Abstract Interpretation, Data-Flow Analysis and Lattices in a Nutshell

Christoph Mallon

27. Dezember 2013

Terms

- Abstract interpretation: Provable system to determine sound overapproximation of behavior of programs
- Data-Flow Analysis: Mechanism to calculate said overapproximation
- Lattice: Underlying structure for performing calculations

Definition of Lattice

- *Lattice* $\mathcal{L} = (S, \sqsubseteq)$ of *set* S and *partial order* $\sqsubseteq$
- Set of *upper bounds* of a, b: $U = \{s | a, b \sqsubseteq s\} \neq \emptyset$
- There is a *unique least upper bound*: $\exists u_l \in U \, \forall u \in U : u_l \sqsubseteq u$
  - *Join* $a \sqcup b := u_l$
  - Not necessarily all upper bounds comparable to each other
  - But all upper bounds comparable to least upper bound
- Analogous definition for $\sqsupseteq$ and $\sqcap$ (*Meet*) with *greatest lower bound*
- *Bottom* and *Top* elements *exist*: $\forall s \in S : \bot \sqsubseteq s \sqsubseteq \top$
  - Follows from least upper bound/greatest lower bound
  - $\bot \sqcup x = x$, $\bot \sqcap x = \bot$
  - $\top \sqcap x = x$, $\top \sqcup x = \top$
- Often $S \subset 2^T$ of set T, $\sqsubseteq \in \{\subset, \supset\}$
  - *Power-set* lattice: $S = 2^T$
  - *Flat* lattice: $S = T_\bot^\top = \{\emptyset, T\} \cup \bigcup_{t \in T} \{\{t\}\}$

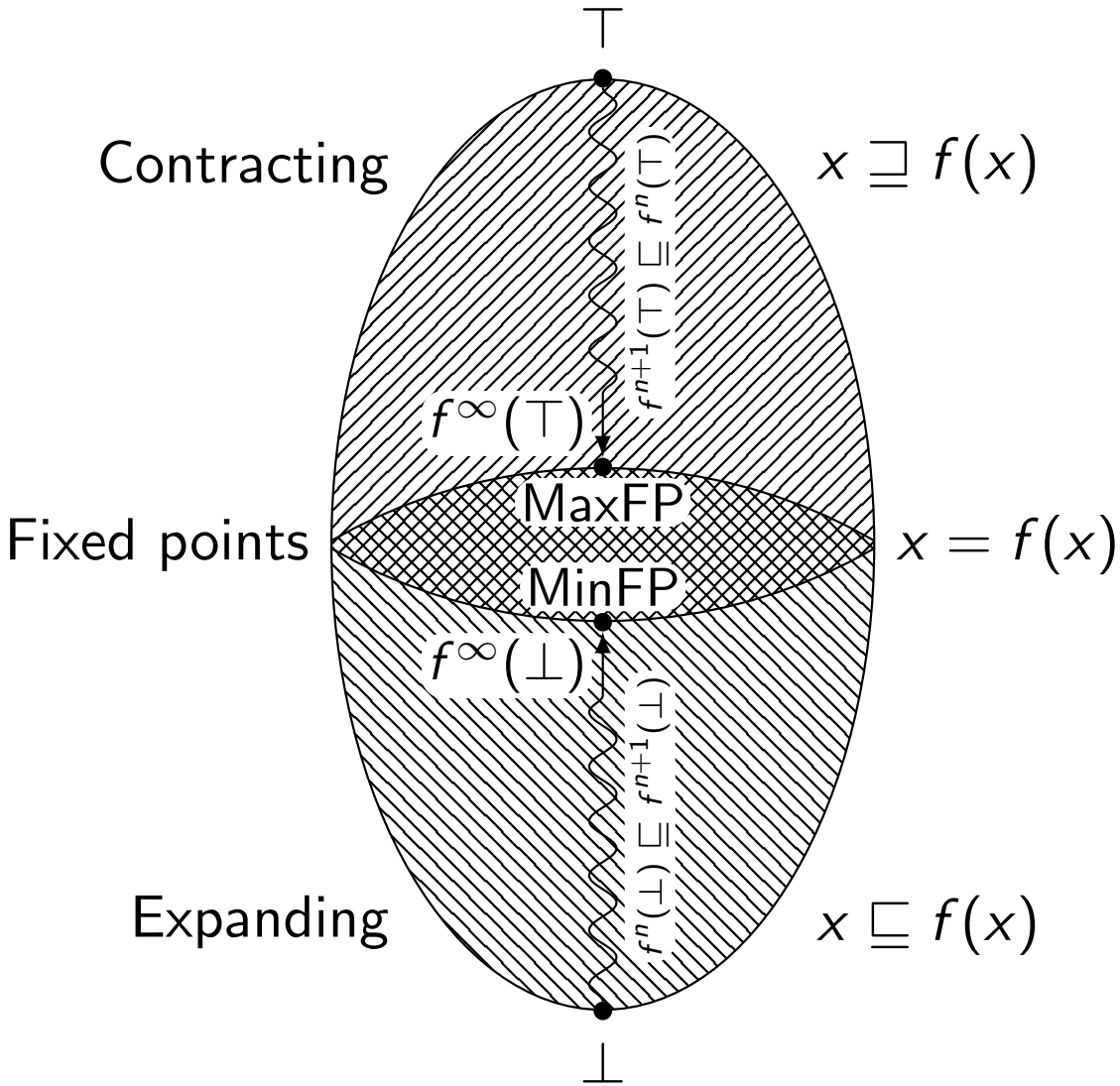
Data-Flow Analysis and Fixed-Point Iteration

