



60023

**Type Systems for
Programming Languages
Imperial College London**

Contents

1	Introduction	2
2	Lambda Calculus	3
2.1	Introduction to Lambda Calculus	3
2.2	Reduction Strategies	5
2.2.1	Head Reduction	5
2.2.2	Call By Name / Lazy	5
2.2.3	Call By Value	5
2.2.4	Normal Order	5
2.2.5	Applicative Order	5
2.2.6	Computability	6
2.3	Normal Forms	6
2.4	Approximation Semantics	8
2.4.1	Properties of Approximants	9
3	Credit	10

Chapter 1

Introduction

Chapter 2

Lambda Calculus

2.1 Introduction to Lambda Calculus

λ -Terms	Definition 2.1.1
<p>Given the set of term-variables $\mathcal{V} = \{x, y, z, \dots\}$, a λ-term is defined by the grammar:</p> $M, N ::= \underset{\text{variable}}{x} \quad \quad \underset{\text{abstraction}}{(\lambda x.M)} \quad \quad \underset{\text{application}}{(M \ N)}$ <p>We can also describe this using an <i>inference system</i>:</p> $\frac{}{x \in \Lambda} (x \in \mathcal{V}) \quad \frac{M \in \Lambda}{(\lambda x.M) \in \Lambda} ((x \in \mathcal{V})) \quad \frac{M \in \Lambda \quad N \in \Lambda}{(M \ N) \in \Lambda}$ <ul style="list-style-type: none"> • In a lambda term $M \cdot N$, M is in the <i>function position</i> and N is an <i>argument</i> • The leftmost, outer brackets can be ommitted ($M \ N \ (P \ Q) = ((M \ N) \ (P \ Q))$) • Abstractions can be abbreviated $\lambda xyz.M = (\lambda x.(\lambda y.(\lambda z.M)))$ • Computation is expressed through term substitution. 	

Free Variables	Definition 2.1.2	Bound Variables	Definition 2.1.3
$\begin{aligned} fv(x) &= \{x\} \\ fv(\lambda y.M) &= fb(M) \setminus \{y\} \\ fb(M \ N) &= fv(M) \cup fv(N) \end{aligned}$ <p>A λ-term M is closed if $fv(M) = \emptyset$.</p>		$\begin{aligned} bv(x) &= \emptyset \\ bv(\lambda y.M) &= bv(M) \cup y \\ bv(M \ N) &= bv(M) \cup bv(N) \end{aligned}$	

We can define term substitution inductively as:
Where $P[N/x]$ means replace x by N in λ -term P .

This definition can result in variable capture, for example:

$$\begin{aligned} x[N/x] &= N \\ y[N/x] &= y \\ (P \ Q)[N/x] &= P[N/x] \ Q[N/x] \\ (\lambda y.M)[N/x] &= \lambda y.(M[N/x]) \text{ where } y \neq x \\ (\lambda x.M)[N/x] &= \lambda x.M \end{aligned}$$

$$(\lambda x.y \ x)[y/x] = \lambda x.x \ x$$

Here the free y was substituted for another free variable x , however has been captured by the bound x in the abstraction.

Barendregt's convention	Definition 2.1.4
<p>Given some $(\lambda x.M)N$ we can assume:</p> $\begin{aligned} x \notin fv(N) & \quad x \text{ is not free in } N \\ \forall y \in bv(M). [y \notin fv(N)] & \quad \text{All bound variables in } M \text{ are not free in } N \end{aligned}$ <p>We can always rename the bound variables of a term, this is a fundamental feature to the degree that α-conversion rarely plays a role and terms are considered modulo α-conversion.</p>	

Equivalence Relation	Definition 2.1.5
A binary relation that is reflexive, symmetric and transitive.	

