AAA616: Program Analysis

Lecture 7 — The Octagon Abstract Domain

Hakjoo Oh
2016 Fall

Reference

- Antoine Miné. The Octagon Abstract Domain. Higher-Order and Symbolic Computation. Volume 19 Issue 1, March 2006

Numerical Abstract Domains

Infer numerical properties of program variables: e.g.,

- division by zero,
- array index out of bounds,
- integer overflow, etc.

Well-known numerical domains:

- interval domain: $x \in [l, u]$
- octagon domain: $\pm x \pm y \leq c$
- polyhedron domain (affine inequalities): $a_1x_1 + \cdots + a_nx_n \leq c$
- Karr's domain (affine equalities): $a_1x_1 + \cdots + a_nx_n = c$
- congruence domain: $x \in a\mathbb{Z} + b$

The octagon domain is a restriction of the polyhedron domain where each constraint involves at most two variables and unit coefficients.

Interval vs. Octagon

```
i = 0;
p = 0;

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

Interval analysis

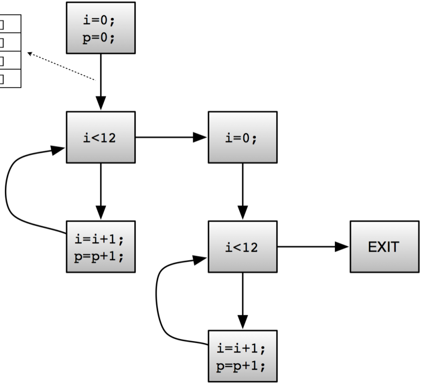
i	[12,12]
p	[0,+oo]

Octagon analysis

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

Example

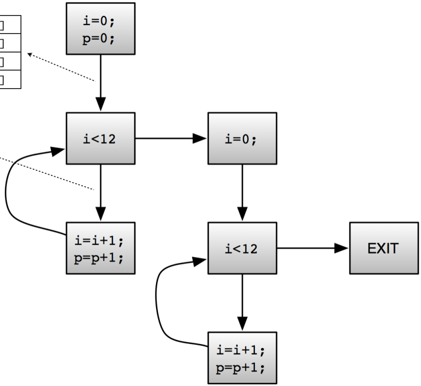
i	[0,0]
p	[0,0]
p-i	[0,0]
p+i	[0,0]



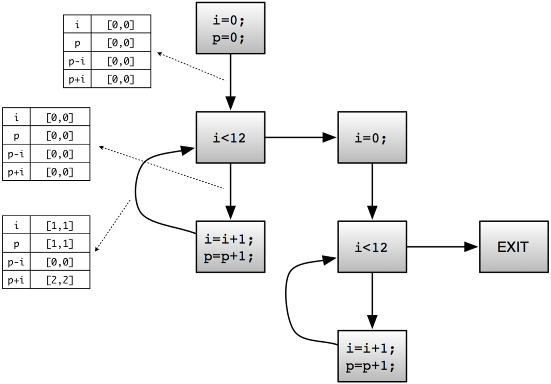
Example

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

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

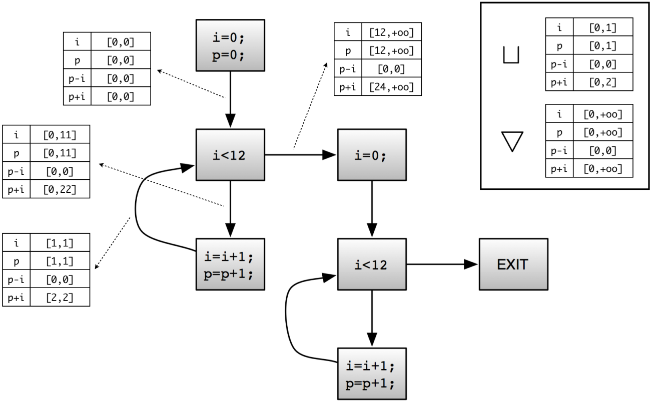


Example



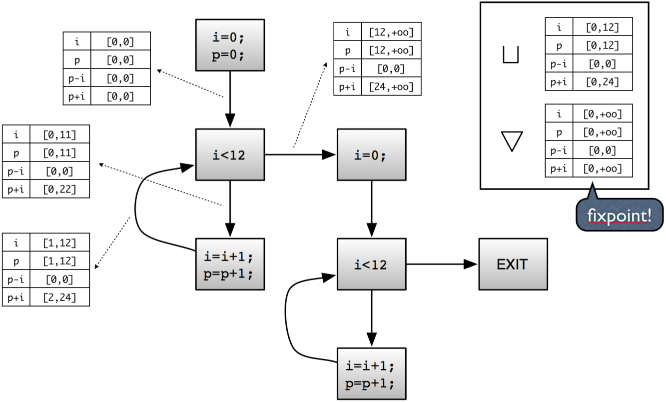
i [0,0]
p [0,0]
p-i [0,0]
p+i [0,0]
i [0,0]
p [0,0]
p-i [0,0]
p+i [0,0]
i [1,1]
p [1,1]
p-i [0,0]
p+i [2,2]
i=0;
p=0;
i<12
i=i+1;
p=p+1;
i=0;
i<12
EXIT
i=i+1;
p=p+1;

Example



i=0;
p=0;
i<12
i=0;
i=i+1;
p=p+1;
i<12
EXIT
i=i+1;
p=p+1;
i [0,0]
p [0,0]
p-i [0,0]
p+i [0,0]
i [0,11]
p [0,11]
p-i [0,0]
p+i [0,22]
i [1,1]
p [1,1]
p-i [0,0]
p+i [2,2]
