Derivatives and distances in probabilistic coherence spaces

Second PIHOC Meeting, Bologna

Thomas Ehrhard, IRIF, CNRS and Univ Paris Diderot

7 Feb 2019

This talk

Try to use the regularity of morphisms in the cartesian closed category $\mathbf{Pcoh}_!$ (they are power series with coefficients ≥ 0 with > 0 radius of convergence).

In particular: these functions have differentials. What is their probabilistic operational meaning? Two (related) answers:

- Expected computation time
- Operational distance

in a probabilistic functional programming language: pPCF.

Syntax of probabilistic pPCF

Syntax:

$$\sigma, \tau, \ldots := \iota \mid \sigma \Rightarrow \tau$$
$$M, N, P \ldots := \underline{n} \mid \mathsf{succ}(M) \mid \mathsf{pred}\, M \mid x \mid \mathsf{coin}(r) \mid \mathsf{let}(x, M, N)$$
$$\mid \mathsf{if}(M, N, P) \mid (M)\, N \mid \lambda x^{\sigma}\, M \mid \mathsf{fix}(M)$$

with $r \in [0, 1]$.

Typing rules

$$\frac{}{\Gamma \vdash \underline{n} : \iota} \qquad \frac{}{\Gamma, x : \sigma \vdash x : \sigma}$$

$$\frac{\Gamma \vdash M : \iota}{\Gamma \vdash \mathsf{succ}(M) : \iota} \qquad \frac{\Gamma \vdash M : \iota}{\Gamma \vdash \mathsf{pred}\, M : \iota}$$

$$\frac{\Gamma \vdash M : \iota \qquad \Gamma, x : \iota \vdash N : \sigma}{\Gamma \vdash \mathsf{let}(x, M, N) : \sigma}$$

Only for M of type ι !

$$\frac{\Gamma, x : \sigma \vdash M : \tau}{\Gamma \vdash \lambda x^{\sigma}\, M : \sigma \Rightarrow \tau} \qquad \frac{\Gamma \vdash M : \sigma \Rightarrow \tau \qquad \Gamma \vdash N : \sigma}{\Gamma \vdash (M)\, N : \tau}$$

$$\frac{\Gamma \vdash M : \sigma \Rightarrow \sigma}{\Gamma \vdash \mathsf{fix}(M) : \sigma} \qquad \frac{r \in [0,1]}{\Gamma \vdash \mathsf{coin}(r) : \iota}$$

Intuitions on pPCF

$\vdash M : \iota$ means that M represents a subprobability distribution on the integers.

Example: coin$(1/3)$ weights $\underline{0}$ with probability $1/3$, $\underline{1}$ with probability $2/3$ and $\underline{n+2}$ with probability 0.

let(x, M, N): samples an integer according to M and feeds N through x with the obtained value (a $\underline{n}$ for some $n \in \mathbb{N}$).

$\vdash M : \iota \Rightarrow \iota$ means that M is an (generally non linear) sub-probability distribution transformer.

Reduction rules

We define a weak head reduction strategy.

Deterministic reduction rules

$$\frac{}{(\lambda x^\sigma\, M)\, N \to_{\mathsf{d}} M\,[N/x]} \qquad \frac{}{\mathsf{fix}(M) \to_{\mathsf{d}} (M)\,\mathsf{fix}(M)}$$

$$\frac{}{\mathsf{succ}(\underline{n}) \to_{\mathsf{d}} \underline{n+1}} \qquad \frac{}{\mathsf{if}(\underline{0}, M, N) \to_{\mathsf{d}} M}$$

$$\frac{}{\mathsf{if}(\underline{n+1}, M, N) \to_{\mathsf{d}} N} \qquad \frac{}{\mathsf{let}(x, \underline{n}, N) \to_{\mathsf{d}} N\,[\underline{n}/x]}$$

Probabilistic reductions

$$\frac{M \to_{\mathsf{d}} M'}{M \xrightarrow{1} M'} \qquad \frac{}{\mathsf{coin}(p) \xrightarrow{p} \underline{0}} \qquad \frac{}{\mathsf{coin}(p) \xrightarrow{1-p} \underline{1}}$$

$$\frac{M \xrightarrow{p} M'}{(M)\,N \xrightarrow{p} (M')\,N} \qquad \frac{M \xrightarrow{p} M'}{\mathsf{succ}(M) \xrightarrow{p} \mathsf{succ}(M')}$$

$$\frac{M \xrightarrow{p} M'}{\mathsf{let}(x, M, N) \xrightarrow{p} \mathsf{let}(x, M', N)}$$

$$\frac{M \xrightarrow{p} M'}{\mathsf{if}(M, N, P) \xrightarrow{p} \mathsf{if}(M', N, P)}$$

Probability of reduction

Given M such that $\vdash M : \iota$, we can consider all possible reductions from M to a given integer constant $\underline{n}$:

$$M = M_0 \xrightarrow{p_1} M_1 \xrightarrow{p_2} \cdots \xrightarrow{p_k} M_k = \underline{n}$$

Summing up the probabilities $\prod_{i=1}^{k} p_i$ of all these paths we get the probability that M reduces to $\underline{n}$, denoted $\Pr(M \downarrow \underline{n})$.

