Higher order differentiation and Taylor expansion

Thomas Ehrhard and Laurent Regnier

Université Paris 7 – PPS and Université d'Aix-Marseille – IML

Linearity

- linearity = additivity: $f(x+y) = f(x) + f(y)$
  $\rightarrow$ differential calculus (first order)

Linearity

- linearity = additivity: $f(x+y) = f(x) + f(y)$
  $\rightarrow$ differential calculus (first order)
- linearity = consumption of resource (e.g. automaton)
  $\rightarrow$ linear logic (higher order)

Linearity

- linearity = additivity: $f(x+y) = f(x) + f(y)$
  $\rightarrow$ differential calculus (first order)
- linearity = consumption of resource (e.g. automaton)
  $\rightarrow$ linear logic (higher order)

Needs an operational semantics for counting the uses of resources

Operational semantics

CBV

$$\bar{2}\,(\mathrm{I}_1\,\mathrm{I}_2)\,0 \leadsto \bar{2}\,\mathrm{I}_2\,0 \leadsto \mathrm{I}_2\,(\mathrm{I}_2\,0) \leadsto \mathrm{I}_2\,0 \leadsto 0$$

I_1 used once

Operational semantics

CBV

$$\bar{2}\,(\mathrm{I}_1\,\mathrm{I}_2)\,0 \rightsquigarrow \bar{2}\,\mathrm{I}_2\,0 \rightsquigarrow \mathrm{I}_2\,(\mathrm{I}_2\,0) \rightsquigarrow \mathrm{I}_2\,0 \rightsquigarrow 0$$

I_1 used once

CBN

$$\bar{2}\,(\mathrm{I}_1\,\mathrm{I}_2)\,0 \rightsquigarrow (\mathrm{I}_1\,\mathrm{I}_2)\,((\mathrm{I}_1\,\mathrm{I}_2)\,0) \rightsquigarrow \mathrm{I}_2\,((\mathrm{I}_1\,\mathrm{I}_2)\,0) \rightsquigarrow (\mathrm{I}_1\,\mathrm{I}_2)\,0 \rightsquigarrow \mathrm{I}_2\,0 \rightsquigarrow 0$$

I_1 used twice

Operational semantics

CBV

$$\bar{2}\,(\mathrm{I}_1\,\mathrm{I}_2)\,0 \rightsquigarrow \bar{2}\,\mathrm{I}_2\,0 \rightsquigarrow \mathrm{I}_2\,(\mathrm{I}_2\,0) \rightsquigarrow \mathrm{I}_2\,0 \rightsquigarrow 0$$

I_1 used once

CBN

$$\bar{2}\,(\mathrm{I}_1\,\mathrm{I}_2)\,0 \rightsquigarrow (\mathrm{I}_1\,\mathrm{I}_2)\,((\mathrm{I}_1\,\mathrm{I}_2)\,0) \rightsquigarrow \mathrm{I}_2\,((\mathrm{I}_1\,\mathrm{I}_2)\,0) \rightsquigarrow (\mathrm{I}_1\,\mathrm{I}_2)\,0 \rightsquigarrow \mathrm{I}_2\,0 \rightsquigarrow 0$$

I_1 used twice

We will stick to CBN (e.g., Krivine's machine)

Sum as non deterministic choice

$$X_1 + X_2 \rightsquigarrow X_i$$
$$P[X_1 + X_2] \sim P[X_i]$$

Sum as non deterministic choice

$$X_1 + X_2 \rightsquigarrow X_i$$
$$P[X_1 + X_2] \sim P[X_i]$$

P linear: has one (linear) occurence of x

$$P\{(X_1 + X_2)/x\} = P[X_1 + X_2] \sim P[X_i]$$
$$P\{X_1/x\} + P\{X_2/x\} \sim P\{X_i/x\} = P[X_i]$$

Sum as non deterministic choice

$$X_1 + X_2 \rightsquigarrow X_i$$
$$P[X_1 + X_2] \sim P[X_i]$$

P linear: has one (linear) occurence of x

$$P\{(X_1 + X_2)/x\} = P[X_1 + X_2] \sim P[X_i]$$
$$P\{X_1/x\} + P\{X_2/x\} \sim P\{X_i/x\} = P[X_i]$$

