Abstract interpretation
Mathematical Tools

Inspiré des notes de cours du MPRI d'Antoine Miné, CNRS/ENS Ulm

Introduction

Concrete semantics

$$\mathcal{S}_i \in \mathcal{D} = \mathcal{P}(\{\texttt{I},\texttt{X}\} \to \mathbb{Z})$$

$(\mathcal{S}_0)$ `assume X in [0, 1000];`	$\mathcal{S}_0 = \{(i,x) : i,x \in \mathbb{Z}\} = \top$
$(\mathcal{S}_1)$ `I := 0`	$\mathcal{S}_1 = \{(i,x) \in \mathcal{S}_0 : x \in [0,1000]\} = F_1(\mathcal{S}_0)$
$(\mathcal{S}_2)$	$\mathcal{S}_2 = \{(0,x) : \exists i, (i,x) \in \mathcal{S}_1\} = F_2(\mathcal{S}_1)$
`while` $(\mathcal{S}_3)$ `I < X do`	$\mathcal{S}_3 = \mathcal{S}_2 \cup \mathcal{S}_5$
$(\mathcal{S}_4)$`I := I + 2;`	$\mathcal{S}_4 = \{(i,x) \in \mathcal{S}_3 : i < x\} = F_4(\mathcal{S}_3)$
$(\mathcal{S}_5)$	$\mathcal{S}_5 = \{(i+2,x) : (i,x) \in \mathcal{S}_4\} = F_5(\mathcal{S}_4)$
$(\mathcal{S}_6)$	$\mathcal{S}_6 = \{(i,x) \in \mathcal{S}_3 : i \geq x\} = F_6(\mathcal{S}_3)$

- smallest solution of a system of equations
- strongest invariant (and an inductive invariant)
- not computable in general

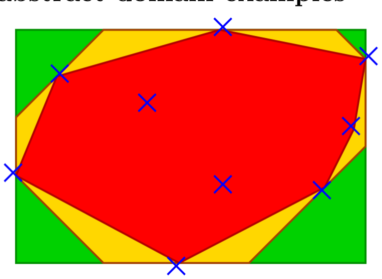
Abstract semantics

$$\mathcal{S}_i^\sharp \in \mathcal{D}^\sharp$$

$(\mathcal{S}_0)$ `assume X in [0, 1000];`	$\mathcal{S}_0^\sharp = \top^\sharp$
$(\mathcal{S}_1)$ `I := 0`	$\mathcal{S}_1^\sharp = F_1^\sharp(\mathcal{S}_0^\sharp)$
$(\mathcal{S}_2)$	$\mathcal{S}_2^\sharp = F_2^\sharp(\mathcal{S}_1^\sharp)$
`while` $(\mathcal{S}_3)$ `I < X do`	$\mathcal{S}_3^\sharp = \mathcal{S}_2^\sharp \cup^\sharp \mathcal{S}_5^\sharp$
$(\mathcal{S}_4)$`I := I + 2;`	$\mathcal{S}_4^\sharp = F_4^\sharp(\mathcal{S}_3^\sharp)$
$(\mathcal{S}_5)$	$\mathcal{S}_5^\sharp = F_5^\sharp(\mathcal{S}_4^\sharp)$
$(\mathcal{S}_6)$	$\mathcal{S}_6^\sharp = F_6^\sharp(\mathcal{S}_3^\sharp)$

- $\mathcal{D}^\sharp$ subset of properties of interest (with a machine repr.))
- $F^\sharp : \mathcal{D}^\sharp \to \mathcal{D}^\sharp$ over-approximates the effect of $F : \mathcal{D} \to \mathcal{D}$ in $\mathcal{D}^\sharp$ (with effective algorithms).

Numeric abstract domain examples



concrete sets $\mathcal{D}$	$\{(0,3),(5.5,0),(12,7),\dots\}$	not comp.
abs. polyhedra $\mathcal{D}_p^\sharp$	$6X + 11Y \geq 33 \wedge \dots$	exp. cost
abs. octogons $\mathcal{D}_o^\sharp$	$X + Y \geq 3 \wedge Y \wedge 0 \wedge \dots$	cubic cost
abs. intervals $\mathcal{D}_i^\sharp$	$X \in [0,12] \wedge Y \in [0,8]$	linear cost

Trade-off between cost and expressiveness/precision.