i [12,+oo]
p [12,+oo]
p-i [0,0]
p+i [24,+oo]
⊔
i [0,1]
p [0,1]
p-i [0,0]
p+i [0,2]
▽
i [0,+oo]
p [0,+oo]
p-i [0,0]
p+i [0,+oo]

Example


i=0;
p=0;
i<12
i=i+1;
p=p+1;
i=0;
i<12
i=i+1;
p=p+1;
EXIT
i [0,0] p [0,0] p-i [0,0] p+i [0,0]
i [0,11] p [0,11] p-i [0,0] p+i [0,22]
i [1,12] p [1,12] p-i [0,0] p+i [2,24]
i [12,+oo] p [12,+oo] p-i [0,0] p+i [24,+oo]
⊔
i [0,12] p [0,12] p-i [0,0] p+i [0,24]
▽
i [0,+oo] p [0,+oo] p-i [0,0] p+i [0,+oo]
fixpoint!

Octagon

- A finite set $\mathbf{V} = \{V_1, \ldots, V_n\}$ of variables.
- An environment $\rho \in (\mathbf{V} \to \mathbb{I})$ $(\rho \in \mathbb{I}^n)$, where $\mathbb{I}$ can be $\mathbb{Z}, \mathbb{Q},$ or $\mathbb{R}$.
- An *octagonal constraint* is a constraint of the form $\pm V_i \pm V_j \leq c$.
- An *octagon* is the set of points satisfying a conjunction of octagonal constraints.

Potential Constraints

- A potential constraint (i.e., difference constraint): $V_i - V_j \leq c$.
- Let C be a set of potential constraints. C can be represented by a potential graph $\mathbf{G} = (\mathbf{V}, \hookrightarrow)$.
  - $(\hookrightarrow) \subseteq \mathbf{V} \times \mathbf{V} \times \mathbb{I}$

$$V_i \hookrightarrow_c V_j \iff (V_j - V_i \leq c) \in C$$

  - Assume that, for every V_i, V_j, there is at most one arc from V_i to V_j.
- A potential set of C is the set of points in $\mathbb{I}^n$ that satisfy C.

Difference Bound Matrices (DBMs)

A DBM $\mathbf{m}$ is a $n \times n$ square matrix, where n is the number of program variables, with elements in $\bar{\mathbb{I}} = \mathbb{I} \cup \{+\infty\}$.

- $\mathbf{m}_{ij} = \begin{cases} c & (V_i - V_j \leq c) \in C \\ +\infty & o.w. \end{cases}$
- $\mathbf{DBM} = \bar{\mathbb{I}}^{n \times n}$: the set of all DBMS.