P bilinear: has 2 (linear) occurrences of x

$$P\{(X_1 + X_2)/x\} \sim P[X_1 + X_2, X_1 + X_2] \sim P[X_i, X_j]$$
$$P\{X_1/x\} + P\{X_2/x\} \sim P\{X_i/x\} = P[X_i, X_i]$$

Sum as non deterministic choice

Remark

Can keep sums without commiting choices, e.g. instead of

$$X_1 + X_2 \rightsquigarrow X_i$$
$$P[X_1 + X_2] \sim P[X_i]$$

have

$$P[X_1 + X_2] \sim P[X_1] + P[X_2]$$

Linearity in lambda-calculus

Lambda-calculus application is not linear:

$$(P_1 + P_2)Q = P_1 Q + P_2 Q$$

but

$$P(Q_1 + Q_2) \neq PQ_1 + PQ_2$$

(e.g. composition $f \circ g$ in an additive category is only left linear)

Differentiation

- differentiation = approximation by linear map

Differentiation

- differentiation = approximation by linear map
- differentiation = forcing an input to be linear

Differentiation

- differentiation = approximation by linear map
- differentiation = forcing an input to be linear

Linear substitution

$$\frac{\partial P}{\partial x} \cdot Q = \text{“}P \text{ in which } Q \text{ is substituted \textit{linearly} for } x\text{”}$$

Differentiation

- differentiation = approximation by linear map
- differentiation = forcing an input to be linear

Linear substitution

$$\frac{\partial P}{\partial x} \cdot Q = \text{“}P \text{ in which } Q \text{ is substituted } \textit{linearly} \text{ for } x\text{”}$$

Example:

$$\text{if } P = x^2 = x.x \text{ then}$$
$$\frac{\partial P}{\partial x} \cdot Q = Q.x + x.Q \quad (= 2x.Q)$$

Differential lambda-calculus

Syntax

$$T = x \mid \lambda x T \mid T_1\, T_2 \mid 0 \mid T_1 + T_2 \mid \mathrm{D}\, T_1 \cdot T_2$$

Associativity, commutativity

Linearity

$$\lambda x\,(P_1 + P_2) = \lambda x\, P_1 + \lambda x\, P_2$$
$$(P_1 + P_2)Q = P_1\, Q + P_2\, Q$$
$$\mathrm{D}(P_1 + P_2) \cdot Q = \mathrm{D} P_1 \cdot Q + \mathrm{D} P_2 \cdot Q$$
$$\mathrm{D} P \cdot (Q_1 + Q_2) = \mathrm{D} P \cdot Q_1 + \mathrm{D} P \cdot Q_2$$

Notation

- Differential calculus: if $f : \mathbb{R}^n \mapsto \mathbb{R}$ then $\mathrm{D}f(x) : \mathbb{R}^n \mapsto \mathbb{R}$ linear

$$\mathrm{D}f : \mathbb{R}^n \to (\mathbb{R}^n \multimap \mathbb{R})$$
$$\mathrm{D}f(x).u = \text{diff of } f \text{ at } x \text{ in direction } u$$

Notation

- Differential calculus: if $f : \mathbb{R}^n \mapsto \mathbb{R}$ then $\mathrm{D}f(x) : \mathbb{R}^n \mapsto \mathbb{R}$ linear

$$\mathrm{D}f : \mathbb{R}^n \to (\mathbb{R}^n \multimap \mathbb{R}) \cong \mathbb{R}^n \multimap (\mathbb{R}^n \to \mathbb{R})$$

$$\mathrm{D}f(x).u = \text{diff of } f \text{ at } x \text{ in direction } u$$

Notation

- Differential calculus: if $f : \mathbb{R}^n \mapsto \mathbb{R}$ then $\mathrm{D}f(x) : \mathbb{R}^n \mapsto \mathbb{R}$ linear

$$\mathrm{D}f : \mathbb{R}^n \to (\mathbb{R}^n \multimap \mathbb{R}) \cong \mathbb{R}^n \multimap (\mathbb{R}^n \to \mathbb{R})$$

$$\mathrm{D}f(x).u = \text{diff of } f \text{ at } x \text{ in direction } u$$

- Differential *lambda*-calculus: if $P : E \to F$ then

$$\mathrm{D}P : E \multimap (E \to F)$$

$$(\mathrm{D}P \cdot Q)X = \text{diff of } P \text{ in direction } Q \text{ at } X$$

