Introduction to Abstract Interpretation

ECE 584

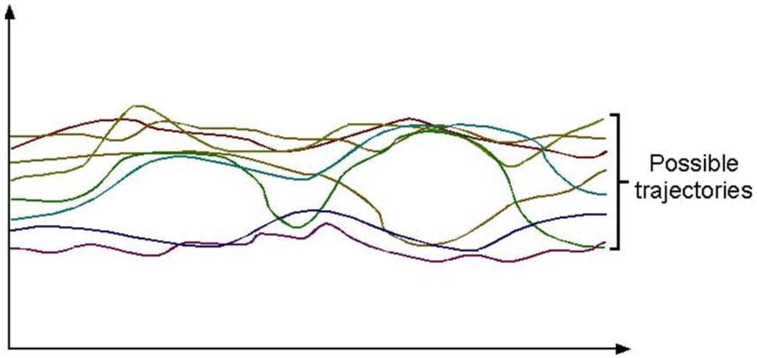
Sayan Mitra

Lecture 18

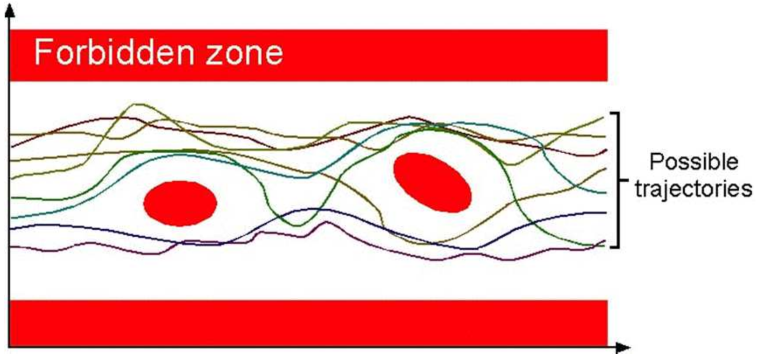
References

- Patrick Cousot, Radhia Cousot: Abstract Interpretation: A Unified Lattice Model for Static Analysis of Programs by Construction or Approximation of Fixpoints. POPL 1977: 238-252
- **16.399: Abstract Interpretation at MIT** http://web.mit.edu/afs/athena.mit.edu/course/16/16.399/www/
- Notes on Abstract Interpretation by Alexandru Salcianu, Nov 2001
- ASTREE tools: http://www.astree.ens.fr/

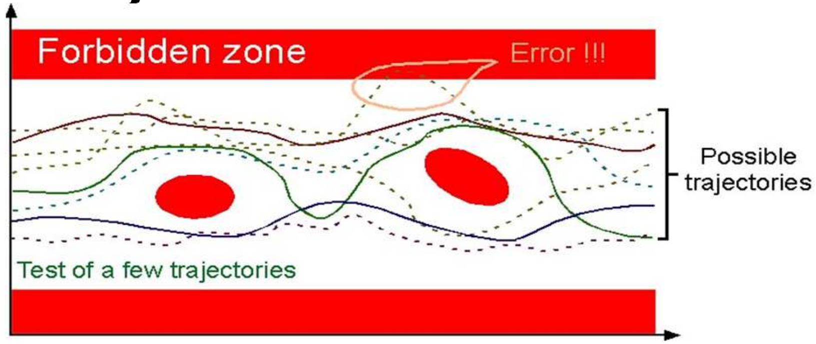
Abstract interpretation in a nut shell (Cousot)



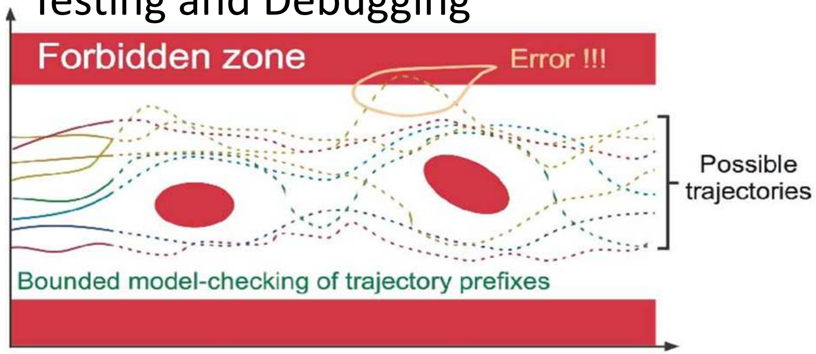
Concrete Executions/Semantics, usually all answering nontrivial questions is undecidable



