Announcements

• Reminder: My office hours: Fridays, 10:30am - noon, CoRE 318 or on the hallway next to CoRE 301

• Solutions for first homework problem sets have been posted.

• Anybody who wants this course is now registered?
**Lambda calculus**

- formalism for studying ways in which functions can be formed, combined, and used for computation

- **computation** is defined as rewriting rules (operational semantics)

- the syntactic notion of computation was developed first; a mathematical semantics followed much later

Examples:

\[
\begin{align*}
f(x) &= x + 2 \\
(\lambda x. (+ x 2)) 1 &= 1 + 2 = 3 \\
(\lambda x.\ (\lambda y. y)) &= \text{arguments and returned “values” can be functions} \\
(\lambda x. xx) &= \text{untyped lambda calculus} \\
f(x) &= x(x)
\end{align*}
\]
Lambda calculus and functional programming

Lambda calculus is the theoretical foundation of pure functional programming (no side effects, referential transparency)

Functional programming: functions are first class citizens

• can be a return value
• can be passed as arguments
• can be put into a data structure
• value of an expression can be a function

(((\lambda x.x) (\lambda x.1)) (\lambda y.y))
Lambda calculus

**λ-terms** (wffs) are inductively defined.

A λ-terms is:

- a variable \( x \)
- \( (\lambda x. M) \) where \( x \) is a variable and \( M \) is λ-term (abstraction)
- \( (M N) \) where \( M \) and \( N \) are λ-terms (application)

**Abbreviations** (Notational conveniences):

- function application is left associative
  \[(f g z) \text{ is } ((f g) z)\]
- function application has precedence over function abstraction — “function body” extends as far to the right as possible
  \[\lambda x.yz \text{ is } (\lambda x.(yz))\]
- “multiple” arguments
  \[\lambda xy.z \text{ is } (\lambda x.(\lambda y.z))\]
Currying functions

\[ f : D^n \rightarrow D \quad \text{can be transformed to} \]
\[ f : D \rightarrow (D \rightarrow \cdots (D \rightarrow D)) \cdots \]

Examples:

\[ f(x, y) = x + y \quad f : \mathbb{N} \times \mathbb{N} \rightarrow \mathbb{N} \]
\[ f'(x) = g_x(y) \quad f' : \mathbb{N} \rightarrow (\mathbb{N} \rightarrow \mathbb{N}) \]
\[ g_x : \mathbb{N} \rightarrow \mathbb{N} \]
\[ g_x(y) = x + y \]
\[ g_2 : \mathbb{N} \rightarrow \mathbb{N} \]
\[ g_2(y) = 2 + y \]

\(g_x(y)\) “freezes” the first argument value \(x\).
This is sometimes referred to as partial evaluation.

Here is an example:
\[ (\lambda x.\lambda y.x+y) \ 2 \ 3 \ = \]
\[ (\lambda y.2+y) \ 3 \ = \]
\[ 2+3 \ = \ 5 \]

To simplify discussion, all \(n\)-ary functions are curried, i.e., are represented by a sequence of one-place functions.
Examples

What are the $\lambda$-terms represented by

$\lambda xyz. x(yz)t$ is

$(\lambda x. (\lambda y. (\lambda z. ((x(yz))t))))$ 

$(\lambda x. (\lambda y. yx)x)$ is

$(\lambda x. ((\lambda y. yx)x))$
Free and bound variables

Abstraction ($\lambda x. M$) “binds” variable $x$ in “body” $M$. You can think of this as a declaration of variable $x$ with scope $M$.

Let $M, N$ be $\lambda$-terms and $x$ is a variable. The set of free variables of $M$, free($M$), is defined inductively as follows:

- $\text{free}(x) = \{x\}$
- $\text{free}(M N) = \text{free}(M) \cup \text{free}(N)$
- $\text{free}(\lambda x. M) = \text{free}(M) - \{x\}$
Free and bound variables

Note:

- a variable can occur free and bound in a λ-term.

  See example above

\[
\lambda x. \lambda y. (\lambda z. xyz) y
\]

  "free" is relative to a λ-subterm

- formal definition of bound: exercise
Function application as substitution

The result of applying an abstraction \((\lambda x. M)\) to an argument \(N\) is formalized by a special form of textual substitution.

\[(\lambda x. M)N \equiv [N/x]M\]

Informally: \(N\) replaces all free occurrences of \(x\) in \(M\).

What can go wrong?