α -Conversion	Definition 2.1.6	α -Equivalence	Definition 2.1.7
$(\lambda x.M)N \rightarrow_\alpha (\lambda z.M[z/x])N$ where z is a new Renaming bound variables within a term.		$N \rightarrow_\alpha M \wedge M \rightarrow_\alpha N \Leftrightarrow M =_\alpha N$ Terms that can be made equal by α -conversion are α -Equivalent	

β -Conversion	Definition 2.1.8
$\begin{array}{ccc} (\lambda x.M)N & \rightarrow_\beta & M[N/x] \\ \text{Reducible Expression/Redex} & & \text{Contractum/Reduct} \end{array}$ <p>The <i>one-step</i> reduction \rightarrow_β can be defined with contextual closure rules:</p> $M \rightarrow_\beta N \Rightarrow \begin{cases} \lambda x.M & \rightarrow_\beta \lambda x.N \\ P M & \rightarrow_\beta P N \\ M P & \rightarrow_\beta N P \end{cases}$ <p>\rightarrow^*_β or \rightarrow_β is the transitive closure of \rightarrow_β. We can also define this using an inference system:</p> $\begin{aligned} (\beta) : \frac{}{(\lambda x.M)N \rightarrow_\beta M[N/x]} \quad (\text{Appl-L}) : \frac{M \rightarrow_\beta N}{M P \rightarrow_\beta N P} \quad (\text{Appl-R}) : \frac{M \rightarrow_\beta N}{P M \rightarrow_\beta P N} \\ (\text{Abstr}) : \frac{M \rightarrow_\beta N}{\lambda x.M \rightarrow_\beta \lambda x.N} \\ (\text{Inherit}_r) : \frac{M \rightarrow_\beta N}{M \rightarrow^*_\beta N} \quad (\text{Refl}) : \frac{}{M \rightarrow^*_\beta M} \quad (\text{Trans}_r) : \frac{M \rightarrow^*_\beta N \quad N \rightarrow^*_\beta P}{M \rightarrow^*_\beta P} \\ (\text{Inherit}_l) : \frac{M \rightarrow^*_\beta N}{M =_\beta N} \quad (\text{Symm}) : \frac{M =_\beta N}{N =_\beta M} \quad (\text{Trans}_{eq}) : \frac{M =_\beta N \quad N =_\beta P}{M =_\beta P} \end{aligned}$ <p>β-reduction is confluent/satisfies the Church-Rosser property:</p> $\forall N, M, P. [M \rightarrow^*_\beta N \wedge M \rightarrow^*_\beta P \Rightarrow \exists Q. [N \rightarrow^*_\beta Q \wedge P \rightarrow^*_\beta Q]]$	

β -conversion does not conform to *Barendregt's convention*, for example:

$$\begin{aligned} (\lambda xy.xy)(\lambda xy.xy) &\rightarrow (\lambda xy.xy)[(\lambda xy.xy)/x] = \lambda y.(\lambda xy.xy)y \\ &\rightarrow \lambda y.(\lambda xy.xy)[y/x] = \lambda y.(\lambda y.yy) \end{aligned}$$

We can avoid this by alpha converting the term to $\lambda y.(\lambda xz.xz)y$ before β -conversion.

η -Reduction	Definition 2.1.9
<p>Given $x \notin fv(M)$ then $\lambda x.M x \rightarrow_\eta M$</p> <p>$\eta$-reduction can be used for eta equivalence. If $f x = g x$ then we can eta reduce both to $f = g$.</p> <ul style="list-style-type: none"> • Eta reduction is a common lint provided by hlint for haskell. 	

2.2 Reduction Strategies

Evaluation Context	Definition 2.2.1
<p>A term with a single hole \square :</p> $C ::= \square \mid C M \mid M C \mid \lambda x.C$ <p>$C[M]$ is the term obtained from context C by replacing the <i>hole</i> \square with M.</p> <ul style="list-style-type: none"> This allows any variables to be captured. <p>The one step β-reduction rule can be defined for any evaluation context as:</p> $C_N[(\lambda x.M)N] \rightarrow C_N[M[N/x]]$	

2.2.1 Head Reduction

$$\frac{}{(\lambda x.M)N \rightarrow_H M[N/x]} \quad \frac{M \rightarrow_H N}{\lambda x.M \rightarrow_H \lambda x.N} \quad \frac{M \rightarrow_H N}{M P \rightarrow_H N P}$$

Reduce the leftmost term, if this is an abstraction, reduce the inside of the abstraction.

