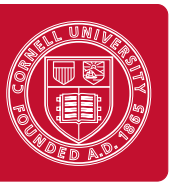
CS 4110 – Programming Languages and Logics
Lecture #24: Featherweight Java Properties



Based on material by Stephen Chong, Greg Morrisett, Andrew Myers, George Necula, and Radu Rugina

In this lecture, we will develop a proof of type soundness for Featherweight Java in the usual way, as a corollary of progress and preservation. The details of these proofs will be a little different than the ones we have seen before, however, due to the presence of subtyping and casts.

1 Preservation

The proof of preservation relies on several supporting lemmas.

Lemma (Method Typing). *If* $mtype(m, C) = \overline{D} \to D$ *and* $mbody(m, C) = (\overline{x}, e)$ *then there exists types* C' *and* D' *such that* $\overline{x} : \overline{D}, \texttt{this} : C' \vdash e : D'$ *and* $D' \leq D$.

Lemma (Substitution). *If* $\Gamma, \overline{x} : \overline{B} \vdash e : C$ *and* $\Gamma \vdash \overline{u} : \overline{B'}$ *with* $\overline{B'} \leq \overline{B}$ *then there exists* C' *such that* $\Gamma \vdash [\overline{x} \mapsto \overline{u}]e : C'$ *and* $C' \leq C$.

Lemma (Weakening). *If* $\Gamma \vdash e : C$ *then* $\Gamma, x : B \vdash e : C$.

Lemma (Decomposition). *If* $\Gamma \vdash E[e] : C$ *then there exists a type* B *such that* $\Gamma \vdash e : B$

Lemma (Context). *If* $\Gamma \vdash E[e] : C$ *and* $\Gamma \vdash e : B$ *and* $\Gamma \vdash e' : B'$ *with* $B' \leq B$ *then there exists a type* C' *such that* $\Gamma \vdash E[e'] : C'$ *and* $C' \leq C$.

Lemma (Preservation). *If* $\Gamma \vdash e : C$ *and* $e \to e'$ *then there exists a type* C' *such that* $\Gamma \vdash e' : C'$ *and* $C' \leq C$.

Proof. By induction on $e \to e'$, with a case analysis of the last rule used in the derivation.

Case E-CONTEXT: $e = E[e_1]$ and $e_1 \to e_1'$ and $e' = E[e_1']$

> By the decomposition lemma we have that there exists a type B such that $\Gamma \vdash e_1 : B$. By the induction hypothesis applied to e_1 we have that there exists a type B' such that $\Gamma \vdash e_1' : B'$ and $B' \leq B$. Then, by the context lemma we have that there exists a type C' such that $\Gamma \vdash E[e_1'] : C'$ and $C' \leq C$, as required.

Case E-PROJ: $e = \texttt{new}\ C_0(\overline{v}).f_i$ and $e' = v_i$ with $fields(C_0) = \overline{C\ f}$

> As the typing rules for Featherweight Java are syntax-directed, the last rule used in the derivation of $\Gamma \vdash e : C$ must have been T-FIELD. Therefore we must also have a derivation $\Gamma \vdash \texttt{new}\ C_0(\overline{v}) : D_0$ with $fields(D_0) = \overline{D\ g}$ and $C = D_i$. By a similar argument, the last rule used in this derivation must have been T-NEW and so $D_0 = C_0$ and we have derivations $\Gamma \vdash \overline{v} : \overline{B}$ with $\overline{B} \leq \overline{D}$. From $D_0 = C_0$ (and as *fields* is a function) we have $\overline{C\ f} = \overline{D\ g}$, and hence $C = C_i$. Thus, $\Gamma \vdash v_i : B_i$ with $B_i \leq C_i$, as required.

Case E-INVK: $e = (\texttt{new}\ C_0(\overline{v})).m(\overline{u})$ and $e' = [\overline{x} \mapsto \overline{u}, \texttt{this} \mapsto \texttt{new}\ C_0(\overline{v})]e$ with $mbody(m, C_0) = (\overline{x}, e)$

By similar reasoning as in the previous case, the last two rules in the derivation of $\Gamma \vdash e : C$ must have been T-INVK and T-NEW with $\Gamma \vdash \texttt{new}\ C_0(\overline{v}) : C_0$ and $\Gamma \vdash \overline{u} : \overline{B}$ and $mtype(m, C_0) = \overline{C} \to C$ with $\overline{B} \leq \overline{C}$. By the method typing lemma, there exist types C_0' and C' such that $\overline{x} : \overline{C}, \texttt{this} : C_0' \vdash e : C'$. By the substitution lemma we have $\vdash [\overline{x} \mapsto \overline{u}, \texttt{this} \mapsto \texttt{new}\ C_0(\overline{v})]e : C''$ with $C'' \leq C'$. By weakening we have $\Gamma \vdash [\overline{x} \mapsto \overline{u}, \texttt{this} \mapsto \texttt{new}\ C_0(\overline{v})]e : C''$. The required result follows as $C'' \leq C$ by S-TRANS.

