michelepasqua.github.io/
michele.pasqua@univr.it

UNIVERSITÀ di VERONA
Dipartimento di INFORMATICA

Principles and Applications of Abstract Interpretation

Abstract Domains

Michele Pasqua, PhD

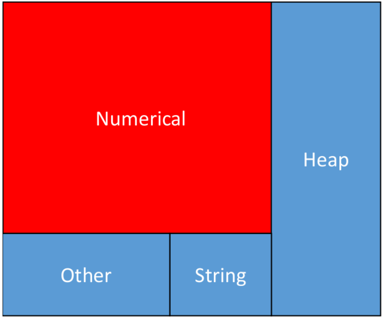
Verona, IT

October 2024

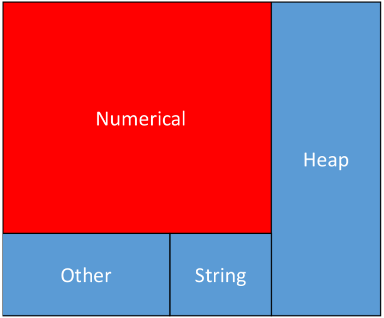
PhD Course @ UniVR 2024/2025

Almost all domains are for trace properties

Almost all domains are for trace properties



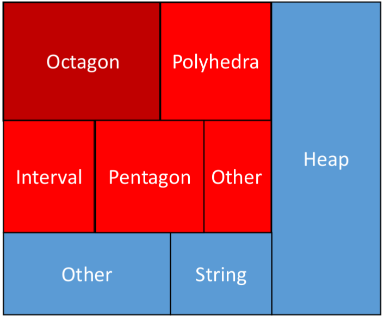
Almost all domains are for trace properties



Buffer overflow
Division by zero
Integer overflow
Alias analysis
Floating-point errors

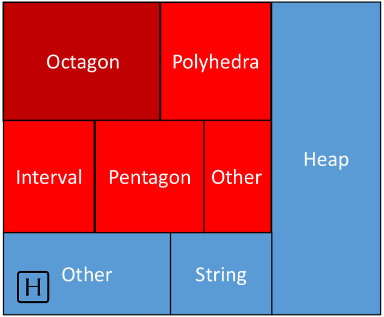
UNIVERSITÀ di VERONA
Dipartimento di INFORMATICA

Almost all domains are for trace properties



Buffer overflow
Division by zero
Integer overflow
Alias analysis
Floating-point errors

Almost all domains are for trace properties ... almost all



Buffer overflow
Division by zero
Integer overflow
Alias analysis
Floating-point errors
Information flows
Data races

Numerical Abstractions

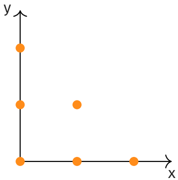
We aim to abstract elements in $\wp(\mathbb{M}) = \wp(\mathbb{L} \to \mathbb{V})$ e.g., $\mathbb{V} = \mathbb{Z}$

- Find an invariant for all values of all variables in $X \in \wp(\mathbb{M})$

We aim to abstract elements in $\wp(\mathbb{M}) = \wp(\mathbb{L} \to \mathbb{V})$ e.g., $\mathbb{V} = \mathbb{Z}$

- Find an invariant for all values of all variables in $X \in \wp(\mathbb{M})$

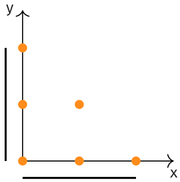
For instance, with $X = \{\mathbb{m} \in \mathbb{M} \mid \mathbb{m}(\mathsf{x}), \mathbb{m}(\mathsf{y}) \in [0, 2] \wedge \mathsf{x} + \mathsf{y} \leq 2\}$



We aim to abstract elements in $\wp(\mathbb{M}) = \wp(\mathbb{L} \rightarrow \mathbb{V})$ e.g., $\mathbb{V} = \mathbb{Z}$

- Find an invariant for all values of all variables in $X \in \wp(\mathbb{M})$

For instance, with $X = \{\mathbb{m} \in \mathbb{M} \mid \mathbb{m}(\mathtt{x}), \mathbb{m}(\mathtt{y}) \in [0, 2] \wedge \mathtt{x} + \mathtt{y} \leq 2\}$



