Differentials and Distances in Probabilistic Coherence Spaces

Thomas Ehrhard
CNRS, IRIF, Université de Paris, France
https://www.irif.fr/~ehrhard/
ehrhard@irif.fr

Abstract

In probabilistic coherence spaces, a denotational model of probabilistic functional languages, morphisms are analytic and therefore smooth. We explore two related applications of the corresponding derivatives. First we show how derivatives allow to compute the expectation of execution time in the weak head reduction of probabilistic PCF (pPCF). Next we apply a general notion of "local" differential of morphisms to the proof of a Lipschitz property of these morphisms allowing in turn to relate the observational distance on pPCF terms to a distance the model is naturally equipped with. This suggests that extending probabilistic programming languages with derivatives, in the spirit of the differential lambda-calculus, could be quite meaningful.

2012 ACM Subject Classification Theory of computation → Lambda calculus; Theory of computation → Probabilistic computation; Theory of computation → Abstract machines; Theory of computation → Linear logic

Keywords and phrases Denotational semantics, probabilistic coherence spaces, differentials of programs

Digital Object Identifier 10.4230/LIPIcs.FSCD.2019.17

Acknowledgements We thank Raphaëlle Crubillé, Paul-André Melliès, Michele Pagani and Christine Tasson for many enlightening discussions on this work. We also thank the referees for their precious comments and suggestions.

Introduction

Currently available denotational models of probabilistic functional programming (with full recursion, and thus partial computations) can be divided in three classes.

- *Game* based models, first proposed in [6] and further developed by various authors (see [2] for an example of this approach). From their deterministic ancestors they typically inherit good definability features.
- Models based on Scott continuous functions on domains endowed with additional probability related structures. Among these models we can mention *Kegelspitzen* [13] (domains equipped with an algebraic convex structure) and *ω-quasi Borel spaces* [15] (domains equipped with a generalized notion of measurability), this latter semantics, as far as we understand the situation, requiring the use of an adapted probabilistic powerdomain construction.
- Models based on (a generalization of) Berry stable functions. The first category of this kind was that of *probabilistic coherence spaces* (PCSs) and power series with non-negative coefficients (the Kleisli category of the model of Linear Logic developed in [5]) for which we could prove adequacy and full abstraction with respect to a probabilistic version of PCF [10]. We extended this idea to "continuous data types" (such as $\mathbb{R}$) by substituting PCSs with *positive cones* and power series with functions featuring

an hereditary monotonicity property that we called *stability*[1] and [3] showed that this extension is actually conservative (stable functions on PCSs, which are special positive cones, are exactly power series).

The main feature of this latter semantics is the extreme regularity of its morphisms. Being power series, they must be smooth. Nevertheless, the category **Pcoh** is not a model of differential linear logic in the sense of [9]. This is due to the fact that general addition of morphisms is not possible (only sub-convex linear combinations are available) thus preventing, *e.g.*, the Leibniz rule to hold in the way it is presented in differential LL. Also a morphism $X \to Y$ in the Kleisli category $\mathbf{Pcoh}_!$ can be considered as a function from the *closed unit ball* of the cone P associated with X to the closed unit ball of the cone Q associated with Y. From a differential point of view such a morphism is well behaved only in the interior of the unit ball. On the border derivatives can typically take infinite values.

Contents

We already used the analyticity of the morphisms of $\mathbf{Pcoh}_!$ to prove full abstraction results [10]. We provide here two more corollaries of this properties, involving now also derivatives. For both results, we consider a paradigmatic probabilistic purely functional programming language[2] which is a probabilistic extension of Scott and Plotkin's PCF. This language pPCF features a single data type ι of integers, a simple probabilistic choice operator $\mathsf{coin}(r) : \iota$ which flips a coin with probability r to get $\underline{0}$ and $1-r$ to get $\underline{1}$. To make probabilistic programming possible, this language has a $\mathsf{let}(x, M, N)$ construct restricted to M of type ι which allows to sample an integer according to the sub-probability distribution represented by M. The operational semantics is presented by a deterministic "stack machine" which is an environment-free version of Krivine machine parameterized by a choice sequence $\in \mathcal{C}_0 = \{0,1\}^{<\omega}$, presented as a partial *evaluation function*. We adopt a standard discrete probability approach, considering $\mathcal{C}_0$ as our basic sample space and the evaluation function as defining a (total) probability density function on $\mathcal{C}_0$. We also introduce an extension $\mathsf{pPCF}_{\mathsf{lab}}$ of pPCF where terms can be labeled by elements of a set $\mathcal{L}$ of labels, making it possible to count the use of labeled subterms during a reduction. Evaluation for this extended calculus gives rise to a random variable (r.v.) on $\mathcal{C}_0$ ranging in the set $\mathcal{M}_{\mathrm{fin}}(\mathcal{L})$ of finite multisets of elements of $\mathcal{L}$. The number of uses of terms labeled by a given $l \in \mathcal{L}$ (which is a measure of the computation time) is then an $\mathbb{N}$-valued r.v. the expectation of which we want to evaluate. We prove that, for a given labeled closed term M of type ι, this expectation can be computed by taking a derivative of the interpretation of this term in the model $\mathbf{Pcoh}_!$ and provide a concrete example of computation of such expectations. This result can be considered as a probabilistic version of [7, 8]. The fact that derivatives can become infinite on the border of the unit ball corresponds then to the fact that this expectation of "computation time" can be infinite.

In the second application, we consider the contextual distance on pPCF terms generalizing Morris equivalence as studied in [4] for instance. The probabilistic features of the language make this distance too discriminating, putting *e.g.* terms $\mathsf{coin}(0)$ and $\mathsf{coin}(\varepsilon)$ at distance 1 for all $\varepsilon > 0$ (*probability amplification*). Any cone (and hence any PCS) is equipped with a norm and hence a canonically defined metric. Using a *locally defined* notion of differential

[1] Because, when reformulated in the domain-theoretic framework of Girard's coherence spaces, this condition exactly characterizes Berry's stable functions.

[2] One distinctive feature of our approach is to not consider probabilities as an effect.

of morphisms in $\mathbf{Pcoh}_!$, we prove that these morphisms enjoy a Lipschitz property on all balls of radius $p < 1$, with a Lipschitz constant $1/(1-p)$ (thus tending towards ∞ when p tends towards 1). Modifying the definition of the operational distance by not considering all possible contexts, but only those which "perturb" the tested terms by allowing them to diverge with probability $1-p$, we upper bound this p-tamed distance by the distance of the model with a ratio $p/(1-p)$. Being in some sense defined wrt. *linear* semantic contexts, the denotational distance does not suffer from the probability amplification phenomenon. This suggests that p-tamed distances might be more suitable than ordinary contextual distances to reason on probabilistic programs.

Notations

We use $\mathbb{R}_{\geq 0}$ for the set of real numbers x such that $x \geq 0$, and we set $\overline{\mathbb{R}_{\geq 0}} = \mathbb{R}_{\geq 0} \cup \{+\infty\}$. Given two sets S and I we use S^I for the set of functions $I \to S$, often considered as I-indexed families $\vec{s}$ of elements of S (the purpose of the arrow is to stress the fact that this object is such a family), the indexing set I being usually easily derivable from the context. The elements of such a family $\vec{s}$ are denoted s_i or $s(i)$ depending on the context. Given $i \in I$ we use $\bar{i}$ for the function $I \to \mathbb{R}_{\geq 0}$ such that $\bar{i}(i) = 1$ and $\bar{i}(j) = 0$ if $j \neq i$. We use $\mathcal{M}_{\mathrm{fin}}(I)$ for the set of finite multisets of elements of I. Such a multiset is a function $\mu : I \to \mathbb{N}$ such that $\mathsf{supp}(\mu) = \{i \in I \mid \mu(i) \neq 0\}$ is finite. We use additive notations for operations on multisets (0 for the empty multiset, $\mu + \nu$ for their pointwise sum). We use $[i_1, \ldots, i_k]$ for the multiset μ such that $\mu(i) = \#\{j \in \mathbb{N} \mid i_j = i\}$. If $\mu, \nu \in \mathcal{M}_{\mathrm{fin}}(I)$ with $\mu \leq \nu$ (pointwise order), we set $\binom{\nu}{\mu} = \prod_{i \in I} \binom{\nu(i)}{\mu(i)}$ where $\binom{n}{m} = \frac{n!}{m!(n-m)!}$ is the usual binomial coefficient. We use $I^{<\omega}$ for the set of finite sequences $\langle i_1, \ldots, i_k \rangle$ of elements of I and $\alpha\,\beta$ for the concatenation of such sequences. We use $\langle\rangle$ for the empty sequence.

1 Probabilistic coherence spaces (PCS)

For the general theory of PCSs we refer to [5, 10]. We recall briefly the basic definitions and provide a characterization of these objects. PCSs are particular cones (a notion borrowed from [14]) as we used them in [10], so we start with a few words about these more general structures to which we plan to extend the constructions of this paper.

1.1 A few words about cones

A (positive) *pre-cone* is a cancellative[3] commutative $\mathbb{R}_{\geq 0}$-semi-module P equipped with a norm $\|\_\|_P$, that is a map $P \to \mathbb{R}_{\geq 0}$, such that $\|r\,x\|_P = r\,\|x\|_P$ for $r \in \mathbb{R}_{\geq 0}$, $\|x+y\|_P \leq \|x\|_P + \|y\|_P$ and $\|x\|_P = 0 \Rightarrow x = 0$. It is moreover assumed that $\|x\|_P \leq \|x+y\|_P$, this condition expressing that the elements of P are positive. Given $x, y \in P$, one says that x is less than y (notation $x \leq y$) if there exists $z \in P$ such that $x + z = y$. By the cancellation property, if such a z exists, it is unique and we denote it as $y - x$. This subtraction obeys usual algebraic laws (when it is defined). Notice that if $x, y \in P$ satisfy $x + y = 0$ then since $\|x\|_P \leq \|x+y\|_P$, we have $x = 0$ (and of course also $y = 0$). Therefore, if $x \leq y$ and $y \leq x$ then $x = y$ and so $\leq$ is an order relation.

A (positive) *cone* is a positive pre-cone P whose unit ball $\mathcal{B}P = \{x \in P \mid \|x\|_P \leq 1\}$ is ω-order-complete in the sense that any increasing sequence of elements of $\mathcal{B}P$ has a least upper bound in $\mathcal{B}P$. In [10] we show how a notion of *stable* function on cones can be defined, which gives rise to a cartesian closed category.

[3] Meaning that $x + y = x' + y \Rightarrow x = x'$.

The following construction will be crucial in Section 3.2. Given a cone P and $x \in \mathcal{B}P$, we define the *local cone at* x as the set $P_x = \{u \in P \mid \exists \varepsilon > 0\ x + \varepsilon u \in \mathcal{B}P\}$. Equipped with the algebraic operations inherited from P, this set is clearly a $\mathbb{R}_{\geq 0}$-semi-ring. We equip it with the following norm: $\|u\|_{P_x} = \inf\{\varepsilon^{-1} \mid \varepsilon > 0 \text{ and } x + \varepsilon u \in \mathcal{B}P\}$ and then it is easy to check that P_x is indeed a cone. It is reduced to 0 exactly when x is maximal in $\mathcal{B}P$. In that case one has $\|x\|_P = 1$ but notice that the converse is not true in general.

