

# Modular Analysis

Joonhyup Lee

May 22, 2024

## Abstract

We present a framework of modular static analysis that analyzes open program fragments in advance and complete the whole analysis later when the fragments are closed. The framework is defined for a call-by-value lambda calculus extended with constructs for defining and linking first-class modules (a collection of bindings) that support recursive bindings for values, modules, and functors(module functions). Thanks to the abstract interpretation framework, our modular analysis framework’s key is how to define the semantics, called “shadow” semantics of open program fragments so that the computation with free variables should be captured in advance and be completed later at link-time. A modular static analysis is abstractions of this shadow semantics and of the linking operation. The safety of the framework is proven in Coq. Two instances of the framework are presented: value analyses for an applicative language(control-flow, closure analysis) and for an imperative language.

## 1 Problem

By modular analysis we mean a static analysis technology that analyzes programs in the compositional way. Modular analysis partially analyzes program fragments separately and then complete the analysis when all the fragments are available and linked.

Designing a modular analysis decides on two factors: how to build sound partial semantic summaries of program fragments and how to complete them at link time. A particular modular analysis strikes a balance between the two factors. Depending on how much analysis work is invested on each factor, modular analyses range from one extreme, a trivial one (wait until the whole code is available and do the whole-program analysis) to another extreme, non-trivial yet hardly-automatic ones (abduct assumptions about free variables and check them at link-time).

However, designing a modular analysis has been elusive. There is a lack of a general framework by which we can design sound modular analyses with varying balance and accuracy of our choice. We need a general framework by which static analysis designers can design non-trivial sound modular analyses.

## 2 Framework Sketch

Thanks to the abstract interpretation framework, the key in our modular analysis framework is how to define the semantics (we call it “shadow” semantics) of open program fragments so that the computation with free variables should be captured in advance and be completed later at link-time. A modular static analysis is then nothing but abstractions of this shadow semantics and of the linking operation.

### 2.1 Examples of Shadow Semantics and its Abstraction

Shadow semantics is the semantics for the computation involving free variables. We give examples of shadows for program fragments written in OCaml.

**Example 1: Mapping an Unknown Function Over a Known List**

```
let rec map = fun f -> fun l ->
  match 0l with
  | [] -> []
  | hd :: tl -> 1(3f 4hd) :: 2(map f tl)
in map 5g 6(1 :: 2 :: 3 :: [])
```

The shadow semantics of the above fragment is

$$\{\text{Call}(G, 1) :: \text{Call}(G, 2) :: \text{Call}(G, 3) :: []\} \text{ where } G = \text{Read}(\text{Init}, g)$$

Shadows such as *Call*, *Read* correspond to semantic operations, like function application and reading from the environment. The *Init* shadow is the unknown initial environment that the fragment will execute under.

Computing sound and finite approximations of the shadow semantics of program fragments corresponds to building sound partial semantic summaries in advance. For the above example, we may use an abstraction that stores summaries of each input *environment* and output *value* per program point.

**Notation.** We write  $p_i$  for each program point  $i$ .  $p_i.in$  denotes the input to  $p_i$ , and  $p_i.out$  denotes the output from  $p_i$ . For brevity, we omit the selector in  $p_i.out$  and write  $p_i$ . We write  $[a, b]_{itv}$  to denote the interval from integer  $a$  to  $b$ .

Now, the abstract shadow

$$\{p_1 :: p_2\}$$

is a sound approximation of the concrete shadow. To illustrate why, here are abstract shadows from some program points involved in the map computation.

$$\begin{array}{ll} p_1.out : \{\text{Call}^\#(p_3, p_4)\} & p_4.out : [1, 3]_{itv} \\ p_2.out : \{[], p_1 :: p_2\} & p_5.in : \{\text{Init}^\#, \text{map} \mapsto \{\langle \lambda f. \dots \rangle\}\} \\ p_3.out : \{\text{Read}^\#(p_5.in, g)\} & \end{array}$$

Starting from  $p_1 :: p_2$ , we can reach the concrete shadow by expanding each program point according to the above table.

### Example 2: Mapping an Unknown Function Over an Unknown List

`let rec map = ... in map 5g 6lst`

A more interesting case is when even the list is replaced with a free variable. The shadow semantics of the above fragment is

$$\{\llbracket \cdot \rrbracket, \text{Call}(G, Hd(Lst)) :: \llbracket \cdot \rrbracket, \text{Call}(G, Hd(Lst)) :: \text{Call}(G, Hd(Tl(Lst))) :: \llbracket \cdot \rrbracket, \dots, \text{Call}(G, Hd(Lst)) :: \dots :: \text{Call}(G, Hd(Tl^n(Lst))) :: \llbracket \cdot \rrbracket, \dots\}$$

where

$$\begin{aligned} G &= \text{Read}(\text{Init}, g) & Lst &= \text{Read}(\text{Init}, \text{lst}) \\ Hd(l) &= \text{Match}(l, ::, 0) & Tl(l) &= \text{Match}(l, ::, 1) \end{aligned}$$

The same abstraction as in Example 1 results in the same abstract shadow

$$\{p_1 :: p_2\}$$

but with different shadows for each program point.

$$\begin{aligned} p_0.out &: \{\text{Read}^\#(p_6.in, \text{lst}), & p_3.out &: \{\text{Read}^\#(p_5.in, g)\} \\ & \text{Match}^\#(p_0, ::, 1)\} & p_4.out &: \{\text{Match}^\#(p_0, ::, 0)\} \\ p_1.out &: \{\text{Call}^\#(p_3, p_4)\} & p_5.in &: \{\text{Init}^\#, \text{map} \mapsto \{\langle \lambda f \dots \rangle\}\} \\ p_2.out &: \{\llbracket \cdot \rrbracket, p_1 :: p_2\} & p_6.in &: \{\text{Init}^\#, \text{map} \mapsto \{\langle \lambda f \dots \rangle\}\} \end{aligned}$$

### Example 3: Unknown Function Modifying the Memory

`let x = 1ref 1 in 2(4g 5x); 3!x`

The shadow semantics can be extended to support imperative features.

`ref e (allocation)    x := e (update)    !x (dereference)`

We only have to add a memory component, and a shadow for *dereferencing*.

