Today

This first part of this lecture is about subtyping

- Let more terms type-check without adding any new operational behavior
  - But at end consider coercions
- Continue using STLC as our core model
- Complementary to type variables which we will do later
  - Parametric polymorphism ($\forall \theta$), a.k.a. generics
  - First-class ADTs ($\exists$
- Even later: OOP, dynamic dispatch, inheritance vs. subtyping

Motto: Subtyping is not a matter of opinion!

Records

We'll use records to motivate subtyping:

\[
\begin{align*}
e &::= \ldots | \{l_1 = e_1, \ldots, l_n = e_n\} | e.l \\
\tau &::= \ldots | \{l_1 : \tau_1, \ldots, l_n : \tau_n\} \\
v &::= \ldots | \{l_1 = v_1, \ldots, l_n = v_n\} \\
\{l_1 = v_1, \ldots, l_n = v_n\} &\rightarrow v_i \\
e_i &\rightarrow e_i' \\
e.l &\rightarrow e.l \\
\end{align*}
\]

True or False?

\[
\begin{align*}
\Gamma \vdash e_1 : \tau_1 \quad \ldots \quad \Gamma \vdash e_n : \tau_n \\
\Gamma \vdash \{l_1 = e_1, \ldots, l_n = e_n\} : \{l_1 : \tau_1, \ldots, l_n : \tau_n\} \\
\end{align*}
\]

Should this typecheck?

\[
(\lambda x : \{l_1:int, l_2:int\}. x.l_1 + x.l_2)\{l_1=3, l_2=4, l_3=5\}
\]

Right now, it doesn’t, but it won’t get stuck

Suggests width subtyping:

\[
\begin{align*}
\tau_1 &\leq \tau_2 \\
\end{align*}
\]

And one new type-checking rule: Subsumption

\[
\begin{align*}
\frac{\Gamma \vdash e : \tau' \quad \tau' \leq \tau}{\Gamma \vdash e : \tau}
\end{align*}
\]
Now it type-checks

\[ \vdash \text{3 : int} \quad \vdash \text{4 : int} \quad \vdash \text{5 : int} \]

\[ \vdash : \{(\text{l1:int}, \text{l2:int})\} \vdash \text{x.l1 + x.l2 : int} \]

\[ \vdash \lambda x: \{(\text{l1:int}, \text{l2:int})\}.x.l1 + x.l2 : \{(\text{l1:int}, \text{l2:int})\} \rightarrow \ldots \]

Instantiation of Subsumption is highlighted (pardon formatting)

The derivation of the subtyping fact
\{l1:int, l2:int, l3:int\} ≤ \{l1:int, l2:int\} would continue, using rules for the τ₁ ≤ τ₂ judgment

▶ But here we just use the one axiom we have so far

Clean division of responsibility:

▶ Where to use subsumption

▶ How to show two types are subtypes

Permutation

Does this program type-check? Does it get stuck?

\{(Ax: \{l1:int, l2:int\}. x.l1 + x.l2{\text{l2}=3; l1=4}\}

Suggests permutation subtyping:

\{l1:τ₁₁, ..., l₁₋₁:τ₁₋₁, l₁:τ₁, ..., lₙ:τₙ\} ≤ \{l₁:τ₁₁, ..., l₁:τ₁, l₁₋₁:τ₁₋₁, ..., lₙ:τₙ\}

Example with width and permutation: Show

\( \vdash \{l1=7, l2=8, l3=9\} : \{l2:int, l1:int\} \)

It’s no longer clear there is an (efficient, sound, complete) type-checking algorithm

▶ They sometimes exist and sometimes don’t

▶ Here they do

Transitivity

Subtyping is always transitive, so add a rule for that:

\[
\frac{\tau₁ ≤ τ₂ \quad \tau₂ ≤ \tau₃}{\tau₁ ≤ \tau₃}
\]

Or just use the subtyping rule multiple times. Or both.

