Programming Languages and Compilers (CS 421)

Sasa Misailovic 4110 SC, UIUC



https://courses.engr.illinois.edu/cs421/fa2017/CS421A

Based in part on slides by Mattox Beckman, as updated by Vikram Adve, Gul Agha, and Elsa L Gunter

10/4/2017

Terminology

- Type: A type t defines a set of possible data values
 - E.g. short in C is $\{x \mid 2^{15} 1 \ge x \ge -2^{15}\}$
 - A value in this set is said to have type t

Type system: rules of a language assigning types to expressions

Why Data Types?

- Data types play a key role in:
 - Data abstraction in the design of programs
 - Type checking in the analysis of programs
 - Compile-time code generation in the translation and execution of programs
 - Data layout (how many words; which are data and which are pointers) dictated by type

Types as Specifications

- Types describe properties
- Different type systems describe different properties:
 - Data is read-write versus read-only
 - Operation has authority to access data
 - Data came from "right" source
 - Operation might or could not raise an exception
- Common type systems focus on types describing same data layout and access methods

Sound Type System

- Type: A type t defines a set of possible data values
 - E.g. short in C is $\{x \mid 2^{15} 1 \ge x \ge -2^{15}\}$
 - A value in this set is said to have type t
- Type system: rules of a language assigning types to expressions
- If an expression is assigned type t, and it evaluates to a value v, then v is in the set of values defined by t
- SML, OCAML, Scheme and Ada have sound type systems
- Most implementations of C and C++ do not

Strongly Typed Language

- When no application of an operator to arguments can lead to a run-time type error, language is strongly typed
 Eg: I + 2.3;;
- Depends on definition of "type error"

Strongly Typed Language

- C++ claimed to be "strongly typed", but
 - Union types allow creating a value at one type and using it at another
 - Type coercions may cause unexpected (undesirable) effects
 - No array bounds check (in fact, no runtime checks at all)
- SML, OCAML "strongly typed" but still must do dynamic array bounds checks, runtime type case analysis, and other checks

Static vs Dynamic Types

- **Static type:** type assigned to an expression at compile time
- Dynamic type: type assigned to a storage location at run time
- Statically typed language: static type assigned to every expression at compile time
- Dynamically typed language: type of an expression determined at run time

Type Checking

- When is op(arg1,...,argn) allowed?
- Type checking assures that operations are applied to the right number of arguments of the right types
 - Right type may mean same type as was specified, or may mean that there is a predefined implicit coercion that will be applied
- Used to resolve overloaded operations

Type Checking

- Type checking may be done statically at compile time or dynamically at run time
- Dynamically typed (aka untyped) languages (eg LISP, Prolog, JavaScript) do only dynamic type checking
- Statically typed languages can do most type checking statically

Dynamic Type Checking

- Performed at run-time before each operation is applied
- Types of variables and operations left unspecified until run-time
 - Same variable may be used at different types

Dynamic Type Checking

 Data object must contain type information
 Errors aren't detected until violating application is executed (maybe years after the code was written)

Static Type Checking

- Performed after parsing, before code generation
- Type of every variable and signature of every operator must be known at compile time

Static Type Checking

- Can eliminate need to store type information in data object if no dynamic type checking is needed
- Catches many programming errors at earliest point
- Can't check types that depend on dynamically computed values
 - Eg: array bounds

Static Type Checking

- Typically places restrictions on languages
 - Garbage collection
 - References instead of pointers
 - All variables initialized when created
 - Variable only used at one type
 - Union types allow for work-arounds, but effectively introduce dynamic type checks

Type Declarations

- Type declarations: explicit assignment of types to variables (signatures to functions) in the code of a program
 - Must be checked in a strongly typed language
 - Often not necessary for strong typing or even static typing (depends on the type system)

Type Inference

- Type inference: A program analysis to assign a type to an expression from the program context of the expression
 - Fully static type inference first introduced by Robin Miller in ML
 - Haskel, OCAML, SML all use type inference
 Records are a problem for type inference

Format of Type Judgments

A type judgement has the form

 $\Gamma \mid - \exp : \tau$

- Γ is a typing environment
 - Supplies the types of variables (and function names when function names are not variables)
 - Γ is a set of the form $\{x:\sigma,\ldots\}$
 - For any x at most one σ such that $(x : \sigma \in \Gamma)$
- exp is a program expression
- τ is a type to be assigned to exp
- pronounced "turnstyle", or "entails" (or "satisfies" or, informally, "shows") 10/4/2017

