Overview

Selected advanced topics:

- reduced products of abstract domains
- disjunctive abstract domains
- inter-procedural analysis

Practical session:
- implement a reduced product
- help with the project
Idea

Theory:

- the set of abstract domains is a lattice,
- ordered by abstraction, which is a partial order, i.e.:
  \[(C, \leq) \xleftarrow{\gamma} (A, \sqsubseteq)\] denotes that \(C\) is more concrete than \(A\).
  
  (every property of \(A\) can also be represented exactly in \(C\))

- there is a least upper bound \(\sqcup\) for arbitrary sets of domains and a greatest lower bound \(\sqcap\).

Application: reduced product

Effective construction for the least upper bound \(A_1 \sqcap A_2\), able to represent properties expressible in either \(A_1\) or \(A_2\)

Benefit

We can design more precise analyses by combining existing abstractions.
Abstract domain lattice
Reminder: interval abstraction

\[ \mathcal{P}(\mathbb{Z}) \overset{\alpha_i}{\leftrightarrow} \{ [a, b] \mid a \leq b \} \cup \{ \perp \} \]

- \( \alpha_i(S) \overset{\text{def}}{=} [\min S, \max S] \)
- \( \gamma_i([a, b]) \overset{\text{def}}{=} \{ x \in \mathbb{Z} \mid a \leq x \leq b \} \)
Reminder: sign abstraction

\[ \mathcal{P}(\mathbb{Z}) \]

\[ \begin{align*}
\gamma_s(\bot) & \overset{\text{def}}{=} \emptyset \\
\gamma_s(0) & \overset{\text{def}}{=} \{0\} \\
\gamma_s(\geq 0) & \overset{\text{def}}{=} \mathbb{N} \\
\gamma_s(\leq 0) & \overset{\text{def}}{=} -\mathbb{N} \\
\gamma_s(\top) & \overset{\text{def}}{=} \mathbb{Z}
\end{align*} \]

\[ \alpha_s(S) \overset{\text{def}}{=} \begin{cases} 
\bot & \text{if } S = \emptyset \\
0 & \text{if } S = \{0\} \\
\geq 0 & \text{else, if } \forall s \in S, s \geq 0 \\
\leq 0 & \text{else, if } \forall s \in S, s \leq 0 \\
\top & \text{otherwise}
\end{cases} \]
Composing abstractions

We can compose Galois connections:

If \((X_1, \subseteq_1) \xrightarrow{\gamma_1} (X_2, \subseteq_2) \xrightarrow{\gamma_2} (X_3, \subseteq_3)\), then
\((X_1, \subseteq_1) \xrightarrow{\gamma_1 \circ \gamma_2} (X_3, \subseteq_3)\).

Proof: \((\alpha_2 \circ \alpha_1)(c) \subseteq_3 a \iff \alpha_1(c) \subseteq_2 \gamma_2(a) \iff c \subseteq_1 (\gamma_1 \circ \gamma_2)(a)\)
Parity domain

\[\mathcal{P}(\mathbb{Z})\]

\[\gamma_p(\bot) \overset{\text{def}}{=} \emptyset\]
\[\gamma_p(\text{even}) \overset{\text{def}}{=} 2\mathbb{Z}\]
\[\gamma_p(\text{odd}) \overset{\text{def}}{=} 2\mathbb{Z} + 1\]
\[\gamma_p(\top) \overset{\text{def}}{=} \mathbb{Z}\]

\[\alpha_p(S) \overset{\text{def}}{=} \begin{cases} 
\bot & \text{if } S = \emptyset \\
\text{even} & \text{else, if } S \subseteq 2\mathbb{Z} \\
\text{odd} & \text{else, if } S \subseteq 2\mathbb{Z} + 1 \\
\top & \text{otherwise}
\end{cases}\]
Part of the abstraction lattice for $\mathcal{P}(\mathbb{Z})$

- Signs are more abstract than intervals; every sign information can be expressed as an interval.
- Parities and intervals are incomparable; no common property, except $\perp$ and $\top$.
- $\mathcal{P}(\mathbb{Z})$ is the most concrete domain.
- $\{\top\}$ is the most abstract domain.
- $\text{intervals} \land \text{parities}$ is the coarsest abstract domain more precise than intervals and parities.
Reduced product construction
Simple product

Algebraic structure:

Given two domains \((\mathcal{D}_1^\# , \sqsubseteq_1)\) and \((\mathcal{D}_2^\# , \sqsubseteq_2)\), we use pairs of abstract elements to represent conjunctions of properties.

- \(\mathcal{D}_1^\# \times_2 \triangleq \mathcal{D}_1^\# \times \mathcal{D}_2^\#\)
- \(\gamma_1 \times_2 (X_1^\#, X_2^\#) \triangleq \gamma_1(X_1^\#) \cap \gamma_2(X_2^\#)\)
- \(\alpha_1 \times_2 (S) \triangleq (\alpha_1(S), \alpha_2(S))\)
- \((X_1^\#, X_2^\#) \sqsubseteq_1 \times_2 (Y_1^\#, Y_2^\#) \iff X_1^\# \sqsubseteq_1 Y_1^\# \text{ and } X_2^\# \sqsubseteq_2 Y_2^\#\)

Abstract operators in \(\mathcal{D}^\#\):

Applied in parallel (independently) in each abstract domain:

- \((X_1^\#, X_2^\#) \cup_1 \times_2 (Y_1^\#, Y_2^\#) \triangleq (X_1^\# \cup_1 Y_1^\#, X_2^\# \cup_2 Y_2^\#), ;\)
- \((X_1^\#, X_2^\#) \triangledown_1 \times_2 (Y_1^\#, Y_2^\#) \triangleq (X_1^\# \triangledown_1 Y_1^\#, X_2^\# \triangledown_2 Y_2^\#);\)
- \(S^\#[s]_1 \times_2 (X_1^\#, X_2^\#) \triangleq (S^\#[s]_1(X_1^\#), S^\#[s]_2(X_2^\#)).\)
Reduced products

Simple products: limitations

\[
V \leftarrow 1;
\]
\[
\text{while } V \leq 10 \text{ do } V \leftarrow V + 2 \text{ done};
\]
\[
\bullet \text{ if } V \geq 12 \text{ then } \bullet V \leftarrow 0 \bullet;
\]

