

A Simple Abstract Interpretation Framework for Modular Analysis

JOONHYUP LEE and KWANGKEUN YI

1 SYNTAX AND SEMANTICS

1.1 Abstract Syntax

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

Fig. 1. Abstract syntax of the language.

1.2 Operational Semantics

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

Fig. 2. Definition of the semantic domains.

1.3 Reconciling with Conventional Backpatching

The semantics in Figure 3 makes sense due to similarity with a conventional backpatching semantics as presented in Figure 5. We have defined a relation \sim that satisfies:

$$\sim \subseteq \text{WVal} \times (\text{MVal} \times \text{Mem} \times \mathcal{P}(\text{Loc})) \quad \bullet \sim (\bullet, \emptyset, \emptyset)$$

Authors' address: Joonhyup Lee; Kwangkeun Yi.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2024 Association for Computing Machinery.

XXXX-XXXX/2024/2-ART \$15.00

<https://doi.org/10.1145/nnnnnnnn.nnnnnnnn>

$$\begin{array}{c}
\boxed{\sigma \vdash e \Downarrow v} \\
\\
\text{ID} \quad \frac{\sigma(x) = v}{\sigma \vdash x \Downarrow v} \quad \text{RECI} \quad \frac{\sigma(x) = \mu\ell.v}{\sigma \vdash x \Downarrow v[\mu\ell.v/\ell]} \quad \text{FN} \quad \frac{}{\sigma \vdash \lambda x.e \Downarrow \langle \lambda x.e, \sigma \rangle} \quad \text{APP} \quad \frac{\sigma \vdash e_1 \Downarrow \langle \lambda x.e, \sigma_1 \rangle \quad \sigma \vdash e_2 \Downarrow v_2 \quad (x, v_2) :: \sigma_1 \vdash e \Downarrow v}{\sigma \vdash e_1 e_2 \Downarrow v} \\
\\
\text{LINK} \quad \frac{\sigma \vdash e_1 \Downarrow \sigma_1 \quad \sigma_1 \vdash e_2 \Downarrow v}{\sigma \vdash e_1 \bowtie e_2 \Downarrow v} \quad \text{EMPTY} \quad \frac{}{\sigma \vdash \varepsilon \Downarrow \bullet} \quad \text{BIND} \quad \frac{\ell \notin \text{FLoc}(\sigma) \quad (x, \ell) :: \sigma \vdash e_1 \Downarrow v_1 \quad (x, \mu\ell.v_1) :: \sigma \vdash e_2 \Downarrow \sigma_2}{\sigma \vdash x = e_1; e_2 \Downarrow (x, \mu\ell.v_1) :: \sigma_2}
\end{array}$$

Fig. 3. The big-step operational semantics.

Environment	σ	\in	MEnv	
Memory	m	\in	$\text{Mem} \triangleq \text{Loc} \xrightarrow{\text{fin}} \text{MVal}$	
Value	v	\in	$\text{MVal} \triangleq \text{MEnv} + \text{Var} \times \text{Expr} \times \text{MEnv}$	
Environment	σ	\rightarrow	\bullet	empty stack
		$ $	$(x, \ell) :: \sigma$	location binding
Value	v	\rightarrow	σ	exported environment
		$ $	$\langle \lambda x.e, \sigma \rangle$	closure

Fig. 4. Definition of the semantic domains with memory.

and the following theorem:

THEOREM 1.1 (EQUIVALENCE OF SEMANTICS). For all $\sigma \in \text{Env}$, $\sigma' \in \text{MEnv} \times \text{Mem} \times \mathcal{P}(\text{Loc})$, $v \in \text{Val}$, $v' \in \text{MVal} \times \text{Mem} \times \mathcal{P}(\text{Loc})$, we have:

$$\begin{aligned}
\sigma \sim \sigma' \text{ and } \sigma \vdash e \Downarrow v &\Rightarrow \exists v' : v \sim v' \text{ and } \sigma' \vdash e \Downarrow v' \\
\sigma \sim \sigma' \text{ and } \sigma' \vdash e \Downarrow v' &\Rightarrow \exists v : v \sim v' \text{ and } \sigma \vdash e \Downarrow v
\end{aligned}$$

The actual definition for \sim can be found in the appendix.