Axioms - Constants

 $\Gamma \mid -n : int$ (assuming *n* is an integer constant)

 Γ |- true : bool Γ |- false : bool

These rules are true with any typing environment
 Γ, n are meta-variables

Axioms – Variables (Monomorphic Rule)

Notation: Let $\Gamma(x) = \sigma$ if $x : \sigma \in \Gamma$ Note: if such σ exits, its unique

Variable axiom:

$$\Gamma \mid -x : \sigma$$
 if $\Gamma(x) = \sigma$

Simple Rules - Arithmetic

Primitive operators ($\bigoplus \in \{+, -, *, ...\}$): $\Gamma \mid -e_1:\tau_1 \quad \Gamma \mid -e_2:\tau_2 \quad (\oplus):\tau_1 \to \tau_2 \to \tau_3$ $\Gamma \mid - \mathbf{e}_1 \oplus \mathbf{e}_2 : \tau_3$ Relations ($\sim \in \{ <, >, =, <=, >= \}$): $\Gamma \mid - \mathbf{e}_1 : \tau \quad \Gamma \mid - \mathbf{e}_2 : \tau$ $\Gamma \mid - e_1 \sim e_2$:bool

For the moment, think τ is int

What do we need to show first?

$\{x:int\} | - x + 2 = 3 : bool$

What do we need for the left side?

$$\frac{\{x : int\} |-x + 2 : int \\ \{x:int\} |-x + 2 = 3 : bool }$$
Rel

How to finish?

$$\frac{\{x:int\} |- x:int \{x:int\} |- 2:int AO \\ \{x:int\} |- x + 2:int \{x:int\} |- 3:int \\ \{x:int\} |- x + 2 = 3:bool \ Rel$$

Complete Proof (type derivation)



Simple Rules - Booleans

Connectives

 $\Gamma \mid -e_{1} : bool \qquad \Gamma \mid -e_{2} : bool$ $\Gamma \mid -e_{1} & \& & e_{2} : bool$ $\Gamma \mid -e_{1} : bool \qquad \Gamma \mid -e_{2} : bool$ $\Gamma \mid -e_{1} \mid \mid e_{2} : bool$

Type Variables in Rules

If then else rule: $\Gamma \mid - e_1 : bool \quad \Gamma \mid - e_2 : \tau \quad \Gamma \mid - e_3 : \tau$

 $\Gamma \mid$ - (if e_1 then e_2 else e_3): τ

- τ is a type variable (meta-variable)
- Can take any type at all
- All instances in a rule application must get same type
- Then branch, else branch and if_then_else must all have same type

Function Application

Application rule:

$$\Gamma \mid - e_1 : \tau_1 \to \tau_2 \quad \Gamma \mid - e_2 : \tau_1$$
$$\Gamma \mid - (e_1 \ e_2) : \tau_2$$

If you have a function expression e_1 of type $\tau_1 \rightarrow \tau_2$ applied to an argument e_2 of type τ_1 , the resulting expression e_1e_2 has type τ_2

Fun Rule

- Rules describe types, but also how the environment <a>\[may change
- Can only do what rule allows!
- fun rule:

$$\{x:\tau_1\} + \Gamma \mid -e:\tau_2$$

$$\Gamma \mid -fun \ x \rightarrow e:\tau_1 \rightarrow \tau_2$$

Fun Examples

$${x : \tau_1} + \Gamma - e : \tau_2$$

 $\Gamma \mid -\text{ fun } x \rightarrow e : \tau_1 \rightarrow \tau_2$

$$\{y : int \} + \Gamma \mid -y + 3 : int$$
$$\Gamma \mid - fun \ y \ -> y + 3 : int \rightarrow int$$

 $\begin{array}{l} \{f: \operatorname{int} \to \operatorname{bool}\} + \Gamma \mid -f 2 :: [\operatorname{true}] : \operatorname{bool} \operatorname{list} \\ \Gamma \mid -(\operatorname{fun} f -> f 2 :: [\operatorname{true}]) \\ : (\operatorname{int} \to \operatorname{bool}) \to \operatorname{bool} \operatorname{list} \end{array}$

(Monomorphic) Let and Let Rec

Iet rule:

$$\frac{\Gamma |-e_{1}:\tau_{1} \{x:\tau_{1}\} + \Gamma |-e_{2}:\tau_{2}}{\Gamma |-(\text{let } x = e_{1} \text{ in } e_{2}):\tau_{2}}$$

let rec rule:

 $\{x: \tau_1\} + \Gamma \mid -e_1:\tau_1 \qquad \{x: \tau_1\} + \Gamma \mid -e_2:\tau_2$ $\Gamma \mid -(\text{let rec } x = e_1 \text{ in } e_2): \tau_2$

let rec rule: Example $\{x: \tau_1\} + \Gamma \mid -e_1:\tau_1 \qquad \{x: \tau_1\} + \Gamma \mid -e_2:\tau_2$ Γ |- (let rec x = e₁ in e₂): τ_2 Which rule do we apply? 7 |-(let rec one = 1 :: one in)|let x = 2 in fun y -> (x :: y :: one)) : int \rightarrow int list

Example

Let rec rule: (2) {one : int list} |-(let x = 2 in)fun y -> (x :: y :: one)) {one : int list} |-(1 :: one) : int list : int \rightarrow int list |-(let rec one = 1 :: one in)|let x = 2 in fun y -> (x :: y :: one)) : int \rightarrow int list

Proof of I

Application rule:

$$\frac{\Gamma \mid -e_1 : \tau_1 \rightarrow \tau_2 \quad \Gamma \mid -e_2 : \tau_1}{\Gamma \mid -(e_1 \mid e_2) : \tau_2}$$

Which rule?

{one : int list} |- (| :: one) : int list

Proof of I

Application rule:

 $\Gamma \mid - \mathbf{e}_1 : \tau_1 \to \tau_2 \quad \Gamma \mid - \mathbf{e}_2 : \tau_1$

$$\Gamma |- (e_1 e_2) : \tau_2$$

Application

3
(3)
(4)
{one : int list} |((::) |): int list→ int list
one : int list
one : int list} |- (| :: one) : int list

Proof of 3

Constants Rule Constants Rule

 $\{ \text{one : int list} \} | - \qquad \{ \text{one : int list} \} | - \\ (::) : \text{int} \rightarrow \text{int list} \rightarrow \text{int list} \qquad I : \text{int} \\ \{ \text{one : int list} \} | - ((::) |) : \text{int list} \rightarrow \text{int list} \\ \end{cases}$


Rule for variables

{one : int list} |- one : int list

(5) {x:int; one : int list} |fun y -> Constant (x :: y :: one)) $\{\text{one}: \text{int list}\} \mid -2: \text{int} : \text{int} \rightarrow \text{int list} \}$ $\{\text{one}: \text{int list}\} \mid - (\text{let } x = 2 \text{ in})$ fun y -> (x :: y :: one)) : int \rightarrow int list

? {x:int; one : int list} |- fun y -> (x :: y :: one)) : int → int list

? $\{y:int; x:int; one : int list\} \mid - (x :: y :: one) : int list$ $\{x:int; one : int list\} \mid - fun y \rightarrow (x :: y :: one))$ $: int \rightarrow int list$



Constant

Variable

{...} |- (::)



{y:int; ...} |- ((::) y) {...; one: int list} |-:int list int list one: int list

{y:int; x:int; one : int list} |- (y :: one) : **int list**

Curry - Howard Isomorphism

- Type Systems are logics; logics are type systems
- Types are propositions; propositions are types
- Terms are proofs; proofs are terms
- Function space arrow corresponds to implication; application corresponds to modus ponens

Curry - Howard Isomorphism

Modus Ponens



Application

$$\Gamma \mid - e_{1} : \alpha \rightarrow \beta \quad \Gamma \mid - e_{2} : \alpha$$
$$\Gamma \mid - (e_{1} e_{2}) : \beta$$

Mea Culpa

- The above system can't handle polymorphism as in OCAML
- No type variables in type language (only meta-variable in the logic)
- Would need:
 - Object level type variables and some kind of type quantification
 - Iet and let rec rules to introduce polymorphism
 - Explicit rule to eliminate (instantiate) polymorphism

Support for Polymorphic Types

- Monomorpic Types (τ):
 - Basic Types: int, bool, float, string, unit, ...
 - Type Variables: α , β , γ , δ , ϵ
 - Compound Types: $\alpha \rightarrow \beta$, int * string, bool list, ...
- Polymorphic Types:
 - Monomorphic types τ
 - Universally quantified monomorphic types
 - $\forall \alpha_1, \ldots, \alpha_n \cdot \tau$
 - Can think of τ as same as \forall . τ

