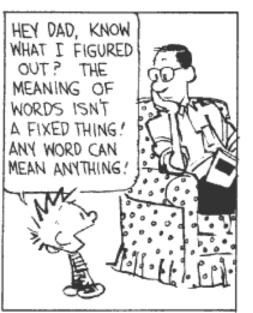
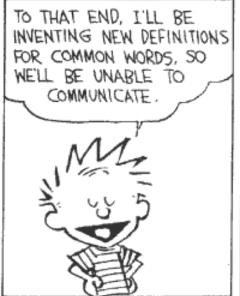
Abstract Interpretation

(Galois, Collections, Widening)



BY GIVING WORDS NEW MEANINGS, ORDINARY ENGLISH CAN BECOME AN EXCLUSIONARY CODE! TWO GENERATIONS CAN BE DIVIDED BY THE SAME LANGUAGE!







One Slide Summary

- In abstract interpretation, the abstraction function β and concretization function γ form a Galois connection: they are almost inverses.
- To abstract the state σ at each program point we use a collecting semantics (the abstract domain holds sets of states). This shows the link between abstract interpretation and model checking.
- This will result in recursively-defined equations. We use the fixed point theorem to solve them. This shows the link between abstract interpretation and dataflow analysis.
- Widening operators help accelerate convergence.

Abstract Interpretation

- We have an abstract domain A
 - e.g., A = { positive, negative, zero }
 - An abstraction function $\beta: \mathbb{Z} \to A$
 - Z is our concrete domain
 - A concretization function $\gamma: A \to \mathcal{P}(\mathbb{Z})$
- Positive + Positive = ???
- Positive + Negative = ???
- Positive / Zero = ???

"The Usual Combinations"



POPL 2025 (one month ago!)

Calculational Design of Hyperlogics by Abstract Interpretation

- we design a generic, structural, fixpoint abstract interpreter parameterized by an algebraic abstract domain describing finite and infinite computations that can be instantiated for various operational, denotational, or relational program semantics.

The Best of Abstract Interpretations

- we put forward a program logic parameterized on an abstract domain A which infers triples $[pre]_A p [post]_A$. These triples encode that the inductive abstract interpretation of p on A with abstract input $pre \in A$ gives $post \in A$ as abstract output and this is the best possible in A.

Linear and Non-linear Relational Analyses for Quantum Program Optimization

- Abstract interpretation has previously been applied to the problems of simulating circuits [9] or verifying properties of circuits including separability [42] and projection-based assertions [21, 58]. The work of [58] gave an abstract domain of products of subspaces on subsets of qubits and applied this to the verification of assertions in static (but large) quantum circuits.

Today's Plan

- We introduced abstract interpretation
- An abstraction mapping from concrete to abstract values
 - Has a concretization mapping which forms a Galois connection

- We'll look a bit more at Galois connections
- We'll lift AI from expressions to programs
- ... and we'll discuss the mythic "widening"

Why Galois Connections?

- We have an abstract domain A
 - An abstraction function $\beta: \mathbb{Z} \to A$
 - Induces $\alpha : \mathcal{P}(\mathbb{Z}) \to \mathsf{A}$ and $\gamma : \mathsf{A} \to \mathcal{P}(\mathbb{Z})$
- We argued that for correctness

$$\gamma(a_1 \text{ op } a_2) \supseteq \gamma(a_1) \text{ op } \gamma(a_2)$$

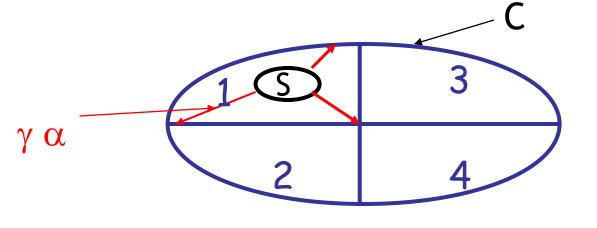
- We wish for the set on the left to be as small as possible
- To reduce the loss of information through abstraction
- For each set $S \subseteq C$, define $\alpha(S)$ as follows:
 - Pick smallest S' that includes S and is in the image of γ
 - Define $\alpha(S) = \gamma^{-1}(S')$
 - Then we define: $a_1 \underline{op} a_2 = \alpha(\gamma(a_1) \underline{op} \gamma(a_2))$
- Then α and γ form a Galois connection

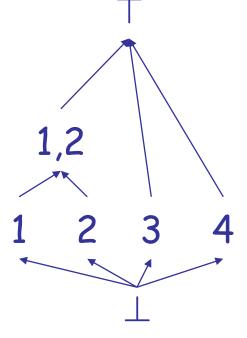
Galois Connections

- A Galois connection between complete lattices A and $\mathcal{P}(C)$ is a pair of functions α and γ such that:
 - γ and α are monotonic
 - (with the \subseteq ordering on $\mathcal{P}(C)$)
 - α (γ (a)) = a for all a \in A

- γ (α (S)) \supseteq S

for all $S \in \mathcal{P}(C)$





More on Galois Connections

- All Galois connections are monotonic
- In a Galois connection one function uniquely and absolutely determines the other



Évariste Galois

Abstract Interpretation for Imperative Programs

- So far we abstracted the value of expressions
- Now we want to abstract the state at each point in the program
- First we define the concrete semantics that we are abstracting
 - We'll use a collecting semantics