In any case, type-checking is no longer syntax-directed: There may be 0, 1, or many different derivations of \(\Gamma \vdash e : \tau\)

▶ And also potentially many ways to show \(\tau₁ ≤ τ₂\)

Hopefully we could define an algorithm and prove it “answers yes” if and only if there exists a derivation

Digression continued

With width subtyping alone, the strategy is easy

With permutation subtyping alone, it’s easy but have to “alphabetize”

With both, it’s not easy...

\[ f₁ : \{l₁ : \text{int}\} → \text{int} \quad f₂ : \{l₂ : \text{int}\} → \text{int} \]

\[ x₁ = \{l₁ = 0, l₂ = 0\} \quad x₂ = \{l₂ = 0, l₃ = 0\} \]

\[ f₁(x₁) \quad f₂(x₁) \quad f₂(x₂) \]

Can use dictionary-passing (look up offset at run-time) and maybe optimize away (some) lookups

Named types can avoid this, but make code less flexible

So far

▶ A new subtyping judgement and a new typing rule subsumption

▶ Width, permutation, and transitivity

\[
\frac{\tau₁ ≤ τ₂}{\{l₁:τ₁₁, ..., lₙ:τₙ, l:τ\} ≤ \{l₁:τ₁₁, ..., lₙ:τₙ\}}
\]

\[
\frac{\{l₁:τ₁₁, ..., l₁₋₁:τ₁₋₁, l₁:τ₁, ..., lₙ:τₙ\} ≤ \{l₁:τ₁₁, ..., l₁:τ₁, l₁₋₁:τ₁₋₁, ..., lₙ:τₙ\}}{\tau₁ ≤ τ₂ \quad \tau₂ ≤ \tau₃}
\]

Now: This is all much more useful if we extend subtyping so it can be used on “parts” of larger types:

▶ Example: Can’t yet use subsumption on a record field’s type

▶ Example: There are no supertypes yet of \(τ₁ → τ₂\)

Digression continued

With width subtyping alone, the strategy is easy

With permutation subtyping alone, it’s easy but have to “alphabetize”

With both, it’s not easy...

\[ f₁ : \{l₁ : \text{int}\} → \text{int} \quad f₂ : \{l₂ : \text{int}\} → \text{int} \]

\[ x₁ = \{l₁ = 0, l₂ = 0\} \quad x₂ = \{l₂ = 0, l₃ = 0\} \]

\[ f₁(x₁) \quad f₂(x₁) \quad f₂(x₂) \]

Can use dictionary-passing (look up offset at run-time) and maybe optimize away (some) lookups

Named types can avoid this, but make code less flexible
Depth

Does this program type-check? Does it get stuck?

\((\lambda x : \{l_1 : \{l_3 : \text{int}\}, l_2 : \text{int}\}. x.l_1.l_3 + x.l_2)\{l_1 = \{l_3 = 3, l_4 = 9\}, l_2 = 4\}\)

Suggests depth subtyping

\[ \tau_1 \leq \tau'_1 \]

\[ \{l_1 : \tau_1, \ldots, l_i : \tau_i, \ldots, l_n : \tau_n\} \leq \{l_1 : \tau'_1, \ldots, l_i : \tau'_i, \ldots, l_n : \tau_n\} \]

(With permutation subtyping, can just have depth on left-most field)

Soundness of this rule depends crucially on fields being immutable!

\( \triangleright \) Depth subtyping is unsound in the presence of mutation

\( \triangleright \) Trade-off between power (mutation) and sound expressiveness (depth subtyping)

Function subtyping

Given our rich subtyping on records (and/or other primitives), how do we extend it to other types, notably \( \tau_1 \rightarrow \tau_2 \)?

For example, we’d like \( \text{int} \rightarrow \{l_1 : \text{int}, l_2 : \text{int}\} \leq \text{int} \rightarrow \{l_1 : \text{int}\} \) so we can pass a function of the subtype somewhere expecting a function of the supertype

\[ \tau_1 \rightarrow \tau_2 \leq \tau_3 \rightarrow \tau_4 \]

For a function to have type \( \tau_3 \rightarrow \tau_4 \) it must return something of type \( \tau_3 \) (including subtypes) whenever given something of type \( \tau_3 \) (including subtypes). A function assuming less than \( \tau_3 \) will do, but not one assuming more. A function returning more than \( \tau_4 \) but not one returning less.

Function subtyping, cont’d

\[ \tau_3 \leq \tau_1 \quad \tau_2 \leq \tau_4 \quad \text{Also want: } \tau \leq \tau \]

Example: \( \lambda x : \{l_1 : \text{int}, l_2 : \text{int}\}. \{l_1 = x.l_2, l_2 = x.l_1\} \)

can have type \( \{l_1 : \text{int}, l_2 : \text{int}, l_3 : \text{int}\} \rightarrow \{l_1 : \text{int}\} \)

but not \( \{l_1 : \text{int}\} \rightarrow \{l_1 : \text{int}\} \)

Jargon: Function types are contravariant in their argument and covariant in their result

\( \triangleright \) Depth subtyping means immutable records are covariant in their fields

This is unintuitive enough that you, a friend, or a manager, will some day be convinced that functions can be covariant in their arguments. THIS IS ALWAYS WRONG (UNSOUND).