Support for Polymorphic Types

- Free variables of monomorphic type just type variables that occur in it
 - Write FreeVars(τ)
- Free variables of polymorphic type removes variables that are universally quantified
 - FreeVars($\forall \alpha_1, \ldots, \alpha_n \cdot \tau$) = FreeVars(τ) { $\alpha_1, \ldots, \alpha_n$ }
- FreeVars(Γ) = all FreeVars of types in range of Γ

Monomorphic to Polymorphic

Given:

- type environment
- monomorphic type τ
- τ shares type variables with Γ
- Want most polymorphic type for τ that doesn' t break sharing type variables with Γ
- Gen(τ , Γ) = $\forall \alpha_1, \dots, \alpha_n \cdot \tau$ where
 - $\{\alpha_1, \ldots, \alpha_n\} = \text{freeVars}(\tau) \text{freeVars}(\Gamma)$

Polymorphic Typing Rules

A type judgement has the form

 $\Gamma \mid$ - exp : τ

- Γ uses polymorphic types
- τ still monomorphic
- Most rules stay same (except use more general typing environments). Rules that change:
 - Variables
 - Let and Let Rec
 - Allow polymorphic constants
- Worth noting functions again

Polymorphic Let and Let Rec

- Interpretation of the second state of the
- Interval in the second sec

Polymorphic Variables (Identifiers)



Fun Rule Stays the Same

fun rule:

$$\{x:\tau_1\} + \Gamma \mid -e:\tau_2$$
$$\Gamma \mid - \text{ fun } x \rightarrow e:\tau_1 \rightarrow \tau_2$$

Types τ_1 , τ_2 monomorphic

Function argument must always be used at same type in function body

Polymorphic Example

- Assume additional constants:
- hd : $\forall \alpha \cdot \alpha$ list -> α
- tl: $\forall \alpha . \alpha$ list -> α list
- is_empty : $\forall \alpha \cdot \alpha$ list -> bool
- ::: $\forall \alpha . \alpha \rightarrow \alpha$ list -> α list
- [] : ∀ α . α list

Polymorphic Example

Show:

?

{} |- let rec length =

fun I -> if is_empty I then 0 else I + length (tl I)

in

length ((::) 2 []) + length((::) true []) : int

10/4/2017

Polymorphic Example: Let Rec Rule (Repeat)

Show: (1) (2) {length: $\forall \alpha. \alpha \text{ list -> int}$ } {length: α list -> int} |- fun lst -> ... |- length ((::) 2 []) + length((::) true []) : int $: \alpha$ list -> int {} |- let rec length = fun |st -> if is empty |st then 0else I + length (tl lst) in length ((::) 2 []) + length((::) true []) : int

10/4/2017

Polymorphic Example (1)

Show:

?

{length:α list -> int} |fun lst -> if is_empty lst then 0
 else I + length (tl lst)
: α list -> int

Polymorphic Example (1): Fun Rule

Show: (3) {length: α list -> int, lst: α list } |if is empty lst then 0 else length (hd l) + length (tl lst) : int {length: α list -> int} |fun |st -> if is empty |st then 0else I + length (tl lst) : α list -> int

Polymorphic Example (3)

Let Γ ={length:α list -> int, lst: α list }
Show

?

Γ|- if is_empty | then 0 else | + length (tl lst) : int

Polymorphic Example (3):IfThenElse

Let Γ ={length:α list -> int, lst: α list }
Show

Polymorphic Example (4)

Let Γ ={length:α list -> int, lst: α list }
Show

?

Γ |- is_empty lst : **bool**

Polymorphic Example (4): Application

Let Γ ={length:α list -> int, lst: α list }
Show



Polymorphic Example (4)

Let Γ ={length:α list -> int, lst: α list }
Show

By Const since α list -> bool is instance of $\forall \alpha. \alpha$ list -> bool ?

