The Untyped λ-Calculus

MPRI course

Logique Linéaire et Paradigmes Logiques du Calcul,

Year 2023-24, part 4, lecture 1

Beniamino Accattoli

Inria & École Polytechnique, PARTOUT team

The Untyped λ-Calculus

Terms:

$$t, u, r \quad ::= \quad x \mid \lambda x.t \mid tu$$

Application associates to the left.

tur stands for $(tu)r$.

Abstraction has precedence over application.

$\lambda x.tu$ stands for $\lambda x.(tu)$.

Substitution

Meta-level substitution is noted $t\{x \leftarrow u\}$.

It α-renames to not capture variables, for instance:

$$(\lambda x.yx)\{y \leftarrow xx\} = \lambda z.xxz.$$

Contexts

Contexts (= terms with a (single) hole $\langle\cdot\rangle$):

$$C \quad := \quad \langle\cdot\rangle \quad | \quad C\,t \quad | \quad t\,C \quad | \quad \lambda x.C$$

Plugging (= filling the hole):

$$\begin{array}{rcl} \langle\cdot\rangle\langle u\rangle & := & u \\ (C\ t)\langle u\rangle & := & C\langle u\rangle\ t \end{array} \qquad \begin{array}{rcl} (\lambda x.C)\langle u\rangle & := & \lambda x.C\langle u\rangle \\ (t\ C)\ \langle u\rangle & := & t\ C\langle u\rangle \end{array}$$

Plugging can capture variables: $(\lambda x.\langle\cdot\rangle)\langle xy\rangle = \lambda x.xy$.

Approaching the λ-Calculus

There are two main ways to look at the λ-calculus.

Rewriting $\sim \beta$ as a computational step:

$$\beta\text{-reduction} \qquad (\lambda x.t)u \to_\beta t\{x{\leftarrow}u\}$$

Equational $\sim \beta$ as an equivalence:

$$\beta\text{-conversion} \qquad (\lambda x.t)u =_\beta t\{x{\leftarrow}u\}$$

Outline

Definition of β-Reduction

β-reduction can be applied anywhere in a term.

Precise inductive definition:

$$\frac{}{(\lambda x.t)u \rightarrow_\beta t\{x \leftarrow u\}} \text{ (root } \beta\text{)} \qquad \frac{t \rightarrow_\beta u}{tr \rightarrow_\beta ur} \text{ (@l)}$$

$$\frac{t \rightarrow_\beta u}{\lambda x.t \rightarrow_\beta \lambda x.u} \ (\lambda) \qquad \frac{t \rightarrow_\beta u}{rt \rightarrow_\beta ru} \text{ (@r)}$$

Contextual Definition of β-Reduction

Contexts:

$$C \quad := \quad \langle\cdot\rangle \quad | \quad C\,t \quad | \quad t\,C \quad | \quad \lambda x.C$$

β-Reduction, contextual definition:

Root rule	Contextual closure
$(\lambda x.t)u \mapsto_\beta t\{x{\leftarrow}u\}$	$C\langle t\rangle \to_\beta C\langle u\rangle \quad \text{if } t \mapsto_\beta u$

The definition works because of capture:

$\lambda y.((\lambda x.x)y) \to_\beta \lambda y.y$ with $C = \lambda y.\langle\cdot\rangle$ and $(\lambda x.x)y \mapsto_\beta y$.

Terminology

A sub-term of the form $(\lambda x.t)u$ is called a β-redex.

A term without β-redexes is a normal form.

Shape of normal forms:

$$\lambda x_1. \ldots . \lambda x_n. y\, t_1 \ldots t_k$$

with $n, k \geq 0$ and where $t_1, \ldots, t_k$ are themselves normal.

Predicate for Normal Forms

Normal forms can also be described by a normal predicate.

It requires an auxiliary neutral predicate.

$$\frac{}{x \text{ is neutral}} \qquad \frac{t \text{ is neutral} \qquad u \text{ is normal}}{tu \text{ is neutral}}$$

$$\frac{t \text{ is neutral}}{t \text{ is normal}} \qquad \frac{t \text{ is normal}}{\lambda x.t \text{ is normal}}$$

Shape of neutral terms:

$$x t_1 \ldots t_k$$

with $k \geq 0$ and where $t_1, \ldots, t_k$ are normal.

Typical Traits of β-Reduction 2

Divergence:

$$\Omega := (\lambda x.xx)(\lambda y.yy) \to_\beta (\lambda y.yy)(\lambda y.yy) \to_\beta \dots$$

Divergence and normalization may co-exist:

$$y \;\; {}_\beta\!\!\leftarrow (\lambda x.y)\Omega \to_\beta (\lambda x.y)\Omega \to_\beta \dots$$

Terminology and Notations

t is weakly normalizing $:=$ t has a reduction sequence to nf.

t is strongly normalizing $:=$ t has no diverging reduction.

t is strongly divergent $:=$ t has no reduction sequence to nf.

Weak β Reduction

Functional languages use a weak form of β.

$$\frac{}{(\lambda x.t)u \to_w t\{x{\leftarrow}u\}} \qquad \frac{t \to_w t'}{tu \to_w t'u} \qquad \frac{t \to_w t'}{ut \to_w ut'}$$

Key point:
Function bodies are not evaluated (before the function is applied).