In the above program, the environment and memory at  $p_1$ ,  $p_2$ , and  $p_3$  are

	Environment	Memory
$p_1$	$\langle \emptyset, \text{Init} \rangle$	$\langle \emptyset, \text{Init} \rangle$
$p_2$	$\langle \{x \mapsto a\}, \text{Init} \rangle$	$\langle \{a \mapsto 1\}, \text{Init} \rangle$ where $a \in \text{Addr}$
$p_3$	$\langle \{x \mapsto a\}, \text{Init} \rangle$	$\langle \{a \mapsto \text{Deref}(C, a)\}, C \rangle$ where $C = \text{Call}(G, a m)$

where

$$G = \text{Read}(\text{Init}, g) \text{ and } m = \langle \{a \mapsto 1\}, \text{Init} \rangle$$

The notation  $\text{Call}(G, a|m)$  is inspired from conditional probability, and means that the shadow  $G$  is called with argument  $a$  *given* memory  $m$ .

The memory is a pair of (1) the set of bindings from addresses to values, and (2) the shadow that keeps track of unknown modifications to the memory. Here, the unknown function call  $\text{Call}(G, a|m)$  modifies what was stored in  $a$ .

**Notation.** We write  $a_i^\#$  to abstract the concrete location allocated at  $p_i$ .  $p_i.in.\sigma^\#$  denotes the input environment at  $p_i$ , and  $p_i.in.m^\#$  denotes the input memory. Likewise,  $p_i.out.v^\#$  denotes the output value at  $p_i$ , and  $p_i.out.m^\#$  denotes the output memory. The selectors  $.out$ ,  $.\sigma^\#$ ,  $.v^\#$ , and  $.m^\#$  are omitted.

Using the same naïve abstraction as before, the abstract result is

$$\{\text{Deref}^\#(p_2, p_1)\} \quad (\text{abbreviated form of } \text{Deref}^\#(p_2.out.m^\#, p_1.out.v^\#))$$

where

$$\begin{aligned} p_2.out.v^\# &: \{\text{Call}^\#(p_4, p_5|p_5)\} & p_4.out.v^\# &: \{\text{Read}^\#(p_4.in, g)\} \\ p_2.out.m^\# &: \{\text{Call}^\#(p_4, p_5|p_5), & p_5.in.\sigma^\# &: \{\text{Init}^\#, x \mapsto \{a_1^\#\}\} \\ & a_1^\# \mapsto \{\text{Deref}^\#(p_2, p_1)\}\} & p_5.out.v^\# &: \{a_1^\#, \text{Read}^\#(p_5.in, x)\} \\ p_4.in.\sigma^\# &: \{\text{Init}^\#, x \mapsto \{a_1^\#\}\} & p_5.out.m^\# &: \{\text{Init}^\#, a_1^\# \mapsto [1, 1]_{itv}\} \end{aligned}$$

#### Example 4: Reading and Writing to Shadow Addresses

$$\text{let } t = (!x)^1 \text{ in } (x := (!y)^2)^3; (y := t)^4$$

The addresses that are dereferenced and updated may also be shadows. Thus, the memory must store bindings from *shadow addresses* to values. A shadow address is a *set* consisting of one or less actual address and zero or more shadows that are aliases of the actual address.

Upon exit at each program point marked above,

	Environment	Possible Memories
$p_1$	$\langle \{t \mapsto *X\}, \text{Init} \rangle$	$\langle \{\{X\} \mapsto *X\}, \text{Init} \rangle$
$p_2$	$\langle \{t \mapsto *X\}, \text{Init} \rangle$	$\langle \{\{X, Y\} \mapsto *X\}, \text{Init} \rangle$ or $\langle \{\{X\} \mapsto *X, \{Y\} \mapsto *Y\}, \text{Init} \rangle$
$p_3$	$\langle \{t \mapsto *X\}, \text{Init} \rangle$	$\langle \{\{X, Y\} \mapsto *X\}, \text{Init} \rangle$ or $\langle \{\{X\} \mapsto *Y, \{Y\} \mapsto *Y\}, \text{Init} \rangle$
$p_4$	$\langle \{t \mapsto *X\}, \text{Init} \rangle$	$\langle \{\{X, Y\} \mapsto *X\}, \text{Init} \rangle$ or $\langle \{\{X\} \mapsto *Y, \{Y\} \mapsto *X\}, \text{Init} \rangle$

where

$$X = \text{Read}(\text{Init}, x) \quad Y = \text{Read}(\text{Init}, y) \quad *S = \text{Deref}(\text{Init}, S) \text{ for shadow } S$$

Unlike the previous example, there are multiple memories possible after execution. Each memory represents different aliasing assumptions. For example,  $\langle \{\{X, Y\} \mapsto \_ \}, \_ \rangle$  encodes the assumption that  $x$  and  $y$  are aliases.

Computing an abstraction of the shadow memory is an interesting problem. One must retain some information about the aliasing *invariant* that is maintained throughout the execution of the program to avoid degenerating to a meaningless analysis.

One traditional way is by assuming a partition of variables that distinguishes variables that must-not-alias. For the above program, we may assume the invariant that  $\mathbf{x}$  and  $\mathbf{y}$  must not alias.

**Notation.** The memory is abstracted by a set of abstract shadows and abstract bindings of the form  $*x \mapsto (\text{set of abstract values})$ . The abstract shadow  $\text{Deref}^\#(p_i, x)$  concretizes to  $\text{Deref}(S, \sigma(x))$ , where  $S$  is abstracted by  $p_i.in.m^\#$  and  $\sigma$  is abstracted by  $p_i.in.\sigma^\#$ .

Assuming a flow-sensitive analysis that allows strong updates when the updated variable can have only one concrete address,

	$\sigma^\#$	$m^\#$
$p_2.in$	$\{\mathbf{t} \mapsto \{*X^\#\}, \text{Init}^\#\}$	$\{\mathbf{x} \mapsto \{*X^\#\}, \text{Init}^\#\}$
$p_3.in$	$\{\mathbf{t} \mapsto \{*X^\#\}, \text{Init}^\#\}$	$\{\mathbf{x} \mapsto \{*X^\#\}, \mathbf{y} \mapsto \{*Y^\#\}, \text{Init}^\#\}$
$p_4.in$	$\{\mathbf{t} \mapsto \{*X^\#\}, \text{Init}^\#\}$	$\{\mathbf{x} \mapsto \{*Y^\#\}, \mathbf{y} \mapsto \{*Y^\#\}, \text{Init}^\#\}$
$p_4.out$	—	$\{\mathbf{x} \mapsto \{*Y^\#\}, \mathbf{y} \mapsto \{*X^\#\}, \text{Init}^\#\}$

where

$$*X^\# = \text{Deref}^\#(p_1, \mathbf{x}) \quad *Y^\# = \text{Deref}^\#(p_2, \mathbf{y})$$

The aliasing assumption is used at  $p_2$ , where dereferencing  $\mathbf{y}$  doesn't include the value pointed by  $\mathbf{x}$ . Strong updates are possible, since  $\mathbf{x}$  concretizes only to  $\text{Read}(\text{Init}, \mathbf{x})$ , and  $\mathbf{y}$  concretizes only to  $\text{Read}(\text{Init}, \mathbf{y})$  throughout the analysis.