Observational distance

Given M and N such that $\vdash M : \sigma$ and $\vdash N : \sigma$, one defines the *observational distance* $\mathrm{d}_{\mathrm{obs}}(M, N)$ between M and N as the sup of all

$$|\Pr((C)\,M \downarrow \underline{0}) - \Pr((C)\,N \downarrow \underline{0})|$$

for all possible "contexts" which are closed terms C such that $\vdash C : \sigma \Rightarrow \iota$.

M and N are *observationally equivalent* if this "distance" is 0.

$\mathrm{d}_{\mathrm{obs}}(\_, \_)$ is a distance on the observational classes of closed terms of type σ (for any type σ).

Part I

Reminder — Probabilistic coherence spaces: an "analytic" denotational model

If $u, u' \in (\mathbb{R}_{\geq 0})^I$ then $\langle u, u' \rangle = \sum_{i \in I} u_i u'_i \in \overline{\mathbb{R}_{\geq 0}}$.
If $P \subseteq \overline{\mathbb{R}_{\geq 0}}^I$ then

$$P^\perp = \{u' \in \overline{\mathbb{R}_{\geq 0}}^I \mid \forall u \in P \ \langle u, u' \rangle \leq 1\}.$$

A probabilistic coherence space (PCS) is a pair $X = (|X|, \mathrm{P}X)$ where $\mathrm{P}X \subseteq \overline{\mathbb{R}_{\geq 0}}^{|X|}$ such that

- $\mathrm{P}X^{\perp\perp} = \mathrm{P}X$
- $\forall a \in |X| \, \exists x \in \mathrm{P}X \ x_a > 0$
- $\forall a \in |X| \, \exists m \in \mathbb{R}_{\geq 0} \forall x \in \mathrm{P}X \ x_a \leq m$

So actually $\mathrm{P}X \subseteq (\mathbb{R}_{\geq 0})^{|X|}$.

Dual of X: $X^\perp = (|X|, \mathsf{P}X^\perp)$ so that $X^{\perp\perp} = X$.

Examples of PCS's.

- $1 = (\{*\}, [0,1])$ with $1^\perp = 1$ for $0 < r < \infty$.
- $\mathbf{Bool} = 1 \oplus 1 = (\{\mathbf{t}, \mathbf{f}\}, \{(x_\mathbf{t}, x_\mathbf{f}) \in \mathbb{R}^2_{\geq 0} \mid x_\mathbf{t} + x_\mathbf{f} \leq 1\})$
- $\mathbf{Bool}^\perp = 1 \,\&\, 1 = (\{\mathbf{t}, \mathbf{f}\}, \{(x_\mathbf{t}, x_\mathbf{f}) \in \mathbb{R}^2_{\geq 0} \mid x_\mathbf{t}, x_\mathbf{f} \leq 1\})$
- $\mathsf{N} = (\mathbb{N}, \{x \in (\mathbb{R}_{\geq 0})^\mathbb{N} \mid \sum_{i=0}^\infty x_i \leq 1\})$

Linear morphisms in PCS's

Linear morphisms from X to Y: if $t \in (\mathbb{R}_{\geq 0})^{|X| \times |Y|}$ (a matrix) and $x \in \mathsf{P}X$ (a vector) then we can apply the matrix to the vector:

$$t\,x \in (\mathbb{R}_{\geq 0})^{|Y|} \text{ with } (t\,x)_b = \sum_{a \in |X|} t_{a,b} x_a$$

Then t is a linear morphism from X to Y ($t \in \mathbf{Pcoh}(X, Y)$) if $\forall x \in \mathsf{P}X\ \ t\,x \in \mathsf{P}Y$. This defines a category **Pcoh**, a model of LL (with all fixpoints of types and term fixpoint operators at all types).

Fact

$\mathbf{Pcoh}(X, Y)$ *is a PCS structure, that is* $\mathbf{Pcoh}(X, Y) = \mathsf{P}(X \multimap Y)$ *for a PCS* $X \multimap Y$ *with* $|X \multimap Y| = |X| \times |Y|$.