1.2 Basic definitions on PCSs

Given an at most countable set I and $u, u' \in \overline{\mathbb{R}_{\geq 0}}^I$, we set $\langle u, u' \rangle = \sum_{i \in I} u_i u'_i \in \overline{\mathbb{R}_{\geq 0}}$. Given $P \subseteq \overline{\mathbb{R}_{\geq 0}}^I$, we define $P^\perp \subseteq \overline{\mathbb{R}_{\geq 0}}^I$ as $P^\perp = \{u' \in \overline{\mathbb{R}_{\geq 0}}^I \mid \forall u \in P\ \langle u, u' \rangle \leq 1\}$. Observe that if P satisfies $\forall a \in I\, \exists x \in P\ x_a > 0$ and $\forall a \in I\, \exists m \in \mathbb{R}_{\geq 0} \forall x \in P\ x_a \leq m$ then $P^\perp \in (\mathbb{R}_{\geq 0})^I$ and $P^\perp$ satisfies the same two properties.

A probabilistic pre-coherence space (pre-PCS) is a pair $X = (|X|, \mathsf{P}X)$ where $|X|$ is an at most countable set[4] and $\mathsf{P}X \subseteq \overline{\mathbb{R}_{\geq 0}}^{|X|}$ satisfies $\mathsf{P}X^{\perp\perp} = \mathsf{P}X$. A probabilistic coherence space (PCS) is a pre-PCS X such that $\forall a \in |X|\, \exists x \in \mathsf{P}X\ x_a > 0$ and $\forall a \in |X|\, \exists m \in \mathbb{R}_{\geq 0} \forall x \in \mathsf{P}X\ x_a \leq m$ so that $\mathsf{P}X \subseteq (\mathbb{R}_{\geq 0})^{|X|}$.

Given any PCS X we can define a cone $\overline{\mathsf{P}}X$ as follows:

$$\overline{\mathsf{P}}X = \{x \in (\mathbb{R}_{\geq 0})^{|X|} \mid \exists \varepsilon > 0\ \varepsilon x \in \mathsf{P}X\}$$

that we equip with the following norm: $\|x\|_{\overline{\mathsf{P}}X} = \inf\{r > 0 \mid x \in r\,\mathsf{P}X\}$ and then it is easy to check that $\mathcal{B}(\overline{\mathsf{P}}X) = \mathsf{P}X$. We simply denote this norm as $\|\_\|_X$.

Given $t \in \overline{\mathbb{R}_{\geq 0}}^{I \times J}$ considered as a matrix (where I and J are at most countable sets) and $u \in \overline{\mathbb{R}_{\geq 0}}^I$, we define $t\,u \in \overline{\mathbb{R}_{\geq 0}}^J$ by $(t\,u)_j = \sum_{i \in I} t_{i,j} u_i$ (usual formula for applying a matrix to a vector), and if $s \in \overline{\mathbb{R}_{\geq 0}}^{J \times K}$ we define the product $s\,t \in \overline{\mathbb{R}_{\geq 0}}^{I \times K}$ of the matrix s and t as usual by $(s\,t)_{i,k} = \sum_{j \in J} t_{i,j} s_{j,k}$. This is an associative operation.

Let X and Y be PCSs, a morphism from X to Y is a matrix $t \in (\mathbb{R}_{\geq 0})^{|X| \times |Y|}$ such that $\forall x \in \mathsf{P}X\ t\,x \in \mathsf{P}Y$. It is clear that the identity matrix is a morphism from X to X and that the matrix product of two morphisms is a morphism and therefore, PCS equipped with this notion of morphism form a category **Pcoh**.

The condition $t \in \mathbf{Pcoh}(X, Y)$ is equivalent to $\forall x \in \mathsf{P}X\, \forall y' \in \mathsf{P}Y^\perp\ \langle t\,x, y' \rangle \leq 1$ but $\langle t\,x, y' \rangle = \langle t, x \otimes y' \rangle$ where $(x \otimes y')_{(a,b)} = x_a y'_b$. This strongly suggests to introduce a construction $X \otimes Z$, given two PCSs X and Z, by setting $|X \otimes Z| = |X| \times |Z|$ and $\mathsf{P}(X \otimes Z) = \{x \otimes z \mid x \in \mathsf{P}X \text{ and } z \in \mathsf{P}Z\}^{\perp\perp}$ where $(x \otimes z)_{(a,c)} = x_a z_c$. Then it is easy to see that $X \otimes Z$ is not only a pre-PCS, but actually a PCS and that we have equipped in that way the category **Pcoh** with a symmetric monoidal structure for which it is $*$-autonomous wrt. a dualizing object $\perp = 1 = (\{*\}, [0, 1])$ (it is at the same time the unit of $\otimes$ and $X^\perp \simeq (X \multimap \perp)$ up to a trivial iso).

The category **Pcoh** is cartesian: if $(X_i)_{i \in I}$ is an at most countable family of PCSs, then $(\&_{i \in I} X_i, (\pi_i)_{i \in I})$ is the cartesian product of the X_is, with $|\&_{i \in I} X_i| = \cup_{i \in I} \{i\} \times |X_i|$, $(\pi_i)_{(j,a),a'} = 1$ if $i = j$ and $a = a'$ and $(\pi_i)_{(j,a),a'} = 0$ otherwise, and $x \in \mathsf{P}(\&_{i \in I} X_i)$ if $\pi_i\, x \in \mathsf{P}X_i$ for each $i \in I$ (for $x \in (\mathbb{R}_{\geq 0})^{|\&_{i \in I} X_i|}$). Given $t_i \in \mathbf{Pcoh}(Y, X_i)$, the unique morphism $t = \langle t_i \rangle_{i \in I} \in \mathbf{Pcoh}(Y, \&_{i \in I} X_i)$ such that $\pi_i\, t = t_i$ is simply defined by $t_{b,(i,a)} = (t_i)_{a,b}$. The

[4] This restriction is not technically necessary, but very meaningful from a philosophic point of view; the non countable case should be handled via measurable spaces and then one has to consider more general objects as in [10] for instance.

dual operation $\oplus_{i\in I} X_i$, which is a coproduct, is characterized by $|\oplus_{i\in I} X_i| = \cup_{i\in I}\{i\}\times|X_i|$ and $x \in \mathsf{P}(\oplus_{i\in I} X_i)$ and $\sum_{i\in I}\|\pi_i\, x\|_{X_i} \leq 1$. A particular case is $\mathsf{N} = \oplus_{n\in\mathbb{N}} X_n$ where $X_n = 1$ for each n. So that $|\mathsf{N}| = \mathbb{N}$ and $x \in (\mathbb{R}_{\geq 0})^{\mathbb{N}}$ belongs to PN if $\sum_{n\in\mathbb{N}} x_n \leq 1$ (that is, x is a sub-probability distribution on $\mathbb{N}$). There are successor and predecessor morphisms $\overline{\mathsf{suc}}, \overline{\mathsf{pred}} \in \mathbf{Pcoh}(\mathsf{N},\mathsf{N})$ given by $\overline{\mathsf{suc}}_{n,n'} = \delta_{n+1,n'}$ and $\overline{\mathsf{pred}}_{n,n'} = 1$ if $n = n' = 0$ or $n = n'+1$ (and $\overline{\mathsf{pred}}_{n,n'} = 0$ in all other cases). An element of $\mathbf{Pcoh}(\mathsf{N},\mathsf{N})$ is a (sub)stochastic matrix and our model should be understood as this kind of representation of programs.

As to the exponentials, one sets $|!X| = \mathcal{M}_{\mathrm{fin}}(|X|)$ and $\mathsf{P}(!X) = \{x^! \mid x \in \mathsf{P}X\}^{\perp\perp}$ where, given $\mu \in \mathcal{M}_{\mathrm{fin}}(|X|)$, $x^!_\mu = x^\mu = \prod_{a\in|X|} x_a^{\mu(a)}$. Then given $t \in \mathbf{Pcoh}(X,Y)$, one defines $!t \in \mathbf{Pcoh}(!X,!Y)$ in such a way that $!t\, x^! = (t\, x)^!$ (the precise definition is not relevant here; it is completely determined by this equation). We do not need here to specify the monoidal comonad structure of this exponential. The resulting cartesian closed category[5] $\mathbf{Pcoh}_!$ can be seen as a category of functions (actually, of stable functions as proved in [3]). Indeed, a morphism $t \in \mathbf{Pcoh}_!(X,Y) = \mathbf{Pcoh}(!X,Y) = \mathsf{P}(!X \multimap Y)$ is completely characterized by the associated function $\widehat{t}\colon \mathsf{P}X \to \mathsf{P}Y$ such that $\widehat{t}(x) = t\, x^! = \left(\sum_{\mu\in|!X|} t_{\mu,b} x^\mu\right)_{b\in|Y|}$ so that we consider morphisms as power series (they are in particular monotonic and Scott continuous functions $\mathsf{P}X \to \mathsf{P}Y$). In this cartesian closed category, the product of a family $(X_i)_{i\in I}$ is $\&_{i\in I} X_i$ (written X^I if $X_i = X$ for all i), which is compatible with our viewpoint on morphisms as functions since $\mathsf{P}(\&_{i\in I} X_i) = \prod_{i\in I} \mathsf{P}X_i$ up to trivial iso. The object of morphisms from X to Y is $!X \multimap Y$ with evaluation mapping $(t,x) \in \mathsf{P}(!X\multimap Y)\times\mathsf{P}X$ to $\widehat{t}(x)$ that we simply denote as $t(x)$ from now on. The well defined function $\mathsf{P}(!X\multimap X)\to\mathsf{P}X$ which maps t to $\sup_{n\in\mathbb{N}} t^n(0)$ is a morphism of $\mathbf{Pcoh}_!$ (and thus can be described as a power series in the vector $t = (t_{m,a})_{m\in\mathcal{M}_{\mathrm{fin}}(|X|),a\in|X|}$) by standard categorical considerations using cartesian closeness: it provides us with fixed point operators at all types.

2 Probabilistic PCF, time expectation and derivatives

We introduce now the probabilistic functional programming language considered in this paper. The operational semantics is presented using elementary probability theoretic tools.

2.1 The core language

The types and terms are given by

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

See Fig. 1 for the typing rules, with typing contexts $\Gamma = (x_1:\sigma_1,\ldots,x_n:\sigma_n)$.

2.1.1 Denotational semantics

We survey briefly the interpretation of pPCF in PCSs thoroughly described in [10]. Types are interpreted by $[\![\iota]\!] = \mathsf{N}$ and $[\![\sigma\Rightarrow\tau]\!] = ![\![\sigma]\!] \multimap [\![\tau]\!]$. Given $M \in \mathsf{pPCF}$ such that $\Gamma \vdash M : \sigma$ (with $\Gamma = (x_1:\sigma_1,\ldots,x_k:\sigma_k)$) one defines $[\![M]\!]_\Gamma \in \mathbf{Pcoh}_!(\&_{i=1}^k [\![\sigma_i]\!], [\![\sigma]\!])$ (a

[5] This is the Kleisli category of "!" which has actually a comonad structure that we do not make explicit here, again we refer to [5, 10].

