: positive coefficients over AV terms can be removed

Theorem ([Chen et al. ESOP'11])

Any AV inequality

$$\sum_i a_i x_i + \sum_{i \neq p} b_i |x_i| + b_p |x_p| \leq c$$

where $b_p > 0$, can be reformulated as a conjunction of two AV inequalities

$$\begin{cases} \sum_i a_i x_i + \sum_{i \neq p} b_i |x_i| + b_p x_p \leq c \\ \sum_i a_i x_i + \sum_{i \neq p} b_i |x_i| - b_p x_p \leq c \end{cases}$$

Domain representation

Concise representation: 3 parts

- octagonal constraints: $\pm x \pm y \leq a$
- absolute value on one variable: $-|x| \pm y \leq b, \pm x - |y| \leq c$
- absolute value on two variables: $-|x| - |y| \leq d$

x				$\leq$	a_1
$-x$				$\leq$	a_2
		y		$\leq$	a_3
		$-y$		$\leq$	a_4
x		$+y$		$\leq$	a_5
x		$-y$		$\leq$	a_6
$-x$		$+y$		$\leq$	a_7
$-x$		$-y$		$\leq$	a_8
	$-\|x\|$			$\leq$	b_1
			$-\|y\|$	$\leq$	b_2
	$-\|x\|$	$+y$		$\leq$	b_3
	$-\|x\|$	$-y$		$\leq$	b_4
x			$-\|y\|$	$\leq$	b_5
$-x$			$-\|y\|$	$\leq$	b_6
	$-\|x\|$		$-\|y\|$	$\leq$	c_1

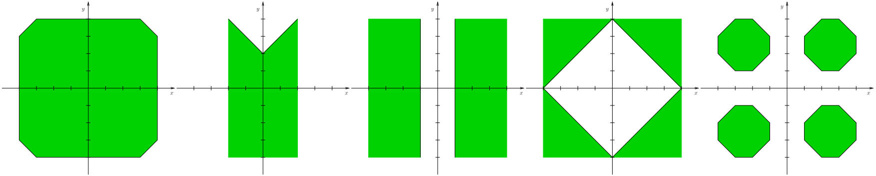
	DBM							
	x	$-x$	$\|x\|$	$-\|x\|$	y	$-y$	$\|y\|$	$-\|y\|$
x		$2a_2$						
$-x$	$2a_1$							
$\|x\|$				$2b_1$				
$-\|x\|$								
y	a_6	a_8		b_4		$2a_4$		
$-y$	a_5	a_7		b_3	$2a_3$			
$\|y\|$	b_5	b_6		c_1				$2b_2$
$-\|y\|$								

Domain representation

Concise representation: 3 parts

- octagonal constraints: $\pm x \pm y \leq a$
- absolute value on one variable: $-|x| \pm y \leq b, \pm x - |y| \leq c$
- absolute value on two variables: $-|x| - |y| \leq d$

Geometric shape : non-convex



octagons $-|x| \pm y \leq c$ $-|x| \leq c$ $-|x| - |y| \leq c$ AV octagons

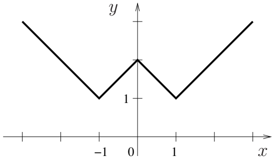
Domain representation

Expressiveness limitation: $-|x| - |y| \le c$

- $|\cdot|$ applies to only (single) variables rather than expressions

An example: $y = ||x| - 1| + 1$, i.e.,

$$y = \begin{cases} -x & \text{if } x \le -1 \\ x + 2 & \text{if } -1 \le x \le 0 \\ 2 - x & \text{if } 0 \le x \le 1 \\ x & \text{if } x \ge 1 \end{cases}$$



Expressiveness lifting

- introduce new auxiliary variables to denote expressions inside the AV function
- e.g., $\{y = |\nu| + 1, \nu = |x| - 1\}$

Domain operation

Closure:

x vs. y

x				$\leq a_1$
$-x$				$\leq a_2$
		y		$\leq a_3$
		$-y$		$\leq a_4$
x		$+y$		$\leq a_5$
x		$-y$		$\leq a_6$
$-x$		$+y$		$\leq a_7$
$-x$		$-y$		$\leq a_8$
	$-\lvert x\rvert$			$\leq b_1$
			$-\lvert y\rvert$	$\leq b_2$
	$-\lvert x\rvert$	$+y$		$\leq b_3$
	$-\lvert x\rvert$	$-y$		$\leq b_4$
x			$-\lvert y\rvert$	$\leq b_5$
$-x$			$-\lvert y\rvert$	$\leq b_6$
	$-\lvert x\rvert$		$-\lvert y\rvert$	$\leq c_1$