- Initialize with $\bot$
  - Best possible value, but unsafe
- Initialisation for start $\iota = \top$
  - Alternatively at end, if backwards analysis
- Monotonous transfer functions
  - $a \sqsubseteq b \implies f(a) \sqsubseteq f(b)$

TODO

Fixed-Point Iteration Graphically

The Fried-Egg Picture



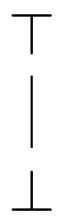
Lifting to Tuples

- Compare elementwise
- $(x_1, \ldots, x_n) \sqsubseteq (y_1, \ldots, y_n) :\Leftrightarrow x_1 \sqsubseteq y_1 \wedge \cdots \wedge x_n \sqsubseteq y_n$
- If order is broken in one place, the two tuples are uncomparable

Reasons for Overapproximation

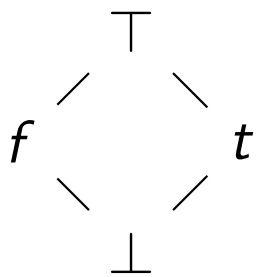
- Merge at each step (MFP) instead of only at end of paths (MOP)
  - MFP: Minimal Fixed Point
  - MOP: Merge Over all Paths
  - MFP *computable*, MOP generally *undecidable*
- Abstraction does not cover full semantics of program
  - Can be alleviated by *domain cooperation*
- Lattice not full powerset of concrete behavior
  - Reduces space and time requirements
  - E.g. flat lattice $\mathbb{Z}_{\perp}^{\top}$ for constant folding instead of $2^{\mathbb{Z}}$

Example: Trivial Lattice



- Example: Reachability
- $T := \{reachable\}$
- $S := 2^T = \{\emptyset, T\}$
- $\sqsubseteq := \subset$
- Implies: $\bot = \emptyset, \top = T, \sqcup = \cup$
- Efficient implementation: bit vector
- Alternative calculation: Depth-first search
- Dual lattice:
  - $T = \{unreachable\}$
  - $\sqsubseteq := \supset$
  - $\bot = T, \top = \emptyset, \sqcup = \cap$