2.2.2 Call By Name / Lazy

$$\frac{}{(\lambda x.M)N \rightarrow_N M[N/x]} \quad \frac{M \rightarrow_N N}{M P \rightarrow_N N P}$$

Reduce the leftmost term. Do not reduce unless a term is applied (lazy evaluation).

We can also express reduction strategy with an evaluation context:

$$C_N ::= \square \mid C_N M \quad \text{where } \rightarrow_\beta^N \text{ is defined as } C_N[(\lambda x.M)N] \rightarrow C_N[M[N/x]]$$

Note that there is only ever one redex to contract.

2.2.3 Call By Value

Given V denotes abstractions and variables (values):

$$\frac{}{(\lambda x.M)V \rightarrow_V M[V/x]} \quad \frac{M \rightarrow_V N}{M P \rightarrow_V N P} \quad \frac{M \rightarrow_V N}{V M \rightarrow_V V N}$$

We can apply values, the leftmost term that is not a value is reduced first.

We can also express reduction strategy with an evaluation context:

$$C_V ::= \square \mid C_V M \mid V C_V \quad \text{where } \rightarrow_\beta^V \text{ is defined as } C_V[(\lambda x.M)V] \rightarrow C_V[M[V/x]]$$

Note that there is only ever one redex to contract.

2.2.4 Normal Order

$$\frac{}{(\lambda x.M)N \rightarrow_N M[N/x]} \quad \frac{M \rightarrow_N N}{M P \rightarrow_N N P} \quad \frac{M \rightarrow_N N}{P M \rightarrow_N P N} (P \text{ contains no redexes}) \quad \frac{M \rightarrow_N N}{\lambda x.M \rightarrow_N \lambda x.N}$$

Reduce the leftmost term until it contains no redexes (then continue to other terms), can reduce the inside of an abstraction.

2.2.5 Applicative Order

$$\frac{}{(\lambda x.M)N \rightarrow_A M[N/x]} (M, N \text{ contain no redexes}) \quad \frac{M \rightarrow_A N}{M P \rightarrow_a N P}$$

$$\frac{M \rightarrow_A N}{P M \rightarrow_A P N} (P \text{ contains no redex}) \quad \frac{M \rightarrow_A N}{\lambda x.M \rightarrow_A \lambda x.N}$$

2.2.6 Computability

SKI Combinator Calculus	Definition 2.2.2
$\mathcal{S} = \lambda xyz.xz(yz) \quad \mathcal{K} = \lambda xy.x \quad \mathcal{I} = \lambda x.x$ <p>Any operation in lambda calculus can be encoded (by <i>abstraction elimination</i>) into the SKI calculus as a binary tree with leaves of symbols \mathcal{S}, \mathcal{K} & \mathcal{I}.</p>	

It is possible to encode all Turing Machines within *lambda*-calculus and vice versa. This makes λ -calculus (along with Turing Machines) a model for what is computable.

Church-Turing thesis	Extra Fun! 2.2.1
<p>The Church-Turing thesis equivocates the computational power of Turing machines and the lambda calculus. (Wikipedia)</p>	

It is possible to write terms that do not terminate under β -reduction:

$$(\lambda x.xx) (\lambda x.xx) \rightarrow_{\beta} (xx)[(\lambda x.xx)/x] = (\lambda x.xx) (\lambda x.xx)$$

We can also apply functions continuously.