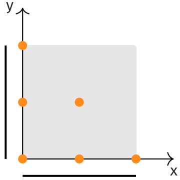
intervals

We aim to abstract elements in $\wp(\mathbb{M}) = \wp(\mathbb{L} \to \mathbb{V})$

e.g., $\mathbb{V} = \mathbb{Z}$

- Find an invariant for all values of all variables in $X \in \wp(\mathbb{M})$

For instance, with $X = \{\mathbb{m} \in \mathbb{M} \mid \mathbb{m}(\mathtt{x}), \mathbb{m}(\mathtt{y}) \in [0, 2] \wedge \mathtt{x} + \mathtt{y} \leq 2\}$



intervals

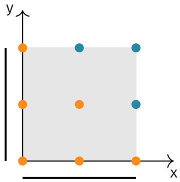
UNIVERSITÀ di VERONA Dipartimento di INFORMATICA

We aim to abstract elements in $\wp(\mathbb{M}) = \wp(\mathbb{L} \to \mathbb{V})$ e.g., $\mathbb{V} = \mathbb{Z}$

- Find an invariant for all values of all variables in $X \in \wp(\mathbb{M})$

For instance, with $X = \{\mathbb{m} \in \mathbb{M} \mid \mathbb{m}(\mathrm{x}), \mathbb{m}(\mathrm{y}) \in [0, 2] \wedge \mathrm{x} + \mathrm{y} \leq 2\}$

non-relational



intervals

UNIVERSITÀ di VERONA
Dipartimento di INFORMATICA

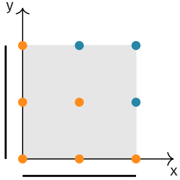
We aim to abstract elements in $\wp(\mathbb{M}) = \wp(\mathbb{L} \to \mathbb{V})$

e.g., $\mathbb{V} = \mathbb{Z}$

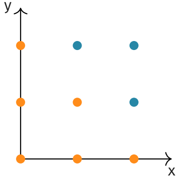
- Find an invariant for all values of all variables in $X \in \wp(\mathbb{M})$

For instance, with $X = \{\mathbb{m} \in \mathbb{M} \mid \mathbb{m}(\mathtt{x}), \mathbb{m}(\mathtt{y}) \in [0,2] \wedge \mathtt{x} + \mathtt{y} \leq 2\}$

non-relational



intervals

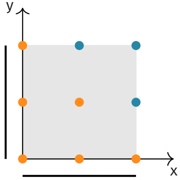


We aim to abstract elements in $\wp(\mathbb{M}) = \wp(\mathbb{L} \to \mathbb{V})$ e.g., $\mathbb{V} = \mathbb{Z}$

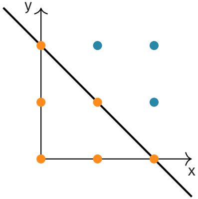
- Find an invariant for all values of all variables in $X \in \wp(\mathbb{M})$

For instance, with $X = \{\mathbb{m} \in \mathbb{M} \mid \mathbb{m}(x), \mathbb{m}(y) \in [0, 2] \wedge x + y \leq 2\}$

non-relational



intervals



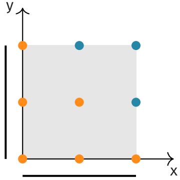
linear inequalities

We aim to abstract elements in $\wp(\mathbb{M}) = \wp(\mathbb{L} \rightarrow \mathbb{V})$ e.g., $\mathbb{V} = \mathbb{Z}$

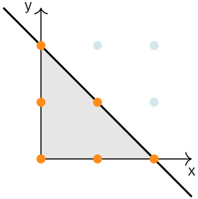
- Find an invariant for all values of all variables in $X \in \wp(\mathbb{M})$