$$\begin{array}{c}
\boxed{\sigma, m, L \vdash e \Downarrow v, m', L'} \\
\\
\text{ID} \quad \frac{\sigma(x) = \ell \quad m(\ell) = v}{\sigma, m, L \vdash x \Downarrow v, m, L} \quad \text{FN} \quad \frac{}{\sigma, m, L \vdash \lambda x.e \Downarrow \langle \lambda x.e, \sigma \rangle, m, L} \\
\\
\text{APP} \quad \frac{\sigma, m, L \vdash e_1 \Downarrow \langle \lambda x.e, \sigma_1 \rangle, m_1, L_1 \quad \sigma, m_1, L_1 \vdash e_2 \Downarrow v_2, m_2, L_2 \quad \ell \notin \text{dom}(m_2) \cup L_2 \quad (x, \ell) :: \sigma_1, m_2[\ell \mapsto v_2], L_2 \vdash e \Downarrow v, m', L'}{\sigma, m, L \vdash e_1 e_2 \Downarrow v, m', L'} \\
\\
\text{LINK} \quad \frac{\sigma, m, L \vdash e_1 \Downarrow \sigma_1, m_1, L_1 \quad \sigma_1, m_1, L_1 \vdash e_2 \Downarrow v, m', L'}{\sigma, m, L \vdash e_1 \times e_2 \Downarrow v, m', L'} \quad \text{EMPTY} \quad \frac{}{\sigma, m, L \vdash \varepsilon \Downarrow \bullet, m, L} \\
\\
\text{BIND} \quad \frac{\ell \notin \text{dom}(m) \cup L \quad (x, \ell) :: \sigma, m, L \cup \{\ell\} \vdash e_1 \Downarrow v_1, m_1, L_1 \quad (x, \ell) :: \sigma, m_1[\ell \mapsto v_1], L_1 \vdash e_2 \Downarrow \sigma_2, m', L'}{\sigma, m, L \vdash x = e_1; e_2 \Downarrow (x, \ell) :: \sigma_2, m', L'}
\end{array}$$

Fig. 5. The big-step operational semantics with memory.

2 GENERATING AND RESOLVING EVENTS

Now we formulate the semantics for generating events.

Environment	σ	\rightarrow	\dots	
			$[E]$	answer to an event
Value	v	\rightarrow	\dots	
			E	answer to an event
Event	E	\rightarrow	Init	initial environment
			Read(E, x)	read event
			Call(E, v)	call event

Fig. 6. Definition of the semantic domains with events. All other semantic domains are equal to Figure 2.

We redefine how to read weak values given an environment.

$$\begin{aligned}
 \bullet(x) &\triangleq \perp & ((x, w) :: \sigma)(x) &\triangleq w \\
 ((x, \ell) :: \sigma)(x) &\triangleq \perp & ((x', _) :: \sigma)(x) &\triangleq \sigma(x) \quad (x \neq x') \\
 [E](x) &\triangleq \text{Read}(E, x)
 \end{aligned}$$

Then we need to add only one rule to the semantics in Figure 3 for the semantics to incorporate events.

$$\frac{\text{APPEVENT} \quad \sigma \vdash e_1 \Downarrow E \quad \sigma \vdash e_2 \Downarrow v}{\sigma \vdash e_1 e_2 \Downarrow \text{Call}(E, v)}$$

Now we need to formulate the *concrete linking* rules. The concrete linking rule $\sigma_0 \bowtie w$, given an answer σ_0 to the Init event, resolves all events within w to obtain a set of final results.

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

$$\text{eval}(e, \sigma) \triangleq \{v \mid \sigma \vdash e \Downarrow v\} \quad \text{eval}(e, \Sigma) \triangleq \bigcup_{\sigma \in \Sigma} \text{eval}(\sigma, e) \quad \sigma_0 \bowtie W \triangleq \bigcup_{w \in W} (\sigma_0 \bowtie w)$$

Then the following holds:

THEOREM 2.1 (SOUNDNESS OF CONCRETE LINKING). Given $e \in \text{Expr}, \sigma \in \text{Env}, v \in \text{Val}$,

$$\forall \sigma_0 \in \text{Env} : \text{eval}(e, \sigma_0 \bowtie \sigma) \subseteq \sigma_0 \bowtie \text{eval}(e, \sigma)$$

$$\bowtie \in \text{Env} \rightarrow \text{Event} \rightarrow \mathcal{P}(\text{Val})$$

$$\sigma_0 \bowtie \text{Init} \triangleq \{\sigma_0\}$$

$$\sigma_0 \bowtie \text{Read}(E, x) \triangleq \{V \mid \Sigma \in \sigma_0 \bowtie E \text{ and } \Sigma(x) = V\} \cup$$

$$\{V[\mu\ell.V/\ell] \mid \Sigma \in \sigma_0 \bowtie E \text{ and } \Sigma(x) = \mu\ell.V\}$$

$$\sigma_0 \bowtie \text{Call}(E, v) \triangleq \{V' \mid \langle \lambda x.e, \Sigma \rangle \in \sigma_0 \bowtie E \text{ and } V \in \sigma_0 \bowtie v \text{ and } (x, V) :: \Sigma \vdash e \Downarrow V'\} \cup$$

$$\{\text{Call}(E', V) \mid E' \in \sigma_0 \bowtie E \text{ and } V \in \sigma_0 \bowtie v\}$$