$$\begin{aligned} \lambda f.(\lambda x.f(x x))(\lambda x.f(x x)) &\rightarrow_{\beta} \lambda f.f(x x)[(\lambda x.f(x x))/x] &&= \lambda f.f((\lambda x.f(x x))(\lambda x.f(x x))) \\ &\rightarrow_{\beta} \lambda f.f(f((\lambda x.f(x x))(\lambda x.f(x x)))) \\ &\rightarrow_{\beta} \lambda f.f(f(f((\lambda x.f(x x))(\lambda x.f(x x)))))) \\ &\vdots \\ &\rightarrow_{\beta} \lambda f.f(f(f(f(f(\dots))))) \end{aligned}$$

This term is a *fixed point constructor*.

Fixed-Point Theorem	Definition 2.2.3
$\forall M.\exists N.[M N =_{\beta} N]$ <p>Take $N = Y M$ where $Y = \lambda f.(\lambda x.f(x x))(\lambda x.f(x x))$:</p> $\begin{aligned} Y M &\triangleq \lambda f.(\lambda x.f(x x))(\lambda x.f(x x)) M \\ &\rightarrow_{\beta} (\lambda x.M(x x))(\lambda x, M(x x)) \\ &\rightarrow_{\beta} (\lambda x.M(x x))(\lambda x, M(x x)) \\ M(Y M) &\triangleq M(\lambda f.(\lambda x.f(x x))(\lambda x.f(x x)) M) \\ &\rightarrow_{\beta} M((\lambda x.M(x x))(\lambda x, M(x x))) \end{aligned}$ <p>Hence $M(Y M) =_{\beta} Y M$ meaning that Y is the fixed point constructor of M</p>	

2.3 Normal Forms

Normal Form	Definition 2.3.1
<p>A λ-term is in normal form if it does not contain a <i>redex</i>.</p> $N ::= x \mid \lambda x.N \mid xN_1 \dots N_n \text{ where } (n \geq 0)$ <p>No β or η reductions are possible</p>	

Head Normal Form		Definition 2.3.2
A λ -term is in head normal form if it is an abstraction with a body that is not <i>reducible</i> .		
$H ::= x \mid \lambda x.H \mid xM_1 \dots M_n$ where $n \geq 1 \wedge M_i \in \Lambda$		
This will mean the term is of the form x or $\lambda x_1 \dots x_n.yM_1 \dots M_m$		
<ul style="list-style-type: none"> • y is the <i>head-variable</i> • If a term has a head-normal form, then head-reduction on the term terminates. 		
Head Normalisable	Definition 2.3.3	Strongly Normalisable
A term M is head normalisable if it has a head-normal form.		Definition 2.3.4
$M \rightarrow_{\beta}^* N$ where N is in head normal form		A term M is strongly normalisable if all reduction sequences starting from M are finite.
Meaningless		Definition 2.3.5
A term without a head-normal form is meaningless as it can never interact with any context (can never apply it to some argument).		
Normal Forms		Example Question 2.3.1
Determine the normality of the following terms:		
<ol style="list-style-type: none"> 1. $\lambda f.(\lambda x.f(x\ x))\ (\lambda x.f(x\ x))$ 2. $(\lambda x.x\ x)\ (\lambda x.x\ x)$ 3. $\mathcal{S}\ \mathcal{K}$ 4. $(\lambda ab.b)\ ((\lambda x.x\ x)\ (\lambda x.x\ x))$ 		
<ol style="list-style-type: none"> 1. Not in either head normal form or normal form (contains a redex). <div style="text-align: center;"> $\lambda f.(\lambda x.f(x\ x))(\lambda x.f(x\ x))$ $\rightarrow_{\beta} \lambda f.f((\lambda x.f(x\ x))\ (\lambda x.f(x\ x)))$ </div> <p>However the β-reduction is in head normal form (head-variable is f).</p> 2. It is a redex, so its not in a normal form. It does not have a normal form as it reduces to itself, so all reducts contain a redex. It has no head-normal form. 3. Hence the original λ-term is not normal form, but it can be normalised. <div style="margin-left: 40px;"> $\begin{aligned} & \mathcal{S}\ \mathcal{K} && \text{Must expand } \mathcal{S} \text{ and } \mathcal{K} \\ = & (\lambda xyz.xz(yz))\ (\lambda xy.x) && \text{Is a redex} \\ \rightarrow_{\beta} & (\lambda xyz.xz(yz))\ (\lambda xy.x) && \text{We rename } y \text{ as per barendregt's convention} \\ =_{\alpha} & (\lambda xyz.xz(yz))\ (\lambda xa.x) \\ \rightarrow_{\beta} & (\lambda yz.(\lambda xa.x)z(yz)) \\ \rightarrow_{\beta} & (\lambda yz.(\lambda a.z)(yz)) \\ \rightarrow_{\beta} & (\lambda yz.z) \end{aligned}$ </div> <p>As all possible redexes are contracted it is <i>strongly normalisable</i>.</p> 4. Contracting the outermost redex results in normal form ter $\lambda b.b$. However contracting the inner term yields itself. Hence it is normalisable, but not <i>strongly normalisable</i>. 		