$\mathsf{P}(\mathsf{N} \multimap \mathsf{N})$: sub-stochastic matrices on $\mathbb{N} \times \mathbb{N}$.

Non-linear morphisms in PCS's

The non-linear morphisms $X \to Y$ are the elements of $\mathsf{P}(!X \multimap Y)$ where $!X$ is a PCS with $|!X| = \mathcal{M}_{\mathrm{fin}}(|X|)$ (finite multisets).
If $x \in \mathsf{P}X$ and $m \in |!X|$ define

$$x^m = \prod_{a \in |X|} x_a^{m(a)} .$$

Then $t \in \mathsf{P}(!X \multimap Y) = \mathbf{Pcoh}_!(X, Y)$ is characterized by

$$\forall x \in \mathsf{P}X \ \widehat{t}(x) = \left(\sum_{m \in |!X|} t_{m,b} x^m \right)_{b \in |Y|} \in \mathsf{P}Y$$

$\mathbf{Pcoh}_!$ is a model of probabilistic pPCF. Morphisms are functions: $\widehat{t \circ s} = \widehat{t} \circ \widehat{s}$ and $(\forall x \in \mathsf{P}X \ \widehat{s}(x) = \widehat{t}(x)) \Rightarrow s = t$.

pPCF interpretation

Fact

$\mathbf{Pcoh}_!$ *is a cartesian closed category with an object of integers* N *and least fixpoints operators* $(X \Rightarrow X) \to X$ *for all* X*. An thus it is a model of* pPCF*.*

$[\![\iota]\!] = \mathsf{N}$ and $[\![\sigma \Rightarrow \tau]\!] = ![\![\sigma]\!] \multimap [\![\tau]\!]$
If $\Gamma \vdash M : \sigma$ with $\Gamma = (x_1 : \sigma_1, \dots, x_k : \sigma_k)$ then

$$[\![M]\!]_\Gamma \in \mathbf{Pcoh}_!([\![\sigma_1]\!] \mathbin{\&} \cdots \mathbin{\&} [\![\sigma_k]\!], [\![\sigma]\!])$$

so $[\![M]\!]_\Gamma$ can be seen as a function $\prod_{i=1}^{k} \mathsf{P}[\![\sigma_i]\!] \to \mathsf{P}[\![\sigma]\!]$.

Why no negative coefficients in power series?

Seems crucial for combining fixpoints and power series. Assume *e.g.* that we admit the "weak parallel or" function

$$\mathsf{wpor} : [0,1] \times [0,1] \to [0,1] \qquad (u, v) \mapsto u + v - uv$$

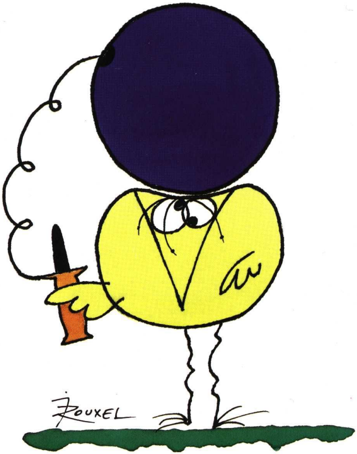
Spawns two threads, stops as soon as one of them stops.

Add it to pPCF. Then we can define $\vdash P : 1 \Rightarrow 1$ by

$$P = \mathsf{fix}(\lambda f^{\iota \Rightarrow \iota}\, \lambda x^{\iota}\, \mathsf{wpor}(x, (f)\, x))\,.$$

Spawns an unbounded number of copies of x, stops as soon as one of them does. Then $[\![P]\!](0) = 0$ and $[\![P]\!](u) = 1$ for $u > 0$.
Scott continuous, but far from being analytic!

Les devises Shadok



EN ESSAYANT CONTINUELLEMENT
ON FINIT PAR RÉUSSIR. DONC:
PLUS ÇA RATE, PLUS ON A
DE CHANCES QUE ÇA MARCHE.