$+$

y vs. z

y				$\leq a_3$
$-y$				$\leq a_4$
		z		$\leq a_3'$
		$-z$		$\leq a_4'$
y		$+z$		$\leq a_5'$
y		$-z$		$\leq a_6'$
$-y$		$+z$		$\leq a_7'$
$-y$		$-z$		$\leq a_8'$
	$-\lvert y\rvert$			$\leq b_2$
			$-\lvert z\rvert$	$\leq b_2'$
	$-\lvert y\rvert$	$+z$		$\leq b_3'$
	$-\lvert y\rvert$	$-z$		$\leq b_4'$
y			$-\lvert z\rvert$	$\leq b_5'$
$-y$			$-\lvert z\rvert$	$\leq b_6'$
	$-\lvert y\rvert$		$-\lvert z\rvert$	$\leq c_1'$

$\Rightarrow$

x vs. z

x				$\leq ?$
$-x$				$\leq ?$
		z		$\leq ?$
		$-z$		$\leq ?$
x		$+z$		$\leq ?$
x		$-z$		$\leq ?$
$-x$		$+z$		$\leq ?$
$-x$		$-z$		$\leq ?$
	$-\lvert x\rvert$			$\leq ?$
			$-\lvert z\rvert$	$\leq ?$
	$-\lvert x\rvert$	$+z$		$\leq ?$
	$-\lvert x\rvert$	$-z$		$\leq ?$
x			$-\lvert z\rvert$	$\leq ?$
$-x$			$-\lvert z\rvert$	$\leq ?$
	$-\lvert x\rvert$		$-\lvert z\rvert$	$\leq ?$

Domain operation

A trivial **strong** closure: via orthant enumeration (over 2^n orthants)

- ask $-|x| + z \leq ?$ in each orthant via Floyd-Warshall algorithm
- the final answer will be the greatest result of all orthants

$$2^4 orthants \begin{cases} \begin{array}{cccc} x & y & z & w \\ \hline + & + & + & + \\ + & + & + & - \\ + & + & - & + \\ + & + & - & - \\ & \dots & & \\ - & - & - & - \\ \hline \end{array} \end{cases}$$

Complexity: $\mathcal{O}(2^n \times n^3)$

Domain operation

A **weak** closure: `WeakCloVia3Sign()` of complexity $\mathcal{O}(n^3)$

```
1: for k ← 0 to |V| − 1
2:     for i ← 0 to |V| − 1
3:         for j ← 0 to |V| − 1
4:             Combine AVO_ik and AVO_kj to tighten AVO_ij by orthant enumeration;
                                                          /* only 8 orthants*/
```

- enumerating the signs of 3 variables each time
- as precise as strong closure for 3 variables
- but weaker than strong closure for more than 3 variables

Example

$\{y \le 24, -|y| + x \le 10, -s - |x| \le 36, -|s| - z \le 8, -z - y \le 84, s + y \le 80\}$,

- strong closure: $x - z \le 112$
- `WeakCloVia3Sign()` : $x - z \le 142$

Domain operation

Another cheaper **weak** closure: `WeakCloVia1Sign()` of complexity $\mathcal{O}(n^3)$

> **for** $k \leftarrow 0$ **to** $|V|-1$
>   **for** $i \leftarrow 0$ **to** $|V|-1$
>   **for** $j \leftarrow 0$ **to** $|V|-1$
>   Combine AVO_{ik} and AVO_{kj} to tighten AVO_{ij} when $x_k \geq 0$;
>   Combine AVO_{ik} and AVO_{kj} to tighten AVO_{ij} when $x_k \leq 0$;
> /* only 2 orthants*/

- enumerating the signs of 1 variables each time
- weaker than the previous weak closure `WeakCloVia3Sign()`

Example

$\{y - x \leq 24, -z - |x| \leq 6, x - z \leq 16, y - |z| \leq 10, y - z \leq 50\}$,

- `WeakCloVia3Sign()`: $y - z \leq 40$
- `WeakCloVia1Sign()`: $y - z \leq 50$

Domain operation

Other domain operations for static analysis

- transfer functions (such as branch tests and assignments)
- join
- meet
- extrapolation (such as widening and narrowing)
- projection
- emptiness test
- inclusion

Implementation

- in the numerical abstract domain library APRON [Jeannet Miné 09]

Supporting strict inequalities

Supporting strict inequalities

- representation: maintain a boolean matrix S of the same size as the AVO matrix M

$$S_{ij} \stackrel{\text{def}}{=} \begin{cases} 0 & \text{if } V''_j - V''_i < M_{ij} \\ 1 & \text{if } V''_j - V''_i \leq M_{ij} \end{cases}$$

