Coherent differentiation

DiλLL meeting, CIRM

May 16, 2024

Thomas Ehrhard
IRIF, CNRS, Inria and Université Paris Cité

Linear Logic

LL: most (all?) denotational models have an underlying structure similar to *linear algebra*

- tensor product
- linear function spaces
- direct product and coproduct
- duality.

There are also non linear morphisms.

Exponential storage modality: connects the linear and non-linear categories.

Dereliction: we can forget that a function is linear.

Differentiation in LL

Introduces a converse operation.

- dereliction: forget linearity of a morphism
  linear $\rightsquigarrow$ non-linear
- differentiation: best linear approximation of a morphism
  non-linear $\rightsquigarrow$ linear

reformulating logically the standard laws of the differential calculus.

$\rightsquigarrow$ the differential λ-calculus:

$$\frac{\Gamma \vdash M : A \Rightarrow B \qquad \Gamma \vdash N : A}{\Gamma \vdash \mathrm{D}M \cdot N : A \Rightarrow B}$$

And $\mathrm{D}M \cdot N$ is linear in N (and also in M).

Intuition

The derivative of M should be $M' : A \Rightarrow (A \multimap B)$. Then intuitively

$$\mathrm{D}M \cdot N = \lambda x : A \cdot (M'\,x)(N)$$

Differential reduction

To define the operational semantics of the differential λ-calculus we define by induction on M a term $\frac{\partial M}{\partial x} \cdot N$ with

$$\Gamma, x : A \vdash M : B \text{ and } \Gamma \vdash N : A \quad \Rightarrow \quad \Gamma, x : A \vdash \frac{\partial M}{\partial x} \cdot N : B$$

Differential reduction:

$$\mathrm{D}(\lambda x : A \cdot M) \cdot N \to \lambda x : A \cdot \left(\frac{\partial M}{\partial x} \cdot N \right)$$

similar to β-reduction

$$(\lambda x : A \cdot M) N \to M \left[N/x \right] .$$

Strong non-determinism of DiLL

In the inductive definition:

$$\frac{\partial (M)P}{\partial x} \cdot N = \left(\frac{\partial M}{\partial x} \cdot N\right) P + \left(\mathsf{D} M \cdot \left(\frac{\partial P}{\partial x} \cdot N\right)\right) P$$

combines Leibniz and chain rule.

Requires apparently a deduction rule

$$\frac{\Gamma \vdash M : A \qquad \Gamma \vdash N : A}{\Gamma \vdash M + N : A}\ (+)$$

Sums come in even if not officially invited

$$M = \lambda x : \mathsf{bool} \cdot (x \wedge \neg x) : \mathsf{bool} \to \mathsf{bool}$$

$$\mathsf{D}M \cdot u = \lambda x : \mathsf{bool} \cdot (u \wedge \neg x + x \wedge \neg u)$$

$$(\mathsf{D}(\mathsf{D}M \cdot u) \cdot v)P = u \wedge \neg v + v \wedge \neg u$$

$$(\mathsf{D}(\mathsf{D}M \cdot \mathsf{true}) \cdot \mathsf{false})P = \mathsf{true} + \mathsf{false}$$

$\rightsquigarrow$ models of DiLL are additive categories

Leibniz $\rightsquigarrow$ the models $\mathcal{L}$ of DiLL are *additive* categories:

- $\mathcal{L}(X, Y)$ is a commutative monoid (with additive notations) for each objects X, Y of $\mathcal{L}$
- morphism composition is bilinear.

Remark

If $\mathcal{L}$ is cartesian and additive then the cartesian product is also a coproduct, the terminal object is initial: $\& = \oplus$.

$\rightsquigarrow$ some LL degeneracy = non-determinism.

But...

... many interesting models of LL *are not* additive categories.

Remark

One of the main new ideas brought by LL is that the linear/non-linear dichotomy does not require additivity.

In *probabilistic coherence spaces* (**Pcoh**) non-linear morphisms are obviously differentiable: they are analytic functions,

and **Pcoh** is not an additive category.

Question

Analytic functions have derivatives: what is the status of derivatives in such models?

At first sight they seem to live outside...

Examples

In **Pcoh** the type 1 (= unit) of LL is interpreted as

$$[0,1] \subseteq \mathbb{R}$$

A non-linear morphism, that is, an element of $\mathbf{Pcoh}_!(1,1)$, is an analytic function

$$\begin{aligned} f : [0,1] &\to [0,1] \\ x &\mapsto \sum_{n=0}^{\infty} a_n x^n \end{aligned}$$

for a (uniquely determined) sequence $(a_n)_{n\in\mathbb{N}}$ of elements of $\mathbb{R}_{\geq 0}$ such that $\sum_{n\in\mathbb{N}} a_n \leq 1$.

Problem

$f'(x) = \sum_{n=0}^{\infty}(n+1)a_{n+1}x^n$ no reason to have $f' \in \mathbf{Pcoh}_!(1,1)$.

