

Structural type inference in Java-like languages

Martin Plümicke

Baden-Wuerttemberg Cooperative State University
Stuttgart/Horb

25. Februar 2016

Overview

Motivating Example

Language

The Algorithm

- The function TYPE

- The function construct

- The function solve

Summary

Motivating Example

Type inference example

```
class A {  
    m (v) {  
        return v.elementAt(0);  
    }  
}
```

Motivating Example

Type inference example

```
class A {  
    m (v) {  
        return v.elementAt(0);  
    }  
}
```

No type inference result for m

Motivating Example

Type inference example

```
import java.util.Vector;

class A {

    m (v) {
        return v.elementAt(0);
    }
}
```

Motivating Example

Type inference example

```
import java.util.Vector;

class A {

    m (v) {
        return v.elementAt(0);
    }
}
```

Result: $m : \text{Vector}\langle T \rangle \rightarrow T.$

Motivating Example

Type inference example

```
class A {  
    m (v) {  
        return v.elementAt(0);  
    }  
}
```

Motivating Example

Type inference example

```
interface I<T> { T elementAt(int x); }
```

```
class A {  
    m (v) {  
        return v.elementAt(0);  
    }  
}
```

Result: $m : I<T> \rightarrow T.$

Input language (extended Featherweight Java)

$L ::= \text{class } C \text{ extends } (CT)^* \{ \bar{f}; \bar{M} \}$

$M ::= m(\bar{x}) \{ \text{return } e; \}$

$e ::= x \mid e.f \mid e.m(\bar{e}) \mid \text{new } NCT(\bar{e}) \mid (CT)e$

$NCT ::= CT \mid C \langle \overline{TVar = CT} \rangle$

$CT ::= C \langle \overline{CT} \rangle$

Output language

$L_t ::= I^* CL_t$

$CL_t ::= \text{class } C \langle \overline{\text{TVar}} \rangle [\overline{\text{CONS}}] \text{ extends } (CT)^* \{ \overline{\text{T f}}; \overline{M_t} \}$

$\text{CONS} ::= T \text{ extends } T$

$T ::= CT \mid \text{TVar}$

$MH ::= T \text{ m}(\overline{\text{T}} \overline{\text{x}});$

$M_t ::= MH \{ \text{return } e_t; \}$

$e_t ::= x : T \mid e.f : T \mid e.m(\overline{e}) : T \mid \text{new NCT}(\overline{e}) : CT \mid (CT)e : CT$

$I ::= \text{interface } I \langle \overline{\text{TVar}} \rangle \{ \overline{\text{T f}}; \overline{MH} \}$

The Algorithm

TI: $\text{TypeAssumptions} \times L \rightarrow L_t$

TI $(Ass, cl) =$

let

$(cl_t, C) = \mathbf{TYPE}(Ass, cl)$

$(l_1 \dots l_m cl_t) = \mathbf{construct}(cl_t, C)$

in

$(l_1 \dots l_m \mathbf{solve}(cl_t))$

The function TYPE

TYPE: $\text{TypeAssumptions} \times \mathbf{L} \rightarrow \mathbf{L}_t \times \text{ConstraintsSet}$

The function TYPE

TYPE: $\text{TypeAssumptions} \times L \rightarrow L_t \times \text{ConstraintsSet}$

TYPE(*Ass*, **class** *A* **extends** \bar{B} { \bar{f} ; \bar{M} }) = **let**

fass := { **this.f** : δ_A^f | $f \in \bar{f}$ } \cup { **this.f** : T | $T f \in \text{fields}(\bar{B})$ }

The function TYPE

TYPE: $\text{TypeAssumptions} \times \mathbf{L} \rightarrow \mathbf{L}_t \times \text{ConstraintsSet}$

TYPE(*Ass*, **class A extends** $\bar{\mathbf{B}} \{ \bar{\mathbf{f}}; \bar{\mathbf{M}} \}$) = **let**

fass := $\{ \text{this.f} : \delta_A^f \mid f \in \bar{\mathbf{f}} \} \cup \{ \text{this.f} : T \mid T f \in \text{fields}(\bar{\mathbf{B}}) \}$

mass := $\{ \text{this.m} : \bar{\alpha}_A^m \rightarrow \gamma_A^m \mid m(\bar{\mathbf{x}}) \{ \text{return } e; \} \in \bar{\mathbf{M}} \}$