Example: Boolean Lattice

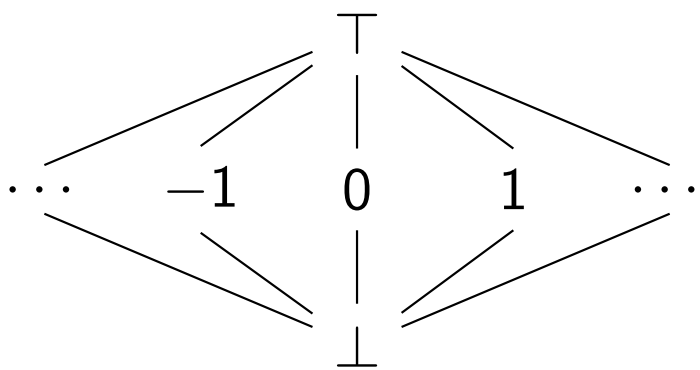


- Example: Branch conditions
- $T := \{f, t\}$
- $S := 2^T = \{\emptyset, \{f\}, \{t\}, T\}$
- $\sqsubseteq := \subset$
- Implies: $\bot = \emptyset, \top = T, \sqcup = \cup$
- Beware: $f \not\sqsubseteq t, f \not\sqsupseteq t$
- Efficient implementation: bit vector
  - $\bot \mapsto 00$
  - $f \mapsto 01$
  - $t \mapsto 10$
  - $\top \mapsto 11$
  - $\sqcup \mapsto \vee$
  - $a \sqsubseteq b \mapsto a \wedge \overline{b} = 0$

Comparison: Trivial and Boolean Lattice

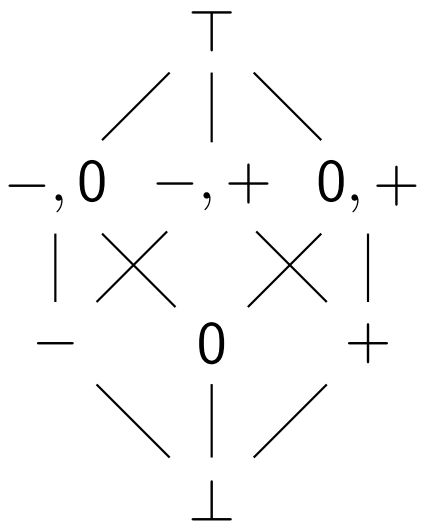
TODO

Example: Flat Lattice



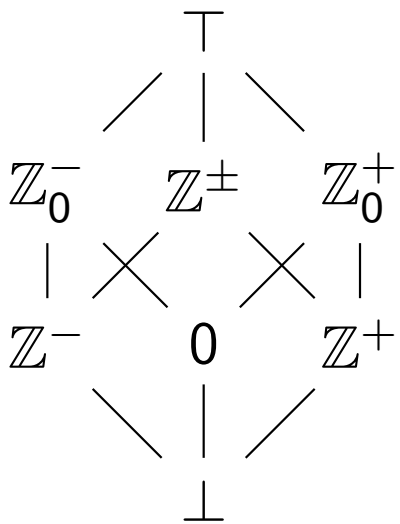
- Example: Constant propagation
- $T := \mathbb{Z}$
- $S := \{\emptyset, \ldots, \{-1\}, \{0\}, \{1\}, \ldots, \mathbb{Z}\}$
- $\sqsubseteq := \subset$
- Implies: $\bot = \emptyset, \top = \mathbb{Z}, \sqcup = \cup$
- $\{0\} \cup \{1\} = \{0, 1\} \notin S$
  - Next higher element is $\top$
- Beware: $-1 \not\sqsubseteq 0 \not\sqsubseteq 1, -1 \not\sqsupseteq 0 \not\sqsupseteq 1$

Example: Power-Set Lattice



- Example: Sign analysis
- $T := \{-, 0, +\}$
- $S := 2^T$
- $\sqsubseteq := \subset$
- Implies: $\bot = \emptyset, \top = T, \sqcup = \cup$
- Efficient implementation: bit vector
  - Analogous to boolean lattice, 3 bits
- TODO γ

Alternative Formulation of Sign Lattice



- Example: Sign analysis
- $T := \mathbb{Z}$
- $S := \{\emptyset, \mathbb{Z}^-, 0, \mathbb{Z}^+, \mathbb{Z}_0^-, \mathbb{Z}^\pm, \mathbb{Z}_0^+, \mathbb{Z}\}$
- $\sqsubseteq := \subset$
- Implies: $\bot = \emptyset, \top = \mathbb{Z}, \sqcup = \cup$
- Same implementation as $-, 0, +$
- Trivial γ
- Not a power-set lattice

Example: Power-Set Lattice

TODO dominance lattice, dual non-dominance lattice

Example: Partition Lattice

TODO GVN

Comparison: Power Set and Partition Lattice

TODO