- The potential set described by $\mathbf{m}$:

$$\gamma^{Pot}(\mathbf{m}) = \{(v_1, \ldots, v_n) \in \mathbb{I}^n \mid \forall i, j. v_j - v_i \leq \mathbf{m}_{ij}\}$$

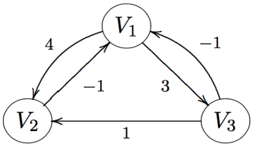
Example

(a)
$$\begin{cases} V_2 - V_1 \le 4 \\ V_1 - V_2 \le -1 \\ V_3 - V_1 \le 3 \\ V_1 - V_3 \le -1 \\ V_2 - V_3 \le 1 \end{cases}$$

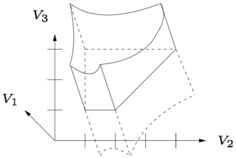
(b)

i \ j	1	2	3
1	$+\infty$	4	3
2	-1	$+\infty$	$+\infty$
3	-1	1	$+\infty$

(c)



(d)



Encoding Octagonal Constraints as Potential Constraints

- $\mathbf{V} = \{V_1, \ldots, V_n\}$: the set of program variables.
- Define $\mathbf{V}' = \{V'_1, \ldots, V'_{2n}\}$, where each $V_i \in \mathbf{V}$ has both a positive form V'_{2n-1} and a negative form V'_{2n}
  - $V'_{2n-1} = V_i$ and $V'_{2n} = -V_n$
- A conjunction of octagonal constraints on $\mathbf{V}$ can be represented as a conjunction of potential constraints on $\mathbf{V}'$.
  - $2n \times 2n$ DBM with elements in $\bar{\mathbb{I}}$
  - $\forall i, V'_{2i-1} = -V'_{2i}$ holds for any DBM that encodes octagonal constraints

the constraint	is represented as
$V_i - V_j \leq c \quad (i \neq j)$	$V'_{2i-1} - V'_{2j-1} \leq c$ and $V'_{2j} - V'_{2i} \leq c$
$V_i + V_j \leq c \quad (i \neq j)$	$V'_{2i-1} - V'_{2j} \leq c$ and $V'_{2j-1} - V'_{2i} \leq c$
$-V_i - V_j \leq c \quad (i \neq j)$	$V'_{2i} - V'_{2j-1} \leq c$ and $V'_{2j} - V'_{2i-1} \leq c$
$V_i \leq c$	$V'_{2i-1} - V'_{2i} \leq 2c$
$V_i \geq c$	$V'_{2i} - V'_{2i-1} \leq -2c$

Concretization

Given a DBM $\mathbf{m}$ of dimension $2n$, the octagon described by $\mathbf{m}$ is defined as follows:

$$\gamma^{Oct} : \mathbf{DBM} \to \wp(\mathbf{V} \to \mathbb{I})$$

$$\gamma^{Oct}(\mathbf{m}) = \{(v_1, \ldots, v_n) \in \mathbb{I}^n \mid (v_1, -v_1, \ldots, v_n, -v_n) \in \gamma^{Pot}(\mathbf{m})\}$$

Coherence

- A DBM $\mathbf{m}$ must be coherent if it encodes a set of octagonal constraints:

$$\mathbf{m} \text{ is coherent } \iff \forall i, j.\ \mathbf{m}_{ij} = \mathbf{m}_{\bar{j}\bar{i}}$$

  where $\bar{\cdot}$ switches between the positive and negative forms of a variable:

$$\bar{i} = \begin{cases} i+1 & \text{if } i \text{ is odd} \\ i-1 & \text{if } i \text{ is even} \end{cases}$$

- Let **CDBM** be the set of all coherent DBMs.

