

Type Systems

Seminar 6

Mehmet Ali Arslan

Slides mostly “borrowed” from Martin Odersky

Plan

PREVIOUSLY: unit, sequencing, let, pairs, sums

TODAY:

1. recursion
2. state
3. ???

NEXT: exceptions?

NEXT: polymorphic (not so simple) typing

Recursion

Recursion in untyped λ -calculus

- We apply a *fixed-point combinator* to a *generator function*
- $\text{fix} = \lambda f. (\lambda x. f (\lambda y. x x y)) (\lambda x. f (\lambda y. x x y))$

Example - Factorial

- `g = λfct. (λn. if realeq n c0 then c1
else (times n(fct (prd (prd n))));`
- `factorial = fix g;`
- whiteboard...

Recursion in STLC

- the term $\lambda y. x x y$ in *fix* can not be typed in STLC
- so we add it as a primitive

Example

```
ff = λie:Nat→Bool.  
    λx:Nat.  
      if iszero x then true  
      else if iszero (pred x) then false  
      else ie (pred (pred x));  
  
iseven = fix ff;  
  
iseven 7;
```

New syntactic forms

$t ::= \dots$
 $\text{fix } t$

terms

fixed point of t

New evaluation rules

$t \longrightarrow t'$

$$\text{fix } (\lambda x:T_1.t_2) \longrightarrow [x \mapsto (\text{fix } (\lambda x:T_1.t_2))]t_2 \quad (\text{E-FIXBETA})$$

$$\frac{t_1 \longrightarrow t'_1}{\text{fix } t_1 \longrightarrow \text{fix } t'_1} \quad (\text{E-FIX})$$

New typing rules

$\Gamma \vdash t : T$

$$\frac{\Gamma \vdash t_1 : T_1 \rightarrow T_1}{\Gamma \vdash \text{fix } t_1 : T_1}$$

(T-FIX)

A more convenient form

`letrec x:T1=t1 in t2` $\stackrel{\text{def}}{=} \text{let } x = \text{fix } (\lambda x:T_1.t_1) \text{ in } t_2$

```
letrec iseven : Nat → Bool =
  λx:Nat.
    if iszero x then true
    else if iszero (pred x) then false
    else iseven (pred (pred x))
in
  iseven 7;
```

References

Mutability

- ▶ In most programming languages, *variables* are mutable — i.e., a variable provides both
 - ▶ a name that refers to a previously calculated value, and
 - ▶ the possibility of *overwriting* this value with another (which will be referred to by the same name)
- ▶ In some languages (e.g., OCaml), these features are separate:
 - ▶ variables are only for naming — the binding between a variable and its value is immutable
 - ▶ introduce a new class of *mutable values* (called *reference cells* or *references*)
 - ▶ at any given moment, a reference holds a value (and can be *dereferenced* to obtain this value)
 - ▶ a new value may be *assigned* to a reference

We choose OCaml's style, which is easier to work with formally.

So a variable of type `T` in most languages (except OCaml) will correspond to a `Ref T` (actually, a `Ref(Option T)`) here.

Basic Examples

```
r = ref 5
```

```
!r
```

```
r := 7
```

```
(r:=succ(!r); !r)
```

```
(r:=succ(!r); r:=succ(!r); r:=succ(!r);  
r:=succ(!r); !r)
```

Basic Examples

```
r = ref 5
```

```
!r
```

```
r := 7
```

```
(r:=succ(!r); !r)
```

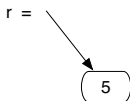
```
(r:=succ(!r); r:=succ(!r); r:=succ(!r);  
r:=succ(!r); !r)
```

i.e.,

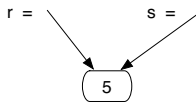
```
((((r:=succ(!r); r:=succ(!r)); r:=succ(!r));  
r:=succ(!r)); !r)
```

Aliasing

A value of type `Ref T` is a *pointer* to a cell holding a value of type `T`.



If this value is “copied” by assigning it to another variable, the cell pointed to is not copied.



So we can change `r` by assigning to `s`:

```
(s:=6; !r)
```

Aliasing all around us

Reference cells are not the only language feature that introduces the possibility of aliasing.

- ▶ object references
- ▶ explicit pointers in C
- ▶ arrays
- ▶ communication channels
- ▶ I/O devices (disks, etc.)

