Differentials and distances in probabilistic coherence spaces

Thomas Ehrhard

DIFFERENTIALS AND DISTANCES IN PROBABILISTIC COHERENCE SPACES

THOMAS EHRHARD

Université de Paris, IRIF, CNRS, F-75013 Paris, France
e-mail address: 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.

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 [DH00] and further developed by various authors (see [CCPW18] 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* [KP17] (domains equipped with an algebraic convex structure) and ω-*quasi Borel spaces* [VKS19] (domains equipped with a generalized notion of measurability).
- 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 [DE11]) for which we could prove adequacy and full abstraction with respect to a probabilistic version of PCF [EPT18, Ehr20]. 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

Key words and phrases: Denotational semantics, probabilistic coherence spaces, differentials of programs, observational equivalence and distances.

This work has been partly funded by the ANR PRC project *Probabilistic Programming Semanics* (PPS) ANR-19-CE48-0014.

Preprint submitted to
Logical Methods in Computer Science

hereditary monotonicity property that we called *stability*[1] and [Cru18] 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 [Ehr18]. 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 [EPT18]. We provide here two more corollaries of this property, involving 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 $\underline{\mathsf{coin}}(r) : \iota$ 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 $\underline{\mathsf{let}}(x, M, N)$ 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 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 of a term M (closed and of ground type) during a reduction of M. 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 [dC09, dC18]. 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 [CL17] for instance. The probabilistic features of the language make this distance too discriminating, putting *e.g.* terms $\underline{\mathsf{coin}}(0)$ and $\underline{\mathsf{coin}}(\varepsilon)$ 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[3]. Using a *locally defined* notion of differential of morphisms in $\mathbf{Pcoh}_!$, we prove that these morphisms enjoy a Lipschitz

[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.

[3] See Remark 3.4 for the definition of this distance for general cones.

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. We use the notation $\vec{s}$ (with an arrow) when we want to stress the fact that the considered object is considered as an indexed 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 (to avoid accumulations of subscripts). Given $i \in I$ we use e_i for the function $I \to \mathbb{R}_{\geq 0}$ such that $\mathsf{e}_i(i) = 1$ and $\mathsf{e}_i(j) = 0$ if $j \neq i$. In other words $\mathsf{e}_i(j) = \delta_{i,j}$, the Kronecker symbol. We use $\mathcal{M}_{\mathrm{fin}}(I)$ for the set of finite multisets of elements of I. 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 $[i_1, \dots, i_k]$ 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. Given $\mu \in \mathcal{M}_{\mathrm{fin}}(I)$ and $i \in I$ we write $i \in \mu$ if $\mu(i) \neq 0$ and we set $\mathsf{supp}(\mu) = \{i \in I \mid i \in \mu\}$.

We use $I^{<\omega}$ for the set of finite sequences $\langle i_1, \dots, i_k \rangle$ 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 [DE11, EPT18] where the reader will find a more detailed presentation, including motivating examples. Here, we recall only the basic definitions and provide a characterization of these objects. So this section should not be considered as an introduction to PCSs: for such an introduction the reader is advised to have a look at the articles mentioned above. PCSs are particular *positive cones*, a notion borrowed from [Sel04]) that we used in [EPT18] to extend the probabilistic semantics of PCS to continuous data-types such as the real line.

1.1. A few words about cones.

A (positive) *pre-cone* is a cancellative[4] 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 cancellativeness 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.

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

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 [EPT18] we show how a notion of *stable* function on cones can be defined, which gives rise to a cartesian closed category and in [Ehr20] we explore the category of cones and linear and Scott-continuous functions.

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[5] 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 $\forall a \in |X| \, \exists m \in \mathbb{R}_{\geq 0} \forall x \in \mathsf{P}X \ x_a \leq m$ or equivalently

$$\forall a \in |X| \quad 0 < \sup_{x \in \mathsf{P}X} x_a < \infty$$

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 $\|\_\|_X$, so that $\|x\|_X = \sup_{x' \in \mathsf{P}X^\perp} \langle x, x' \rangle$.

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 $t \cdot u \in \overline{\mathbb{R}_{\geq 0}}^J$ by $(t \cdot u)_j = \sum_{i \in I} t_{i,j} u_i$ (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 $\forall x \in \mathsf{P}X \ t \cdot x \in \mathsf{P}Y$. 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, PCSs equipped with this notion of morphism form a category **Pcoh**.

The condition $t \in \mathbf{Pcoh}(X, Y)$ is equivalent to $\forall x \in \mathsf{P}X \, \forall y' \in \mathsf{P}Y^\perp \ \langle t \cdot x, y' \rangle \leq 1$ but $\langle t \cdot x, y' \rangle = \langle t, x \otimes y' \rangle$ where $(x \otimes y')_{(a,b)} = x_a y'_b$. We define $X \multimap Y = (|X| \times |Y|, \{t \in (\mathbb{R}_{\geq 0})^{|X \multimap Y|} \mid \forall x \in \mathsf{P}X \ t \cdot x \in \mathsf{P}Y\})$: this is a pre-PCS by this observation, and checking that it is indeed a PCS is easy.

We define then $X \otimes Y = (X \multimap Y^\perp)^\perp$; this is a PCS which satisfies $\mathsf{P}(X \otimes Z) = \{x \otimes z \mid x \in \mathsf{P}X \text{ and } z \in \mathsf{P}Z\}^{\perp\perp}$ where $(x \otimes z)_{(a,c)} = x_a z_c$. Then it is easy to see that we have equipped in that way the category **Pcoh** with a symmetric monoidal structure for which it

[5] 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 [EPT18] for instance.

is $*$-autonomous wrt. the dualizing object $\perp = 1 = (\{*\}, [0,1])$ which is also the unit of $\otimes$. The $*$-autonomy follows easily from the observation that $(X \multimap \perp) \simeq P^\perp$.