Lattice Structure

The set of DBMs forms a complete lattice ($\mathbb{I} \neq \mathbb{Q}$):

$$(\mathbf{CDBM}, \sqsubseteq, \sqcup, \sqcap, \bot, \top)$$

- $\top$ is a DBM such that $\top_{ij} = +\infty$
- $\bot$ is a new smallest element
  - $\forall \mathbf{m}.\ \bot \sqsubseteq \mathbf{m}$
  - $\forall \mathbf{m}.\ \bot \sqcup \mathbf{m} = \mathbf{m} \sqcup \bot = \mathbf{m}$
  - $\forall \mathbf{m}.\ \bot \sqcap \mathbf{m} = \mathbf{m} \sqcap \bot = \mathbf{m}$
- $\forall \mathbf{m}, \mathbf{n}.\ \mathbf{m} \sqsubseteq \mathbf{n} \iff \forall i, j.\ \mathbf{m}_{ij} \leq \mathbf{n}_{ij}\ (\mathbf{m}, \mathbf{n} \neq \bot)$
- $\forall \mathbf{m}, \mathbf{n}.\ (\mathbf{m} \sqcup \mathbf{n})_{ij} = \max(\mathbf{m}_{ij}, \mathbf{n}_{ij})\ (\mathbf{m}, \mathbf{n} \neq \bot)$
- $\forall \mathbf{m}, \mathbf{n}.\ (\mathbf{m} \sqcap \mathbf{n})_{ij} = \min(\mathbf{m}_{ij}, \mathbf{n}_{ij})\ (\mathbf{m}, \mathbf{n} \neq \bot)$

Galois Connection

$$\wp(\mathbf{V} \to \mathbb{I}) \xleftrightarrow[\alpha^{Oct}]{\gamma^{Oct}} \textbf{CDBM}$$

$$\alpha^{Oct}(\emptyset) = \bot$$

$$\left(\alpha^{Oct}(R)\right)_{ij} = \begin{cases} \max\{\rho(V_l) - \rho(V_k) \mid \rho \in R\} & i = 2k-1, j = 2l-1 \\ & \text{or } i = 2l, j = 2k \\ \max\{\rho(V_l) + \rho(V_k) \mid \rho \in R\} & i = 2k, j = 2l-1 \\ \max\{-\rho(V_l) - \rho(V_k) \mid \rho \in R\} & i = 2k-1, j = 2l \end{cases}$$

Normalization

Different, incomparable DBMs may represent the same potential set:

(a)

		j: 1	2	3
i	1	$+\infty$	4	3
	2	-1	$+\infty$	$+\infty$
	3	-1	1	$+\infty$

(b)

		j: 1	2	3
i	1	0	5	3
	2	-1	$+\infty$	$+\infty$
	3	-1	1	$+\infty$

(c)

		j: 1	2	3
i	1	0	4	3
	2	-1	0	$+\infty$
	3	-1	1	0

Shortest-Path Closure

The shortest-path closure $\mathbf{m}^*$ of $\mathbf{m}$ is defined as follows:

$$\begin{cases} \mathbf{m}^*_{ii} \stackrel{\text{def}}{=} 0 \\ \mathbf{m}^*_{ij} \stackrel{\text{def}}{=} \min\limits_{\substack{\text{all path from } i \text{ to } j \\ \langle i = i_1, i_2, \ldots, i_m = j \rangle}} \sum_{k=1}^{m-1} \mathbf{m}_{i_k i_{k+1}} & \text{if } i \neq j \end{cases}$$