Example

```
c = ref 0
incc = λx:Unit. (c := succ (!c); !c)
decc = λx:Unit. (c := pred (!c); !c)
incc unit
decc unit
o = {i = incc, d = decc}
```

Syntax

`t ::=`

`unit`

`x`

`λx:T.t`

`t t`

`ref t`

`!t`

`t:=t`

terms

unit constant

variable

abstraction

application

reference creation

dereference

assignment

... plus other familiar types, in examples.

Typing Rules

$$\frac{\Gamma \vdash t_1 : T_1}{\Gamma \vdash \text{ref } t_1 : \text{Ref } T_1} \quad (\text{T-REF})$$

$$\frac{\Gamma \vdash t_1 : \text{Ref } T_1}{\Gamma \vdash !t_1 : T_1} \quad (\text{T-DEREF})$$

$$\frac{\Gamma \vdash t_1 : \text{Ref } T_1 \quad \Gamma \vdash t_2 : T_1}{\Gamma \vdash t_1 := t_2 : \text{Unit}} \quad (\text{T-ASSIGN})$$

Final example

```
NatArray = Ref (Nat→Nat);
```

```
newarray = λ_:Unit. ref (λn:Nat.0);  
          : Unit → NatArray
```

```
lookup = λa:NatArray. λn:Nat. (!a) n;  
         : NatArray → Nat → Nat
```

```
update = λa:NatArray. λm:Nat. λv:Nat.  
         let oldf = !a in  
         a := (λn:Nat. if equal m n then v else oldf n);  
         : NatArray → Nat → Nat → Unit
```

Evaluation

What is the *value* of the expression `ref 0`?

Evaluation

What is the *value* of the expression `ref 0`?

Crucial observation: evaluating `ref 0` must *do* something.

Otherwise,

```
r = ref 0
s = ref 0
```

and

```
r = ref 0
s = r
```

would behave the same.

Evaluation

What is the *value* of the expression `ref 0`?

Crucial observation: evaluating `ref 0` must *do* something.

Otherwise,

```
r = ref 0
s = ref 0
```

and

```
r = ref 0
s = r
```

would behave the same.

Specifically, evaluating `ref 0` should *allocate some storage* and yield a *reference* (or *pointer*) to that storage.

Evaluation

What is the *value* of the expression `ref 0`?

Crucial observation: evaluating `ref 0` must *do* something.

Otherwise,

```
r = ref 0
s = ref 0
```

and

```
r = ref 0
s = r
```

would behave the same.

Specifically, evaluating `ref 0` should *allocate some storage* and yield a *reference* (or *pointer*) to that storage.

So what is a reference?

The Store

A reference names a *location* in the *store* (also known as the *heap* or just the *memory*).

What is the store?

The Store

A reference names a *location* in the *store* (also known as the *heap* or just the *memory*).

What is the store?

- ▶ *Concretely*: An array of 8-bit bytes, indexed by 32-bit integers.

The Store

A reference names a *location* in the *store* (also known as the *heap* or just the *memory*).

What is the store?

- ▶ *Concretely*: An array of 8-bit bytes, indexed by 32-bit integers.
- ▶ *More abstractly*: an array of *values*

The Store

A reference names a *location* in the *store* (also known as the *heap* or just the *memory*).

What is the store?

- ▶ *Concretely*: An array of 8-bit bytes, indexed by 32-bit integers.
- ▶ *More abstractly*: an array of *values*
- ▶ *Even more abstractly*: a partial function from *locations* to *values*.

Locations

Syntax of values:

$v ::=$

`unit`

`$\lambda x:T.t$`

`/`

values

unit constant

abstraction value

store location

... and since all values are terms...

Syntax of Terms

$t ::=$

`unit`

`x`

`$\lambda x:T.t$`

`t t`

`ref t`

`!t`

`t:=t`

`/`

terms

unit constant

variable

abstraction

application

reference creation

dereference

assignment

store location

Aside

Does this mean we are going to allow programmers to write explicit locations in their programs??

No: This is just a modeling trick. We are enriching the “source language” to include some run-time structures, so that we can continue to formalize evaluation as a relation between source terms.

Aside: If we formalize evaluation in the big-step style, then we can add locations to the set of values (results of evaluation) without adding them to the set of terms.

Evaluation

The result of evaluating a term now depends on the store in which it is evaluated. Moreover, the result of evaluating a term is not just a value — we must also keep track of the changes that get made to the store.

I.e., the evaluation relation should now map a term and a store to a reduced term and a new store.

$$t \mid \mu \longrightarrow t' \mid \mu'$$

We use the metavariable μ to range over stores.