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 $\pi_i \cdot x \in \mathsf{P}X_i$ 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 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 $\mathbb{N}$). For each $n \in \mathbb{N}$ we have $\mathsf{e}_n \in \mathsf{PN}$ which is the distribution concentrated on the integer n. There are successor and predecessor morphisms $\mathsf{suc}, \mathsf{pred} \in \mathbf{Pcoh}(\mathsf{N}, \mathsf{N})$ given by $\mathsf{suc}_{n,n'} = \delta_{n+1,n'}$ and $\mathsf{pred}_{n,n'} = 1$ if $n = n' = 0$ or $n = n' + 1$ (and $\mathsf{pred}_{n,n'} = 0$ in all other cases). An element of $\mathbf{Pcoh}(\mathsf{N}, \mathsf{N})$ is a (sub)stochastic matrix and the very idea of this model is to represent programs as transformations of this kind, and their generalizations.

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 $!t \cdot x^! = (t \cdot x)^!$ (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[6] $\mathbf{Pcoh}_!$ can be seen as a category of functions (actually, of stable functions as proved in [Cru18]). 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 $\widehat{t} : \mathsf{P}X \to \mathsf{P}Y$ such that $\widehat{t}(x) = t \cdot x^! = \left(\sum_{\mu\in|!X|} t_{\mu,b} x^\mu\right)_{b\in|Y|}$ 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.

[6] This is the Kleisli category of "!" which has actually a comonad structure that we do not make explicit here, again we refer to [DE11, EPT18].

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

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

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

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

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

Figure 1: Typing rules for pPCF terms and stacks

2.1. **The core language.** The types and terms are given by

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

See Fig. 1 for the typing rules, with typing contexts $\Gamma = (x_1 : \sigma_1, \ldots, x_n : \sigma_n)$; notice that this figures includes the typing rules for the stacks that we introduce below. It is important to keep in mind that it would not make sense to extend the construction $\underline{\mathsf{let}}(z, M, N)$ to terms M which are not of type ι. This construction uses essentially the fact that the type ι is a *positive* formula of linear logic, see [ET16].

2.1.1. *Denotational semantics.* We survey briefly the interpretation of pPCF in PCSs thoroughly described in [EPT18]. 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 "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. These functions are given by

$$\begin{aligned} [\![\underline{n}]\!]_\Gamma(\vec{u}) &= \mathsf{e}_n \\ [\![x_i]\!]_\Gamma(\vec{u}) &= u_i \\ [\![\underline{\mathsf{coin}}(r)]\!]_\Gamma(\vec{u}) &= r\, \mathsf{e}_0 + (1 - r)\, \mathsf{e}_1 \\ [\![\underline{\mathsf{succ}}(M)]\!]_\Gamma(\vec{u}) &= \mathsf{suc} \cdot [\![M]\!]_\Gamma(\vec{u}) = \sum_{n \in \mathbb{N}} [\![M]\!]_\Gamma(\vec{u})_n \mathsf{e}_{n+1} \\ [\![\underline{\mathsf{pred}}(M)]\!]_\Gamma(\vec{u}) &= \mathsf{pred} \cdot [\![M]\!]_\Gamma(\vec{u}) = [\![M]\!]_\Gamma(\vec{u})_0 \mathsf{e}_0 + \sum_{n \in \mathbb{N}} [\![M]\!]_\Gamma(\vec{u})_{n+1} \mathsf{e}_n \\ [\![\underline{\mathsf{let}}(x, M, N)]\!]_\Gamma(\vec{u}) &= \sum_{n \in \mathbb{N}} [\![M]\!]_\Gamma(\vec{u})_n\, [\![N\, [\underline{n}/x]]\!]_\Gamma(\vec{u}) \\ [\![\underline{\mathsf{if}}(M, N, P)]\!]_\Gamma(\vec{u}) &= [\![M]\!]_\Gamma(\vec{u})_0\, [\![N]\!]_\Gamma(\vec{u}) + \left(\sum_{n \in \mathbb{N}} [\![M]\!]_\Gamma(\vec{u})_{n+1} \right) [\![P]\!]_\Gamma(\vec{u}) \\ [\![(M)N]\!]_\Gamma(\vec{u}) &= ([\![M]\!]_\Gamma(\vec{u}))([\![N]\!]_\Gamma(\vec{u})) \\ [\![\underline{\mathsf{fix}}(M)]\!]_\Gamma(\vec{u}) &= \sup_{n \in \mathbb{N}} ([\![M]\!]_\Gamma(\vec{u}))^n(0) \end{aligned}$$

and, assuming that $\Gamma, x : \sigma \vdash M : \tau$ and $\vec{u} \in \prod_{i=1}^{k} \mathsf{P}[\![\sigma_i]\!]$, $[\![\lambda x^\sigma\, M]\!]_\Gamma(\vec{u})$ is the element t of $\mathsf{P}(![\![\sigma]\!] \multimap [\![\tau]\!])$ characterized by $\forall u \in \mathsf{P}[\![\sigma]\!]\ \widehat{t}(u) = [\![M]\!]_{\Gamma,x:\sigma}(\vec{u}, u)$.

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 branching in pPCF because of the $\underline{\mathsf{coin}}(r)$ construct (see [EPT18]). 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 $\underline{\mathsf{coin}}(r)$ 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 the stack π takes a term of type σ and returns an integer) and the typing rules are given in Fig. 1.

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 $\mathsf{Ev} : \mathsf{S} \times \mathcal{C}_0 \to [0,1]$ in Fig. 2[7] where we use the functions

$$\nu_0(r) = r \tag{2.1}$$

$$\nu_1(r) = 1 - r\,. \tag{2.2}$$

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

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

so that $\langle 1 \rangle$ carries the weight of divergence. We prefer this option rather than adding an error element to $\mathcal{C}_0$ which would be more natural from a programming language point of view, but less standard from the viewpoint of probability theory. This choice is arbitrary

[7] Notice that all the equations defining Ev in Fig. 2 are well-typed in the sense that if the state of the LHS of an equation is well-typed, so is its RHS.

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

Figure 2: The pPCF Krivine Machine

and has no impact on the result we prove because all the events of interest for us will be subsets of $\langle 0\rangle\mathcal{C}_0 \subset \mathcal{C}_0$.

Let $\mathbb{P}_s$ be the associated probability measure. 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 [EPT18] (see *e.g.* [BLGS16] for more details on the connection between these two kinds of operational semantics). So the Adequacy Theorem of [EPT18] can be expressed as follows.

Theorem 2.1. *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 $\mathbb{P}_{\langle M,\varepsilon\rangle}(\langle M,\varepsilon\rangle \downarrow \underline{0})$.

