Structure of Programming Languages: Type Safety for STLC with Constants, Booleans, and Conditionals

This is similar to the proof from Lecture 10, extended for cases for handling booleans and conditionals. The additions are highlighted in green.

Syntax

\[
\begin{align*}
e & ::= c \mid \lambda x. e \mid x \mid e \ e \mid \text{true} \mid \text{false} \mid \text{if} \ e \ e_2 \ e_3 \\
v & ::= c \mid \lambda x. e \mid \text{true} \mid \text{false} \\
\tau & ::= \text{int} \mid \tau \rightarrow \tau \mid \text{bool} \\
\Gamma & ::= \cdot \mid \Gamma, x : \tau
\end{align*}
\]

Evaluation Rules (a.k.a. Dynamic Semantics)

\[
ie \rightarrow e'
\]

\[
\begin{align*}
\text{E-Apply} & : (\lambda x. e) \ v \rightarrow e[v/x] \\
\text{E-If1} & : e_1 \rightarrow e'_1 \\
\text{E-If2} & : e_1 \rightarrow e'_1 \ e_2 \\
\text{E-If3} & : e_2 \rightarrow e'_2
\end{align*}
\]

Typing Rules (a.k.a. Static Semantics)

\[
\Gamma \vdash e : \tau
\]

\[
\begin{align*}
\text{T-Const} & : \Gamma \vdash c : \text{int} \\
\text{T-Var} & : \Gamma \vdash x : \Gamma(x) \\
\text{T-Fun} & : \Gamma, x : \tau_1 \vdash e : \tau_2 \quad x \not\in \text{Dom}(\Gamma) \\
\text{T-App} & : \Gamma \vdash e_1 : \tau_2 \rightarrow \tau_1 \quad \Gamma \vdash e_2 : \tau_2 \\
\text{T-If} & : \Gamma \vdash e_1 : \text{bool} \quad \Gamma \vdash e_2 : \tau \quad \Gamma \vdash e_3 : \tau \\
\text{T-True} & : \Gamma \vdash \text{true} : \text{bool} \\
\text{T-False} & : \Gamma \vdash \text{false} : \text{bool}
\end{align*}
\]
Type Soundness

**Theorem** (Type Soundness). If $\cdot \vdash e : \tau$ and $e \rightarrow^* e'$, then either $e'$ is a value or there exists an $e''$ such that $e' \rightarrow e''$.

**Proof**

The Type Soundness Theorem follows as a simple corollary to the Progress and Preservation Theorems stated and proven below: Given the Preservation Theorem, a trivial induction on the number of steps taken to reach $e'$ from $e$ establishes that $\cdot \vdash e' : \tau$. Then the Progress Theorem ensures $e'$ is a value or can step to some $e''$.

We need the following lemma for our proof of Progress, below.

**Lemma** (Canonical Forms). If $\cdot \vdash v : \tau$, then

i If $\tau$ is int, then $v$ is a constant, i.e., some $c$.

ii If $\tau$ is $\tau_1 \rightarrow \tau_2$, then $v$ is a lambda, i.e., $\lambda x. e$ for some $x$ and $e$.

iii If $\tau$ is bool, then $v$ is either true or false.

**Canonical Forms.** The proof is by inspection of the typing rules.

i If $\tau$ is int, then the only rule which lets us give a value this type is $T$-CONST.

ii If $\tau$ is $\tau_1 \rightarrow \tau_2$, then the only rule which lets us give a value this type is $T$-FUN.

iii If $\tau$ is bool, then the only rules which let us have a value this type are $T$-TRUE and $T$-FALSE, therefore the value is either true or false.

**Theorem** (Progress). If $\cdot \vdash e : \tau$, then either $e$ is a value or there exists some $e'$ such that $e \rightarrow e'$.

**Progress.** The proof is by induction on (the height of) the derivation of $\cdot \vdash e : \tau$, proceeding by cases on the bottommost rule used in the derivation.

$T$-CONST $e$ is a constant, which is a value, so we are done.

$T$-VAR Impossible, as $\Gamma$ is $\cdot$.

