Live Variable Analysis

A variable is *live* at the exit from a label if there exists a path from the label to a use of the variable that does not re-define the variable. The *Live Variables Analysis* will determine:

*For each program point, which variables may be live at the exit from the point.*

This analysis might be used as the basis for *Dead Code Elimination*. If the variable is not live at the exit from a label then, if the elementary block is an assignment to the variable, the elementary block can be eliminated.
Parity Analysis

A variable has *even* or *odd* parity at a label if we can guarantee that its value is *even* (e) or *odd* (o) for any execution of this label (not necessarily the same actual value). The *Parity Analysis* will determine:

> For each program point, what is the parity of each variable.

This analysis might be used as the basis for . . . (saving a bit?).

LV Analysis: Property Space

\[
\text{kill}_{LV} : \text{Block}_* \rightarrow \mathcal{P}(\text{Var}_*)
\]

\[
\text{gen}_{LV} : \text{Block}_* \rightarrow \mathcal{P}(\text{Var}_*)
\]

\[
\text{LV}_{\text{entry}} : \text{Lab}_* \rightarrow \mathcal{P}(\text{Var}_*)
\]

\[
\text{LV}_{\text{exit}} : \text{Lab}_* \rightarrow \mathcal{P}(\text{Var}_*)
\]

Important fact: Information we are interested in is in \(\mathcal{P}(\text{Var}_*)\).
Parity Information

The \textit{LV} Analysis associates to labels information – concretely the set of live variables, i.e. a set in $\mathcal{P}(\text{Var}_*)$. This is modified by local \textit{transfer functions} and \textit{collected} globally according to $\text{flow}$.

For \textit{Parity} we have identify the abstract properties to work with.

- Sets in $\mathcal{P}(\text{Var}_* \times \{e, o\})$ or maybe $\mathcal{P}(\text{Var}_* \times \{e, o, ?\})$, e.g. $\{(x, e), (x, o), (y, e)\} \equiv \{(x, ?), (y, e)\}$.
- Functions in $\text{Var}_* \rightarrow \{e, o\}$ or better $\text{Var}_* \rightarrow \{e, o, ?\}$. e.g. $\{x \mapsto ?, y \mapsto e\}$.
- represented as value tables, e.g. $\begin{array}{c|c} x & y \\ \hline ? & e \end{array}$

Questions: How to modify parity information locally and how to combine it, e.g. maybe $\{(x, e), (x, o), (y, e)\} \cup \{(x, e), (y, e)\}$. 

LV Equations and Transfer Functions

\[
\begin{align*}
\text{LV}_{\text{exit}}(\ell) &= \begin{cases} \emptyset, \text{if } \ell \in \text{final}(S_*) \\ \bigcup \{\text{LV}_{\text{entry}}(\ell') \mid (\ell', \ell) \in \text{flow}^R(S_*)\}, \text{otherwise} \end{cases} \\
\text{LV}_{\text{entry}}(\ell) &= (\text{LV}_{\text{exit}}(\ell) \setminus \text{kill}_{\text{LV}}([B]^{\ell}) \cup \text{gen}_{\text{LV}}([B]^{\ell})) \\
\text{where } [B]^{\ell} &\in \text{blocks}(S_*)
\end{align*}
\]

with

\[
\begin{align*}
\text{kill}_{\text{LV}}([x := a]^{\ell}) &= \{x\} \\
\text{kill}_{\text{LV}}([\text{skip}]^{\ell}) &= \emptyset \\
\text{kill}_{\text{LV}}([b]^{\ell}) &= \emptyset \\
\text{gen}_{\text{LV}}([x := a]^{\ell}) &= \text{FV}(a) \\
\text{gen}_{\text{LV}}([\text{skip}]^{\ell}) &= \emptyset \\
\text{gen}_{\text{LV}}([b]^{\ell}) &= \text{FV}(b)
\end{align*}
\]
Simplification, Abstraction, Approximation

Designing a Program Analysis needs to establish correctness.

Doing this for each program property, cf. Live Variable, might be cumbersome, so we are looking for a general way to construct correct and efficient frameworks; more or less automatically.