We shall introduce several versions of PCF in the sequel, with associated machines.

- In Section 2.2 we introduce $\mathsf{pPCF}_{\mathsf{lab}}$ where terms can be labeled by elements of $\mathcal{L}$ and the machine $\mathsf{Ev}_{\mathsf{lab}}$ which returns a multiset of elements of $\mathcal{L}$ counting how many times labeled subterms arrive in head position during the evaluation.
- In Section 2.4 we introduce $\mathsf{pPCF}_{\mathsf{lc}}$ which includes a labeled version of the $\underline{\mathsf{coin}}(\_)$ construct, and the associated machine $\mathsf{Ev}_{\mathsf{lc}}$ which returns an element of $\mathbb{R}_{\geq 0}$ (a probability actually).
- In the same section we introduce as an auxiliary tool the machine $\mathsf{Ev}_{\mathsf{lc}}^{\eta}$ which differs from the previous one by the fact that it returns an element of $\mathcal{C}_0$.
- We also introduce a machine $\mathsf{Ev}_{\mathsf{lc}}^{-\eta}$ which is a kind of inverse of the previous one.

In the proofs these machines will often be used with an additional integer parameter for indexing the execution steps.

It is important to notice that the $\mathsf{pPCF}_{\mathsf{lc}}$ language and the associated machines are only an *intermediate step* in the proof of Theorem 2.11, the main result of this section, whose statement does not mention them at all.

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 for l belonging to a fixed set $\mathcal{L}$ of labels. The typing rule for this new construct is simply

$$\frac{\Gamma \vdash N : \sigma}{\Gamma \vdash N^l : \sigma}$$

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,

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

When applied to $\langle M, \varepsilon\rangle$, this function counts how often labeled subterms of M arrive in head position during the reduction; these numbers, represented altogether as a multiset of elements of $\mathcal{L}$, depend of course on the random tape provided as argument together with the state.

Let $\mathcal{D}_{\mathsf{lab}}(s)$ be the set of α's such that $\mathsf{Ev}_{\mathsf{lab}}(s, \alpha)$ is defined. Define $\mathsf{strip}(s) \in \mathsf{S}$ as s stripped from its labels.

Lemma 2.2. *For any $s \in \mathsf{S}_{\mathsf{lab}}$ we have $\mathcal{D}_{\mathsf{lab}}(s) = \mathcal{D}(\mathsf{strip}(s))$.*

Proof. Simple inspection of the definition of the two involved functions. More precisely, an easy induction on n shows that

$$\forall n \in \mathbb{N} \quad \mathsf{Ev}(\mathsf{strip}(s), \alpha, n) \neq \uparrow \Leftrightarrow \exists k \geq n \; \mathsf{Ev}_{\mathsf{lab}}(s, \alpha, k) \neq \uparrow .$$

□

We define a r.v.[8] $\mathsf{Ev}_{\mathsf{lab}}(s) : \mathcal{C}_0 \to \mathcal{M}_{\mathrm{fin}}(\mathcal{L})$ by

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

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

$$\begin{aligned}
\mathbb{E}(\mathsf{Ev}_{\mathsf{lab}}(s)_l) &= \sum_{n \in \mathbb{N}} n \, \mathbb{P}_s(\mathsf{Ev}_{\mathsf{lab}}(s)_l = n) \\
&= \sum_{n \in \mathbb{N}} n \sum_{\substack{\mu \in \mathcal{M}_{\mathrm{fin}}(\mathcal{L}) \\ \mu(l) = n}} \mathbb{P}_s(\mathsf{Ev}_{\mathsf{lab}}(s) = \mu) \\
&= \sum_{\mu \in \mathcal{M}_{\mathrm{fin}}(L)} \mu(l) \mathbb{P}_s(\mathsf{Ev}_{\mathsf{lab}}(s) = \mu) \, .
\end{aligned} \tag{2.3}$$

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

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 $\mathsf{strip}(s) \downarrow \underline{0}$). We have

$$\mathbb{E}(\mathsf{Ev}_{\mathsf{lab}}(s)_l \mid \mathsf{strip}(s) \downarrow \underline{0}) = \frac{\mathbb{E}(\mathsf{Ev}_{\mathsf{lab}}(s)_l)}{\mathbb{P}_{\mathsf{strip}(s)}(\mathsf{strip}(s) \downarrow \underline{0})}$$

as the r.v. $\mathsf{Ev}_{\mathsf{lab}}(s)_l$ vanishes outside the event $s \downarrow \underline{0}$ since $\mathcal{D}_{\mathsf{lab}}(s) = \mathcal{D}(\mathsf{strip}(s))$.

2.3. **A bird's eye view of the proof.** 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.

For this purpose we will replace in M each N^l (where N has type σ) with $\underline{\mathsf{if}}(x_l, N, \Omega^\sigma)$ 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 $\Omega^\sigma = \underline{\mathsf{fix}}(\lambda x^\sigma\, x)$ (an ever-looping term). We will obtain in that way in Section 2.5 a term $\mathsf{sp}_{\vec{x}} M$ such that

$$[\![\mathsf{sp}_{\vec{x}} M]\!]_{\vec{x}} \in \mathbf{Pcoh}_!(\mathsf{N}^L, \mathsf{N})\,.$$

We will consider this function as an analytic function $(\mathsf{PN})^L \to \mathsf{PN}$ which therefore induces an analytic function

$$\begin{aligned} f : [0,1]^L &\to [0,1] \\ \vec{r} &\mapsto [\![\mathsf{sp}_{\vec{x}} M]\!]((r_l \mathsf{e}_0)_{l \in L})_0 \end{aligned}$$

(where $\vec{r}\mathsf{e}_0 = (r_l\, \mathsf{e}_0)_{l \in L} \in \mathsf{PN}^L$ for $\vec{r} \in [0,1]^L$). We will prove that the expectation of the number of uses of subterms of M labeled by l is

$$\frac{\partial f(\vec{r})}{\partial r_l}(1, \dots, 1)\,.$$

Notice that in the partial derivative above, the r_l's are bound by the partial derivative itself and by the fact that it is evaluated at $(1, \dots, 1)$.