For instance: f defined by $f(u) = 1 - \sqrt{1 - u^2}$ belongs to $\mathbf{Pcoh}_!(1, 1)$ but

$$f'(u) = \frac{u}{\sqrt{1 - u}}$$

is not even defined on the whole of $[0, 1]$ and is not bounded on $[0, 1)$. . . and we cannot reject f because it is the interpretation of a program: $f = [\![M]\!]$ where (in a CBN programming language)

$$\begin{aligned}\texttt{let rec } & M\,(x : \mathsf{unit}) : \mathsf{unit} \\ & = \texttt{if } \mathsf{bernouilli}(\frac{1}{2}) \texttt{ then } (x; x) \texttt{ else } (M\,x; M\,x)\end{aligned}$$

$$f(u) = \frac{1}{2}u^2 + \frac{1}{2}f(u)^2 \quad \text{i.e.} \quad f(u)^2 - 2f(u) + u^2 = 0$$

Other example: Von Neumann rectifier

In a CBN functional language:

$$\begin{aligned}\texttt{let rec}\ M\,(x : \texttt{bool}) : \texttt{bool} \\ = \texttt{if}\ x\ \texttt{then}\ \texttt{if x then}\ M\,x \\ \texttt{else}\ \mathsf{true} \\ \texttt{else if x then}\ \mathsf{false} \\ \texttt{else}\ M\,x\end{aligned}$$

In LL, bool $= 1 \oplus 1$.

In probabilistic coherence spaces,

$$[\![\mathsf{bool}]\!] = 1 \oplus 1$$

$|1 \oplus 1| = \{1, 2\}$ and $\mathsf{P}(1 \oplus 1) = \{u \in \mathbb{R}^2_{\geq 0} \mid u_1 + u_2 \leq 1\}$.
Then $[\![M]\!] = f : \mathsf{P}(1 \oplus 1) \to \mathsf{P}(1 \oplus 1)$ such that

$$\begin{aligned} f(u) &= u_1^2 f(u) + u_1 u_2 e_1 + u_2 u_1 e_2 + u_1^2 f(u) \\ f(u_0, 0) &= f(0, u_1) = (0, 0) \quad \text{by minimality} \end{aligned}$$

so if $u_1 + u_2 = 1$ and $u_1 u_2 \neq 0$, $u_1^2 + u_2^2 + 2u_1 u_2 = 1$

$$f(u) = \frac{1}{1 - u_1^2 - u_2^2}(u_1 u_2, u_2 u_1) = (1/2, 1/2)$$

whence the name: it rectifies any (non completely) biased coin!

We have $f(u) = \frac{1}{2}(g(u), g(u))$ where

$$g(u) = \frac{2u_1 u_2}{1 - u_1^2 - u_2^2}$$

so that

$$g'(u) \cdot v = 2\,\frac{(1 + u_1^2 - u_2^2)u_2 v_1 + (1 + u_2^2 - u_1^2)u_1 v_2}{(1 - u_1^2 - u_2^2)^2}$$

and one can check that

$$u + v \in \mathrm{P}(1 \oplus 1) \Rightarrow f'(u) \cdot v \in \mathrm{P}(1 \oplus 1)$$

Key observation, part 1

If $f \in \mathbf{Pcoh}_!(1 \oplus 1, 1 \oplus 1)$ and

$$u, v \in \mathrm{P}(1 \oplus 1) \text{ satisfy } u + v \in \mathrm{P}(1 \oplus 1)$$

then by the Taylor formula at x

$$f(u+v) = f(u) + f'(u) \cdot v + \frac{1}{2} f''(u) \cdot (v, v) + \cdots \in \mathrm{P}(1 \oplus 1)$$

and all the partial derivatives involved have only ≥ 0 coefficients in the monomials $u_1^p u_2^q$, so we have

$$f(u) + f'(u) \cdot v \in \mathrm{P}(1 \oplus 1)\,.$$

So if we set $S = \{(u,v) \in \mathsf{P}(1 \oplus 1)^2 \mid u + v \in \mathsf{P}(1 \oplus 1)\}$ we can define

$$\begin{aligned} \mathsf{T}f : S &\to S \\ (u,v) &\mapsto (f(u), f'(u) \cdot v) \end{aligned}$$

formally similar to a "tangent bundle functor".

Key observation, part 2

S can be seen as an object of **Pcoh** and

$$\forall f \in \mathbf{Pcoh}_!(1 \oplus 1, 1 \oplus 1) \quad \mathsf{T}f \in \mathbf{Pcoh}_!(S, S)$$

Can be extended to *all the objects* of **Pcoh**.

$\rightsquigarrow$ Coherent differentiation

Coherent differentiation can be presented in 2 settings, very similar to those of JS Lemay's talk:

- monoidal categories with a storage modality
- general (cartesian) categories.

I will focus on a special case of the first setting which can be used in most concrete situations we know.

An important case of coherent differentiation: the elementary situation

Summability in a linear category

Assume:

- $\mathcal{L}$ is a SMCC, tensor $X \otimes Y$, tensor unit 1, internal hom $\mathcal{L}(Z \otimes X, Y) \simeq \mathcal{L}(Z, X \multimap Y)$;
- $\mathcal{L}$ has 0-morphisms $0 \in \mathcal{L}(X, Y)$ with $0\,t = 0$ and $t\,0 = 0$;
- $\mathcal{L}$ is cartesian, cartesian product $\&_{i \in I} X_i$ with projections pr_i and if $(t_i \in \mathcal{L}(Y, X_i))_{i \in I}$ then $\langle t_i \rangle_{i \in I} \in \mathcal{L}(Y, \&_{i \in I} X_i)$.