For instance, with $X = \{\mathbb{m} \in \mathbb{M} \mid \mathbb{m}(\mathtt{x}), \mathbb{m}(\mathtt{y}) \in [0, 2] \wedge \mathtt{x} + \mathtt{y} \leq 2\}$

non-relational



intervals



linear inequalities

relational

Forget relations between variables by applying the cartesian abstraction

Forget relations between variables by applying the cartesian abstraction

Upper closure operator on $\wp(\mathbb{M})$:

$$\eta_{nr} \triangleq \gamma_{nr} \circ \alpha_{nr} = \lambda X \,.\, \{\mathbb{m} \in \mathbb{M} \mid \forall \mathtt{x} \in \mathbb{X}\, \exists \mathbb{m}' \in X \,.\, \mathbb{m}(\mathtt{x}) = \mathbb{m}'(\mathtt{x})\}$$

Forget relations between variables by applying the cartesian abstraction

Upper closure operator on $\wp(\mathbb{M})$:

$$\eta_{nr} \triangleq \gamma_{nr} \circ \alpha_{nr} = \lambda X\,.\,\{\mathbb{m} \in \mathbb{M} \mid \forall \mathbb{x} \in \mathbb{X}\, \exists \mathbb{m}' \in X\,.\,\mathbb{m}(\mathbb{x}) = \mathbb{m}'(\mathbb{x})\}$$

A domain is non-relational if $\eta_{nr} \circ \gamma = \gamma$

- The domain cannot distinguish between X and X' when $\eta_{nr}(X) = \eta_{nr}(X')$

Forget relations between variables by applying the cartesian abstraction

Upper closure operator on $\wp(\mathbb{M})$:

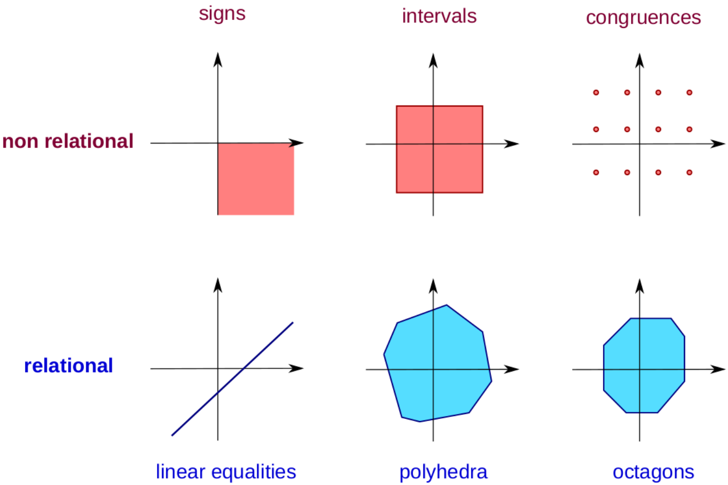
$$\eta_{nr} \triangleq \gamma_{nr} \circ \alpha_{nr} = \lambda X.\ \{\mathbb{m} \in \mathbb{M} \mid \forall \mathbb{x} \in \mathbb{X}\ \exists \mathbb{m}' \in X.\ \mathbb{m}(\mathbb{x}) = \mathbb{m}'(\mathbb{x})\}$$

A domain is non-relational if $\eta_{nr} \circ \gamma = \gamma$

- The domain cannot distinguish between X and X' when $\eta_{nr}(X) = \eta_{nr}(X')$

All domains seen so far (sign, intervals, constants) are non-relational

Numerical domain examples



© A. Miné

Prove relational invariants, of course

Prove relational invariants, of course

Prove some complex non-relational invariants with relational loop invariants

Prove relational invariants, of course

Prove some complex non-relational invariants with relational loop invariants

```
x = 0;
i = 1;
while (i < 5000) {
   if (rand(0,1) == 1)
      { x = x + 1 }
      { x = x - 1 };
   i = i + 1
}
```

The need for relational abstractions

Prove relational invariants, of course

Prove some complex non-relational invariants with relational loop invariants

```
1 x = 0;
2 i = 1;
3 while (i < 5000) {
    4 if (rand(0,1) == 1)
        { 5 x = x + 1 }
        { 6 x = x - 1 };
    7 i = i + 1
8 }
```

A non-relational analysis at 8 finds: $\mathtt{i} = 5000 \wedge \mathtt{x} \in \mathbb{Z}$