From the 1970s the work of Cousot and Cousot on Abstract Interpretation provides a tools to do this. They demonstrated that numerous analyses can be obtained this way.

The central element is the simplification of the concrete semantics in order to obtain an abstract one as an optimal approximation.

Concrete Semantics vs Abstract Semantics

code

m := 1;
while n > 1 do
  m := m*n;
  n := n-1;
endwhile

concrete model

abstract model

(1 0 0)
(0 1 0)
(0 0 1)

property

property #
Cast-out-of-Nines

**Plausibility check** for arithmetic calculations, for example:

\[ 123 \times 457 + 76543 = ? = 123654 \]

Perform operations \( n \mod 9 \) (enough to consider digits’ sum)

\[ 6 \times 7 + 7 = 42 + 7 = 6 + 7 = 4 \neq 3 \]

This holds because elementary facts like:

\[
\begin{align*}
(a \pm b) \mod 9 &= (a \mod 9 \pm b \mod 9) \mod 9 \\
(a \times b) \mod 9 &= (a \mod 9 \times b \mod 9) \mod 9 \\
(10 \times a \pm b) \mod 9 &= (a \pm b) \mod 9
\end{align*}
\]

Note that there are **false positives**, cf also [1] and [2].

---

**Approximation and Correctness**

Data-flow analyses can be re-formulated in a different scenario where correctness is guaranteed by construction.

Classically, the theory of **Abstract Interpretation** allows us to

- construct simplified a (computable) abstract semantics
- construct approximate solutions
- obtain the correctness of the approximate solutions

Abstract Interpretation also uses other techniques, like **widening/narrowing**, which we will not cover here.
Notions of Approximation

Assume that we have a “solution” $s$ to a problem. What counts as a (good) approximation $s^*$ to $s$?

In order theoretic structures we are looking for Safe Approximations

$$s^* \sqsubseteq s \quad \text{or} \quad s \sqsubseteq s^*$$

In quantitative, vector space structures we want Close Approximations

$$\|s - s^*\| = \min_x \|s - x\|$$

Example: Function Approximation

Concrete and abstract domain are step-functions on $[a, b]$. The set of (real-valued) step-function $T_n$ is based on the sub-division of the interval into $n$ sub-intervals.

The concrete function needs $n$ data points, its abstraction or approximation should need less, i.e. from $\mathbb{R}^n$ to $\mathbb{R}^m$ with $m < m$. 
Close Approximations

Approximate $f \in \mathbb{R}^{16}$ by “least square” simplifications in $\mathbb{R}^{8}$, in $\mathbb{R}^{4}$, in $\mathbb{R}^{2}$ or even in $\mathbb{R}$.

Safe Approximations

Approximate $f \in \mathbb{R}^{16}$ by over/under approximation in $\mathbb{R}^{8}$, in $\mathbb{R}^{4}$, in $\mathbb{R}^{2}$ or even in $\mathbb{R}$.
Abstract Interpretation

In Program Analysis (cf. Monotone Frameworks) our property spaces are (complete) lattice.

Aim: Find abstract descriptions on which computations are easier; then relate the concrete and abstract solutions.

Definition
Let \( C = (C, \leq_C) \) and \( D = (D, \leq_D) \) be two partially ordered sets. If there are two functions \( \alpha : C \to D \) and \( \gamma : D \to C \) such that for all \( c \in C \) and all \( d \in D \):

\[
c \leq_C \gamma(d) \text{ iff } \alpha(c) \leq_D d,
\]
then \( (C, \alpha, \gamma, D) \) form a Galois connection.