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

$$\sigma_0 \bowtie \bullet \triangleq \{\bullet\}$$

$$\sigma_0 \bowtie (x, \ell) :: \sigma \triangleq \{(x, \ell) :: \Sigma \mid \Sigma \in \sigma_0 \bowtie \sigma\}$$

$$\sigma_0 \bowtie (x, w) :: \sigma \triangleq \{(x, W) :: \Sigma \mid W \in \sigma_0 \bowtie w \text{ and } \Sigma \in \sigma_0 \bowtie \sigma\}$$

$$\sigma_0 \bowtie [E] \triangleq \{\Sigma \in \text{Env} \mid \Sigma \in \sigma_0 \bowtie E\} \cup$$

$$\{[E'] \mid E' \in \sigma_0 \bowtie E\}$$

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

$$\sigma_0 \bowtie \langle \lambda x.e, \sigma \rangle \triangleq \{\langle \lambda x.e, \Sigma \rangle \mid \Sigma \in \sigma_0 \bowtie \sigma\}$$

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

$$\sigma_0 \bowtie \mu\ell.v \triangleq \{\mu\ell.V \mid V \in \sigma_0 \bowtie v\}$$

Fig. 7. Definition for concrete linking.

3 TYPING

The definitions for types are in Figure 8 and the typing rules are in Figure 9. The definitions for subtyping are in Figure 10.

Types	$\tau \rightarrow \Gamma$	module type
	$\mid \tau \rightarrow \tau$	function type
Typing Environment	$\Gamma \rightarrow \bullet$	empty environment
	$\mid (x, \tau) :: \Gamma$	type binding

Fig. 8. Definition of types.

T-ID	T-FN	T-APP
$\frac{\Gamma(x) = \tau}{\Gamma \vdash x : \tau}$	$\frac{(x, \tau_1) :: \Gamma \vdash e : \tau_2}{\Gamma \vdash \lambda x. e : \tau_1 \rightarrow \tau_2}$	$\frac{\Gamma \vdash e_1 : \tau_1 \rightarrow \tau \quad \Gamma \vdash e_2 : \tau_2 \quad \tau_1 \geq \tau_2}{\Gamma \vdash e_1 e_2 : \tau}$
T-LINK	T-NIL	T-BIND
$\frac{\Gamma \vdash e_1 : \Gamma_1 \quad \Gamma_1 \vdash e_2 : \tau_2}{\Gamma \vdash e_1 \bowtie e_2 : \tau_2}$	$\frac{}{\Gamma \vdash \varepsilon : \bullet}$	$\frac{\Gamma \vdash e_1 : \tau_1 \quad (x, \tau_1) :: \Gamma \vdash e_2 : \Gamma_2}{\Gamma \vdash x = e_1; e_2 : (x, \tau_1) :: \Gamma_2}$

Fig. 9. The typing judgment.

NIL	CONSFREE	CONSBOUND	ARROW
$\frac{}{\bullet \geq \bullet}$	$\frac{x \notin \text{dom}(\Gamma) \quad \Gamma \geq \Gamma'}{\Gamma \geq (x, \tau) :: \Gamma'}$	$\frac{\Gamma(x) \geq \tau \quad \Gamma - x \geq \Gamma'}{\Gamma \geq (x, \tau) :: \Gamma'}$	$\frac{\tau_2 \geq \tau_1 \quad \tau'_1 \geq \tau'_2}{\tau_1 \rightarrow \tau'_1 \geq \tau_2 \rightarrow \tau'_2}$

Fig. 10. The subtype relation.

3.1 Type Safety

THEOREM 3.1 (TYPE SAFETY). For all $e \in \text{Expr}$, if $\bullet \vdash e : \tau$ for some τ , then there exists some $v \in \text{Val}$ such that $\bullet \vdash e \Downarrow v$.