Somewhere to the north, a PL professor JUMPED UP AND DOWN about this.

Covariance and Contravariance

Given types \( S \) and \( T \) such that \( S \leq T \) (also written as \( S < : T \)):

\( \triangleright \) Covariant: \( S \) and \( T \) are said to be covariant when the more specific type, \( S \) can be used where the more generic type, \( T \), is specified. This applies to functions, i.e., a function that returns \( S \) can be used in the same context as a function that returns \( T \).

\( \triangleright \) Contravariant: \( S \) and \( T \) are contra variant when the more generic type, \( T \) can be used where the more specific type, \( S \), is specified. A function that takes an argument of type \( T \) can be used in the same context as a function that takes an argument of type \( S \).

\( \triangleright \) Invariant: The type specified is the only one that can be used.

Covariant Example (C++)

```cpp
class X {}
class Y : public X {}
class Z : public Y {}

class A {
public:
    virtual Y *foo() { return new Y(); }
};

class B : public A {
public:
    virtual Z *foo() { return new Z(); }
};
```

Here we have three classes X, Y, and Z which we will return from a virtual function in classes A and B. This code is valid because B::foo is returning a narrower type than A::foo because Z is a subtype of Y.

Covariant Example (C++)

But what happens if we make B::foo return a wider type?

```cpp
class X {}
class Y : public X {}
class Z : public Y {}

class A {
public:
    virtual Y *foo() { return new Y(); }
};

class B : public A {
public:
    virtual X *foo() { return new X(); } // 12
};
```
Covariant Example (C++)