Contextually:

$$\textsc{Weak ctxs} \quad W ::= \langle\cdot\rangle \mid Wu \mid uW$$

Root rule	Contextual closure
$(\lambda x.t)u \mapsto_\beta t\{x{\leftarrow}u\}$	$W\langle t\rangle \to_w W\langle u\rangle \quad \text{if } t \mapsto_\beta u$

Closed Terms

Functional languages also evaluate only closed terms.

Weak β + closed terms $\Rightarrow$ normal forms $=$ abstractions.

Abstractions are constructors, also called values.

Call-by-Value

Functional languages are often call-by-value.

β-reduction is restricted to values, noted v:

$$(\lambda x.t)v \quad \to_{\beta_v} \quad t\{x{\leftarrow}v\} \ .$$

For instance, in CbV $(\lambda x.y)\Omega$ can only diverge.

That is, $(\lambda x.y)\Omega \not\to_{\beta_v} y$ because $\Omega = \delta\delta$ is not a value.

Strong λ-Calculus and Proof Assistants

The unrestricted case (under λ, possibly open) is also called strong.

Used in proof assistants with dependent types (e.g. Coq or Agda).

Dependent types may contain terms (with β-redexes) into types.

Type checking requires to normalize types, and then check equality.

Weak vs Strong

The term marking the weak/strong divide is $\lambda x.\Omega$.

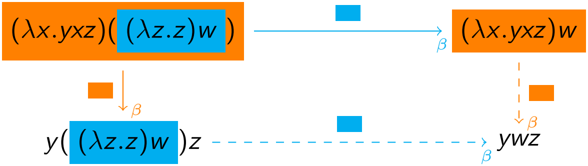
It is normal in the weak case, and divergent in the strong case.

Outline

The λ-Calculus is Confluent

β-reduction is non-deterministic, for instance:

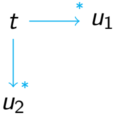
Its non-determinism is harmless, for instance:



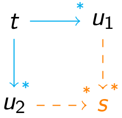
β-reduction is confluent.

Confluence

A rewriting system $(S, \to)$ if confluent when:



implies $\exists s$ s.t.



Corollary: normal forms, when they exist, are unique.

Confluence in general is difficult to prove.

Diamond Property

Confluence is easy if the system has the diamond property:

$$\begin{array}{ccc} t & \longrightarrow & u_1 \\ \downarrow & & \\ u_2 & & \end{array} \quad \text{and } u_1 \neq u_2 \text{ implies } \exists s \text{ s.t.} \quad \begin{array}{ccc} t & \longrightarrow & u_1 \\ \downarrow & & \downarrow \\ u_2 & \dashrightarrow & s \end{array}$$

Diamond Property

Exercise 1: prove that the diamond property implies confluence.

Exercise 2: prove the following lemma

Lemma

Let $\to$ be diamond and $t \to^{k} u$ with u $\to$-normal.

1. *Uniform normalization: no $\to$ reduction sequence from t can be longer than k.*
2. *Random descent: all $\to$ reduction sequences from t to normal form have length k.*

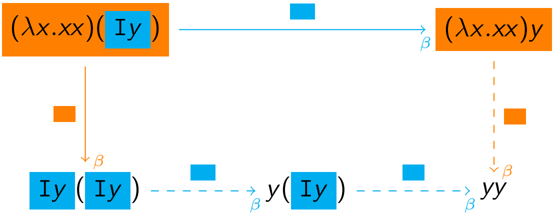
Diamond

Roughly, non-determinism is only apparent.

Essentially, diamond $=$ lax determinism.

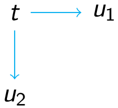
Diamond Property

λ-calculus is not diamond, because of duplication:

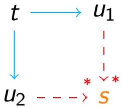


Local Confluence

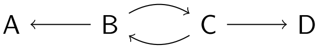
Local confluence is the weaker property:



implies $\exists s$ s.t.



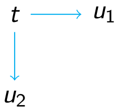
Local confluence does not imply confluence. Counter-example:



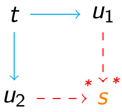
Local Confluence

Lemma (Newman)

Local confluence



implies $\exists s$ *s.t.*



plus strong normalization imply confluence.

But the λ-calculus is not strongly normalizing: consider Ω.
Exercise: prove the lemma.

Proof of Confluence

Confluence of β is usually proved via parallel β reduction $\Rightarrow_\beta$.

Elegant Tait-Martin Löf technique (diamond for $\Rightarrow_\beta$).

Omitted here, first theorem about β in every course.

Theorem with the highest number of formalized proofs.

Weak Evaluation and Confluence

The weak λ-calculus is not confluent:

$$(\lambda x.\lambda y.yx)(\mathtt{II}) \longrightarrow_\beta \lambda y.y(\mathtt{II})$$

$$\downarrow_\beta$$

$$(\lambda x.\lambda y.yx)\mathtt{I} \dashrightarrow_\beta \lambda y.y\mathtt{I}$$

But $\lambda y.y(\underline{\mathtt{II}}) \not\to_\beta \lambda y.y\mathtt{I}$ in the weak case

Both $\lambda y.y(\underline{\mathtt{II}})$ and $\lambda y.y\mathtt{I}$ are normal in the weak λ-calculus.