We don't assume $\mathcal{L}$ to be additive.

Remark (some sums do always exist)

$\langle \mathsf{Id}_1, 0 \rangle, \langle 0, \mathsf{Id}_1 \rangle \in \mathcal{L}(1, 1 \,\&\, 1)$
have a sum $\langle \mathsf{Id}_1, 0 \rangle + \langle 0, \mathsf{Id}_1 \rangle = \langle \mathsf{Id}_1, \mathsf{Id}_1 \rangle \in \mathcal{L}(1, 1 \,\&\, 1)$.

Main idea of CD: this restricted kind of sum is sufficient for differentiation.

The functor of summable pairs

$\mathbf{S} : \mathcal{L} \to \mathcal{L}$, "representable functor" given by

$$\mathbf{S}X = (\mathbb{D} \multimap X) \quad \text{where} \quad \mathbb{D} = 1 \,\&\, 1$$

Intuition

A "point" of $\mathbf{S}X$ is a pair (u_0, u_1) of two points of X whose sum $u_0 + u_1$ is well defined.

- $\pi_0 = (\overline{\pi}_0 \multimap X) \in \mathcal{L}(\mathbf{S}X, X)$ fst component of pairs
- $\pi_1 = (\overline{\pi}_1 \multimap X) \in \mathcal{L}(\mathbf{S}X, X)$ snd component of pairs
- $\sigma = (\Delta \multimap X) \in \mathcal{L}(\mathbf{S}X, X)$ sum of pairs.

where

$\overline{\pi}_0 = \langle \mathsf{Id}_1, 0 \rangle,\ \overline{\pi}_1 = \langle 0, \mathsf{Id}_1 \rangle,\ \Delta = \langle \mathsf{Id}_1, \mathsf{Id}_1 \rangle \in \mathcal{L}(1, \mathbb{D} = 1 \,\&\, 1)$

NB: to simplify the presentation, $X \otimes 1 = (1 \multimap X) = X$.

We need $\overline{\pi}_0, \overline{\pi}_1 \in \mathcal{L}(1, \mathbb{D})$ to be jointly epic so that $\pi_0, \pi_1 \in \mathcal{L}(\mathbf{S}X, X)$ are jointly monic: this is a property of $\mathcal{L}$ which holds very often.

Definition (summability and sum of morphisms)

$f_0, f_1 \in \mathcal{L}(Y, X)$ are summable if there is $h \in \mathcal{L}(Y, \mathbf{S}X)$ such that $\pi_i \ h = f_i$ $(i = 0, 1)$.

This h is unique: $\langle\!\langle f_0, f_1 \rangle\!\rangle = h$ (witness of summability).

$f_0 + f_1 = \sigma \langle\!\langle f_0, f_1 \rangle\!\rangle$.

In other words: there is $h \in \mathcal{L}(X \otimes \mathbb{D}, Y)$ such that

$$f_i = h\,(X \otimes \overline{\pi}_i) \quad \text{for} \quad i = 0, 1\,.$$

Witness Property

We also assume that if $h_1, h_2 \in \mathcal{L}(Y, \mathbf{S}X)$ and the associated sums $\sigma\, h_1, \sigma\, h_2 \in \mathcal{L}(X, Y)$ are summable, then h_1 and h_2 are summable.

Fact

Equipped with 0 *and* $+$*, each* $\mathcal{L}(X, Y)$ *is a commutative partial monoid.*

Composition is compatible with this structure.

Partial commutative monoids

Remark

The kind of commutative partial monoid we obtain here is the following:

- if $x + y$ is defined and $(x + y) + z$ are defined, then so are $y + z$ and $x + (y + z)$, and $(x + y) + z = x + (y + z)$
- if $x + y$ is defined then so is $y + x$, and $x + y = y + x$.

$\{0, 1\}$ (with usual addition of $\mathbb{N}$) is such a partial monoid, but not $\{-1, 0, 1\}$: $(-1 + 1) + 1$ is defined but $1 + 1$ is not.

Such partial commutative monoids feature some kind of positivity.

There are also more liberal definitions of commutative partial monoids.

Summability depends on $\mathcal{L}$

In each such SMCC $\mathcal{L}$, the functor **S** tells us which summations are possible in the homsets of $\mathcal{L}$ and tells us how to compute the sums.

- If $\mathcal{L}$ has biproducts, then $\mathbb{D} = 1 \,\&\, 1 = 1 \oplus 1$ and hence $\mathbf{S}X = X \,\&\, X$: all sums are allowed (already well known).
- In **Pcoh**: $\mathsf{P}(\mathbf{S}X) \simeq \{(u, v) \in \mathsf{P}X^2 \mid u + v \in \mathsf{P}X\}$ where $u + v$ is the pointwise sum of $u, v \in (\mathbb{R}_{\geq 0})^{|X|}$.
- In **Coh** (Girard's coherence spaces): $\mathsf{Cl}(\mathbf{S}X) = \{(u, v) \in \mathsf{Cl}(X)^2 \mid u \cap v = \emptyset \text{ and } u \cup v \in \mathsf{Cl}(X)\}$.

Another example comes later. . .