Analysis in the product domain of intervals and parities:

<table>
<thead>
<tr>
<th>intervals</th>
<th>parities</th>
<th>product: intervals (\times) parities</th>
</tr>
</thead>
<tbody>
<tr>
<td>(V \in [11, 12])</td>
<td>(V) odd</td>
<td>((V \in [11, 12]) \land (V) odd)</td>
</tr>
<tr>
<td>(V = 12)</td>
<td>(V) odd</td>
<td>((V = 12) \land (V) odd)</td>
</tr>
<tr>
<td>(V = 0)</td>
<td>(V) even</td>
<td>((V = 0) \land (V) even)</td>
</tr>
</tbody>
</table>

Identical to two separate analyses:

- at \(\bullet\), we get \((V = 12) \land (V\) odd\), which represents \(\emptyset\);
- at \(\bullet\), we apply \(V \leftarrow 0\) independently on intervals and parities, which gives \((V = 0) \land (V\) even\), instead of \(\emptyset\)!

\(\implies\) huge loss of precision
Fully reduced product

Idea: propagate information between domains

Given Galois connections \((\alpha_1, \gamma_1)\) and \((\alpha_2, \gamma_2)\) over \(D_1^\#\) and \(D_2^\#\), we define a reduction operator \(\rho\) as:

\[\rho : D_1^\# \times D_2^\# \rightarrow D_1^\# \times D_2^\#\]

\[\rho(X_1^\#, X_2^\#) \overset{\text{def}}{=} (\alpha_1(\gamma_1(X_1^\#) \cap \gamma_2(X_2^\#)), \alpha_2(\gamma_1(X_1^\#) \cap \gamma_2(X_2^\#)))\]

i.e., the best representation of \(\gamma_1 \times_2(X_1^\#, X_2^\#)\) in both domains

Application:
use \(\rho\) to transfer information between domains after abstract operations:

- \((X_1^\#, X_2^\#) \cup_{1 \times 2} (Y_1^\#, Y_2^\#) \overset{\text{def}}{=} \rho(X_1^\# \cup_1 Y_1^\#, X_2^\# \cup_2 Y_2^\#),\)
- \(S^\#[[s]]_{1 \times 2}(X_1^\#, X_2^\#) \overset{\text{def}}{=} \rho(S^\#[[s]]_1(X_1^\#), S^\#[[s]]_2(X_2^\#)).\)

Warning:
\(\rho\) should not be used on fixpoint iterates with widening \(\triangledown(X_{n+1} \overset{\text{def}}{=} \rho(X_n \triangledown F(X_n)))\)
\(\implies\) this could prevent the convergence in \(D_1^\# \times D_2^\#\)!
Analysis with reduction: exemple

\[
\begin{align*}
V & \leftarrow 1; \\
\textbf{while } V \leq 10 & \textbf{ do } V \leftarrow V + 2 \textbf{ done; } \\
\bullet & \textbf{ if } V \geq 12 \textbf{ then } V \leftarrow 0
\end{align*}
\]

Reduction \( \rho \) between intervals and parities \( \rho([a, b], p) \):
First refine interval bounds \([a, b]\) using parity information \(p\), then refine the parity information using the refined bounds:

- let \( a' = a + 1 \) if \( a \not\in \gamma_p(p) \), \( a' = a \) otherwise;
- let \( b' = b - 1 \) if \( b \not\in \gamma_p(p) \), \( b' = b \) otherwise;
- if \( a' > b' \), return \((\bot, \bot)\);
- if \( a' = b' \), return \(([a', b'], \alpha_p(a))\);
- otherwise, return \(([a', b'], p)\).

Example:
At \( \bullet \), \( \rho([11, 12], \text{odd}) = ([11, 11], \text{odd}) \)
\( \implies \) the “then” branch is not reachable.
Partial reduction

The optimal reduction $\rho$ is well-defined but:

- $\rho$ assumes we have Galois connections;
- there is no general effective algorithm to compute $\rho$.

(similar to the case of optimal operators, defined as $F^\# \overset{\text{def}}{=} \alpha \circ F \circ \gamma$)

**Partial reduction:**

Practical definition, when the optimal reduction is not available:

- $\rho(X_1^\#, X_2^\#) = (Y_1^\#, Y_2^\#)$ is a **partial reduction** if:
  - $Y_1^\# \sqsubseteq_1 X_1^\#$ and $Y_2^\# \sqsubseteq_2 X_2^\#$ (improvement)
  - $\gamma_{1 \times 2}(Y_1^\#, Y_2^\#) = \gamma_{1 \times 2}(X_1^\#, X_2^\#)$ (soundness)

**Example:**

$$
\rho(X_1^\#, X_2^\#) \overset{\text{def}}{=} \begin{cases} 
(\bot_1, \bot_2) & \text{if } X_1^\# = \bot_1 \text{ or } X_2^\# = \bot_2 \\
(X_1^\#, X_2^\#) & \text{otherwise}
\end{cases}
$$

In practice, an analyzer contains many abstract domains (for expressiveness) with limited reductions between them (for efficiency).
Disjunctive domains
**Motivation**

**Remark:** most domains abstract *convex sets* (conjunctions of constraints)

$$\Rightarrow \cup^\# \text{ causes a loss of precision!}$$

---

The need for non-convex invariants

```plaintext
X ← rand(10, 20);
Y ← rand(0, 1);
if Y > 0 then X ← −X;
• Z ← 100/X
```

**Concrete semantics:**

At $\bullet$, $X \in [-20, -10] \cup [10, 20]$

$$\Rightarrow \text{there is no division by zero}$$

**Abstract analysis:**

Convex analyses (intervals, polyhedra) will find $X \in [-20, 20]$

(with intervals, $[-20, -10] \cup^\# [10, 20] = [-20, 20]$)

$$\Rightarrow \text{possible division by zero}$$

 fals alarm}
Principle:
generic constructions to lift any numeric abstract domain
to a domain able to represent disjunctions exactly