Weak Evaluation and Confluence

The weak λ-calculus is not confluent:

$$\begin{array}{ccc} (\lambda x.\lambda y.yx)(\texttt{II}) & \longrightarrow_\beta & \lambda y.y(\texttt{II}) \\ \downarrow_\beta & & \\ (\lambda x.\lambda y.yx)\texttt{I} & \dashrightarrow_\beta & \lambda y.y\texttt{I} \end{array}$$

Problem: redexes are weak, but substitution acts under abstraction.

Ad-hoc solutions exist.

Outline

β-Conversion

The equational perspective forgets about the dynamic aspect.

It focuses on β-conversion $=_\beta$ and its extensions.

Outline

Definition of β-conversion

$$\frac{}{(\lambda x.t)u =_\beta t\{x \leftarrow u\}}\ (\text{ax}) \qquad \frac{t =_\beta u}{C\langle t\rangle =_\beta C\langle u\rangle}\ (\text{ctx})$$

$$\frac{}{t =_\beta t}\ (\text{ref}) \qquad \frac{t =_\beta u}{u =_\beta t}\ (\text{sym}) \qquad \frac{t =_\beta u \qquad u =_\beta r}{t =_\beta r}\ (\text{tra})$$

β-Conversion and Normal Forms

β-conversion extends β-reduction.

Are normal forms still unique in β-equivalence classes? Yes.

The proof showcases the link between equations and reductions.

Towards the Consistency of β-Conversion

Alternative definition of β-conversion based on β-reduction:

$$\frac{t \to_\beta^* u}{t =_\beta u}\ (\text{lift}) \qquad \frac{t =_\beta u}{u =_\beta t}\ (\text{sym}) \qquad \frac{t =_\beta u \qquad u =_\beta r}{t =_\beta r}\ (\text{tra})$$

Reflexivity and context closure are inherited from $\to_\beta^*$.

Checking β-conversion can be reduced to β-reduction.

Church-Rosser

Proposition (Church-Rosser property)

If $t =_\beta u$ *then there exists* r *such that* $t \to_\beta{}^* r$ *and* $u \to_\beta{}^* r$*.*

Proof.

By induction on the reduction-based definition of $t =_\beta u$. Cases:

Lifting: if $t \to_\beta{}^* u$ then the statement holds with $r := u$.

Symmetry: if $u =_\beta t$ then the *i.h.* gives the statement.

Transitivity: let $t =_\beta u$ and $u =_\beta r$;

By *i.h.* $\exists$ p, p' s.t. $t \to_\beta{}^* p$, $u \to_\beta{}^* p$, $u \to_\beta{}^* p'$, and $r \to_\beta{}^* p'$;

By confluence on u, $\exists q$ s.t. $p \to_\beta{}^* q$ and $p' \to_\beta{}^* q$;

Then $t \to_\beta{}^* p \to_\beta{}^* q$ and $r \to_\beta{}^* p' \to_\beta{}^* q$. □

Uniqueness of Normal Forms, Equationally

Corollary

No λ-term is β-convertible to two distinct normal forms.

Proof.

Let t be β-convertible to two distinct normal forms u and r.

Since $u =_\beta r$, by Church-Rosser they reduce to a common term.

But u and r are normal, so they cannot reduce,

and they are distinct—absurd. $\square$

Consistency of β-Conversion

Definition

A relation R between λ-terms is:

Consistent if R does not equate all terms;

Inconsistent otherwise.

Corollary

β-conversion is consistent.

Proof.

It does not equate different normal forms. □

Consistency From Confluence

The used proof technique can be made abstract.

For any $\rightarrow_x$:

$$\rightarrow_x \text{ confluent}$$

$$\Rightarrow$$

$$\rightarrow_x \text{ Church-Rosser}$$

$$\Rightarrow$$

$$=_x \text{ consistent.}$$

Outline

λ-Theories

Program equiv. $\sim$ equational theory extending $=_\beta$ $\sim$ λ-theory.

A λ-theory, noted $\mathcal{T}$ or $=_\mathcal{T}$, is an equiv. relation on terms s.t.:

$$\frac{t =_\beta u}{t =_\mathcal{T} u}\,(\beta) \qquad\qquad \frac{t =_\mathcal{T} u}{C\langle t\rangle =_\mathcal{T} C\langle u\rangle}\,(\mathrm{ctx})$$

The smallest λ-theory is β-conversion.

Motivating λ-Theories

Consider $\Omega_3 := \delta_3\delta_3$ where $\delta_3 := \lambda x.xxx$.

Note that $\Omega_3 \to_\beta \delta_3\delta_3\delta_3 = \Omega_3\delta_3 \to_\beta \Omega_3\delta_3\delta_3 \to_\beta \ldots$

Motivating λ-Theories

Ω_3 is strongly divergent, like Ω.

They are in some sense equivalent.

But they are not related by β-reduction.

Motivating λ-Theories

More generally:
Natural to consider all strongly divergent terms as equivalent.

Not naturally captured via rewriting.

Equationally:
The smallest λ-theory $\mathcal{D}$ identifying all strongly divergent terms.