In order to reduce this problem to Theorem 2.1, we will introduce the intermediate language $\mathsf{pPCF}_{\mathsf{lc}}$ which will allow to see each of these parameters r_l as the probability of yielding $\underline{0}$ for a biased coin construct $\mathsf{lcoin}(l, r_l)$. This calculus will be executed in a further "Krivine machine" $\mathsf{Ev}_{\mathsf{lc}}$ which has as many random tapes as there are elements in L (plus one for the plain $\underline{\mathsf{coin}}(\_)$ constructs already occurring in M).

This intermediate language will be used as follows: given a closed labeled term M whose all labels belong to $L \subseteq \mathcal{L}$ and a family of probabilities $\vec{r} \in [0,1]^L$ we will define in Section 2.5 a term $\mathsf{lc}_{\vec{r}}(M)$ of $\mathsf{pPCF}_{\mathsf{lc}}$ which is M where each labeled subterm N^l is replaced with $\underline{\mathsf{if}}(\mathsf{lcoin}(l, r_l), \mathsf{lc}_{\vec{r}}(N), \Omega^\sigma)$ where Ω^σ. The term $\mathsf{lc}_{\vec{r}}(M)$ defined in that way is closed, the r_l's being parameters and not variables in the sense of the λ-calculus. The probability $p(\vec{r})$ of convergence of $\mathsf{lc}_{\vec{r}}(M)$ depends on $(\mu(l))_{l \in I}$ where $\mu(l) \in \mathbb{N}$ is the number of times an l-labeled subterm of M arrives in head-position during the evaluation of M: this is the meaning of Lemma 2.8. More precisely $\mu(l)$ is the exponent of r_l's in this probability. The main feature of $\mathsf{sp}_{\vec{x}}(M)$, exploited in Section 2.6, is that $p(\vec{r})$ is obtained by applying the semantics of $\mathsf{sp}_{\vec{x}}(M)$ to $\vec{r}$ (or more precisely to $(r_l \mathsf{e}_0)_{l \in L} \in \mathsf{PN}^L$) — the proof of this fact uses Theorem 2.1. The last step will consist in observing that, by taking the derivative of this probability wrt. the variables x_l, one obtains the expectation of the number of times an l-labeled term arrives in head-position during the evaluation of M; this is due to the fact that the $\mu(l)$ are exponents in the expression of $p(\vec{r})$ and that these exponents become coefficients by differentiation.

2.4. **Probabilistic PCF with labeled coins.** Let $\mathsf{pPCF}_{\mathsf{lc}}$ be pPCF extended with a construct $\mathsf{lcoin}(l, r)$ typed as

$$\frac{r \in [0,1] \cap \mathbb{Q} \text{ and } l \in \mathcal{L}}{\Gamma \vdash \mathsf{lcoin}(l, r) : \iota}$$

This language features the usual $\underline{\mathsf{coin}}(r)$ 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 use the similar notations $\mathsf{pPCF}_{\mathsf{lab}}(L)$ and $\mathsf{S}_{\mathsf{lab}}(L)$ for the sets of labeled terms and stacks (see Section 2.2) whose all labels belong to L.

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 $\underline{\mathsf{coin}}(r)$, we use only the first parameter in $\mathcal{C}_0$), with the additional parameters $\vec{\beta}$ passed unchanged in the recursive calls, apart for the the following new rules:

$$\mathsf{Ev}_{\mathsf{lc}}(\langle \mathsf{lcoin}(l, r), \pi\rangle, \alpha, \vec{\beta}) = \mathsf{Ev}_{\mathsf{lc}}(\langle \underline{i}, \pi\rangle, \alpha, \vec{\beta}\,[\gamma/l]) \cdot \nu_i(r) \quad \text{if } \beta(l) = \langle i\rangle\gamma$$

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 $\mathsf{Ev}_{\mathsf{lc}}(s, \_, \_)$.

We define a version $\mathsf{Ev}^{\eta}_{\mathsf{lc}}(s, \_, \_)$ of the machine $\mathsf{Ev}_{\mathsf{lc}}(s, \_, \_)$ which returns an element of $\mathcal{C}_0$ instead of an element of $\mathbb{R}_{\geq 0}$. The definition is the same up to the following rules:

$$\begin{aligned}
\mathsf{Ev}^{\eta}_{\mathsf{lc}}(\langle \mathsf{lcoin}(l, r), \pi\rangle, \alpha, \vec{\beta}) &= \langle i\rangle \mathsf{Ev}^{\eta}_{\mathsf{lc}}(\langle \underline{i}, \pi\rangle, \alpha, \vec{\beta}\,[\gamma/l]) \quad \text{if } \beta(l) = \langle i\rangle\gamma \\
\mathsf{Ev}^{\eta}_{\mathsf{lc}}(\langle \underline{\mathsf{coin}}(r), \pi\rangle, \langle i\rangle\alpha, \vec{\beta}) &= \langle i\rangle \mathsf{Ev}^{\eta}_{\mathsf{lc}}(\langle \underline{i}, \pi\rangle, \alpha, \vec{\beta}) \\
\mathsf{Ev}^{\eta}_{\mathsf{lc}}(\langle \underline{0}, \varepsilon\rangle, \langle\rangle, \vec{\langle\rangle}) &= \langle\rangle\,.
\end{aligned}$$

When defined, $\mathsf{Ev}^{\eta}_{\mathsf{lc}}(\langle \mathsf{lcoin}(l, r), \pi\rangle, \alpha, \vec{\beta})$ is a shuffle of the β_l's and of α (in the order in which the corresponding elements of the tape are read during the execution). Being defined by similar recursions, the functions $\mathsf{Ev}_{\mathsf{lc}}(s, \_, \_)$ and $\mathsf{Ev}^{\eta}_{\mathsf{lc}}(s, \_, \_)$ have the same domain. Then we have

$$\mathsf{Ev}_{\mathsf{lc}}(s, \alpha, \vec{\beta}) = \mathsf{Ev}(\mathsf{strip}(s), \mathsf{Ev}^{\eta}_{\mathsf{lc}}(s, \alpha, \vec{\beta})) \tag{2.4}$$

where "=" should be understood as Kleene equality (either both sides are undefined or both sides are defined and equal).