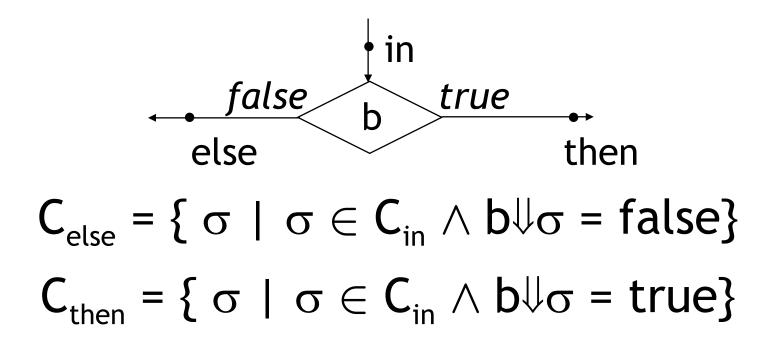
Collecting Semantics

- Recall
 - A state $\sigma \in \Sigma$. Any state σ has type $\text{Var} \to \mathbb{Z}$
 - States vary from program point to program point
- We introduce a set of program points: <u>labels</u>
- We want to answer questions like:
 - Is x always positive at label i?
 - Is x always greater or equal to y at label j?
- To answer these questions we'll construct
 - $C \in Contexts$. C has type Labels $\to \mathcal{P}(\Sigma)$
 - For each label i, C(i) = all possible states at label i
 - This is called the collecting semantics of the program
 - This is software model checkers approximate (e.g., using BDDs to store $\mathcal{P}(\Sigma)$ efficiently)

- We first define relations between the collecting semantics at different labels
 - We do it for unstructured CFGs (cf. HW6!)
 - Can do it for IMP with careful notion of program points
- Define a label on each edge in the CFG
- For assignment

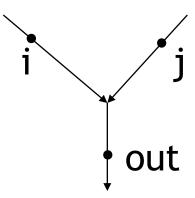
$$c_j = \{\sigma[x := n] \mid \sigma \in C_i \land e \forall \sigma = n\}$$

For conditionals



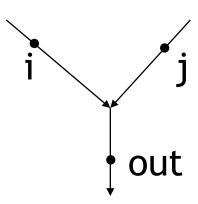
Assumes b has no side effects (as in IMP or HW6)