Example of term interpretations

- $[\![x_i]\!]_\Gamma(\vec{u}) = u_i$
- $[\![\underline{n}]\!]_\Gamma(\vec{u}) = e_n$ where $(e_n)_j = \delta_{n,j} \in \mathsf{PN}$.
- $[\![\mathrm{if}(M, N, P)]\!]_\Gamma =$
  $[\![M]\!]_\Gamma(\vec{u})_0 [\![N]\!]_\Gamma(\vec{u}) + (\sum_{n=1}^{\infty} [\![M]\!]_\Gamma(\vec{u})_n) [\![P]\!]_\Gamma(\vec{u})$
- $[\![\mathrm{coin}(p)]\!]_\Gamma(\vec{u}) = pe_0 + (1-p)e_1$
- $[\![\mathrm{let}(x, M, N)]\!]_\Gamma(\vec{u}) = \sum_{n=0}^{\infty} [\![M]\!]_\Gamma(\vec{u})_n [\![N]\!]_{\Gamma, x:\iota}(\vec{u}, e_n)$

Main properties of this interpretation

Fact

For all M with $\vdash M : \iota$ and $n \in \mathbb{N}$, we have $\mathrm{Pr}(M \downarrow \underline{n}) = [\![M]\!]_n$.

As a consequence

Fact (adequacy, Danos and E.)

For all M, N such that $\vdash M : \sigma$ and $\vdash N : \sigma$, we have
$[\![M]\!] = [\![N]\!] \Rightarrow \mathrm{d_{obs}}(M, N) = 0$

And also

Fact (full abstraction, Pagani, Tasson and E.)

The converse implication.

Part II

Derivatives and execution (on an example)

In a pPCF extension with unit type 1, for $r \in [0,1]$

$$M_r = \mathrm{fix}(\lambda f^{1\Rightarrow 1}\, \lambda x^1\, \mathrm{if}(\mathrm{coin}(r),\ (f)\,x;(f)\,x,\ \ x;x))$$

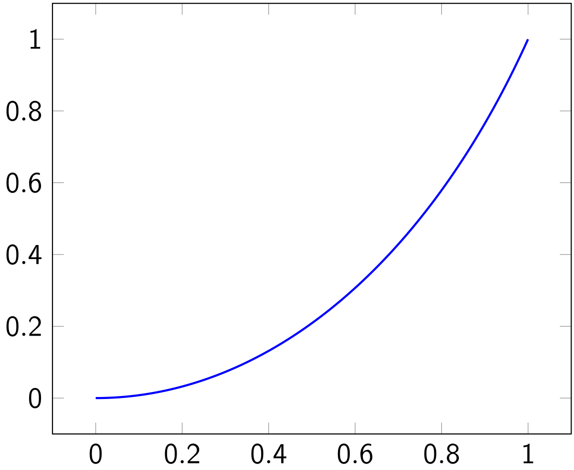
where $()$ is the unique value of type 1 and ";" is the "unit conditional" ($M;N$ reduces to $()$ if both M and N do).

Then $[\![M_r]\!]$ is a *monotonic* function $\varphi_r : [0,1] \to [0,1]$ minimal such that $\varphi_r(u) = (1-r)u^2 + r\varphi_r(u)^2$. Hence

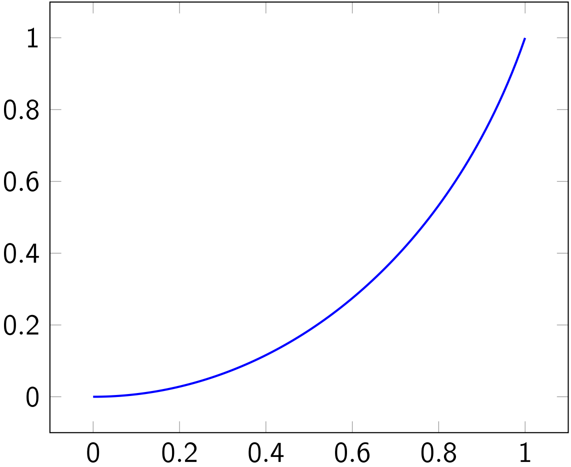
$$\varphi_r(u) = \begin{cases} \frac{1-\sqrt{1-4r(1-r)u^2}}{2r} & \text{if } u > 0 \\ u^2 & \text{if } r = 0\,. \end{cases}$$

NB: by adequacy, $\varphi_r(1)$ is $\mathrm{Pr}((M_r)\,()\downarrow())$

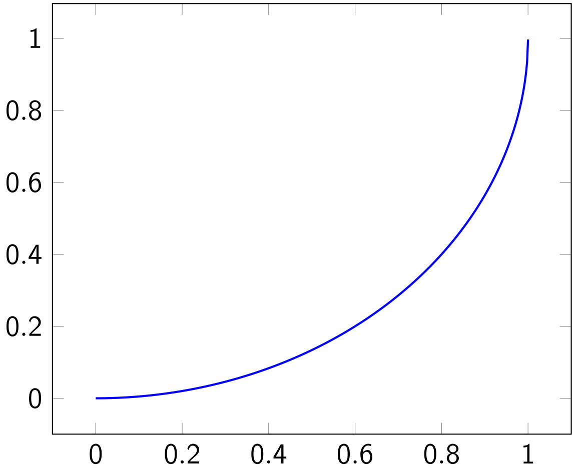
$r = 0.2$ — graph of $\varphi_{0.2}(u) = (1 - \sqrt{1 - 0.64u^2})/0.4$