SKETCH. We prove this through unary logical relations and induction on the typing judgment.

Value Relation

$$\begin{aligned}
 \mathcal{V}[\bullet] &\triangleq \text{Env} \\
 \mathcal{V}[(x, \tau) :: \Gamma] &\triangleq \{\sigma \mid \sigma(x) \in \mathcal{V}[\tau] \text{ and } \sigma \in \mathcal{V}[\Gamma - x]\} \\
 \mathcal{V}[\tau_1 \rightarrow \tau_2] &\triangleq \{\langle \lambda x. e, \sigma \rangle \mid \forall v \in \mathcal{V}[\tau_1] : (e, (x, v) :: \sigma) \in \mathcal{E}[\tau_2]\}
 \end{aligned}$$

Expression Relation

$$\mathcal{E}[\tau] \triangleq \{(e, \sigma) \mid \exists v \in \mathcal{V}[\tau] : \sigma \vdash e \Downarrow v\}$$

Semantic Typing

$$\Gamma \models e : \tau \triangleq \forall \sigma \in \mathcal{V}[\Gamma] : (e, \sigma) \in \mathcal{E}[\tau]$$

We prove

$$\Gamma \vdash e : \tau \Rightarrow \Gamma \models e : \tau$$

by induction on \vdash . □

3.2 Type Inference

When modules are first-class, type variables can go in the place of type environments.

First we define the syntax for type constraints.

Type Variable	α	\in	TyVar	
Path	p	\rightarrow	ϵ	empty string
			px	concatenation with identifier
Types	τ	\rightarrow	$\Gamma \mid \tau \rightarrow \tau$	module/function types
Type Environment	Γ	\rightarrow	\bullet	empty environment
			$(x, \tau) :: \Gamma$	binding
			$\alpha.p$	type variable
			$[\] . p$	type read from hole
Type Constraint	u	\rightarrow	$\tau \doteq \tau$	equality constraint
			$\tau \dot{\geq} \tau$	subtyping constraint
Set of Constraints	U	\subseteq	$\{u \mid u \text{ type constraint}\}$	

Fig. 11. Definition of type constraints.

Next we define the type access operation $\tau(x)$:

$$\begin{aligned}
 \bullet(x) &\triangleq \perp & (\alpha.p)(x) &\triangleq \alpha.px \\
 ((x, \tau) :: _)(x) &\triangleq \tau & ([\] . p)(x) &\triangleq [\] . px \\
 ((x', _) :: \Gamma)(x) &\triangleq \Gamma(x) & \text{when } x' \neq x & (_ \rightarrow _)(x) \triangleq \perp
 \end{aligned}$$

Now we can define the constraint generation algorithm $V(\Gamma, e, \alpha)$. Note that the **let** $U = _ \text{ in } _$ notation returns \perp if the right hand side is \perp . Also note that we write α for $\alpha.\epsilon$ as well.

$$V(\Gamma, e, \alpha) = U$$

$$\begin{aligned}
 V(\Gamma, \epsilon, \alpha) &\triangleq \{\alpha \doteq \bullet\} & V(\Gamma, e_1 \bowtie e_2, \alpha) &\triangleq \text{let } \alpha_1 = \text{fresh in} \\
 V(\Gamma, x, \alpha) &\triangleq \text{let } \tau = \Gamma(x) \text{ in} & & \text{let } U_1 = V(\Gamma, e_1, \alpha_1) \text{ in} \\
 & \{\alpha \doteq \tau\} & & \text{let } U_2 = V(\alpha_1, e_2, \alpha) \text{ in} \\
 V(\Gamma, \lambda x. e, \alpha) &\triangleq \text{let } \alpha_1, \alpha_2 = \text{fresh in} & & U_1 \cup U_2 \\
 & \text{let } U = V((x, \alpha_1) :: \Gamma, e, \alpha_2) \text{ in} & V(\Gamma, x = e_1; e_2, \alpha) &\triangleq \text{let } \alpha_1, \alpha_2 = \text{fresh in} \\
 & \{\alpha \doteq \alpha_1 \rightarrow \alpha_2\} \cup U & & \text{let } U_1 = V(\Gamma, e_1, \alpha_1) \text{ in} \\
 V(\Gamma, e_1 \ e_2, \alpha) &\triangleq \text{let } \alpha_1, \alpha_2, \alpha_3 = \text{fresh in} & & \text{let } U_2 = V((x, \alpha_1) :: \Gamma, e_2, \alpha_2) \text{ in} \\
 & \text{let } U_1 = V(\Gamma, e_1, \alpha_1) \text{ in} & & \{\alpha \doteq (x, \alpha_1) :: \alpha_2\} \cup U_1 \cup U_2 \\
 & \text{let } U_2 = V(\Gamma, e_2, \alpha_2) \text{ in} & & \\
 & \{\alpha_1 \doteq \alpha_3 \rightarrow \alpha, \alpha_3 \dot{\geq} \alpha_2\} \cup U_1 \cup U_2
 \end{aligned}$$

We want to prove that the constraint generation algorithm is correct.