Notation

- Differential calculus: if $f : \mathbb{R}^n \mapsto \mathbb{R}$ then $\mathrm{D}f(x) : \mathbb{R}^n \mapsto \mathbb{R}$ linear

$$\mathrm{D}f : \mathbb{R}^n \to (\mathbb{R}^n \multimap \mathbb{R}) \cong \mathbb{R}^n \multimap (\mathbb{R}^n \to \mathbb{R})$$

$$\mathrm{D}f(x).u = \text{diff of } f \text{ at } x \text{ in direction } u$$

- Differential *lambda*-calculus: if $P : E \to F$ then

$$\mathrm{D}P : E \multimap (E \to F)$$

$$(\mathrm{D}P \cdot Q)X = \text{diff of } P \text{ in direction } Q \text{ at } X$$

$\rightsquigarrow \mathrm{D}P \cdot Q$ as a linearization of PQ

Partial derivative = linear substitution

$$\frac{\partial x}{\partial x} \cdot Q = Q \qquad\qquad \frac{\partial y}{\partial x} \cdot Q = 0$$

$$\frac{\partial \lambda y\, P}{\partial x} \cdot Q = \lambda y\, \frac{\partial P}{\partial x} \cdot Q \qquad\qquad (x \neq y)$$

Partial derivative = linear substitution

$$\frac{\partial x}{\partial x} \cdot Q = Q \qquad\qquad \frac{\partial y}{\partial x} \cdot Q = 0$$

$$\frac{\partial \lambda y\, P}{\partial x} \cdot Q = \lambda y\, \frac{\partial P}{\partial x} \cdot Q \qquad (x \neq y)$$

$$\frac{\partial \mathrm{D} P_1 \,.\, P_2}{\partial x} \cdot Q =$$

Partial derivative = linear substitution

$$\frac{\partial x}{\partial x} \cdot Q = Q \qquad \frac{\partial y}{\partial x} \cdot Q = 0$$

$$\frac{\partial \lambda y\, P}{\partial x} \cdot Q = \lambda y\, \frac{\partial P}{\partial x} \cdot Q \qquad (x \neq y)$$

$$\frac{\partial \mathrm{D} P_1 \;.\; P_2}{\partial x} \cdot Q = \mathrm{D}\left(\frac{\partial P_1}{\partial x} \cdot Q\right) . \; P_2 + \mathrm{D} P_1 \;.\; \left(\frac{\partial P_2}{\partial x} \cdot Q\right)$$

Partial derivative = linear substitution

$$\frac{\partial x}{\partial x} \cdot Q = Q \qquad\qquad \frac{\partial y}{\partial x} \cdot Q = 0$$

$$\frac{\partial \lambda y\, P}{\partial x} \cdot Q = \lambda y\, \frac{\partial P}{\partial x} \cdot Q \qquad (x \neq y)$$

$$\frac{\partial \mathrm{D} P_1 \,.\, P_2}{\partial x} \cdot Q = \mathrm{D}\left(\frac{\partial P_1}{\partial x} \cdot Q\right) .\, P_2 + \mathrm{D} P_1 \,.\, \left(\frac{\partial P_2}{\partial x} \cdot Q\right)$$

$$\frac{\partial P_1\, P_2}{\partial x} \cdot Q =$$

Partial derivative = linear substitution

$$\frac{\partial x}{\partial x} \cdot Q = Q \qquad\qquad \frac{\partial y}{\partial x} \cdot Q = 0$$

$$\frac{\partial \lambda y\, P}{\partial x} \cdot Q = \lambda y\, \frac{\partial P}{\partial x} \cdot Q \qquad (x \neq y)$$

$$\frac{\partial \mathrm{D} P_1 \,.\, P_2}{\partial x} \cdot Q = \mathrm{D}\left(\frac{\partial P_1}{\partial x} \cdot Q\right) . \, P_2 + \mathrm{D} P_1 \,.\, \left(\frac{\partial P_2}{\partial x} \cdot Q\right)$$

$$\frac{\partial P_1\, P_2}{\partial x} \cdot Q = \left(\frac{\partial P_1}{\partial x} \cdot Q\right) P_2 +$$