- operations: over the pair (M_{ij}, S_{ij})
  - ordering: $(M_{ij}, S_{ij}) \sqsubseteq (M'_{ij}, S'_{ij}) \stackrel{\text{def}}{\iff} (M_{ij} < M'_{ij} \vee (M_{ij} = M'_{ij} \wedge S_{ij} \leq S'_{ij}))$
  - emptiness test: $\exists i, M_{ii} < 0 \vee (S_{ii} = 0 \wedge M_{ii} = 0)$
  - propagation: $(M_{ik}, S_{ik}) + (M_{kj}, S_{kj}) \stackrel{\text{def}}{=} (M_{ik} + M_{kj}, S_{ik} \& S_{kj})$
  - ...

Example analyses

An example[a]

- involving non-convex constraints (due to disjunctions, the usage of the AV function) as well as strict inequalities

```
static void p_line16_primary (...)  {
    real  dx, dy, x, y, slope;
    ...
    if (dx == 0.0 && dy == 0.0)
        return;
 ① if (fabs(dy) > fabs(dx)) {
     ② slope = dx / dy;
        ...
    } else {
     ③ slope = dy / dx;
        ...
    } }
```

Loc	AV octagons
①	$-\lvert dx\rvert - \lvert dy\rvert < 0$
②	$-\lvert dx\rvert - \lvert dy\rvert < 0 \wedge$ $\lvert dx\rvert - \lvert dy\rvert < 0 \wedge$ $-\lvert dy\rvert < 0$
③	$-\lvert dx\rvert - \lvert dy\rvert < 0 \wedge$ $-\lvert dx\rvert + \lvert dy\rvert \leq 0 \wedge$ $-\lvert dx\rvert < 0$

[a] extracted from the XTide package and used in the Donut domain [Ghorbal et al. 12]

Experiments

Preliminary experimental results

NECLA Benchmarks: Division-by-zero False Alarms [Ghorbal et al. 12]

- show commonly used practices that developers use to protect a division-by-zero
- extracted from available free C source code of various projects
- "involve non-convex tests (using for instance disjunctions or the AV function), strict inequalities tests, ..."

program	donut domain		octagons		AV octagons	
	invariants	♯FP	invariants	♯FP	invariants	♯FP
motiv(if)	$dy \neq 0$	0	$dy \in [-\infty, +\infty]$	1	$\lvert dy\rvert > 0$	0
motiv(else)	$dx \neq 0$	0	$dx \in [-\infty, +\infty]$	1	$\lvert dx\rvert > 0$	0
gpc	$den \notin [-0.1, 0.1]$	0	$den \in [-\infty, +\infty]$	1	$\lvert den\rvert > 0.1$	0
goc	$d \notin [-0.09, 0.09]$	0	$d \in [-\infty, +\infty]$	1	$\lvert d\rvert \geq 0.1$	0
x2	$Dx \neq 0$	0	$Dx \in [-\infty, +\infty]$	1	$\lvert Dx\rvert > 0$	0
xcor	$usemax \notin [1, 10]$	1	$usemax \geq 0$	1	$usemax > 0$	0

Preliminary experimental results

Experiments on ASTRÉE

- a set of large embedded industrial C codes
- compare octagons and AVO (disabling disjunctive domains in ASTRÉE)

code	size (KLoc)	octagons		AV octagons		result comparison	
		time (s)	♯alarm	time (s)	♯alarm	♯alarm reduction	time increase
$P1$	154	6216	881	7687	881	0	23.66%
$P2$	186	6460	1114	7854	1114	0	21.58%
$P3$	103	1112	403	2123	403	0	90.92%
$P4$	493	17195	4912	38180	4912	0	122.04%
$P5$	661	18949	7075	43660	7070	5	130.41%
$P6$	616	34639	8192	70541	8180	12	103.65%
$P7$	2428	99853	10980	217506	10959	21	117.83%
$P8$	3	517	0	581	0	0	12.38%
$P9$	18	534	16	670	16	0	25.47%
$P10$	26	1065	102	1133	102	0	6.38%

Conclusion

Summary

- **the AVO domain**: extending octagons with absolute value
  - to infer invariants in the form of

    $$\{\pm x \pm y \leq a, \pm x \pm |y| \leq b, \pm|x| \pm |y| \leq c\}$$

  - **more precise** than octagon domain but with the same magnitude of complexity $\mathcal{O}(n^3)$
  - **non-convexity** expressiveness
  - support **strict** inequalities

Future Work

- more choices for closure algorithm
  - is the strong closure problem NP-hard?
- consider AV octagonal constraints with integers as constant terms