The closure $\mathbf{m}^*$ of $\mathbf{m}$ is the smallest DBM representing $\gamma^{Pot}(\mathbf{m})$:

$$\forall X \in \textbf{DBM}.\ \gamma^{Pot}(\mathbf{m}) = \gamma^{Pot}(X) \implies \mathbf{m}^* \sqsubseteq X$$

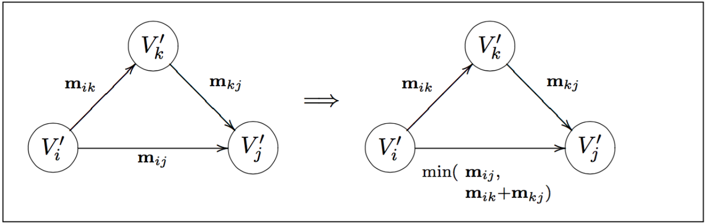
Floyd-Warshall Algorithm:

$$\begin{cases} \mathbf{m}^0 \stackrel{\text{def}}{=} \mathbf{m} \\ \mathbf{m}^k_{ij} \stackrel{\text{def}}{=} \min(\mathbf{m}^{k-1}_{ij}, \mathbf{m}^{k-1}_{ik} + \mathbf{m}^{k-1}_{kj}) & \text{if } 1 \leq i, j, k \leq n \\ \mathbf{m}^*_{ij} \stackrel{\text{def}}{=} \begin{cases} \mathbf{m}^n_{ij} & \text{if } i \neq j \\ 0 & \text{if } i = j \end{cases} \end{cases}$$

Implicit Constraints

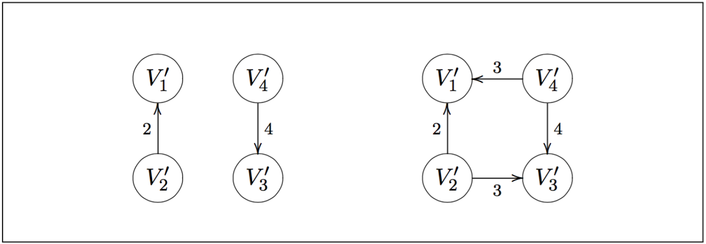
Closing a DBM makes implicit constraints explicit:

$$V_j - V_k \le c \wedge V_k - V_l \le d \implies V_j - V_i \le c + d$$



Strong Closure

The closure $\mathbf{m}^*$ of $\mathbf{m}$ may not be the smallest DBM representing $\gamma^{Oct}(\mathbf{m})$:



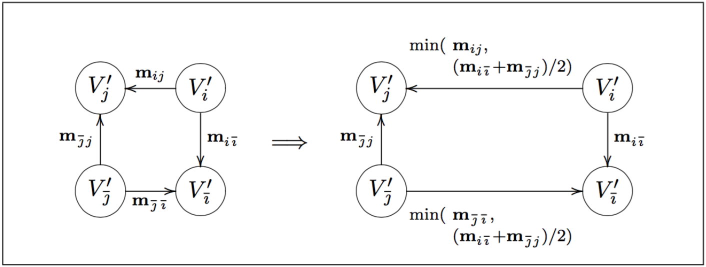
Strong Closure

The strong closure $\mathbf{m}^\bullet$ of $\mathbf{m}$ is the smallest DBM representing γ^{Oct}. $\mathbf{m}$ is strongly closed iff:

$$\begin{array}{ll} \forall i,j,k. & \mathbf{m}_{ij} \leq \mathbf{m}_{ik} + \mathbf{m}_{kj} \\ \forall i,j. & \mathbf{m}_{ij} \leq (\mathbf{m}_{i\bar{\imath}} + \mathbf{m}_{\bar{\jmath}j})/2 \\ \forall i, & \mathbf{m}_{ii} = 0 \end{array}$$

The encoding of octagonal constraints implies

$$V'_j - V'_{\bar{\jmath}} \leq c \wedge V'_{\bar{\imath}} - V'_i \leq d \implies V'_j - V'_i \leq (c+d)/2$$



Abstract Transfer Functions

Abstract counterparts of concrete transfer functions:

- union, intersection
- assignment
- test (guard)

Soundness condition:

$$\boldsymbol{F} \circ \boldsymbol{\gamma} \sqsubseteq \boldsymbol{\gamma} \circ \hat{\boldsymbol{F}}$$

Union

The union of two octagons may not be an octagon. $\sqcup$ gives a sound approximation.

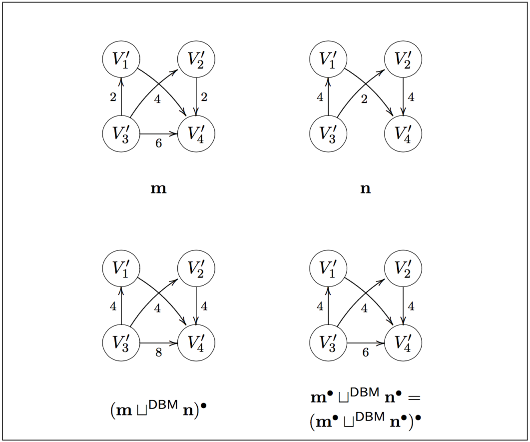


Figure 11. Abstract union of octagons, based on $\sqcup^{\mathsf{DBM}}$. DBMs should be strongly closed for best precision. This also ensures that the result is strongly closed.