 $\Gamma |- is_empty : \alpha list -> bool \qquad \Gamma |- lst : \alpha list$ $\Gamma |- is_empty lst : bool$

Polymorphic Example (4)

- Let Γ ={length:α list -> int, l: α list }
 Show
- By Const since α list -> bool is By Variable instance of $\forall \alpha$. α list -> bool $\Gamma(\text{lst}) = \alpha$ list

 $\Gamma|- is_empty : \alpha list -> bool \qquad \Gamma|- lst : \alpha list \\ \Gamma|- is_empty lst : bool$ This finishes (4) Polymorphic Example (3):IfThenElse (Repeat)

Let Γ ={length:α list -> int, lst: α list }
Show

$$\Gamma$$
|- if is_empty | then 0
else | + length (tl lst) : int

Polymorphic Example (5):Const

- Let Γ ={length:α list -> int, lst: α list }
 Show
- By Const Rule

 Γ |- 0:int

Polymorphic Example (6):Arith Op

- Let Γ ={length:α list -> int, lst: α list }
 Show
- By Variable(7) $\Gamma|$ length $\Gamma|$ (tl lst)By Const: α list -> int: α list -> int<td: : α list $\Gamma|$ I: int $\Gamma|$ length (tl lst) : int $\Gamma|$ I + length (tl lst) : int

Polymorphic Example (7):App Rule

- Let $\Gamma = \{ \text{length}: \alpha \text{ list -> int, lst: } \alpha \text{ list } \}$
- Show

By Const

 $\Gamma|$ - (tl lst) : α list -> α list

By Variable

Γ |-lst: α list

 Γ |- (tl lst) : α list

By Const since α list -> α list is instance of $\forall \alpha. \alpha \text{ list -> } \alpha \text{ list}$

Polymorphic Example: Let Rec Rule (Repeat)

in

length ((::) 2 []) + length((::) true []) : int

10/5/2017

Polymorphic Example: (2) by ArithOp

Let Γ' = {length: ∀α. α list -> int}
 Show:

(8) (9)

Γ' |- Γ' |length ((::) 2 []) :int length((::) true []) : int
{length: α. α list -> int}
|- length ((::) 2 []) + length((::) true []) : int

Polymorphic Example: (8)AppRule

Let Γ' = {length:∀α. α list -> int}
 Show:

$\Gamma' \mid - \text{ length : int list ->int } \Gamma' \mid - ((::)2 []) : \text{ int list}$ $\Gamma' \mid - \text{ length } ((::) 2 []) : \text{ int}$

Polymorphic Example: (8)AppRule

• Let $\Gamma' = \{ \text{length}: \forall \alpha. \alpha \text{ list -> int} \}$

Show:

By Var since int list -> int is instance of $\forall \alpha. \alpha \text{ list -> int}$

(10) $\Gamma' \mid - \text{ length : int list ->int}$ $\Gamma' \mid - ((::)2 []):int list$

Γ' |- length ((::) 2 []) : int
Polymorphic Example: (10)AppRule

- Let $\Gamma' = \{ \text{length}: \forall \alpha. \alpha \text{ list -> int} \}$
- Show:
- By Const since α list is instance of $\forall \alpha. \ \alpha$ list

(||)

Polymorphic Example: (11)AppRule

- Let $\Gamma' = \{ \text{length:} \forall \alpha. \alpha \text{ list -> int} \}$
- Show:
- By Const since α list
 - is instance of
 - $\forall \alpha. \alpha \text{ list}$

By Const Γ' |- 2 : int

Γ' |- (::) : int -> int list -> int list

Γ' |- ((::) 2) : int list -> int list

Polymorphic Example: (9)AppRule

Let Γ' = {length:∀α. α list -> int}
Show:

Polymorphic Example: (9)AppRule

- Let $\Gamma' = \{ \text{length:} \forall \alpha. \alpha \text{ list -> int} \}$
- Show:
- By Var since bool list -> int is instance of $\forall \alpha. \alpha$ list -> int

(12) $\Gamma' \mid \Gamma' \mid$ length : **bool list ->int** ((::) true []) :**bool list** $\Gamma' \mid -$ length ((::) true []) :**int**

Polymorphic Example: (12)AppRule

- Let $\Gamma' = \{ \text{length}: \forall \alpha. \alpha \text{ list -> int} \}$
- Show:
- By Const since α list is instance of $\forall \alpha$. α list

(13) $\Gamma' |-((::)true):$ bool list ->bool list $\Gamma' |- []:$ bool list $\Gamma' |- ((::) true []) :$ bool list

Polymorphic Example: (13)AppRule

• Let $\Gamma' = \{ \text{length}: \forall \alpha. \alpha \text{ list -> int} \}$

Show:

By Const since bool list

is instance of $\forall \alpha$. α list

By Const $\overline{\Gamma}$ |-(::):bool ->bool list ->bool list true : **boo**

 $\Gamma' \mid -((::) \text{ true}) : \text{bool list -> bool list}$

Γ'Ι-