Relating Concrete and Abstract Properties

\[ C \]
\[ \gamma(d) \]
\[ c \]
\[ \alpha(c) \]
\[ d \]

e.g. \( \mathcal{P}(\mathbb{Z}) \)

e.g. \( \mathcal{P}({-, 0, +}) \)
Galois Connections

Definition
Let $\mathcal{C} = (\mathcal{C}, \leq \mathcal{C})$ and $\mathcal{D} = (\mathcal{D}, \leq \mathcal{D})$ be two partially ordered sets with two order-preserving functions $\alpha : \mathcal{C} \mapsto \mathcal{D}$ and $\gamma : \mathcal{D} \mapsto \mathcal{C}$. Then $(\mathcal{C}, \alpha, \gamma, \mathcal{D})$ form a Galois connection iff

(i) $\alpha \circ \gamma$ is reductive i.e. $\forall d \in \mathcal{D}, \alpha \circ \gamma(d) \leq \mathcal{D} d$,

(ii) $\gamma \circ \alpha$ is extensive i.e. $\forall c \in \mathcal{C}, c \leq \mathcal{C} \gamma \circ \alpha(c)$.

Proposition
Let $(\mathcal{C}, \alpha, \gamma, \mathcal{D})$ be a Galois connection. Then $\alpha$ and $\gamma$ are quasi-inverse, i.e.

(i) $\alpha \circ \gamma \circ \alpha = \alpha$ and (ii) $\gamma \circ \alpha \circ \gamma = \gamma$

Uniqueness and Duality
Given an abstraction $\alpha$ there is a unique concretisation $\gamma$.

Proposition
Let $(\mathcal{C}, \alpha, \gamma, \mathcal{D})$ be a Galois connection, then

(i) $\alpha$ uniquely determines $\gamma$ by

$$\gamma(d) = \bigsqcup \{c \mid \alpha(c) \leq \mathcal{D} d\},$$

and $\gamma$ uniquely determines $\alpha$ via

$$\alpha(c) = \bigsqcap \{d \mid c \leq \mathcal{C} \gamma(d)\}.$$

(ii) $\alpha$ is completely additive and $\gamma$ is completely multiplicative, and $\alpha(\bot) = \bot$ and $\gamma(\top) = \top$.

For a proof see e.g. [3] Lemma 4.22.
**Correctness and Optimality**

**Proposition**

Given $\alpha : \mathcal{P}(\mathbb{Z}) \to \mathcal{D}$ and $\gamma : \mathcal{D} \to \mathcal{P}(\mathbb{Z})$ a Galois connection with $\mathcal{D}$ some property lattice. Consider an operation $\text{op} : \mathbb{Z} \to \mathbb{Z}$ on $\mathbb{Z}$ which is lifted to $\widehat{\text{op}} : \mathcal{P}(\mathbb{Z}) \to \mathcal{P}(\mathbb{Z})$ via

$$\widehat{\text{op}}(Z) = \{ \text{op}(x) \mid x \in X \},$$

then $\text{op}^\# : \mathcal{D} \to \mathcal{D}$ defined as $\text{op}^\# = \alpha \circ \widehat{\text{op}} \circ \gamma$ is the most precise function on $\mathcal{D}$ satisfying for all $Z \subseteq \mathbb{Z}$:

$$\alpha(\widehat{\text{op}}(Z)) \subseteq \text{op}^\#(\alpha(Z))$$

It is enough to consider so-called Galois Insertions. See [1] Lemma 2.3.2.

---

**General Construction**

The general construction of correct (and optimal) abstractions $f^\#$ of concrete function $f$ is as follows:

$$
\begin{array}{ccc}
A & \xleftarrow{\alpha} & A^# \\
\downarrow{f} & \quad & \downarrow{f^#} \\
B & \xleftarrow{\alpha'} & B^#
\end{array}
$$

Correct approximation:

$$\alpha' \circ f \leq^\# f^\# \circ \alpha.$$

Induced semantics:

$$f^\# = \alpha' \circ f \circ \gamma.$$
Abstract Multiplication

How can we justify or obtain correct abstract versions of various operations, e.g. multiplication?

<table>
<thead>
<tr>
<th>× #</th>
<th>⊥</th>
<th>even</th>
<th>odd</th>
<th>⊤</th>
</tr>
</thead>
<tbody>
<tr>
<td>⊥</td>
<td>⊥</td>
<td>⊥</td>
<td>⊥</td>
<td>⊥</td>
</tr>
<tr>
<td>even</td>
<td>⊥</td>
<td>even</td>
<td>even</td>
<td>⊥</td>
</tr>
<tr>
<td>odd</td>
<td>⊥</td>
<td>even</td>
<td>odd</td>
<td>⊤</td>
</tr>
<tr>
<td>⊤</td>
<td>⊥</td>
<td>even</td>
<td>⊤</td>
<td>⊤</td>
</tr>
</tbody>
</table>

Abstract Interpretation – introduced by Patrick Cousot and Radhia Cousot in 1977 – allows to “compute” abstractions which are correct by construction.