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

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

$$\frac{\Gamma, x : \sigma \vdash M : \tau}{\Gamma \vdash \lambda x^{\sigma}\, M : \sigma \Rightarrow \tau} \qquad \frac{\Gamma \vdash M : \sigma \Rightarrow \tau \qquad \Gamma \vdash N : \sigma}{\Gamma \vdash (M)N : \tau} \qquad \frac{\Gamma \vdash M : \sigma \Rightarrow \sigma}{\Gamma \vdash \mathsf{fix}(M) : \sigma} \qquad \frac{r \in [0,1] \cap \mathbb{Q}}{\Gamma \vdash \mathsf{coin}(r) : \iota}$$

$$\frac{}{\iota \vdash \varepsilon} \qquad \frac{\vdash M : \sigma \qquad \tau \vdash \pi}{\sigma \Rightarrow \tau \vdash \mathsf{arg}(M) \cdot \pi} \qquad \frac{\iota \vdash \pi}{\iota \vdash \mathsf{succ} \cdot \pi} \qquad \frac{\iota \vdash \pi}{\iota \vdash \mathsf{pred} \cdot \pi}$$

$$\frac{\vdash N : \sigma \qquad \vdash P : \sigma \qquad \sigma \vdash \pi}{\iota \vdash \mathsf{if}(N, P) \cdot \pi} \qquad \frac{x : \iota \vdash N : \sigma \qquad \sigma \vdash \pi}{\iota \vdash \mathsf{let}(x, N) \cdot \pi}$$

Figure 1 Typing rules for pPCF terms and stacks.

"Kleisli morphism") that we see as a function $\prod_{i=1}^{k} \mathsf{P}[\![\sigma_i]\!] \to \mathsf{P}[\![\sigma]\!]$ as explained in Section 1.2. For instance $[\![x_i]\!]_\Gamma(\vec{u}) = u_i$, $[\![\underline{n}]\!]_\Gamma(\vec{u}) = \overline{n}$ (remember that $\overline{n} \in \mathsf{PN}$ is defined by $\overline{n}_i = \delta_{n,i}$), $[\![\mathsf{succ}(M)]\!]_\Gamma(\vec{u}) = \overline{\mathsf{suc}}\,[\![M]\!]_\Gamma(\vec{u})$ and similarly for $\mathsf{pred}(M)$, more importantly

$$[\![\mathsf{coin}(r)]\!]_\Gamma(\vec{u}) = r\,\overline{0} + (1-r)\,\overline{1} \qquad [\![\mathsf{let}(x, M, N)]\!]_\Gamma = \sum_{n \in \mathbb{N}} [\![M]\!]_\Gamma(\vec{u})_n\, [\![N\,[\underline{n}/x]]\!]_\Gamma(\vec{u})$$

$$[\![\mathsf{if}(M, N, P)]\!]_\Gamma(\vec{u}) = [\![M]\!]_\Gamma(\vec{u})_0\, [\![N]\!]_\Gamma(\vec{u}) + \left(\sum_{n \in \mathbb{N}} [\![M]\!]_\Gamma(\vec{u})_{n+1}\right) [\![P]\!]_\Gamma(\vec{u})\,.$$

Application and λ-abstraction are interpreted as usual in a cartesian closed category (in particular $[\![(M)N]\!]_\Gamma(\vec{u}) = ([\![M]\!]_\Gamma(\vec{u}))([\![N]\!]_\Gamma(\vec{u}))$). Last $[\![\mathsf{fix}(M)]\!]_\Gamma(\vec{u}) = \sup_{n \in \mathbb{N}} ([\![M]\!]_\Gamma(\vec{u}))^n(0)$.

2.1.2 Operational semantics

In former papers we have presented the operational semantics of pPCF as a discrete Markov chain on states which are the closed terms of pPCF. This Markov chain implements the standard weak head reduction strategy of PCF which is deterministic for ordinary PCF but features branchings in pPCF because of the $\mathsf{coin}(r)$ construct (see [10]). Here we prefer another, though strictly equivalent, presentation of this operational semantics, based on an environment-free Krivine Machine (thus handling states which are pairs made of a closed term and a closed stack) further parameterized by an element of $\{0,1\}^{<\omega}$ to be understood as a "random tape" prescribing the values taken by the $\mathsf{coin}(r)$ terms during the execution of states. We present this machine as a partial function taking a state s, a random tape α and returning an element of $[0,1]$ to be understood as the probability that the sequence α of 0/1 choices occurs during the execution of s. We allow only execution of ground type states and accept $\underline{0}$ as the only terminating value: a completely arbitrary choice, sufficient for our purpose in this paper. Also, we insist that a terminating computation from (s, α) completely consumes the random tape α. These choices allow to fit within a completely standard discrete probability setting.

Given an extension Λ of pPCF (with the same format for typing rules), we define the associated language of stacks (called Λ-stacks).

$$\pi := \varepsilon \mid \mathsf{arg}(M) \cdot \pi \mid \mathsf{succ} \cdot \pi \mid \mathsf{pred} \cdot \pi \mid \mathsf{if}(N, P) \cdot \pi \mid \mathsf{let}(x, N) \cdot \pi$$

where M and N range over Λ. A stack typing judgment is of shape $\sigma \vdash \pi$ (meaning that it takes a term of type σ and returns an integer) and the typing rules are given in Fig. 1.

$$\begin{array}{ll}
\mathsf{Ev}(\langle \mathsf{let}(x,M,N),\pi\rangle,\alpha) = \mathsf{Ev}(\langle M,\mathsf{let}(x,N)\cdot\pi\rangle,\alpha) & \mathsf{Ev}(\langle \lambda x^{\sigma}\, M,\mathsf{arg}(N)\cdot\pi\rangle,\alpha) = \mathsf{Ev}(\langle M\,[N/x]\,,\pi\rangle,\alpha)\\
\mathsf{Ev}(\langle \underline{n},\mathsf{let}(x,N)\cdot\pi\rangle,\alpha) = \mathsf{Ev}(\langle N\,[\underline{n}/x]\,,\pi\rangle,\alpha) & \mathsf{Ev}(\langle \mathsf{fix}(M),\pi\rangle,\alpha) = \mathsf{Ev}(\langle M,\mathsf{arg}(\mathsf{fix}(M))\cdot\pi\rangle,\alpha)\\
\mathsf{Ev}(\langle \mathsf{if}(M,N,P),\pi\rangle) = \mathsf{Ev}(\langle M,\mathsf{if}(N,P)\cdot\pi\rangle,\alpha) & \mathsf{Ev}(\langle \mathsf{coin}(r),\pi\rangle,\langle 0\rangle\alpha) = \mathsf{Ev}(\langle \underline{0},\pi\rangle,\alpha)\cdot r\\
\mathsf{Ev}(\langle \underline{0},\mathsf{if}(N,P)\cdot\pi\rangle,\alpha) = \mathsf{Ev}(\langle N,\pi\rangle,\alpha) & \mathsf{Ev}(\langle \mathsf{coin}(r),\pi\rangle,\langle 1\rangle\alpha) = \mathsf{Ev}(\langle \underline{1},\pi\rangle,\alpha)\cdot(1-r)\\
\mathsf{Ev}(\langle \underline{n+1},\mathsf{if}(N,P)\cdot\pi\rangle,\alpha) = \mathsf{Ev}(\langle P,\pi\rangle,\alpha) & \mathsf{Ev}(\langle \underline{0},\varepsilon\rangle,\langle\rangle) = 1
\end{array}$$

Figure 2 The pPCF Krivine Machine.

A *state* is a pair $\langle M,\pi\rangle$ (where we say that M is *in head position*) such that $\vdash M:\sigma$ and $\sigma\vdash\pi$ for some (uniquely determined) type σ, let S be the set of states. Let $\mathcal{C}_0=\{0,1\}^{<\omega}$ be the set of finite lists of booleans (random tapes), we define a *partial* function $\mathsf{Ev}:\mathsf{S}\times\mathcal{C}_0\to\mathbb{R}_{\geq 0}$ in Fig. 2. Let $\mathcal{D}(s)$ be the set of all $\alpha\in\mathcal{C}_0$ such that $\mathsf{Ev}(s,\alpha)$ is defined. When $\alpha\in\mathcal{D}(s)$, the number $\mathsf{Ev}(s,\alpha)\in[0,1]$ is the probability that the random tape α occurs during the execution. When all coins are fair (all the values of the parameters r are $1/2$), this probability is $2^{-\mathsf{len}(\alpha)}$. The sum of these (possibly infinitely many) probabilities is ≤ 1. For fitting within a standard probabilistic setting, we define a total probability distribution $\mathsf{Ev}(s):\mathcal{C}_0\to[0,1]$ as follows

$$\mathsf{Ev}(s)(\alpha)=\begin{cases}\mathsf{Ev}(s,\beta) & \text{if } \alpha=\langle 0\rangle\beta \text{ and } \beta\in\mathcal{D}(s)\\ 1-\sum_{\beta\in\mathcal{D}(s)}\mathsf{Ev}(s,\beta) & \text{if } \alpha=\langle 1\rangle\\ 0 & \text{in all other cases}\end{cases}$$

Let $\mathbb{P}_s$ be the associated probability measure[6] (we are in a discrete setting so simply $\mathbb{P}_s(A)=\sum_{\alpha\in A}\mathsf{Ev}(s)(\alpha)$ for all $A\subseteq\mathcal{C}_0$).

The event $(s\downarrow\underline{0})=\langle 0\rangle\mathcal{D}(s)$ is the set of all random tapes (up to 0-prefixing) making s reduce to $\underline{0}$. Its probability is $\mathbb{P}_s(s\downarrow\underline{0})=\sum_{\beta\in\mathcal{D}(s)}\mathsf{Ev}(s,\beta)$. In the case $s=\langle M,\varepsilon\rangle$ (with $\vdash M:\iota$) this probability is *exactly the same* as the probability of M to reduce to $\underline{0}$ in the Markov chain setting of [10] (see *e.g.* [1] for more details on the connection between these two kinds of operational semantics). So the Adequacy Theorem of [10] can be expressed as follows.

▶ **Theorem 1.** *Let $M\in\mathsf{pPCF}$ with $\vdash M:\iota$. Then $[\![M]\!]_0=\mathbb{P}_{\langle M,\varepsilon\rangle}(\langle M,\varepsilon\rangle\downarrow\underline{0})$.*

We use sometimes $\mathbb{P}(M\downarrow\underline{0})$ as an abbreviation for $\mathbb{P}_{\langle M,\varepsilon\rangle}(\langle M,\varepsilon\rangle\downarrow\underline{0})$.