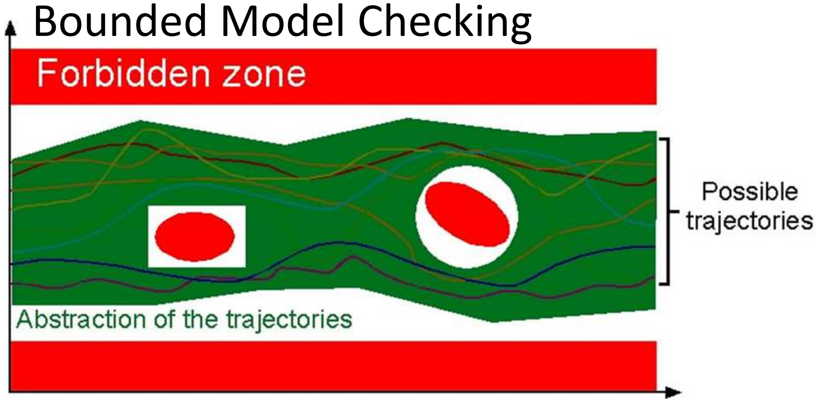
Safety Verification



Testing and Debugging



Bounded Model Checking



Abstract interpretation

Outline

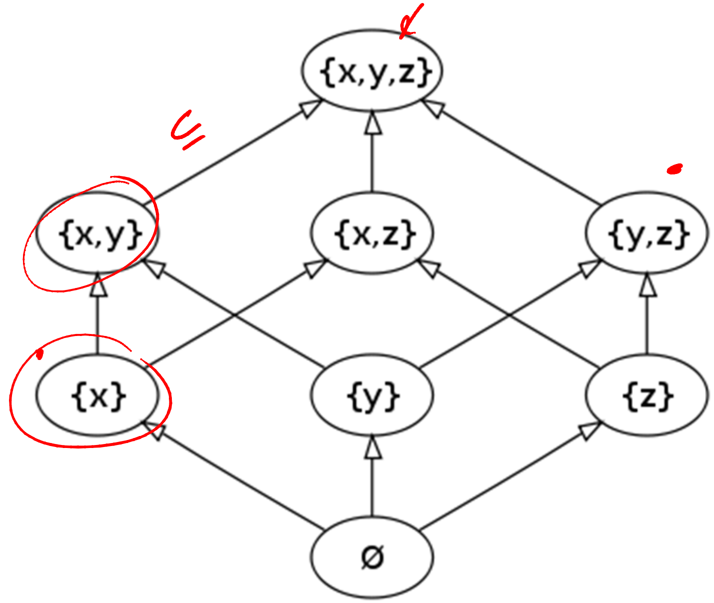
- Background
  - Lattices and Galois connections
- Property lattice
- Basic and refined analysis
- Fixed point computations
  - Widening

Abstract Interpretation in more detail

- Abstract interpretation is a technique for approximating a basic analysis with a refined analysis that sacrifices precision for speed
- Abstract interpretation (AI) relates basic analysis and refined analysis using a Galois Connection between property lattices
- The basic analysis may be too hard to compute, but hopefully successive refinements give an analysis that is computable
- Widening/narrowing for terminating fixed point computations

A thing or two about lattices

- Concrete System $\langle Q, Q_0, D \rangle, D \subseteq Q \times Q$
- Execution $q_0, q_1, q_2, \ldots$
- A property space L is a **complete lattice** $\langle L, \sqsubseteq \rangle$
  - A poset $\langle L, \sqsubseteq \rangle$ is a complete lattice if every subset $A \subseteq L$ has both a greatest lower bound (inf, glb, meet, $\bigwedge A, \sqcap A$) and a least upper bound (sup, lub, join, $\bigvee A, \sqcup A$)
- A lattice has **ascending chain property** if there is no infinite ascending chain