Parity (again)

Consider concrete $C = \mathcal{P}(\mathbb{Z})$ and abstract $D = \mathcal{P}\{\text{even, odd}\}$.

The abstraction $\alpha : C \rightarrow D$ is given by for $X \subseteq \mathbb{Z}$:

\[
\begin{align*}
\alpha(\emptyset) &= ⊥ = \emptyset \\
\alpha(X) &= \text{even} \text{ iff } \forall x \in X \exists k : x = 2k \\
\alpha(X) &= \text{odd} \text{ iff } \forall x \in X \exists k : x = 2k + 1 \\
\alpha(X) &= ⊤ = \{\text{even, odd}\} \text{ otherwise}
\end{align*}
\]

The concretisation $\gamma : D \rightarrow C$ then needs to be:

\[
\begin{align*}
\gamma(⊥) &= \emptyset \\
\gamma(\text{even}) &= \{x \in \mathbb{Z} | \exists k : x = 2k\} = E \\
\gamma(\text{odd}) &= \{x \in \mathbb{Z} | \exists k : x = 2k + 1\} = O \\
\gamma(⊤) &= ⊤ = \mathbb{Z} \text{ otherwise}
\end{align*}
\]
Parity: From $\times$ to $\times^\#$

To construct $\times^\#$ using $\alpha$ and $\gamma$ we need to lift $\times: \mathbb{Z} \times \mathbb{Z} \to \mathbb{Z}$ to $\hat{\times}: \mathcal{P}(\mathbb{Z}) \times \mathcal{P}(\mathbb{Z}) \to \mathcal{P}(\mathbb{Z})$. Obviosly, for $X = \{x\} \subseteq \mathbb{Z}$ and $Y = \{y\} \subseteq \mathbb{Z}$:

$$X \hat{\times} Y = \{x \times y \mid x \in X \text{ and } y \in Y\}$$

Defining the abstract multiplication $\times^\# = \alpha \circ (\hat{\times}) \circ (\gamma, \gamma)$:

- $\gamma(\text{even}) = E$, then $E \hat{\times} E = E' \subset E$, and $\alpha(E') = \text{even}$
- $\gamma(\text{odd}) = O$, then $E \hat{\times} O = E$ and $\alpha(E) = \text{even}$
- etc.

Therefore, $\text{even} \times^\# \text{even} = \text{even}$, $\text{even} \times^\# \text{odd} = \text{even}$, etc.

Concrete Semantics $\rightarrow$ and Abstract Semantics $\Rightarrow$

Imagine some programming language, e.g. WHILE. Its concrete semantics identifies values in $\mathcal{V}$ (e.g. states) and specifies how a program $S$ transforms $v_1$ into $v_2$; we may write this as

$$S \vdash v_1 \rightarrow v_2$$

A program analysis or abstract semantics identifies the set $\mathcal{L}$ of properties and how a program $S$ transforms $l_1$ in to $l_2$

$$S \vdash l_1 \Rightarrow l_2$$

Unlike for general semantics, it is customary to require $\Rightarrow$ to be deterministic and thus define a function; this allows us to write:

$$f_S(l_1) = l_2 \text{ to mean } S \vdash l_1 \Rightarrow l_2.$$
Situation in While

We have SOS transitions $\langle S, s \rangle \Rightarrow \langle S', s' \rangle$ with $S$ and $S'$ programs and $s, s' \in \text{State} = (\text{Var} \rightarrow \mathbb{Z})$, e.g.