Example constructions:
- powerset completion
  unordered “soup” of abstract elements
- state partitioning
  abstract elements keyed to selected subsets of environments
- path-sensitive analyses
  partition with respect to the history of execution

each construction has its strength and weakness
they can be combined during an analysis to exploit the best of each
Powerset completion
Disjunctive domains

Powerset completion

Given: \((\mathcal{E}^\#, \subseteq, \gamma, \cup^#, \cap^#, \triangledown, S^#[ stat ]\) )

abstract domain \(\mathcal{E}^\#\)

ordered by \(\subseteq\), which also acts as a sound abstraction of \(\subseteq\) (i.e., \(\subseteq^\# = \subseteq\))

with concretization \(\gamma: \mathcal{E}^\# \rightarrow \mathcal{P}(\mathcal{E})\)

sound abstractions \(\cup^#, \cap^#, S^#[ stat ]\) of \(\cup, \cap, S[ stat ]\), and a widening \(\triangledown\)

Construct: \((\hat{\mathcal{E}}^\#, \hat{\subseteq}, \hat{\gamma}, \hat{\cup^#}, \hat{\cap^#}, \hat{\triangledown}, \hat{S}^#[ stat ]\) )

- \(\hat{\mathcal{E}}^\# \overset{\text{def}}{=} P_{\text{finite}}(\mathcal{E}^\#)\) (finite sets of abstract elements)
- \(\hat{\gamma}(A^#) \overset{\text{def}}{=} \cup \{ \gamma(X^#) \mid X^# \in A^# \} \) (join of concretizations)

Example: using the interval domain as \(\mathcal{E}^\#\)

\(\hat{\gamma}(\{[-10, -5], [2, 4], [0, 0], [2, 3]\}) = [-10, -5] \cup \{0\} \cup [2, 4]\)
Ordering

**Issue:** how can we compare two elements of $\hat{E}^#$?

- $\hat{\gamma}$ is generally not injective
  - there is no canonical representation for $\hat{\gamma}(A^#)$
- testing $\hat{\gamma}(A^#) = \hat{\gamma}(B^#)$ or $\hat{\gamma}(A^#) \subseteq \hat{\gamma}(B^#)$ is difficult

**Example:** powerset completion of the interval domain

- $A^# = \{\{0\} \times \{0\}, [0, 1] \times \{1\}\}$
- $B^# = \{\{0\} \times \{0\}, \{0\} \times \{1\}, \{1\} \times \{1\}\}$
- $C^# = \{\{0\} \times [0, 1], [0, 1] \times \{1\}\}$
- $\hat{\gamma}(A^#) = \hat{\gamma}(B^#) = \hat{\gamma}(C^#)$

$B^#$ is more costly to represent: it requires three abstract elements instead of two
$C^#$ is a covering and not a partition (red $\cap$ blue $= \{0\} \times \{1\} \neq \emptyset$)
Solution: sound approximation of $\subseteq$

\[ A^\# \hat{\subseteq} B^\# \iff \forall X^\# \in A^\#: \exists Y^\# \in B^\#: X^\# \subseteq Y^\# \] (Hoare powerdomain order)

- $\hat{\subseteq}$ is a partial order (when $\subseteq$ is)
- $\hat{\subseteq}$ is a sound approximation of $\subseteq$ (when $\subseteq$ is)
- $A^\# \hat{\subseteq} B^\# \implies \hat{\gamma}(A^\#) \subseteq \hat{\gamma}(B^\#)$ but the converse may not hold
- testing $\hat{\subseteq}$ reduces to testing $\subseteq$ finitely many times

Example: powerset completion of the interval domain

\[ \hat{\gamma}(A^\#) = \hat{\gamma}(B^\#) = \hat{\gamma}(C^\#) \]

\[ B^\# \hat{\subseteq} A^\# \hat{\subseteq} C^\# \]
Abstract operators

- \( \hat{S}[^{\text{stat}}] A^\# \overset{\text{def}}{=} \{ S[^{\text{stat}}] X^\# \mid X^\# \in A^\# \} \)
  - apply \text{stat} on each abstract element independently

- \( A^\# \hat{\cup}^\# B^\# \overset{\text{def}}{=} A^\# \cup B^\#
  - keep elements from both arguments without applying any abstract operation
  - \( \hat{\cup}^\# \) is exact

- \( A^\# \hat{\cap}^\# B^\# \overset{\text{def}}{=} \{ X^\# \cap^\# Y^\# \mid X^\# \in A^\#, Y^\# \in B^\# \} \)
  - \( \hat{\cap}^\# \) is exact if \( \cap^\# \) is (as \( \cup \) and \( \cap \) are distributive)

Galois connection:

in general, there is no abstraction function \( \hat{\alpha} \) corresponding to \( \hat{\gamma} \)

Example: powerset completion \( \hat{\mathcal{E}}^\# \) of the interval domain \( \mathcal{E}^\# \)
given the disc \( S \overset{\text{def}}{=} \{ (x, y) \mid x^2 + y^2 \leq 1 \} \)
\( \alpha(S) = [-1, 1] \times [-1, 1] \) (optimal interval abstraction)
but there is no best abstraction in \( \hat{\mathcal{E}}^\# \)
Dynamic approximation

**Issue:** the size $|A^#|$ of elements $A^# \in \hat{E}^#$ is unbounded

- every application of $\hat{\cup}^#$ adds some more elements
- $\implies$ efficiency and convergence problems

**Solution:** to reduce the size of elements

- redundancy removal

  \[
  simplify(A^#) \overset{\text{def}}{=} \{ X^# \in A^# \mid \forall Y^# \neq X^# \in A^# : X^# \not\subseteq Y^# \}
  \]
  
  no loss of precision: $\gamma(simplify(A^#)) = \gamma(A^#)$

- collapse: join elements in $E^#$

  \[
  collapse(A^#) \overset{\text{def}}{=} \{ \hat{\cup}^# \{ X^# \in A^# \} \}
  \]

  large loss of precision, but very effective: $|collapse(A^#)| = 1$

- partial collapse: limit $|A^#|$ to a fixed size $k$ by $\hat{\cup}^#$

  but how to choose which elements to merge? no easy solution!
**Widening**

**Issue:** for loops, abstract iterations \((A_n^\#)_{n \in \mathbb{N}}\) may not converge

- the size of \(A_n^\#\) may grow arbitrarily large
- even if \(|A_n^\#|\) is stable, some elements in \(A_n^\#\) may not converge
  if \(E^\#\) has infinite increasing sequences

\[ \Rightarrow \text{we need a widening} \hat{\nabla} \]