## 2.2 Examples of Linking and its Abstraction

The shadow semantics become actual when the involved free variables are known at link time. The linking semantics, the semantics of the link operation, defines this actualization operation. For the above fragments, let's consider that a closing fragment is available.

### Example 5: Module Exporting a User-Defined Function

```
let g = 7fun x -> x + 1
```

This is a module which returns the environment  $\sigma = \{g \mapsto \langle \lambda \mathbf{x}. \mathbf{x} + 1, \emptyset \rangle\}$ . Linking this with the concrete shadow in Example 1 gives:

$$\{2 :: 3 :: 4 :: []\}$$

Computing a sound and finite approximation of the linking semantics corresponds to completing partial analysis summaries at link time when the analysis results for the involved free variable are available. For the above example,

$$\sigma^\# = \{g \mapsto \{\langle \lambda \mathbf{x}. \mathbf{x} + 1, p_7.in \rangle\}\} \text{ where } p_7.in : \emptyset$$

is the abstract shadow.

Applying a sound abstract version of the linking operator will result in

$$\begin{array}{ll}
p_1.out : [2, 4]_{itv} & p_4.out : [1, 3]_{itv} \\
p_2.out : \{\square, p_1 :: p_2\} & p_5.in : \sigma^\# \\
p_3.out : \{\langle \lambda x.x + 1, p_7.in \rangle\} &
\end{array}$$

### Example 6: Module Exporting a Foreign Function

```
external g : int -> int = "incr"
```

The function `g` might also be a foreign function. The return value of this module is  $\sigma = \{g \mapsto \text{Prim}(\text{incr})\}$ , where `Prim` stands for a *primitive* value. Linking this with the concrete shadow in Example 1 gives:

$$\{\text{PrimCall}(\text{incr}, 1) :: \text{PrimCall}(\text{incr}, 2) :: \text{PrimCall}(\text{incr}, 3) :: \square\}$$

The sound abstraction for  $\sigma$  is

$$\sigma^\# = \{g \mapsto \{\text{Prim}(\text{incr})\}\}$$

Applying a sound abstract version of the linking operator will result in

$$\begin{array}{ll}
p_1.out : \{\text{PrimCall}^\#(\text{incr}, p_4)\} & p_4.out : [1, 3]_{itv} \\
p_2.out : \{\square, p_1 :: p_2\} & p_5.in : \sigma^\# \\
p_3.out : \{\text{Prim}(\text{incr})\} &
\end{array}$$

### Example 7: Partial Resolution of Shadows

```
let lst = 71 :: 8(92 :: 10(113 :: 12[]) )
```

The return value of this module is  $\sigma = \langle \{\text{lst} \mapsto 1 :: 2 :: 3 :: \square\}, \text{Init} \rangle$ . Linking this with the concrete shadow in Example 2 gives:

$$\{\square, \text{Call}(G, 1) :: \square, \text{Call}(G, 1) :: \text{Call}(G, 2) :: \square, \text{Call}(G, 1) :: \text{Call}(G, 2) :: \text{Call}(G, 3) :: \square\}$$

where

$$G = \text{Read}(\text{Init}, g)$$

The reason for this imprecision is because some information is forgotten. For example, the output of Example 2 can be  $\square$  if and only if `Lst` is matched with  $\square$ . To prevent this, shadows may be augmented with constraints, such as `Matched(Lst,  $\square$ )`, but we elide this detail for presentation.

The sound abstraction for  $\sigma$  is

$$\sigma^\# = \{\text{Init}^\#, \text{lst} \mapsto \{p_7 :: p_8\}\} \text{ where } p_7.out : [1, 1]_{itv}, p_8.out : \{p_9 :: p_{10}\} \dots$$

The result of linking  $\sigma^\#$  to the abstract shadow in Example 2 is equivalent to the result of Example 1. This means that the shadows can be filled in *incrementally*.

### Example 8: Linking into Memory

```
let x = 5ref 1
let y = 6ref 2
```

Assume we want to link this module with Example 4. The environment and memory exported by this module are

$$\sigma = \{x \mapsto a_5, y \mapsto a_6\} \quad m = \{\{a_5\} \mapsto 1, \{a_6\} \mapsto 2\} \quad \text{where } a_5 \neq a_6$$

Linking is extended, where linking an input environment and memory with a shadow results in an output value and memory.

Since the result of linking into  $X = \text{Read}(\text{Init}, x)$  and  $Y = \text{Read}(\text{Init}, y)$  are different, only the memories where  $X$  and  $Y$  are kept separate are resolved. The result of concrete linking is thus:

	Environment	Possible Memories
$p_4$	$\{t \mapsto 1\}$	$\{\{a_5\} \mapsto 2, \{a_6\} \mapsto 1\}$

To link with the abstract memory in Example 4, the module must be analyzed under the same assumptions. Namely,  $x$  and  $y$  should be always assumed to be not aliasing. Under this assumption,

$$\sigma^\# = \{x \mapsto \{a_5^\#\}, y \mapsto \{a_6^\#\}\} \quad m^\# = \{*x \mapsto [1, 1]_{itv}, *y \mapsto [2, 2]_{itv}\}$$

and linking results in the expected behavior.

## 2.3 Paper Overview

We present our framework for a call-by-value lambda calculus extended with constructs for defining and linking first-class modules (a collection of bindings) that support recursive bindings for values, modules, and functors (module functions).

The framework shows two points: how to define the shadow semantics and what to prove for the soundness of consequent modular analysis. The safety of the framework is proven in Coq. We present two instances of the framework: for high-order applicative language we show modular closure analysis design, and for imperative languages we show modular [TODO] analysis design.