But what happens if we make B::foo return a wider type?

```c++
g++ -Wall -ansi -pedantic -c t.cpp
t.cpp:12:14: error: return type of virtual function 'foo'
    is not covariant with the return type of
    the function it overrides ('X *' is not derived from 'Y *')
virtual X *foo() { return new X(); }
^ 1 error generated.
```

```c++
int main()

class A {
public:
    virtual Y *foo() { return new Y(); }
};
```

```c++
class X {};
class Y : public X {};
class Z : public Y {};
class A {
public:
    virtual void foo(Y &y) { }
};

class B : public A {
public:
    virtual void foo(Y &y) { }
    virtual void foo(Z &z) { }
};
```

```
return 0;
```

```
int main()

class X {};
class Y : public X {};
class Z : public Y {};
class A {
public:
    virtual void foo(Y &y) { }
};

class B : public A {
public:
    virtual void foo(Y &y) { }
    virtual void foo(Z &z) { }
};
```

```
return 0;
```

```
int main()

class X {};
class Y : public X {};
class Z : public Y {};
class A {
public:
    virtual void foo(Y &y) { }
};

class B : public A {
public:
    virtual void foo(Y &y) { }
    virtual void foo(Z &z) { }
};
```

```
return 0;
```

```
int main()

class X {};
class Y : public X {};
class Z : public Y {};
class A {
public:
    virtual void foo(Y &y) { }
};

class B : public A {
public:
    virtual void foo(Y &y) { }
    virtual void foo(Z &z) { }
};
```

```
return 0;
```

```
int main()

class X {};
class Y : public X {};
class Z : public Y {};
class A {
public:
    virtual void foo(Y &y) { }
};

class B : public A {
public:
    virtual void foo(Y &y) { }
    virtual void foo(Z &z) { }
};
```

```
return 0;
```

```
int main()

class X {};
class Y : public X {};
class Z : public Y {};
class A {
public:
    virtual void foo(Y &y) { }
};

class B : public A {
public:
    virtual void foo(Y &y) { }
    virtual void foo(Z &z) { }
};
```

```
return 0;
```

```
int main()

class X {};
class Y : public X {};
class Z : public Y {};
class A {
public:
    virtual void foo(Y &y) { }
};

class B : public A {
public:
    virtual void foo(Y &y) { }
    virtual void foo(Z &z) { }
};
```

```
return 0;
```

```
int main()

class X {};
class Y : public X {};
class Z : public Y {};
class A {
public:
    virtual void foo(Y &y) { }
};

class B : public A {
public:
    virtual void foo(Y &y) { }
    virtual void foo(Z &z) { }
};
```

```
return 0;
```

```
int main()

class X {};
class Y : public X {};
class Z : public Y {};
class A {
public:
    virtual void foo(Y &y) { }
};

class B : public A {
public:
    virtual void foo(Y &y) { }
    virtual void foo(Z &z) { }
};
```

```
return 0;
```

```
int main()

class X {};
class Y : public X {};
class Z : public Y {};
class A {
public:
    virtual void foo(Y &y) { }
};

class B : public A {
public:
    virtual void foo(Y &y) { }
    virtual void foo(Z &z) { }
};
```

```
return 0;
```

```
int main()

class X {};
class Y : public X {};
class Z : public Y {};
class A {
public:
    virtual void foo(Y &y) { }
};

class B : public A {
public:
    virtual void foo(Y &y) { }
    virtual void foo(Z &z) { }
};
```

```
return 0;
```

```
int main()

class X {};
class Y : public X {};
class Z : public Y {};
class A {
public:
    virtual void foo(Y &y) { }
};

class B : public A {
public:
    virtual void foo(Y &y) { }
    virtual void foo(Z &z) { }
};
```

```
return 0;
```

```
int main()

class X {};
class Y : public X {};
class Z : public Y {};
class A {
public:
    virtual void foo(Y &y) { }
};

class B : public A {
public:
    virtual void foo(Y &y) { }
    virtual void foo(Z &z) { }
};
```

```
return 0;
```

```
int main()

class X {};
class Y : public X {};
class Z : public Y {};
class A {
public:
    virtual void foo(Y &y) { }
};

class B : public A {
public:
    virtual void foo(Y &y) { }
    virtual void foo(Z &z) { }
};
```

```
return 0;
```

```
int main()

class X {};
class Y : public X {};
class Z : public Y {};
class A {
public:
    virtual void foo(Y &y) { }
};

class B : public A {
public:
    virtual void foo(Y &y) { }
    virtual void foo(Z &z) { }
};
```

```
return 0;
```

```
int main()

class X {};
class Y : public X {};
class Z : public Y {};
class A {
public:
    virtual void foo(Y &y) { }
};

class B : public A {
public:
    virtual void foo(Y &y) { }
    virtual void foo(Z &z) { }
};
```

```
return 0;
```

```
int main()

class X {};
class Y : public X {};
class Z : public Y {};
class A {
public:
    virtual void foo(Y &y) { }
};

class B : public A {
public:
    virtual void foo(Y &y) { }
    virtual void foo(Z &z) { }
};
```

```
return 0;
```

```
int main()

class X {};
class Y : public X {};
class Z : public Y {};
class A {
public:
    virtual void foo(Y &y) { }
};

class B : public A {
public:
    virtual void foo(Y &y) { }
    virtual void foo(Z &z) { }
};
```

```
return 0;
```

```
int main()

class X {};
class Y : public X {};
class Z : public Y {};
class A {
public:
    virtual void foo(Y &y) { }
};

class B : public A {
public:
    virtual void foo(Y &y) { }
    virtual void foo(Z &z) { }
};
```

```
return 0;
```

```
int main()

class X {};
class Y : public X {};
class Z : public Y {};
class A {
public:
    virtual void foo(Y &y) { }
};

class B : public A {
public:
    virtual void foo(Y &y) { }
    virtual void foo(Z &z) { }
};
```
Ah, Canonical Forms

That’s because Canonical Forms is where the action is:
- If \( \vdash v : \{ l_1 : \tau_1, \ldots, l_n : \tau_n \} \), then \( v \) is a record with fields \( l_1, \ldots, l_n \)
- If \( \vdash v : \tau_1 \rightarrow \tau_2 \), then \( v \) is a function

We need these for the “interesting” cases of Progress

Now have to use induction on the typing derivation (may end with many subsumptions) and induction on the subtyping derivation (e.g., “going up the derivation” only adds fields)
- Canonical Forms is typically trivial without subtyping; now it requires some work

Note: Without subtyping, Preservation is a little “cleaner” via induction on \( e \rightarrow e' \), but with subtyping it’s much cleaner via induction on the typing derivation
- That’s why we did it that way

A matter of opinion?

If subsumption makes well-typed terms get stuck, it is wrong

We might allow less subsumption (e.g., for efficiency), but we shall not allow more than is sound

But we have been discussing “subset semantics” in which \( e : \tau \) and \( \tau \leq \tau' \) means \( e \) is a \( \tau' \)
- There are “fewer” values of type \( \tau \) than of type \( \tau' \), but not really

Very tempting to go beyond this, but you must be very careful . . .