$\langle z := 2 \times z, [z \mapsto 2] \rangle \Rightarrow [z \mapsto 4]$ 

translates to just an evaluation of the state:

$z := 2 \times z \vdash [z \mapsto 2] \rightarrow [z \mapsto 4]$ 

The fact that this also holds for the (abstract) parity means:

$z := 2 \times z \vdash \text{even}(z) \rightsquigarrow \text{even}(z)$

and also $z := 2 \times z \vdash \text{odd}(z) \rightsquigarrow \text{even}(z)$.

Correctness Relation

Every program analysis should be correct with respect to the semantics.

For a class of (so-called first-order) program analyses this is established by directly relating properties to values using a correctness relation:

$\triangleright : \mathcal{V} \times \mathcal{L} \rightarrow \{\text{tt}, \text{ff}\}$ or $\triangleright \subseteq \mathcal{V} \times \mathcal{L}$

The intention is that "$v \triangleright l$" formalises our claim that the value $v$ is described by the property $l$ (or $v$ abstracts to $l$).
Preservation of Correctness

One has to prove that $\triangleright$ is preserved under computation. This may be formulated as the implication:

$$
\begin{align*}
&v_1 \triangleright l_1 \\
&S \vdash v_1 \rightarrow v_2 \\
&S \vdash l_1 \leadsto l_2 \\
\Rightarrow & \quad v_2 \triangleright l_2
\end{align*}
$$

This property is also expressed by the following diagram:

$$
\begin{array}{c}
S \vdash v_1 \rightarrow v_2 \\
\vdash \Rightarrow \vdash \\
S \vdash l_1 \leadsto l_2
\end{array}
$$

Correctness of Parity

<table>
<thead>
<tr>
<th>0 $\triangleright$ even</th>
<th>1 $\triangleright$ odd</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 $\triangleright$ even</td>
<td>3 $\triangleright$ odd</td>
</tr>
<tr>
<td>4 $\triangleright$ even</td>
<td>5 $\triangleright$ odd</td>
</tr>
<tr>
<td>$\cdots$</td>
<td>$\cdots$</td>
</tr>
</tbody>
</table>

- $z := 2 \times z \vdash [z \mapsto 1] \rightarrow [z \mapsto 2]$ \quad $\text{odd}(z) \leadsto \text{even}(z)$
- $z := 2 \times z \vdash [z \mapsto 2] \rightarrow [z \mapsto 4]$ \quad $\text{even}(z) \leadsto \text{even}(z)$
- $z := 2 \times z \vdash [z \mapsto 3] \rightarrow [z \mapsto 6]$ \quad $\text{odd}(z) \leadsto \text{even}(z)$
- $\cdots$ | $\cdots$

Thus it is correct: \( p \equiv z := 2 \times z \) always produces an \textbf{even} \( z \).
Abstract Interpretation and Correctness

The theory of Abstract Interpretation comes to life when we augment the set of properties $\mathcal{L}$ with a preorder (better: lattice) structure and elate this to the correctness relation $\triangleright$.

The most common scenario is when $\mathcal{L} = (\mathcal{L}, \sqsubseteq, \sqcup, \sqcap, \bot, \top)$ is a complete lattice with partial ordering $\sqsubseteq$.

We then impose the following relationship between $\triangleright$ and $\mathcal{L}$:

\[ v \triangleright l_1 \land l_1 \sqsubseteq l_2 \Rightarrow v \triangleright l_2 \]  
(1)

\[ \forall l \in \mathcal{L}' \subseteq \mathcal{L} : v \triangleright l \Rightarrow v \triangleright \bigcap \mathcal{L}' \]  
(2)

Condition (1)

Consider the first of these conditions:

\[ v \triangleright l_1 \land l_1 \sqsubseteq l_2 \Rightarrow v \triangleright l_2 \]

- The condition says that the smaller the property is with respect to the partial order, the better (i.e. precise) it is.

- This is an “arbitrary” decision in the sense that we could instead have decided that the larger the property is, the better it is, as is indeed the case in much of the literature on Data Flow Analysis; luckily the principle of duality from lattice theory tells us that this difference is only cosmetic.
Condition (2)

Looking at the second condition describing correctness:

$$\forall l \in L' \subseteq L : v \triangleright l \Rightarrow v \triangleright \bigcap L'$$

- The second condition says that there is always a best property for describing a value. This is important for having to perform only one analysis (using the best property, i.e. the greatest lower bound of the candidates) instead of several analyses (one for each of the candidates).