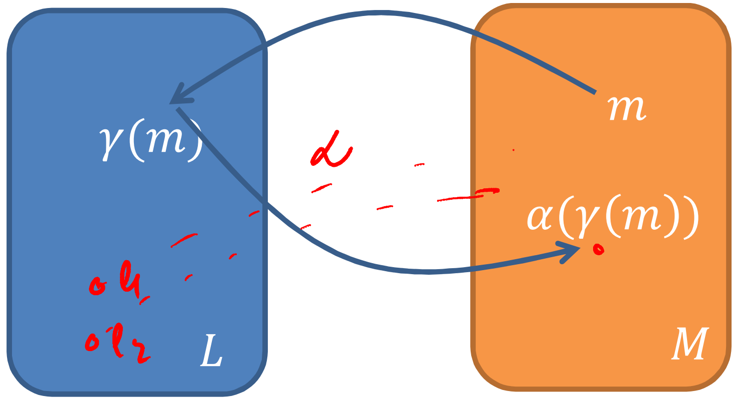
http://code.google.com/p/python-lattice/

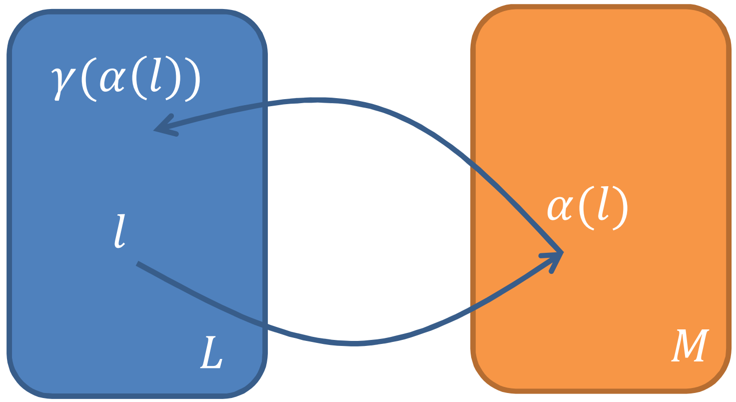
$L \triangleq Val(\{x,y,z\})$ $M \triangleq Val(\{x,y\})$

Galois Connections

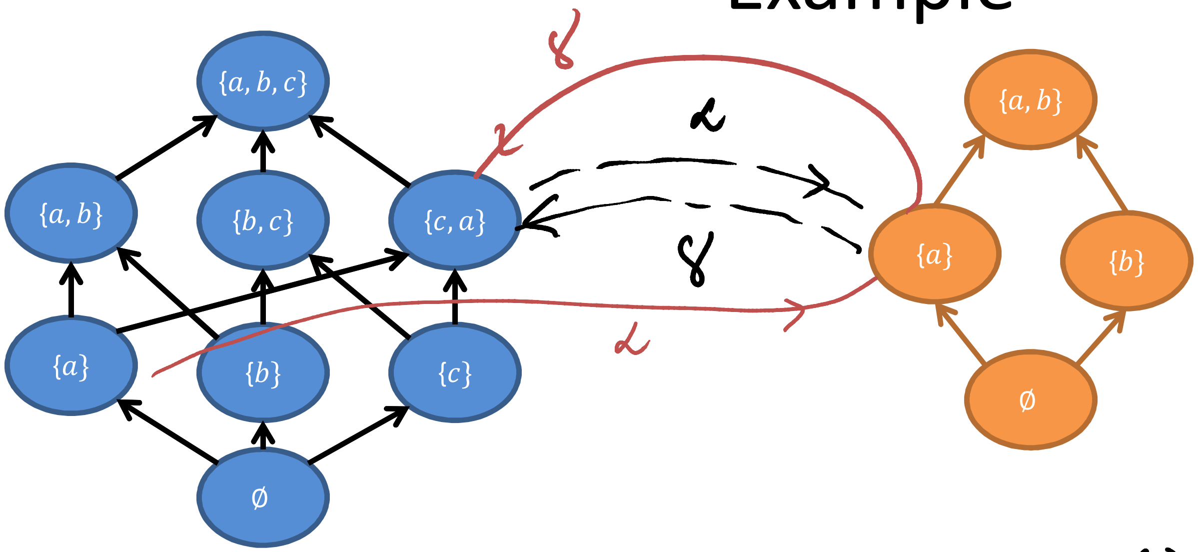
- $\langle L, \alpha, \gamma, M \rangle$ is a Galois conneciton between the lattice $\langle L, \sqsubseteq_L \rangle$ and $\langle M, \sqsubseteq_M \rangle$ iff
  - $\alpha: L \to M, \gamma: \cancel{L} \to M$ $M \to L$
    Abstraction and concretization
  - α, γ are monotonic
  - $\alpha \circ \gamma \sqsubseteq_M \lambda\, m.\, m$
    Concretization does not lose precision
  - $\gamma \circ \alpha \sqsupseteq_L \lambda\, l.\, l$
    Abstraction may lose precision but remains correct