Prove relational invariants, of course

Prove some complex non-relational invariants with relational loop invariants

```
1 x = 0;
2 i = 1;
3 while (i < 5000) {
    4 if (rand(0,1) == 1)
        { 5 x = x + 1 }
        { 6 x = x - 1 };
    7 i = i + 1
8 }
```

A non-relational analysis at 8 finds: $\mathtt{i} = 5000 \wedge \mathtt{x} \in \mathbb{Z}$

The best invariant at 8 is: $\mathtt{i} = 5000 \wedge \mathtt{x} \in [-4999, 4999] \wedge \mathtt{x} \equiv_2 1$

The need for relational abstractions

Prove relational invariants, of course

Prove some complex non-relational invariants with relational loop invariants

```
1 x = 0;
2 i = 1;
3 while (i < 5000) {
    4 if (rand(0,1) == 1)
       { 5 x = x + 1 }
       { 6 x = x - 1 };
    7 i = i + 1
8 }
```

A non-relational analysis at 8 finds: $\mathtt{i} = 5000 \wedge \mathtt{x} \in \mathbb{Z}$

The best invariant at 8 is: $\mathtt{i} = 5000 \wedge \mathtt{x} \in [-4999, 4999] \wedge \mathtt{x} \equiv_2 1$

- Find the relational loop invariant $\text{-}\mathtt{i} < \mathtt{x} < \mathtt{i} \wedge \mathtt{x} + \mathtt{i} \equiv_2 1$ at 4
- Filter out by applying the negation of the guard at 3

Relational Numerical Abstractions

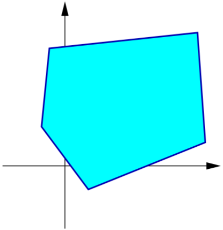
We look for invariants of the form:

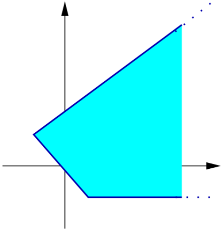
$$\bigwedge_{j} \left(\sum_{i=1}^{n} c_{ij} \cdot \mathsf{x}_i \geq d_j \right) \qquad c, d \in \mathbb{V}$$

We look for invariants of the form:

$$\bigwedge_j \left(\sum_{i=1}^{n} c_{ij} \cdot \mathsf{x}_i \geq d_j \right) \qquad c, d \in \mathbb{V}$$

Polyhedra domain $\mathbb{M}^\sharp \triangleq \{\text{closed convex polyhedra of } \mathbb{X} \to \mathbb{V}\}$





© A. Miné

Invariants are systems of affine inequalities, that can be expressed in matrix form $\langle \mathbf{M}, \vec{C} \rangle$

$$\gamma(\langle \mathbf{M}, \vec{C} \rangle) = \{\vec{X} \in \mathbb{V}^n \mid \mathbf{M} \times \vec{X} \geq \vec{C}\} \simeq \left\{\sum_{i=1}^{n} c_{ij} \cdot \mathsf{x}_i \geq d_j\right\}$$

given $\mathbf{M} \in \mathbb{V}^{m \times n}$ and $\vec{C} \in \mathbb{V}^m$

we assume $\mathbb{V} \in \{\mathbb{Q}, \mathbb{R}\}$

Invariants are systems of affine inequalities, that can be expressed in matrix form $\langle \mathbf{M}, \vec{C} \rangle$

$$\gamma(\langle \mathbf{M}, \vec{C} \rangle) = \{\vec{X} \in \mathbb{V}^n \mid \mathbf{M} \times \vec{X} \geq \vec{C}\} \simeq \left\{\sum_{i=1}^{n} c_{ij} \cdot \mathrm{x}_i \geq d_j\right\}$$

given $\mathbf{M} \in \mathbb{V}^{m \times n}$ and $\vec{C} \in \mathbb{V}^m$

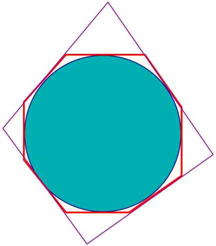
we assume $\mathbb{V} \in \{\mathbb{Q}, \mathbb{R}\}$