But first we need to emphasize a really nice property of our current setup: Types never affect run-time behavior

Erasure

A program type-checks or does not. If it does, it evaluates just like in the untyped \( \lambda \)-calculus. More formally, we have:

1. Our language with types (e.g., \( \lambda x : \tau \cdot e, A_\tau + \tau_\tau(e) \), etc.) and a semantics
2. Our language without types (e.g., \( \lambda x \cdot e, A(e) \), etc.) and a different (but very similar) semantics
3. An erasure metafunction from first language to second
4. An equivalence theorem: Erasure commutes with evaluation

This useful (for reasoning and efficiency) fact will be less obvious (but true) with parametric polymorphism

Implementing Coercions

If coercion \( C \) (e.g., float_of_int) “witnesses” \( \tau \leq \tau' \) (e.g., \( \text{int} \leq \text{float} \)), then we insert \( C \) where \( \tau \) is subsumed to \( \tau' \)

So translation to the untyped language depends on where subsumption is used. So it’s from typing derivations to programs.

But typing derivations aren’t unique: uh-oh

Example 1:
- Suppose \( \text{int} \leq \text{float} \) and \( \tau \leq \text{string} \)
- Consider \( \vdash \text{print_string}(34) : \text{unit} \)

Example 2:
- Suppose \( \text{int} \leq \{ l_1 : \text{int} \} \)
- Consider \( 34 == 34 \), where \( == \) is equality on ints or pointers

Coercion Semantics

Wouldn’t it be great if . . .
- \( \text{int} \leq \text{float} \)
- \( \text{int} \leq \{ l_1 : \text{int} \} \)
- \( \tau \leq \text{string} \)
- we could “overload the cast operator”

For these proposed \( \tau \leq \tau' \) relationships, we need a run-time action to turn a \( \tau \) into a \( \tau' \)
- Called a coercion

Could use float_of_int and similar but programmers whine about it

Coherence

Coercions need to be coherent, meaning they don’t have these problems

More formally, programs are deterministic even though type checking is not—any typing derivation for \( e \) translates to an equivalent program

Alternately, can make (complicated) rules about where subsumption occurs and which subtyping rules take precedence
- Hard to understand, remember, implement correctly

It’s a mess . . .
### C++

**Semi-Example: Multiple inheritance a la C++**

```cpp
class C2 {};
class C3 {};
class C1 : public C2, public C3 {};
class D {
    public:
        int f(class C2) { return 0; }
        int f(class C3) { return 1; }
    };
int main() { return D().f(C1()); }
```

Note: A compile-time error "ambiguous call"

Note: Same in Java with interfaces (“reference is ambiguous”)

### Downcasts

Can’t deny downcasts exist, but here are some bad things about them:

- Types don’t erase – you need to represent \( \tau \) and \( e_1 \)’s type at run-time. (Hidden data fields)
- Breaks abstractions: Before, passing \( \{ l_1 = 3, l_2 = 4 \} \) to a function taking \( \{ l_1 : \text{int} \} \) hid the \( l_2 \) field, so you know it doesn’t change or affect the callee

Some better alternatives:

- Use ML-style datatypes — the programmer decides which data should have tags
- Use parametric polymorphism — the right way to do container types (not downcasting results)

### Upcasts and Downcasts

- “Subset” subtyping allows "upcasts"
- “Coercive subtyping” allows casts with run-time effect
- What about “downcasts”?

That is, should we have something like:

```cpp
if_hastype(\tau, e_1)\ then\ x. e_2\ else\ e_3\n```

Roughly, if at run-time \( e_1 \) has type \( \tau \) (or a subtype), then bind it to \( x \) and evaluate \( e_2 \). Else evaluate \( e_3 \). Avoids having exceptions.

> Not hard to formalize

### Parametric Polymorphism

Done with subtyping.

Now: Parametric polymorphism

When type inference determines than an expression is valid for any type it is automatically made polymorphic. In OCaml, the polymorphic types in type expressions are denoted ‘\( a \), ‘\( b \), ‘\( c \) and so on. For example, the following function reverses the order of the elements in a 2-tuple (pair) and can be applied to pairs of values of any type:

```ocaml
# let rev2 (x, y) = (y, x);;