For a join



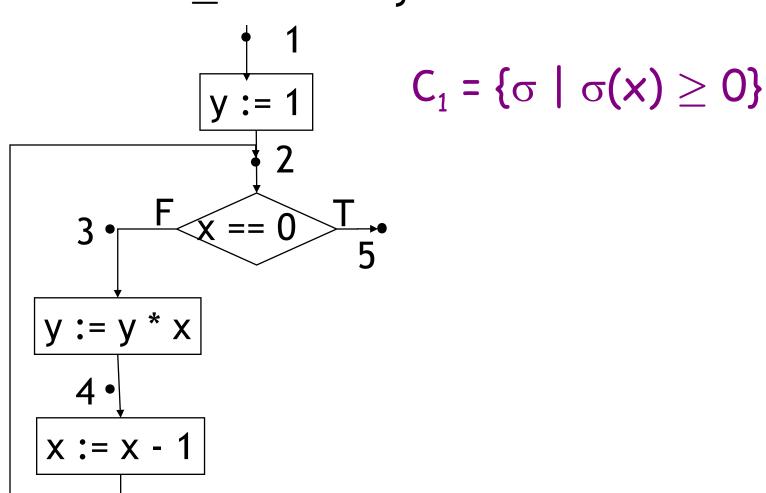


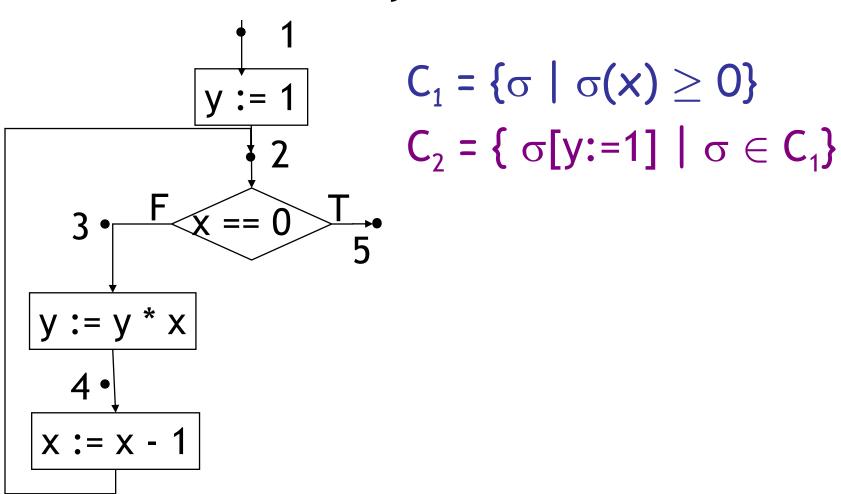
For a join

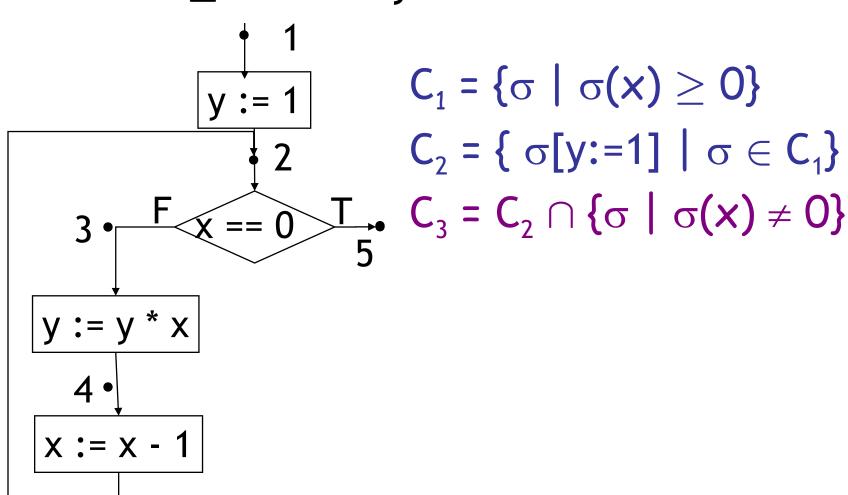


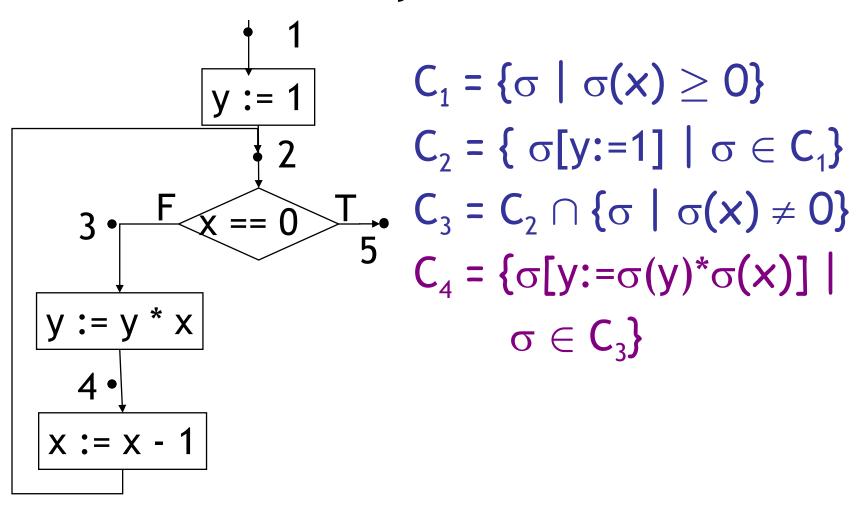
$$C_{out} = C_i \cup C_j$$

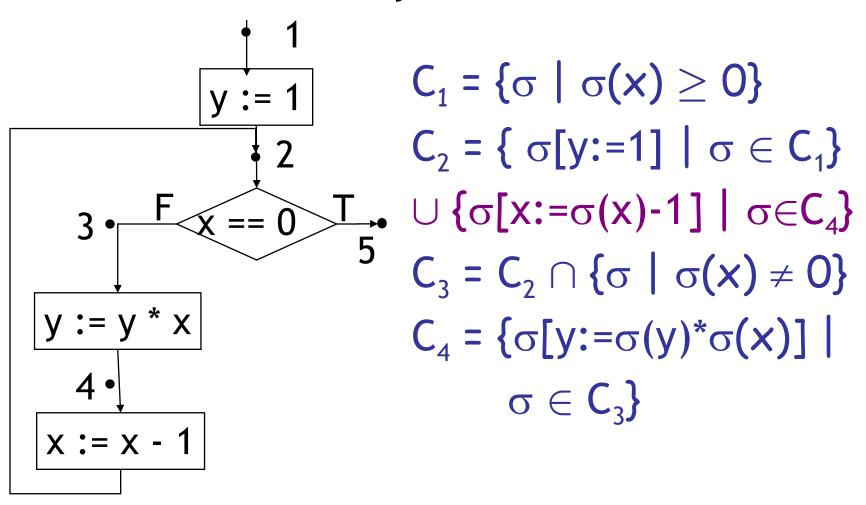
- Verify that these relations are monotonic
 - If we increase a C_x all other C_v can only increase

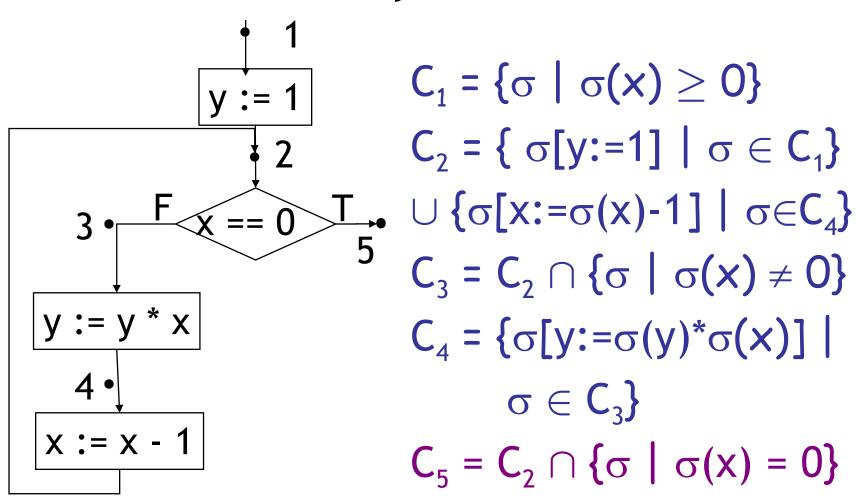










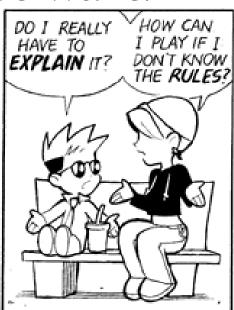


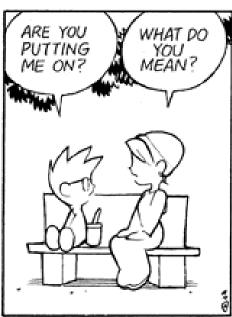
Why Does This Work?