Comonoid structure of $\mathbb{D} = 1 \,\&\, 1$

When $\mathcal{L}$ satisfies these conditions, $\mathbb{D} = 1 \,\&\, 1$ has a structure of commutative comonoid

$$\mathsf{pr}_0 : \mathbb{D} \to 1 \quad \textsf{fst projection of } \&$$
$$\widetilde{\mathsf{L}} : \mathbb{D} \to \mathbb{D} \otimes \mathbb{D}$$

such that

$$\widetilde{\mathsf{L}}\, \overline{\pi}_0 = \overline{\pi}_0 \otimes \overline{\pi}_0$$
$$\widetilde{\mathsf{L}}\, \overline{\pi}_1 = \overline{\pi}_0 \otimes \overline{\pi}_1 + \overline{\pi}_1 \otimes \overline{\pi}_1$$

NB: this sum is well defined (by the Witness Property).

Since $\overline{\pi}_0$ and $\overline{\pi}_1$ are jointly epic, these equations fully characterize $\widetilde{\mathsf{L}}$.

Monad structure of $\mathbf{S}$

This comonoid structure of $\mathbb{D} = 1 \mathbin{\&} 1$ induces a monad structure on the functor $\mathbf{S} = (\mathbb{D} \multimap \_)$.

- Unit: $\iota_0 = (\mathsf{pr}_0 \multimap X) \in \mathcal{L}(X, \mathbf{S}X)$
- Multiplication: $\tau = \left(\widetilde{\mathsf{L}} \multimap X\right) \in \mathcal{L}(\mathbf{S}^2 X, \mathbf{S}X)$

Formally completely similar to the monad structure of the tangent functor in tangent categories.

Intuitively $\iota_0(u) = (u, 0)$ and
$\tau((u_{00}, u_{01}), (u_{10}, u_{11})) = (u_{00}, u_{01} + u_{10})$

This monad has also a tensorial strength and is commutative.

So far, nothing about differentiation.

As in differential categories, we need a storage modality to speak about differentiation.

Exponential

Assume that $\mathcal{L}$ is equipped with a storage modality, that is

- a comonad $(!, \mathsf{der}, \mathsf{dig})$
- with a strong symmetric monoidal structure $(\mathcal{L}, \&) \to (\mathcal{L}, \otimes)$: there are well-behaved isos $1 \to !\top$ and $!X \otimes !Y \to !(X \;\&\; Y)$, the Seely isos.

Then the Kleisli category $\mathcal{L}_!$ is intuitively the category of non-linear (smooth, analytic. . .) morphisms that we will differentiate.

- $\mathsf{Obj}(\mathcal{L}_!) = \mathsf{Obj}(\mathcal{L})$
- $\mathcal{L}_!(X, Y) = \mathcal{L}(!X, Y)$.

Elementary differential structure

Definition

An elementary differential structure on $\mathcal{L}$ is a !-coalgebra structure $\widetilde{\partial}$ on $\mathbb{D} = 1 \,\&\, 1$:

$$\widetilde{\partial} : \mathbb{D} \to !\mathbb{D}$$

such that pr_0 and $\widetilde{\mathsf{L}}$ are coalgebra morphisms, using the fact that 1 and $\mathbb{D} \otimes \mathbb{D}$ are canonically !-coalgebras.

Remark (CD is everywhere. . .)

If $(\mathcal{L}, !\_)$ is a Lafont category (ie. $!\_$ is the cofree symmetric comonoid functor) there is exactly one elementary differential structure, induced by $(\mathsf{pr}_0, \widetilde{\mathsf{L}})$.

Generalizes a result mentioned by JS Lemay (for differential categories).

What is the link with differentiation?

Using $\widetilde{\partial}$ we can define a natural transformation
$\partial_X : !\mathbf{S}X = !(\mathbb{D} \multimap X) \ \to \ \mathbf{S}!X = (\mathbb{D} \multimap !X)$,

Curry transpose of

$$\begin{array}{c} !(\mathbb{D} \multimap X) \otimes \mathbb{D} \\ \big\downarrow \mathsf{Id} \otimes \widetilde{\partial} \\ !(\mathbb{D} \multimap X) \otimes !\mathbb{D} \\ \big\downarrow \mu^2 \\ !((\mathbb{D} \multimap X) \otimes \mathbb{D}) \\ \big\downarrow !\mathsf{ev} \\ !X \end{array}$$

μ^2: lax monoidality $\otimes \to \otimes$, derived from the strong monoidality $\& \to \otimes$.   ev: SMCC evaluation morphism.

Fact (extending $\mathbf{S}$ to $\mathcal{L}_!$ thanks to $\partial \rightsquigarrow$ differentiation functor)

$\partial_X : !\mathbf{S}X \to \mathbf{S}!X$ *is a distributive law*

- *between the functor* $\mathbf{S}$ *and the comonad* $!\_$
- *and between the monad* $\mathbf{S}$ *and the functor* $!\_$

satisfying additional properties (Leibniz, Schwarz etc).