$T$-FUN $e$ is $\lambda x. e'$, which is a value, so we are done.
T-App \( e \) is \( e_1 \ e_2 \).

By inversion, \( \cdot \vdash e_1 : \tau' \rightarrow \tau \) and \( \cdot \vdash e_2 : \tau' \) for some \( \tau' \).

If \( e_1 \) is not a value, then \( \cdot \vdash e_1 : \tau' \rightarrow \tau \) and the induction hypothesis ensures \( e_1 \rightarrow e'_1 \) for some \( e'_1 \). Therefore, by E-App1, \( e_1 \ e_2 \rightarrow e'_1 \ e_2 \).

Else \( e_1 \) is a value. If \( e_2 \) is not a value, then \( \cdot \vdash e_2 : \tau' \) and our induction hypothesis ensures \( e_2 \rightarrow e'_2 \) for some \( e'_2 \). Therefore, by E-App2, \( e_1 \ e_2 \rightarrow e_1 \ e'_2 \).

Else \( e_1 \) and \( e_2 \) are values. Then \( \cdot \vdash e_1 : \tau' \rightarrow \tau \) and the Canonical Forms Lemma ensures \( e_1 \) is some \( \lambda x. \ e' \). And \( (\lambda x. \ e') \ e_2 \rightarrow e'[e_2/x] \) by E-Apply, so \( e_1 \ e_2 \) can take a step.

T-True \( e \) is true, which is a value, so we are done.

T-False \( e \) is false, which is a value, so we are done.

T-If \( e \) is if \( e_1 \ e_2 \ e_3 \).

By inversion, \( \Gamma \vdash e_1 : \text{bool}, \Gamma \vdash e_2 : \tau' \), and \( \Gamma \vdash e_3 : \tau' \) for some \( \tau' \).

If \( e_1 \) is not a value, then \( \cdot \vdash e_1 : \text{bool} \), and the induction hypothesis ensures \( e_1 \rightarrow e'_1 \) for some \( e'_1 \). Therefore, by E-If1, \( \text{if} \ e_1 \ e_2 \ e_3 \rightarrow \text{if} \ e'_1 \ e_2 \ e_3 \).

Else \( e_1 \) is a value. Then \( \cdot \vdash e_1 : \tau' \rightarrow \tau \) and the Canonical Forms Lemma ensures \( e_1 \) is either true or false. By E-If2, if \( \text{true} \ e_2 \ e_3 \) takes a step to \( e_2 \). By E-If3, if \( \text{false} \ e_2 \ e_3 \) takes a step to \( e_3 \).

We will need the following lemma for our proof of Preservation, below. Actually, in the proof of Preservation, we need only a Substitution Lemma where \( \Gamma \) is \( \cdot \), but proving the Substitution Lemma itself requires the stronger induction hypothesis using any \( \Gamma \).

**Lemma (Substitution).** If \( \Gamma, x: \tau' \vdash e : \tau \) and \( \Gamma \vdash e' : \tau' \), then \( \Gamma \vdash e[e'/x] : \tau \).

To prove this lemma, we will need the following two technical lemmas, which we will assume without proof (they’re not that difficult).

**Lemma (Weakening).** If \( \Gamma \vdash e : \tau \) and \( x \notin \text{Dom}(\Gamma) \), then \( \Gamma, x: \tau' \vdash e : \tau \).

**Lemma (Exchange).** If \( \Gamma, x: \tau_1, y: \tau_2 \vdash e : \tau \) and \( y \neq x \), then \( \Gamma, y: \tau_2, x: \tau_1 \vdash e : \tau \).

Now we prove Substitution.

**Substitution.** The proof is by induction on the derivation of \( \Gamma, x: \tau' \vdash e : \tau \). There are four cases. In all cases, we know \( \Gamma \vdash e' : \tau' \) by assumption.

T-Const \( e \) is \( c \), so \( c[e'/x] \) is \( c \). By T-Const, \( \Gamma \vdash c : \text{int} \).
Theorem (Preservation). If \( \cdot \vdash e : \tau \) and \( e \rightarrow e' \), then \( \cdot \vdash e' : \tau \).
Preservation. The proof is by induction on the derivation of $\cdot \vdash e : \tau$. There are four cases.