Example: Assume we have constants and arithmetic operation “+” in our lambda calculus

\[(\lambda a. \lambda b. a+b)2 \ x \ \equiv \ (\lambda b.2+b)x \ \equiv \ 2+x\]

What about:

\[(\lambda a. \lambda b. a+b)b \ 3 \ \equiv \ (\lambda b.b+b)3 \ \equiv \ 3+3 \ \equiv \ 6\]
Function application as substitution

We need capture free substitution

(1) If the free variables of N have no bound occurrence in M, then \([N/x]M\) is formed by replacing all free occurrences of x in M by N.

Example: \((\lambda b.2+b) \ x \cong [x/b]2+b \cong 2+x\)

(2) Otherwise, if variable y is free in N and bound in M, replace the binding and bound occurrences of y in M by new (fresh) variable z. Repeat until case (1) applies.

\((\lambda a.\lambda b.a+b) \ b \cong \lambda a.\lambda z.a+z) \ b \cong (\lambda z.b+z)\)
Function application as substitution

Examples:

\[
[u/x] x \equiv u \quad \text{u not bound in } x
\]
\[
[u/x]\lambda x.xu \equiv \lambda x.xu \quad \text{x not free in } \lambda x.xu
\]
\[
[u/x]\lambda u.x \equiv [u/x]\lambda z.x \quad \text{u is bound in } \lambda u.x
\]
\[
[u/x]\lambda u.u \equiv [u/x]\lambda z.z
\]
\[
\equiv \lambda z.z
\]
**Capture–free substitution**

Let $x, y, z$ denote variables and $M, N, P$ arbitrary $\lambda$-terms.

\[
\begin{align*}
[N/y] x &= x & \text{if } y \neq x \\
        &= N & \text{if } y = x
\end{align*}
\]

\[
\begin{align*}
[N/x](\lambda x. M) &= (\lambda x. M) \\
[N/x](\lambda y. M) &= \lambda y.[N/x]M & \text{where } x \neq y, \\
                    &= \lambda z.[N/x][z/y]M & y \notin \text{free}(N), \\
                    &= \lambda z.[N/x][z/y]M & y \in \text{free}(N), \\
                    &= \lambda z.[N/x][z/y]M & z \notin \text{free}(M) \\
                    &= \lambda z.[N/x][z/y]M & z \notin \text{free}(N)
\end{align*}
\]

\[
[N/x](M P) = ([N/x]M [N/x]P)
\]
Substitution — Examples

\[
[y/x] ((\lambda z.zx)(\lambda x.x)) =
((y/x) (\lambda z.zx) (y/x)(\lambda x.x)) =
(\lambda z.zy)(\lambda x.x)
\]

\[
(\lambda x.xy/x)((\lambda y.xy)z) =
(((\lambda x.xy/x)(\lambda y.xy)[\lambda x.xy/x]z) =
((\lambda z.[\lambda x.xy/x][z/y]xy)(\lambda z.[\lambda x.xy/x]xz)z) =
((\lambda z.(\lambda x.xy)z)z)
\]
Function application

Computation in the lambda calculus is based on the concept or reduction (rewriting rules). The goal is to “simplify” an expression until it can no longer be further simplified.

\[(\lambda x. M)N \Rightarrow_\beta [N/x]M \quad (\beta\text{-reduction})\]

\[(\lambda x. M) \Rightarrow_\alpha \lambda y.[y/x]M \quad (\alpha\text{-reduction})\]

if \(y \notin free(M)\)

\[(\lambda x. M)x \Rightarrow_\eta M \quad (\eta\text{-reduction})\]

if \(x \notin free(M)\)

Note:

- An equivalence relation can be defined based on \(\equiv\)-convertable \(\lambda\)-terms. “Reduction” rules really work both ways, but we are interested in reducing the complexity of \(\lambda\)-term (\(\rightarrow\) direction).

- \(\alpha\)-reduction does not reduce the complexity.

- \(\beta\)-reduction: corresponds to application, models computation.

- \(\eta\)-reduction: one can remove (or introduce) levels of indirections in function applications.
Reduction

• A subterm of the form \(((\lambda x.M)N)\) is called a **redex** (reduction expression).

• A reduction is any sequence of \(\beta\)–reductions and \(\alpha\)–reductions.

• A term that cannot be \(\beta\)–reduced is said to be in \(\beta\)–normal form (**normal form**).

• A subterm that is an abstraction or a variable is said to be in **head normal form**.