(In)Consistency

Sanity check for a λ-theory $\mathcal{T}$:
$\mathcal{T}$ does not extend β-conversion too much.

Danger:
Closure properties (ctx, equiv. rel.) end up identifying all terms.

Example of Inconsistent λ-Theory

Recall: Surprisingly, $\mathcal{D}$ is inconsistent.

Sign that the λ-calculus:

Is simple to define,

And yet it is a complex framework.

Example of Inconsistent λ-Theory

Proposition

$\mathcal{D}$ is inconsistent.

Proof.

We prove that $t =_{\mathcal{D}} u$ for any two terms t and u.

Note that $\lambda x.xt\Omega =_{\mathcal{D}} \lambda x.xu\Omega$, because Ω is strongly divergent.

$(\lambda x.xt\Omega)(\lambda y.\lambda z.y) =_{\mathcal{D}} (\lambda x.xu\Omega)(\lambda y.\lambda z.y)$ by ctx closure.

Finally (remember that by definition a λ-theory contains $=_\beta$):

$$\begin{array}{lll} t & =_\beta & (\lambda x.xt\Omega)(\lambda y.\lambda z.y) \\ & =_{\mathcal{D}} & (\lambda x.xu\Omega)(\lambda y.\lambda z.y) \quad =_\beta \quad u. \quad \square \end{array}$$

Example of Inconsistent λ-Theory

Such an inconsistency was first noted by Barendregt (1970s).

It is the starting point of his reference book (1984).

It can also be understood via Godel's partial recursive functions.

Encoding Recursive Functions

We shall not see the details of the encoding.

A natural number $n \in \mathbb{N}$ is represented as Church numeral $\underline{n}$.

All you need to know is $\underline{0} := \lambda x.\lambda y.y$.

Representing Total Functions

For a total rec. function $f : \mathbb{N} \to \mathbb{N}$ there exists a λ-term t_f s.t.:

$$t_f\,\underline{n} =_\beta \underline{f(n)}$$

for every $n \in \mathbb{N}$, where $\underline{n}$ is, say, the church encoding of n.

Church numerals are normal, so the results of evaluation is normal.

Representing Partial Functions

Partial functions: $f(n)$ may be undefined, noted $f(n) = \bot$.

Let $f_\bot$ be the everywhere undefined function $f_\bot(n) = \bot$ forall n.

Representing Partial Functions

Natural approach to represent (un)defined in the λ-calculus:

Being defined := reducing to normal form.

Being undefined := $\underbrace{\text{no reductions to normal form}}_{=\,\text{strongly diverging term}}$.

There is a problem, pointed out by Barendregt in the 1970s.

Representing Partial Functions

$x\Omega$ is strongly diverging $\Rightarrow$ it represents undefined.

$\lambda x.x\Omega$ represents the everywhere undefined function $f_\perp$.

Now, $f_\perp(0) = \perp$, while applying $\lambda x.x\Omega$ to $\underline{0} := \lambda x.\lambda y.y$ one has:

$$(\lambda x.x\Omega)(\lambda x.\lambda y.y) \to_\beta (\lambda x.\lambda y.y)\Omega \to_\beta \lambda y.y$$

Problem: $\lambda y.y$ is not divergent, i.e. it does not represent $\perp$.

Representing Partial Functions

Essence of the problem:
Undefined is not stable by composition and substitution.

Consequence: being undefined $\neq$ being strongly divergent.

Back to β-Conversion

$\mathcal{D}$ is not a good λ-theory.

Is β-conversion the λ-theory of reference?

Denotational Models

A denotational model of the λ-calculus is given by:

A mathematical object M, together with

An interpretation $[\![\cdot]\!]$ from λ-terms to elements of M,

Such that the induced equality on λ-terms is a λ-theory.

Defined as $t =_M u$ if $[\![t]\!] = [\![u]\!]$.

Puzzling Fact

Surprisingly, there are no models inducing β-conversion $=_\beta$.

All models induce extensions of $=_\beta$.

β-conversion is not as good as it might seem.

A Layered Approach

Black and white view of divergence/normalization is misleading.

A different, layered approach is needed.

Historically, the change happened via head reduction.

Outline

Outline

Idea

Better representation of partial recursive functions:

Being defined := reducing to head normal form.

Being undefined := no reduction to head normal form.

Introduced by Barendregt and Wadsworth in the 1970s.

Head Reduction

Head = reducing only on the left of applications.

$$\frac{}{(\lambda x.t)u \rightarrow_h t\{x \leftarrow u\}}$$

$$\frac{t \rightarrow_h u \qquad t \neq \lambda x.t'}{tr \rightarrow_h ur}$$

$$\frac{t \rightarrow_h u}{\lambda x.t \rightarrow_h \lambda x.u}$$

Head reduction $\rightarrow_h$ is deterministic.

Head Reduction

Examples:

$$\begin{array}{rcl} \delta\mathtt{I} & \to_h & \mathtt{II} \\ \lambda x.\delta\mathtt{I} & \to_h & \lambda x.\mathtt{II} \\ (\lambda x.\delta\mathtt{I})t & \not\to_h & (\lambda x.\mathtt{II})t \\ (\lambda x.\delta\mathtt{I})t & \to_h & \delta\mathtt{I} \end{array}$$

Note that:

$$y(\delta\mathtt{I}) \not\to_h x(\mathtt{II})$$

because $y(\delta\mathtt{I})$ is head normal.

