Program Analysis

https://www.cse.iitb.ac.in/~karkare/cs618/

Foundations of Data Flow Analysis

Amey Karkare
Dept of Computer Science and Engg
IIT Kanpur
Visiting IIT Bombay
karkare@cse.iitk.ac.in
karkare@cse.iitb.ac.in

Taxonomy of Dataflow Problems

- Categorized along several dimensions
  - the information they are designed to provide
  - the direction of flow
  - confluence operator
- Four kinds of dataflow problems, distinguished by
  - the operator used for confluence or divergence
  - data flows backward or forward

2

Taxonomy of Dataflow Problems

Confluence→ Direction↓	$\cup$	$\cap$
Forward	Reaching Definition	Available Expressions
Backward	Live Variables	Very Busy Expressions

3

When does Data Flow Analysis Works?

- Suitable initial values and boundary conditions
- Suitable domain of values
  - Bounded, Finite
- Suitable meet operator
- Suitable flow functions
  - monotonic, closed under composition
- But what is “SUITABLE” ?

4

Why Data Flow Analysis Works?

- Suitable initial … dary condition…
- Suita…
  - B…
- Su…
- Sui…
  - mon… …mposition
- But what is "SUITABLE" ?

Lattice Theory

5

Partially Ordered Sets

- Posets
- S : a set
- ≤ : a relation
- (S, ≤) is a poset if ∀ x, y ,z ∈ S
  - x ≤ x (reflexive)
  - x ≤ y and y ≤ x ⇒ x = y (antisymmetric)
  - x ≤ y and y ≤ z ⇒ x ≤ z (transitive)

6

Chain

- Linear Ordering
- Poset where every pair of elements is comparable
- x1 ≤ x2 ≤ … ≤ xk is a chain of length k
- We are interested in chains of **finite** length

7

Observation

- Any finite nonempty subset of a poset has minimal and maximal elements
- Any finite nonempty chain has unique minimum and maximum elements

8

Semilattice

- Set S and meet $\wedge$
- $\forall$ x, y, z $\in$ S
  - x $\wedge$ x = x (idempotent)
  - x $\wedge$ y = y $\wedge$ x (commutative)
  - x $\wedge$ (y $\wedge$ z) = (x $\wedge$ y) $\wedge$ z (associative)
- Partial order for semilattice
  - x $\leq$ y if and only if x $\wedge$ y = x
  - Reflexive, antisymmetric, transitive

9

Border Element

- Top Element ($\top$)
  - $\forall x \in S, x \wedge \top = \top \wedge x = x$
- (Optional) Bot Element ($\perp$)
  - $\forall x \in S, x \wedge \perp = \perp \wedge x = \perp$

10

Familiar (semi)lattices

- Powerset for a set S, 2^S
- Meet $\wedge$ is $\cap$
- Partial Order is $\subseteq$
- Top element is S
- Bottom element is $\emptyset$

11

Familiar (semi)lattices

- Powerset for a set S, 2^S
- Meet $\wedge$ is $\cup$
- Partial Order is $\supseteq$
- Top element is $\emptyset$
- Bottom element is S, the universal set

12

Greatest Lower bound

- glb of x and y is an element g s.t.
  - $g \leq x$
  - $g \leq y$
  - If $z \leq x$ and $z \leq y$ then $z \leq g$
- $x \wedge y$ is glb of x and y (Prove!)