2.2 Probabilistic PCF with labels and the associated random variables

In order to count the number of times a given subterm N of a closed term M of type ι is used (that is, arrives in head position) during the execution of $\langle M,\varepsilon\rangle$ in the Krivine machine of Section 2.1.2, we extend pPCF into $\mathsf{pPCF}_{\mathsf{lab}}$ by adding a term labeling construct N^l. The typing rule for this new construct is simply $\dfrac{\Gamma\vdash N:\sigma}{\Gamma\vdash N^l:\sigma}$. Of course $\mathsf{pPCF}_{\mathsf{lab}}$-stacks involve now such labeled terms but their syntax is not extended otherwise; let $\mathsf{S}_{\mathsf{lab}}$ be the corresponding set of states. Then we define a partial function $\mathsf{Ev}_{\mathsf{lab}}:\mathsf{S}_{\mathsf{lab}}\times\mathcal{C}_0\to\mathcal{M}_{\mathrm{fin}}(\mathcal{L})$ exactly as Ev apart for the following cases,

$$\begin{array}{ll}
\mathsf{Ev}_{\mathsf{lab}}(\langle M^l,\pi\rangle,\alpha)=\mathsf{Ev}_{\mathsf{lab}}(\langle M,\pi\rangle,\alpha)+[l] & \\
\mathsf{Ev}_{\mathsf{lab}}(\langle \mathsf{coin}(r),\pi\rangle,\langle i\rangle\alpha)=\mathsf{Ev}_{\mathsf{lab}}(\langle \underline{i},\pi\rangle,\alpha) & \mathsf{Ev}_{\mathsf{lab}}(\langle \underline{0},\varepsilon\rangle,\langle\rangle)=0 \quad \text{the empty multiset.}
\end{array}$$

[6] The choice of accumulating on $\langle 1\rangle$ all the complementary probability is completely arbitrary and has no impact on the result we prove because all the events of interest for us will be subsets of $\langle 0\rangle\mathcal{C}_0\subset\mathcal{C}_0$.

When applied to $\langle M, \varepsilon\rangle$, this function counts how often labeled subterms of M arrive in head position during the reduction; this number depends of course on the random tape provided as argument together with the state. The result is a finite multiset of labels.

Let $\mathcal{D}_{\mathsf{lab}}(s)$ be the set of αs such that $\mathsf{Ev}_{\mathsf{lab}}(s,\alpha)$ is defined. Defining $\underline{s} \in \mathsf{S}$ as s stripped from its labels, we clearly have $\mathcal{D}_{\mathsf{lab}}(s) = \mathcal{D}(\underline{s})$. We define a r.v.[7] $\mathsf{Ev}_{\mathsf{lab}}(s) : \mathcal{C}_0 \to \mathcal{M}_{\mathrm{fin}}(\mathcal{L})$ by

$$\mathsf{Ev}_{\mathsf{lab}}(s)(\alpha) = \begin{cases} \mathsf{Ev}_{\mathsf{lab}}(s,\beta) & \text{if } \alpha = \langle 0\rangle\beta \text{ and } \beta \in \mathcal{D}(s) \\ 0 & \text{in all other cases.} \end{cases}$$

Let $l \in \mathcal{L}$ and let $\mathsf{Ev}_{\mathsf{lab}}(s)_l : \mathcal{C}_0 \to \mathbb{N}$ be the *integer* r.v. defined by $\mathsf{Ev}_{\mathsf{lab}}(s)_l(\alpha) = \mathsf{Ev}_{\mathsf{lab}}(s)(\alpha)(l)$. Its expectation is

$$\begin{aligned} \mathbb{E}(\mathsf{Ev}_{\mathsf{lab}}(s)_l) &= \sum_{n\in\mathbb{N}} n\,\mathbb{P}_s(\mathsf{Ev}_{\mathsf{lab}}(s)_l = n) = \sum_{n\in\mathbb{N}} n \sum_{\substack{\mu\in\mathcal{M}_{\mathrm{fin}}(\mathcal{L}) \\ \mu(l)=n}} \mathbb{P}_s(\mathsf{Ev}_{\mathsf{lab}}(s) = \mu) \\ &= \sum_{\mu\in\mathcal{M}_{\mathrm{fin}}(L)} \mu(l)\mathbb{P}_s(\mathsf{Ev}_{\mathsf{lab}}(s) = \mu)\,. \end{aligned} \tag{1}$$

This is the expected number of occurrences of l-labeled subterms of s arriving in head position during successful executions of s. It is more meaningful to condition this expectation under convergence of the execution of s (that is, under the event $\underline{s} \downarrow \underline{0}$). We have $\mathbb{E}(\mathsf{Ev}_{\mathsf{lab}}(s)_l \mid \underline{s} \downarrow \underline{0}) = \mathbb{E}(\mathsf{Ev}_{\mathsf{lab}}(s)_l)/\mathbb{P}_{\underline{s}}(\underline{s} \downarrow \underline{0})$ as the r.v. $\mathsf{Ev}_{\mathsf{lab}}(s)_l$ vanishes outside the event $s \downarrow \underline{0}$ since $\mathcal{D}_{\mathsf{lab}}(s) = \mathcal{D}(\underline{s})$.

Our goal now is to extract this expectation from the denotational semantics of a term M such that $\vdash M : \iota$, which contains labeled subterms, or rather of a term suitably definable from M. The general idea is to replace in M each N^l (where N has type σ) with $\mathsf{if}(x_l, N, \Omega^\sigma)$ where $\vec{x} = (x_l)_{l\in L}$ (for some finite subset L of $\mathcal{L}$ containing all the labels occurring in M) is a family of pairwise distinct variables of type ι and $\Omega^\sigma = \mathsf{fix}(\lambda x^\sigma\, x)$. We obtain in that way a term $\mathsf{sp}_{\vec{x}}M$ whose semantics $[\![\mathsf{sp}_{\vec{x}}M]\!]_{\vec{x}}$ is an element of $\mathbf{Pcoh}_!(\mathsf{N}^L, \mathsf{N})$ that we can consider as an analytic function $(\mathsf{PN})^L \to \mathsf{PN}$ and which therefore induces an analytic function $f : [0,1]^L \to [0,1]$ by $f(\vec{r}) = [\![M']\!]((r_l\overline{0})_{l\in L})_0$ (where $\vec{r}\,\overline{0} = (r_l\,\overline{0})_{l\in L} \in \mathsf{PN}^L$ for $\vec{r} \in [0,1]^L$). Our main claim is that the expectation of the number of uses of subterms labeled by l is $\frac{\partial f(\vec{r})}{\partial r_l}(1,\dots,1)$.

In order to reduce this problem to Theorem 1, we need a further "Krivine machine" with has as many random tapes as elements of L (plus one for the plain $\mathsf{coin}(\_)$ constructs occurring in M).

2.3 Probabilistic PCF with labeled coins

Let $\mathsf{pPCF}_{\mathsf{lc}}$ be pPCF extended with a construct $\mathsf{lcoin}(l,r)$ typed as $\dfrac{r \in [0,1]\cap\mathbb{Q} \text{ and } l \in \mathcal{L}}{\Gamma \vdash \mathsf{lcoin}(l,r) : \iota}$

This language features the usual $\mathsf{coin}(r)$ construct for probabilistic choice as well as a supply of identical constructs labeled by $\mathcal{L}$ that we will use to simulate the counting of Section 2.2. Of course $\mathsf{pPCF}_{\mathsf{lc}}$-stacks involve now terms with labeled coins but their syntax is not extended otherwise; let S_{lc} be the corresponding set of states. We use $\mathsf{lab}(M)$ for the set of labels occurring in M (and similarly $\mathsf{lab}(s)$ for $s \in \mathsf{S}_{\mathsf{lc}}$). Given a *finite* subset L of $\mathcal{L}$, we use $\mathsf{pPCF}_{\mathsf{lc}}(L)$ for the set of terms M such that $\mathsf{lab}(M) \subseteq L$ and we define similarly $\mathsf{S}_{\mathsf{lc}}(L)$. We also use the similar notations $\mathsf{pPCF}_{\mathsf{lab}}(L)$ and $\mathsf{S}_{\mathsf{lab}}(L)$.

[7] That is, simply, a function since we are in a discrete probability setting.

The partial function $\mathsf{Ev}_{\mathsf{lc}} : \mathsf{S}_{\mathsf{lc}}(L) \times \mathcal{C}_0 \times \mathcal{C}_0^L \to \mathbb{R}_{\geq 0}$ is defined exactly as Ev (for the unlabeled $\mathsf{coin}(r)$, we use only the first parameter in $\mathcal{C}_0$), extended by the following rules:

$$\mathsf{Ev}_{\mathsf{lc}}(\langle \mathsf{lcoin}(l,r), \pi\rangle, \alpha, \vec{\beta}) = \begin{cases} \mathsf{Ev}_{\mathsf{lc}}(\langle \underline{0}, \pi\rangle, \alpha, \vec{\beta}\,[\gamma/l]) \cdot r & \text{if } \beta(l) = \langle 0\rangle\gamma \\ \mathsf{Ev}_{\mathsf{lc}}(\langle \underline{1}, \pi\rangle, \alpha, \vec{\beta}\,[\gamma/l]) \cdot (1-r) & \text{if } \beta(l) = \langle 1\rangle\gamma \end{cases}$$

where $\vec{\beta} = (\beta(l))_{l\in L}$ stands for an L-indexed family of elements of $\mathcal{C}_0$ and $\vec{\beta}\,[\gamma/l]$ is the family $\vec{\delta}$ such that $\delta(l') = \beta(l')$ if $l' \neq l$ and $\delta(l) = \gamma$. We define $\mathcal{D}_{\mathsf{lc}}(s) \subseteq \mathcal{C}_0 \times \mathcal{C}_0^L$ as the domain of the partial function $\mathsf{Ev}_{\mathsf{lc}}(s, \_, \_)$. Let $\underline{s} \in \mathsf{S}$ be obtained by stripping s from its labels (so that $\underline{\mathsf{lcoin}(l,r)} = \mathsf{coin}(r)$). And $\underline{M} \in \mathsf{pPCF}$ is defined similarly.

▶ **Lemma 2.** *For all* $s \in \mathsf{S}_{\mathsf{lc}}(L)$

$$\mathbb{P}_{\underline{s}}(\underline{s} \downarrow \underline{0}) = \sum_{(\alpha,\vec{\beta})\in\mathcal{D}_{\mathsf{lc}}(s)} \mathsf{Ev}_{\mathsf{lc}}(s, \alpha, \vec{\beta})\,.$$

Proof. (Sketch) With each $(\alpha, \vec{\beta}) \in \mathcal{D}_{\mathsf{lc}}(s)$ we can associate a uniquely defined $\eta_s(\alpha, \vec{\beta}) \in \mathcal{D}(\underline{s})$ which is a shuffle of α and of the $\beta(l)$'s (for $l \in L$) such that $\mathsf{Ev}_{\mathsf{lc}}(s, \alpha, \vec{\beta}) = \mathsf{Ev}(s, \eta_s(\alpha, \vec{\beta}))$, uniquely determined by the run of $(s, \alpha, \vec{\beta})$ in the "machine" $\mathsf{Ev}_{\mathsf{lab}}$. This mapping η_s (which is defined much like $\mathsf{Ev}_{\mathsf{lc}}(s, \_, \_)$) is easily seen to be bijective. ◀

2.3.1 Spying labeled terms in pPCF

We arrive to the last step, which consists in turning a *closed* labeled term M (with labels in the finite set L) into the already mentioned term $\mathsf{sp}_{\vec{x}}(M)$, defined in such a way that $[\![\underline{\mathsf{lc}_{\vec{r}}(M)}]\!]$ has a simple expression in terms of $\mathsf{sp}_{\vec{x}}(M)$ (Lemma 5), allowing to relate the coefficients of the power series interpreting $\mathsf{sp}_{\vec{x}}(M)$ in terms of probability of reduction of the machine $\mathsf{Ev}_{\mathsf{lab}}$ with given resulting multisets of labels (Equation (2)). This in turn is the key to the proof of Theorem 6.