- : (\text{int} * \text{int}) \to (\text{int} * \text{int}) = <fun>
```

### Goal

Understand what this interface means and why it matters:

- Different lists with elements of different types
- New reusable functions outside of library, e.g.:
  ```ocaml
  val twocons : ‘a -> ‘a -> ‘a list
  val decons : ‘a list -> (‘a * ‘a list) option
  val length : ‘a list -> int
  val map : (‘a -> ‘b) -> ‘a list -> ‘b list
  ```

From two perspectives:

1. Library: Implement code to this partial specification
2. Client: Use code written to this partial specification

### What The Client Likes

1. Library is reusable. Can make:
   - Different lists with elements of different types
   - New reusable functions outside of library, e.g.:

   ```ocaml
   val twocons : ‘a -> ‘a -> ‘a list
   val decons : ‘a list -> (‘a * ‘a list) option
   val length : ‘a list -> int
   val map : (‘a -> ‘b) -> ‘a list -> ‘b list
   ```

2. Easier, faster, more reliable than subtyping
   - No downcast to write, run, maybe-fail (cf. Java 1.4 Vector)

3. Library must "behave the same" for all "type instantiations"!
   - ‘\( a \) and ‘\( b \) held abstract from library
   - E.g., with built-in lists: If foo has type ‘\( a \) list -> int, then foo \{ [1;2;3] \} and foo \{ [5;4];(7;2);(9;2) \} are totally equivalent!
   (Never true with downcasts)
   - In theory, means less (re-)integration testing
   - Proof is beyond this course, but not much
What the Library Likes

1. Reusability — For same reasons client likes it
2. Abstraction of mylist from clients
   - Clients must "behave the same" for all equivalent implementations, even if "hidden definition" of 'a mylist changes
   - Clients typechecked knowing only there exists a type constructor mylist
   - Unlike Java, C++, R5RS Scheme, no way to downcast a t mylist to, e.g., a pair

Informal semantics

1. \( \Lambda \alpha. e \): A value that, when used, runs \( e \) (with some \( \tau \) for \( \alpha \))
   - To type-check \( e \), know \( \alpha \) is one type, but not which type
2. \( e[\tau] \): Evaluate \( e \) to some \( \Lambda \alpha. e' \) and then run \( e' \)
   - With \( \tau \) for \( \alpha \), but the choice of \( \tau \) is irrelevant at run-time
   - \( \tau \) used for type-checking and proof of Preservation
3. Types can use type variables \( \alpha, \beta \), etc., but only if they're in scope (just like term variables)
   - Type-checking will be \( \Delta; \Gamma \vdash e : \tau \) using \( \Delta \) to know what type variables are in scope in \( e \)
   - In universal type \( \forall \alpha. \tau \), can also use \( \alpha \) in \( \tau \)

Syntax

\[
e ::= c \mid x \mid \lambda x: \tau. e \mid ee \mid \Lambda \alpha. e \mid e[\tau]
\]
\[
\tau ::= int \mid \tau \rightarrow \tau \mid \alpha \mid \forall \alpha. \tau
\]
\[
v ::= c \mid \lambda x: \tau. e \mid \Lambda \alpha. e
\]
\[
\Gamma ::= \cdot \mid \Gamma, x: \tau
\]
\[
\Delta ::= \cdot \mid \Delta, x: \tau
\]

New things:
- Terms: Type abstraction and type application
- Types: Type variables and universal types
- Type contexts: what type variables are in scope

Operational semantics

Small-step, CBV, left-to-right operational semantics:
- Note: \( \Lambda \alpha. e \) is a value

\[
e \rightarrow e'
\]

Old:
\[
e_1 e_2 \rightarrow e'_1 e'_2
e_2 \rightarrow e'_2\]
\[
v e_2 \rightarrow v' e'_2\]
\[
(\lambda x: \tau. e) v \rightarrow e[v/x]
\]

New:
\[
e[\tau] \rightarrow e'[\tau](\Lambda \alpha. e)[\tau] \rightarrow e[\tau/\alpha]
\]

Plus now have 3 different kinds of substitution, all defined in straightforward capture-avoiding way:
- \( e_1[e_2/x] \) (old)
- \( e[\tau'/\alpha] \) (new)
- \( \tau[\tau'/\alpha] \) (new)

Example

Example (using addition):
\[
(\Lambda \alpha. \Lambda \beta. \lambda x: \alpha. \lambda f: \alpha \rightarrow \beta. f \ x) [\text{int}] [\text{int}] 3 (\lambda y: \text{int}. y + y)
\]
\[
\rightarrow (\Lambda \beta. \lambda x: \text{int}. \lambda f: \text{int} \rightarrow \beta. f \ x) [\text{int}] 3 (\lambda y: \text{int}. y + y)
\]
\[