Widenings for powerset domains are difficult to design

**Example widening:** collapse after a fixed number \(N\) of iterations

\[
A_{n+1}^\# \overset{\text{def}}{=} A_n^\# \hat{\nabla} B_n^\# \overset{\text{def}}{=} \begin{cases} 
\text{simplify}(A_n^\# \cup^\# B_n^\#) & \text{if } n < N \\
\text{collapse}(A_n^\#) \nabla \text{collapse}(B_n^\#) & \text{otherwise}
\end{cases}
\]

(this is very naïve, see Bagnara et al. STTT06 for more interesting widenings)
State partitioning
State partitioning

**Principle:**
- partition *a priori* $\mathcal{E}$ into finitely many sets
- abstract each partition of $\mathcal{E}$ independently using an element of $\mathcal{E}^\#$

**Abstract domain:**

Given an abstract partition $P^\# \subseteq \mathcal{E}^\#$, i.e., a set such that:

- $P^\#$ is finite
- $\bigcup \{ \gamma(X^\#) \mid X^\# \in P^\# \} = \mathcal{E}$

for generality, we have in fact a covering, not a partitioning of $\mathcal{E}$ i.e., we can have $X^\# \neq Y^\# \in P^\#$ with $\gamma(X^\#) \cap \gamma(Y^\#) \neq \emptyset$

We define $\tilde{\mathcal{E}}^\# \overset{\text{def}}{=} P^\# \rightarrow \mathcal{E}^\#$ representable in memory, as $P^\#$ is finite
Example: \( \mathcal{E}^\# \) is the interval domain

\[
P^\# = \{ P_1, P_2, P_3, P_4, P_5 \}
\]

where

\[
P_1 = [-\infty, 0] \times [-\infty, +\infty]
\]

\[
P_2 = [0, 10] \times [0, +\infty]
\]

\[
P_3 = [0, 10] \times [-\infty, 0]
\]

\[
P_4 = [10, +\infty] \times [0, +\infty]
\]

\[
P_5 = [10, +\infty] \times [-\infty, 0]
\]

\[
X^\# = [P_1 \mapsto [-6, -5] \times [5, 6], 
P_2 \mapsto \bot,
P_3 \mapsto [9, 10] \times [-\infty, -1],
P_4 \mapsto \bot,
P_5 \mapsto [10, 12] \times [-3, -1]]
\]

\[
\mathcal{E}^\# \overset{\text{def}}{=} P^\# \rightarrow \mathcal{E}^\#
\]

\[
\tilde{\gamma}(A^\#) \overset{\text{def}}{=} \bigcup \{ \gamma(A^\#(X^\#)) \cap \gamma(X^\#) \mid X^\# \in P^\# \}
\]

\[
A^\# \sqsubseteq B^\# \overset{\text{def}}{\iff} \forall X^\# \in P^\#: A^\#(X^\#) \subseteq B^\#(X^\#)
\]  
(point-wise order)

\[
\tilde{\alpha}(S) \overset{\text{def}}{=} \lambda X^\# \in P^\#. \alpha(S \cap \gamma(X^\#))
\]

if \( \mathcal{E}^\# \) enjoys a Galois connection, so does \( \tilde{\mathcal{E}}^\# \)
Abstract operators

**Abstract operators:** point-wise extension from \( \mathcal{E}^\# \) to \( P^\# \rightarrow \mathcal{E}^\# \)

- \( A \cup^\# B \overset{\text{def}}{=} \lambda X^\# \in P^\#.A(X^#) \cup^# B(X^#) \)
- \( A \cap^\# B \overset{\text{def}}{=} \lambda X^\# \in P^\#.A(X^#) \cap^# B(X^#) \)
- \( A \triangledown B \overset{\text{def}}{=} \lambda X^\# \in P^\#.A(X^#) \triangledown B(X^#) \)
- \( \tilde{S}^\#[ e \leq 0?] A^# \overset{\text{def}}{=} \lambda X^\# \in P^#.S^#[ e \leq 0?] A^#(X^#) \)
- \( \tilde{S}^#[ V \leftarrow e ] A^# \) is more complex

any \( S^#[ V \leftarrow e ] A^#(X^#) \) may escape its partition \( X^# \); we must cut them at partition borders and glue the pieces falling into the same partition

example: \( X \leftarrow X + 2 \)

\[
\tilde{S}^#[ V \leftarrow e ] A^# \overset{\text{def}}{=} \lambda X^#. \cup^# \{ X^# \cap^# S^#[ V \leftarrow e ] A(Y^#) \mid Y^# \in P^# \}
\]
Example analysis

Example

\[
\begin{align*}
X & \leftarrow \text{rand}(10, 20); \\
Y & \leftarrow \text{rand}(0, 1); \\
\text{if } Y > 0 \text{ then } X & \leftarrow -X; \\
\bullet \quad Z & \leftarrow 100/X
\end{align*}
\]

Analysis:

\bullet \ \mathcal{E}^\# \text{ is the interval domain}

\bullet \ \text{partition with respect to the sign of } X

\begin{align*}
P^\# & \overset{\text{def}}{=} \{ X^+, X^- \} \text{ where} \\
X^+ & \overset{\text{def}}{=} [0, +\infty] \times \mathbb{Z} \times \mathbb{Z} \text{ and } X^- \overset{\text{def}}{=} [\neg \infty, 0] \times \mathbb{Z} \times \mathbb{Z}
\end{align*}

\bullet \ \text{at } \bullet \ \text{we find:}

\begin{align*}
X^+ & \mapsto [X \in [10, 20], Y \mapsto [0, 0], Z \mapsto [0, 0]] \\
X^- & \mapsto [X \in [-20, -10], Y \mapsto [1, 1], Z \mapsto [0, 0]]
\end{align*}