Given $\vec{r} = (r_l)_{l\in L} \in (\mathbb{Q} \cap [0,1])^L$, we define a (type preserving) translation $\mathsf{lc}_{\vec{r}} : \mathsf{pPCF}_{\mathsf{lab}}(L) \to \mathsf{pPCF}_{\mathsf{lc}}$ by induction on terms. For all term constructs but labeled terms, the transformation does nothing (for instance $\mathsf{lc}_{\vec{r}}(x) = x$, $\mathsf{lc}_{\vec{r}}(\lambda x^\sigma\, M) = \lambda x^\sigma\, \mathsf{lc}_{\vec{r}}(M)$ etc), the only non trivial case being $\mathsf{lc}_{\vec{r}}(M^l) = \mathsf{if}(\mathsf{lcoin}(l, r_l), \mathsf{lc}_{\vec{r}}(M), \Omega^\sigma)$ where σ is the type[8] of M.

▶ **Lemma 3.** *Let* $s \in \mathsf{S}_{\mathsf{lab}}(L)$. *Then* $\mathcal{D}_{\mathsf{lab}}(s) = \mathcal{D}(\underline{s})$, $\mathcal{D}_{\mathsf{lc}}(\mathsf{lc}_{\vec{r}}(s)) = \{(\alpha, (\langle 0\rangle^{\mathsf{Ev}_{\mathsf{lab}}(s,\alpha)(l)})_{l\in L}) \mid \alpha \in \mathcal{D}(\underline{s})\}$ *and* $\mathsf{Ev}_{\mathsf{lc}}(\mathsf{lc}_{\vec{r}}(s), \alpha, \langle 0\rangle^{\mathsf{Ev}_{\mathsf{lab}}(s,\alpha)(l)}) = \mathbb{P}_{\underline{s}}(\{\langle 0\rangle\alpha\})(\vec{r})^{\mathsf{Ev}_{\mathsf{lab}}(s,\alpha)}$.

Of course $\langle 0\rangle^n$ stands for the sequence $\langle 0, \ldots, 0\rangle$ (with n occurrences of 0). The proof is by induction on the length of α and boils down to the observation that $\mathcal{D}(\langle \Omega^\sigma, \pi\rangle) = \emptyset$ for any (well typed) stack π. Remember that $\mathbb{P}_{\underline{s}}(\{\langle 0\rangle\alpha\}) = \mathsf{Ev}(\underline{s}, \alpha)$ and that $(\vec{r})^\mu = \prod_{l\in l} r_l^{\mu(l)}$ for all $\mu \in \mathcal{M}_{\mathrm{fin}}(L)$.

We consider a last type preserving translation from $\mathsf{pPCF}_{\mathsf{lab}}(L)$ to pPCF: let $\vec{x}$ be a L-indexed family of pairwise distinct variables (that we identify with the typing context $(x_l : \iota)_{l\in L}$). If $M \in \mathsf{pPCF}_{\mathsf{lab}}(L)$ with $\Gamma \vdash M : \sigma$ (assuming that no free variable of M occurs in $\vec{x}$) we define $\mathsf{sp}_{\vec{x}}(M)$ with $\Gamma, \vec{x} \vdash \mathsf{sp}_{\vec{x}}(M) : \sigma$ by induction on M. The unique non trivial case is $\mathsf{sp}_{\vec{x}}(M^l) = \mathsf{if}(x_l, \mathsf{sp}_{\vec{x}}(M), \Omega^\sigma)$ where σ is the type of M.

[8] *A priori* this type is known only if we know the type of the free variables of M, so to be more precise this translation should be specified in a given typing context; this can easily be fixed by adding a further parameter to lc at the price of heavier notations.

▶ **Lemma 4.** *Let $M \in \mathsf{pPCF}_{\mathsf{lab}}(L)$ with $\vdash M : \sigma$. If $\vec{\rho} \in \mathcal{M}_{\mathrm{fin}}(\mathbb{N})^L = \mathcal{M}_{\mathrm{fin}}(L \times \mathbb{N})$ and $a \in |[\![\sigma]\!]|$ satisfy $([\![\mathsf{sp}_{\vec{x}}(M)]\!]_{\vec{x}})_{(\vec{\rho},a)} \neq 0$ then $\rho_l(n) \neq 0 \Rightarrow n = 0$.*

The proof is a simple induction on M (of course we also have to consider open terms) and uses the fact that $[\![\Omega^\sigma]\!] = 0$.

Given $\mu \in \mathcal{M}_{\mathrm{fin}}(L)$, we use $\mu\,[0]$ for the element ρ of $\mathcal{M}_{\mathrm{fin}}(\mathbb{N})^L$ such that $\rho_l(n) = \mu(l)$ if $n = 0$ and $\rho_l(n) = 0$ otherwise.

▶ **Lemma 5.** *Let $\vec{r} \in (\mathbb{Q} \cap [0,1])^L$ and $M \in \mathsf{pPCF}_{\mathsf{lab}}(L)$ with $\vdash M : \sigma$. Then $[\![\mathsf{sp}_{\vec{x}}(M)]\!]_{\vec{x}}(\vec{r}\,\overline{0}) = [\![\underline{\mathsf{lc}_{\vec{r}}(M)}]\!]$.*

Easy induction on M based on the fact that $[\![\mathsf{coin}(r)]\!] = r\overline{0} + (1-r)\overline{1}$ (again, one needs a more general statement involving open terms).

By Lemma 5, $[\![\underline{\mathsf{lc}_{\vec{r}}(M)}]\!]_0 = \sum_{\mu \in \mathcal{M}_{\mathrm{fin}}(L)} ([\![\mathsf{sp}_{\vec{x}}(M)]\!]_{\vec{x}})_{(\mu\,[0],0)} (\vec{r})^\mu$. By Theorem 1, we have

$$\begin{aligned}
[\![\underline{\mathsf{lc}_{\vec{r}}(M)}]\!]_0 &= \mathbb{P}_{\underline{\mathsf{lc}_{\vec{r}}(\langle M,\varepsilon\rangle)}}(\underline{\mathsf{lc}_{\vec{r}}(\langle M,\varepsilon\rangle)} \downarrow \underline{0}) \\
&= \sum_{(\alpha,\vec{\beta}) \in \mathcal{D}_{\mathsf{lc}}(\mathsf{lc}_{\vec{r}}(\langle M,\varepsilon\rangle))} \mathsf{Ev}_{\mathsf{lc}}(\mathsf{lc}_{\vec{r}}(\langle M,\varepsilon\rangle), \alpha, \vec{\beta}) \quad \text{by Lemma 2} \\
&= \sum_{\alpha \in \mathcal{D}(\underline{\langle M,\varepsilon\rangle})} \mathsf{Ev}(\underline{\langle M,\varepsilon\rangle}, \alpha) \prod_{l \in L} r_l^{\mathsf{Ev}_{\mathsf{lab}}(\langle M,\varepsilon\rangle,\alpha)(l)} \quad \text{by Lemma 3} \\
&= \sum_{\mu \in \mathcal{M}_{\mathrm{fin}}(L)} \left(\sum_{\substack{\alpha \in \langle 0\rangle \mathcal{C}_0 \\ \mathsf{Ev}_{\mathsf{lab}}(\langle M,\varepsilon\rangle)(\alpha) = \mu}} \mathsf{Ev}(\underline{\langle M,\varepsilon\rangle})(\alpha) \right) (\vec{r})^\mu
\end{aligned}$$

and since this holds for all $\vec{r} \in (\mathbb{Q} \cap [0,1])^L$, we must have, for all $\mu \in \mathcal{M}_{\mathrm{fin}}(L)$,

$$([\![\mathsf{sp}_{\vec{x}}(M)]\!]_{\vec{x}})_{(\mu\,[0],0)} = \sum_{\substack{\alpha \in \langle 0\rangle \mathcal{C}_0 \\ \mathsf{Ev}_{\mathsf{lab}}(\langle M,\varepsilon\rangle)(\alpha) = \mu}} \mathsf{Ev}(\underline{\langle M,\varepsilon\rangle})(\alpha) = \mathbb{P}_{\underline{\langle M,\varepsilon\rangle}}(\mathsf{Ev}_{\mathsf{lab}}(\langle M,\varepsilon\rangle) = \mu) \tag{2}$$

Let $l \in L$, we have

$$\begin{aligned}
\mathbb{E}(\mathsf{Ev}_{\mathsf{lab}}(\langle M,\varepsilon\rangle)_l) &= \sum_{\mu \in \mathcal{M}_{\mathrm{fin}}(L)} \mu(l) \mathbb{P}_{\langle M,\varepsilon\rangle}(\mathsf{Ev}_{\mathsf{lab}}(\langle M,\varepsilon\rangle) = \mu) \quad \text{by Equation (1)} \\
&= \sum_{\mu \in \mathcal{M}_{\mathrm{fin}}(L)} \mu(l) [\![\mathsf{sp}_{\vec{x}}(M)]\!]_{\vec{x}})_{(\mu\,[0],0)} \quad \text{by Equation (2)} \\
&= \frac{\partial [\![\mathsf{sp}_{\vec{x}} M]\!]_{\vec{x}}(\vec{r}\,\overline{0})_0}{\partial r_l}(1,\ldots,1)\,.
\end{aligned}$$

Indeed, given $\vec{r} \in [0,1]^L$ one has $[\![\mathsf{sp}_{\vec{x}} M]\!]_{\vec{x}}(\vec{r}\,\overline{0})_0 = \sum_{\mu \in \mathcal{M}_{\mathrm{fin}}(L)} [\![\mathsf{sp}_{\vec{x}}(M)]\!]_{\vec{x}})_{(\mu\,[0],0)} \vec{r}^\mu$ and $\frac{\partial \vec{r}^\mu}{\partial r_l}(1,...,1) = \mu(l)$, whence the last equation.

▶ **Theorem 6.** *Let $M \in \mathsf{pPCF}_{\mathsf{lab}}(L)$ with $\vdash M : \iota$. Then*

$$\mathbb{E}(\mathsf{Ev}_{\mathsf{lab}}(\langle M,\varepsilon\rangle)_l \mid \langle \underline{M},\varepsilon\rangle \downarrow \underline{0}) = \frac{\partial [\![\mathsf{sp}_{\vec{x}} M]\!](\vec{r}\,\overline{0})}{\partial r_l}(1,\ldots,1) / [\![\underline{M}]\!]_0\,.$$

▶ **Example 7.** The point of this formula is that we can apply it to algebraic expressions of the semantics of the program. Consider the following term M_q (for $q \in \mathbb{Q} \cap [0,1]$) such that $\vdash M_q : \iota \Rightarrow \iota$: $M_q = \mathsf{fix}(\lambda f^{\iota \Rightarrow \iota}\, \lambda x^\iota\, \mathsf{if}(\mathsf{coin}(q), \mathsf{if}((f)x, \mathsf{if}((f)x, \underline{0}, \Omega^\iota), \Omega^\iota), \mathsf{if}(x, \mathsf{if}(x, \underline{0}, \Omega^\iota), \Omega^\iota)))$, we study $(M_q)\underline{0}^l$ (for a fixed label $l \in \mathcal{L}$). So in this example, "time" means "number of uses