Partial derivative = linear substitution

$$\frac{\partial x}{\partial x} \cdot Q = Q \qquad\qquad \frac{\partial y}{\partial x} \cdot Q = 0$$

$$\frac{\partial \lambda y\, P}{\partial x} \cdot Q = \lambda y\, \frac{\partial P}{\partial x} \cdot Q \qquad\qquad (x \neq y)$$

$$\frac{\partial \mathrm{D} P_1 \,.\, P_2}{\partial x} \cdot Q = \mathrm{D}\left(\frac{\partial P_1}{\partial x} \cdot Q\right) .\, P_2 + \mathrm{D} P_1 \,.\, \left(\frac{\partial P_2}{\partial x} \cdot Q\right)$$

$$\frac{\partial P_1\, P_2}{\partial x} \cdot Q = \left(\frac{\partial P_1}{\partial x} \cdot Q\right) P_2 + P_1 \left(\frac{\partial P_2}{\partial x} \cdot Q\right)$$

Partial derivative = linear substitution

$$\frac{\partial x}{\partial x} \cdot Q = Q \qquad\qquad \frac{\partial y}{\partial x} \cdot Q = 0$$

$$\frac{\partial \lambda y\, P}{\partial x} \cdot Q = \lambda y\, \frac{\partial P}{\partial x} \cdot Q \qquad (x \neq y)$$

$$\frac{\partial \mathrm{D} P_1 \mathbin{.} P_2}{\partial x} \cdot Q = \mathrm{D}\left(\frac{\partial P_1}{\partial x} \cdot Q\right) \mathbin{.} P_2 + \mathrm{D} P_1 \mathbin{.} \left(\frac{\partial P_2}{\partial x} \cdot Q\right)$$

$$\frac{\partial P_1\, P_2}{\partial x} \cdot Q = \left(\frac{\partial P_1}{\partial x} \cdot Q\right) P_2 + \left(\mathrm{D} P_1 \mathbin{.} \left(\frac{\partial P_2}{\partial x} \cdot Q\right)\right) P_2$$

The chain rule

If $F : \mathbb{R}^p \mapsto \mathbb{R}$ and $U : \mathbb{R}^n \mapsto \mathbb{R}^p$ are differentiable then (with standard notations):

$$\frac{\partial F \circ U}{\partial x_i}(\vec{x}) = \mathrm{D}F(U(\vec{x})) \, . \, \frac{\partial U}{\partial x_i}(\vec{x})$$

If P_1 doesn't depend on x so that $\frac{\partial P_1}{\partial x} \cdot Q = 0$ then (with l-diff notations):

$$\frac{\partial P_1 P_2}{\partial x} \cdot Q = \left(\mathrm{D}P_1 \, . \, \left(\frac{\partial P_2}{\partial x} \cdot Q \right) \right) P_2$$

Dynamics

Reduction

$$(\lambda x\, P)Q \rightsquigarrow P\{Q/x\}$$

$$\mathrm{D}\lambda x\, P \,.\, Q \rightsquigarrow \lambda x\, \frac{\partial P}{\partial x} \cdot Q$$

Taylor expansion

Differential calculus

$$f(x) = \sum_{n \geq 0} \frac{1}{n!} \mathrm{D}^n f(0) \, . \, x^n$$

Taylor expansion

Differential calculus

$$f(x) = \sum_{n \geq 0} \frac{1}{n!} \mathrm{D}^n f(0) \, . \, x^n$$

Differential *lambda*-calculus

$$PQ = \sum_{n \geq 0} \frac{1}{n!} \left(\mathrm{D}^n P . Q^n \right) 0$$

Taylor expansion

Differential calculus

$$f(x) = \sum_{n \geq 0} \frac{1}{n!} \mathrm{D}^n f(0) \, . \, x^n$$

Differential *lambda*-calculus

$$PQ = \sum_{n \geq 0} \frac{1}{n!} \left(\mathrm{D}^n P . Q^n\right) 0$$

$$P\{Q/x\} = \sum_{n \geq 0} \frac{1}{n!} \left(\frac{\partial^n P}{\partial x^n} \cdot Q^n \right) \{0/x\}$$