- We just made a system of recursive equations that are defined largely in terms of themselves
 - e.g., $C_2 = F(C_4)$, $C_4 = G(C_3)$, $C_3 = H(C_2)$
- Why do we have any reason to believe that this will get us what we want?



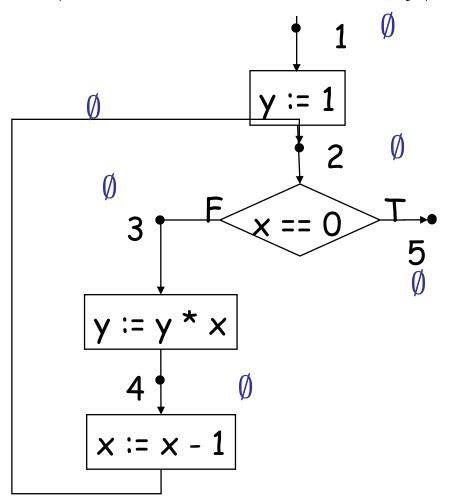






The Collecting Semantics

- We have an equation with the unknown C
 - The equation is defined by a monotonic and continuous function on domain Labels $\to \mathcal{P}(\Sigma)$
- We can use the least fixed-point theorem
 - Start with $C^0(L)=\emptyset$ (aka $C^0 = \lambda L.\emptyset$)
 - Apply the relations between C_i and C_j to get C¹_i
 from C⁰_i
 - Stop when all $C^k = C^{k-1}$
 - Problem: we'll go on forever for most programs
 - But we know the fixed point exists



$$C_{1} = \{\sigma \mid \sigma(x) \geq 0\}$$

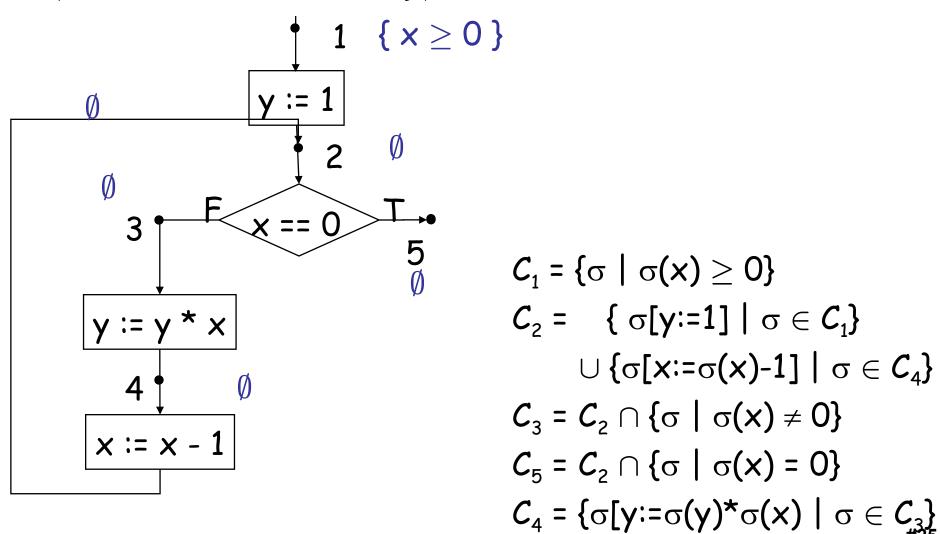
$$C_{2} = \{\sigma[y:=1] \mid \sigma \in C_{1}\}$$

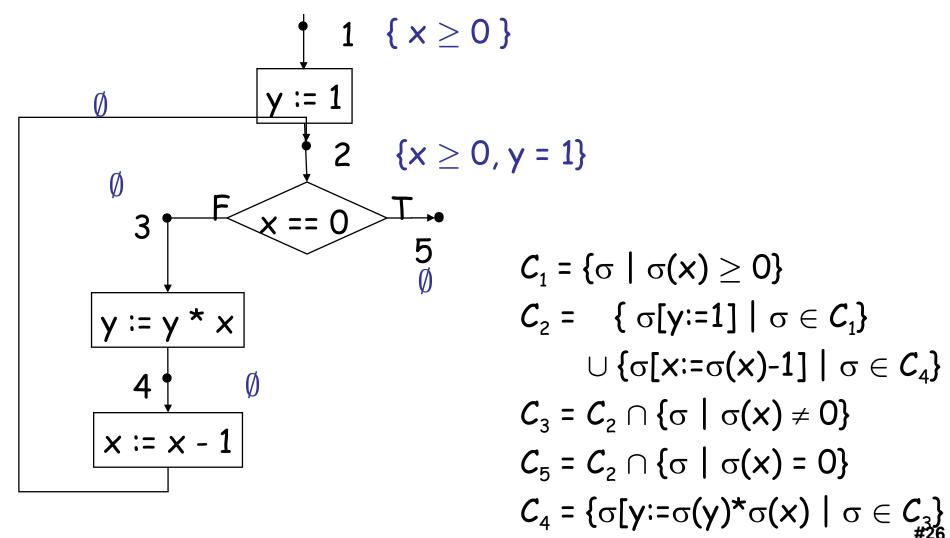
$$\cup \{\sigma[x:=\sigma(x)-1] \mid \sigma \in C_{4}\}$$