\implies \text{no division by zero}
Path partitioning
Path sensitivity

**Principle:** partition wrt. the history of computation

- keep different abstract elements for different execution paths
  e.g., different branches taken, different loop iterations
- avoid merging with $\bigcup^*$ elements at control-flow joins
  at the end of `if · · · then · · · else`, or at loop head

**Intuition:** as a program transformation

\[
X \leftarrow \text{rand}(-50, 50);
\]

\[
\text{if } X \geq 0 \text{ then}
Y \leftarrow X + 10
\]

\[
\text{else}
Y \leftarrow X - 10;
\]

\[
\text{assert } Y \neq 0
\]

\[
X \leftarrow \text{rand}(-50, 50);
\]

\[
\text{if } X \geq 0 \text{ then}
Y \leftarrow X + 10;
\]

\[
\text{assert } Y \neq 0
\]

\[
\text{else}
Y \leftarrow X - 10;
\]

\[
\text{assert } Y \neq 0
\]

the `assert` is tested in the context of each branch
instead of after the control-flow join
the interval domain can prove the assertion on the right, but not on the left
Abstract domain

Formalization: we consider here only if · · · then · · · else

- \( L \) denote syntactic labels of if · · · then · · · else instructions
- history abstraction \( H \overset{\text{def}}{=} L \rightarrow \{\text{true}, \text{false}, \bot\} \)

\( H \in H \) indicates the outcome of the last time we executed each test:
- \( H(\ell) = \text{true} \): we took the then branch
- \( H(\ell) = \text{false} \): we took the else branch
- \( H(\ell) = \bot \): we never executed the test

Notes:
- \( H \) can remember the outcome of several successive tests
  \( \ell_1 : \text{if} \cdots \text{then} \cdots \text{else}; \ell_2 : \text{if} \cdots \text{then} \cdots \text{else} \)
- for tests in loops, \( H \) remembers only the last outcome
  while \( \cdots \) do \( \ell : \text{if} \cdots \text{then} \cdots \text{else} \)
- we could extend \( H \) to longer histories with \( H = (L \rightarrow \{\text{true}, \text{false}, \bot\})^* \)
- we could extend \( H \) to track loop iterations with \( H = L \rightarrow \mathbb{N} \)

- \( \tilde{E}^\# \overset{\text{def}}{=} H \rightarrow E^\# \)

use a different abstract element for each abstract history
Abstract operators

- \( \tilde{\mathcal{E}}^\# \overset{\text{def}}{=} \mathbb{H} \to \mathcal{E}^\# \)
- \( \tilde{\gamma}(A^\#) = \bigcup \{ \gamma(A^\#(H)) \mid H \in \mathbb{H} \} \)
- \( \sqsubseteq, \sqcup^\#, \sqcap^\#, \sqtriangleset \) are point-wise
- \( \hat{\mathcal{S}}^\#[ V \leftarrow e ] \) and \( \hat{\mathcal{S}}^\#[ e \leq 0? ] \) are point-wise
- \( \tilde{\mathcal{S}}^\#[ \ell : \text{if } c \text{ then } s_1 \text{ else } s_2 ] \) \( A^\# \) is more complex
  - we merge all information about \( \ell \)
    \[ C^\# = \lambda H. A^\#(H[\ell \mapsto \text{true}]) \cup^\# A^\#(H[\ell \mapsto \text{false}]) \cup^\# A^\#(H[\ell \mapsto \bot]) \]
  - we compute the then branch, where \( H(\ell) = \text{true} \)
    \[ T'\# = \hat{\mathcal{S}}^\#[ s_1 ] (\hat{\mathcal{S}}^\#[ c? ] T\#) \]
    \[ T\# = \lambda H. C^\#(H) \text{ if } H(\ell) = \text{true}, \bot \text{ otherwise} \]
  - we compute the else branch, where \( H(\ell) = \text{false} \)
    \[ F'\# = \hat{\mathcal{S}}^\#[ s_2 ] (\hat{\mathcal{S}}^\#[ \neg c? ] F\#) \]
    \[ F\# = \lambda H. C^\#(H) \text{ if } H(\ell) = \text{false}, \bot \text{ otherwise} \]
  - we join both branches: \( T'\# \sqcup^\# F'\# \)
    the join is exact as \( \forall H \in \mathbb{H} : \) either \( T'\#(H) = \bot \) or \( F'\#(H) = \bot \)

\( \implies \) we get a semantic by induction on the syntax of the original program
Complex example

Concrete semantics: table-based interpolation based on the value of $X$

- Look-up index $I$ in the interpolation table: $TX[I] \leq X \leq TX[I + 1]$
- Interpolate from value $TY[I]$ when $X = TX[I]$ with slope $TS[I]$

Analysis: in the interval domain

- Without partitioning:
  $$Y \in [\min TY, \max TY] + (X - [\min TX, \max TX]) \times [\min TS, \max TS]$$
- Partitioning with respect to the number of loop iterations:
  $$Y \in \bigcup_{I \in [0,N]} TY[I] + ([0, TX[I + 1] - TX[I]] \times TS[I]$$
  More precise as it keeps the relation between table indices
Inter-procedural analyses
Overview

- **Analysis on the control-flow graph**
  - reduce function calls and returns to *gotos*
  - useful for the project!

- **Inlining**
  - simple and precise
  - but not efficient and may not terminate

- **Call-site and call-stack abstraction**
  - terminates even for recursive programs
  - parametric cost-precision trade-off

- **Tabulated abstraction**
  - optimal reuse of analysis partial results

- We also mentioned summary-based abstractions last week, leveraging relational domains for modular bottom-up analysis

In general, these different abstractions give incomparable results; there is no clear winner
Analysis on the control-flow graph
Inter-procedural control-flow graphs

Extend control-flow graphs:

- one subgraph for each function
- additional arcs to denote function calls and returns

we get one big graph without procedures nor calls, only gotos

\[\Rightarrow\] reduced to a classic analysis based on equation systems
but difficult to use in a denotational-style analysis by induction on the syntax

**Note:** to simplify, we assume here no local variable and no function argument:

- locals and arguments are transformed into globals
- only possible if there are no recursive calls
## Example: Control-flow graph