$\cup \{ \text{this.m} : \bar{\text{aty}} \rightarrow \text{rty} \mid \text{mtype}(m, \bar{\mathbf{B}}) = \bar{\text{aty}} \rightarrow \text{rty} \}$

The function TYPE

TYPE: $\text{TypeAssumptions} \times \mathbf{L} \rightarrow \mathbf{L}_t \times \text{ConstraintsSet}$

TYPE(*Ass*, **class** *A* **extends** $\bar{\mathbf{B}}$ { $\bar{\mathbf{f}}$; $\bar{\mathbf{M}}$ }) = **let**

fass := { **this.f** : $\delta_A^f \mid f \in \bar{\mathbf{f}}$ } \cup { **this.f** : $\mathbf{T} \mid \mathbf{T} f \in \text{fields}(\bar{\mathbf{B}})$ }

mass := { **this.m** : $\bar{\alpha}_A^m \rightarrow \gamma_A^m \mid m(\bar{\mathbf{x}})\{\text{return } e;\} \in \bar{\mathbf{M}}$ }

\cup { **this.m** : $\bar{\text{aty}} \rightarrow \text{rty} \mid \text{mtype}(m, \bar{\mathbf{B}}) = \bar{\text{aty}} \rightarrow \text{rty}$ }

AssAll = *Ass* \cup *fass* \cup *mass* \cup { **this** : *A* }

The function TYPE

TYPE: $\text{TypeAssumptions} \times \mathbf{L} \rightarrow \mathbf{L}_t \times \text{ConstraintsSet}$

TYPE(*Ass*, **class** *A* **extends** \bar{B} { \bar{f} ; \bar{M} }) = **let**

fass := { **this.f** : δ_A^f | $f \in \bar{f}$ } \cup { **this.f** : T | $T f \in \text{fields}(\bar{B})$ }

mass := { **this.m** : $\alpha_A^m \rightarrow \gamma_A^m$ | $m(\bar{x})\{ \text{return } e; \} \in \bar{M}$ }

\cup { **this.m** : $\overline{aty} \rightarrow rty$ | $mtype(m, \bar{B}) = \overline{aty} \rightarrow rty$ }

AssAll = *Ass* \cup *fass* \cup *mass* \cup { **this** : *A* }

For $m(\bar{x})\{ \text{return } e; \} \in \bar{M}$ {

Ass = *AssAll* \cup { x_j : $\alpha_A^{m,j}$ | $\bar{x} = x_1 \dots x_{n_i}$ }

(e_t : *rty*, *C'*) = **TYPEExpr**(*Ass*, *e*)

C = *C* \cup *C'*[$\gamma_A^m \mapsto rty$]

$\bar{M}_t = \{ rty \ m(\overline{\alpha_A^m} \ x)\{ \text{return } e_t; \} \mid m(\bar{x})\{ \text{return } e; \} \in \bar{M} \}$

The function TYPE

TYPE: $\text{TypeAssumptions} \times \mathbf{L} \rightarrow \mathbf{L}_t \times \text{ConstraintsSet}$

TYPE(*Ass*, **class A extends** $\overline{\mathbf{B}}$ { $\overline{\mathbf{f}}$; $\overline{\mathbf{M}}$ }) = **let**

fass := { **this.f** : δ_A^f | $f \in \overline{\mathbf{f}}$ } \cup { **this.f** : T | T f \in *fields*($\overline{\mathbf{B}}$) }

mass := { **this.m** : $\overline{\alpha}_A^m \rightarrow \gamma_A^m$ | $m(\overline{\mathbf{x}})\{ \text{return } e; \} \in \overline{\mathbf{M}}$ }

\cup { **this.m** : $\overline{aty} \rightarrow rty$ | *mtype*($m, \overline{\mathbf{B}}$) = $\overline{aty} \rightarrow rty$ }

AssAll = *Ass* \cup *fass* \cup *mass* \cup { **this** : A }

For $m(\overline{\mathbf{x}})\{ \text{return } e; \} \in \overline{\mathbf{M}}$ {

Ass = *AssAll* \cup { x_j : $\alpha_A^{m,j}$ | $\overline{\mathbf{x}} = x_1 \dots x_{n_i}$ }

(e_t : *rty*, *C'*) = **TYPEExpr**(*Ass*, **e**)