Does a normal form always exist?

Examples:

\((\lambda x.xx)(\lambda x.xx)\)

\((\lambda x.xxx)(\lambda x.xxx)\)
Reduction — example

\((\lambda xyz. (xz)(yz))(\lambda xy.x) (\lambda xy.x)\)

\((\lambda yz. ((\lambda xy.x)z)(yz))(\lambda xy.x)\)

\((\lambda yz. (\lambda y.z)(yz))(\lambda xy.x)\)

\[ \lambda z.((\lambda xy.x)z) ((\lambda xy.x)z) \]

\[ \lambda z.(\lambda y.z)((\lambda xy.x)z) \]

\[ (\lambda yz.z)(\lambda xy.x) \]

\[ \lambda z.((\lambda xy.x)z)(\lambda y.z) \]

\[ \lambda z.(\lambda y.z)(\lambda y.z) \]

\[ \lambda z.z \]
**Church–Rosser Property**

Informal: If an expression reduces to two normal forms, they must be identical modulo \(\alpha\)-conversion.

**Theorem (CR–I)**

Let \(M, P, Q,\) and \(R\) be \(\lambda\)-terms. If \(M \Rightarrow^* P\) and \(M \Rightarrow^* Q\), then

\[\exists R \text{ such that } P \Rightarrow^* R \text{ and } Q \Rightarrow^* R\]

**Corollary**: If a normal form exists for \(M\), it must be unique. Proof is homework!
Reduction strategies

If P is not in normal form, what redex to choose?

Does it make a difference (remember CR–I)?

Yes, since there are infinite computations possible

\[(\lambda x.y)((\lambda x.xx)(\lambda x.xx))\]

Note:

- It is undecidable whether a \(\lambda\)-term has a normal form, i.e., whether it will reduce to a normal form.
- You can think of the normal form as the “value” to which a \(\lambda\)-term evaluates
- Two \(\lambda\)-terms are equal if they have the same normal form.
Reduction strategies & CR–II

• If A and B are two redexes in a \( \lambda \)-term M and the first occurrence of \( \lambda \) in A is to the left of the first occurrence of \( \lambda \) in B, then A is to the left of B.

• If A is a redex in M, and it is to the left of all other redexes, then A is the leftmost redex of M.

Theorem (CR–II)
Let X and Y be \( \lambda \)-terms.
If \( X \Rightarrow^* Y \) and Y is in normal form, then \( X \Rightarrow^* Y \) using only leftmost redexes.
Choosing always the left-most redex is called **normal order**.

- **normal order** specifies the redex that is chosen next for a given \( \lambda \)-term. **c-b-n** and **c-b-v** specify what to do with a redex of the form \((\lambda x.M)N\).
  - **c-b-n**: use \( \beta \)-reduction (don’t “touch” \( N \)).
  - **c-b-v**: need a notion of value. Only after “\( N \)” is reduced to a value, the \( \beta \)-reduction is performed. What’s a value? Let’s pick head-normal form.

Question: Reduce the following \( \lambda \)-term using **c-b-n** and **c-b-v**

\[
(\lambda z.y)(\lambda y.(\lambda x.xx)(\lambda x.xx))
\]

Scott refers to **c-b-v** as **applicative-order evaluation**, **c-b-n** is called **normal-order evaluation** of a redex. **lazy evaluation** is a specific combination of both approaches, where arguments are evaluated at most once (memoization).
call–by–name vs. call–by–value

left-most redex, call–by–name:

- guaranteed to reach a normal form if it exists.
- can result in some parameter being evaluated several times or never.

left-most redex, call–by–value:

- efficient — assuming parameter is used at least once.
- can lead to nonterminating computation.

Functional languages typically use call–by–value because of its efficiency.
Programming in lambda calculus

The lambda calculus has very few constructs and it is therefore easy to reason about it.

Question: Is the lambda calculus too simple, i.e., can we express all computable functions in the lambda calculus?

Remember: Computation in the lambda calculus is a sequence of applications of reduction rules (mostly $\beta$–reductions).

Logical constants and operations:

true $\equiv \lambda a.\lambda b.a$

false $\equiv \lambda a.\lambda b.b$

cond $\equiv \lambda m.\lambda n.\lambda p.((p\ m)\ n)$

not $\equiv \lambda x.((x\ false)\ true)$

and $\equiv \lambda x.\lambda y.((x\ y)\ false)$

or $\equiv \lambda x.\lambda y.((x\ true)\ y)$
What about data structures?