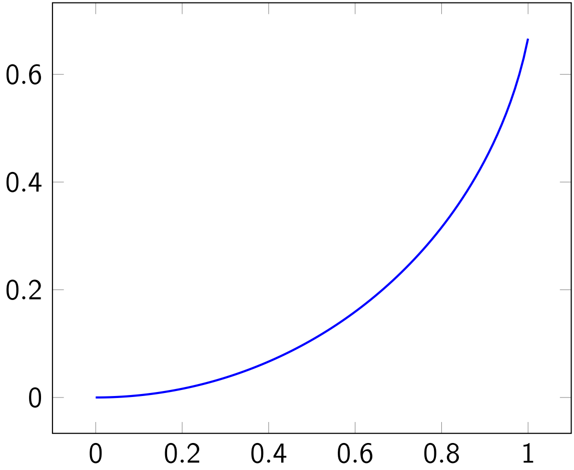
$r = 0.3$ — graph of $\varphi_{0.3}(u) = (1 - \sqrt{1 - 0.84u^2})/0.6$ (steeper slope at $u = 1$)



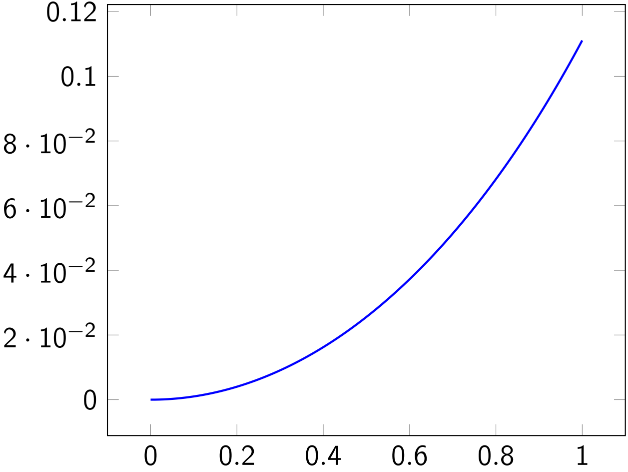
$r = 0.5$ — graph of $\varphi_{0.5}(u) = 1 - \sqrt{1 - u^2}$ (vertical slope at $u = 1$)



$r = 0.6$ — graph of $\varphi_{0.6}(u) = (1 - \sqrt{1 - 0.96u^2})/1.2$ ($\varphi_{0.6}(u) < 1$ but less steep slope at $u = 1$)



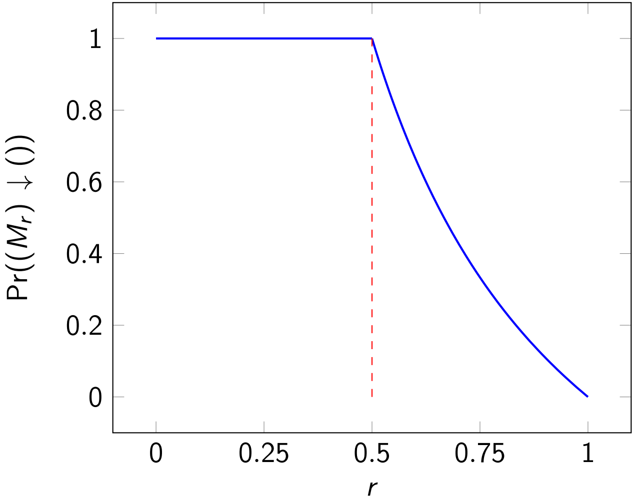
$r = 0.9$ — graph of $\varphi_{0.9}(u) = (1 - \sqrt{1 - 0.36u^2})/1.8$



Probability of termination of $(M_r)\,()$ for $r \in [0,1]$:

$$\begin{aligned}
\varphi_r(1) &= \frac{1 - \sqrt{1 - 4r(1-r)}}{2r} \\
&= \frac{1 - |1 - 2r|}{2r} \\
&= \begin{cases} 1 & \text{if } r \leq 1/2 \\ \frac{1-r}{r} & \text{if } r > 1/2 \end{cases}
\end{aligned}$$

Graph of $\varphi_r(1)$ for $0 \leq r \leq 1$: probability of termination of $(M_r)\,()$.



$$\varphi_r(u) = \sum_{n=0}^{\infty} a_n(r) u^n$$

Fact

$a_n(r) \in [0,1]$ is the probability that the execution of $(M_r)\,()$ uses n times the argument $()$ in its reduction to $()$.

$$\varphi_r'(1) = \lim_{u \to 1^-} \frac{\varphi_r(1) - \varphi_r(u)}{1-u} = \sum_{n=0}^{\infty} n\, a_n(r)$$

Fact

$\varphi_r'(1)/\varphi_r(1)$ is the conditional expectation of this execution time, under the condition that the computation terminates.