$\forall l \;\; \gamma \circ \alpha(l) \sqsupseteq l$





Example



$\alpha(\{a\}) = \{a\}$ $\alpha(\{a,b\}) = \{a,b\}$ $\alpha(\{a,c\}) = \{a\}$ $\alpha(\{a,b,c\}) = \{a,b\}$

$\gamma(\{a\}) = \{a,c\}$ $\gamma(\{b\}) = \{b,c\}$ $\gamma(\{a,b\}) = \{a,b,c\}$ $\gamma(\emptyset) = \{c\}$

$v, q \in Q$

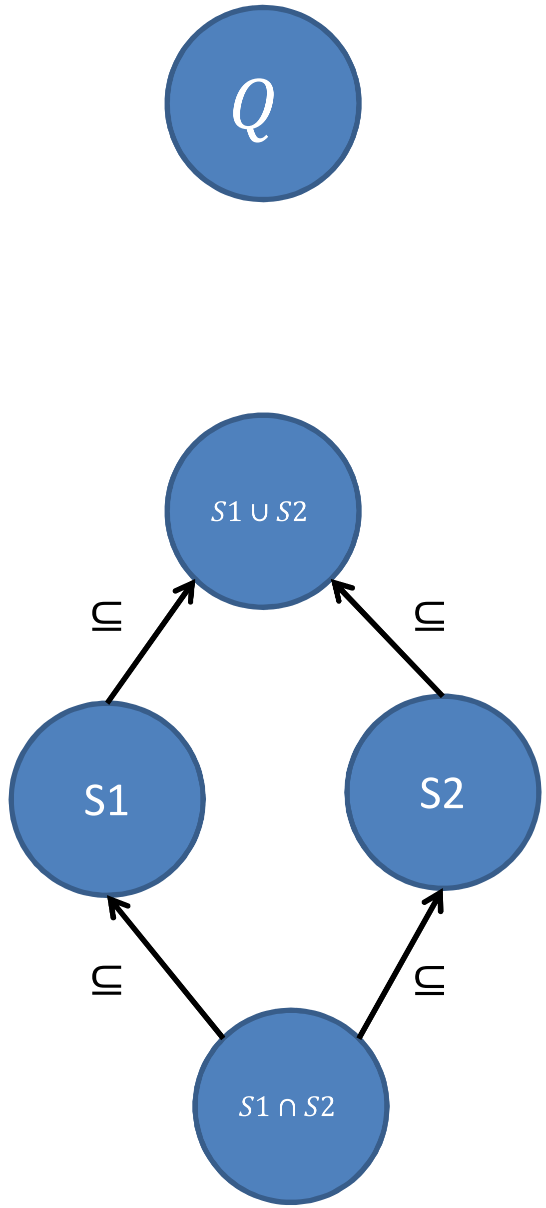
$\langle Q, Q_0, \rightarrow \rangle$

Property Lattice

- For $l_1, l_2 \in L\ with\ l_1 \sqsubseteq l_2$ think of l_1 being **no weaker than** l_2
- A relation $R \subseteq Q \times L$ is a correctness relation iff
  - $\forall v, l_1, l_2, (v\ R\ l_1) \wedge (l_1 \sqsubseteq l_2) \rightarrow (v \overset{R}{\sqsubseteq} l_2)$
  - l_1 approximates *v* and there is an "upper" approximation l_2 means l_2 approximates v
  - $\forall v, \forall L' \subseteq L, \big(\forall l \in L', (v\ R\ l)\big) \rightarrow v\ R\ (\wedge L')$
  - If *v* is approximated by several properties in L' we take the smallest (most precise) of them
- In property lattice smaller means more precise, the top element T is the least precise and approximates all the other properties
- **Lemma.** $R \subseteq Q \times L$ is correctness relation, then
  - $(v\ R\ \top)$
  - $(v\ R\ l_1) \wedge (v\ R\ l_2) \rightarrow v\ R\ (l_1 \sqcap l_2)$
- v satisfies both l_1 AND l_2 then we obtain more precise $l_1 \sqcap l_2$
- v satisfies l_1 OR l_2 then we obtain weaker $l_1 \sqcup l_2$