If $t \in \mathcal{L}_!(X, Y) = \mathcal{L}(!X, Y)$ *seen as a non-linear morphism* $X \to Y$ *then*

$$\mathbf{T}t = (\mathbf{S}t)\, \partial_X \in \mathcal{L}_!(\mathbf{S}X, \mathbf{S}Y)$$

can be understood intuitively as mapping $(x, u) \in \mathbf{S}X$ *(that is* $x, u \in X$ *summable) to* $(t(x), t'(x) \cdot u) \in \mathbf{S}Y$*, a summable pair.*

$\mathbf{T}$ *is a functor on* $\mathcal{L}_!$ *(chain rule): the extension of* $\mathbf{S}$ *to* $\mathcal{L}_!$ *and the monad structure of* $\mathbf{S}$ *is preserved by this extension (derivatives commute with sums).*

The simplest non-trivial example I know

Semi-coherence spaces (SCS)

A simplified version of Girard's coherence spaces implicitely present in F. Lamarche (1995).

$E = (|E|, \frown_E)$ where $|E|$ is a set (web) and $\frown_E$ is a binary and symmetric relation on $|E|$ (not required to be reflexive nor anti-reflexive).

$\mathrm{Cl}(E) = \{x \subseteq |E| \mid \forall a, a' \in x \; a \frown_E a'\}$.

The dual of E: $|E^\perp| = |E|$ and $a \frown_{E^\perp} b$ if $\neg a \frown_E b$.

Remark

If $a \in |E|$, we must have $a \frown_E a$ of $a \frown_{E^\perp} a$, but we cannot have both.

If $u \in \mathrm{Cl}(E)$ and $u' \in \mathrm{Cl}(E^\perp)$ then $u \cap u' = \emptyset$.

The category **Scoh**

$E \multimap F$ defined by $|E \multimap F| = |E| \times |F|$ and

$$(a, b) \frown_{E \multimap F} (a', b') \quad \text{if} \quad a \frown_E b \Rightarrow a' \frown_F b' \,.$$

Category **Scoh**: objects are the semi-coherence spaces and $\mathbf{Scoh}(E, F) = \mathrm{Cl}(E \multimap F) \subseteq |E| \times |F|$.

Composition: relational composition. Identity: diagonal relation.

The SMC structure of SCS

The webs are defind exactly as in **Rel**.

- $|1| = \{*\}$ with $* \frown_1 *$
- $|E \otimes F| = |E| \times |F|$ and $(a, b) \frown_{E \otimes F} (a', b')$ if $a \frown_E a'$ and $b \frown_F b'$
- SMCC: $\mathbf{Scoh}(G \otimes E, F) \simeq \mathbf{Scoh}(G, E \multimap F)$ trivially maps t to $\{(c, (a, b)) \mid ((c, a), b) \in t\}$.

Cartesian product

- $|\&_{i\in I} E_i| = \bigcup_{i\in I}\{i\} \times |E_i|$
- $(i,a) \frown_{\&_{i\in I} E_i} (j,b)$ if $i = j \Rightarrow a \frown_{E_i} b$.
- So that in particular $\mathrm{Cl}(\&_{i\in I} E_i) \simeq \prod_{i\in I} \mathrm{Cl}(E_i)$.

Fact

$|\mathbb{D} = 1 \,\&\, 1| = \{0,1\}$ *with* $i \frown_{\mathbb{D}} j$ *for all* $i,j \in \{0,1\}$, *so that* $\mathrm{Cl}(\mathbb{D}) = \mathcal{P}(\{0,1\})$.

$\overline{\pi}_0 = \langle \mathrm{Id}_1, 0\rangle = \{(*,0)\}$ *and* $\overline{\pi}_1 = \{(*,1)\}$ *are trivially jointly epic.*

$\mathrm{Cl}(\mathbf{S}E) = \mathrm{Cl}(\mathbb{D} \multimap E) \simeq \{(u_0,u_1) \in \mathrm{Cl}(E)^2 \mid u_0 \cup u_1 \in \mathrm{Cl}(E)\}$

$s_0, s_1 \in \mathbf{Scoh}(E,F)$ *are summable iff* $s_0 \cup s_1 \in \mathbf{Scoh}(E,F)$ *and then* $s_0 + s_1 = s_0 \cup s_1$.

Booleans in **Scoh**

Scoh is a model of classical LL: $E^{\perp\perp} = E$.

So **Scoh** has coproducts, in particular $1 \oplus 1 = (1^{\perp} \mathbin{\&} 1^{\perp})^{\perp}$ (notice that $1^{\perp} \neq 1$ contrarily to Girard's CS).

$\mathsf{Cl}(1 \oplus 1) = \{\emptyset, \{0\}, \{1\}\}$ so $\{0\}$ and $\{1\}$ are not summable in $1 \oplus 1$ (though they are summable in $1 \mathbin{\&} 1$): the category **Scoh** is not additive.

Remark (SCS morphisms are not stable)

Contrarily to Girard's CS, SCS accept the *parallel or* program.

Comonoid structure of $\mathbb{D} = 1 \mathbin{\&} 1$

Remember

- $|1| = \{*\}$ and $* \frown_1 *$
- $|\mathbb{D}| = \{0,1\}$ with $i \frown_{1\&1} j$ for all $i,j \in \{0,1\}$, so that $\mathsf{Cl}(\mathbb{D}) = \mathcal{P}(\{0,1\})$.