**data structures:**

**pairs** can be represented as

\[ [M . N] \equiv \lambda z.((z M) N) \]

**first** \(\equiv \lambda x.(x \text{ true})\) \hspace{1cm} (car)

**second** \(\equiv \lambda x.(x \text{ false})\) \hspace{1cm} (cdr)

**build** \(\equiv \lambda x.\lambda y.\lambda z.((z x) y)\) \hspace{1cm} (cons)

What do we need to represent lists?

- *empty list*
- *isEmpty? function*
Programming in lambda calculus

What about arithmetic constants and operations?

There are many options here. Let’s look at the system proposed by Church:

\[
0 \equiv \lambda fx.x \\
1 \equiv \lambda fx.(f x) \\
2 \equiv \lambda fx.(f (f x)) \\
\ldots \\
n \equiv \lambda fx.(f(f(\ldots(f x)\ldots)) \equiv \lambda fx.(f^n x)
\]

The natural number \( n \) is represented as a function that applies a function \( f \) \( n \)–times to its argument \( x \).

\[
\text{succ} \equiv \lambda m.(\lambda fx.(f (m f x))) \\
\text{add} \equiv \lambda mn.(\lambda fx.((m f) (n f x))) \\
\text{mult} \equiv \lambda mn.(\lambda fx.((m (n f)) x)) \\
\text{isZero?} \equiv \lambda m.((m (\text{true false})) \text{ true})
\]

Note: \text{pred} function can also be encoded, but the encoding is rather complicated
Programming in lambda calculus

Examples:

\[(\text{mult} \ 2 \ 3) =\]
\[((\lambda mn. (\text{fx}. ((m \ (n \ f)) \ x))) \ 2 \ 3) =\]
\[\lambda f_0 x_0. ((2 \ 3 f_0) \ x_0) =\]
\[\lambda f_0 x_0. ((2 ((\text{fx}. (f \ (f \ f))) \ f_0)) \ x_0) =\]
\[\lambda f_0 x_0. ((2 (\lambda x. (f_0 \ (f_0 \ (f_0 \ x)))))) \ x_0) =\]
\[\lambda f_0 x_0. (((2 \ \lambda x_1. (f_0^3 x_1)) \ x_0) =\]
\[\lambda f_0 x_0. (\lambda x. (((\lambda x_1. (f_0^3 x_1)) \ (\lambda x_1. (f_0^3 x_1)) \ x_0)) \ x_0) =\]
\[\lambda f_0 x_0. (\lambda x. ((\lambda x_1. (f_0^3 x_1)) \ (f_0^3 x) \ x_0)) =\]
\[\lambda f_0 x_0. (((\lambda x. (f_0^3 (f_0^3 x))) \ x_0) =\]
\[\lambda f_0 x_0. (f_0^3 (f_0^3 x)) =\]
\[\lambda f x. (f^6 x) = 6\]

\[(\text{isZero?} \ 0) =\]
\[((\lambda f x. x) (\lambda y. \text{false})) \ \text{true}) =\]
\[((\lambda x. x) \ \text{true}) = \text{true}\]

\[(\text{isZero?} \ n) \quad \text{where} \ n > 0 ?\]
Recursion in lambda calculus

Does this make sense?

\[ f \equiv \ldots f \ldots \]

In lambda calculus, such an equation does not define a term. How to find a \( \lambda \)-term that does “satisfy” the recursive definition?

Example:

\[ \text{add} \equiv \lambda mn. \\
(\text{cond } m (\text{add} (\text{succ } m) (\text{pred } n)) (\text{isZero? } n)) \]

Just to make things easier to read, we will write instead:

\[ \text{add} \equiv \lambda mn. \\
\text{if } (\text{isZero? } n) \text{ then } m \text{ else } (\text{add} (\text{succ } m) (\text{pred } n)) \]

This is not a valid definition of a \( \lambda \)-term. What about this one?

\[ \text{add} \equiv \lambda f. (\lambda mn. \\
\text{if } (\text{isZero? } n) \text{ then } m \text{ else } (f (\text{succ } m) (\text{pred } n))) \]

**Claim**: The fixed point of the above function is what we are looking for.
Function fixed points

The fixed points of a function $g$ is the set of values $fix_g = \{x \mid x = g(x)\}$.