Intersection

The intersection of two octagons is always an octagon. $\sqcap$ gives the exact intersection of two octagons.

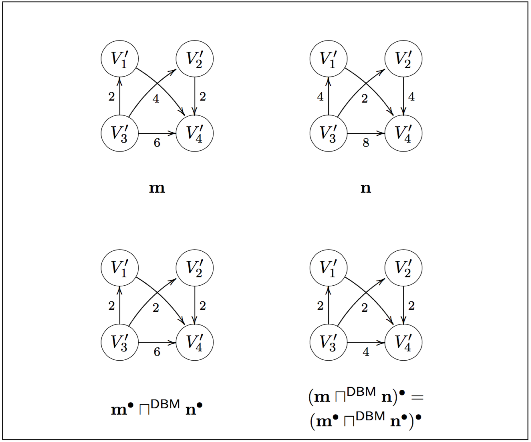


Figure 12. Exact intersection of octagons, based on $\sqcap^{\mathsf{DBM}}$. The arguments do not need to be strongly closed, and the result is seldom strongly closed.

Assignment (the forget operator)

Concrete semantics:

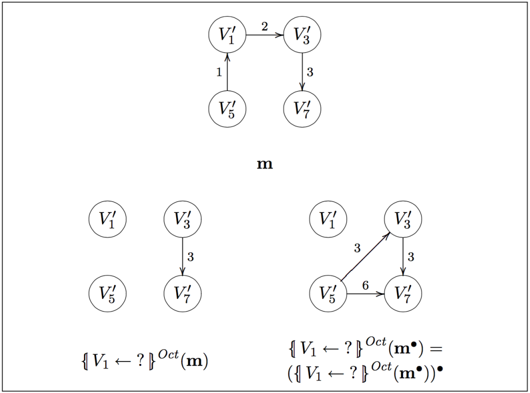
$$\begin{array}{rcl} \{\!|\, V_f \leftarrow ?\, |\!\}(R) & \stackrel{\text{def}}{=} & \{\, \rho[V_f \mapsto v] \mid \rho \in R,\ v \in \mathbb{I} \,\} \\ & = & \{\, \rho \mid \exists v \in \mathbb{I},\ \rho[V_f \mapsto v] \in R \,\} \end{array}$$

Abstract semantics:

$$(\{\!|\, V_f \leftarrow ?\, |\!\}^{Oct}(\mathbf{m}))_{ij} \stackrel{\text{def}}{=} \begin{cases} \mathbf{m}_{ij} & \text{if } i \neq 2f-1, 2f \text{ and } j \neq 2f-1, 2f \\ 0 & \text{if } i = j = 2f-1 \text{ or } i = j = 2f \\ +\infty & \text{otherwise} \end{cases}$$

Assignment (the forget operator)

When the argument is strongly closed, the result is exact:



Assignments

- $(\{\!\!|\, V_{j_0} \leftarrow [a,b] \,|\!\!\}^{Oct}_{exact}(\mathbf{m}))_{ij} \stackrel{\text{def}}{=}$

$$\begin{cases} -2a & \text{if } i = 2j_0 - 1,\ j = 2j_0 \\ 2b & \text{if } i = 2j_0,\ j = 2j_0 - 1 \\ (\{\!\!|\, V_{j_0} \leftarrow ?\,|\!\!\}^{Oct}(\mathbf{m}^\bullet))_{ij} & \text{otherwise} \end{cases}$$

- $(\{\!\!|\, V_{j_0} \leftarrow V_{j_0} + [a,b] \,|\!\!\}^{Oct}_{exact}(\mathbf{m}))_{ij} \stackrel{\text{def}}{=}$

$$\begin{cases} \mathbf{m}_{ij} - a & \text{if } i = 2j_0 - 1,\ j \neq 2j_0 - 1, 2j_0 \\ & \text{or } j = 2j_0,\ i \neq 2j_0 - 1, 2j_0 \\ \mathbf{m}_{ij} + b & \text{if } i \neq 2j_0 - 1, 2j_0,\ j = 2j_0 - 1 \\ & \text{or } j \neq 2j_0 - 1, 2j_0,\ i = 2j_0 \\ \mathbf{m}_{ij} - 2a & \text{if } i = 2j_0 - 1,\ j = 2j_0 \\ \mathbf{m}_{ij} + 2b & \text{if } i = 2j_0,\ j = 2j_0 - 1 \\ \mathbf{m}_{ij} & \text{otherwise} \end{cases}$$

- $(\{\!\!|\, V_{j_0} \leftarrow V_{i_0} + [a,b] \,|\!\!\}^{Oct}_{exact}(\mathbf{m}))_{ij} \stackrel{\text{def}}{=}$