C = *C* \cup *C'*[$\gamma_A^m \mapsto rty$]

$\overline{\mathbf{M}}_t = \{ rty \ m(\overline{\alpha}_A^m \ \overline{\mathbf{x}})\{ \text{return } e_t; \} \mid m(\overline{\mathbf{x}})\{ \text{return } e; \} \in \overline{\mathbf{M}} \}$

in

(**class** A **extends** $\overline{\mathbf{B}}$ { $\overline{\delta}_A$ f; $\overline{\mathbf{M}}_t$ }, *C*)

Type constraints

- ▶ $c \leq c'$ means c is a subtype of c' .
- ▶ $\phi(c, f, c')$ means c provides a field f with type c' .
- ▶ $\mu(c, m, \bar{c}, (c', \bar{c}'))$ means c provides a method m applicable to arguments of type \bar{c} , with return type c' and parameters of type \bar{c}' .
 $\mu(c, m, \bar{c}, (c', \bar{c}'))$ implicitly includes the constraints $\bar{c} \leq \bar{c}'$.

TYPEExpr for Method-call

TYPEExpr($Ass, e_0.m(\bar{e})$) = **let**
 $(e_{0_t} : ty_0, C_0)$ = **TYPEExpr**(Ass, e_0)
 $(\bar{e}_t : \bar{ty}, \bar{C})$ = **TYPEExpr**(Ass, \bar{e})

TYPEExpr for Method-call

TYPEExpr(*Ass*, $e_0.m(\bar{e})$) = **let**
 $(e_{0_t} : ty_0, C_0)$ = **TYPEExpr**(*Ass*, e_0)
 $(\overline{e_t : ty}, \overline{C})$ = **TYPEExpr**(*Ass*, \bar{e})

in

if (*ty₀ is a type variable*) **then**

$((e_{0_t} : ty_0).m(\overline{e_t : ty}) : \gamma_{ty_0}^m, (C_0 \cup \overline{C}) \cup \{ \mu(ty_0, m, \overline{ty}, (\gamma_{ty_0}^m, \overline{\beta_{ty_0}^m})) \})$

TYPEExpr for Method-call

TYPEExpr(*Ass*, $e_0.m(\bar{e})$) = **let**
 $(e_{0_t} : ty_0, C_0)$ = **TYPEExpr**(*Ass*, e_0)
 $(\overline{e_t : ty}, \overline{C})$ = **TYPEExpr**(*Ass*, \bar{e})

in

if (*ty*₀ is a type variable) **then**

$((e_{0_t} : ty_0).m(\overline{e_t : ty}) : \gamma_{ty_0}^m, (C_0 \cup \overline{C}) \cup \{ \mu(ty_0, m, \overline{ty}, (\gamma_{ty_0}^m, \overline{\beta_{ty_0}^m})) \})$

else if (*ty*₀ ∈ *Ass*) && (*mtype*(*m*, *ty*₀) = $\overline{aty} \rightarrow rty$) **then**

$((e_{0_t} : ty_0).m(\overline{e_t : ty}) : rty, (C_0 \cup \overline{C}) \cup \{ \overline{ty} \ll aty \})$

The function construct

1. For any $\phi(\text{ty1}, \mathbf{f}, \delta)$ and $\mu(\text{ty2}, \mathbf{m}, \bar{\alpha}, (\gamma, \bar{\beta})) \in C$ interfaces are generated:

```
interface ty1 < ...,  $\delta$ , ... > {  
     $\delta$  f;  
}
```

```
interface ty2 < ...,  $\gamma$ ,  $\bar{\beta}$ , ... > {  
     $\gamma$  m( $\bar{\beta}$  x1);  
}
```

2. fresh type variables ν_1 , ν_2 and constraints are introduced:

$$\nu_1 \triangleleft \text{ty1} \langle \dots \rangle,$$
$$\nu_2 \triangleleft \text{ty2} \langle \dots \rangle$$

3. the type variables ν_1 , ν_2 are introduced as generics in the class:

```
class A < ...,  $\nu_1$ ,  $\nu_2$ , ... > extends ...
```

The function solve

solve: $L_t \rightarrow L_t$

solve(class $A < \bar{T} > [\overline{ty \leftarrow ty'}]$ extends $\bar{B} \{ \overline{ty \ f}; \overline{M_t} \}$) =

class $A < \dots > [\dots]$ extends $\bar{B} \{ \overline{\sigma(ty) \ f}; \overline{\sigma(M_t)} \}$,

with

▶ $\sigma = \mathbf{TypeUnify}(\overline{ty \leftarrow ty'}) :$

For the pairs $\overline{ty \leftarrow ty'}$ a substitution σ is determined such that:

$$\overline{\sigma(ty)} \leq^* \overline{\sigma(ty')}$$

▶ \leq^* is the subtyping relation.