Evaluation

An assignment $t_1 := t_2$ first evaluates t_1 and t_2 until they become values...

$$\frac{t_1 \mid \mu \longrightarrow t'_1 \mid \mu'}{t_1 := t_2 \mid \mu \longrightarrow t'_1 := t_2 \mid \mu'} \quad (\text{E-ASSIGN1})$$

$$\frac{t_2 \mid \mu \longrightarrow t'_2 \mid \mu'}{v_1 := t_2 \mid \mu \longrightarrow v_1 := t'_2 \mid \mu'} \quad (\text{E-ASSIGN2})$$

... and then returns `unit` and updates the store:

$$l := v_2 \mid \mu \longrightarrow \text{unit} \mid [l \mapsto v_2]\mu \quad (\text{E-ASSIGN})$$

A term of the form `ref t1` first evaluates inside `t1` until it becomes a value...

$$\frac{t_1 \mid \mu \longrightarrow t'_1 \mid \mu'}{\text{ref } t_1 \mid \mu \longrightarrow \text{ref } t'_1 \mid \mu'} \quad (\text{E-REF})$$

... and then chooses (allocates) a fresh location `l`, augments the store with a binding from `l` to `v1`, and returns `l`:

$$\frac{l \notin \text{dom}(\mu)}{\text{ref } v_1 \mid \mu \longrightarrow l \mid (\mu, l \mapsto v_1)} \quad (\text{E-REFV})$$

A term $!t_1$ first evaluates in t_1 until it becomes a value...

$$\frac{t_1 \mid \mu \longrightarrow t'_1 \mid \mu'}{!t_1 \mid \mu \longrightarrow !t'_1 \mid \mu'} \quad (\text{E-DEREF})$$

... and then looks up this value (which must be a location, if the original term was well typed) and returns its contents in the current store:

$$\frac{\mu(l) = v}{!l \mid \mu \longrightarrow v \mid \mu} \quad (\text{E-DEREFLOC})$$

Evaluation rules for function abstraction and application are augmented with stores, but don't do anything with them directly.

$$\frac{t_1 \mid \mu \longrightarrow t'_1 \mid \mu'}{t_1 \ t_2 \mid \mu \longrightarrow t'_1 \ t_2 \mid \mu'} \quad (\text{E-APP1})$$

$$\frac{t_2 \mid \mu \longrightarrow t'_2 \mid \mu'}{v_1 \ t_2 \mid \mu \longrightarrow v_1 \ t'_2 \mid \mu'} \quad (\text{E-APP2})$$

$$(\lambda x:T_{11}.t_{12}) \ v_2 \mid \mu \longrightarrow [x \mapsto v_2]t_{12} \mid \mu \quad (\text{E-APPABS})$$

Aside: garbage collection

Note that we are not modeling garbage collection — the store just grows without bound.

Aside: pointer arithmetic

We can't do any!

Store Typings

Typing Locations

Q: What is the *type* of a *location*?

Typing Locations

Q: What is the *type* of a *location*?

A: It depends on the store!

E.g., in the store $(l_1 \mapsto \text{unit}, l_2 \mapsto \text{unit})$, the term $!l_2$ has type `Unit`.

But in the store $(l_1 \mapsto \text{unit}, l_2 \mapsto \lambda x:\text{Unit}.x)$, the term $!l_2$ has type `Unit` \rightarrow `Unit`.

Typing Locations — first try

Roughly:

$$\frac{\Gamma \vdash \mu(l) : T_1}{\Gamma \vdash l : \text{Ref } T_1}$$

Typing Locations — first try

Roughly:

$$\frac{\Gamma \vdash \mu(l) : T_1}{\Gamma \vdash l : \text{Ref } T_1}$$

More precisely:

$$\frac{\Gamma \mid \mu \vdash \mu(l) : T_1}{\Gamma \mid \mu \vdash l : \text{Ref } T_1}$$

I.e., typing is now a *four*-place relation (between contexts, *stores*, terms, and types).

Problem

However, this rule is not completely satisfactory. For one thing, it can make typing derivations very large!