Case E-CAST: $e = (C)\ (\texttt{new}\ C_0(\overline{v}))$ and $e' = \texttt{new}\ C_0(\overline{v})$ with $C_0 \leq C$

By similar reasoning as the previous cases, the last two rules in the derivation of $\Gamma \vdash e : C$ must have been T-UCAST and T-NEW with $\Gamma \vdash \texttt{new}\ C_0(\overline{v}) : C_0$. The result is immediate as $C_0 \leq C$.

□

2 Progress

The proof of progress also relies on a few supporting lemmas.

Lemma (Canonical Forms)**.** *If* $\vdash v : C$ *then* $v = \texttt{new}\ C(\overline{v})$.

Lemma (Inversion)**.**

1. *If* $\vdash (\texttt{new}\ C(\overline{v})).f_i : C_i$ *then* $\texttt{fields}(C) = \overline{C\ f}$ *and* $f_i \in \overline{f}$.
2. *If* $\vdash (\texttt{new}\ C(\overline{v})).m(\overline{u}) : C$ *then* $\texttt{mbody}(m, C) = (\overline{x}, e)$ *and* $|\overline{u}| = |\overline{e}|$.

Lemma (Progress)**.** *Let* e *be an expression such that* $\vdash e : C$*. Then either:*

1. e *is a value,*
2. *there exists an expression* e' *such that* $e \to e'$*, or*
3. $e = E[(B)\ (\texttt{new}\ A(\overline{v}))]$ *with* $A \not\leq B$.

Proof. By induction on $\vdash e : C$, with a case analysis on the last rule used in the derivation.

Case T-VAR: $e = x$ with $\emptyset(x) = C$

Can't happen, as $\emptyset(x)$ is undefined.

Case T-FIELD: $e = e_0.f$ with $\vdash e_0 : C_0$ and $fields(C_0) = \overline{C\ f}$ and $C = C_i$

By the induction hypothesis applied to e_0 we have that either e_0 is a value, there exists e_0' such that $e_0 \to e_0'$, or there exists E such that $e_0 = E_0[(B)\ (\texttt{new}\ A(\overline{v}))]$ with $A \not\leq B$. We analyze each of these subcases:

1. If e_0 is a value then by the canonical forms lemma, $e_0 = \texttt{new}\ C_0(\overline{v})$ and by the inversion lemma $f \in \overline{f}$. By E-PROJ we have $e \to v_i$.
2. Alternatively, if there exists an expression such that $e_0 \to e_0'$ then by E-CONTEXT we have $e = E[e_0] \to E[e_0']$ where $E = [\cdot].f$.
3. Otherwise, if $e_0 = E_0[(B)\ (\texttt{new}\ A(\overline{v}))]$ with $A \not\leq B$ then we have $e = E[(B)\ (\texttt{new}\ A(\overline{v}))]$ where $E = [\cdot].f$, which finishes the case.

Case T-INVK: $e = e_0.m(\overline{e})$ with $\vdash e_0 : C_0$ and $\mathit{mtype}(m, C_0) = \overline{B} \to C$ and $\vdash \overline{e} : \overline{A}$ and $\overline{A} \leq \overline{B}$

By the induction hypothesis applied to e_0 we have that either e_0 is a value, there exists e_0' such that $e_0 \to e_0'$, or there exists E such that $e_0 = E_0[(B)\ (\texttt{new}\ A(\overline{v}))]$ with $A \not\leq B$. We analyze each of these subcases:

1. If e_0 is a value then by the canonical forms lemma, $e_0 = \texttt{new}\ C_0(\overline{v})$. If $\overline{e}$ is a list of values $\overline{u}$, then by the inversion lemma we have $|\overline{u}| = |\overline{x}|$ where $\mathit{mbody}(m, C_0) = (\overline{x}, e_0')$. By E-INVK we have $e \to [\overline{x} \mapsto \overline{u}, \texttt{this} \mapsto \texttt{new}\ C_0(\overline{v})]e_0'$. Otherwise, let i be the least index of an expression in $\overline{e}$ that is not a value. By the induction hypothesis applied to e_i we have that e_i is a value, or there exists e_i' such that $e_i \to e_i'$ or there exists E_i such that $e_i = E_i[(B)\ (\texttt{new}\ A(\overline{v}))]$ and $A \not\leq B$. In the first subsubcase, then we have a contradiction to our assumption that i is the index of the least expression in $\overline{e}$ that is not a value. Otherwise let $E = (\texttt{new}\ C_0(\overline{v})).m(e_1, \ldots, e_{i-1}, E_i, e_{i+1}, \ldots |\overline{e}|)$. In the second subcase, we have $e = E[e_i] \to E[e_i']$ by E-CONTEXT. In the third subcase, we have $e = E[(B)\ (\texttt{new}\ A(\overline{v}))]$ with $A \not\leq B$.
2. Alternatively, if there exists an expression such that $e_0 \to e_0'$ then by E-CONTEXT we have $E[e_0] \to E[e_0']$ where $E = [\cdot].m(\overline{e})$.
3. Otherwise, if $e_0 = E_0[(B)\ (\texttt{new}\ A(\overline{v}))]$ with $A \not\leq B$ then we have $e = E[(B)\ (\texttt{new}\ A(\overline{v}))]$ where $E = [\cdot].m(\overline{e})$, which finishes the case.