Example (input syntax)

```
class A {  
    mt(x, y, z) {  
        return x.sub(y).add(z);  
    }  
}
```


Result of TYPE

```
class A {
```

```
    mt( $\alpha_A^{mt,1}$  x,  $\alpha_A^{mt,2}$  y,  $\alpha_A^{mt,3}$  z) {
```

```
        return e_t;
```

```
    }  
}
```

```
mt(x, y, z) {  
    return x.sub(y).add(z);  
}
```

with $e_t = [[[x : \alpha_A^{mt,1}].sub([y : \alpha_A^{mt,2}]) : \gamma_{\alpha_A^{mt,1}}^{sub}].add(z : \alpha_A^{mt,3}) : \gamma_{\alpha_A^{mt,1}}^{add}]]$

Result of TYPE

```
mt(x, y, z) {  
    return x.sub(y).add(z);  
}
```

```
class A {
```

```
    mt( $\alpha_A^{mt,1}$  x,  $\alpha_A^{mt,2}$  y,  $\alpha_A^{mt,3}$  z) {  
         $\gamma_{\alpha_A^{mt,1}}^{add}$   
         $\gamma_{\alpha_A^{mt,1}}^{sub}$ 
```

```
        return e_t;
```

```
    }  
}
```

with $e_t = [[[x : \alpha_A^{mt,1}].sub([y : \alpha_A^{mt,2}]) : \gamma_{\alpha_A^{mt,1}}^{sub}].add(z : \alpha_A^{mt,3}) : \gamma_{\alpha_A^{mt,1}}^{add}]]$

and

$$C = \{ \mu(\alpha_A^{mt,1}, \text{sub}, \alpha_A^{mt,2}, (\gamma_{\alpha_A^{mt,1}}^{sub}, \beta_{\alpha_A^{mt,1}}^{sub,1})), \\ \mu(\gamma_{\alpha_A^{mt,1}}^{sub}, \text{add}, \alpha_A^{mt,3}, (\gamma_{\alpha_A^{mt,1}}^{add}, \beta_{\alpha_A^{mt,1}}^{add,1})) \}$$

The result of construct (output syntax)

```
interface  $\alpha_A^{mt,1}$  <Gamma_m, Beta_m> { Gamma_m sub(Beta_m x); }
```

```
interface  $\gamma_{\alpha_A}^{sub,1}$  <Gamma_n, Beta_n> { Gamma_n add(Beta_n x); }
```

```
class A < $\nu_1, \nu_2, \nu_3, \nu_4, \nu_5, \nu_6, \nu_7$ >  
  [ $\nu_3 < \nu_5, \nu_4 < \nu_7, \nu_1 < \alpha_A^{mt,1} < \nu_2, \nu_5 >, \nu_2 < \gamma_{\alpha_A}^{sub,1} < \nu_6, \nu_7 >$ ] {  
     $\nu_6$  mt( $\nu_1$  x,  $\nu_3$  y,  $\nu_4$  z) {  
      return x.sub(y).add(z);  
    }  
  }  
}
```

The result of construct (output syntax)

```
interface  $\alpha_A^{mt,1}$  <Gamma_m, Beta_m> { Gamma_m sub(Beta_m x); }
```

```
interface  $\gamma_{\alpha_A}^{sub,mt,1}$  <Gamma_n, Beta_n> { Gamma_n add(Beta_n x); }
```

```
class A < $\nu_1, \nu_2, \nu_3, \nu_4, \nu_5, \nu_6, \nu_7$ >  
  [ $\nu_3 < \nu_5, \nu_4 < \nu_7, \nu_1 < \alpha_A^{mt,1} < \nu_2, \nu_5 >, \nu_2 < \gamma_{\alpha_A}^{sub,mt,1} < \nu_6, \nu_7 >$ ]{  
   $\nu_6$  mt( $\nu_1$  x,  $\nu_3$  y,  $\nu_4$  z) {  
    return x.sub(y).add(z);  
  }  
}
```

The application of **solve** changes nothing.