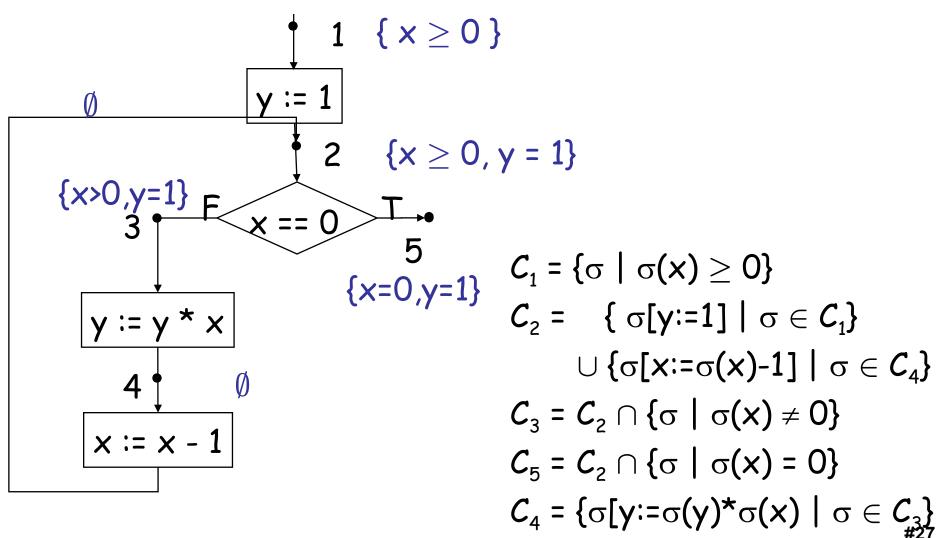
$$C_{3} = C_{2} \cap \{\sigma \mid \sigma(x) \neq 0\}$$

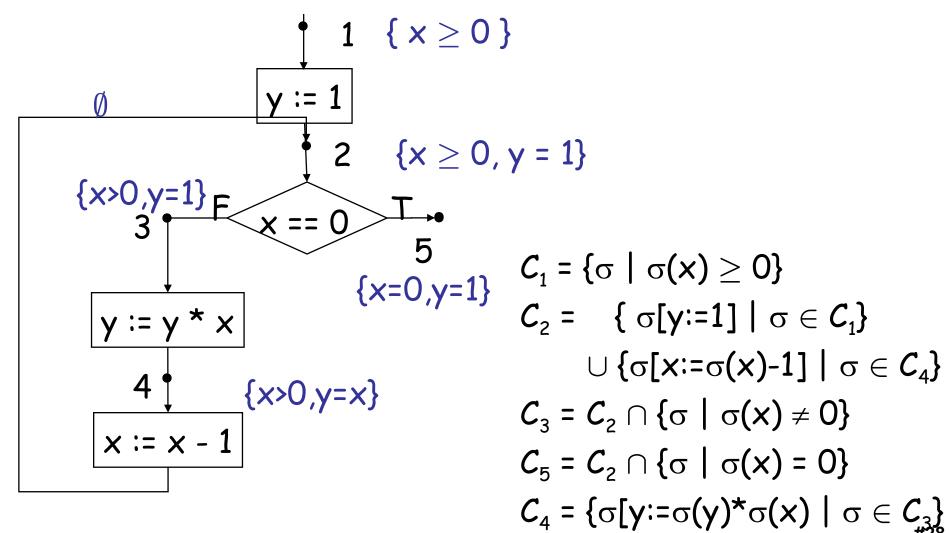
$$C_{5} = C_{2} \cap \{\sigma \mid \sigma(x) = 0\}$$