Example 1

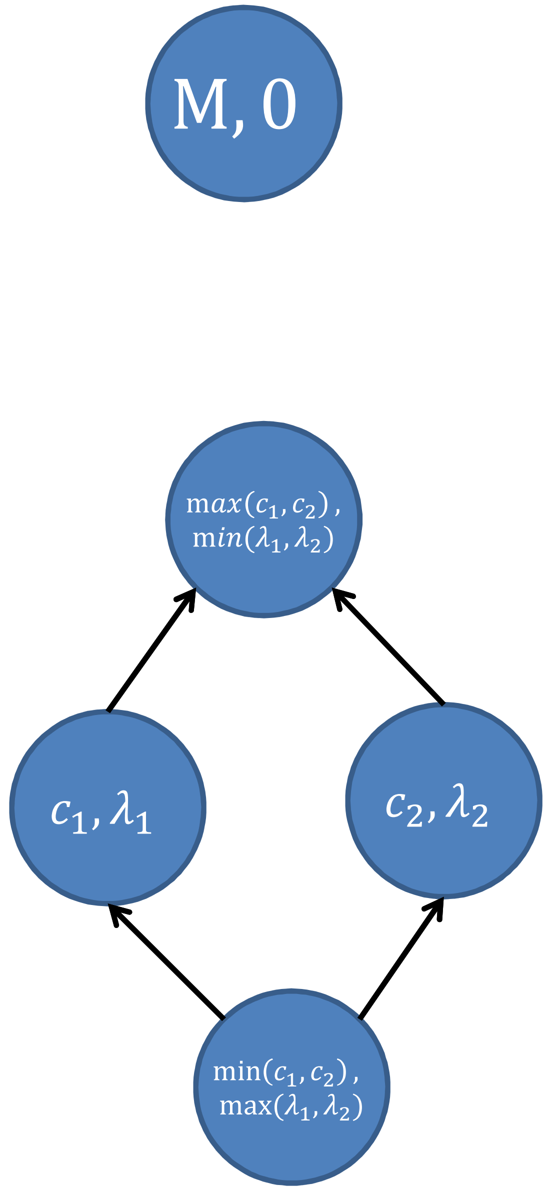
- Invariant properties of A
  - $\text{Reach}_A \subseteq S2 \cap S1$
  - S1 is an invariant ($\text{Reach}_A \subseteq S1$)
  - S2 is an invariant ($\text{Reach}_A \subseteq S2$)
  - $\text{Reach}_A \subseteq S2 \cup S1$
- If Q is finite then the lattice has ascending chain property



Example 2

- Global exponential Stability
  - $|x(t)| \leq \min(c_1, c_2)\, e^{-\max(\lambda_1,\lambda_2)}|x(0)|$
  - $|x(t)| \leq c_1 e^{-\lambda_1 t}|x(0)|$ $\leftarrow$
  - $|x(t)| \leq c_2 e^{-\lambda_2 t}|x(0)|$
  - $|x(t)| \leq \max(c_1, c_2)\, e^{-min(\lambda_1,\lambda_2)}|x(0)|$

- If c and λ come from set of bounded integers [M] then this property lattice has ascending chain property



Basic Analysis

- Given an execution $q_0, q_1, q_2, \ldots q_k$ of the concrete system and a property lattice $\langle L, \sqsubseteq \rangle$ and a property relation R, how to compute the property satisfied by q_k?
- Suppose we have an abstract initial value $l_0 \in L$ such that $\overset{q}{\cancel{v}}_0\ R\ l_0$
- Consider each concrete execution path p from q_0 that reachs q_k?
  - Follow the corresponding abstract path from l_0 to arrive at the abstract state $l_{k,p}$
  - Join the $l'_{k,p}$~~s~~
- **Correctness conditions**
  - $v_0\ R\ l_0$
  - $\forall v_1, v_2, l_1 l_2, (v_1 \to v_2) \wedge (v_1 R\ \ l_1) \wedge (f_L(l_1) = l_2) \to v_2 R\ l_2$
- Usually this join over paths cannot be computed, therefore, approximate

$$
\begin{tikzcd}
l_1 \arrow[r, "f_L"] \arrow[d, leftrightarrow, "R"] & l_2 \arrow[d, leftrightarrow, "R"] \\
q_1 \arrow[r] & q_2
\end{tikzcd}
$$