Case T-NEW: $e = \texttt{new}\ C(\overline{e})$ and $\mathit{fields}(C) = \overline{C\ f}$ and $\vdash \overline{e} : \overline{B}$ and $\overline{B} \leq \overline{C}$

If $\overline{e}$ is a list of values $\overline{u}$, then e is a value. Otherwise, let i be the least index of an expression in $\overline{e}$ that is not a value. By the induction hypothesis applied to e_i we have that e_i is a value, or there exists e_i' such that $e_i \to e_i'$ or there exists E_i such that $e_i = E_i[(B)\ (\texttt{new}\ A(\overline{v}))]$ and $A \not\leq B$. We analyze each of these subcases:

1. If e_i is a value then we have a contradiction to our assumption that i is the index of the least expression in $\overline{e}$ that is not a value.
2. If there exists e_i' such that $e_i \to e_i'$ then let $E = (\texttt{new}\ C(e_1, \ldots, e_{i-1}, E_i, e_{i+1}, \ldots, |\overline{e}|)$. By E-CONTEXT we have $e = E[e_i] \to E[e_i']$.
3. Otherwise, if there exists E_i with $e_i = E_i[(B)\ (\texttt{new}\ A(\overline{v}))]$ and $A \not\leq B$ then let $E = (\texttt{new}\ C(e_1, \ldots, e_{i-1}, E_i, e_{i+1}, \ldots, |\overline{e}|)$. By construction we have $e = E[(B)\ (\texttt{new}\ A(\overline{v}))]$, which finishes the case.

Case T-UCAST: $e = (C)\ e$ with $\vdash e_0 : D$ and $D \leq C$

By the induction hypothesis applied to e_0 we have that either e_0 is a value, there exists e_0'

such that $e_0 \to e_0'$, or there exists E such that $e_0 = E_0[(B)\ (\texttt{new}\ A(\overline{v}))]$ with $A \not\leq B$. We analyze each of these subcases:

1. If e_0 is a value then by the canonical forms lemma, $e_0 = \texttt{new}\ D(\overline{v})$. By E-CAST we have $e \to \texttt{new}\ D(\overline{v})$.
2. Alternatively, if there exists an expression such that $e_0 \to e_0'$ then by E-CONTEXT we have $e = E[e_0] \to E[e_0']$ where $E = (C)\ [\cdot]$.
3. Otherwise, if $e_0 = E_0[(B)\ (\texttt{new}\ A(\overline{v}))]$ with $A \not\leq B$ then we have $e = E[(B)\ (\texttt{new}\ A(\overline{v}))]$ where $E = (C)\ [\cdot]$, which finishes the case.

Case T-DCAST: $e = (C)\ e$ with $\vdash e_0 : D$ and $C \leq D$ and $C \neq D$

By the induction hypothesis applied to e_0 we have that either e_0 is a value, there exists e_0' such that $e_0 \to e_0'$, or there exists E such that $e_0 = E_0[(B)\ (\texttt{new}\ A(\overline{v}))]$ with $A \not\leq B$. We analyze each of these subcases:

1. If e_0 is a value then by the canonical forms lemma we have that $e = \texttt{new}\ D(\overline{v})$. Let $E = [\cdot]$. We immediately $e = E[(C)\ \texttt{new}\ C(\overline{v})]$ with $D \not\leq C$.
2. Alternatively, if there exists an expression such that $e_0 \to e_0'$ then by E-CONTEXT we have $e = E[e_0] \to E[e_0']$ where $E = (C)\ [\cdot]$.
3. Otherwise, if $e_0 = E_0[(B)\ (\texttt{new}\ A(\overline{v}))]$ with $A \not\leq B$ then we have $e = E[(B)\ (\texttt{new}\ A(\overline{v}))]$ where $E = (C)\ [\cdot]$, which finishes the case.

Case T-SCAST: similar to the previous case.

□