To prove Equation (2.4), one considers step-indexed versions of the involved machines, equipped with a further parameter in $\mathbb{N}$. For instance the modified $\mathsf{Ev}(s, \alpha, n)$ will be a total function $\mathsf{S}_{\mathsf{lc}}(L) \times (\mathcal{C}_0 \cup \{\uparrow\}) \times \mathbb{N} \to \mathbb{R}_{\geq 0} \cup \{\uparrow\}$ where $\uparrow$ stands for non-terminated computations. Here are a few cases of this modified definition, the others being similar:

$$\begin{aligned}
\mathsf{Ev}(s,\alpha,0) &= \uparrow \\
\mathsf{Ev}(\langle \underline{\mathsf{let}}(x,M,N),\pi\rangle,\alpha,n+1) &= \mathsf{Ev}(\langle M,\mathsf{let}(x,N)\cdot\pi\rangle,\alpha,n) \\
\mathsf{Ev}(\langle \underline{\mathsf{coin}}(r),\pi\rangle,\langle 0\rangle\alpha,n+1) &= \mathsf{Ev}(\langle \underline{i},\pi\rangle,\alpha,n)\cdot\nu_i(r) \\
\mathsf{Ev}(\langle \underline{\mathsf{coin}}(r),\pi\rangle,\langle\rangle,n+1) &= \uparrow \\
\mathsf{Ev}(\langle \underline{\mathsf{coin}}(r),\pi\rangle,\uparrow,n+1) &= \uparrow \\
\mathsf{Ev}(\langle \underline{0},\varepsilon\rangle,\langle\rangle,n+1) &= 1
\end{aligned} \tag{2.5}$$

where of course multiplication is extended by $r\uparrow = \uparrow r = \uparrow$ and similarly for concatenation.

Remark 2.3. One obvious and important feature of this definition, shared by all the forthcoming definitions based on similar step-indexing, is that if $\mathsf{Ev}(s,\alpha,n) \neq \uparrow$, we have $\mathsf{Ev}(s,\alpha,k) = \mathsf{Ev}(s,\alpha,n) \neq \uparrow$ for all $k \geq n$. This property will be referred to as "monotonicity of step-indexing".

Then $\mathsf{Ev}(s,\alpha)$ is defined, and has value r, iff $\mathsf{Ev}(s,\alpha,n) = r$ for some n (and then the same will hold for any greater n).

Thanks to this step-indexing, the proof of Equation (2.4) boils down to the following lemma.

Lemma 2.4. *For all $n \in \mathbb{N}$, one has*

$$\forall n \in \mathbb{N} \quad \mathsf{Ev}_{\mathsf{lc}}(s,\alpha,\vec{\beta},n) = \mathsf{Ev}(\mathsf{strip}(s),\mathsf{Ev}^{\eta}_{\mathsf{lc}}(s,\alpha,\vec{\beta},n),n)\,.$$

Proof. Assume that the property holds for all integer $p < n$ and let us prove it for n. One reasons by cases on the shape of s considering only a few cases, the others being similar. Notice that the equation is obvious if $n = 0$ since then both hand-sides are $= \uparrow$ so we can assume $n > 0$.

- $s = \langle \underline{\mathsf{let}}(x,M,N),\pi\rangle$. By definition of the machine $\mathsf{Ev}_{\mathsf{lc}}$ we have $\mathsf{Ev}_{\mathsf{lc}}(s,\alpha,\vec{\beta},n) = \mathsf{Ev}_{\mathsf{lc}}(t,\alpha,\vec{\beta},n-1)$ and $\mathsf{Ev}(\mathsf{strip}(s),\mathsf{Ev}^{\eta}_{\mathsf{lc}}(s,\alpha,\vec{\beta},n),n) = \mathsf{Ev}(\mathsf{strip}(t),\mathsf{Ev}^{\eta}_{\mathsf{lc}}(t,\alpha,\vec{\beta},n-1),n-1)$ where $t = \langle M,\mathsf{let}(x,N)\cdot\pi\rangle$ and the inductive hypothesis applies.
- $s = \langle \underline{\mathsf{coin}}(r),\pi\rangle$. We have $\mathsf{Ev}_{\mathsf{lc}}(s,\alpha,\vec{\beta},n) = \uparrow = \mathsf{Ev}(\mathsf{strip}(s),\mathsf{Ev}^{\eta}_{\mathsf{lc}}(s,\alpha,\vec{\beta},n),n)$ if $\alpha = \langle\rangle$ or $\alpha = \uparrow$. We have $\mathsf{Ev}_{\mathsf{lc}}(s,\langle i\rangle\alpha,\vec{\beta},n) = \mathsf{Ev}_{\mathsf{lc}}(t,\alpha,\vec{\beta},n-1)\cdot\nu_i(r)$ and, setting $t = \langle \underline{i},\pi\rangle$,
$$\begin{aligned}
\mathsf{Ev}(\mathsf{strip}(s),\mathsf{Ev}^{\eta}_{\mathsf{lc}}(s,\langle i\rangle\alpha,\vec{\beta},n),n) &= \mathsf{Ev}(\mathsf{strip}(s),\langle i\rangle\mathsf{Ev}^{\eta}_{\mathsf{lc}}(t,\alpha,\vec{\beta},n-1),n) \\
&= \mathsf{Ev}(\mathsf{strip}(t),\mathsf{Ev}^{\eta}_{\mathsf{lc}}(t,\alpha,\vec{\beta},n-1),n-1)\cdot\nu_i(r)
\end{aligned}$$
and the inductive hypothesis applies.
- As a last example assume that $s = \langle \mathsf{lcoin}(l,r),\pi\rangle$. We have $\mathsf{Ev}_{\mathsf{lc}}(s,\alpha,\vec{\beta},n) = \uparrow = \mathsf{Ev}(\mathsf{strip}(s),\mathsf{Ev}^{\eta}_{\mathsf{lc}}(s,\alpha,\vec{\beta},n),n)$ if $\beta(l) = \langle\rangle$ or $\beta(l) = \uparrow$. If $\beta(l) = \langle i\rangle\gamma$ we have $\mathsf{Ev}_{\mathsf{lc}}(s,\alpha,\vec{\beta},n) = \mathsf{Ev}_{\mathsf{lc}}(\langle \underline{i},\pi\rangle,\alpha,\vec{\beta}\,[\gamma/l]\,,n-1)\cdot\nu_i(r)$ and
$$\begin{aligned}
\mathsf{Ev}(\mathsf{strip}(s),\mathsf{Ev}^{\eta}_{\mathsf{lc}}(s,\alpha,\vec{\beta},n),n) &= \mathsf{Ev}(\langle \underline{\mathsf{coin}}(r),\mathsf{strip}(\pi)\rangle,\langle i\rangle\mathsf{Ev}^{\eta}_{\mathsf{lc}}(\langle \underline{i},\pi\rangle,\alpha,\vec{\beta}\,[\gamma/l]\,,n-1),n) \\
&= \mathsf{Ev}(\langle \underline{i},\mathsf{strip}(\pi)\rangle,\mathsf{Ev}^{\eta}_{\mathsf{lc}}(\langle \underline{i},\pi\rangle,\alpha,\vec{\beta}\,[\gamma/l]\,,n-1),n-1)\cdot\nu_i(r)
\end{aligned}$$
and the inductive hypothesis applies.