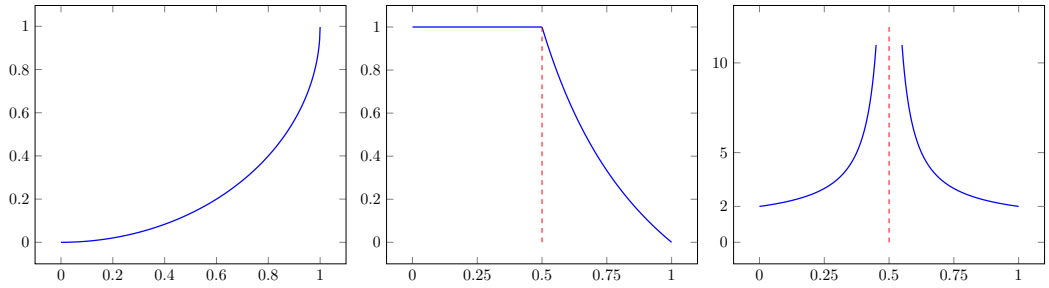

Figure 3 Plot of $\varphi_{0.5}(u)$ with u on the x-axis (vertical slope at $u = 1$). Plots of $\varphi_q(1)$ and $\mathbb{E}(\mathsf{Ev}_{\mathsf{lab}}(\langle (M_q)\underline{0}^l, \varepsilon\rangle)_l \mid \langle (M_q)\underline{0}, \varepsilon\rangle \downarrow \underline{0})$ with q on the x-axis. See Example 7.

of the parameter $\underline{0}$". For all $v \in \mathsf{PN}$, we have $[\![M_q]\!](v) = \varphi_q(v_0)\,\overline{0}$ where $\varphi_q : [0,1] \to [0,1]$ is such that $\varphi_q(u)$ is the least element of $[0,1]$ which satisfies $\varphi_q(u) = (1-q)\,u^2 + q\,\varphi_q(u)^2$. So $\varphi_q(u) = (1 - \sqrt{1 - 4q(1-q)u^2})/2q$ if $q > 0$ and $\varphi_0(u) = u^2$, the choice between the two solutions of the quadratic equation being determined by the fact that the resulting function φ_q must be monotonic in u. So by Theorem 1 (for $q \in (0,1]$)

$$\mathbb{P}((M_q)\underline{0} \downarrow \underline{0}) = \varphi_q(1) = \frac{1 - |2q-1|}{2q} = \begin{cases} 1 & \text{if } q \leq 1/2 \\ \frac{1-q}{q} & \text{if } q > 1/2\,. \end{cases} \tag{3}$$

Observe that we have also $\mathbb{P}(M_0 \downarrow \underline{0}) = \varphi_0(1) = 1$ so that Equation (3) holds for all $q \in [0,1]$ (the corresponding curve is the second one in Fig. 3). Then by Theorem 6 we have $\mathbb{E}(\mathsf{Ev}_{\mathsf{lab}}(\langle (M_q)\underline{0}^l, \varepsilon\rangle)_l \mid \langle (M_q)\underline{0}, \varepsilon\rangle \downarrow \underline{0}) = \varphi_q'(1)/\varphi_q(1)$. Since $\varphi_q(u) = (1-q)\,u^2 + q\,\varphi_q(u)^2$ we have $\varphi_q'(u) = 2(1-q)u + 2q\varphi_q'(u)\varphi_q(u)$ and hence $\varphi_q'(1) = 2(1-q)/(1-2q\varphi_q(1))$, so that $\varphi_q'(1)/\varphi_q(1) = 2(1-q)/(1-2q)$ if $q < 1/2$, $\varphi_{1/2}'(1)/\varphi_q(1) = \infty$ and $\varphi_q'(1)/\varphi_q(1) = 2(1-q)/(2q-1)$ if $q > 1/2$ (using the expression of $\varphi_q(1)$ given by Equation (3)), see the third curve in Fig. 3. For $q > 1/2$ notice that the conditional time expectation *and* the probability of convergence decrease when q tends to 1. When q is very close to 1, $(M_q)\underline{0}$ has a very low probability to terminate, but when it does, it uses its argument only twice. For $q = 1/2$ we have almost sure termination with an infinite expected computation time.

3 Differentials and distances

3.1 Order theoretic characterization of PCSs

The following simple lemma will prove quite useful in the sequel. It is proven in [12] in a rather sketchy way, we provide here a detailed proof for further references. We say that a partially ordered set S is ω-complete if any increasing sequence of elements of S has a least upper bound.

▶ **Lemma 8.** *Let I be a countable set and let $P \subseteq (\mathbb{R}_{\geq 0})^I$. Then (I,P) is a probabilistic coherence space iff the following properties hold (equipping P with the product order).*

1. *P is downwards closed and closed under barycentric combinations*
2. *P is ω-complete*
3. *and for all $a \in I$ there is $\varepsilon > 0$ such that $\varepsilon e_a \in P$ and $P_a \subseteq [0, 1/\varepsilon]$.*

Proof. The $\Rightarrow$ implication is easy (see [5]), we prove the converse, which uses the Hahn-Banach theorem in finite dimension. Let $y \in (\mathbb{R}_{\geq 0})^I$ such that $y \notin P$. We must prove that there exists $x' \in P^\perp$ such that $\langle y, x'\rangle > 1$ and $\forall x \in P\, \langle x, x'\rangle \leq 1$. Given $J \subseteq I$ and $z \in (\mathbb{R}_{\geq 0})^I$, let $z|_J$ be the element of $(\mathbb{R}_{\geq 0})^I$ which takes value z_j for $j \in J$ and 0 for $j \notin J$.

Then y is the lub of the increasing sequence $\{y|_{\{i_1,\dots,i_n\}} \mid n \in \mathbb{N}\}$ (where $i_1, i_2, \dots$ is any enumeration of I) and hence there must be some $n \in \mathbb{N}$ such that $y|_{\{i_1,\dots,i_n\}} \notin P$. Therefore it suffices to prove the result for I finite, what we assume now. Let $Q = \{x \in \mathbb{R}^I \mid (|x_i|)_{i\in I} \in P\}$ which is a convex subset of $\mathbb{R}^I$. Let $t_0 = \sup\{t \in \mathbb{R}_{\geq 0} \mid ty \in P\}$. By our closeness assumption on P, we have $t_0 y \in P$ and therefore $t_0 < 1$. Let $h : \mathbb{R}y \to \mathbb{R}$ be defined by $h(ty) = t/t_0$ ($t_0 \neq 0$ by our assumption (3) about P and because I is finite). Let $q : \mathbb{R}^I \to \mathbb{R}_{\geq 0}$ be the gauge of Q, which is the semi-norm given by $q(z) = \inf\{\varepsilon > 0 \mid z \in \varepsilon Q\}$. It is actually a norm by our assumptions on P. Observe that $h(z) \leq q(z)$ for all $z \in \mathbb{R}y$: this boils down to showing that $t \leq t_0 q(ty) = |t|\, t_0 q(y)$ for all $t \in \mathbb{R}$ which is clear since $t_0 q(y) = 1$ by definition of these numbers. Hence, by the Hahn-Banach Theorem, there exists a linear $l : \mathbb{R}^I \to \mathbb{R}$ which is $\leq q$ and coincides with h on $\mathbb{R}y$. Let $y' \in \mathbb{R}^I$ be such that $\langle z, y'\rangle = l(z)$ for all $z \in \mathbb{R}^I$ (using again the finiteness of I). Let $x' \in (\mathbb{R}_{\geq 0})^I$ be defined by $x'_i = |y'_i|$. It is clear that $\langle y, x'\rangle > 1$: since $y \in (\mathbb{R}_{\geq 0})^I$ we have $\langle y, x'\rangle \geq \langle y, y'\rangle = l(y) = h(y) = 1/t_0 > 1$. Let $N = \{i \in I \mid y'_i < 0\}$. Given $z \in P$, let $\bar{z} \in \mathbb{R}^I$ be given by $\bar{z}_i = -z_i$ if $i \in N$ and $\bar{z}_i = z_i$ otherwise. Then $\langle z, x'\rangle = \langle \bar{z}, y'\rangle = l(\bar{z}) \leq 1$ since $\bar{z} \in Q$ (by definition of Q and because $z \in P$). It follows that $x' \in P^\perp$. ◀

3.2 Local PCS and derivatives

Let X be a PCS and let $x \in \mathsf{P}X$. We define a new PCS X_x as follows. First we set $|X_x| = \{a \in |X| \mid \exists \varepsilon > 0\ x + \varepsilon e_a \in \mathsf{P}X\}$ and then $\mathsf{P}(X_x) = \{u \in (\mathbb{R}_{\geq 0})^{|X_x|} \mid x + u \in \mathsf{P}X\}$. There is a slight abuse of notation here: u is not an element of $(\mathbb{R}_{\geq 0})^{|X|}$, but we consider it as such by simply extending it with 0 values to the elements of $|X| \setminus |X_x|$. Observe also that, given $u \in \mathsf{P}X$, if $x + u \in \mathsf{P}X$, then we *must have* $u \in \mathsf{P}(X_x)$, in the sense that u necessarily vanishes outside $|X_x|$. It is clear that $(|X_x|, \mathsf{P}(X_x))$ satisfies the conditions of Lemma 8 and therefore X_x is actually a PCS, called the *local PCS of X at x*.

Let $t \in \mathbf{Pcoh}_!(X, Y)$ and let $x \in \mathsf{P}X$. Given $u \in \mathsf{P}(X_x)$, we know that $x + u \in \mathsf{P}X$ and hence we can compute $t(x + u) \in \mathsf{P}Y$: $t(x + u)_b = \sum_{\mu\in!X|} t_{\mu,b}(x + u)^\mu = \sum_{\mu\in!X|} t_{\mu,b} \sum_{\nu\leq\mu} \binom{\mu}{\nu} x^{\mu-\nu}u^\nu$. Upon considering only the u-constant and the u-linear parts of this summation (and remembering that actually $u \in \mathsf{P}(X_x)$), we get $t(x) + \sum_{a\in|X|} u_a \sum_{\mu\in!X|}(\mu(a)+1)t_{\mu+[a],b}x^\mu \leq t(x+u) \in \mathsf{P}Y$. Given $a \in |X_x|$ and $b \in |Y_{t(x)}|$, we set $t'(x)_{a,b} = \sum_{\mu\in!X|}(\mu(a) + 1)t_{\mu+[a],b}x^\mu$ and we have proven that actually $t'(x) \in \mathsf{P}(X_x, Y_{t(x)})$. By definition, this linear morphism $t'(x)$ is the *derivative (or differential, or Jacobian) of t at x*[9]. It is uniquely characterized by the fact that, for all $x \in \mathsf{P}X$ and $u \in \mathsf{P}X_x$, we have

$$t(x + u) = t(x) + t'(x)\, u + \widetilde{t}(x, u) \tag{4}$$

where $\widetilde{t}$ is a power series in x and u whose all terms have global degree ≥ 2 in u.

As a typical example, consider the case where $Y = !X$ and $t = \delta = \mathsf{Id}_{!X} \in \mathbf{Pcoh}_!(X, !X)$, so that $\delta(x) = x^!$. Given $a \in |X_x|$ and $\nu \in [!X_{x^!}]$, we have

$$\delta'(x)_{a,\nu} = \sum_{\mu\in!X|} (\mu(a) + 1)\delta_{\mu+[a],\nu}x^\mu = \begin{cases} 0 & \text{if } \nu(a) = 0 \\ \nu(a)x^{\nu-[a]} & \text{if } \nu(a) > 0\,. \end{cases}$$

We know that $\delta'(x) \in \mathsf{P}(X_x \multimap !X_{x^!})$ so that $\delta'(x)$ is a "local version" of DiLL's coderelliction [9]. Observe for instance that $\delta'(0)$ satisfies $\delta'(0)_{a,\nu} = \delta_{\nu,[a]}$ and therefore coincides with the ordinary definition of coderelliction.