<table>
<thead>
<tr>
<th>Function</th>
<th>Code</th>
</tr>
</thead>
</table>
| **main** | \( R \leftarrow -1; \)  
\( X \leftarrow \text{rand}(5, 10); f(); \)  
\( X \leftarrow 80; f() \) |
| **f**    | \( R \leftarrow 2 \times X; \)  
if \( R > 100 \) then \( R \leftarrow 0 \) |

Create one control-flow graph for each function.
Inter-procedural analyses
Analysis on the control-flow graph

Example: Control-flow graph

```
main:
R ← −1;
X ← rand(5, 10); f();
X ← 80; f()

f:
R ← 2 × X;
if R > 100 then R ← 0
```

Example

```
main:
R = −1
X = rand(5, 10)
f()
X = 80
f()

f:
R = 2X
R > 100
R = 0
R ≤ 100
```

replace call instructions with gotos
Example: Equation system

- each variable \( S_i \) denotes a set of environments at a control location \( i \)
- we can derive an abstract version of the system
  e.g.: \( S_f^\#_2 = S_f^\#[ R \leftarrow 2X ] S_f^\#_1, \ S_f^\#_6 = S_f^\#_4 \cup^\# S_f^\#_5, \) etc.
- we can solve the abstract system, using widenings to terminate
  c.f. project

\[
\begin{align*}
S_{\text{main},1} &= \top \\
S_{\text{main},2} &= S[R \leftarrow 1] S_{\text{main},1} \\
S_{\text{main},3} &= S[X \leftarrow \text{rand}(5, 10)] S_{\text{main},2} \\
S_{\text{main},4} &= S_f,6 \\
S_{\text{main},5} &= S[X \leftarrow 80] S_{\text{main},4} \\
S_{\text{main},6} &= S_f,6 \\
S_{f,1} &= S_{\text{main},3} \cup S_{\text{main},5} \\
S_{f,2} &= S[R \leftarrow 2X] S_{f,1} \\
S_{f,3} &= S[R > 100] S_{f,2} \\
S_{f,4} &= S[R \leftarrow 0] S_{f,3} \\
S_{f,5} &= S[R \leq 100] S_{f,2} \\
S_{f,6} &= S_{f,4} \cup S_{f,5}
\end{align*}
\]
Example: Equation system

using intervals we get the following solution:

\[
\begin{align*}
S_{\text{main},1} & : X, R \in \mathbb{Z} \\
S_{\text{main},2} & : X \in \mathbb{Z}, R = -1 \\
S_{\text{main},3} & : X \in [5, 10], R = -1 \\
S_{\text{main},4} & : X \in [5, 80], R \in [0, 100] \\
S_{\text{main},5} & : X = 80, R \in [0, 100] \\
S_{\text{main},6} & : X \in [5, 80], R \in [0, 100]
\end{align*}
\]

\[
\begin{align*}
S_{f,1} & : X \in [5, 80], R \in [-1, 100] \\
S_{f,2} & : X \in [5, 80], R \in [10, 160] \\
S_{f,3} & : X \in [5, 80], R \in [101, 160] \\
S_{f,4} & : X \in [5, 80], R = 0 \\
S_{f,5} & : X \in [5, 80], R \in [10, 100] \\
S_{f,6} & : X \in [5, 80], R \in [0, 100]
\end{align*}
\]
Imprecision

In fact, in our example, $R = 0$ holds at the end of the program but we find $R \in [0, 100]$! $\implies$ the analysis is imprecise

**Explanation:** the control-flow graph adds impossible executions paths
General case: concrete semantics
Procedures

**Syntax:**

- $\mathcal{F}$ finite set of procedure names
- $body : \mathcal{F} \rightarrow stat$: procedure bodies
- $main \in stat$: entry point body
- $\mathcal{V}_G$: set of **global** variables
- $\mathcal{V}_f$: set of **local** variables for procedure $f \in \mathcal{F}$
  - procedure $f$ can only access $\mathcal{V}_f \cup \mathcal{V}_G$
  - $main$ has no local variable and can only access $\mathcal{V}_G$
- $stat ::= f(\text{expr}_1, \ldots, \text{expr}_{|\mathcal{V}_f|}) | \cdots$

**procedure call**, $f \in \mathcal{F}$, setting all its local variables

  - local variables double as **procedure arguments**
  - no special mechanism to return a value (a global variable can be used)
Concrete environments

Notes:

- when \( f \) calls \( g \), we must remember the value of \( f \)'s locals \( \forall_f \) in the semantics of \( g \) and restore them when returning
- several copies of each \( V \in \forall_f \) may exist at a given time due to recursive calls, i.e.: cycles in the call graph

\[ \implies \text{concrete environments use per-variable stacks} \]

Stacks:

\[ S \equiv \mathbb{Z}^* \text{ (finite sequences of integers)} \]

- \( \text{push}(v, s) \equiv v \cdot s \) (\( v, v' \in \mathbb{Z}, s, s' \in S \))
- \( \text{pop}(s) \equiv s' \text{ when } \exists v: s = v \cdot s' \), undefined otherwise
- \( \text{peek}(s) \equiv v \text{ when } \exists s': s = v \cdot s' \), undefined otherwise
- \( \text{set}(v, s) \equiv v \cdot s' \text{ when } \exists v': s = v' \cdot s' \), undefined otherwise

Environments:

\[ E \equiv (\bigcup_{f \in F} \forall_f \cup \forall_G) \to S \]

for \( \forall_G \), stacks are not necessary but simplify the presentation

traditionally, there is a single global stack for all local variables using per-variable stacks instead also makes the presentation simpler
Concrete semantics:

on $\mathcal{E} \overset{\text{def}}{=} (\bigcup_{f \in F} \mathcal{V}_f \cup \mathcal{V}_G) \rightarrow S$

variable reads and updates only consider the top of the stack; procedure calls push and pop local variables