$$\begin{cases} -a & \text{if } i = 2j_0 - 1,\ j = 2i_0 - 1 \\ & \text{or } i = 2i_0,\ j = 2j_0 \\ b & \text{if } i = 2i_0 - 1,\ j = 2j_0 - 1 \\ & \text{or } i = 2j_0,\ j = 2i_0 \\ (\{\!\!|\, V_{j_0} \leftarrow ?\,|\!\!\}^{Oct}(\mathbf{m}^\bullet))_{ij} & \text{otherwise} \end{cases}$$

- $(\{\!\!|\, V_{j_0} \leftarrow -V_{j_0} \,|\!\!\}^{Oct}_{exact}(\mathbf{m}))_{ij} \stackrel{\text{def}}{=}$

$$\begin{cases} \mathbf{m}_{\bar{\imath} j} & \text{if } i \in \{2j_0 - 1, 2j_0\} \text{ and } j \notin \{2j_0 - 1, 2j_0\} \\ \mathbf{m}_{i \bar{\jmath}} & \text{if } i \notin \{2j_0 - 1, 2j_0\} \text{ and } j \in \{2j_0 - 1, 2j_0\} \\ \mathbf{m}_{\bar{\imath} \bar{\jmath}} & \text{if } i \in \{2j_0 - 1, 2j_0\} \text{ and } j \in \{2j_0 - 1, 2j_0\} \\ \mathbf{m}_{ij} & \text{if } i \notin \{2j_0 - 1, 2j_0\} \text{ and } j \notin \{2j_0 - 1, 2j_0\} \end{cases}$$

- $\{\!\!|\, V_{j_0} \leftarrow -V_{i_0} \,|\!\!\}^{Oct}_{exact} \stackrel{\text{def}}{=} \{\!\!|\, V_{j_0} \leftarrow -V_{j_0} \,|\!\!\}^{Oct}_{exact} \circ \{\!\!|\, V_{j_0} \leftarrow V_{i_0} \,|\!\!\}^{Oct}_{exact}$
- $\{\!\!|\, V_{j_0} \leftarrow -V_{j_0} + [a,b] \,|\!\!\}^{Oct}_{exact} \stackrel{\text{def}}{=}$

$$\{\!\!|\, V_{j_0} \leftarrow V_{j_0} + [a,b] \,|\!\!\}^{Oct}_{exact} \circ \{\!\!|\, V_{j_0} \leftarrow -V_{j_0} \,|\!\!\}^{Oct}_{exact}$$

- $\{\!\!|\, V_{j_0} \leftarrow -V_{i_0} + [a,b] \,|\!\!\}^{Oct}_{exact} \stackrel{\text{def}}{=}$

$$\{\!\!|\, V_{j_0} \leftarrow V_{j_0} + [a,b] \,|\!\!\}^{Oct}_{exact} \circ \{\!\!|\, V_{j_0} \leftarrow -V_{i_0} \,|\!\!\}^{Oct}_{exact}$$

Variable Clustering

- A collection $\mathbf{\Pi} : \wp(\wp(\mathbf{V}))$ of variable clusters such that

$$\bigcup_{\pi \in \mathbf{\Pi}} \pi = \mathbf{V}$$

- The complete lattice:

$$\mathbb{O}_{\mathbf{\Pi}} = \prod_{\pi \in \mathbf{\Pi}} \mathbb{O}_{\pi}$$

  where $\mathbb{O}_{\pi}$ is the lattice of Octagon for variables in π.
- Challenge: How to choose a good $\mathbf{\Pi}$?
  - Heo et al. "Learning a Variable-Clustering Strategy for Octagon from Labeled Data Generated by a Static Analysis". SAS 2016