Taylor expansion

Differential calculus

$$f(x) = \sum_{n \geq 0} \frac{1}{n!} \mathrm{D}^n f(0) \,.\, x^n$$

Differential *lambda*-calculus

$$PQ = \sum_{n \geq 0} \frac{1}{n!} \left(\mathrm{D}^n P . Q^n\right) 0$$

$$P\{Q/x\} = \sum_{n \geq 0} \frac{1}{n!} \left(\frac{\partial^n P}{\partial x^n} \cdot Q^n\right) \{0/x\}$$

And now let's move to the *full Taylor expansion*

Resource calculus

Syntax

$$s = x \mid \lambda x\, s \mid \langle s \rangle\, B$$
$$B = 1 \mid s_1 \ldots s_n$$

$+$ linearity conditions

Resource calculus

Syntax

$$s = x \mid \lambda x\, s \mid \langle s \rangle\, B$$
$$B = 1 \mid s_1 \ldots s_n$$

+ linearity conditions

Remark

$$\langle s \rangle\, B = (\mathrm{D}^n s \,.\, B)0 \quad \text{where } \deg(B) = n$$

Linear substitution

$$\frac{\partial x}{\partial x} \cdot t = t \qquad \frac{\partial y}{\partial x} \cdot t = 0$$

$$\frac{\partial \lambda y\, s}{\partial x} \cdot t = \lambda y\, \frac{\partial s}{\partial x} \cdot t$$

$$\frac{\partial \langle s \rangle B}{\partial x} \cdot t = \left\langle \frac{\partial s}{\partial x} \cdot t \right\rangle B + \langle s \rangle \frac{\partial B}{\partial x} \cdot t$$

$$\frac{\partial s_1 \ldots s_n}{\partial x} \cdot t = \sum_{i=1}^{n} s_1 \ldots \left(\frac{\partial s_i}{\partial x} \cdot t \right) \ldots s_n$$

Reduction

$$\langle \lambda x\, s \rangle\, s_1 \dots s_n \rightsquigarrow 0 \qquad \text{if } \deg_x(s) \neq n$$

$$\rightsquigarrow \frac{\partial^n s}{\partial x^n} \cdot s_1 \dots s_n = \sum_{\sigma \in \mathfrak{S}_n} s[s_{\sigma 1} \dots s_{\sigma n}]$$

Uniformity

Uniform resource term = lambda-term approximant

Approximations

$$
\begin{aligned}
\lambda x\, x\, x &\rightsquigarrow \lambda x\ \langle x \rangle\, x^n \\
\lambda f\, f\, (\lambda g\, g \lambda z\, z) &\rightsquigarrow \lambda f\ \langle f \rangle\, 1 \\
&\rightsquigarrow \lambda f\ \langle f \rangle\, \lambda g\, \langle g \rangle\, 1 \\
&\rightsquigarrow \lambda f\ \langle f \rangle\, \lambda g\, \langle g \rangle\, 1\ .\ \lambda g\, \langle g \rangle\, \lambda z z \\
&\rightsquigarrow \lambda f\ \langle f \rangle\, \lambda g\, \langle g \rangle\, 1\ .\ \lambda g\, \langle g \rangle\, \lambda z\, z\ .\ \lambda g\, \langle g \rangle\, (\lambda z\, z)^2
\end{aligned}
$$

Reduction vs beta-reduction

Theorem

If $S \rightarrow_\beta S_0$ *(normal form) and* $s_0 \in \mathcal{T}(S_0)$ *then there is a unique* $s \in \mathcal{T}(S)$ *s.t.* $s \rightsquigarrow s_0$.

$$\begin{array}{ccc} S & \xrightarrow{\quad * \quad} & s \in \mathcal{T}(S) \\ {\scriptstyle\beta}\big\downarrow & & \big\downarrow \\ S_0 & \xrightarrow[\quad * \quad]{} & s_0 \in \mathcal{T}(S_0) \end{array}$$

Full Taylor expansion

Let S be a lambda-term;

$$S^* = \sum_{s \in \mathcal{T}(S)} \frac{1}{m(s)} s$$

$m(s)$ is the *multiplicity coefficient* of s:

Full Taylor expansion

Let S be a lambda-term;

$$S^* = \sum_{s \in \mathcal{T}(S)} \frac{1}{m(s)} s$$

$m(s)$ is the *multiplicity coefficient* of s:

Theorem

$$(PQ)^* = \sum_{n \geq 0} \frac{1}{n!} \langle P^* \rangle \, Q^{*\,n}$$