



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 $mtype(m, C) = \overline{D} \rightarrow D$ and $mbody(m, C) = (\overline{x}, e)$ then there exists types C' and D' such that $\overline{x} : \overline{D}$, $this : C' \vdash e : D'$ 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 $e \rightarrow e'$ then there exists a type C' such that $\Gamma \vdash e' : C'$ and $C' \leq C$.*

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

Case E-Context: $e = E[e_1]$ and $e_1 \rightarrow e'_1$ 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: $e = \text{new } C_0(\overline{v}).f_i$ and $e' = v_i$ with $fields(C_0) = \overline{C} \overline{f}$

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 $\Gamma \vdash \text{new } C_0(\overline{v}) : D_0$ with $fields(D_0) = \overline{D} \overline{g}$ 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 $\overline{C} \overline{f} = \overline{D} \overline{g}$, and hence $C = C_i$. Thus, $\Gamma \vdash v_i : B_i$ with $B_i \leq C_i$, as required.

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

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 $\Gamma \vdash \text{new } C_0(\bar{v}) : C_0$ and $\Gamma \vdash \bar{u} : \bar{B}$ and $mtype(m, C_0) = \bar{C} \rightarrow C$ with $\bar{B} \leq \bar{C}$. By the method typing lemma, there exist types C'_0 and C' such that $\bar{x} : \bar{C}, \text{this} : C'_0 \vdash e : C'$. By the substitution lemma we have $\vdash [\bar{x} \mapsto \bar{u}, \text{this} \mapsto \text{new } C_0(\bar{v})]e : C''$ with $C'' \leq C'$. By weakening we have $\Gamma \vdash [\bar{x} \mapsto \bar{u}, \text{this} \mapsto \text{new } C_0(\bar{v})]e : C''$. The required result follows as $C'' \leq C$ by S-TRANS.

Case E-Cast: $e = (C) (\text{new } C_0(\bar{v}))$ and $e' = \text{new } C_0(\bar{v})$ 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 $\Gamma \vdash \text{new } C_0(\bar{v}) : C_0$. 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 $v = \text{new } C(\bar{v})$.*

Lemma (Inversion).

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

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 $e \rightarrow e'$, or
3. $e = E[(B) (\text{new } A(\bar{v}))]$ 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 $fields(C_0) = \overline{C} \bar{f}$ 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 $e_0 \rightarrow e'_0$, or there exists E such that $e_0 = E[(B) (\text{new } A(\bar{v}))]$ with $A \not\leq B$. We analyze each of these subcases:

1. If e_0 is a value then by the canonical forms lemma, $e_0 = \text{new } C_0(\bar{v})$ and by the inversion lemma $f \in \bar{f}$. By E-PROJ we have $e \rightarrow v_i$.

2. Alternatively, if there exists an expression such that $e_0 \rightarrow e'_0$ then by E-CONTEXT we have $e = E[e_0] \rightarrow E[e'_0]$ where $E = [\cdot].f$.
3. Otherwise, if $e_0 = E_0[(B) (\text{new } A(\bar{v}))]$ with $A \not\leq B$ then we have $e = E[(B) (\text{new } A(\bar{v}))]$ where $E = [\cdot].f$, which finishes the case.

Case T-Invk: $e = e_0.m(\bar{e})$ with $\vdash e_0 : C_0$ and $mtype(m, C_0) = \bar{B} \rightarrow C$ and $\vdash \bar{e} : \bar{A}$ and $\bar{A} \leq \bar{B}$

By the induction hypothesis applied to e_0 we have that either e_0 is a value, there exists e'_0 such that $e_0 \rightarrow e'_0$, or there exists E such that $e_0 = E_0[(B) (\text{new } A(\bar{v}))]$ with $A \not\leq B$. We analyze each of these subcases:

1. If e_0 is a value then by the canonical forms lemma, $e_0 = \text{new } C_0(\bar{v})$. If \bar{e} is a list of values \bar{u} , then by the inversion lemma we have $|\bar{u}| = |\bar{x}|$ where $mbody(m, C_0) = (\bar{x}, e'_0)$. By E-INVK we have $e \rightarrow [\bar{x} \mapsto \bar{u}, \text{this} \mapsto \text{new } C_0(\bar{v})]e'_0$. Otherwise, let i be the least index of an expression in \bar{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 $e_i \rightarrow e'_i$ or there exists E_i such that $e_i = E_i[(B) (\text{new } A(\bar{v}))]$ 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 \bar{e} that is not a value. Otherwise let $E = (\text{new } C_0(\bar{v})).m(e_1, \dots, e_{i-1}, E_i, e_{i+1}, \dots, |\bar{e}|)$. In the second subcase, we have $e = E[e_i] \rightarrow E[e'_i]$ by E-CONTEXT. In the third subcase, we have $e = E[(B) (\text{new } A(\bar{v}))]$ with $A \not\leq B$.
2. Alternatively, if there exists an expression such that $e_0 \rightarrow e'_0$ then by E-CONTEXT we have $E[e_0] \rightarrow E[e'_0]$ where $E = [\cdot].m(\bar{e})$.
3. Otherwise, if $e_0 = E_0[(B) (\text{new } A(\bar{v}))]$ with $A \not\leq B$ then we have $e = E[(B) (\text{new } A(\bar{v}))]$ where $E = [\cdot].m(\bar{e})$, which finishes the case.

Case T-New: $e = \text{new } C(\bar{e})$ and $fields(C) = \bar{C} \bar{f}$ and $\vdash \bar{e} : \bar{B}$ and $\bar{B} \leq \bar{C}$

If \bar{e} is a list of values \bar{u} , then e is a value. Otherwise, let i be the least index of an expression in \bar{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 $e_i \rightarrow e'_i$ or there exists E_i such that $e_i = E_i[(B) (\text{new } A(\bar{v}))]$ 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 \bar{e} that is not a value.
2. If there exists e'_i such that $e_i \rightarrow e'_i$ then let $E = (\text{new } C(e_1, \dots, e_{i-1}, E_i, e_{i+1}, \dots, |\bar{e}|))$. By E-CONTEXT we have $e = E[e_i] \rightarrow E[e'_i]$.
3. Otherwise, if there exists E_i with $e_i = E_i[(B) (\text{new } A(\bar{v}))]$ and $A \not\leq B$ then let $E = (\text{new } C(e_1, \dots, e_{i-1}, E_i, e_{i+1}, \dots, |\bar{e}|))$. By construction we have $e = E[(B) (\text{new } A(\bar{v}))]$, 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 $e_0 \rightarrow e'_0$, or there exists E such that $e_0 = E_0[(B) (\text{new } A(\bar{v}))]$ with $A \not\leq B$. We analyze each of these subcases:

1. If e_0 is a value then by the canonical forms lemma, $e_0 = \text{new } D(\bar{v})$. By E-CAST we have $e \rightarrow \text{new } D(\bar{v})$.

2. Alternatively, if there exists an expression such that $e_0 \rightarrow e'_0$ then by E-CONTEXT we have $e = E[e_0] \rightarrow E[e'_0]$ where $E = (C) [\cdot]$.
3. Otherwise, if $e_0 = E_0[(B) (\mathbf{new} A(\bar{v}))]$ with $A \not\leq B$ then we have $e = E[(B) (\mathbf{new} A(\bar{v}))]$ 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 $e_0 \rightarrow e'_0$, or there exists E such that $e_0 = E_0[(B) (\mathbf{new} A(\bar{v}))]$ 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 $e = \mathbf{new} D(\bar{v})$. Let $E = [\cdot]$. We immediately have $e = E[(C) \mathbf{new} C(\bar{v})]$ with $D \not\leq C$.
2. Alternatively, if there exists an expression such that $e_0 \rightarrow e'_0$ then by E-CONTEXT we have $e = E[e_0] \rightarrow E[e'_0]$ where $E = (C) [\cdot]$.
3. Otherwise, if $e_0 = E_0[(B) (\mathbf{new} A(\bar{v}))]$ with $A \not\leq B$ then we have $e = E[(B) (\mathbf{new} A(\bar{v}))]$ where $E = (C) [\cdot]$, which finishes the case.

Case T-SCast: similar to the previous case.

□