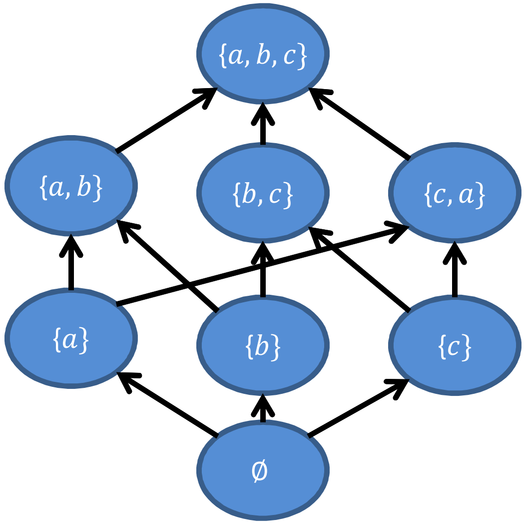
Refining the Analysis

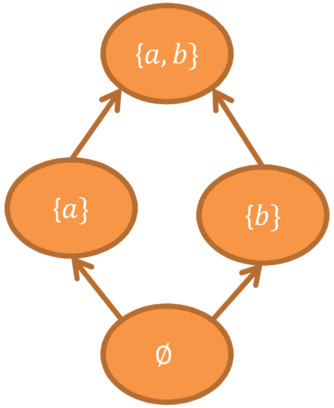
- For invariance, we want to compute a fixpoint of the f_L
- Usually the fixed point computation is too difficult or impossible
  - Analysis converges too slowly
  - L does not have ascending chain property
- **Use a smaller, more approximate property lattice M** such that there is a **Galois connection** $\langle L, \alpha, \gamma, M \rangle$

More facts about Galois Connections

- If $\langle L, \alpha, \gamma, M \rangle$ is a Galois Connection then α uniquely determines γ
  - $\gamma(m) = \sqcup \{ l \mid \alpha(l) \sqsubseteq m \}$
- γ uniquely determines α
  - $\alpha(l) = \sqcap \{ m \mid \gamma(m) \sqsubseteq l \}$
- If α is completely additive $\forall L' \subseteq L, \alpha(\sqcup L') = \sqcup \{\alpha(l) | l \in L'\}$ then there is a γ such that $\langle L, \alpha, \gamma, M \rangle$ is a GC
- If γ is completely multiplicative $\forall M' \subseteq M \; \gamma(\sqcap M') = \sqcap \{\gamma(m) | m \in M'\}$ then there is a α such that $\langle L, \alpha, \gamma, M \rangle$ is a GC
- If GC then α is completely additive and γ is completely multiplicative

Example cont.





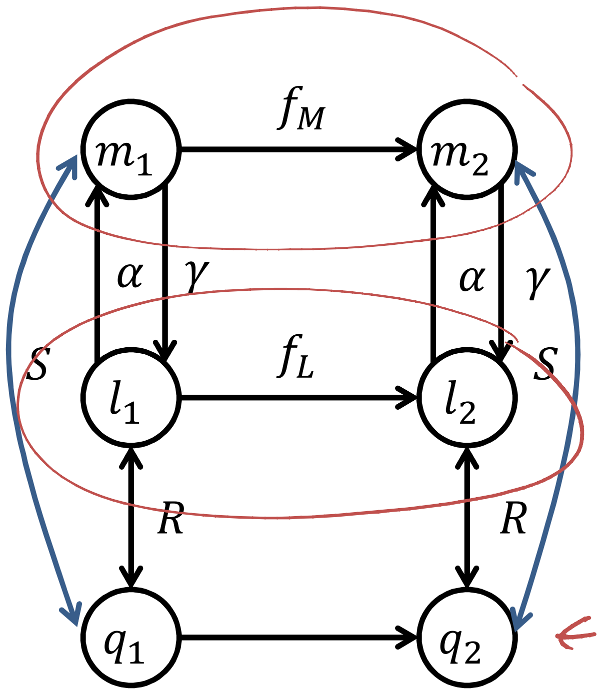
- Abstraction defined as:
- $\alpha(\{a\}) = \alpha(\{a,c\}) = \{a\}$
- $\alpha(\{b\}) = \alpha(\{b,c\}) = \{b\}$
- $\alpha(\{a,b\}) = \alpha(\{a,b,c\}) = \{a,b\}$
- $\alpha(\emptyset) = \emptyset$