- $E[V] \rho \overset{\text{def}}{=} \text{peek}(\rho(V))$
- $S[V \leftarrow e] R \overset{\text{def}}{=} \{ \rho[V \mapsto \text{set}(x, \rho(V))] \mid \rho \in R, x \in E[e] \rho \}$
- $S[f(e_{V_1}, \ldots, e_{V_n})] R = R_3$, where:
  - $R_1 \overset{\text{def}}{=} \{ \rho[\forall V \in \mathcal{V}_f : V \mapsto \text{push}(x_V, \rho(V))] \mid \rho \in R, \forall V \in \mathcal{V}_f : x_V \in E[e_V] \rho \}$
    (evaluate each argument $e_V$ and push its value $x_V$ on the stack $\rho(V)$)
  - $R_2 \overset{\text{def}}{=} S[\text{body}(f)] R_1$
    (evaluate the procedure body)
  - $R_3 \overset{\text{def}}{=} \{ \rho[\forall V \in \mathcal{V}_f : V \mapsto \text{pop}(\rho(V))] \mid \rho \in R_2 \}$
    (pop local variables)
- initial environment: $\rho_0 \overset{\text{def}}{=} \lambda V \in \mathcal{V}_G.0$

other statements are unchanged
Semantic inlining
Semantic inlining

**Naïve abstract procedure call:** mimic the concrete semantics

- **assign abstract variables** to stack positions:
  \[ \mathcal{V}^\# \overset{\text{def}}{=} \mathcal{V}_G \cup \left( \bigcup_{f \in \mathcal{F}} \mathcal{V}_f \times \mathbb{N} \right) \]
  \( \mathcal{V}^\# \) is infinite, but each abstract environment uses finitely many variables

- **\( \mathcal{E}_\mathcal{V}^\# \)** abstracts \( \mathcal{P}(\mathcal{V} \rightarrow \mathbb{Z}) \), for any finite \( \mathcal{V} \subseteq \mathcal{V}^\# \)
  - \( \mathcal{V} \in \mathcal{V}_f \) denotes \( (\mathcal{V}, 0) \) in \( \mathcal{V}^\# \)
  - **push** \( \mathcal{V} \): shift variables, replacing \( (\mathcal{V}, i) \) with \( (\mathcal{V}, i+1) \), then add \( (\mathcal{V}, 0) \)
  - **pop** \( \mathcal{V} \): remove \( (\mathcal{V}, 0) \) and shift each \( (\mathcal{V}, i) \) to \( (\mathcal{V}, i-1) \)

- **\( S^\#[f(e_1, \ldots, e_n)] \mathcal{X}^\# \)** is then reduced to:
  \[
  \begin{align*}
  X_1^\# &= S^\#[\text{push} \ V_1; \ldots; \text{push} \ V_n] \mathcal{X}^\# \\
  X_2^\# &= S^\#[V_1 \leftarrow e_1; \ldots; V_n \leftarrow e_n] X_1^\# \\
  X_3^\# &= S^\#[\text{body}(f)] X_2^\# \\
  X_4^\# &= S^\#[\text{pop} \ V_1; \ldots; \text{pop} \ V_n] X_3^\#
  \end{align*}
  \]
  (add fresh variables for \( \mathcal{V}_f \))
  (bind arguments to locals)
  (execute the procedure body)
  (delete local variables)

**Limitations:**
- does not terminate in case of unbounded recursivity
- requires many abstract variables to represent the stacks
- procedures must be re-analyzed for every call
  full context-sensitivity: precise but costly
Example

**Example**

```plaintext
main :
R ← -1;
f(rand(5, 10));
f(80)
```

```plaintext
f(X) :
R ← 2 × X;
if R > 100 then R ← 0
```

**Analysis using intervals**

- after the first call to `f`, we get $R \in [10, 20]$
- after the second call to `f`, we get $R = 0$
Call-site abstraction
Call-site abstraction

Abstracting stacks: into a fixed, bounded set $\mathbb{V}^\#$ of variables

- $\mathbb{V}^\# \overset{\text{def}}{=} \bigcup_{f \in \mathcal{F}} \{ \mathbb{V}, \hat{\mathbb{V}} \mid \mathbb{V} \in \mathbb{V}_f \} \cup \mathbb{V}_G$
  - two copies of each local variable
    - $\mathbb{V}$ abstracts the value at the top of the stack (current call)
    - $\hat{\mathbb{V}}$ abstracts the rest of the stack

- $S^\#[\text{push } \mathbb{V}] X^\# \overset{\text{def}}{=} X^\# \cup^\# S^\#[\hat{\mathbb{V}} \leftarrow \mathbb{V}] X^\#$
- $S^\#[\text{pop } \mathbb{V}] X^\# \overset{\text{def}}{=} X^\# \cup^\# S^\#[\mathbb{V} \leftarrow \hat{\mathbb{V}}] X^\#$
  - weak updates, similar to array manipulation
  - no need to create and delete variables dynamically

- assignments and tests always access $\mathbb{V}$, not $\hat{\mathbb{V}}$
  $\implies$ strong update (precise)

Note: when there is no recursivity, $\hat{\mathbb{V}}$, push and pop can be omitted
Call-site abstraction

**Principle:** merge all the contexts in which each function is called

- we maintain two global maps $\mathcal{F} \rightarrow \mathcal{E}^\#$:
  - $C^\#(f)$: abstracts the environments when calling $f$
  - $R^\#(f)$: abstracts the environments when returning from $f$

- gather environments from all possible calls to $f$, disregarding the call sites

- during the analysis, when encountering a call $S^\#[ body(f) ] X^\#$:
  - we return $R^\#(f)$
  - but we also replace $C^\#$ with $C^#[ f \mapsto C^\#(f) \cup^\# X^\# ]$

- $R^\#(f)$ is computed from $C^\#(f)$ as
  $$ R^\#(f) = S^\#[ body(f) ] (C^\#(f)) $$
Call-site abstraction

**Fixpoint:**

There may be circular dependencies between $C^\#$ and $R^\#$

e.g., in $f(2); f(3)$, the input for $f(3)$ depends on the output from $f(2)$

$\implies$ we compute a fixpoint for $C^\#$ by iteration:

- Initially, $\forall f: C^\#(f) = R^\#(f) = \bot$
- Analyze `main`
- While $\exists f: C^\#(f)$ not stable
  - Apply widening $\nabla$ to the iterates of $C^\#(f)$
  - Update $R^\#(f) = S^\#[ body(f) ] C^\#(f)$
  - Analyze `main` and all the procedures again
    (this may modify some $C^\#(g)$)