Head Normal Forms

Head normal forms have shape:

$$\lambda x_1. \ldots .\lambda x_n . y\, t_1 \ldots t_k$$

where y may be one of the x_i, and $n, k \geq 0$.

Arguments of the head variable y need not to be normal.

Barendregt Counter-Example, Revisited

Recall the undefined in Barendregt's counter-example: $x\Omega$.

It is head normal, so in the refined theory it is defined.

Undefined-as-head-divergent is stable by substitution.

If t has no head normal form, then neither does $t\{x{\leftarrow}u\}$.

Careful: other good definitions of (un)defined exist.

The λ-Theory $\mathcal{H}$

A λ-theory $\mathcal{T}$ is head-collapsing if:

$t =_{\mathcal{T}} u$ for any two $\to_h$-diverging terms t and u.

We note $\mathcal{H}$ the smallest head-collapsing λ-theory.

Consistency

Theorem

$\mathcal{H}$ is consistent.

The proof is non-trivial.

We first need to develop some tools.

Contextual Definition

Defining head reduction using contexts requires an auxiliary notion.

Applicative contexts	A	$::=$	$\langle\cdot\rangle \mid At$
Head contexts	H	$::=$	$A \mid \lambda x.H$

Root rule

$$(\lambda x.t)u \mapsto_{\beta} t\{x{\leftarrow}u\}$$

Contextual closure

$$\frac{t \mapsto_{\beta} t'}{H\langle t\rangle \rightarrow_{h} H\langle t'\rangle}$$

Contextual Definition

Applicative contexts	A	$::=$	$\langle\cdot\rangle \mid At$
Head contexts	H	$::=$	$A \mid \lambda x.H$

Root rule

$$(\lambda x.t)u \mapsto_\beta t\{x{\leftarrow}u\}$$

Contextual closure

$$\frac{t \mapsto_\beta t'}{H\langle t\rangle \to_h H\langle t'\rangle}$$

In proof nets:

Head contexts are out of all !-boxes.

Reminder: boxes are associated to arguments.

Contextual Definition

The need for applicative context comes from this clause:

$$\frac{t \to_h u \qquad t \neq \lambda x.t'}{tr \to_h ur}$$

Where one needs to be sure that t is not an abstraction.

Lax Head Reduction

Removing $t \neq \lambda x.t'$ gives a simpler, more general notion.

Lax Head contexts $\quad \overline{H} ::= \langle\cdot\rangle \mid \overline{H}t \mid \lambda x.\overline{H}$

Root rule

$$(\lambda x.t)u \mapsto_\beta t\{x \leftarrow u\}$$

Contextual closure

$$\frac{t \mapsto_\beta t'}{\overline{H}\langle t\rangle \to_{\overline{h}} \overline{H}\langle t'\rangle}$$

In proof nets:
Lax head contexts are still out of all !-boxes.

Lax Head Reduction

Lax head reduction is non-deterministic:

$$
\begin{array}{ccc}
(\lambda x.\delta \mathtt{I})t & \longrightarrow_{\overline{h}} & \delta \mathtt{I} \\
\downarrow_{\overline{h}} & & \\
(\lambda x.\mathtt{II})t & &
\end{array}
$$

Proposition

Lax head reduction $\rightarrow_{\overline{h}}$ *has the diamond property.*

Lax Head Reduction

Apart from determinism, $\rightarrow_h$ and $\rightarrow_{\overline{h}}$ have the same properties.

We shall then refer to both as head reduction.

Head *vs* Weak and Strong

Head reduction is in between weak and strong β-reduction.

Not weak, because it goes under abstraction.

Not fully strong, because it doe not go into arguments.

Weak Head Reduction

There is a weak variant of head reduction.

$$\frac{}{(\lambda x.t)u \to_{wh} t\{x\leftarrow u\}} \qquad \frac{t \to_{wh} u}{t r \to_{wh} u r}$$

It is also called call-by-name reduction.

Defined/Undefined := reducing to/not having weak head nf.

The obtained λ-theory $\mathcal{W}$ is consistent.

Iterated Head Reduction

There is also a strong variant of head reduction.

Obtained by iterating head reduction into arguments on head nfs.

From Before: Predicate for Normal Forms

Normal forms can also be described by a normal predicate.

It requires an auxiliary neutral predicate.

$$\frac{}{x \text{ is neutral}} \qquad \frac{t \text{ is neutral} \qquad u \text{ is normal}}{tu \text{ is neutral}}$$

$$\frac{t \text{ is neutral}}{t \text{ is normal}} \qquad \frac{t \text{ is normal}}{\lambda x.t \text{ is normal}}$$

Shape of neutral terms:

$$x\, t_1 \ldots t_k$$

with $k \geq 0$ and where $t_1, \ldots, t_k$ are normal.