[9] But unlike our models of Differential LL, this derivative is only defined locally; this is slightly reminiscent of what happens in differential geometry.

▶ **Proposition 9** (Chain Rule). *Let $s \in \mathbf{Pcoh}_!(X,Y)$ and $t \in \mathbf{Pcoh}_!(Y,Z)$. Let $x \in \mathsf{P}X$ and $u \in \mathsf{P}X_x$. Then we have $(t \circ s)'(x)\, u = t'(s(x))\, s'(x)\, u$.*

Proof. It suffices to write

$$\begin{aligned}(t \circ s)(x+u) &= t(s(x+u)) = t(s(x) + s'(x)\, u + \widetilde{s}(x,u)) \\ &= t(s(x)) + t'(s(x))\,(s'(x)\, u + \widetilde{s}(x,u))) + \widetilde{t}(s(x), s'(x)\, u + \widetilde{s}(x,u)) \\ &= t(s(x)) + t'(s(x))\,(s'(x)\, u) + t'(s(x))\,(\widetilde{s}(x,u)) + \widetilde{t}(s(x), s'(x)\, u + \widetilde{s}(x,u))\end{aligned}$$

by linearity of $t'(s(x))$ which proves our contention by the observation that, in the power series $t'(s(x))\,(\widetilde{s}(x,u)) + \widetilde{t}(s(x), s'(x)\, u + \widetilde{s}(x,u))$, u appears with global degree ≥ 2 by what we know on $\widetilde{s}$ and $\widetilde{t}$. ◀

3.3 Glb's, lub's and distance

Since we are working with probabilistic coherence spaces, we could deal directly with families of real numbers and define these operations more concretely. We prefer not to do so to have a more canonical presentation which can be generalized to cones such as those considered in [10].

Given $x, y \in \mathsf{P}X$, observe that $x \wedge y \in \mathsf{P}X$, where $(x \wedge y)_a = \min(x_a, y_a)$, and that $x \wedge y$ is the glb of x and y in $\mathsf{P}X$ (with its standard ordering). It follows that x and y have also a lub $x \vee y \in \overline{\mathsf{P}}X$ which is given by $x \vee y = x + y - (x \wedge y)$ (and of course $(x \vee y)_a = \max(x_a, y_a)$).

Let us prove that $x + y - (x \wedge y)$ is actually the lub of x and y. First, $x \leq x + y - (x \wedge y)$ simply because $x \wedge y \leq y$. Next, let $z \in \overline{\mathsf{P}}X$ be such that $x \leq z$ and $y \leq z$. We must prove that $x + y - (x \wedge y) \leq z$, that is $x + y \leq z + (x \wedge y) = (z + x) \wedge (z + y)$, which is clear since $x + y \leq z + x, z + y$. We have used the fact that $+$ distributes over $\wedge$ so let us prove this last fairly standard property: $z + (x \wedge y) = (z + x) \wedge (z + y)$. The "$\leq$" inequation is obvious (monotonicity of $+$) so let us prove the converse, which amounts to $x \wedge y \geq (z+x) \wedge (z+y) - z$ (observe that indeed that $z \leq (z + x) \wedge (z + y)$). This in turn boils down to proving that $x \geq (z + x) \wedge (z + y) - z$ (and similarly for y) which results from $x + z \geq (z + x) \wedge (z + y)$ and we are done.

We define the distance between x and y by $\mathsf{d}_X(x,y) = \|x - (x \wedge y)\|_X + \|y - (x \wedge y)\|_X$. The only non obvious fact to check for proving that this is actually a distance is the triangular inequality, so let $x, y, z \in \mathsf{P}X$. We have $x - (x \wedge z) \leq x - (x \wedge y \wedge z) = x - (x \wedge y) + (x \wedge y) - (x \wedge y \wedge z)$ and hence $\|x - (x \wedge z)\|_X \leq \|x - (x \wedge y)\|_X + \|(x \wedge y) - (x \wedge y \wedge z)\|_X$. Now we have $(x \wedge y) \vee (y \wedge z) \leq y$, that is $(x \wedge y) + (y \wedge z) - (x \wedge y \wedge z) \leq y$, that is $(x \wedge y) - (x \wedge y \wedge z) \leq y - (y \wedge z)$. It follows that $\|x - (x \wedge z)\|_X \leq \|x - (x \wedge y)\|_X + \|y - (y \wedge z)\|_X$ and symmetrically $\|z - (x \wedge z)\|_X \leq \|z - (z \wedge y)\|_X + \|y - (y \wedge x)\|_X$ and summing up we get, as expected $\mathsf{d}_X(x,z) \leq \mathsf{d}_X(x,y) + \mathsf{d}_X(y,z)$.

3.4 A Lipschitz property

First of all, observe that, if $w \in \overline{\mathsf{P}}(X \multimap Y)$ and $x \in \overline{\mathsf{P}}X$, we have $\|w\, x\|_Y \leq \|w\|_{X \multimap Y}\, \|x\|_X$. Indeed $\frac{w}{\|w\|_{X\multimap Y}} \in \mathsf{P}(X \multimap Y)$ and $\frac{x}{\|x\|_X} \in \mathsf{P}X$, therefore $\frac{w}{\|w\|_{X\multimap Y}}\, \frac{x}{\|x\|_X} \in \mathsf{P}Y$ and our contention follows.

Let $p \in [0,1)$. If $x \in \mathsf{P}X$ and $\|x\|_X \leq p$, observe that, for any $u \in \mathsf{P}X$, one has $\|x + (1-p)u\|_X \leq \|x\|_X + (1-p)\|u\|_X \leq 1$ and hence $(1-p)u \in \mathsf{P}(X_x)$. Therefore, given $w \in \mathsf{P}(X_x \multimap Y)$, we have $\|w\,(1-p)u\|_Y \leq 1$ for all $u \in \mathsf{P}X$ and hence $(1-p)w \in \mathsf{P}(X \multimap Y)$.

Let $t \in \mathsf{P}(!X \multimap 1)$. We have seen that, for all $x \in \mathsf{P}X$ we have $t'(x) \in \mathsf{P}(X_x \multimap 1_{t(x)}) \subseteq \mathsf{P}(X_x \multimap 1)$. Therefore, if we assume that $\|x\|_X \leq p$, we have

$$(1-p)t'(x) \in \mathsf{P}(X \multimap 1) = \mathsf{P}X^{\perp}\,. \tag{5}$$

Let $x \leq y \in \mathsf{P}X$ be such that $\|y\|_X \leq p$. Observe that $2-p > 1$ and that $x+(2-p)(y-x) = y+(1-p)(y-x) \in \mathsf{P}X$ (because $\|y\|_X \leq p$ and $y-x \in \mathsf{P}X$). We consider the function $h : [0, 2-p] \to [0,1]$ defined by $h(\theta) = t(x+\theta(y-x))$, which is clearly analytic on $[0, 2-p)$. More precisely, one has $h(\theta) = \sum_{n=0}^{\infty} c_n \theta^n$ for some sequence of non-negative real numbers c_n such that $\sum_{n=0}^{\infty} c_n (2-p)^n \leq 1$.

Therefore the derivative of h is well defined on $[0,1] \subset [0, 2-p)$ and one has $h'(\theta) = t'(x+\theta(y-x))\,(y-x) \leq \frac{\|y-x\|_X}{1-p}$ by (5), using Proposition 9. We have

$$0 \leq t(y) - t(x) = h(1) - h(0) = \int_0^1 h'(\theta)\, d\theta \leq \frac{\|y-x\|_X}{1-p}\,. \tag{6}$$

Let now $x, y \in \mathsf{P}X$ be such that $\|x\|_X, \|y\|_X \leq p$ (we don't assume any more that they are comparable). We have $|t(x) - t(y)| = |t(x) - t(x \wedge y) + t(x \wedge y) - t(y)| \leq |t(x) - t(x \wedge y)| + |t(y) - t(x \wedge y)| \leq \frac{1}{1-p}(\|x - (x \wedge y)\|_X + \|y - (x \wedge y)\|_X) = \frac{\mathsf{d}_X(x,y)}{1-p}$ by (6) since $x \wedge y \leq x, y$.

▶ **Theorem 10.** *Let $t \in \mathsf{P}(!X \multimap 1)$. Given $p \in [0,1)$, the function t is Lipschitz with Lipschitz constant $\frac{1}{1-p}$ on $\{x \in \mathsf{P}X \mid \|x\|_X \leq p\}$ when $\mathsf{P}X$ is equipped with the distance d_X, that is*

$$\forall x, y \in \mathsf{P}X \quad \|x\|_X, \|y\|_X \leq p \Rightarrow |t(x) - t(y)| \leq \frac{\mathsf{d}_X(x,y)}{1-p}\,.$$

4 Application to the observational distance in pPCF

Given a term M such that $\vdash M : \iota$, remember that we use $\mathbb{P}(M \downarrow \underline{0})$ for the probability of M to reduce to $\underline{0}$ in the probabilistic reduction system of [10], so that $\mathbb{P}(M \downarrow \underline{0}) = \mathbb{P}_{\langle M, \varepsilon\rangle}(\langle M, \varepsilon\rangle \downarrow \underline{0})$ with the (admittedly heavy) notations of Section 2. Remember that $\mathbb{P}(M \downarrow \underline{0}) = [\![M]\!]_0$ by the Adequacy Theorem of [10].

Given a type σ and two pPCF terms M, M' such that $\vdash M : \sigma$ and $\vdash M' : \sigma$, we define the *observational distance* $\mathsf{d}_{\mathsf{obs}}(M, M')$ between M and M' as the sup of all the $|\mathbb{P}((C)M \downarrow \underline{0}) - \mathbb{P}((C)M' \downarrow \underline{0})|$ taken over terms C such that $\vdash C : \iota$ (testing contexts).

If $\varepsilon \in [0,1] \cap \mathbb{Q}$ we have $\mathsf{d}_{\mathsf{obs}}(\mathsf{coin}(0), \mathsf{coin}(\varepsilon)) = 1$ as soon as $\varepsilon > 0$. It suffices indeed to consider the context $C = \mathsf{fix}\, f^{\iota \Rightarrow \iota}\, \lambda x^{\iota}\, \mathsf{if}(x, (f)x, z \cdot \underline{0})$. The semantics $[\![C]\!] \in \mathsf{P}(!\mathsf{N} \multimap \mathsf{N})$ is a function $c : \mathsf{PN} \to \mathsf{PN}$ such that $\forall u \in \mathsf{PN}\ c(u) = u_0 c(u) + (\sum_{i=1}^{\infty} u_i)\overline{0}$ and which is minimal (for the order relation of $\mathsf{P}(!\mathsf{N} \multimap \mathsf{N})$). If follows that

$$c(u) = \begin{cases} 0 & \text{if } u_0 = 1 \\ \frac{1}{1-u_0} \sum_{i=1}^{\infty} u_i & \text{otherwise}\,. \end{cases}$$

Then $c((1-\varepsilon)\overline{0} + \varepsilon\overline{1}) = 0$ if $\varepsilon = 0$ and $c((1-\varepsilon)\overline{0} + \varepsilon\overline{1}) = 1$ is $\varepsilon > 0$. This is a well known phenomenon called "probability amplification" in stochastic programming.