$$C_{4} = \{\sigma[y:=\sigma(y)*\sigma(x) \mid \sigma \in C_{3}\}$$

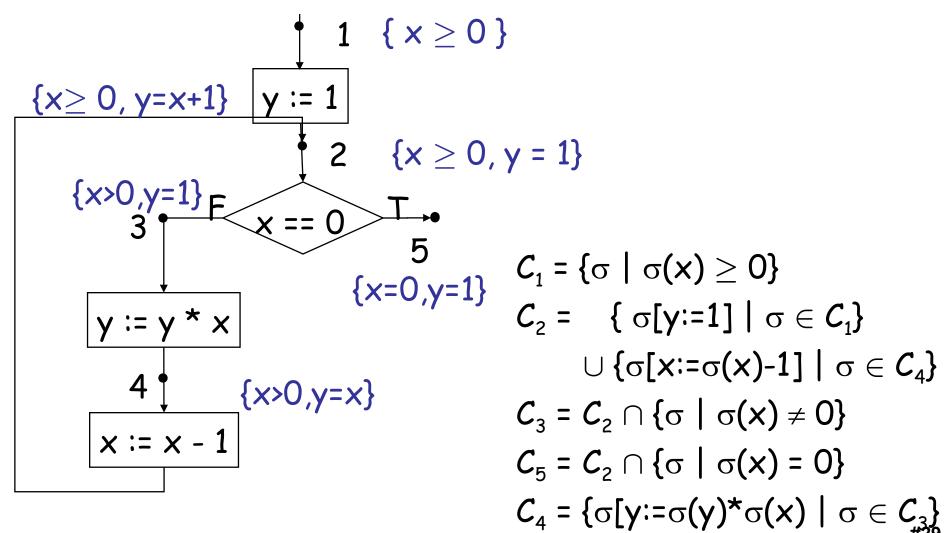


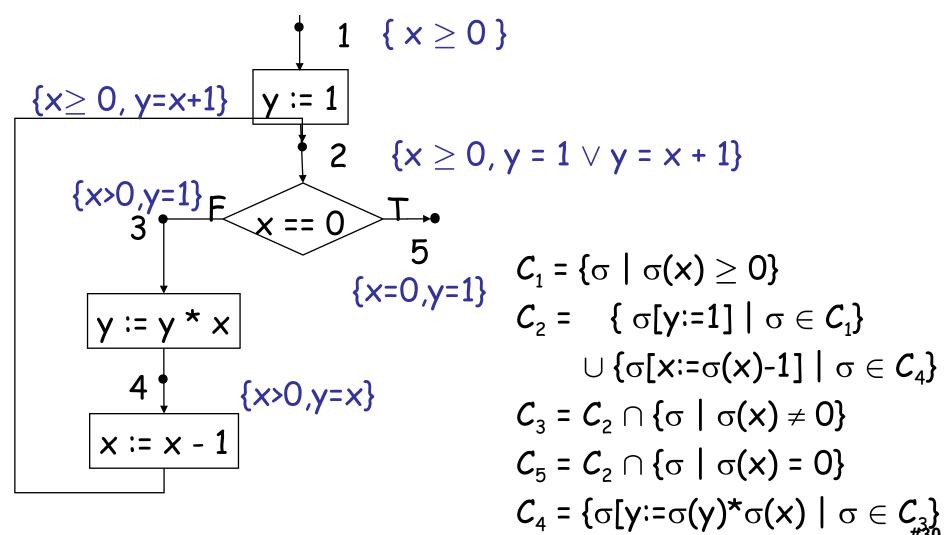






(assume x ≥ 0 initially)





Spanish Novels

• This 1605 Spanish novel, El ingenioso hidalgo de la Mancha, is considered foundational to Western literature and is the second most-translated book in the world. In it, a noble believes he is a knight-errant and travels with a witty farmer, along the way making social commentary about individuals and society.

Sanskrit Epics (Memory Test)

 This Sanskrit epic is one of the two great canon stories of India, and is attributed to the Hindu sage Valmiki. It covers dharma and human values while explaining the protagonist's attempt to recover his wife, Sita, who has been taken by the demons of Lanka. It is heavy on allegory and philosophy. Archery, including an epic use of the brahmastra, is often involved.

Botany

 Give the common name for the acer genus of trees and shrubs. Most species are deciduous, and many are renowned for their autumn leaf colors. Their distinctive seeds spin as they fall. They are popular in bonsai and in syrup production. Their wood is used in bowling pins, MLB baseball bats, and recurve bows. It is also a tonewood, and most violins, violas, cellos, double basses and electric guitars include components made of it.

Computer Science

 This American Turing-award winner is known for work on the B and C programming languages, the Unix and Plan 9 operating systems, regular expressions in text editors, UTF-8, and chess endgames. He co-developed the Go language at Google. Almost all programs that use regular expressions today use his notation for them (and he invented an algorithm for converting REs to NFAs).

Abstract Interpretation

- Pick a complete lattice A (abstractions for $\mathcal{P}(\Sigma)$)
 - Along with a monotonic abstraction $\alpha: \mathcal{P}(\Sigma) \to A$
 - Alternatively, pick $\beta : \Sigma \to A$
 - This uniquely defines its Galois connection γ
- Take the relations between C_i and move them to the abstract domain:

a: Label
$$\rightarrow$$
 A

Assignment

Concrete:
$$C_j = \{\sigma[x := n] \mid \sigma \in C_i \land e \forall \sigma = n\}$$

Abstract: $a_j = \alpha \{\sigma[x := n] \mid \sigma \in \gamma(a_i) \land e \forall \sigma = n\}$

Abstract Interpretation

Conditional

Concrete:
$$C_j = \{ \sigma \mid \sigma \in C_i \land b \forall \sigma = \text{false} \}$$
 and $C_k = \{ \sigma \mid \sigma \in C_i \land b \forall \sigma = \text{true} \}$
Abstract: $a_j = \alpha \{ \sigma \mid \sigma \in \gamma(a_i) \land b \forall \sigma = \text{false} \}$ and $a_k = \alpha \{ \sigma \mid \sigma \in \gamma(a_i) \land b \forall \sigma = \text{true} \}$

Join

Concrete:
$$C_k = C_i \cup C_j$$

Abstract: $a_k = \alpha \ (\gamma(a_i) \cup \gamma(a_i)) = \text{lub } \{a_i, a_i\}$