$\implies$ using $\nabla$, the analysis always terminates in finite time

We can be more efficient and avoid re-analyzing procedures when not needed

e.g., use a workset algorithm, track procedure dependencies, etc.
**Example**

**main**: 

\[ R ← −1; \]
\[ f(\text{rand}(5, 10)); \]
\[ f(80) \]

\[ f(X) : \]
\[ R ← 2 \times X; \]
\[ \text{if } R > 100 \text{ then } R ← 0 \]

**Analysis**: using intervals  (without widening as there is no dependency)

- first analysis of **main**: we get \( \bot \)  (as \( R^\sharp(f) = \bot \))
  but \( C^\sharp(f) = [R \mapsto [-1, -1], X \mapsto [5, 10]] \)

- first analysis of **f**: \( R^\sharp(f) = [R \mapsto [10, 20], X \mapsto [5, 10]] \)

- second analysis of **main**: we get
  \( C^\sharp(f) = [R \mapsto [-1, 20], X \mapsto [5, 80]] \)

- second analysis of **f**: \( R^\sharp(f) = [R \mapsto [0, 100], X \mapsto [5, 80]] \)

- final analysis of **main**, we find \( R \in [0, 100] \) at the program end
  less precise than \( R = 0 \) found by semantic inlining
**Variants:** $k$–limiting, $k$ is a constant

- **stack:**
  assign a distinct variable for the $k$ highest levels of $V$
  abstract the lower (unbounded) stack part with $\hat{V}$
  more precise than keeping only the top of the stack separately

- **context-sensitivity:**
  each syntactic call has a unique call-site $\ell \in \mathcal{L}$
  a call stack is a sequence of nested call sites: $c \in \mathcal{L}^*$
  an abstract call stack remembers the last $k$ call sites: $c^\# \in \mathcal{L}^k$
  the $C^\#$ and $R^\#$ maps now distinguish abstract call stacks
  $C^\#, R^\# : \mathcal{L}^k \rightarrow \mathcal{E}^\#$
  more precise than a partitioning by function only

larger $k$ give more precision but less efficiency
Example: context-sensitivity

**Example**

```plaintext
main:
  R ← −1;
  ℓ₁ : f(rand(5,10));
  ℓ₂ : f(80)

f(X):
  R ← 2 × X;
  if R > 100 then R ← 0
```

Analysis: using intervals and $k = 1$

- $C^#(ℓ₁) = [R \mapsto [-1,1], X \mapsto [5,10]]$
  $\implies R^#(ℓ₁) = [R \mapsto [10,20], X \mapsto [5,10]]$

- $C^#(ℓ₂) = [R \mapsto [10,20], X \mapsto [80,80]]$
  $\implies R^#(ℓ₂) = [R \mapsto [0,0], X \mapsto [80,80]]$

- at the end of the analysis, we get $R = 0$
  more precise than $R \in [0,100]$ found without context-sensitivity
Tabulation abstraction
Cardinal power

**Principle:**
the semantic of a function is $S[[\text{body}(f)]] : \mathcal{P}(\mathcal{E}) \rightarrow \mathcal{P}(\mathcal{E})$

$\Rightarrow$ abstract it as an abstract function in $\mathcal{E}^\# \rightarrow \mathcal{E}^\#$

we use a partial function as the image of most abstract elements is not useful

**Analysis:** tabulated analysis

- use a global partial map $F^\# : \mathcal{F} \times \mathcal{E}^\# \rightarrow \mathcal{E}^\#$
- $F^\#$ is initially empty, and is filled on-demand
- when encountering $S^\#[[\text{body}(f)]] X^\#$
  - return $F^#(f, X^\#)$ if defined
  - else, compute $S^\#[[\text{body}(f)]] X^\#$, store it in $F^#(f, X^\#)$ and return it

**Optimizations:** trade precision for efficiency

- if $X^\# \subseteq Y^\#$ and $F^#(f, X^\#)$ is not defined, we can use $F^#(f, Y^\#)$ instead
- if the size of $F^#$ grows too large, use $F^#(f, \top)$ instead
  - sound, and ensures that the analysis terminates in finite time
Example

Example

\[
\begin{align*}
\text{main :} & \quad f(X) : \\
R & \leftarrow -1; \\
f(\text{rand}(5, 10)); & \quad R \leftarrow 2 \times X; \\
f(80) & \quad \text{if } R > 100 \text{ then } R \leftarrow 0
\end{align*}
\]

Analysis using intervals

\begin{itemize}
\item \(F\# = \)
\begin{align*}
[ (f, [R \mapsto [-1, -1], X \mapsto [5, 10]]) & \mapsto [R \mapsto [10, 20], X \mapsto [5, 10]], \\
(f, [R \mapsto [10, 20], X \mapsto [80, 80]]) & \mapsto [R \mapsto [0, 0], X \mapsto [80, 80]] ]
\end{align*}
\end{itemize}

at the end of the analysis, we get again \(R = 0\)

here, the function partitioning gives the same result as the call-site partitioning
Dynamic partitioning: complex example

Example: McCarthy's 91 function

\[
\text{main} : \\
\quad \text{Mc} (\text{rand}(0, +\infty))
\]

\[
\text{Mc}(n) : \\
\quad \text{if } n > 100 \text{ then } r \leftarrow n - 10 \\
\quad \text{else } \text{Mc}(n + 11); \text{Mc}(r)
\]

- In the concrete, when terminating:
  \[ r = n - 10 \text{ when } n > 101, \text{ and } r = 91 \text{ wen } n \in [0, 101] \]

- Using a widening \( \nabla \) to choose tabulated abstract values \( F^{\#}(f, X^{\#}) \)
  
  we find:
  
  \[
  \begin{align*}
  & n \in [0, 72] \quad \Rightarrow \quad r = 91 \\
  & n \in [73, 90] \quad \Rightarrow \quad r \in [91, 101] \\
  & n \in [91, 101] \quad \Rightarrow \quad r = 91 \\
  & n \in [102, 111] \quad \Rightarrow \quad r \in [91, 101] \\
  & n \in [112, +\infty] \quad \Rightarrow \quad r \in [91, +\infty]
  \end{align*}
  \]

(source: Bourdoncle, JFP 1992)