E.g., if

$$\begin{aligned}(\mu = & l_1 \mapsto \lambda x:\text{Nat}. 999, \\ & l_2 \mapsto \lambda x:\text{Nat}. !l_1 (!l_1 x), \\ & l_3 \mapsto \lambda x:\text{Nat}. !l_2 (!l_2 x), \\ & l_4 \mapsto \lambda x:\text{Nat}. !l_3 (!l_3 x), \\ & l_5 \mapsto \lambda x:\text{Nat}. !l_4 (!l_4 x)),\end{aligned}$$

then how big is the typing derivation for $!l_5$?

Problem!

But wait... it gets worse. Suppose

$$\begin{aligned} (\mu = l_1 \mapsto \lambda x:\text{Nat}. !l_2 x, \\ l_2 \mapsto \lambda x:\text{Nat}. !l_1 x), \end{aligned}$$

Now how big is the typing derivation for $!l_2$?

Store Typings

Observation: The typing rules we have chosen for references guarantee that a given location in the store is *always* used to hold values of the *same* type.

These intended types can be collected into a *store typing* — a partial function from locations to types.

E.g., for

$$\begin{aligned}\mu = (&l_1 \mapsto \lambda x:\text{Nat}. 999, \\ &l_2 \mapsto \lambda x:\text{Nat}. !l_1 (!l_1 x), \\ &l_3 \mapsto \lambda x:\text{Nat}. !l_2 (!l_2 x), \\ &l_4 \mapsto \lambda x:\text{Nat}. !l_3 (!l_3 x), \\ &l_5 \mapsto \lambda x:\text{Nat}. !l_4 (!l_4 x)),\end{aligned}$$

A reasonable store typing would be

$$\begin{aligned}\Sigma = (&l_1 \mapsto \text{Nat} \rightarrow \text{Nat}, \\ &l_2 \mapsto \text{Nat} \rightarrow \text{Nat}, \\ &l_3 \mapsto \text{Nat} \rightarrow \text{Nat}, \\ &l_4 \mapsto \text{Nat} \rightarrow \text{Nat}, \\ &l_5 \mapsto \text{Nat} \rightarrow \text{Nat})\end{aligned}$$

Now, suppose we are given a store typing Σ describing the store μ in which we intend to evaluate some term t . Then we can use Σ to look up the types of locations in t instead of calculating them from the values in μ .

$$\frac{\Sigma(l) = T_1}{\Gamma \mid \Sigma \vdash l : \text{Ref } T_1} \quad (\text{T-Loc})$$

I.e., typing is now a four-place relation between between contexts, *store typings*, terms, and types.

Final typing rules

$$\frac{\Sigma(l) = T_1}{\Gamma \mid \Sigma \vdash l : \text{Ref } T_1} \quad (\text{T-LOC})$$

$$\frac{\Gamma \mid \Sigma \vdash t_1 : T_1}{\Gamma \mid \Sigma \vdash \text{ref } t_1 : \text{Ref } T_1} \quad (\text{T-REF})$$

$$\frac{\Gamma \mid \Sigma \vdash t_1 : \text{Ref } T_{11}}{\Gamma \mid \Sigma \vdash !t_1 : T_{11}} \quad (\text{T-DEREF})$$

$$\frac{\Gamma \mid \Sigma \vdash t_1 : \text{Ref } T_{11} \quad \Gamma \mid \Sigma \vdash t_2 : T_{11}}{\Gamma \mid \Sigma \vdash t_1 := t_2 : \text{Unit}} \quad (\text{T-ASSIGN})$$

Q: Where do these store typings come from?

Q: Where do these store typings come from?

A: When we first typecheck a program, there will be no explicit locations, so we can use an empty store typing.

So, when a new location is created during evaluation,

$$\frac{l \notin \text{dom}(\mu)}{\text{ref } v_1 \mid \mu \longrightarrow l \mid (\mu, l \mapsto v_1)} \quad (\text{E-REFV})$$

we can extend the “current store typing” with the type of v_1 .

Safety

Preservation

First attempt: just add stores and store typings in the appropriate places.

Theorem (?): If $\Gamma \mid \Sigma \vdash t : T$ and $t \mid \mu \longrightarrow t' \mid \mu'$, then $\Gamma \mid \Sigma \vdash t' : T$.

Preservation

First attempt: just add stores and store typings in the appropriate places.

Theorem (?): If $\Gamma \mid \Sigma \vdash t : T$ and $t \mid \mu \longrightarrow t' \mid \mu'$, then $\Gamma \mid \Sigma \vdash t' : T$. **Wrong!**

Why is this wrong?

Preservation

First attempt: just add stores and store typings in the appropriate places.