∎

Let $\mathsf{strip}(s) \in \mathsf{S}$ be obtained by stripping s from its labels (so that $\mathsf{strip}(\mathsf{lcoin}(l,r)) = \underline{\mathsf{coin}}(r)$). And $\mathsf{strip}(M) \in \mathsf{pPCF}$ is defined similarly.

Equation (2.4) shows in particular that

$$\forall(\alpha,\vec{\beta}) \in \mathcal{D}_{\mathsf{lc}}(s) \quad \mathsf{Ev}^{\eta}_{\mathsf{lc}}(s,\alpha,\vec{\beta}) \in \mathcal{D}(\mathsf{strip}(s))\,.$$

Lemma 2.5. *The function*

$$\begin{aligned} \eta_s : \mathcal{D}_{\mathsf{lc}}(s) &\to \mathcal{D}(\mathsf{strip}(s)) = \mathcal{D}_{\mathsf{lab}}(s) \\ (\alpha,\vec{\beta}) &\mapsto \mathsf{Ev}^{\eta}_{\mathsf{lc}}(s,\alpha,\vec{\beta}) \end{aligned}$$

is a bijection.

Proof. We provide explicitly an inverse function, defined as another machine $\mathsf{Ev}^{-\eta}_{\mathsf{lc}}(s,\alpha)$. Again we provide only a few cases

$$\begin{aligned} \mathsf{Ev}^{-\eta}_{\mathsf{lc}}(\langle\underline{\mathsf{let}}(x,M,N),\pi\rangle,\delta) &= \mathsf{Ev}^{-\eta}_{\mathsf{lc}}(\langle M,\mathsf{let}(x,N)\cdot\pi\rangle,\delta) \\ \mathsf{Ev}^{-\eta}_{\mathsf{lc}}(\langle\mathsf{lcoin}(l,r),\pi\rangle,\langle i\rangle\delta) &= (\alpha,\vec{\gamma}) \quad \text{if } \mathsf{Ev}^{-\eta}_{\mathsf{lc}}(\langle\underline{i},\pi\rangle,\delta) = (\alpha,\vec{\beta}), \\ &\qquad\quad \gamma(l) = \langle i\rangle\beta(l) \text{ and } \gamma(k) = \beta(k) \text{ for } k \neq l \\ \mathsf{Ev}^{-\eta}_{\mathsf{lc}}(\langle\underline{\mathsf{coin}}(r),\pi\rangle,\langle i\rangle\delta) &= (\langle i\rangle\alpha,\vec{\beta}) \quad \text{if } \mathsf{Ev}^{-\eta}_{\mathsf{lc}}(\langle\underline{i},\pi\rangle,\delta) = (\alpha,\vec{\beta}) \\ \mathsf{Ev}^{-\eta}_{\mathsf{lc}}(\langle\underline{0},\varepsilon\rangle,\langle\rangle) &= (\langle\rangle,\vec{\langle\rangle})\,. \end{aligned}$$

It is clear from this recursion that the partial function $\mathsf{Ev}^{-\eta}_{\mathsf{lc}}(s,\_)$ has $\mathcal{D}(\mathsf{strip}(s))$ as domain. Let us prove that

$$\forall(\alpha,\vec{\beta}) \in \mathcal{D}_{\mathsf{lc}}(s) \quad \mathsf{Ev}^{-\eta}_{\mathsf{lc}}(s,\mathsf{Ev}^{\eta}_{\mathsf{lc}}(s,\alpha,\vec{\beta})) = (\alpha,\vec{\beta})\,.$$

Considering a step-indexed version of $\mathsf{Ev}^{-\eta}_{\mathsf{lc}}$ with an additional parameter in $n \in \mathbb{N}$ defined along the same lines as (2.5), it suffices to prove that

$$\forall n \in \mathbb{N}\,\forall(\alpha,\vec{\beta}) \in \mathcal{D}_{\mathsf{lc}}(s) \quad \mathsf{Ev}^{\eta}_{\mathsf{lc}}(s,\alpha,\vec{\beta},n) \neq\uparrow \Rightarrow \mathsf{Ev}^{-\eta}_{\mathsf{lc}}(s,\mathsf{Ev}^{\eta}_{\mathsf{lc}}(s,\alpha,\vec{\beta},n),n) = (\alpha,\vec{\beta})\,. \tag{2.6}$$

The proof is by induction on n. So assume that the property holds for all integer $p < n$ and let us prove it for n. Assume that $\mathsf{Ev}^{\eta}_{\mathsf{lc}}(s,\alpha,\vec{\beta},n) \neq\uparrow$, which implies that $n > 0$. As usual we consider only a few interesting cases. All other cases are similar to the first one (and similarly trivial).