Leftmost(-Outermost) Strategy

Iterated head reduction = leftmost-outermost reduction.

$$\frac{}{(\lambda x.t)u \rightarrow_{lo} t\{x \leftarrow u\}} \qquad \frac{t \rightarrow_{lo} u \qquad t \neq \lambda x.t'}{tr \rightarrow_{lo} ur}$$

$$\frac{t \rightarrow_{lo} u}{\lambda x.t \rightarrow_{lo} \lambda x.u} \qquad \boxed{\frac{r \text{ is neutral} \qquad t \rightarrow_{lo} u}{rt \rightarrow_{lo} ru}}$$

Note that 3 clauses are exactly those for head reduction.

Leftmost(-Outermost) Strategy

We shall show:
$\rightarrow_{lo}$ diverging terms = strongly diverging terms.

Then $\rightarrow_{lo}$ diverging terms cannot represent undefined.

Back to Layered Normalization

Head reduction introduces a layerization of normalization.

Naturally leads to infinite normal forms.

Not available in the black and white approach.

Layered Normalization and Infinite Normal Forms

Let $\delta_y := \lambda x.y(xx)$ and $\Omega_y := \delta_y \delta_y$. Since:

$$\Omega_y \;=\; \delta_y \delta_y \;\to_h\; y(\delta_y \delta_y) \;=\; y\Omega_y$$

Ω_y β-diverges, producing an infinite sequence of occurrences of y:

$$y(y(y \ldots$$

$y\Omega_y$ is head normal, so that Ω_y is defined in $\mathcal{H}$.

A β-undefined term becomes a hereditarily head-defined one.

Enlarging the set of meaningful programs.

Weak Head Layers

Weak head reduction introduces a further layerization.

Weak Head Layers

Let $\delta_\lambda := \lambda x.\lambda y.xx$ and $\Omega_\lambda := \delta_\lambda \delta_\lambda$. Since:

$$\Omega_\lambda \;=\; \delta_\lambda \delta_\lambda \;\to_{wh}\; \lambda y.\delta_\lambda \delta_\lambda \;=\; \lambda y.\Omega_\lambda$$

Ω_λ $\to_h$-diverges, producing infinitely many abstractions $\lambda y.$:

$$\lambda y.\lambda y.\lambda y.\ldots$$

$\lambda y.\Omega_\lambda$ is weak head normal, so that Ω_λ is defined in $\mathcal{W}$.

A $\to_h$-undefined term becomes a hereditarily $\to_{wh}$-defined one.

Enlarging again the set of meaningful programs.

Outline

Head Factorization

Head reduction can also be motivated via rewriting.

Let $\rightarrow_{\neg h}$ denote a β step that is not head.

Theorem (Head Factorization)
If $t \rightarrow_{\beta}^{*} u$ then $t \rightarrow_{h}^{*} r \rightarrow_{\neg h}^{*} u$ for some r.

Roughly, it means that head steps are more relevant.

Head Factorization

Theorem (Head Factorization)

If $t \to_{\beta}^{*} u$ then $t \to_{h}^{*} r \to_{\neg h}^{*} u$ for some r.

The opposite factorization $t \to_{\neg h}^{*} r \to_{h}^{*} u$ does not hold.

Consider:

$$(\lambda x.xy(xy))(\lambda z.z) \quad \to_h \quad (\lambda z.z)y((\lambda z.z)y) \quad \to_{\neg h} \quad (\lambda z.z)yy$$

It cannot be re-organized as to have $\to_{\neg h}$ before $\to_h$.

Head Factorization

The proof of head factorization is non-trivial.

Based on parallel reduction, akin to confluence

Factorization also hold for weak head and leftmost reduction.

Relatively to their respective dual reductions.

Normalizing Strategies

A strategy $\rightarrow_x$ is normalizing if it normalizes whenever possible.

In the strong λ-calculus, leftmost reduction is normalizing.

This is a key theorem.

Proving the Leftmost Normalization Theorem

There is an elegant proof of the leftmost normalization theorem.

It rests on three abstract properties.

Two of them require the dual $\rightarrow_{\neg lo}$ of $\rightarrow_{lo}$.

Abstract Properties for Untyped Normalization

Fullness: if $t \rightarrow_{\beta} u$ then $t \rightarrow_{lo} r$ for some r.

If there is a redex then there is a leftmost redex, immediate.

Persistence: if $t \rightarrow_{lo} u$ and $t \rightarrow_{\neg lo} r$ then $r \rightarrow_{lo} p$ for some p.

Non-leftmost redexes cannot erase the leftmost redex, immediate.

Factorization: if $t \rightarrow_{\beta}^{*} u$ then $t \rightarrow_{lo}^{*} r \rightarrow_{\neg lo}^{*} u$ for some r.

As for head factorization, the proof is non-trivial.

Leftmost Normalization Theorem

Theorem (Leftmost (untyped) normalization)

Let $t \to_\beta{}^* u$ with u normal. Then $t \to_{lo}{}^* u$.

Proof.

If $t \to_\beta^* u$ then by factorization $t \to_{lo}{}^* r \to_{\neg lo}{}^* u$ for some r.

Let us show that r is $\to_{lo}$ normal.

If by contradiction it is not, then $r \to_{lo} p$.

By iterated persistence on $r \to_{\neg lo}{}^* u$ we have $u \to_{lo} q$ for some q.

But u is normal by hypohtesis—absurd. Then r is $\to_{lo}$-normal.