Theorem (?): If $\Gamma \mid \Sigma \vdash t : T$ and $t \mid \mu \longrightarrow t' \mid \mu'$, then $\Gamma \mid \Sigma \vdash t' : T$. **Wrong!**

Why is this wrong?

Because Σ and μ here are not constrained to have anything to do with each other!

(Exercise: Construct an example that breaks this statement of preservation.)

Preservation

A store μ is said to be *well typed* with respect to a typing context Γ and a store typing Σ , written $\Gamma \mid \Sigma \vdash \mu$, if $\text{dom}(\mu) = \text{dom}(\Sigma)$ and $\Gamma \mid \Sigma \vdash \mu(l) : \Sigma(l)$ for every $l \in \text{dom}(\mu)$.

Preservation

A store μ is said to be *well typed* with respect to a typing context Γ and a store typing Σ , written $\Gamma \mid \Sigma \vdash \mu$, if $\text{dom}(\mu) = \text{dom}(\Sigma)$ and $\Gamma \mid \Sigma \vdash \mu(l) : \Sigma(l)$ for every $l \in \text{dom}(\mu)$.

Next attempt:

Theorem (?): If

$$\Gamma \mid \Sigma \vdash t : T$$

$$t \mid \mu \longrightarrow t' \mid \mu'$$

$$\Gamma \mid \Sigma \vdash \mu$$

then $\Gamma \mid \Sigma \vdash t' : T$.

Preservation

A store μ is said to be *well typed* with respect to a typing context Γ and a store typing Σ , written $\Gamma \mid \Sigma \vdash \mu$, if $\text{dom}(\mu) = \text{dom}(\Sigma)$ and $\Gamma \mid \Sigma \vdash \mu(l) : \Sigma(l)$ for every $l \in \text{dom}(\mu)$.

Next attempt:

Theorem (?): If

$$\Gamma \mid \Sigma \vdash t : T$$

$$t \mid \mu \longrightarrow t' \mid \mu'$$

$$\Gamma \mid \Sigma \vdash \mu$$

then $\Gamma \mid \Sigma \vdash t' : T$.

Still wrong!

What's wrong now?

Preservation

A store μ is said to be *well typed* with respect to a typing context Γ and a store typing Σ , written $\Gamma \mid \Sigma \vdash \mu$, if $\text{dom}(\mu) = \text{dom}(\Sigma)$ and $\Gamma \mid \Sigma \vdash \mu(l) : \Sigma(l)$ for every $l \in \text{dom}(\mu)$.

Next attempt:

Theorem (?): If

$$\Gamma \mid \Sigma \vdash t : T$$

$$t \mid \mu \longrightarrow t' \mid \mu'$$

$$\Gamma \mid \Sigma \vdash \mu$$

then $\Gamma \mid \Sigma \vdash t' : T$.

Still wrong!

Creation of a new reference cell...

$$\frac{l \notin \text{dom}(\mu)}{\text{ref } v_1 \mid \mu \longrightarrow l \mid (\mu, l \mapsto v_1)} \quad (\text{E-REFV})$$

... breaks the correspondence between the store typing and the store.

Preservation (correct version)

Theorem: If

$$\Gamma \mid \Sigma \vdash t : T$$

$$\Gamma \mid \Sigma \vdash \mu$$

$$t \mid \mu \longrightarrow t' \mid \mu'$$

then, for **some** $\Sigma' \supseteq \Sigma$,

$$\Gamma \mid \Sigma' \vdash t' : T$$

$$\Gamma \mid \Sigma' \vdash \mu'.$$

Preservation (correct version)

Theorem: If

$$\Gamma \mid \Sigma \vdash t : T$$

$$\Gamma \mid \Sigma \vdash \mu$$

$$t \mid \mu \longrightarrow t' \mid \mu'$$

then, for **some** $\Sigma' \supseteq \Sigma$,

$$\Gamma \mid \Sigma' \vdash t' : T$$

$$\Gamma \mid \Sigma' \vdash \mu'.$$

Proof: Easy extension of the preservation proof for λ_{\rightarrow} .

Progress

Theorem: Suppose t is a closed, well-typed term (that is, $\emptyset \mid \Sigma \vdash t : \mathbb{T}$ for some \mathbb{T} and Σ). Then either t is a value or else, for any store μ such that $\emptyset \mid \Sigma \vdash \mu$, there is some term t' and store μ' with $t \mid \mu \longrightarrow t' \mid \mu'$.