Instance Example (input syntax)

```
class myInteger extends  $\alpha_A^{mt,1}$ <myInteger, myInteger>,  
                        $\gamma_{\alpha_A^{mt,1}}^{sub}$ <myInteger, myInteger> {  
  
    Integer i;  
  
    myInteger sub(myInteger x) {  
        return new myInteger(i - x.i);  
    }  
  
    myInteger add(myInteger x) {  
        return new myInteger(i + x.i);  
    }  
}
```

```

class A < $\nu_1, \nu_2, \nu_3, \nu_4, \nu_5, \nu_6, \nu_7$ >
    [ $\nu_3 \ll \nu_5, \nu_4 \ll \nu_7, \nu_1 \ll \alpha_A^{\text{mt},1} \langle \nu_2, \nu_5 \rangle, \nu_2 \ll \gamma_{\alpha_A^{\text{mt},1}}^{\text{sub}} \langle \nu_6, \nu_7 \rangle$ ]{
     $\nu_6$  mt( $\nu_1$  x,  $\nu_3$  y,  $\nu_4$  z) {
        return x.sub(y).add(z);
    }
}

class Main {
    main() {
        return new A< $\nu_1$ =myInteger,  $\nu_2$ =myInteger>()
            .mt(new myInteger(2),
                new myInteger(1),
                new myInteger(3));
    }
}

```

The result of TYPE

$$C = \{ \begin{array}{l} \nu_3 \triangleleft \nu_5, \\ \nu_4 \triangleleft \nu_7, \\ \text{myInteger} \triangleleft \alpha_A^{\text{mt},1} \langle \text{myInteger}, \nu_5 \rangle, \\ \text{myInteger} \triangleleft \gamma_{\alpha_A^{\text{mt},1}}^{\text{sub}} \langle \nu_6, \nu_7 \rangle, \\ \text{myInteger} \triangleleft \nu_3, \\ \text{myInteger} \triangleleft \nu_4 \end{array} \}$$

The result of TYPE

$$C = \{ \begin{array}{l} \nu_3 \triangleleft \nu_5, \\ \nu_4 \triangleleft \nu_7, \\ \text{myInteger} \triangleleft \alpha_A^{\text{mt},1} \langle \text{myInteger}, \nu_5 \rangle, \\ \text{myInteger} \triangleleft \gamma_{\alpha_A^{\text{mt},1}}^{\text{sub}} \langle \nu_6, \nu_7 \rangle, \\ \text{myInteger} \triangleleft \nu_3, \\ \text{myInteger} \triangleleft \nu_4 \end{array} \}$$

construct adds no interfaces, as there is no call of abstract fields or methods.

The result of solve

$$\sigma = \left\{ \begin{array}{l} \nu_5 \mapsto \text{myInteger}, \\ \nu_6 \mapsto \text{myInteger}, \\ \nu_7 \mapsto \text{myInteger}, \\ \nu_3 \mapsto \text{myInteger}, \\ \nu_4 \mapsto \text{myInteger} \end{array} \right\}$$

The result of solve

$$\sigma = \{ \nu_5 \mapsto \text{myInteger}, \\ \nu_6 \mapsto \text{myInteger}, \\ \nu_7 \mapsto \text{myInteger}, \\ \nu_3 \mapsto \text{myInteger}, \\ \nu_4 \mapsto \text{myInteger} \}$$

```
class Main {  
  
    myInteger main() {  
        return new A<myInteger, myInteger, myInteger, myInteger,  
            myInteger, myInteger, myInteger>()  
            .mt(new myInteger(2),  
                new myInteger(1),  
                new myInteger(3));  
    }  
}
```

Summary

Conclusion

We have presented a type inference algorithm for a Java-like language.

- ▶ type-less Java classes independently from any environment
- ▶ separate compilation of Java classes without relying on type informations of other classes.
- ▶ infers structural types, given as generated interfaces

Summary

Conclusion

We have presented a type inference algorithm for a Java-like language.

- ▶ type-less Java classes independently from any environment
- ▶ separate compilation of Java classes without relying on type informations of other classes.
- ▶ infers structural types, given as generated interfaces

Ongoing work

- ▶ complete **construct** and **solve** algorithms
- ▶ soundness and completeness proof