By fullness, r is normal.

By uniqueness of normal forms, $r = u$, that is, $t \to_{lo}{}^* u$. □

Head Normalization Theorem

There also is a head normalization theorem.

As well as a weak head one.

Theorem (Head normalization)

Let $t \to_\beta{}^* u$ with u head normal. Then $\to_h$ terminates on t.

Note $\to_h$ terminates on t and not $t \to_h^* u$.

Consider the following $\to_\beta$-sequence to head normal form:

$$\mathtt{I}(x(\mathtt{II})) \quad \to_\beta \quad \mathtt{I}(x\mathtt{I}) \quad \to_\beta \quad x\mathtt{I}$$

And the head normalization $\mathtt{I}(x(\mathtt{II})) \to_h x(\mathtt{II}) \neq x\mathtt{I}$.

Three Depths

Summing up:

There are three depths of the λ-calculus:

Closed/Weak

$\downarrow$ eval under abs & open terms

Head

$\downarrow$ eval in arguments

Strong

The λ-calculus does not exist, there is a multitude of λ-calculi.

Outline

Solvability

There is a nice characterization of head normalization.

Definition
A term t is solvable if there is a head context H such that $H\langle t\rangle \to_\beta^* \mathtt{I}$, and unsolvable otherwise.

Theorem (Wadsworth)
A term t is solvable if and only if t is head normalizing.

Solvability

Theorem (Wadsworth)

A term t is solvable if and only if t is head normalizing.

Proof.

Direction $\Leftarrow$ is an easy exercise.

For direction $\Rightarrow$, one needs (easy fact, proof omitted):

if $H\langle t\rangle$ is head normalizing then t is head normalizing.

Now, if t is solvable then there exists H s.t. $H\langle t\rangle \to_\beta^* \mathtt{I}$.

By the head normalization thm, $H\langle t\rangle$ is head normalizing.

By the fact above, t is head normalizing. $\square$

Solvability and Divergence

Solvability clarifies the divergent / undefined relationship.

There are two forms of divergence.

Head unremovable, such as Ω.

It corresponds to loops, represents undefined.

Head removable, such as $x\Omega$, removed by $(\lambda x.\langle\cdot\rangle)(\lambda y.\mathtt{I})$.

Aka loops activated under some conditions, not undefined.

Head Contextual Equivalence

There is also an interactive approach to λ-theories, due to Morris.

Definition

t and u are head contextual equivalent, noted $t =_C^h u$,

If for all contexts C we have that

$C\langle t\rangle \rightarrow_\beta^* t'$ with t' in head normal form

if and only if

$C\langle u\rangle \rightarrow_\beta^* u'$ with u' in head normal form.

Head Contextual Equivalence

The head normalization thm induces an alternative definition.

Definition

t and u are head contextual equivalent, noted $t =_C^h u$,

If for all contexts C we have that

$C\langle t\rangle$ is $\to_h$-normalizing if and only if $C\langle u\rangle$ is $\to_h$-normalizing.

Head Contextual Equivalence

In the literature, $=_C^h$ is also called $\mathcal{H}^*$.

Contextual equivalences are hard to study.

Because of the quantification over contexts.

Head Contextual Equivalence

Proposition

Head contextual equivalence $=_C^h$ is a consistent λ-theory.

Proof.

Consistency is easy: $\Omega \neq_C^h \mathtt{I}$ by considering the empty context.

Equiv. relation and contextual closure follow from the definition.

Containement of $\to_\beta$:
It follows from confluence and the following easy fact:

If t is $\to_h$-normal and $t \to_\beta u$ then u is $\to_h$-normal.

□

Outline

Consistency of $\mathcal{H}$

The consistency of $\mathcal{H}^*$ can be used to prove the consistency of $\mathcal{H}$.

By proving that $\mathcal{H} \subseteq \mathcal{H}^*$, that is, that $\mathcal{H}^*$ is head-collapsing.

The proof that $\mathcal{H}^*$ is head-collapsing is non-trivial.

It rests on an interactive property, called light genericity.

Light Genericity

Theorem (Light genericity)

Let t be $\to_h$-divergent. If $C\langle t\rangle$ is $\to_h$-normalizing then $C\langle u\rangle$ is $\to_h$-normalizing for every term u.

Proof.

Omitted (see notes). It uses the head normalization thm. □

Consistency of $\mathcal{H}$

Corollary

1. *$\mathcal{H}^*$ is head-collapsing.*
2. *$\mathcal{H}$ is consistent.*

Proof.

1. Let t and u be $\to_h$-divergent and let C be a context.

   $C\langle t\rangle$ $\to_h$-normalizing

   $\Rightarrow$ by light genericity $C\langle u\rangle$ $\to_h$-normalizing

   By exchanging the role of t and u, we obtain the converse.

   That is, $t =^h_C u$.

2. It follows from Point 1 and the consistency of $\mathcal{H}^*$. $\square$

$\mathcal{H}$ and $\mathcal{H}^*$

Is it the case that $\mathcal{H} = \mathcal{H}^*$? No.