- Assume first that $s = \langle\underline{\mathsf{let}}(x,M,N),\pi\rangle$. Setting $t = \langle M,\mathsf{let}(x,N)\cdot\pi\rangle$ we have $\mathsf{Ev}^{\eta}_{\mathsf{lc}}(t,\alpha,\vec{\beta},n-1) = \mathsf{Ev}^{\eta}_{\mathsf{lc}}(s,\alpha,\vec{\beta},n) \neq\uparrow$. Then
$$\begin{aligned} \mathsf{Ev}^{-\eta}_{\mathsf{lc}}(s,\mathsf{Ev}^{\eta}_{\mathsf{lc}}(s,\alpha,\vec{\beta},n),n) &= \mathsf{Ev}^{-\eta}_{\mathsf{lc}}(t,\mathsf{Ev}^{\eta}_{\mathsf{lc}}(t,\alpha,\vec{\beta},n-1),n-1) \\ &= (\alpha,\vec{\beta}) \end{aligned}$$
by inductive hypothesis.
- Assume now that $s = \langle\underline{\mathsf{coin}}(r),\pi\rangle$. Since $\mathsf{Ev}^{\eta}_{\mathsf{lc}}(s,\alpha,\vec{\beta},n) \neq\uparrow$ we must have $\alpha \neq \langle\rangle$. Let us write $\alpha = \langle i\rangle\gamma$, then we have $\mathsf{Ev}^{\eta}_{\mathsf{lc}}(s,\alpha,\vec{\beta},n) = \langle i\rangle\mathsf{Ev}^{\eta}_{\mathsf{lc}}(t,\gamma,\vec{\beta},n-1)$ where $t = \langle\underline{i},\pi\rangle$ and we have $\mathsf{Ev}^{\eta}_{\mathsf{lc}}(t,\gamma,\vec{\beta},n-1) \neq\uparrow$. By inductive hypothesis it follows that
$$\mathsf{Ev}^{-\eta}_{\mathsf{lc}}(t,\mathsf{Ev}^{\eta}_{\mathsf{lc}}(t,\gamma,\vec{\beta},n),n) = (\gamma,\vec{\beta})$$

Then we have

$$\begin{aligned}\mathsf{Ev}_{\mathsf{lc}}^{-\eta}(s,\mathsf{Ev}_{\mathsf{lc}}^{\eta}(s,\alpha,\vec{\beta},n),n) &= \mathsf{Ev}_{\mathsf{lc}}^{-\eta}(s,\langle i\rangle\mathsf{Ev}_{\mathsf{lc}}^{\eta}(t,\gamma,\vec{\beta},n-1),n)\\ &= (\langle i\rangle\gamma,\vec{\beta})\\ &= (\alpha,\vec{\beta})\end{aligned}$$

by definition of $\mathsf{Ev}_{\mathsf{lc}}^{-\eta}$.

- The case $s=\langle\mathsf{lcoin}(l,r),\pi\rangle$ is similar to the previous one, dealing with $\beta(l)$ instead of α.

Now we prove that for all $\delta\in\mathcal{D}(\mathsf{strip}(s))$ one has $\mathsf{Ev}_{\mathsf{lc}}^{-\eta}(s,\delta)\in\mathcal{D}_{\mathsf{lc}}(s)$ and that

$$\forall\delta\in\mathcal{D}(\mathsf{strip}(s))\quad \mathsf{Ev}_{\mathsf{lc}}^{\eta}(s,\mathsf{Ev}_{\mathsf{lc}}^{-\eta}(s,\delta))=\delta\,.$$

It suffices to prove that

$$\forall n\in\mathbb{N}\,\forall\alpha\in\mathcal{D}(\mathsf{strip}(s))\quad \mathsf{Ev}_{\mathsf{lc}}^{-\eta}(s,\delta,n)\neq\uparrow\Rightarrow\mathsf{Ev}_{\mathsf{lc}}^{\eta}(s,\mathsf{Ev}_{\mathsf{lc}}^{-\eta}(s,\delta,n),n)=\delta$$

using as usual the step-indexed versions of our machines. The proof is by induction on n so assume that the property holds for all $p<n$ and let us prove it for n. Assume that $\mathsf{Ev}_{\mathsf{lc}}^{-\eta}(s,\delta,n)\neq\uparrow$, which implies $n>0$. We reason by cases on s, considering the same cases as above (the other cases, similar to the first one, are similarly trivial).

- Assume first that $s=\langle\underline{\mathsf{let}}(x,M,N),\pi\rangle$. Setting $t=\langle M,\mathsf{let}(x,N)\cdot\pi\rangle$ we know that $\mathsf{Ev}_{\mathsf{lc}}^{-\eta}(t,\delta,n-1)\neq\uparrow$ and hence by inductive hypothesis $\mathsf{Ev}_{\mathsf{lc}}^{\eta}(t,\mathsf{Ev}_{\mathsf{lc}}^{-\eta}(t,\delta,n-1),n-1)=\delta$, proving our contention.
- Assume next that $s=\langle\underline{\mathsf{coin}}(r),\pi\rangle$. Since $\mathsf{Ev}_{\mathsf{lc}}^{-\eta}(s,\delta,n)\neq\uparrow$ we know that $\delta\neq\langle\rangle$. So we can write $\delta=\langle i\rangle\theta$. Then we know that $\mathsf{Ev}_{\mathsf{lc}}^{-\eta}(s,\delta,n)=(\langle i\rangle\alpha,\vec{\beta})$ where $(\alpha,\vec{\beta})=\mathsf{Ev}_{\mathsf{lc}}^{-\eta}(\langle\underline{i},\pi\rangle,\delta,n-1)$. By inductive hypothesis we have $\mathsf{Ev}_{\mathsf{lc}}^{\eta}(\langle\underline{i},\pi\rangle,(\alpha,\vec{\beta}))=\theta$ and therefore

$$\begin{aligned}\mathsf{Ev}_{\mathsf{lc}}^{\eta}(s,\mathsf{Ev}_{\mathsf{lc}}^{-\eta}(s,\delta,n),n) &= \mathsf{Ev}_{\mathsf{lc}}^{\eta}(s,(\langle i\rangle\alpha,\vec{\beta}),n)\\ &= \langle i\rangle\theta\\ &= \delta\,.\end{aligned}$$

- The case $s=\langle\mathsf{lcoin}(l,r),\pi\rangle$ is similar to the previous one, dealing with $\beta(l)$ instead of α. □

Lemma 2.6. *For all $s\in\mathsf{S}_{\mathsf{lc}}(L)$*

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

Proof. We have

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

By equation (2.4) and by the bijective correspondence of Lemma 2.5. □

2.5. **Spying labeled terms in** pPCF**.** 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

$$\mathsf{lc}_{\vec{r}}(M^l) = \underline{\mathsf{if}}(\mathsf{lcoin}(l, r_l), \mathsf{lc}_{\vec{r}}(M), \Omega^\sigma)$$

where σ is the type[9] of M.