Galois connection

$$(\mathcal{D}, \subseteq) \underset{\alpha}{\overset{\gamma}{\leftrightarrows}} (\mathcal{D}^\sharp, \subseteq^\sharp)$$

$$\alpha(X) \subseteq^\sharp Y^\sharp \quad \Leftrightarrow \quad X \subseteq \gamma(Y^\sharp)$$

- $\alpha(X)$ is the best abstraction of X in $\mathcal{D}^\sharp$
- $F^\sharp = \alpha \circ F \circ \gamma$ is the best abstraction of F in $\mathcal{D}^\sharp \to \mathcal{D}^\sharp$

Example. (interval domain $\mathcal{D}_i^\sharp$)

- $[l_1, h_1] \subseteq_i^\sharp [l_2, h_2] \Leftrightarrow l_1 \geq l2 \wedge h_1 \leq h_2$
- $\gamma_i([l,h]) = \{x \in \mathbb{Z} : l \leq x \leq h\}$
- $\alpha_i(X) = [\min X, \max X]$

Resolution by iteration and extrapolation

Problem. the equation system is recursive : $\overrightarrow{\mathcal{S}}^\sharp = \overrightarrow{F}^\sharp(\overrightarrow{\mathcal{S}}^\sharp)$

Solution. resolution by iteration : $\overrightarrow{\mathcal{S}}^{\sharp 0} = \varnothing^\sharp$, $\overrightarrow{\mathcal{S}}^{\sharp i+1} = \overrightarrow{F}^\sharp(\overrightarrow{\mathcal{S}}^{\sharp i})$

e.g. $\mathcal{S}_3^\wedge$: $\texttt{I} \in \varnothing$, $\texttt{I} = 0$, $\texttt{I} \in [0,2]$, $\texttt{I} \in [0,4]$, …, $\texttt{I} \in [0,1000]$

Problem. infinite or very long sequence of iterate in $\mathcal{D}^\sharp$

Solution. extrapolation operator $\triangledown$

e.g. $[0,2] \triangledown [0,4] = [0,+\infty[$

- remove unstable bounds and constraits
- ensures the convergence in finite time
- inductive reasoning (through generalisation)

Order theory

Definition. Set X. $\sqsubseteq \in X \times X$ is a **partial order** iff

1. reflexive: $\forall x \in X, x \sqsubseteq x$
2. antisymmetric: $\forall x, y \in X, x \sqsubseteq y \wedge y \sqsubseteq x \Rightarrow x = y$
3. transitive: $\forall x, y, z \in X, x \sqsubseteq y \wedge y \sqsubseteq x \wedge y \sqsubseteq z \Rightarrow x \sqsubseteq z$

$(X, \sqsubseteq)$ is a **poset** (partially ordered set).
Without antisymmetry, $\sqsubseteq$ is a **preorder**.
Examples. $(\mathbb{Z}, \leq)$, $(\mathcal{P}(X), \subseteq)$, $\forall S, (S, =)$.
Usage.

- logic: ordered by implication $\Rightarrow$
- approximations: $\sqsubseteq$ is an information order
- program verification: program semantics $\sqsubseteq$ specification

Definitions.

- c is an **upper bound of** a **and** b if $a \sqsubseteq c$ and $b \sqsubseteq c$
- c is an **least upper bound of** a **and** b (**lub** or **join**) if
  - c is and upper bound of a and b
  - for every upper bound d of a and b, $c \sqsubseteq d$

  Unique and noted $a \sqcup b$.

Generalized to upper bounds of arbitrary sets : $\sqcup Y$, $Y \subseteq X$.
$a \sqcap b$, $\sqcap Y$ are **greatest lower bounds** (**glb** or **meet**) if $a \sqcap b \sqsubseteq a, b$ and $\forall c, c \sqsubseteq a, b \Rightarrow c \sqsubseteq a \sqcap b$.
$C \subseteq X$ is a **chain in** $(X, \sqsubseteq)$ if it is totally ordered : $\forall x, y \in C, x \sqsubseteq y \vee y \sqsubseteq x$.
Poset $(X, \sqsubseteq)$ is a **complete partial order** (**CPO**) if every chain C (incl. $\varnothing$) has a least upper bound $\sqcup C$.
A CPO has a **least element** $\sqcup\varnothing$, denoted $\perp$.
Examples.