In the example we can compute this derivative. We have

$$\varphi_r(u) = (1-r)u^2 + r\varphi_r(u)^2$$

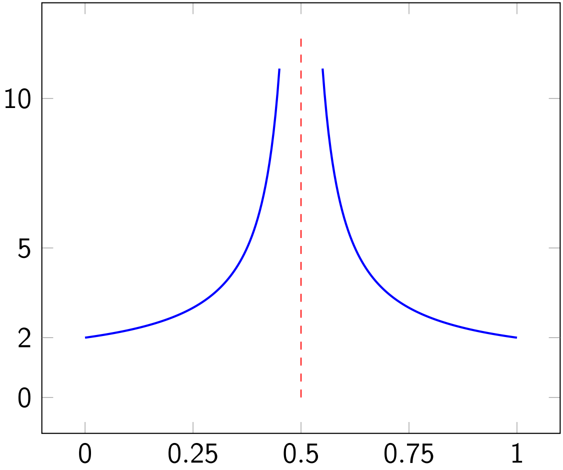
so

$$\varphi_r'(u) = 2(1-r)u + 2r\varphi_r(u)\varphi_r'(u)$$

The conditional expectation of execution time is

$$\frac{\varphi_r'(1)}{\varphi_r(1)} = \frac{2(1-r)}{(1-2r\varphi_r(1))\varphi_r(1)} = \begin{cases} \frac{2(1-r)}{1-2r} & \text{if } 0 \leq r < 1/2 \\ \frac{2r}{2r-1} & \text{if } 1/2 < r \leq 1 \end{cases}$$

Graph of $\varphi_r'(1)/\varphi_r(1)$: conditional expectation of the number of steps in the reduction of $(M_r)()$ for $0 \leq r \leq 1$.



Part III

Lipschitz property and distances

Amplification of probabilities

Fact

$$\forall \varepsilon \in [0,1] \quad \varepsilon > 0 \Rightarrow \mathsf{d}_{\mathsf{obs}}(\mathsf{coin}(0), \mathsf{coin}(\varepsilon)) = 1$$

Take $C = \mathsf{fix}(\lambda f^{\iota \Rightarrow \iota}\, \lambda x^{\iota}\, \mathsf{if}(x, \underline{0}, (f)\, x))$ then

$$\Pr((C)\,\mathsf{coin}(0) \downarrow \underline{0}) = 0$$
$$\Pr((C)\,\mathsf{coin}(\varepsilon) \downarrow \underline{0}) = 1 \quad \forall \varepsilon > 0$$

The denotational distance

There is a "norm" on $\mathsf{P}X$:

$$\|x\|_X = \sup\{\langle x, x'\rangle \mid x' \in \mathsf{P}X^\perp\} \in [0,1]$$

For instance $\|x\|_{\mathsf{N}} = \sum_{n=0}^{\infty} x_n$.

Also $\mathsf{P}X$ is a lattice: $x \wedge y \in \mathsf{P}X$ defined pointwise.
Then

$$\mathsf{d}_X(x,y) = \|x - (x \wedge y)\|_X + \|y - (x \wedge y)\|_X \,.$$

defines a distance on $\mathsf{P}X$.

Remember that $\mathsf{d}_{\mathsf{obs}}(\mathsf{coin}(0), \mathsf{coin}(\varepsilon)) = 1$ if $\varepsilon > 0$.