$\mathbb{D}$ as a comonoid

- counit: $\mathsf{pr}_0 = \{(0,*) \in \mathbf{Scoh}(\mathbb{D},1)\}$
- comultiplication: $\widetilde{\mathsf{L}} \in \mathbf{Scoh}(\mathbb{D}, \mathbb{D} \otimes \mathbb{D})$ given by

$$\widetilde{\mathsf{L}} = \{(0,(0,0))\} \cup \{(1,(0,1)),(1,(1,0))\}$$

The cofree exponential: **Scoh** is Lafont

Much simpler than the exponential of Lamarche who insisted on $|!E| \subseteq \mathcal{P}_{\mathrm{fin}}(|E|)$.

Instead we use finite multisets: $|!E| = \mathcal{M}_{\mathrm{fin}}(|E|)$ and

$$[a_1, \ldots, a_n] \frown_{!E} [b_1, \ldots, b_k] \quad \text{if} \quad \forall i, j \ a_i \frown_E b_j \,.$$

Then $\widetilde{\partial} \in \mathbf{Scoh}(\mathbb{D}, !\mathbb{D})$ is

$$\widetilde{\partial} = \{(i, [i_1, \ldots, i_k]) \mid i, i_1, \ldots, i_k \in \{0, 1\} \text{ and } i = i_1 + \cdots + i_k\}$$

that is

- either $i = 0$ and all the i_j's are $= 0$
- or $i = 1$ and all the i_j's $= 0$ but one which $= 1$.

Induced differentiation

Remember that $\mathbf{S}E = (\mathbb{D} \multimap E)$.

So that $|\mathbf{S}E| = \{0,1\} \times |E|$ and $(i,a) \frown_{\mathbf{S}E} (j,b) \Leftrightarrow a \frown_E b$.

Given $t \in \mathbf{Scoh}(!E, F)$ we get $\mathbf{T}t \in \mathbf{Scoh}(!\mathbf{S}E, \mathbf{S}F)$. Remember that intuitively

$$\mathbf{T}t(u,v) = (t(u), t'(u) \cdot v)\,.$$

Fact

$$\mathbf{T}t = \{([(0,a_1,\dots,(0,a_n)],(0,b)) \mid ([a_1,\dots,a_n],b) \in t\} \\ \cup \{([(0,a_1,\dots,(0,a_n),(1,a)],(1,b)) \mid ([a_1,\dots,a_n,a],b) \in t\}$$

For $(u,v) \in \mathsf{Cl}(\mathbf{S}E)$

$$t'(u) \cdot v = \{b \in |F| \mid \\ \exists a_1,\dots,a_n \in u,\ a \in v\ ([a_1,\dots,a_n,a],b) \in t\} \in \mathsf{Cl}(F)$$

Coherent differential syntax

Try to do something similar to the differential λ-calculus:

- simply-typed λ-calculus
- extended with a differential construct

but without the possibility of freely adding terms.

- The resulting system is more complex
- but it is compatible with the determinism of, say, PCF
- and it is compatible with PCF fixpoints, without restrictions.

Remark (issue with fixpoints and sums)

$$\vdash M = \mathsf{fix}(\lambda x : \mathsf{unit} \cdot (() + x)) : \mathsf{unit}$$

then $M \to () + M \to 2() + M \to \cdots$ and we need ∞ coeffs in the model. Using probabilistic choice instead of $+$ is perfectly ok (cf. **Pcoh**).

Simple types for coherent differential PCF

Types:

- if $d \in \mathbb{N}$ then $\mathsf{T}^d \iota$ is a type;
- if A and B are types then $A \Rightarrow B$ is a type.

If A is a type, we define a type $\mathsf{T}A$:

- $\mathsf{T}\mathsf{T}^d \iota = \mathsf{T}^{d+1} \iota$
- $\mathsf{T}(A \Rightarrow B) = (A \Rightarrow \mathsf{T}B)$.

A type E is sharp if it is not of shape $\mathsf{T}A$, that is $E = (A_1 \Rightarrow \cdots \Rightarrow A_k \Rightarrow \iota)$.

Intuition

A term of type $\mathsf{T}^d E$ denotes a perfect binary "witness-tree" whose leaves are terms of type E, with all branches of length d.

Such a tree is an additive, not a multiplicative construct!

Some (almost standard) term constructions

$d, \nu \in \mathbb{N}$

$$\frac{}{\Gamma \vdash \underline{\nu} : \iota} \qquad \frac{\Gamma \vdash M : \mathsf{T}^d \iota}{\Gamma \vdash \mathsf{succ}^d(M) : \mathsf{T}^{d+1}\iota} \qquad \frac{\Gamma \vdash M : \mathsf{T}^d \iota \qquad \Gamma \vdash P, Q : A}{\Gamma \vdash \mathsf{if}^d(M, P, Q) : \mathsf{T}^d A}$$

$$\frac{\Gamma, x : A \vdash M : B}{\Gamma \vdash \lambda x : A \cdot M : A \Rightarrow B} \qquad \frac{\Gamma \vdash M : A \Rightarrow B \qquad \Gamma \vdash P : A}{\Gamma \vdash (M)P : B}$$