- $(\mathbb{N}, \leq)$ not complete, $(\mathbb{N} \cup \{\infty\}, \leq)$ compl.
- $(\{x \in \mathbb{Q} : 0 \leq x \leq 1\}, \leq)$ not compl. but $(\{x \in \mathbb{R} : 0 \leq x \leq 1\}, \leq)$ compl.
- $\forall Y, (\mathcal{P}(Y), \subseteq)$ compl.

Definition. Poset $(X, \sqsubseteq, \sqcup, \sqcap)$ is a **lattice** with

1. a lub $a \sqcup b$ for every pair of a and b
2. a glb $a \sqcap b$ for every pair of a and b

Examples.

- integer intervals: $(\{[a,b] : a, b \in \mathbb{Z}, a \leq b\} \cup \{\varnothing\}, \subseteq, \sqcup, \cap)$ where $[a,b] \sqcup [c,d] = [\min(a,a'), \max(b,b')]$.
- divisibility: $(\mathbb{N}^\star, |, \gcd, \mathrm{lcm})$

Definition. Poset $(X, \sqsubseteq, \sqcup, \sqcap, \perp, \top)$ is a **complete lattice** with

1. a lub $\sqcup S$ for every set $S \subseteq X$
2. a glb $\sqcap S$ for every set $S \subseteq X$
3. a least element $\perp$
4. a greatest element $\top$

Examples.

- real segment $[0,1]$: $([0,1], \leq, \max, \min)$
- powersets: $(\mathcal{P}(S), \subseteq, \cup, \cap, \varnothing, S)$
- finite lattices
- integer intervals with finite and infinite bounds

Derivation

Complete posets or lattices $(X, \sqsubseteq_X, \dots)$ and $(Y, \sqsubseteq_Y, \dots)$, we can derive new ones by: duality, adding a least element $\perp$ (lifting), product, point-wise lifting by some set S, sublattice.

Fixpoints

Definition. A function $f : (X, \sqsubseteq_X, \dots) \to (Y, \sqsubseteq_Y, \dots)$ is

- **monotonic** if $\forall x, x', x \sqsubseteq_X x' \Rightarrow f(x) \sqsubseteq_Y f(x')$
- **strict** if $f(\perp_X) = \perp_Y$
- **continuous between CPOs** if C chain $\subseteq X, \{f(c) : c \in C\}$ is a chain in Y and $f(\sqcup_X C) = \sqcup_Y \{f(c) : c \in C\}$
- a **complete** $\sqcup$**−morphism between complete lattices** if $\forall S \subseteq X, f(\sqcup_X S) = \sqcup_Y \{f(s) : s \in S\}$
- **extensive** if $X = Y$ and $\forall x, x \sqsubseteq_X f(x)$

Definition. Given $f : (X, \sqsubseteq) \to (X, \sqsubseteq)$

- x is a **fixpoint** of f if $f(x) = x$
- x is a **prefixpoint** of f if $x \sqsubseteq f(x)$
- x is a **postfixpoint** of f if $f(x) \sqsubseteq x$

$$\begin{aligned} \mathrm{fp}(f) &= \{x \in X : f(x) = x\} \\ \mathrm{lfp}_x\, f &= min_{\sqsubseteq}\{y \in \mathrm{fp}(f) : x \sqsubseteq y\} \text{ if exists} \\ \mathrm{lfp}\, f &= \mathrm{lfp}_\perp\, f \end{aligned}$$

Tarski's fixpoint theorem. If $f : X \to X$ is monotonic in a complete lattice X then $\mathrm{fp}(f)$ is a complete lattice.
Kleene fixpoint theorem. If $f : X \to X$ is continuous in a CPO X and $a \sqsubseteq f(a)$ then $\mathrm{lfp}_a(f)$ exists.
Definition. $(S, \sqsubseteq)$ is a **well-ordered set** if:

- $\sqsubseteq$ is a total-order on S
- every $X \subseteq S$ such that $X \neq \varnothing$ has a least element $\sqcap X \in X$

Consequence.