Nevertheless, we can control a tamed version of the observational distance. Given a closed pPCF term C such that $\vdash C : \sigma \Rightarrow \iota$ we define $C^{\langle p\rangle} = \lambda z^{\sigma}\, (C)\mathsf{if}(\mathsf{coin}(p), z, \Omega^{\sigma})$ and a tamed version of the observational distance is defined by

$$\mathsf{d}^{\langle p\rangle}_{\mathsf{obs}}(M, M') = \sup \left\{ \left| \mathbb{P}((C^{\langle p\rangle})M \downarrow \underline{0}) - \mathbb{P}((C^{\langle p\rangle})M' \downarrow \underline{0}) \right| \mid \vdash C : \sigma \Rightarrow \iota \right\}.$$

▶ **Theorem 11.** *Let $p \in [0,1) \cap \mathbb{Q}$. Let M and M' be terms such that $\vdash M : \sigma$ and $\vdash M' : \sigma$. Then we have*

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

Proof.

$$\begin{aligned}\mathsf{d}_{\mathsf{obs}}^{\langle p\rangle}(M,M') &= \sup\{|[\![C]\!](p[\![M]\!])_0 - [\![C]\!](p[\![M']\!])_0| \mid \vdash C:\sigma\Rightarrow\iota\}\\ &\leq \sup\{|t(p[\![M]\!]) - t(p[\![M']\!]|) \mid t\in \mathsf{P}(!\![\![\sigma]\!]\multimap 1)\}\\ &\leq \frac{\mathsf{d}_{[\![\sigma]\!]}(p[\![M]\!],p[\![M']\!])}{1-p} = \frac{p}{1-p}\,\mathsf{d}_{[\![\sigma]\!]}([\![M]\!],[\![M']\!])\,.\end{aligned}$$

by the Adequacy Theorem and by Theorem 10. ◀

Since $p/(1-p) = p+p^2+\cdots$ and $\mathsf{d}_{[\![\sigma]\!]}(\_,\_)$ is an over-approximation of the observational distance restricted to linear contexts, this inequation carries a rather clear operational intuition in terms of execution in a Krivine machine as in Section 2.1.2 (thanks to Paul-André Melliès for this observation). Indeed, using the stacks of Section 2.1.2, a *linear* observational distance on pPCF terms can easily be defined as follows, given terms M and M' such that $\vdash M:\sigma$ and $\vdash M':\sigma$:

$$\mathsf{d}_{\mathsf{lin}}(M,M') = \sup_{\sigma\vdash\pi}\left|\mathbb{P}_{\langle M,\pi\rangle}(\langle M,\pi\rangle\downarrow\underline{0}) - \mathbb{P}_{\langle M',\pi\rangle}(\langle M',\pi\rangle\downarrow\underline{0})\right|\,.$$

In view of Theorem 11 and of the fact that $\mathsf{d}_{\mathsf{lin}}(M,M')\leq \mathsf{d}_{[\![\sigma]\!]}([\![M]\!],[\![M']\!])$ (easy to prove, since each stack can be interpreted as a linear morphism in **Pcoh**), a natural and purely syntactic conjecture seems to be

$$\mathsf{d}_{\mathsf{obs}}^{\langle p\rangle}(M,M') \leq \frac{p}{1-p}\,\mathsf{d}_{\mathsf{lin}}(M,M')\,. \tag{7}$$

This seems easy to prove in the case $\mathbb{P}_{\langle M',\pi\rangle}(\langle M',\pi\rangle\downarrow\underline{0}) = 0$: it suffices to observe that a path which is a successful reduction of $\langle(C^{\langle p\rangle})M,\varepsilon\rangle$ in the "Krivine Machine" of Section 2.1.2 (considered here as a Markov chain) can be decomposed as

$$\begin{aligned}\langle(C^{\langle p\rangle})M,\varepsilon\rangle &\to^* \langle \mathsf{if}(\mathsf{coin}(p),M,\Omega^\sigma),\pi_1(C,M)\rangle \to^* \langle \mathsf{if}(\mathsf{coin}(p),M,\Omega^\sigma),\pi_2(C,M)\rangle\\ &\to^*\cdots\to^* \langle \mathsf{if}(\mathsf{coin}(p),M,\Omega^\sigma),\pi_k(C,M)\rangle \to^* \langle\underline{0},\varepsilon\rangle\end{aligned}$$

where $(\pi_i(C,M))_{i=1}^k$ is a finite sequence of stacks such that $\sigma\vdash\pi_i(M)$ for each i. Notice that this sequence of stacks depends not only on C and M but also on the considered path of the Markov chain.

In the general case, Inequation (7) seems less easy to prove because, for a given common initial context C, the sequences of reductions (and of associated stacks) starting with $\langle(C^{\langle p\rangle})M,\varepsilon\rangle$ and $\langle(C^{\langle p\rangle})M',\varepsilon\rangle$ differ. This divergence has low probability when $\mathsf{d}_{\mathsf{lin}}(M,M')$ is small, but it is not completely clear how to evaluate it. Coinductive methods like probabilistic bisimulation as in the work of Crubillé and Dal Lago are certainly relevant here.

Our Theorem 10 shows that another and more geometric approach, based on a simple denotational model, is also possible to get Theorem 11 which, though weaker than Inequation (7), allows nevertheless to control the p-tamed distance.

We finish the paper by observing that the equivalence relations induced on terms by these observational distances coincide with the ordinary observational distance if $p\neq 0$.

▶ **Theorem 12.** *Assume that $0<p\leq 1$. If $\mathsf{d}_{\mathsf{obs}}^{\langle p\rangle}(M,M')=0$ then $M\sim M'$ (that is, M and M' are observationally equivalent).*

Proof. If $\vdash M:\sigma$ we set $M_p = \mathsf{if}(\mathsf{coin}(p),M,\Omega^\sigma)$. If $\mathsf{d}_{\mathsf{obs}}^{\langle p\rangle}(M,M')=0$ then $M_p\sim M'_p$ by definition of observational equivalence, hence $[\![M_p]\!]=[\![M'_p]\!]$ by our Full Abstraction Theorem [10], but $[\![M_p]\!]=p[\![M]\!]$ and similarly for M'. Since $p\neq 0$ we get $[\![M]\!]=[\![M']\!]$ and hence $M\sim M'$ by adequacy [10]. ◀

So for each $p \in (0,1)$ and for each type σ we can consider $\mathsf{d}^{\langle p \rangle}$ as a distance on the observational classes of closed terms of type σ. We call it the *p-tamed observational distance*. Our Theorem 11 shows that we can control this distance using the denotational distance. For instance we have $\mathsf{d}^{\langle p \rangle}_{\mathsf{obs}}(\mathsf{coin}(0), \mathsf{coin}(\varepsilon)) \leq \frac{2p\varepsilon}{1-p}$ so that $\mathsf{d}^{\langle p \rangle}_{\mathsf{obs}}(\mathsf{coin}(0), \mathsf{coin}(\varepsilon))$ tends to 0 when ε tends to 0.

Conclusion

The two results of this paper are related: both use derivatives wrt. probabilities to evaluate the number of times arguments are used. We think that they provide motivations for investigating further differential extensions of pPCF and related languages in the spirit of [11].

References

1 Johannes Borgström, Ugo Dal Lago, Andrew D. Gordon, and Marcin Szymczak. A lambda-calculus foundation for universal probabilistic programming. In Jacques Garrigue, Gabriele Keller, and Eijiro Sumii, editors, *Proceedings of the 21st ACM SIGPLAN International Conference on Functional Programming, ICFP 2016, Nara, Japan, September 18-22, 2016*, pages 33–46. ACM, 2016.

2 Simon Castellan, Pierre Clairambault, Hugo Paquet, and Glynn Winskel. The concurrent game semantics of Probabilistic PCF. In Anuj Dawar and Erich Grädel, editors, *Proceedings of the 33rd Annual ACM/IEEE Symposium on Logic in Computer Science, LICS 2018, Oxford, UK, July 09-12, 2018*, pages 215–224. ACM, 2018. doi:10.1145/3209108.

3 Raphaëlle Crubillé. Probabilistic Stable Functions on Discrete Cones are Power Series. In Anuj Dawar and Erich Grädel, editors, *Proceedings of the 33rd Annual ACM/IEEE Symposium on Logic in Computer Science, LICS 2018, Oxford, UK, July 09-12, 2018*, pages 275–284. ACM, 2018. doi:10.1145/3209108.3209198.

4 Raphaëlle Crubillé and Ugo Dal Lago. Metric Reasoning About Lambda-Terms: The General Case. In Hongseok Yang, editor, *Programming Languages and Systems - 26th European Symposium on Programming, ESOP 2017, Held as Part of the European Joint Conferences on Theory and Practice of Software, ETAPS 2017, Uppsala, Sweden, April 22-29, 2017, Proceedings*, volume 10201 of *Lecture Notes in Computer Science*, pages 341–367. Springer, 2017. doi:10.1007/978-3-662-54434-1_13.

5 Vincent Danos and Thomas Ehrhard. Probabilistic coherence spaces as a model of higher-order probabilistic computation. *Information and Computation*, 152(1):111–137, 2011.

6 Vincent Danos and Russell Harmer. Probabilistic game semantics. In *Proceedings of the 15th Annual IEEE Symposium on Logic in Computer Science*. IEEE Computer Society, 2000.

7 Daniel de Carvalho. Execution Time of lambda-Terms via Denotational Semantics and Intersection Types. *CoRR*, abs/0905.4251, 2009. arXiv:0905.4251.

8 Daniel de Carvalho. Execution time of λ-terms via denotational semantics and intersection types. *MSCS*, 28(7):1169–1203, 2018.

9 Thomas Ehrhard. An introduction to differential linear logic: proof-nets, models and antiderivatives. *Mathematical Structures in Computer Science*, 28(7):995–1060, 2018. doi:10.1017/S0960129516000372.

10 Thomas Ehrhard, Michele Pagani, and Christine Tasson. Full Abstraction for Probabilistic PCF. *Journal of the ACM*, 65(4):23:1–23:44, 2018. doi:10.1145/3164540.

11 Thomas Ehrhard and Laurent Regnier. The differential lambda-calculus. *Theoretical Computer Science*, 309(1-3):1–41, 2003.

12 Jean-Yves Girard. Between logic and quantic: a tract. In Thomas Ehrhard, Jean-Yves Girard, Paul Ruet, and Philip Scott, editors, *Linear Logic in Computer Science*, volume 316 of *London Mathematical Society Lecture Notes Series*, pages 346–381. Cambridge University Press, 2004.

13 Klaus Keimel and Gordon D. Plotkin. Mixed powerdomains for probability and nondeterminism. *Logical Methods in Computer Science*, 13(1), 2017. doi:10.23638/LMCS-13(1:2)2017.

14 Peter Selinger. Towards a semantics for higher-order quantum computation. In *Proceedings of the 2nd International Workshop on Quantum Programming Languages, Turku, Finland*, number 33 in TUCS General Publication. Turku Centre for Computer Science, 2004.

15 Matthijs Vákár, Ohad Kammar, and Sam Staton. A domain theory for statistical probabilistic programming. *PACMPL*, 3(POPL):36:1–36:29, 2019.