Least Fixed Points In The Abstract Domain

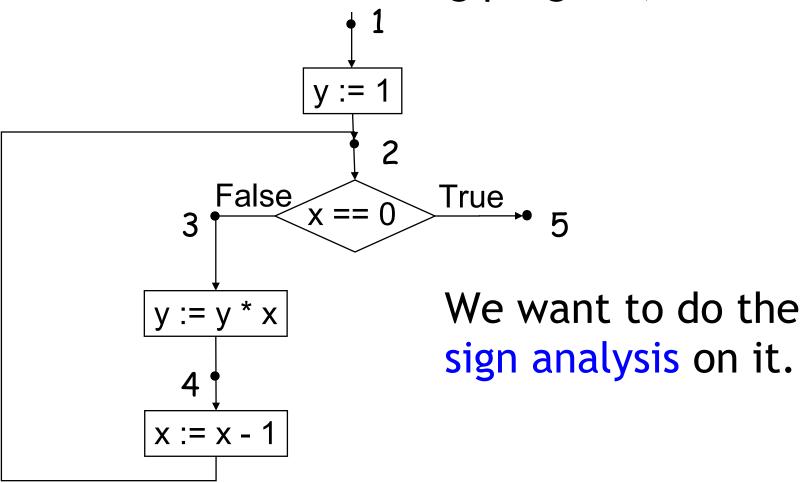
- We have a recursive equation with unknown "a"
 - Defined by a monotonic and continuous function on the domain Labels → A
- We can use the least fixed-point theorem:
 - Start with $a^0 = \lambda L. \perp$ (aka: $a^0(L) = \perp$)
 - Apply the monotonic function to compute a^{k+1} from a^k
 - Stop when $a^{k+1} = a^k$
- Exactly the same computation as for the collecting semantics
 - What is new?
 - "There is nothing new under the sun but there are lots of old things we don't know." Ambrose Bierce

Least Fixed Points In The Abstract Domain

- We have a hope of termination!
- Classic setup: A has only <u>uninteresting</u> chains (finite number of elements in each chain)
 - A has finite height h (= "finite-height lattice")
- The computation takes $O(h \times |Labels|^2)$ steps
 - At each step "a" makes progress on at least one label
 - We can only make progress h times
 - And each time we must compute |Labels| elements
- This is a quadratic analysis: good news
 - This is exactly the same as Kildall's 1973 analysis of dataflow's polynomial termination given a finite-height lattice and monotonic transfer functions.

Abstract Interpretation: Example

Consider the following program, x>0



Abstract Domain for Sign Analysis

Invent the complete sign lattice

$$S = \{ \perp, -, 0, +, \top \}$$

Construct the complete lattice

$$A = \{x, y\} \rightarrow S$$

- With the usual point-wise ordering
- Abstract state gives the sign for x and y
- We start with $a^0 = \lambda L \cdot \lambda v \in \{x,y\} \cdot \bot$
 - aka: $a^{0}(L,v) = \bot$

Let's Do It!

Label		Iterations $ ightarrow$								
1	X	+								
	у	H								
2	X	H								
	У	1								
3	X	1								
	У	1								
4	X	1								
	У	1								
5	X									
	У] #4,

Let's Do It!

Label						Iter	atio	ns —	>		
1	X	+									+
	у	T									T
2	X	1	+			Т					Т
	У		+						T		Т
3	X			+			T				Т
	У			+						T	Т
4	X	1			+			Т			T
	У	1			+			Т			T
5	X						0				0
	У	上					+			Τ	T #2

Notes, Weaknesses, Solutions

We abstracted the state of each variable independently

$$A = \{x, y\} \rightarrow \{\bot, -, 0, +, \top\}$$

- We lost relationships between variables
 - e.g., at a point x and y may always have the same sign
 - In the previous abstraction we get $\{x := \top, y := \top\}$ at label 2 (when in fact y is always positive!)
- We can also abstract the state as a whole

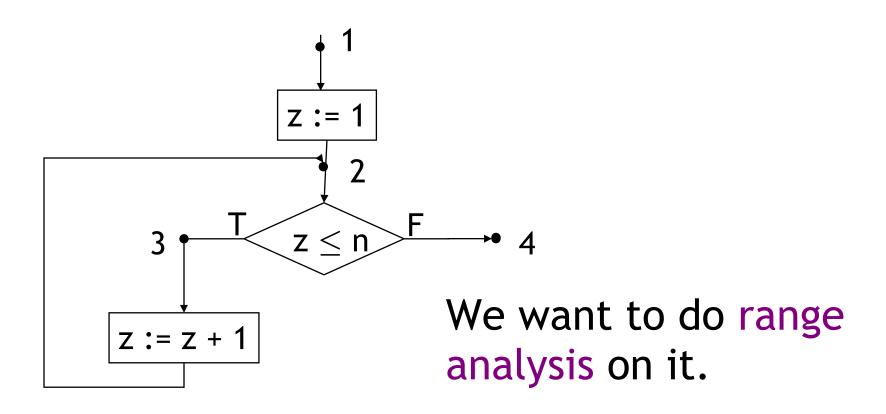
$$A = \mathcal{P}(\{\bot, -, 0, +, \top\} \times \{\bot, -, 0, +, \top\})$$

Other Abstract Domains

- Range analysis
 - Lattice of ranges: R ={ \perp , [n..m], (- ∞ , m], [n, + ∞), \top }
 - It is a complete lattice
 - $[n..m] \sqcup [n'..m'] = [min(n, n')..max(m,m')]$
 - [n..m] □ [n'..m'] = [max(n, n')..min(m, m')]
 - ullet With appropriate care in dealing with ∞
 - $\beta : \mathbb{Z} \to R$ such that $\beta(n) = [n..n]$
 - $\alpha : \mathcal{P}(\mathbb{Z}) \to \mathbb{R}$ such that $\alpha(S) = \text{lub } \{\beta(n) \mid n \in S\} = [\min(S)..\max(S)]$
 - $\gamma : \mathbb{R} \to \mathcal{P}(\mathbb{Z})$ such that $\gamma(r) = \{ n \mid n \in r \}$
- This lattice has infinite-height chains
 - So the abstract interpretation might not terminate!