- Any elt $x \in S$ has a **succesor** $x+1 = \sqcap\{y : x \sqsubset y\}$
- If $\exists y, x = y+1$, x is a **limit** and $x = \sqcup\{y : y \sqsubset x\}$

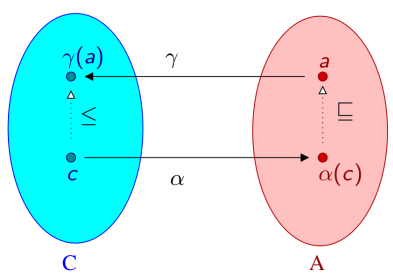
Examples. $(\mathbb{N}, \leq)$, $(\mathbb{N} \cup \{\infty\}, \leq)$
Definition. Given $f : X \to X$ and $a \in X$, the **transfinite iterates of** f are:

$$\begin{cases} x_0 &= a \\ x_n &= f(x_{n-1}) \text{if } n \text{ is a succesor ordinal} \\ x_n &= \sqcup\{x_m : m < n\} \text{if } n \text{ is a limit ordinal} \end{cases}$$

Tarski's fixpoint theorem. If $f : X \to X$ is monotonic in a complete lattice X and $a \sqsubseteq f(a)$ then $\text{lfp}_a(f) = x_\delta$ for some ordinal δ.
Definition. An ascending chain C in $(X, \sqsubseteq)$ is a sequence $c_i \in X$ such that $i \leq j \Rightarrow c_i \leq c_j$.
A poset $(X, \sqsubseteq)$ satisfies the **ascending chain condition** (**ACC**) iff for every ascending chain C, $\exists i \in \mathbb{N}, \forall j \geq i, c_i = c_j$ (similar definition for the **descending chain condition**).
Examples. $(\mathcal{P}(X), \subseteq)$ is ACC/DCC iff X is finite, $(\mathbb{Z} \cup \{\bot\}, \sqsubseteq)$ where $x \sqsubseteq y \Leftrightarrow x = \bot \vee x = y$ is ACC/DCC. $(\mathbb{N}^\star, |)$ is DCC but not ACC.
Kleene finite fixpoint theorem. If $f : X \leftarrow X$ is monotonic in an AAC poset X and $a \sqsubseteq f(a)$ then $\text{lfp}_a f$ exists.
Definition. Given posets $(C, \leq)$ and $(A, \sqsubseteq)$.
$(\alpha : C \to A, \gamma : A \to C)$ is a **Galois connection** iff

$$\forall a \in A, c \in C, \alpha(c) \sqsubseteq a \Leftrightarrow c \leq \gamma(a)$$

denoted $(C, \leq) \underset{\alpha}{\overset{\gamma}{\leftrightarrows}} (A, \sqsubseteq)$



α is the **upper adjoint** or **abstraction**; A is the **abstract domain**.
γ is the **lower adjoint** or **concretization**; C is the **concrete domain**.
Properties. Assume $\forall a, c, \alpha(c) \sqsubseteq a \Leftrightarrow c \leq \gamma(a)$

1. $\gamma \circ \alpha$ is **extensive** : $\forall c, c \leq \gamma(\alpha(c))$
2. $\alpha \circ \gamma$ is **reductive** : $\forall a, \alpha(\gamma(c)) \sqsubseteq a$
3. α is monotonic
4. γ is monotonic
5. $\gamma \circ \alpha \circ \gamma = \gamma$
6. $\alpha \circ \gamma \circ \alpha = \alpha$
7. $\alpha \circ \gamma$ and $\gamma \circ \alpha$ are idempotent

Corollary/Definition. $(\alpha : C \to A, \gamma : A \to C)$ is a Galois connection if:

1. γ is monotonic
2. α is monotonic
3. $\gamma \circ \alpha$ is extensive
4. $\alpha \circ \gamma$ is reductive

Corollary. Given $(C, \leq) \underset{\alpha}{\overset{\gamma}{\leftrightarrows}} (A, \sqsubseteq)$, each adjoint can be **uniquely defined** in term of the other:

1. $\alpha(c) = \sqcap\{a : c \leq \gamma(a)\}$
2. $\gamma(c) = \vee\{c : \alpha(c) \sqsubseteq a\}$

Corollary. Given $(C, \leq) \underset{\alpha}{\overset{\gamma}{\leftrightarrows}} (A, \sqsubseteq)$, then