13

Semi (?)-Lattice

- We can define symmetric concepts:
  - $\geq$ order
  - $\vee$, Join operation
  - Least upper bound (lub)
- A complete lattice has both meet and join
  - Powerset lattice
- We will talk about "meet" semi-lattices only

14

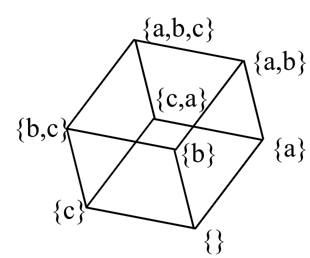
Lattice Diagrams

- Graphical view of posets
- Elements = nodes in the graph
- If x < y then x is depicted lower than y in the diagram
- An edge between x and y (x lower than y) implies x < y and no other element z s.t. x < z < y (i.e. transitivity excluded)

15

Lattice Diagram

$\top = \{a, b, c\}$
$\bot = \emptyset$
$meet = \cap$



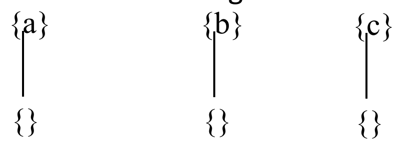
Lattice of superset relation

$x \wedge y$ (glb): the highest *z* for which there are paths downward from both x and y.

16

What if we have a large number of elements?

- Combine simple lattices to build a complex one
- Superset lattices for singletons



- Combine to form superset lattice for multi-element sets

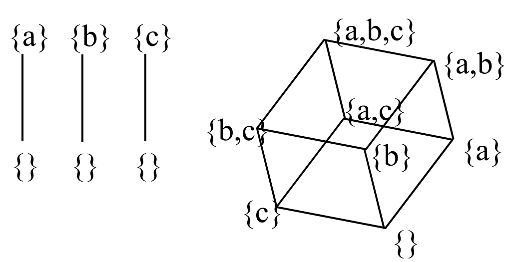
17

Product Lattice

- (S, ∧) is product lattice of (S_1, $∧_1$) and (S_2, $∧_2$)
  - $S = S_1 \times S_2$ (domain)
  - For (a_1,a_2) and (b_1, b_2) ∈ S
    - $(a_1,a_2) \wedge (b_1,b_2) = (a_1 \wedge_1 b_1, a_2 \wedge_2 b_2)$
    - $(a_1, a_2) \leq (b_1, b_2)$ iff $a_1 \leq_1 b_1$ and $a_2 \leq_2 b_2$
  - ≤ relation follows from ∧
- Product of lattices is associative
- $∧_1$, $∧_2$, … are called component lattices

18

Product Lattice



19

Height of a Semilattice

- Length of a chain x1 ≤ x2 ≤ … ≤ xk is k
- K = max over length of all chains in the semilattice
- Height of semilattice = K-1

20

Data Flow Analysis Framework

- (D, S, Λ, F)
- D: direction, Forward or Backward
- (S, Λ): Semilattice Domain and meet
- F: family of transfer functions, S->S

Transfer Functions

- F: family of functions, S -> S. Includes
  - functions suitable for the boundary conditions (constant transfer functions for ENTRY and EXIT nodes)
  - Identity function I: $I(x) = x \quad \forall x \in S$
- Closed under composition
  - $f, g \in F, h(x) = g\big(f(x)\big) \Rightarrow h \in F$

Monotonic Functions

- (S,≤) : a poset
- f: S->S is monotonic iff
  $\forall x, y \in S \quad x \leq y \Rightarrow f(x) \leq f(y)$
- Composition preserves monotonicity
  - If f and g are monotonic, h = f.g, then h is also monotonic

Monotone Frameworks

- (D, S, Λ, F) is monotone if the family F consists of monotonic functions only
  $f \in F, x, y \in S \; x \leq y \Rightarrow f(x) \leq f(y)$
- Equivalently
  $f \in F, x, y \in S \; f(x \wedge y) \leq f(x) \wedge f(y)$
  - Proof: Exercise

A Fixed Point Theorem

- $f: S -> S$ a monotonic function
- (S, Λ) is a finite height semilattice,
- ⊤ is top element
- $f^0(x) = x, f^{i+1}(x) = f\left(f^i(x)\right), i \geq 0$
- The greatest fixed point of f is $f^k(\top)$ where $f^{k+1}(\top) = f^k(\top)$

25

Fixed Point Algorithm

```
// monotonic f on a meet semilattice
x := ⊤ ;

while (x != f(x))    x := f(x);

return x;
```

26