Strict Inclusion and η

Consider η-conversion:

$$\frac{x \notin \texttt{fv}(t)}{\lambda x.tx =_\eta t} \qquad \frac{t =_\eta u}{C\langle t\rangle =_\eta C\langle u\rangle}$$

η-conversion is included in $\mathcal{H}^*$ but not in $\mathcal{H}$.

That is, $\mathcal{H} \subsetneq \mathcal{H}^*$.

Unfortunately, proving these facts requires some detours.

Omitted (see the TD).

Böhm Trees

The λ-theory $\mathcal{B}$ induced by Böhm trees is such that $\mathcal{H} \subsetneq \mathcal{B} \subsetneq \mathcal{H}^*$.

By definition, Böhm trees equate all $\to_h$-divergent terms.
Since they all have trivial Böhm tree $\perp$.

Thus $\mathcal{H} \subseteq \mathcal{B}$.

Böhm Trees

Example of terms equated by $\mathcal{B}$ and not by $\mathcal{H}$: fix-point operators.

Thus $\mathcal{H} \subsetneq \mathcal{B}$.

They all have the same Böhm tree.

We shall show $\mathcal{B} \subseteq \mathcal{H}^*$ in a few slides.

$\mathcal{B}$ does not include η: $x =_\eta \lambda y.xy$ have different Böhm trees.

Thus $\mathcal{B} \subsetneq \mathcal{H}^*$.

Outline

Maximality of $\mathcal{H}^*$

Our final theorem: $\mathcal{H}^*$ is a maximal consistent λ-theory.

Any extension of $\mathcal{H}^*$ is inconsistent.

Head-Adequacy

We need a concept.

A λ-theory $\mathcal{T}$ is head-adequate if:

$t =_{\mathcal{T}} u$ and $t \to_h$-normalizing imply $u \to_h$-normalizing.

Key Property

Theorem

Let $\mathcal{T}$ be a λ-theory that is head-collapsing but not head-adequate. Then $\mathcal{T}$ is inconsistent.

Proof.

$\mathcal{T}$ is not h-adequate $\Rightarrow$ $t =_{\mathcal{T}} u$ with t h-normaliz. and u h-diverg.

By solvability, $\exists H$ such that $H\langle t\rangle \to_\beta^* \mathtt{I}$.

By the def of equational theory, we have $\mathtt{I} =_{\mathcal{T}} H\langle t\rangle =_{\mathcal{T}} H\langle u\rangle$.

Now, let r be a term. Then $r =_{\mathcal{T}} \mathtt{I}r$ because $=_\beta \subseteq \mathcal{T}$.

By the context closure and $\mathtt{I} =_{\mathcal{T}} H\langle u\rangle$, we obtain $\mathtt{I}r =_{\mathcal{T}} H\langle u\rangle r$.

u is h-diverging, thus unsolvable $\Rightarrow$ $H\langle u\rangle$ and $H\langle u\rangle r$ h-diverging.

$\mathcal{T}$ h-collapsing $\Rightarrow$ $H\langle u\rangle r =_{\mathcal{T}} H\langle u\rangle$.

Summing up, $r =_{\mathcal{T}} \mathtt{I}r =_{\mathcal{T}} H\langle u\rangle r =_{\mathcal{T}} H\langle u\rangle$ for all r. $\square$

Corollary 1

Corollary

If $\mathcal{T}$ is a consistent head-collapsing λ-theory then $\mathcal{T} \subseteq \mathcal{H}^*$.

Proof.

Let t and u be such that $t =_{\mathcal{T}} u$. Suppose that $t \neq_{\mathcal{H}^*} u$.

$\Rightarrow$ $C\langle t \rangle$ is $\rightarrow_h$-normalizing and $C\langle u \rangle$ is $\rightarrow_h$-diverging for some C.

By compatibility of $\mathcal{T}$, $C\langle t \rangle =_{\mathcal{T}} C\langle u \rangle$.

Hence $\mathcal{T}$ is not head-adequate. And $\mathcal{T}$ head-collapsing by hyp.

Key property $\Rightarrow$ $\mathcal{T}$ is inconsistent. Absurd! □

Corollary 2

Corollary

$\mathcal{H}^*$ is a maximal consistent λ-theory, and the unique such one that is head-collapsing.

Proof.

Immediate consequence of corollary 1 and the fact that $\mathcal{H}^*$ is head-collapsing. □

Corollary 3

Corollary

$\mathcal{B} \subseteq \mathcal{H}^*$.

Proof.

We know that $\mathcal{H} \subseteq \mathcal{B}$, that is, $\mathcal{B}$ is head-collapsing.

$\mathcal{B}$ is consistent, since Ω and $\mathtt{I}$ have different Böhm trees.

By corollary 2, $\mathcal{B} \subseteq \mathcal{H}^*$. □

Models and $\mathcal{H}^*$

Further property:

$\mathcal{H}^*$ is the λ-theory of the model D_∞.

D_∞ is the first discovered model of the λ-calculus, by Dana Scott (1970s).

Other Contextual Equivalence

Contextual equivalence can be adapted to other strategies.

Weak head and leftmost.

Giving consistent λ-theories.

The leftmost case $=_C^{lo}$ is interesting.

Strong divergence gives a consistent theory, interactively.

leftmost divergence $=$ strong divergence, by the normalization theorem.

THANKS!