On the other hand $[\![\mathsf{coin}(0)]\!] = e_1$ and $[\![\mathsf{coin}(\varepsilon)]\!] = \varepsilon e_0 + (1-\varepsilon)e_1$. Also $e_1 \wedge (\varepsilon e_0 + (1-\varepsilon)e_1) = (1-\varepsilon)e_1$.

$$\begin{aligned} \mathsf{d}_{\mathsf{N}}([\![\mathsf{coin}(0)]\!], [\![\mathsf{coin}(\varepsilon)]\!]) &= \|e_1 - (1-\varepsilon)e_1\|_{\mathsf{N}} \\ &+ \|(1-\varepsilon)e_1 + \varepsilon e_0 - (1-\varepsilon)e_1\|_{\mathsf{N}} \\ &= 2\varepsilon\,. \end{aligned}$$

Remember that by Full Abstraction, if $\vdash M : \sigma$ and $\vdash N : \sigma$,

$$\mathsf{d}_{[\![\sigma]\!]}([\![M]\!], [\![N]\!]) = 0 \Leftrightarrow \mathsf{d}_{\mathsf{obs}}(M, N) = 0$$

we would like to say something not completely trivial in the case $\mathsf{d}_{\mathsf{obs}}(M, N) \neq 0$ by limiting the space of observation contexts (in the spirit of the work of Dal Lago and *al.* on probabilistic distances).

The local PCS

Given $x \in \mathsf{P}X$, there is a PCS X_x such that

$$\mathsf{P}(X_x) = \{u \in \mathsf{P}X \mid x + u \in \mathsf{P}X\}$$

We have $|X_x| = \{a \in |X| \mid \exists \varepsilon > 0\ x + \varepsilon e_a \in \mathsf{P}X\}$.

It is a PCS (not completely obvious).

This is the *local PCS* of X at x.

[$e_a \in (\mathbb{R}_{\geq 0})^{|X|}$ defined by $(e_a)_{a'} = \delta_{a,a'}$.]

Local derivatives

Let $t \in \mathbf{Pcoh}_!(X, Y)$ and $x \in \mathsf{P}X$
Given $u \in \mathsf{P}(X_x)$, we know that $x + u \in \mathsf{P}X$ and hence we can compute $\widehat{t}(x + u) \in \mathsf{P}Y$:

$$\begin{aligned}\widehat{t}(x+u)_b &= \sum_{m \in |!X|} t_{m,b}(x+u)^m \\ &= \sum_{m \in |!X|} t_{m,b} \sum_{p \leq m} \binom{m}{p} x^{m-p} u^p.\end{aligned}$$

where

$$\binom{m}{p} = \prod_{a \in |X|} \binom{m(a)}{p(a)}$$

So keeping only the terms which are constant (p empty) and linear (p singleton) in u:

$$\widehat{t}(x) + \sum_{a \in |X|} u_a \sum_{m \in |!X|} (m(a)+1) t_{m+[a],b} x^m \leq \widehat{t}(x+u) \in \mathrm{P}Y$$

that is

$$\sum_{a \in |X|} \left(\sum_{m \in |!X|} (m(a)+1) t_{m+[a],b} x^m \right) u_a \in \mathrm{P}(X_{\widehat{t}(x)})$$

So we define the "Jacobian matrix" of t at x by

$$t'(x)_{a,b} = \sum_{m \in |!X|} (m(a) + 1) t_{m+[a],b} x^m$$

and we have seen that

$$t'(x) \in \mathrm{P}(X_x, Y_{\widehat{t}(x)}) .$$

Fact (chain rule)

Let $s \in \mathbf{Pcoh}_!(X, Y)$ and $t \in \mathbf{Pcoh}_!(Y, Z)$. Let $x \in \mathrm{P}X$ and $u \in \mathrm{P}X_x$. Then we have $(t \circ s)'(x)\, u = t'(\widehat{s}(x))\, s'(x)\, u$.

A Lipschitz propery

We take some $p \in [0,1)$. Let $t \in \mathbf{Pcoh}_!(X,1)$.

Fact (basic observation)

If $x \in \mathsf{P}X$ and $\|x\|_X \leq p$, then $\forall u \in \mathsf{P}X$, one has

$$\|x + (1-p)u\|_X \leq \|x\|_X + (1-p)\|u\|_X \leq 1$$

so $(1-p)u \in \mathsf{P}(X_x)$.

So if $w \in \mathsf{P}(X_x \multimap Y)$, we have $\forall u \in \mathsf{P}X$ $\|w\,(1-p)u\|_Y \leq 1$ and hence

$$(1-p)w \in \mathsf{P}(X \multimap Y)\,.$$

In particular

$$(1-p)t'(x) \in \mathsf{P}(X \multimap 1)\,.$$

Let $x \leq y \in \mathsf{P}X$ such that $\|y\|_X \leq p$. Observe that $2-p > 1$.

Then $x + (2-p)(y-x) = y + (1-p)(y-x) \in \mathsf{P}X$ since $\|y-x\|_X \leq \|y\|_X \leq p \leq 1$.

