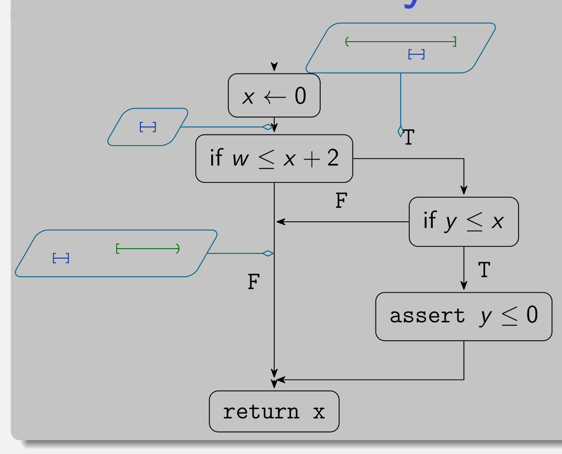
Zonotope Guided Predicate Generation

Kenny Ballou, California State University San Marcos

Research Question

Can we use weakly-relational or Zonotope domains to improve predicate generation for program analysis over simple value analysis?

Data-Flow Analysis



Data-flow Analysis (DFA) computes information about programs in the form of abstract states. These states can be used to represent numerical values using infinite domains such as Intervals, Zones, or Octagons.

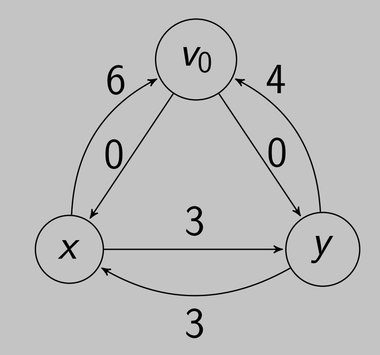
Comparison of invariants between Octagons and Predicates

$O \equiv P$	$O \prec P$	$O \succ P$	$O \prec\succ P$
3768	2514	280	368

Comparing Octagons to Predicates*, we can see that a random Predicates domain is more precise than Octagons second only to equivalent precision.

Comparison numbers from work under review Ballou et al., 2025, "Minimal Comparison of Octagon Abstract Domain".

Zone Domain Graph Representation

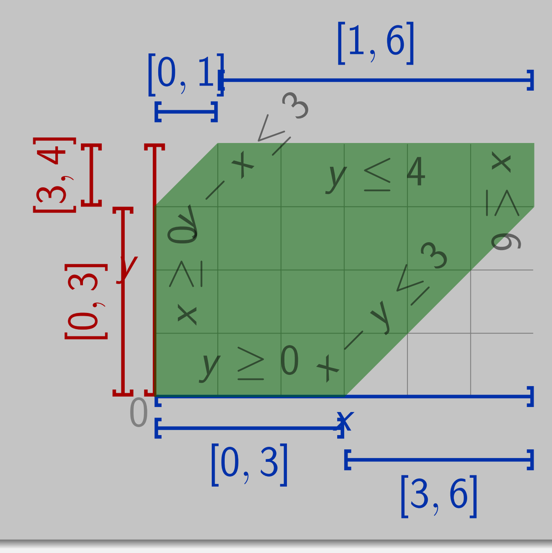


The Zone domain as introduced by Miné† represents the strict unit difference between program variables. The graph to the left represents the following system:

$$
\begin{aligned}
x &\leq 6 \\
x &\geq 0 \\
y &\leq 4 \\
y &\leq 0 \\
x - y &\leq 3 \\
y - x &\leq 3
\end{aligned}
$$

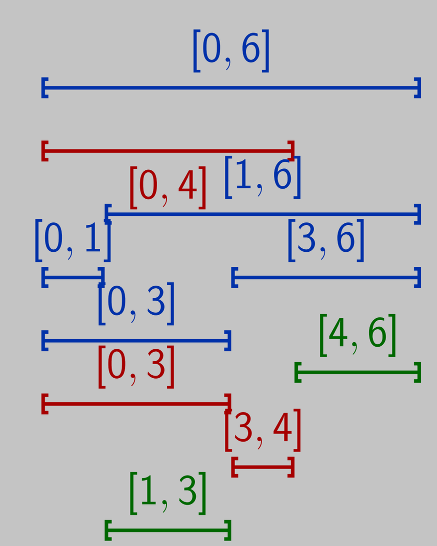
† Miné, 2001, "A New Numerical Abstract Domain Based on Difference-Bound Matrices"

Zones Projection



We can project the region described by relationships of two variables into a plane. Notice, the various *implied* intervals at the intersection points.

Intervals to Predicates



- After parsing and querying weakly-relational domain analysis, we must construct a set of intervals to use for Predicate Generation.
- We require that the resulting set of predicates are *disjoint* and *complete*.‡
- Additionally, we are constrained by the number of predicates. Too many, and the analysis is too slow. Too few, and the analysis is imprecise.

‡ Sherman et al., 2015, "Exploiting Domain and Program Structure to Synthesize Efficient and Precise Data Flow Analyses (T)"

Candidiate Predicate Set

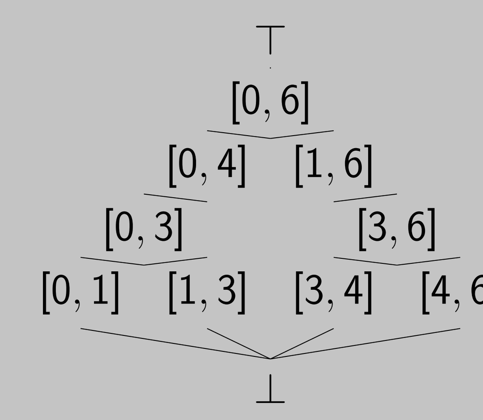
$$\{(-\infty, 0), [0, 1], (1, 3], (3, 4], (4, 6], (6, +\infty)\}$$

Notice, the individual predicates are disjoint and the set covers the entire range of values over integers.

Closed-Normal Form

$$\{(-\infty, -1], [0, 1], [2, 3], [4], [5, 6], [7, +\infty)\}$$

Incomplete Lattice Representation under $\subset$



- One idea for working through the generation of predicates that satisfy the disjoint and complete properties is to leverage a lattice data structure over the incomplete intervals.
- Through this data structure, we should be able to easily interleave the different intervals into a set of predicates of a specified or requested size.

Tasks to Resolve

- Implement a parser of zonotope invariants output
- Implement predicate generator over the incomplete lattice
- Empirical Evaluation

References

[1] Kenny Ballou and Elena Sherman. "Minimal Comparison of Octagon Abstract Domain". In: *Integragted Formal Methods 2025* (Nov. 2025). Under Review.

[2] Antoine Miné. "A New Numerical Abstract Domain Based on Difference-Bound Matrices". In: *Lecture Notes in Computer Science* (2001), pp. 155–172. ISSN: 0302-9743. DOI: 10.1007/3-540-44978-7_10. URL: http://dx.doi.org/10.1007/3-540-44978-7_10.

[3] Elena Sherman and Matthew B. Dwyer. "Exploiting Domain and Program Structure to Synthesize Efficient and Precise Data Flow Analyses (T)". In: *2015 30th IEEE/ACM International Conference on Automated Software Engineering (ASE)* (Nov. 2015). DOI: 10.1109/ase.2015.41. URL: http://dx.doi.org/10.1109/ASE.2015.41.