### 3 Syntax and Semantics

#### 3.1 Abstract Syntax

Identifiers	$x$	$\in$	Var	
Expression	$e$	$\rightarrow$	$x \mid \lambda x.e \mid e \ e$	$\lambda$ -calculus
			$\mid e \bowtie e$	linked expression
			$\mid \varepsilon$	empty module
			$\mid x = e ; e$	(recursive) binding

Figure 1: Abstract syntax of the language.

#### 3.2 Operational Semantics

Environment	$\sigma$	$\in$	$\text{Env} \triangleq \{\bullet\} + \text{Var} \times (\text{Loc} + \text{WVal}) \times \text{Env}$	
Location	$\ell$	$\in$	Loc	
Value	$v$	$\in$	$\text{Val} \triangleq \text{Env} + \text{Var} \times \text{Expr} \times \text{Env}$	
Weak Value	$w$	$\in$	$\text{WVal} \triangleq \text{Val} + \underline{\text{Val}}$	
Environment	$\sigma$	$\rightarrow$	$\bullet$	empty stack
			$\mid (x, w) :: \sigma$	weak value binding
			$\mid (x, \ell) :: \sigma$	free location binding
Value	$v$	$\rightarrow$	$\sigma$	exported environment
			$\mid \langle \lambda x.e, \sigma \rangle$	closure
Weak Value	$w$	$\rightarrow$	$v$	value
			$\mid \mu \ell.v$	recursive value

Figure 2: Definition of the semantic domains.

The big-step operational semantics is *deterministic* up to  $\alpha$ -equivalence. We have also proven the bisimilarity of this semantics with a more conventional one that uses backpatching for arbitrary recursive definitions.



$$\boxed{\sigma \vdash e \Downarrow v}$$

$$\begin{array}{c}
\text{ID} \qquad \text{RECID} \qquad \text{FN} \\
\frac{\sigma(x) = v}{\sigma \vdash x \Downarrow v} \quad \frac{\sigma(x) = \mu\ell.v}{\sigma \vdash x \Downarrow v[\mu\ell.v/\ell]} \quad \frac{}{\sigma \vdash \lambda x.e \Downarrow \langle \lambda x.e, \sigma \rangle}
\end{array}$$

$$\begin{array}{c}
\text{APP} \\
\frac{\sigma \vdash e_1 \Downarrow \langle \lambda x.e, \sigma_1 \rangle \quad \sigma \vdash e_2 \Downarrow v_2 \quad (x, v_2) :: \sigma_1 \vdash e \Downarrow v}{\sigma \vdash e_1 e_2 \Downarrow v}
\end{array}$$

$$\begin{array}{c}
\text{LINK} \qquad \text{EMPTY} \\
\frac{\sigma \vdash e_1 \Downarrow \sigma_1 \quad \sigma_1 \vdash e_2 \Downarrow v}{\sigma \vdash e_1 \times e_2 \Downarrow v} \quad \frac{}{\sigma \vdash \varepsilon \Downarrow \bullet}
\end{array}$$

$$\begin{array}{c}
\text{BIND} \\
\frac{\ell \notin \text{FLoc}(\sigma) \quad (x, \ell) :: \sigma \vdash e_1 \Downarrow v_1 \quad (x, \mu\ell.v_1) :: \sigma \vdash e_2 \Downarrow \sigma_2}{\sigma \vdash x = e_1; e_2 \Downarrow (x, \mu\ell.v_1) :: \sigma_2}
\end{array}$$

Figure 3: The big-step operational semantics.

## 4 Generating and Resolving Shadows

Now we formulate the semantics for generating shadows.

Shadow	$S$	$\rightarrow$	Init	initial environment
			Read( $S, x$ )	read shadow
			Call( $S, v$ )	call shadow
Environment	$\sigma$	$\rightarrow$	$\dots$	
			$[S]$	answer to an shadow
Value	$v$	$\rightarrow$	$\dots$	
			$S$	answer to an shadow

Figure 4: Definition of the semantic domains with shadows. All other semantic domains are equal to Figure 2.

We extend how to read weak values given an environment.

$$\begin{aligned}
 \bullet(x) &\triangleq \perp & ((x', \ell) :: \sigma)(x) &\triangleq (x = x' ? \ell : \sigma(x)) \\
 [S](x) &\triangleq \text{Read}(S, x) & ((x', w) :: \sigma)(x) &\triangleq (x = x' ? w : \sigma(x))
 \end{aligned}$$

Then we need to add only three rules to the semantics in Figure 3 for the semantics to incorporate shadows.

$$\begin{array}{c}
 \text{LINKSHADOW} \quad \frac{\sigma \vdash e_1 \Downarrow S \quad [S] \vdash e_2 \Downarrow v}{\sigma \vdash e_1 \times e_2 \Downarrow v} \quad \text{APPSHADOW} \quad \frac{\sigma \vdash e_1 \Downarrow S \quad \sigma \vdash e_2 \Downarrow v}{\sigma \vdash e_1 e_2 \Downarrow \text{Call}(S, v)} \\
 \\
 \text{BINDSHADOW} \quad \frac{\ell \notin \text{FLoc}(\sigma) \quad (x, \ell) :: \sigma \vdash e_1 \Downarrow v_1 \quad (x, \mu\ell.v_1) :: \sigma \vdash e_1 \Downarrow S_2}{\sigma \vdash x = e_1; e_2 \Downarrow (x, \mu\ell.v_1) :: [S_2]}
 \end{array}$$

Now we need to formulate the *concrete linking* rules. The concrete linking rule  $\sigma_0 \times w$ , given an answer  $\sigma_0$  to the Init shadow, resolves all shadows within  $w$  to obtain a set of final results.