Invariants are systems of affine inequalities, that can be expressed in matrix form $\langle \mathbf{M}, \vec{C} \rangle$

$$\gamma(\langle \mathbf{M}, \vec{C} \rangle) = \{\vec{X} \in \mathbb{V}^n \mid \mathbf{M} \times \vec{X} \geq \vec{C}\} \simeq \left\{\sum_{i=1}^{n} c_{ij} \cdot \mathrm{x}_i \geq d_j\right\}$$

given $\mathbf{M} \in \mathbb{V}^{m \times n}$ and $\vec{C} \in \mathbb{V}^m$

- No best abstraction α (a disc has infinite approximations)
- No bound on the representation (exponential in practice)

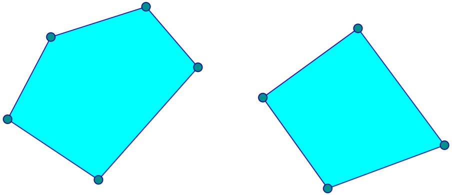


© A. Miné

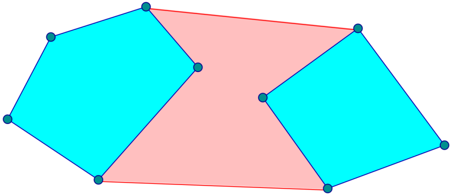
Topological closure of the convex hull of $\gamma(\mathbb{m}_1^\sharp) \cup \gamma(\mathbb{m}_2^\sharp)$

UNIVERSITÀ di VERONA Dipartimento di INFORMATICA

Topological closure of the convex hull of $\gamma(\mathbb{m}_1^\sharp) \cup \gamma(\mathbb{m}_2^\sharp)$



Topological closure of the convex hull of $\gamma(\mathbb{m}_1^\sharp) \cup \gamma(\mathbb{m}_2^\sharp)$



A convex polyhedron containing the polyhedra to join

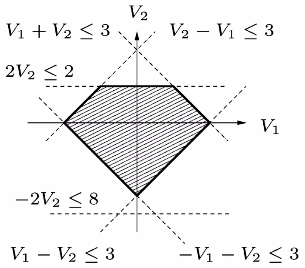
We look for invariants of the form:

$$\bigwedge \left(\pm \mathsf{x}_i \pm \mathsf{x}_j \leq c_k \right)$$

We look for invariants of the form:

$$\bigwedge \left(\pm x_i \pm x_j \le c_k\right)$$

Octagons domain: special case of polyhedra domain (symmetric and bounded)



Simpler matrix representation $\mathbf{M}$

$$\gamma(\mathbf{M}) = \{\vec{X} \in \mathbb{V}^n \mid \forall i,j \in [1,n] \,.\, \vec{X}_j - \vec{X}_i \leq \mathbf{M}_{i,j}\} \simeq \{\pm \mathrm{x}_i \pm \mathrm{x}_j \leq c_k\}$$

given $\mathbf{M} \in (\mathbb{V} \cup \{\infty\})^{2n \times 2n}$

difference bound matrix

we assume $\mathbb{V} \in \{\mathbb{Z}, \mathbb{Q}, \mathbb{R}\}$

Simpler matrix representation **M**

$$\gamma(\mathbf{M}) = \{\vec{X} \in \mathbb{V}^n \mid \forall i, j \in [1, n] \,.\, \vec{X}_j - \vec{X}_i \leq \mathbf{M}_{i,j}\} \simeq \left\{\pm \mathsf{x}_i \pm \mathsf{x}_j \leq c_k\right\}$$

given $\mathbf{M} \in (\mathbb{V} \cup \{\infty\})^{2n \times 2n}$

difference bound matrix

we assume $\mathbb{V} \in \{\mathbb{Z}, \mathbb{Q}, \mathbb{R}\}$

Simpler matrix representation $\mathbf{M}$

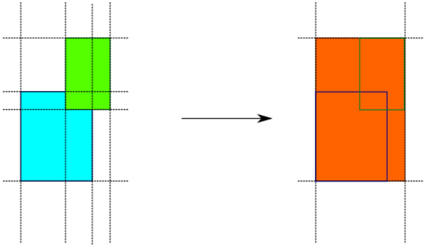
$$\gamma(\mathbf{M}) = \{\vec{X} \in \mathbb{V}^n \mid \forall i,j \in [1,n] \,.\, \vec{X}_j - \vec{X}_i \leq \mathbf{M}_{i,j}\} \simeq \{\pm \mathsf{x}_i \pm \mathsf{x}_j \leq c_k\}$$

given $\mathbf{M} \in (\mathbb{V} \cup \{\infty\})^{2n \times 2n}$ difference bound matrix