$$\frac{\Gamma \vdash M : A \Rightarrow A}{\Gamma \vdash \mathsf{fix}(M) : A}$$

Differential constructions

$$\frac{\Gamma \vdash M : A \Rightarrow B}{\Gamma \vdash \mathsf{T}M : \mathsf{T}A \Rightarrow \mathsf{T}B} \qquad \frac{\Gamma \vdash M : \mathsf{T}^{d+1}A \quad i \in \{0,1\}}{\Gamma \vdash \pi_i^d(M) : \mathsf{T}^d A}$$

$$\frac{\Gamma \vdash M : \mathsf{T}^d A \quad i \in \{0,1\}}{\Gamma \vdash \iota_i^d(M) : \mathsf{T}^{d+1}A} \qquad \frac{\Gamma \vdash M : \mathsf{T}^{d+2}A}{\Gamma \vdash \theta^d(M) : \mathsf{T}^{d+1}A}$$

$$\frac{\Gamma \vdash M : \mathsf{T}^{d+l+2}A}{\Gamma \vdash \mathsf{c}_l^d(M) : \mathsf{T}^{d+l+2}A} \qquad \frac{\Gamma \vdash M_1 : A \quad \Gamma \vdash M_2 : A}{\Gamma \vdash M_1 + M_2 : A}$$

Last rule: to be used only for proving subject reduction!

Differential modification

We define $\partial(x, M)$ by induction on M in such a way that

$$\Gamma, x : A \vdash M : B \quad \Rightarrow \quad \Gamma, x : \mathsf{T}A \vdash \partial(x, M) : \mathsf{T}B$$

Similar to the $\frac{\partial M}{\partial x} \cdot N$ of the differential λ-calculus, but without explicit sums.

- $\partial(x, x) = x, \quad \partial(x, y) = \iota_0^0(y), \quad \partial(x, \underline{\nu}) = \iota_0^0(\underline{\nu})$
- $\partial(x, \mathsf{succ}^d(M)) = \mathsf{succ}^{d+1}(\partial(x, M))$
- $\partial(x, \mathsf{if}^d(M, P, Q)) = \theta^0(\mathsf{c}_d^0(\mathsf{if}^{d+1}(\partial(x, M), \partial(x, P), \partial(x, Q))))$

- $\partial(x, \lambda y : A \cdot M) = \lambda y : A \cdot \partial(x, M)$
- $\partial(x, (M)P) = (\theta^0(\mathsf{T}\partial(x, M)))\partial(x, P)$
- $\partial(x, \mathsf{fix}(M)) = \mathsf{fix}(\theta^0(\mathsf{T}\partial(x, M)))$

Typing: we have $\Gamma, x : A \vdash M : B \Rightarrow B$.

Hence $\Gamma, x : \mathsf{T}A \vdash \partial(x, M) : \mathsf{T}(B \Rightarrow B)$, and remember that $\mathsf{T}(B \Rightarrow B) = (B \Rightarrow \mathsf{T}B)$.

So $\Gamma, x : \mathsf{T}A \vdash \mathsf{T}\partial(x, M) : (\mathsf{T}B \Rightarrow \mathsf{T}^2B) = \mathsf{T}^2(\mathsf{T}B \Rightarrow B)$
and hence
$\Gamma, x : \mathsf{T}A \vdash \theta^0(\mathsf{T}\partial(x, M)) : \mathsf{T}(\mathsf{T}B \Rightarrow B) = (\mathsf{T}B \Rightarrow \mathsf{T}B)$.

Finally $\Gamma, x : \mathsf{T}A \vdash \mathsf{fix}(\theta^0(\mathsf{T}\partial(x, M))) : \mathsf{T}B$.

- $\partial(x, \mathsf{T}M) = \mathsf{c}_0^0(\mathsf{T}\partial(x, M))$
- $\partial(x, \pi_i^d(M)) = \pi_i^{d+1}(\partial(x, M))$
- $\partial(x, \theta^d(M)) = \theta^{d+1}(\partial(x, M))$
- $\partial(x, \mathsf{c}_l^d(M)) = \mathsf{c}_l^{d+1}(\partial(x, M))$

$$\vdots$$

Intuition

In $\theta^d(M)$, θ^d is a tag for a place where a sum should be inserted when a projection will be applied.

Some standard PCF reduction rules

- $\mathsf{succ}^0(\underline{\nu}) \to \underline{\nu + 1}$
- $\mathsf{if}^0(\underline{0}, P, Q) \to P, \quad \mathsf{if}^0(\underline{\nu + 1}, P, Q) \to Q$
- $(\lambda x : A \cdot M)P \to M\,[P/x]$
- $\mathsf{fix}(M) \to (M)\mathsf{fix}(M)$

Some differential reduction rules

- $\mathsf{T}(\lambda x : A \cdot M) \to \lambda x : \mathsf{T}A \cdot \partial(x, M)$
- $\pi_i^d(\iota_j^d(M)) \to 0$ if $i \neq j$
- $\pi_i^d(\iota_i^d(M)) \to M$
- $\pi_{i_1}^d(\cdots \pi_{i_{l+2}}^d(\mathsf{c}_l^d(M))) \to \pi_{i_{l+2}}^d(\pi_{i_1}^d(\cdots \pi_{i_{l+1}}^d(M)))$
- $\pi_0^d(\theta^d(M)) \to \pi_0^d(\pi_0^d(M))$
- $\pi_1^d(\theta^d(M)) \to \pi_1^d(\pi_0^d(M)) + \pi_0^d(\pi_1^d(M))$

it is the only reduction which produces a sum.