$$\boxed{\times \in \text{Env} \rightarrow \text{Shadow} \rightarrow \mathcal{P}(\text{Val})}$$

$$\begin{aligned} \sigma_0 \times \text{Init} &\triangleq \{\sigma_0\} \\ \sigma_0 \times \text{Read}(S, x) &\triangleq \{v_+ | \sigma_+ \in \sigma_0 \times S, \sigma_+(x) = v_+\} \\ &\quad \cup \{v_+ | \mu\ell.v_+/\ell | \sigma_+ \in \sigma_0 \times E, \sigma_+(x) = \mu\ell.v_+\} \\ \sigma_0 \times \text{Call}(S, v) &\triangleq \{v'_+ | \langle \lambda x.e, \sigma_+ \rangle \in \sigma_0 \times E, v_+ \in \sigma_0 \times v, (x, v_+) :: \sigma_+ \vdash e \Downarrow v'_+\} \\ &\quad \cup \{\text{Call}(S_+, v_+) | S_+ \in \sigma_0 \times E, v_+ \in \sigma_0 \times v\} \end{aligned}$$

$$\boxed{\times \in \text{Env} \rightarrow \text{Env} \rightarrow \mathcal{P}(\text{Env})}$$

$$\begin{aligned} \sigma_0 \times \bullet &\triangleq \{\bullet\} \\ \sigma_0 \times (x, \ell) :: \sigma &\triangleq \{(x, \ell) :: \sigma_+ | \sigma_+ \in \sigma_0 \times \sigma\} \\ \sigma_0 \times (x, w) :: \sigma &\triangleq \{(x, w_+) :: \sigma_+ | w_+ \in \sigma_0 \times w, \sigma_+ \in \sigma_0 \times \sigma\} \\ \sigma_0 \times [S] &\triangleq \{\sigma_+ | \sigma_+ \in \sigma_0 \times S\} \cup \{[S_+] | S_+ \in \sigma_0 \times S\} \end{aligned}$$

$$\boxed{\times \in \text{Env} \rightarrow \text{Val} \rightarrow \mathcal{P}(\text{Val})}$$

$$\sigma_0 \times \langle \lambda x.e, \sigma \rangle \triangleq \{\langle \lambda x.e, \sigma_+ \rangle | \sigma_+ \in \sigma_0 \times \sigma\}$$

$$\boxed{\times \in \text{Env} \rightarrow \text{WVal} \rightarrow \mathcal{P}(\text{WVal})}$$

$$\sigma_0 \times \mu\ell.v \triangleq \{\mu\ell'.v_+ | \ell' \notin \text{FLoc}(v) \cup \text{FLoc}(\sigma_0), v_+ \in \sigma_0 \times v[\ell'/\ell]\}$$

Concrete linking makes sense because of the following theorem. First define:

$$\text{eval}(e, \sigma) \triangleq \{v | \sigma \vdash e \Downarrow v\} \quad \text{eval}(e, \Sigma) \triangleq \bigcup_{\sigma \in \Sigma} \text{eval}(e, \sigma) \quad \Sigma_0 \times W \triangleq \bigcup_{\substack{\sigma_0 \in \Sigma_0 \\ w \in W}} (\sigma_0 \times w)$$

Then the following holds:

**Theorem 4.1** (Advance). *Given  $e \in \text{Expr}$ ,  $\Sigma_0, \Sigma \subseteq \text{Env}$ ,*

$$\text{eval}(e, \Sigma_0 \times \Sigma) \subseteq \Sigma_0 \times \text{eval}(e, \Sigma)$$

Now we can formulate modular analysis. A modular analysis consists of two requirements: an abstraction for the semantics with shadows and an abstraction for the semantic linking operator.

**Theorem 4.2** (Modular analysis). *Assume:*

1. *An abstract domain  $\text{WVal}^\#$  and a monotonic  $\gamma \in \text{WVal}^\# \rightarrow \mathcal{P}(\text{WVal})$*
2. *A sound  $\text{eval}^\#$ :  $\Sigma_0 \subseteq \gamma(\sigma_0^\#) \Rightarrow \text{eval}(e, \Sigma_0) \subseteq \gamma(\text{eval}^\#(e, \sigma_0^\#))$*
3. *A sound  $\times^\#$ :  $\Sigma_0 \subseteq \gamma(\sigma_0^\#)$  and  $W \subseteq \gamma(w^\#) \Rightarrow \Sigma_0 \times W \subseteq \gamma(\sigma_0^\# \times^\# w^\#)$*

*then we have:*

$$\Sigma_0 \subseteq \gamma(\sigma_0^\#) \text{ and } \Sigma \subseteq \gamma(\sigma^\#) \Rightarrow \text{eval}(e, \Sigma_0 \times \Sigma) \subseteq \gamma(\sigma_0^\# \times^\# \text{eval}^\#(e, \sigma^\#))$$

**Corollary 4.1** (Modular analysis of linked program).

$$\begin{aligned} \Sigma_0 \subseteq \gamma(\sigma_0^\#) \text{ and } [\text{Init}] \in \gamma(\text{Init}^\#) \Rightarrow \\ \text{eval}(e_1 \times e_2, \Sigma_0) \subseteq \gamma(\text{eval}^\#(e_1, \sigma_0^\#) \times^\# \text{eval}^\#(e_2, \text{Init}^\#)) \end{aligned}$$

## 5 CFA

### 5.1 Collecting semantics

Program point	$p$	$\in$	$\mathbb{P} \triangleq \{\text{finite set of program points}\}$
Labelled expression	$pe$	$\in$	$\mathbb{P} \times \text{Expr}$
Labelled location	$\ell^p$	$\in$	$\mathbb{P} \times \text{Loc}$
Collecting semantics	$t$	$\in$	$\mathbb{T} \triangleq \mathbb{P} \rightarrow \mathcal{P}(\text{Env} + \text{Env} \times \text{Val})$
Labelled expression	$pe$	$\rightarrow$	$\{p : e\}$
Expression	$e$	$\rightarrow$	$x \mid \lambda x. pe \mid pe \text{ } pe \mid pe \rtimes pe \mid \varepsilon \mid x = pe; pe$