\rightarrow (\lambda x: \text{int}. \lambda f: \text{int} \rightarrow \text{int} \ f \ x) 3 (\lambda y: \text{int}. y + y)
\]
\[
\rightarrow (\lambda f: \text{int} \rightarrow \text{int}. f \ x) 3 (\lambda y: \text{int}. y + y)
\]
\[
\rightarrow (\lambda f: \text{int} \rightarrow \text{int}. f \ x) 3 (\lambda y: \text{int}. y + y)
\]
\[
\rightarrow (\lambda y: \text{int}. y + y) 3
\]
\[
\rightarrow 3 + 3
\]
\[
\rightarrow 6
\]
Type System, part 1

Mostly just need to be picky about “no free type variables”

- Typing judgment has the form $\Delta; \Gamma \vdash e : \tau$
  - (whole program $\vdash \vdash e : \tau$)
  - Next slide
- Uses helper judgment $\Delta \vdash \tau$
  - “all free type variables in $\tau$ are in $\Delta$”

Types are boring, but trust me, allowing free type variables is a pernicious source of language/compiler bugs

New:

In ML you can’t do the last one; in System F you can

### Examples

An overly simple polymorphic function...

Let id = $\Lambda \alpha. \lambda x : \alpha. x$

- id has type $\forall \alpha. \alpha \rightarrow \alpha$
- id $[\text{int}]$ has type $\text{int} \rightarrow \text{int}$
- id $[\text{int} \times \text{int}]$ has type $(\text{int} \times \text{int}) \rightarrow (\text{int} \times \text{int})$
- (id $[\forall \alpha. \alpha \rightarrow \alpha]$) id has type $\forall \alpha. \alpha \rightarrow \alpha$

In ML you can’t do the last one; in System F you can

### The Whole Language, Called System F

e ::= c | x | \lambda x: \tau. e | ee | \Lambda \alpha. e | e [\tau]

$\tau ::= \text{int} | \tau \rightarrow \tau | \alpha | \forall \alpha. \tau$

$v ::= c | \lambda x: \tau. e | \Lambda \alpha. e$

$\Gamma ::= \cdot | \Gamma, x: \tau$

$\Delta ::= \cdot | \Delta, \alpha$

The typing derivation is rather tall and painful, but just a syntax-directed derivation by instantiating the typing rules

More Examples

Let apply1 = $\Lambda \alpha. \Lambda \beta. \lambda x : \alpha. \lambda f : \alpha \rightarrow \beta. f x$

- apply1 has type $\forall \alpha. \forall \beta. \alpha \rightarrow (\alpha \rightarrow \beta) \rightarrow \beta$
- $\cdot. \text{int} \rightarrow \text{int} \vdash \text{apply1} [\text{int}][\text{int}] 3 (\lambda y : \text{int}. y + y)$

Let apply2 = $\Lambda \alpha. \lambda x : \alpha. \lambda f : \alpha \rightarrow \beta. f x$

- apply2 has type $\forall \alpha. \forall \beta. \alpha \rightarrow (\forall \beta. (\alpha \rightarrow \beta) \rightarrow \beta) \rightarrow \beta$ (impossible in ML)

- $\cdot. \text{int} \rightarrow \text{string}, \text{int} \rightarrow \text{int} \vdash (\text{let } z = \text{apply2} [\text{int}] \text{ in } z (z \text{ [int]} h \text{ [string]} g) : \text{string}$

Let twice = $\Lambda \alpha. \lambda x : \alpha. \lambda f : \alpha \rightarrow \alpha. f (f x)$

- twice has type $\forall \alpha. \alpha \rightarrow (\alpha \rightarrow \alpha) \rightarrow \alpha$

- Cannot be made more polymorphic
What next?

Having defined System F...

- Metatheory (what properties does it have)
- What (else) is it good for
- How/why ML is more restrictive and implicit

Metatheory

- Safety: Language is type-safe
  - Need a Type Substitution Lemma
- Termination: All programs terminate
  - Surprising — we saw id [τ] id
- Parametricity, a.k.a. theorems for free
  - Example: If θ : α, then \(\nu \beta. (\alpha \times \beta) \rightarrow (\beta \times \alpha)\), then \(e\) is equivalent to \(\lambda \alpha. \beta. \lambda x. \alpha \times \beta. (x, 1, 2)\). Every term with this type is the swap function!!
  - Intuition: \(e\) has no way to make an \(\alpha\) or a \(\beta\) and it cannot tell what \(\alpha\) or \(\beta\) are or raise an exception or diverge...

- Erasure: Types do not affect run-time behavior

Note: Mutation “breaks everything”
- depth subtyping: hw4, termination: hw3, parametricity: hw5

Security from safety?

Example: A process \(e\) should not access files it did not open (fopen can check permissions)

Require an untrusted process \(e\) to type-check as follows:

\[