- The condition has two immediate consequences:
  
  $$v \triangleright \top$$
  
  $$v \triangleright l_1 \land v \triangleright l_2 \Rightarrow v \triangleright (l_1 \cap l_2)$$

Again: Parity Example

The abstract properties **even** and **odd** do themselves not form a lattice \(L\), but we can use – as usual: \(L = \mathcal{P}\{\text{even, odd}\}\), where \{**even**\} represents the definitive fact **even** and \{**odd**\} the precise property **odd**; while the empty set \(\perp = \emptyset\) represents an undefined parity and \(\top = \{\text{even, odd}\}\) stands for any parity.

The conditions imposed on \(\triangleright\) and \(L\) mean in this case:

1. Any parity is always a valid description, e.g.
   
   $$2 \triangleright \{\text{even}\} \land \{\text{even}\} \subseteq \top \Rightarrow 2 \triangleright \top$$

2. The most precise parity is valid, e.g.
   
   $$(2 \triangleright \{\text{even}\} \land 2 \triangleright \top) \Rightarrow 2 \triangleright (\{\text{even}\} \cap \top)$$
   
   i.e. $$(2 \triangleright \{\text{even}\} \land 2 \triangleright \top) \Rightarrow 2 \triangleright \{\text{even}\}$$
Preservation of Correctness via Abstraction

We require that correctness is preserved:

\[ v_1 \triangleright l_1 \land S \vdash v_1 \rightarrow v_2 \land S \vdash l_1 \leadsto l_2 \Rightarrow v_2 \triangleright l_2 \]

With a (semantical transfer) function \( f_S \) we have:

\[ v_1 \triangleright l_1 \land f_S(v_1) = v_2 \land f^*_S(l_1) = l_2 \Rightarrow v_2 \triangleright l_2 \]

This property is also expressed by the following diagram:

\[ \begin{array}{ccc}
\mathcal{V} & \xrightarrow{f_S} & \mathcal{V} \\
\downarrow^\alpha & & \downarrow^\gamma \\
\mathcal{L} & \xrightarrow{f^*_S} & \mathcal{L}
\end{array} \]

Representation and Extraction Functions

We can use a representation function \( \beta : \mathcal{V} \rightarrow \mathcal{L} \) to induce a Galois connection \((\mathcal{P}(\mathcal{V}), \alpha, \gamma, \mathcal{L})\) via

\[
\alpha(\mathcal{V}) = \bigsqcup \{ \beta(v) \mid v \in \mathcal{V} \}
\]

\[
\gamma(l) = \{ v \in \mathcal{V} \mid \beta(v) \sqsubseteq l \}
\]

For \( \mathcal{L} = \mathcal{P}(\mathcal{D}) \) with \( \mathcal{D} \) being some set of “abstract values” we can also use an extraction function, \( \eta : \mathcal{V} \rightarrow \mathcal{D} \) defined as

\[
\alpha(\mathcal{V}) = \{ \eta(v) \mid v \in \mathcal{V} \}
\]

\[
\gamma(\mathcal{D}) = \{ v \mid \eta(v) \in \mathcal{D} \}
\]

in order to construct a Galois connection.
Example: Parity

A representation function $\beta : \mathbb{Z} \rightarrow \mathcal{P}(\{\text{even, odd}\})$ is easily defined by:

$$\beta(n) = \begin{cases} 
\{\text{even}\} & \text{if } \exists k \in \mathbb{Z} \text{ s.t. } n = 2k \\
\{\text{odd}\} & \text{otherwise}
\end{cases}$$

Correctness implies that the abstract properties are dominated by the actual ones, e.g. $\beta(4) = \{\text{even}\} \subseteq \top = \{\text{even, odd}\}$ is acceptable.

This means that we also could use as a representation function

$$\beta(n) = \top = \{\text{even, odd}\}$$

for all $n \in \mathbb{Z}$. Though this would be valid it would also be rather imprecise.

References Abstract Interpretation