Examples:

<table>
<thead>
<tr>
<th>function $g$</th>
<th>$fix_g$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda x.6$</td>
<td>${6}$</td>
</tr>
<tr>
<td>$\lambda x.(6 - x)$</td>
<td>${3}$</td>
</tr>
<tr>
<td>$\lambda x.((x*x) + (x-4))$</td>
<td>${-2, 2}$</td>
</tr>
<tr>
<td>$\lambda x.x$</td>
<td>entire domain of f</td>
</tr>
<tr>
<td>$\lambda x.(x+1)$</td>
<td>${}$</td>
</tr>
</tbody>
</table>

Is there a $\lambda$–term $Y$ that “computes” a fixed point of a function $F = \lambda f.(\ldots f \ldots)$, i.e., $YF = F(YF)$?

YES. $Y$ is called the **fixed point combinator**.

$$ Y \equiv \lambda f.((\lambda x.f(x x)) (\lambda x.f(x x))) $$

$$ YF = ((\lambda f.((\lambda x.f(x x)) (\lambda x.f(x x)))) F) $$

$$ = (\lambda x.F(x x)) (\lambda x.F(x x)) $$

$$ = F( (\lambda x.F(x x)) (\lambda x.F(x x))) $$

$$ = F(YF) $$
The Y–combinator

Example:

\[ F \equiv \lambda f.(\lambda mn. \text{if (isZero? n) then m else (f (succ m) (pred n)))} \]

\[
((YF) 3 2) = \\
(((\lambda f.((\lambda x.f(x x)) (\lambda x.f(x x)))) F) 3 2) = \\
((\lambda mn.if (isZero? n) then m else ((\lambda x.F(x x)) (\lambda x.F(x x)))) 3 2) = \\
\text{if (isZero? 2) then 3 else} \\
(((\lambda x.F(x x)) (\lambda x.F(x x))) (\text{succ m} \ (\text{pred n}))) 3 2) = \\
\text{if (isZero? 2) then 3 else} \\
(((\lambda x.F(x x)) (\lambda x.F(x x))) (\text{succ 3} \ (\text{pred 2}))) = \\
(((\lambda x.F(x x)) (\lambda x.F(x x))) 4 1) = \\
((\lambda x.F(x x)) (\lambda x.F(x x))) 4 1) = \\
\text{if (isZero? 1) then 4 else} \\
(((\lambda x.F(x x)) (\lambda x.F(x x))) (\text{succ 4} \ (\text{pred 1}))) = \\
(((\lambda x.F(x x)) (\lambda x.F(x x))) 5 0) = \\
((\lambda x.F(x x)) (\lambda x.F(x x))) 5 0) = \\
\text{if (isZero? 0) then 5 else} \\
(((\lambda x.F(x x)) (\lambda x.F(x x))) (\text{succ 5} \ (\text{pred 0}))) = 5 \]
The Y–combinator example (cont.)

Note:

• Informally, the Y–combinator allows us to get as many copies of the recursive procedure body as we need. The computation “unrolls” recursive procedure calls one at a time.

• This notion of recursion is purely syntactic.

• Question: Would our Y–combinator work with call–by–value application order?
Call–by–value lambda calculus interpreter

Procedure $eval(P)$:
\[
\text{case } P:\n\begin{align*}
\text{"x" : } & \quad x \\
\text{"(λ x.M)" : } & \quad λ \ x.M \\
\text{"(M N)" : } & \quad e_1 = eval(M); \\
& \quad \text{error if not } e_1 \equiv λx.M_1; \\
& \quad e_2 = eval(N); \\
& \quad e_3 = [e_2/x]M_1; //capture–free subst. \\
& \quad eval(e_3)
\end{align*}
\]
\]
\]
\]
\]
endcase

Note:

- Values are variables or $λ$–abstractions
- You could think of this interpreter as an operational semantics for our call–by–value application order lambda calculus (defining interpreter)
Lambda calculus — final remarks

- We can express all computable functions in our λ-calculus. However, nobody “programs” in lambda calculus. For that we have more “convenient” functional languages.

- All computable functions can be express by the following two combinators, referred to as $S$ and $K$:
  
  - $K \equiv \lambda x y. x$
  
  - $S \equiv \lambda x y z. x (z (y z))$

  Combinatory logic is as powerful as Turing Machines.
Next class

• Revisiting our TINY interpreter

• Type systems

• Using unification to perform polymorphic type checking

Read Scott: Chapter 7.1 - 7.2; ALSU Chapter 6.5