When proving the soundness of this reduction (wrt. the semantics outlined before), one also proves that these sums are well defined in any model.

A Krivine machine

+ many rules expressing some kind of orthogonality between the action of projections $\pi_i^d(\_)$ and the other constructs.

A complete strategy can be presented by means of a "Krivine machine" without environment. A state of the machine is a tuple

$$(\delta, M, s)$$

where

- $\vdash M : \mathsf{T}^d E$ where E is a sharp type (that is, not of the shape $\mathsf{T}A$, that is $E = (A_1 \Rightarrow \cdots A_k \Rightarrow \iota)$),
- $\delta \in \{0, 1\}^d$ to be understood as a sequence of projections (address in the witness-tree M)
- s is a stack such that $s : E \vdash \iota$

The stacks

E: sharp type.

$$\frac{}{() : \iota \vdash \iota} \qquad \frac{s : \iota \vdash \iota}{\mathsf{succ} \cdot s : \iota \vdash \iota}$$

$$\frac{s : E \vdash \iota \qquad \delta \in \{0,1\}^d \qquad \vdash P, Q : \mathsf{T}^d E}{\mathsf{if}(\delta, P, Q) \cdot s : \iota \vdash \iota}$$

$$\frac{\vdash M : A \qquad s : E \vdash \iota}{\mathsf{arg}(M) \cdot s : A \Rightarrow E \vdash \iota} \qquad \frac{s : \mathsf{T}A \Rightarrow E \vdash \iota \qquad i \in \{0,1\}}{\mathsf{T}(i) \cdot s : A \Rightarrow E \vdash \iota}$$

Example of (almost) standard transitions

$d = \mathsf{len}(\delta)$

- $(\delta, \mathsf{succ}^d(M), s) \to (\delta, M, \mathsf{succ} \cdot s)$
- $(\langle\rangle, \underline{\nu}, \mathsf{succ} \cdot s) \to (\langle\rangle, \underline{\nu + 1}, s)$
- $(\varepsilon\delta, \mathsf{if}^d(M, P, Q), s) \to (\delta, M, \mathsf{if}(\varepsilon, P, Q) \cdot s)$
- $(\langle\rangle, \underline{0}, \mathsf{if}(\delta, P, Q) \cdot s) \to (\delta, P, s)$,
- $(\langle\rangle, \underline{\nu + 1}, \mathsf{if}(\delta, P, Q) \cdot s) \to (\delta, Q, s)$
- $(\delta, (M)P, s) \to (\delta, M, \mathsf{arg}(P) \cdot s)$
- $(\delta, \mathsf{fix}(M), s) \to (\delta, M, \mathsf{arg}(\mathsf{fix}(M)) \cdot s)$
- $(\delta, \lambda x : A \cdot M, \mathsf{arg}(P) \cdot s) \to (\delta, M\left[P/x\right], s)$

Differential transitions

- $(\delta i, \mathsf{T}M, s) \to (\delta, M, \mathsf{T}(i) \cdot s)$
- $(\delta, \lambda x : A \cdot M, \mathsf{T}(i) \cdot s) \to (\delta i, \lambda x : \mathsf{T}A \cdot \partial(x, M), s)$
- $(\varepsilon\delta, \pi_i^d(M), s) \to (\varepsilon i \delta, M, s)$
- $(\varepsilon i \delta, \iota_i^d(M), s) \to (\varepsilon\delta, M, s), \quad (\varepsilon i \delta, \iota_{1-i}^d(M), s) \to 0$
- $(\varepsilon\lambda\delta, \mathsf{c}_l^d(M), s) \to (\varepsilon \underset{\to}{\lambda} \delta, M, s) \quad$ with $l + 2 = \mathrm{len}(\lambda)$
  $\underset{\to}{\lambda}$: circular permutation to the right
- $(\varepsilon 0 \delta, \theta^d(M), s) \to (\varepsilon 00 \delta, M, s)$
- $(\varepsilon 1 \delta, \theta^d(M), s) \to (\varepsilon 01 \delta, M, s) + (\varepsilon 10 \delta, M, s)$

Last trick: one can get rid of the $+$ by making the word component of states *writable*. Thank you Guillaume Geoffroy!
$\rightsquigarrow$ a fully deterministic evaluation mechanism.

Conclusion

- We are developing the general theory of Coherent Differentiation with Aymeric Walch (PhD thesis), see his poster.
- In particular, there is a purely "cartesian theory" of CD without references to LL (as in JS Lemay's talk).
- Instead of $\mathbb{D} = 1 \mathbin{\&} 1$, we can take $\mathbb{D} = \overbrace{1 \mathbin{\&} 1 \mathbin{\&} \cdots}^{\mathbb{N}}$ and then $\mathsf{T}M$ becomes a syntactic Faà di Bruno transformation (also in the semantics of course) which generalizes the Taylor expansion $\rightsquigarrow$ Coherent Taylor expansion, Coherent Taylor PCF.
- Is there a Coherent Differential LL? The Elementary Situation suggests that this should be the case. Not clear yet what it is exactly. . .