1. $\forall X \subseteq C$, if $\vee X$ exists, then $\alpha(\vee X) = \sqcup\{\alpha(x) : x \in X\}$
2. $\forall X \subseteq C$, if $\sqcap$ exists, then $\gamma(\sqcap) = \wedge\{\gamma(x) : x \in X\}$

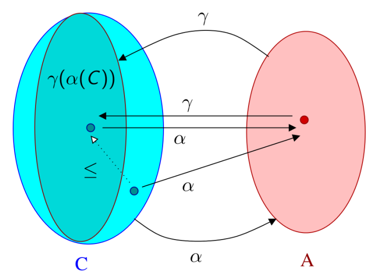
Deriving Galois connections

Corollary. Given $(C, \leq) \underset{\alpha}{\overset{\gamma}{\leftrightarrows}} (A, \sqsubseteq)$ and $(C', \leq') \underset{\alpha}{\overset{\gamma}{\leftrightarrows}} (A', \sqsubseteq')$, we can construct new Galois connections by duality, composition, point-wise lifting by some set S, functional lifting of monotonic operators.

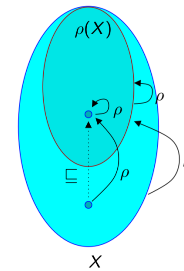
Galois embeddings

Definition. If $(C, \leq) \underset{\alpha}{\overset{\gamma}{\leftrightarrows}} (A, \sqsubseteq)$, (α, γ) is a **Galois embedding** – denoted $(C, \leq) \underset{\alpha}{\overset{\gamma}{\leftrightarrows}} (A, \sqsubseteq)$ – if it satisfies one of the following properties:

1. α is surjective
2. γ is injective
3. $\alpha \circ \gamma = id$



Corollary. A Galois conn. can be made into an embedding by **quotienting** A by the equivalence relation $a \equiv a' \Leftrightarrow \gamma(a) = \gamma(a')$.
Definition. $\rho : X \to X$ is an **upper closure** in the poset $(X, \sqsubseteq)$ if it is monotonic, extensive and idempotent.



Corollary. Given $(C, \leq) \underset{\alpha}{\overset{\gamma}{\leftrightarrows}} (A, \sqsubseteq)$, $\gamma \circ \alpha$ is an upper closure on $(C, \leq)$.
Corollary. Given an upper closure ρ on $(X, \sqsubseteq)$, we have $(X, \sqsubseteq) \underset{\rho}{\overset{id}{\leftrightarrows}} (\rho(X), \sqsubseteq)$.
Remark. We can rephrase abstract interpretation using upper closures instead of Galois connections, but we lose : the notion of **abstract representation** and the ability to have **several distinct** abstract representations for a single concrete object.

Sound, best, and exact abstractions

Definitions. Given $(C, \leq) \underset{\alpha}{\overset{\gamma}{\leftrightarrows}} (A, \sqsubseteq)$

- $a \in A$ is a **sound abstraction of** $c \in C$ if $c \leq \gamma(a)$ or $\alpha(c) \sqsubseteq a$.
- Given $c \in C$, its **best abstraction** is $\alpha(c)$
- $g : A \to A$ is a **sound abstraction of** $f : C \to C$ if $\forall a \in A, (f \circ \gamma)(a) \leq (\gamma \circ f)(a)$
- Given $f : C \xrightarrow{\leq} C$, its **best abstraction** is $\alpha \circ f \circ \gamma$
- $g : A \to A$ is an **exact abstraction of** $f : C \to C$ if $f \circ \gamma = \gamma \circ g$

Corollary. If g and g' abstract respectively f and f' then

- if f and f' are sound abstractions and f is monotonic then $g \circ g'$ is a sound abstraction of $f \circ f'$
- if g, g' are exact abstractions then $g \circ g'$ is an exact abstraction
- if g and g' are best abstractions, then $g \circ g'$ in **not** always a best abstraction

Fixpoint abstraction example theorem. Let $(C, \leq, \vee, \wedge, \bot, \top)$ be a complete lattice, $g : A \to A$ a sound abstraction of a monotonic $f : C \xrightarrow{\leq} C$ and a a postfixpoint of g then a is a sound abstraction of lfp f.