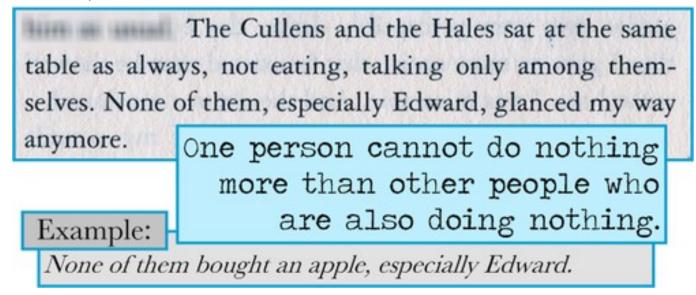
Example of Non-Termination

Consider this (common) program fragment



Example of Non-Termination

- Consider the sequence of abstract states at point 2
 - [1..1], [1..2], [1..3], ...
 - The analysis never terminates
 - Or terminates very late if the loop bound is known statically



Example of Non-Termination

- Consider the sequence of abstract states at point 2
 - [1..1], [1..2], [1..3], ...
 - The analysis never terminates
 - Or terminates very late if the loop bound is known statically
- It is time to approximate even more: widening
- We redefine the join (lub) operator of the lattice to ensure that from [1..1] upon union with [2..2] the result is $[1..+\infty)$ and not [1..2]
- Now the sequence of states is
 - [1..1], $[1, +\infty)$, $[1, +\infty)$ Done (no more infinite chains)

Formal Definition of Widening

(Cousot 16.399 "Abstract Interpretation", 2005)

- A widening $\nabla : (P \times P) \rightarrow P$ on a poset $\langle P, \sqsubseteq \rangle$ satisfies:
 - $\neg \forall x, y \in P$. $x \sqsubseteq (x \bigtriangledown y) \land y \sqsubseteq (x \bigtriangledown y)$
 - For all increasing chains $x^0 \sqsubseteq x^1 \sqsubseteq ...$ the increasing chain $y^0 = ^{def} x^0$, ..., $y^{n+1} = ^{def} y^n \bigtriangledown x^{n+1}$, ... is <u>not</u> strictly increasing.
- Two different main uses:
 - Approximate missing lubs. (Not for us.)
 - Convergence acceleration. (This is the real use.)
 - A widening operator can be used to effectively compute an upper approximation of the least fixpoint of $F \in L \nabla L$ starting from below when L is computer-representable but does not satisfy the ascending chain condition.

Formal Widening Example $[1,1] \nabla [1,2] = [1,+\infty)$

• Range Analysis on z:

L0:
$$z := 1$$
;

L1: while z<99 do

L2:
$$z := z+1$$

L3: done /*z > 99 */

L4:

 $x_{j}^{Li} = def$ the jth iterative attempt to compute an abstract value for z at label Li

Recall lub
$$S = [min(S)..max(S)]$$

lub $\{[2,+\infty),[1,+\infty)\} = \{[1,+\infty)\}$

Original x ⁱ	Widened yi				
$X_{\Gamma 0}^{0} = \top$	$y_{0} = \bot$				
$x^{L1}_0 = [1,1]$	$y_{0}^{L_{0}} = [1,1]$				
$x_{0}^{L2} = [1,1]$	$y_{0}^{L2} = [1,1]$				
$x^{L3}_0 = [2,2]$	$y_{0}^{L3} = [2,2]$				
$x^{L2}_1 = [1,2]$	$y^{L2}_{1} = [1, +\infty)$				
$x^{L3}_1 = [2, +\infty)$	$y^{L3}_1 = [2, +\infty)$				
$X^{L4}_0 = [99, +\infty)$	$y^{\text{L4}}_{0} = [99, +\infty)$				
stable (fewer than 99 iterations!)					

Other Abstract Domains

- Linear relationships between variables
 - A convex <u>polyhedron</u> is a subset of \mathbb{Z}^k whose elements satisfy a number of inequalities:

$$a_1x_1 + a_2x_2 + ... + a_kx_k \ge c_i$$

- This is a complete lattice; linear programming methods compute lubs
- Linear relationships with at most two variables
 - Convex polyhedra but with ≤ 2 variables per constraint
 - Octagons $(x \pm y \ge c)$ have efficient algorithms
- Modulus constraints (e.g. even and odd)

Abstract Chatter

- AI, Dataflow and Software Model Checking
 - The big three (aside from flow-insensitive type systems) for program analyses
- Are in fact quite related:
 - David Schmidt. Data flow analysis is model checking of abstract interpretation. POPL '98.
- AI is usually flow-sensitive (per-label answer)
- Al can be path-sensitive (if your abstract domain includes ∨, for example), which is just where model checking uses BDD's

Abstract Interpretation Conclusions

- Disadvantages:
 - Without care, AI can be slow
 - When the lattices have infinite height and widening heuristics are used the result becomes unpredictable
- Al can be applied to many problems in functional and imperative programming languages
- Al is a very powerful technique that underlies a large number of program analyses
 - Including Dataflow Analysis and Model Checking