2.4 Approximation Semantics

There are many methods of describing the semantics of the λ -calculus.

- Reduction rules with *operational semantics*
- set theory with *denotational semantics*

The approach studied in this module defines semantics in a denotational style, but using a reduction system for its definition.

We introduce an extension to the λ -calculus syntax by adding the constant \perp ,

- \perp means unknown/meaningless/no information
- used to mask sub-terms (typically containing redexes) to allow us to focus on the the *stable* parts of the term that do not change under reduction.

The set of $\Lambda\perp$ -terms is defined as:

$$M, N ::= z \mid \perp \mid \lambda x.M \mid M N$$

β -reduction is extended to \rightarrow_\perp to include:

$$\lambda x.\perp \rightarrow_\perp \perp \quad \text{and} \quad \perp M \rightarrow_\perp \perp$$

The set of normal forms of $\Lambda\perp$ with respect to \rightarrow_\perp is the set \mathcal{A} :

$$A ::= \perp \mid \lambda x.A \ (A \neq \perp) \mid xA_1 \dots A_n$$

Note that $\lambda x.\perp$ is considered a redex.

Approximant

Definition 2.4.1

An approximant is a redex-free $\Lambda\perp$ -normal forms that can contain \perp and are used to represent finite parts of potentially infinitely large λ -terms in head-normal form.

The partial order $\sqsubseteq \subseteq (\Lambda\perp)^2$ is defined as the smallest pre-order (reflexive and transitive) such that:

$$\begin{array}{lll} \perp \sqsubseteq M & M \sqsubseteq M' & \Rightarrow \lambda x.M \sqsubseteq \lambda x.M' \\ x \sqsubseteq x & M_1 \sqsubseteq M'_1 \wedge M_2 \sqsubseteq M'_2 & \Rightarrow M_1 M_2 \sqsubseteq M'_1 M'_2 \end{array}$$

- For $A \in \mathcal{A}, M \in \Lambda$, if $A \sqsubseteq M$ then A is the *direct approximant* of M
- The set of *approximants* of M , $\mathcal{A}(M)$ is defined as:

$$\mathcal{A}(M) \triangleq \{A \in \mathcal{A} \mid \exists M' \in \Lambda. [M \rightarrow_\beta^* M' \wedge A \sqsubseteq M']\}$$

- If A is a *direct approximant* of M , then A and M have the same structure, but some parts A contains \perp (\perp masking part of M).
- Redexes in M are masked by \perp in A (\perp masks the redex, or a larger location that contains the redex).