\text{\texttt{\{fopen : string \rightarrow \alpha, fread : \alpha \rightarrow int\} \rightarrow unit}}
\]

This type ensures that a process won’t “forge a file handle” and pass it to fread

So fread doesn’t need to check (faster), file handles don’t need to be encrypted (safer), etc.

Moral of Example

In simply-typed lambda-calculus, type safety just means not getting stuck

With type abstraction, it enables secure interfaces!

Suppose we (the system library) implement file-handles as ints. Then we instantiate \(\alpha\) with \texttt{int}, but untrusted code cannot tell

Memory safety is a necessary but insufficient condition for language-based enforcement of strong abstractions

Are types used at run-time?

We said polymorphism was about “many types for same term”, but for clarity and easy checking, we changed:

- The syntax via \(\Delta\alpha. e\) and \(e[\tau]\)
- The operational semantics via type substitution
- The type system via \(\Delta\)

Claim: The operational semantics did not “really” change; types need not exist at run-time

More formally: Erasing all types from System F produces an equivalent program in the untyped lambda calculus

Strengthened induction hypothesis: If \(e \rightarrow e_1\) in System F and \(\texttt{erase}(e) \rightarrow e_2\) in untyped lambda-calculus, then \(e_2 = \texttt{erase}(e_1)\)

“Erasure and evaluation commute”

Erasure

Erasure is easy to define:

\[
\begin{align*}
\texttt{erase}(c) &= c \\
\texttt{erase}(x) &= x \\
\texttt{erase}(e_1 e_2) &= \texttt{erase}(e_1) \texttt{erase}(e_2) \\
\texttt{erase}(\lambda x. e) &= \lambda x. \texttt{erase}(e) \\
\texttt{erase}(\Delta x. e) &= \lambda x. \texttt{erase}(e) \\
\texttt{erase}(e[\tau]) &= \texttt{erase}(e) 0
\end{align*}
\]

In pure System F, preserving evaluation order isn’t crucial, but it is with fix, exceptions, mutation, etc.
Connection to reality

System F has been one of the most important theoretical PL models since the 1970s and inspires languages like ML.

But you have seen ML polymorphism and it looks different. In fact, it is an implicitly typed restriction of System F.

These two qualifications ((1) implicit, (2) restriction) are deeply related.

Restrictions

- All types have the form $\forall \alpha_1, \ldots, \alpha_n. \tau$ where $n \geq 0$ and $\tau$ has no $\forall$. (Prenex-quantification; no first-class polymorphism.)

- Only let (rec) variables (e.g., $x$ in \texttt{let} $x = e1$ in $e2$) can have polymorphic types. So $n = 0$ for function arguments, pattern variables, etc. (Let-bound polymorphism)
  - So cannot (always) desugar let to $\lambda$ in ML

- In \texttt{let rec} $f\ x = e1$ in $e2$, the variable $f$ can have type $\forall \alpha_1, \ldots, \alpha_n. \tau_1 \rightarrow \tau_2$ only if every use of $f$ in $e1$ instantiates each $\alpha_i$ with $\alpha_i$. (No polymorphic recursion)

- Let variables can be polymorphic only if $e1$ is a “syntactic value”
  - A variable, constant, function definition, ...
  - Called the “value restriction” (relaxed partially in OCaml)

Why?

ML-style polymorphism can seem weird after you have seen System F. And the restrictions do come up in practice, though tolerable.

- Type inference for System F (given untyped $e$, is there a System F term $e'$ such that $\text{erase}(e') = e$) is undecidable (1995)

- Type inference for ML with polymorphic recursion is undecidable (1992)

- Type inference for ML is decidable and efficient in practice, though pathological programs of size $O(n)$ and run-time $O(n)$ can have types of size $O(2^{2^n})$

- The type inference algorithm is unsound in the presence of ML-style mutation, but value-restriction restores soundness
  - Based on unification

Recovering lost ground?

Extensions to the ML type system to be closer to System F:

- Usually require some type annotations

- Are judged by:
  - Soundness: Do programs still not get stuck?
  - Conservatism: Do all (or most) old ML programs still type-check?
  - Power: Does it accept many more useful programs?
  - Convenience: Are many new types still inferred?