Since $x + (2-p)(y-x) \in \mathsf{P}X$ we can define

$$\begin{aligned} h : [0, 2-p] &\to [0,1] \\ \theta &\mapsto \widehat{t}(x + \theta(y-x)) \end{aligned}$$

We have $h \in \mathbf{Pcoh}_!([0, 2-p], [0,1])$ by compositionality.

Remember that $2 - p > 1$ so $1 \in [0, 2-p)$. By Chain Rule $\forall \theta \in [0,1]$, $h'(\theta) = t'(x + \theta(y-x))\,(y-x)$.

Since $\|x + \theta(y-x)\|_X \leq \|y\|_X \leq p$ we have

$$\forall \theta \in [0,1] \quad (1-p)t'(x + \theta(y-x)) \in \mathsf{P}(X \multimap 1)\,.$$

Hence

$$\forall \theta \in [0,1] \quad h'(\theta) = t'(x + \theta(y-x))\,(y-x) \leq \frac{\|y-x\|_X}{1-p}\,.$$

We have

$$0 \leq \widehat{t}(y) - \widehat{t}(x) = h(1) - h(0)$$
$$= \int_0^1 h'(\theta)\, d\theta \leq \frac{\|y - x\|_X}{1-p}\,.$$

From this we deduce easily

Fact (Lipschitz property of non-linear morphisms)

Let $p \in [0,1)$ and $t \in \mathbf{Pcoh}_!(X,1)$. Let $x, y \in \mathsf{P}X$ with $\|x\|_X, \|y\|_X \leq p$. Then

$$\left|\widehat{t}(x) - \widehat{t}(y)\right| \leq \frac{\mathsf{d}_X(x,y)}{1-p}$$

A syntactic consequence

If $\Gamma \vdash C : \sigma \Rightarrow 1$ let $C^{\langle p \rangle}$ with $\vdash C^{\langle p \rangle} : \sigma \Rightarrow 1$ be

$$C^{\langle p \rangle} = \lambda x^{\sigma}\,(C)\,\mathrm{if}(\mathrm{coin}(p), M, \Omega^{\sigma})\,.$$

So that $[\![(C^{\langle p \rangle})\; M]\!] = \widehat{[\![C]\!]}(p[\![M]\!])$ when $\vdash M : \sigma$. If $\vdash M : \sigma$ and $\vdash N : \sigma$.

Definition (p-tamed observational distance)

$$\mathrm{d}_{\mathrm{obs}}^{\langle p \rangle}(M, N) = \sup\{\left|\Pr(\left(C^{\langle p \rangle}\right) M \downarrow \underline{0}) - \Pr(\left(C^{\langle p \rangle}\right) N \downarrow \underline{0})\right| \;\mid\; \vdash C : \sigma \Rightarrow \iota\}$$

$$
\begin{aligned}
&\mathsf{d}^{\langle p\rangle}_{\mathsf{obs}}(M,N)\\
&= \sup\left\{ \left| \widehat{[\![C]\!]}(p[\![M]\!])_0 - \widehat{[\![C]\!]}(p[\![N]\!])_0 \right| \mid \ \vdash C : \sigma \Rightarrow \iota \right\}\\
&\leq \sup\left\{ \left| \widehat{t}(p[\![M]\!]) - \widehat{t}(p[\![N]\!] \right|) \mid t \in \mathsf{P}(![\![\sigma]\!] \multimap 1) \right\}\\
&\leq \frac{\mathsf{d}_{[\![\sigma]\!]}(p[\![M]\!], p[\![N]\!])}{1-p} = \frac{p}{1-p}\,\mathsf{d}_{[\![\sigma]\!]}([\![M]\!], [\![N]\!])\,.
\end{aligned}
$$

We have proven:

Fact (metric adequacy of PCS's)

If $\vdash M : \sigma$, $\vdash N : \sigma$ *and* $0 \leq p < 1$, *then*

$$\mathrm{d}^{\langle p \rangle}_{\mathrm{obs}}(M, N) \leq \frac{p}{1-p}\mathrm{d}_{[\![\sigma]\!]}([\![M]\!], [\![N]\!])$$

So for instance

$$\mathrm{d}^{\langle p \rangle}_{\mathrm{obs}}(\mathrm{coin}(0), \mathrm{coin}(\varepsilon)) \leq \frac{2\varepsilon p}{1-p}$$

Perspectives

- Understand better what happens on the border of PCS's (elements such that $\|x\| = 1$)
- Extend to "continuous types" (using positive cones and Crubillé's Theorem on stable functions on positive cones)
- A differential pPCF and probabilistic LL? What is the proof-theoretic status of these local derivatives?