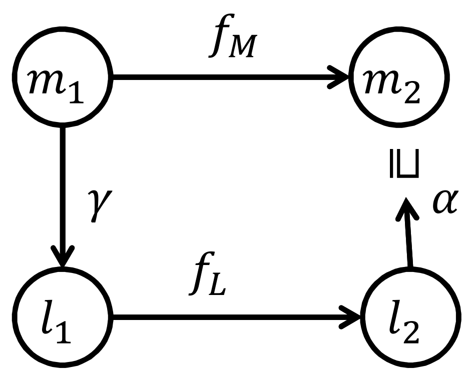
LUb

- Then, $\gamma(\{a\}) = \sqcup \{l \mid \alpha(\{a\}) \sqsubseteq \{a\}\} = \{a\} \sqcup \{a,c\} = \{a,c\}$
- $\gamma(\{b\}) = \sqcup \{l \mid \alpha(\{a\}) \sqsubseteq \{a\}\} = \{b\} \sqcup \{b,c\} = \{b,c\}$
- $\gamma(\{a,b\}) = \{a,b,c\} \sqcup \{a,c\} \sqcup \{a,b\} \ldots = \{a,b,c\}$

Refined Analysis

- New correctness relation $S \subseteq V \times M$ with $v\; S\; m$ iff $v\; R\; \gamma(m)$
- New transition function $f_M : M \to M$ such that $f_M \sqsupseteq \alpha \circ f_L \circ \gamma$
- **Theorem.** S is indeed a correctness relation.
- **Theorem.** S is preserved by any such f_M





S^*

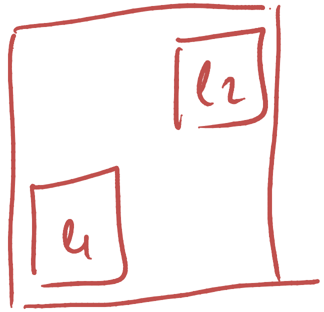
Fixed Point Computation

- To compute the result of an analysis we have to compute the fixpoint of $f_M: M \to M, where\ M$ is a complete lattice
- This is usually done as $f^n{}_M(\bot)$ (ascending sequence)
  - Alternatively from ⊤ as a descending sequence

ascending Kleene sequence

- If AKS stabilizes then it stabilizes at the lfp
  - Always stabilizes if f is continuous $f_M(\sqcup L') = \sqcup \{f(l) | l \in L'\}$
- If AKS stabilizes too slowly then use widening

Widening



- $S_0, S_1 = T(S_0), S_2 = T(S_1) = T^2(S_0), \ldots Tn$
- $S_0, S_1, S_2, \ldots$
- $\nabla : L \times L \rightarrow L$ is a widening operator iff
  - $\forall l_1, l_2, l_1 \sqsubseteq l_1 \nabla l_2$ and $l_2 \sqsubseteq l_1 \nabla l_2$ (coarsens)
  - For all ascending chains $\{l_n\}$ in L the $\{l_n{}^{\nabla}\}$ chain eventually stabilizes
    - $l_0{}^{\nabla} = l_0$
    - $l_n{}^{\nabla} = l_{n-1}{}^{\nabla} \nabla l_n,\ for\ n > 0$
- For a monotone function $f : L \rightarrow L$ on a complete lattice and ∇ define the sequence $f_n{}^{\nabla}$
    - $\perp$ $\quad$ $if\ n = 0$
    - $f_{n-1}{}^{\nabla}$ $\quad$ if $n > 0$ and $f(f_{n-1}{}^{\nabla}) \sqsubseteq f_{n-1}{}^{\nabla}$
    - $f_{n-1}{}^{\nabla} \nabla f(f_{n-1}{}^{\nabla})$ otherwise
- Theorem. $f_n{}^{\nabla}$ stabilizes at $f_m{}^{\nabla} \sqsupseteq$ lfp(f)
  - Safe over-approximation of lfp(f)

Conclusions

$\pm x \pm y \leq c$

- Abstract interpretation is a widely used framework for static analysis of programs and now hybrid systems
  - Invariants
  - Termination
  - Type checking
- Example abstract domains
  - Intervals, Octagons (+/-*x*+/-*y* ≤ *c*), *polyhedra*
- Galois connections can be combined in series or in parallel to get new analyses
- Look at the references to learn more