**T-Const** $e$ is $c$. This case is impossible, as there is no $e'$ such that $c \rightarrow e'$.

**T-Var** $e$ is $x$. This case is impossible, as $x$ cannot be typechecked under the empty context.

**T-Fun** $e$ is $\lambda x. e_b$. This case is impossible, as there is no $e'$ such that $\lambda x. e_b \rightarrow e'$.

**T-App** $e$ is $e_1 e_2$, so $\cdot \vdash e_1 e_2 : \tau$.

By inversion on the typing rule, $\cdot \vdash e_1 : \tau_2 \rightarrow \tau$ and $\cdot \vdash e_2 : \tau_2$ for some $\tau_2$.

There are three possible rules for deriving $e_1 e_2 \rightarrow e'$.

**E-App1** Then $e' = e'_1 e_2$ and $e_1 \rightarrow e'_1$.

By $\cdot \vdash e_1 : \tau_2 \rightarrow \tau$, $e_1 \rightarrow e'_1$, and induction, $\cdot \vdash e'_1 : \tau_2 \rightarrow \tau$.

Using this and $\cdot \vdash e_2 : \tau_2$, T-App lets us derive $\cdot \vdash e'_1 e_2 : \tau$.

**E-App2** Then $e' = e_1 e'_2$ and $e_2 \rightarrow e'_2$.

By $\cdot \vdash e_2 : \tau_2$, $e_2 \rightarrow e'_2$, and induction $\cdot \vdash e'_2 : \tau_2$.

Using this and $\cdot \vdash e_1 : \tau_2 \rightarrow \tau$, T-App lets us derive $\cdot \vdash e_1 e'_2 : \tau$.

**E-Apply** Then $e_1$ is $\lambda x. e_b$ for some $x$ and $e_b$, and $e' = e_b[e_2/x]$.

By inversion of the typing of $\cdot \vdash e_1 : \tau_2 \rightarrow \tau$, we have $\cdot, x : \tau_2 \vdash e_b : \tau$.

This and $\cdot \vdash e_2 : \tau_2$ lets us use the Substitution Lemma to conclude $\cdot \vdash e_b[e_2/x] : \tau$.

**T-True** $e$ is $\text{true}$. This case is impossible, as there is no $e'$ such that $\text{true} \rightarrow e'$.

**T-False** $e$ is $\text{false}$. This case is impossible, as there is no $e'$ such that $\text{false} \rightarrow e'$.

**T-If** $e$ is if $e_1 e_2 e_3$, so $\cdot \vdash$ if $e_1 e_2 e_3 : \tau$.

By inversion on T-If, we have $\Gamma \vdash e_1 : \text{bool}$, $\Gamma \vdash e_2 : \tau$, and $\Gamma \vdash e_3 : \tau$.

There are three possible rules for deriving if $e_1 e_2 e_3 \rightarrow e'$

**E-If1** Then $e = \text{if} e_1 e_2 e_3$ and $e' = \text{if} e'_1 e_2 e_3$. By $\Gamma \vdash e_1 : \text{bool}$ and $e_1 \rightarrow e'_1$, $\Gamma \vdash e'_1 : \text{bool}$. By $\Gamma \vdash e'_1 : \text{bool}$, $\Gamma \vdash e_2 : \tau$, and $\Gamma \vdash e_3 : \tau$ and induction, using T-If lets us derive $\Gamma \vdash e'_1 e_2 e_3 : \tau$, hence $\Gamma \vdash e' : \tau$

**E-If2** Then $e = \text{if} \text{true} e_2 e_3$ and $e' = e_2$. By $\Gamma \vdash e_2 : \tau$ and induction, $\Gamma \vdash e' : \tau$.

**E-If3** Then $e = \text{if} \text{false} e_2 e_3$ and $e' = e_3$. By $\Gamma \vdash e_3 : \tau$ and induction, $\Gamma \vdash e' : \tau$.