First, for $\tau \in \text{Type}$, define the path access operation $\tau(p)$:

$$\begin{aligned}
 \tau(\epsilon) &\triangleq \tau & \tau(px) &\triangleq \tau(p)(x)
 \end{aligned}$$

and define the injection operation $\tau[\Gamma_0]$:

$$\begin{aligned} (\bullet)[\Gamma_0] &\triangleq \bullet & ((x, \tau) :: \Gamma)[\Gamma_0] &\triangleq (x, \tau[\Gamma_0]) :: \Gamma[\Gamma_0] \\ (\alpha.p)[\Gamma_0] &\triangleq \alpha.p & ([\cdot].p)[\Gamma_0] &\triangleq \Gamma_0(p) \\ (\tau_1 \rightarrow \tau_2)[\Gamma_0] &\triangleq \tau_1[\Gamma_0] \rightarrow \tau_2[\Gamma_0] \end{aligned}$$

Let $\text{Subst} \triangleq \text{TyVar} \xrightarrow{\text{fin}} \text{Type}$ be the set of substitutions. For $S \in \text{Subst}$, define:

$$\begin{aligned} S\bullet &\triangleq \bullet & S(\tau_1 \rightarrow \tau_2) &\triangleq S\tau_1 \rightarrow S\tau_2 \\ S(\alpha.p) &\triangleq \alpha.p & \text{when } \alpha \notin \text{dom}(S) & & S(\alpha.p) &\triangleq \tau(p) & \text{when } \alpha \mapsto \tau \in S \\ S([\cdot].p) &\triangleq [\cdot].p \end{aligned}$$

Define:

$$\begin{aligned} (S, \Gamma_0) \models U &\triangleq \forall (\tau_1 \dot{=} \tau_2) \in U : (S\tau_1)[\Gamma_0] = (S\tau_2)[\Gamma_0] \text{ and} \\ &\forall (\tau_1 \dot{\geq} \tau_2) \in U : (S\tau_1)[\Gamma_0] \geq (S\tau_2)[\Gamma_0] \end{aligned}$$

where subtyping rules are the same as Figure 10 and subtyping between type variables are not defined.

Then we can show that:

THEOREM 3.2 (CORRECTNESS OF V). For $e \in \text{Expr}$, $\Gamma, \Gamma_0 \in \text{TyEnv}$, $\alpha \in \text{TyVar}$, $S \in \text{Subst}$:

$$(S, \Gamma_0) \models V(\Gamma, e, \alpha) \Leftrightarrow (S\Gamma)[\Gamma_0] \vdash e : (S\alpha)[\Gamma_0]$$

SKETCH. Structural induction on e . □

Note that by including $[\cdot].p$ in type environments, we can naturally generate constraints about the external environment $[\cdot]$. Also, by injection, we can utilize constraints generated *in advance* to obtain constraints generated from a more informed environment. We extend injection to the output of the constraint-generating algorithm:

$$\begin{aligned} \perp[\Gamma_0] &\triangleq \perp \\ U[\Gamma_0] &\triangleq \{\tau_1[\Gamma_0] \dot{=} \tau_2[\Gamma_0] \mid (\tau_1 \dot{=} \tau_2) \in U\} \cup \\ &\quad \{\tau_1[\Gamma_0] \dot{\geq} \tau_2[\Gamma_0] \mid (\tau_1 \dot{\geq} \tau_2) \in U\} & \text{when all injections succeed} \\ U[\Gamma_0] &\triangleq \perp & \text{when injection fails} \end{aligned}$$

Then we can prove:

THEOREM 3.3 (ADVANCE). For $e \in \text{Expr}$, $\Gamma, \Gamma_0 \in \text{TyEnv}$, $\alpha \in \text{TyVar}$:

$$V(\Gamma[\Gamma_0], e, \alpha) = V(\Gamma, e, \alpha)[\Gamma_0]$$

SKETCH. Structural induction on Γ . □

REFERENCES