Direct Approximants

Example Question 2.4.1

Show the direct approximants for each reduction step of:

1. $\mathcal{S} \mathcal{K}$
2. $\mathcal{S} a \mathcal{K}$

1.

$$\mathcal{S} \mathcal{K} = (\lambda xyz.xz(yz)) (\lambda ab.a) \xrightarrow{\beta} \lambda yz.(\lambda ab.a)z(yz) \xrightarrow{\beta} \lambda yz.(\lambda b.z)(yz) \xrightarrow{\beta} \lambda yz.z$$

$$\{\perp\} \qquad \{\perp\} \qquad \{\perp\} \qquad \{\perp, \lambda yz.z\}$$

2.

$$\begin{array}{lll} \mathcal{S} a \mathcal{K} = & (\lambda xyz.xz(yz)) a (\lambda cd.c) & \{\perp\} \\ \xrightarrow{\beta} & (\lambda yz.az(yz)) (\lambda cd.c) & \{\perp\} \\ \xrightarrow{\beta} & (\lambda z.az((\lambda cd.c)z)) & \{\perp, \lambda z.a\perp\perp, \lambda z.az\perp\} \\ \xrightarrow{\beta} & (\lambda z.az(\lambda d.z)) & \{\perp, \lambda z.a\perp\perp, \lambda z.az\perp, \lambda a\perp(\lambda d.z), \lambda az(\lambda d.z)\} \end{array}$$

Some basic approximants are:

$$\begin{aligned}
\mathcal{A}(\lambda x.x) &= \{\perp, \lambda x.x\} \\
\mathcal{A}(\lambda x.x x) &= \{\perp, \lambda x.x \perp, \lambda x.x x\} \\
\mathcal{A}(\lambda x.x((\lambda y.yy)(\lambda y.yy))) &= \{\perp, \lambda x.x \perp\} \\
\mathcal{A}(\mathcal{S}) = \mathcal{A}(\lambda xyz.xz(yz)) &= \{\perp, \lambda xyz.x \perp \perp, \lambda xyz.x \perp(y \perp), \lambda xyz.x \perp(yz), \lambda xyz.xz \perp, \lambda xyz.xz(y \perp), \lambda xyz.xz(yz)\} \\
\mathcal{A}(\lambda f.(\lambda x.f(x x)) (\lambda x.f(x x))) &= \{\perp, \lambda f.f(\perp), \lambda f.f(f(\perp)), \lambda f.f(f(f(\perp))), \dots\}
\end{aligned}$$

2.4.1 Properties of Approximants

$$(A \in \mathcal{A}(xM_1 \dots M_n) \wedge A \neq \perp \wedge A' \in \mathcal{A}(N)) \Rightarrow AA' \in \mathcal{A}(xM_1 \dots M_n N)$$

Given A is in the approximants of some variable x are lambda terms $M_1 \dots M_n$, and A' in the approximants of N , then AA' is in the approximants of $A A'$ (Applying A to A').

$$(A \in \mathcal{A}(Mz) \wedge z \notin fv(M)) \Rightarrow \left(\begin{array}{l} A = \perp \\ \vee A \equiv A'z \text{ where } z \notin fv(A') \wedge A' \in \mathcal{A}(M) \\ \vee \lambda x.A \in \mathcal{A}(M) \end{array} \right)$$

If A is an approximant of Mz , and z is not free in M , then either:

- A is \perp
- A is some $A'z$, hence by η -reduction, we can see $A' \in \mathcal{A}(M)$ (the z part can be disregarded, and just look at approximants of M).

$$A \sqsubseteq M \wedge M \rightarrow_{\beta}^* N \Rightarrow A \sqsubseteq N$$

If A is ordered before M , and M β -reduces to N , then A is also before N .

$$A \in \mathcal{A}(M) \wedge M \rightarrow_{\beta}^* N \Rightarrow A \in \mathcal{A}(N) \quad A \in \mathcal{A}(N) \wedge M \rightarrow_{\beta}^* N \Rightarrow A \in \mathcal{A}(M)$$

UNFINISHED!!!

Chapter 3

Credit

Content

Based on the Type Systems course taught by Dr Steffen van Bakel.

These notes were written by Oliver Killane.