Step :  $\mathbb{T} \rightarrow \mathbb{T}$

$$\text{Step}(t) \triangleq \bigcup_{p \in \mathbb{P}} \text{step}(t, p)$$

step :  $(\mathbb{T} \times \mathbb{P}) \rightarrow \mathbb{T}$

$$\begin{aligned}
\text{step}(t, p) &\triangleq [p \mapsto \{(\sigma, v) \mid \sigma \in t(p) \text{ and } \sigma(x) = v\}] && \text{when } \{p : x\} \\
&\cup [p \mapsto \{(\sigma, v[\mu\ell^{p'}.v/\ell^{p'}]) \mid \sigma \in t(p) \text{ and } \sigma(x) = \mu\ell^{p'}.v\}] \\
\text{step}(t, p) &\triangleq [p \mapsto \{(\sigma, \langle \lambda x. p', \sigma \rangle) \mid \sigma \in t(p)\}] && \text{when } \{p : \lambda x. p'\} \\
\text{step}(t, p) &\triangleq [p_1 \mapsto \{\sigma \in \text{Env} \mid \sigma \in t(p)\}] && \text{when } \{p : p_1 \text{ } p_2\} \\
&\cup [p_2 \mapsto \{\sigma \in \text{Env} \mid \sigma \in t(p)\}] \\
&\cup \bigcup_{\sigma \in t(p)} \bigcup_{(\sigma, \langle \lambda x. p', \sigma_1 \rangle) \in t(p_1)} [p' \mapsto \{(x, v_2) :: \sigma_1 \mid (\sigma, v_2) \in t(p_2)\}] \\
&\cup [p \mapsto \bigcup_{\sigma \in t(p)} \bigcup_{(\sigma, \langle \lambda x. p', \sigma_1 \rangle) \in t(p_1)} \bigcup_{(\sigma, v_2) \in t(p_2)} \{(\sigma, v) \mid ((x, v_2) :: \sigma_1, v) \in t(p')\}] \\
&\cup [p \mapsto \bigcup_{\sigma \in t(p)} \{(\sigma, \text{Call}(S_1, v_2)) \mid (\sigma, S_1) \in t(p_1) \text{ and } (\sigma, v_2) \in t(p_2)\}] \\
\text{step}(t, p) &\triangleq [p_1 \mapsto \{\sigma \mid \sigma \in t(p)\}] && \text{when } \{p : p_1 \rtimes p_2\} \\
&\cup [p_2 \mapsto \bigcup_{\sigma \in t(p)} \{\sigma_1 \mid (\sigma, \sigma_1) \in t(p_1)\}] \\
&\cup [p \mapsto \bigcup_{\sigma \in t(p)} \bigcup_{(\sigma, \sigma_1) \in t(p_1)} \{(\sigma, v_2) \mid (\sigma_1, v_2) \in t(p_2)\}] \\
\text{step}(t, p) &\triangleq [p \mapsto \{(\sigma, \bullet) \mid \sigma \in t(p)\}] && \text{when } \{p : \varepsilon\} \\
\text{step}(t, p) &\triangleq [p_1 \mapsto \bigcup_{\sigma \in t(p)} \{(x, \ell^{p_1}) :: \sigma \mid \ell \notin \text{FLoc}(\sigma)\}] && \text{when } \{p : x = p_1; p_2\} \\
&\cup [p_2 \mapsto \bigcup_{\sigma \in t(p)} \{(x, \mu\ell^{p_1}.v_1) :: \sigma \mid ((x, \ell^{p_1}) :: \sigma, v_1) \in t(p_1)\}] \\
&\cup [p \mapsto \bigcup_{\sigma \in t(p)} \bigcup_{((x, \ell^{p_1}) :: \sigma, v_1) \in t(p_1)} \{(\sigma, (x, \mu\ell^{p_1}.v_1) :: \sigma_2) \mid ((x, \mu\ell^{p_1}.v_1) :: \sigma, \sigma_2) \in t(p_2)\}]
\end{aligned}$$

The collecting semantics  $\llbracket p_0 \rrbracket \Sigma_0$  computed by

$$\llbracket p_0 \rrbracket \Sigma_0 \triangleq \text{Ifp}(\lambda t. \text{Step}(t) \cup t_{\text{init}}) \quad \text{where } t_{\text{init}} = [p_0 \mapsto \Sigma_0]$$

contains all derivations of the form  $\sigma_0 \vdash p_0 \Downarrow v_0$  for some  $\sigma_0 \in \Sigma_0$  and  $v_0$ . That is,  $(\sigma, v)$  is contained in  $\llbracket p_0 \rrbracket \Sigma_0(p)$  if and only if  $\sigma \vdash p \Downarrow v$  is contained in some derivation for the judgment  $\sigma_0 \vdash p_0 \Downarrow v_0$ .

## 5.2 Abstract semantics

Abstract shadow	$S^\#$	$\in$	$\text{Shadow}^\#$
Abstract environment	$\sigma^\#$	$\in$	$\text{Env}^\# \triangleq (\text{Var} \xrightarrow{\text{fin}} \mathcal{P}(\mathbb{P})) \times \mathcal{P}(\text{Shadow}^\#)$
Abstract closure	$\langle \lambda x.p, p' \rangle$	$\in$	$\text{Clos}^\# \triangleq \text{Var} \times \mathbb{P} \times \mathbb{P}$
Abstract value	$v^\#$	$\in$	$\text{Val}^\# \triangleq \text{Env}^\# \times \mathcal{P}(\text{Clos}^\#)$
Abstract semantics	$t^\#$	$\in$	$\mathbb{T}^\# \triangleq \mathbb{P} \rightarrow \text{Env}^\# \times \text{Val}^\#$
Abstract shadow	$S^\#$	$\rightarrow$	$\text{Init}^\# \mid \text{Read}^\#(p, x) \mid \text{Call}^\#(p, p)$

$$\sigma \preceq (\sigma^\#, t^\#)$$

$$\begin{array}{c} \text{CONC-NIL} \quad \text{CONC-ENIL} \quad \text{CONC-CONSLoc} \\ \hline \bullet \preceq \sigma^\# \quad \frac{S \preceq (\sigma^\#, \emptyset)}{[S] \preceq \sigma^\#} \quad \frac{p \in \sigma^\#.1(x) \quad \sigma \preceq \sigma^\#}{(x, \ell^p) :: \sigma \preceq \sigma^\#} \end{array}$$

$$\begin{array}{c} \text{CONC-CONSWVAL} \\ \hline \frac{p \in \sigma^\#.1(x) \quad w \preceq t^\#(p).2 \quad \sigma \preceq \sigma^\#}{(x, w) :: \sigma \preceq \sigma^\#} \end{array}$$

$$w \preceq (v^\#, t^\#)$$

$$\begin{array}{c} \text{CONC-CLOS} \quad \text{CONC-REC} \\ \hline \frac{\langle \lambda x.p, p' \rangle \in v^\#.2 \quad \sigma \preceq t^\#(p').1}{\langle \lambda x.p, \sigma \rangle \preceq v^\#} \quad \frac{v \preceq t^\#(p).2 \quad v \preceq v^\#}{\mu \ell^p.v \preceq v^\#} \end{array}$$

$$\begin{array}{c} \text{CONC-INIT} \quad \text{CONC-READ} \\ \hline \frac{\text{Init}^\# \in v^\#.1.2}{\text{Init} \preceq v^\#} \quad \frac{\text{Read}^\#(p, x) \in v^\#.1.2 \quad [S] \preceq t^\#(p).1}{\text{Read}(S, x) \preceq v^\#} \end{array}$$

$$\begin{array}{c} \text{CONC-CALL} \\ \hline \frac{\text{Call}^\#(p_1, p_2) \in v^\#.1.2 \quad S \preceq t^\#(p_1).2 \quad v \preceq t^\#(p_2).2}{\text{Call}(S, v) \preceq v^\#} \end{array}$$

Figure 5: The concretization relation between weak values and abstract values.  $t^\#$  is omitted for brevity.