- Best abstraction α does exist
- Quadratic memory cost / Cubic time cost

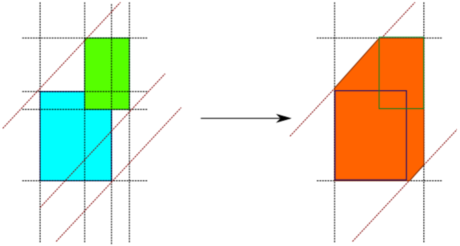
Element-wise maximum

Element-wise maximum



No better approximation than intervals

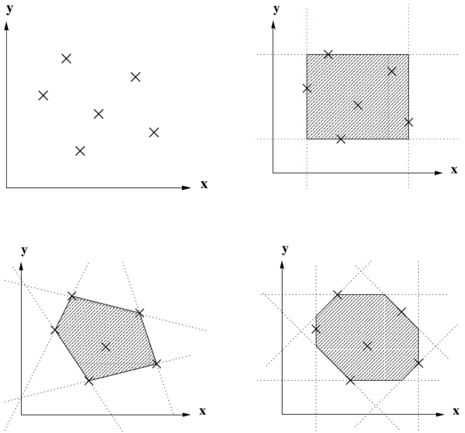
Element-wise maximum with closure (we add new constraints to increase precision)

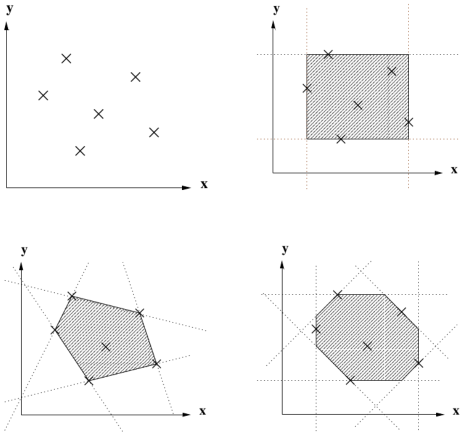


© A. Miné

Better approximation than intervals

Summary



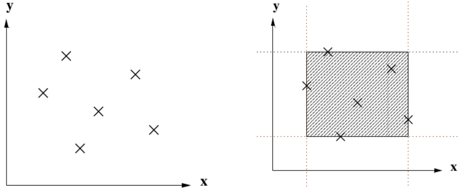


Intervals: *non-relational*

Complexity: *linear*

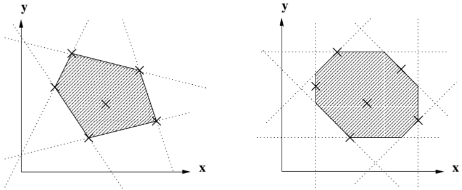
© A. Miné

Summary



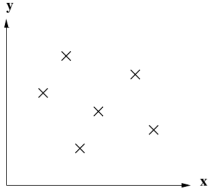
Intervals: *non-relational*

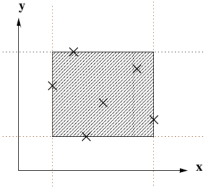
Complexity: *linear*



Octagons: *weakly relational*

Complexity: *quadratic/cubic*



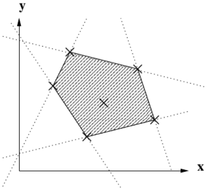


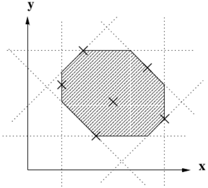
Intervals: *non-relational*

Complexity: *linear*

Polyhedra: *relational*

Complexity: *exponential*





Octagons: *weakly relational*

Complexity: *quadratic/cubic*

Thanks for the attention!

LaTeX is the way