The concretization function  $\gamma \in \mathbb{T}^\# \rightarrow \mathbb{T}$  is defined as:

$$\gamma(t^\#) \triangleq \lambda p. \{\sigma \mid \sigma \preceq (t^\#(p).1, t^\#)\} \cup \{(\sigma, v) \mid v \preceq (t^\#(p).2, t^\#)\}$$

where  $\preceq$  is the concretization relation inductively defined in Figure 5.

Now the abstract semantic function can be given.

$$\boxed{\text{Step}^\# : \mathbb{T}^\# \rightarrow \mathbb{T}^\#}$$

$$\text{Step}^\#(t^\#) \triangleq \bigsqcup_{p \in \mathbb{P}} \text{step}^\#(t^\#, p)$$

$$\boxed{\text{step}^\# : (\mathbb{T}^\# \times \mathbb{P}) \rightarrow \mathbb{T}^\#}$$

$$\begin{aligned} \text{step}^\#(t^\#, p) &\triangleq [p \mapsto \bigsqcup_{p' \in t^\#(p).1.1(x)} (\perp, t^\#(p').2)] && \text{when } \{p : x\} \\ &\sqcup [p \mapsto (\perp, ((\perp, \{\text{Read}^\#(p, x)\}), \emptyset))] && \text{if } t^\#(p).1.2 \neq \emptyset \\ \text{step}^\#(t^\#, p) &\triangleq [p \mapsto (\perp, (\perp, \{\langle \lambda x. p', p \rangle\}))] && \text{when } \{p : \lambda x. p'\} \\ \text{step}^\#(t^\#, p) &\triangleq [p_1 \mapsto (t^\#(p).1, \perp)] && \text{when } \{p : p_1 \times p_2\} \\ &\sqcup [p_2 \mapsto (t^\#(p).1, \perp)] \\ &\sqcup \bigsqcup_{\langle \lambda x. p', p'' \rangle \in t^\#(p_1).2.2} [p' \mapsto (t^\#(p'').1 \sqcup ([x \mapsto \{p_2\}], \emptyset), \perp)] \\ &\sqcup [p \mapsto \bigsqcup_{\langle \lambda x. p', \_ \rangle \in t^\#(p_1).2.2} (\perp, t^\#(p').2)] \\ &\sqcup [p \mapsto (\perp, ((\perp, \{\text{Call}^\#(p_1, p_2)\}), \emptyset))] && \text{if } t^\#(p_1).2.1.2 \neq \emptyset \\ \text{step}^\#(t^\#, p) &\triangleq [p_1 \mapsto (t^\#(p).1, \perp)] && \text{when } \{p : p_1 \times p_2\} \\ &\sqcup [p_2 \mapsto (t^\#(p_1).2.1, \perp)] \\ &\sqcup [p \mapsto (\perp, t^\#(p_2).2)] \\ \text{step}^\#(t^\#, p) &\triangleq \perp && \text{when } \{p : \varepsilon\} \\ \text{step}^\#(t^\#, p) &\triangleq [p_1 \mapsto (t^\#(p).1 \sqcup ([x \mapsto \{p_1\}], \emptyset), \perp)] && \text{when } \{p : x = p_1; p_2\} \\ &\sqcup [p_2 \mapsto (t^\#(p).1 \sqcup ([x \mapsto \{p_1\}], \emptyset), \perp)] \\ &\sqcup [p \mapsto (\perp, (t^\#(p_2).2.1 \sqcup ([x \mapsto \{p_1\}], \emptyset), \emptyset))] \end{aligned}$$

The abstract semantics  $t^\#$  computed by

$$\llbracket p_0 \rrbracket^\#(\sigma_0^\#, t_0^\#) \triangleq \text{lfp}(\lambda t^\#. \text{Step}^\#(t^\#) \sqcup t_{\text{init}}^\#) \quad \text{where } t_{\text{init}}^\# = t_0^\# \sqcup [p_0 \mapsto (\sigma_0^\#, \perp)]$$

is a sound abstraction of  $\llbracket p_0 \rrbracket \Sigma_0$  when  $\Sigma_0 \subseteq \gamma(\sigma_0^\#, t_0^\#)$ .

### 5.3 Abstract linking

Now we define a sound linking operator that abstracts  $\times$ . Assume we have

$$\sigma_0 \preceq (\sigma_0^\#, t_0^\#) \quad t \subseteq \gamma(t^\#)$$

we define:

$$\sigma_0 \times t \triangleq \lambda p. \bigcup_{\sigma \in t(p)} (\sigma_0 \times \sigma) \cup \bigcup_{(\sigma, v) \in t(p)} \{(\sigma_+, v_+) | \sigma_+ \in \sigma_0 \times \sigma \text{ and } v_+ \in \sigma_0 \times v\}$$

We want to define  $\times^\#$  so that the following holds:

$$\sigma_0 \times t \subseteq \gamma((\sigma_0^\#, t_0^\#) \times^\# t^\#)$$

This is equivalent to saying that the linked result  $t_+^\# = (\sigma_0^\#, t_0^\#) \times^\# t^\#$  satisfies:

$$\sigma_0 \preceq (\sigma_0^\#, t_0^\#) \text{ and } w \preceq (v^\#, t^\#) \Rightarrow w_+ \preceq (v_+^\#, t_+^\#)$$

for each  $w_+ \in \sigma_0 \times w$  and  $p \in \mathbb{P}$ , when  $[v^\#, v_+^\#] = [(t^\#(p).1, \emptyset), (t_+^\#(p).1, \emptyset)]$  or  $[t^\#(p).2, t_+^\#(p).2]$ .

The condition for  $t_+^\#$  can be deduced by attempting the proof of the above in advance.

We proceed by induction on the derivation for

$$w_+ \in \sigma_0 \times w$$

and inversion on  $w \preceq (v^\#, t^\#)$ .

When:	$w = \text{Init},$
Have:	$\text{Init}^\# \in v^\#.1.2$
Need:	$v_+^\# \sqsupseteq \sigma_0^\#$ $t_+^\# \sqsupseteq t_0^\#$
When:	$w = \text{Read}(S, x),$
Have:	$\text{Read}^\#(p', x) \in v^\#.1.2$ and $[S] \preceq t^\#(p').1$
Need:	$v_+^\# \sqsupseteq t_+^\#(p'').2$ for $p'' \in t_+^\#(p').1.1(x)$ $v_+^\# \sqsupseteq (([], \{\text{Read}^\#(p', x)\}), \emptyset)$ if $t_+^\#(p').1.2 \neq \emptyset$
When:	$w = \text{Call}(S, v),$
Have:	$\text{Call}^\#(p_1, p_2) \in v^\#.1.2$ and $S \preceq t^\#(p_1).2$ and $v \preceq t^\#(p_2).2$
Need:	$v_+^\# \sqsupseteq t_+^\#(p').2$ for $\langle \lambda x.p', p'' \rangle \in t_+^\#(p_1).2.2$ $v_+^\# \sqsupseteq (([], \{\text{Call}^\#(p_1, p_2)\}), \emptyset)$ if $t_+^\#(p_1).2.1.2 \neq \emptyset$ $t_+^\#(p') \sqsupseteq (t_+^\#(p'').1 \sqcup ([x \mapsto \{p_2\}], \emptyset), \emptyset)$ for $\langle \lambda x.p', p'' \rangle \in t_+^\#(p_1).2.2$ $t_+^\# \sqsupseteq \text{Step}^\#(t_+^\#)$
When:	$w = (x, \ell^{p'}) :: \sigma,$
Have:	$p' \in v^\#.1.1(x)$ and $\sigma \preceq v^\#$
Need:	$v_+^\#.1.1(x) \ni p'$
When:	$w = (x, w') :: \sigma,$
Have:	$p' \in v^\#.1.1(x)$ and $w' \preceq t^\#(p').2$ and $\sigma \preceq v^\#$
Need:	$v_+^\#.1.1(x) \ni p'$
When:	$w = \langle \lambda x.p', \sigma \rangle,$
Have:	$\langle \lambda x.p', p'' \rangle \in v^\#.2$ and $\sigma \preceq t^\#(p'').1$
Need:	$v_+^\#.2 \ni \langle \lambda x.p', p'' \rangle$



The above conditions can be summarized by saying  $t_+^\#$  is a post-fixed point of:

$$\lambda t_+^\#. \text{Step}^\#(t_+^\#) \sqcup \text{Link}^\#(\sigma_0^\#, t^\#, t_+^\#) \sqcup t_0^\#$$

where  $\text{Link}^\#(\sigma_0^\#, t^\#, t_+^\#)$  is the least function that satisfies:

Let $\text{link}^\# = \text{Link}^\#(\sigma_0^\#, t^\#, t_+^\#)$ in For each $p \in \mathbb{P}$ , when $v_+^\#, v_+^\# = (t^\#(p).1, \emptyset), (\text{link}^\#(p).1, \emptyset)$ or when $v_+^\#, v_+^\# = t^\#(p).2, \text{link}^\#(p).2$	
If: $\text{Init}^\# \in v_+^\#.1.2$	
Then: $v_+^\# \sqsupseteq \sigma_0^\#$	
If: $\text{Read}^\#(p', x) \in v_+^\#.1.2$	
Then: $v_+^\# \sqsupseteq t_+^\#(p').2$	for $p'' \in t_+^\#(p').1.1(x)$
$v_+^\# \sqsupseteq ((\sqcup, \{\text{Read}^\#(p', x)\}), \emptyset)$	if $t_+^\#(p').1.2 \neq \emptyset$
If: $\text{Call}^\#(p_1, p_2) \in v_+^\#.1.2$	
Then: $v_+^\# \sqsupseteq t_+^\#(p').2$	for $\langle \lambda x.p', p'' \rangle \in t_+^\#(p_1).2.2$
$v_+^\# \sqsupseteq ((\sqcup, \{\text{Call}^\#(p_1, p_2)\}), \emptyset)$	if $t_+^\#(p_1).2.1.2 \neq \emptyset$
$\text{link}^\#(p') \sqsupseteq (t_+^\#(p'').1 \sqcup ([x \mapsto \{p_2\}], \emptyset), \emptyset)$	for $\langle \lambda x.p', p'' \rangle \in t_+^\#(p_1).2.2$
If: $p' \in v_+^\#.1.1(x)$	
Then: $v_+^\#.1.1(x) \ni p'$	
If: $p' \in v_+^\#.1.1(x)$	
Then: $v_+^\#.1.1(x) \ni p'$	
If: $\langle \lambda x.p', p'' \rangle \in v_+^\#.2$	
Then: $v_+^\#.2 \ni \langle \lambda x.p', p'' \rangle$	

Note that the left-hand side contains only  $\text{link}^\#$  and the right-hand side does not depend on the value of  $\text{link}^\#$ .

Some auxiliary lemmas:

**Lemma 5.1** (Substitution of values).

$$w \preceq (v^\#, t^\#) \text{ and } u \preceq (t^\#(p).2, t^\#) \Rightarrow w[u/\ell^p] \preceq (v^\#, t^\#)$$

**Lemma 5.2** (Sound step<sup>#</sup>).

$$\forall p, t, t^\# : t \subseteq \gamma(t^\#) \Rightarrow \text{step}(t, p) \cup t \subseteq \gamma(\text{step}^\#(t^\#, p) \sqcup t^\#)$$

**Lemma 5.3** (Sound Step<sup>#</sup>).

$$\forall t_{\text{init}}, t^\# : t_{\text{init}} \subseteq \gamma(t^\#) \text{ and } \text{Step}^\#(t^\#) \sqsubseteq t^\# \Rightarrow \text{lfp}(\lambda t. \text{Step}(t) \cup t_{\text{init}}) \subseteq \gamma(t^\#)$$

## 6 Conclusion