In Section 2.6, we will turn a *closed* labeled term M (with labels in the finite set L) into the term $\mathsf{sp}_{\vec{x}}(M)$, defined in such a way that $[\![\mathsf{strip}(\mathsf{lc}_{\vec{r}}(M))]\!]$ has a simple expression in terms of $\mathsf{sp}_{\vec{x}}(M)$ (Lemma 2.10), allowing to interpret 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 (2.10)). This in turn is the key to the proof of Theorem 2.11.

We write $\langle 0\rangle^k$ for the sequence consisting of k occurrences of 0.

Lemma 2.7. *Let $s \in \mathsf{S}_{\mathsf{lab}}(L)$. Then*

$$\mathcal{D}_{\mathsf{lc}}(\mathsf{lc}_{\vec{r}}(s)) = \{(\alpha, (\langle 0\rangle^{\mathsf{Ev}_{\mathsf{lab}}(s,\alpha)(l)})_{l\in L}) \mid \alpha \in \mathcal{D}(\mathsf{strip}(s))\}$$

Remember from Lemma 2.2 that $\mathcal{D}_{\mathsf{lab}}(s) = \mathcal{D}(\mathsf{strip}(s))$.

Proof. We show that for any $n \in \mathbb{N}$ and any $(\alpha, \vec{\beta}) \in \mathcal{C}_0 \times \mathcal{C}_0^L$, one has

$$\begin{aligned} &\forall n \in \mathbb{N}\ \big(\mathsf{Ev}_{\mathsf{lc}}(\mathsf{lc}_{\vec{r}}(s), \alpha, \vec{\beta}, n) \neq \uparrow \\ &\qquad\qquad \Rightarrow \exists k \in \mathbb{N}\ \mathsf{Ev}_{\mathsf{lab}}(s, \alpha, k) \neq \uparrow \text{ and } \forall l \in L\ \beta_l = \langle 0\rangle^{\mathsf{Ev}_{\mathsf{lab}}(s,\alpha,k)(l)}\big) \end{aligned} \tag{2.7}$$

$$\begin{aligned} &\forall n \in \mathbb{N}\ \big(\mathsf{Ev}_{\mathsf{lab}}(s, \alpha, n) \neq \uparrow \text{ and } \forall l \in L\ \beta_l = \langle 0\rangle^{\mathsf{Ev}_{\mathsf{lab}}(s,\alpha,n)(l)} \\ &\qquad\qquad \Rightarrow \exists k \in \mathbb{N}\ \mathsf{Ev}_{\mathsf{lc}}(\mathsf{lc}_{\vec{r}}(s), \alpha, \vec{\beta}, k) \neq \uparrow\big) \end{aligned} \tag{2.8}$$

using as before step-indexed versions of the various machines. But in the present situation we shall not have the same indexing on both sides of implications because the encoding $\mathsf{lc}_{\vec{r}}(s)$ requires additional execution steps.

We prove (2.7) by induction on $n \in \mathbb{N}$. Assume now that the implication holds for all integers $p < n$ and let us prove it for n, so assume that $\mathsf{Ev}_{\mathsf{lc}}(\mathsf{lc}_{\vec{r}}(s), \alpha, \vec{\beta}, n) \neq \uparrow$ which implies $n > 0$. We consider only three cases as to the shape of s, the other cases being completely similar to the first one. We use the following convention: if M is a labeled term we use M' for $\mathsf{lc}_{\vec{r}}(M)$ and similarly for stacks and states.

- Assume first that $s = \langle \underline{\mathsf{let}}(x, M, N), \pi\rangle$ and let $t = \langle M, \mathsf{let}(x, N) \cdot \pi\rangle$. We have $s' = \langle \underline{\mathsf{let}}(x, M', N'), \pi'\rangle$ and $t' = \langle M', \mathsf{let}(x, N') \cdot \pi'\rangle$. We have

  $$\mathsf{Ev}_{\mathsf{lc}}(t', \alpha, \vec{\beta}, n-1) = \mathsf{Ev}_{\mathsf{lc}}(s', \alpha, \vec{\beta}, n) \neq \uparrow$$

  and hence by inductive hypothesis there is $k \in \mathbb{N}$ such that

  $$\mathsf{Ev}_{\mathsf{lab}}(t, \alpha, k) \neq \uparrow \text{ and } \forall l \in L\ \beta_l = \langle 0\rangle^{\mathsf{Ev}_{\mathsf{lab}}(t,\alpha,k)(l)}\,.$$

[9] *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.

It follows that $\mathsf{Ev}_{\mathsf{lab}}(s,\alpha,k{+}1) = \mathsf{Ev}_{\mathsf{lab}}(t,\alpha,k) \neq \uparrow$ and $\mathsf{Ev}_{\mathsf{lab}}(s,\alpha,k{+}1)(l) = \mathsf{Ev}_{\mathsf{lab}}(t,\alpha,k)(l)$. Therefore we have

$$\mathsf{Ev}_{\mathsf{lab}}(s,\alpha,k+1) \neq \uparrow \text{ and } \forall l \in L\ \beta_l = \langle 0\rangle^{\mathsf{Ev}_{\mathsf{lab}}(s,\alpha,k+1)(l)}\,.$$

- Assume now that $s = \langle \underline{\mathsf{coin}}(r), \pi\rangle$ so that $s' = \langle \underline{\mathsf{coin}}(r), \pi'\rangle$. Since $\mathsf{Ev}_{\mathsf{lc}}(s',\alpha,\vec{\beta},n) \neq \uparrow$ we know that $\alpha = \langle i\rangle\gamma$ for some $i \in \{0,1\}$ and that $\mathsf{Ev}_{\mathsf{lc}}(t',\gamma,\vec{\beta},n-1) \neq \uparrow$ where $t = \langle \underline{i}, \pi\rangle$ (and hence $t' = \langle \underline{i}, \pi'\rangle$). By inductive hypothesis there is $k \in \mathbb{N}$ such that

$$\mathsf{Ev}_{\mathsf{lab}}(t,\gamma,k) \neq \uparrow \text{ and } \forall l \in L\ \beta_l = \langle 0\rangle^{\mathsf{Ev}_{\mathsf{lab}}(t,\gamma,k)(l)}\,.$$

  It follows that $\mathsf{Ev}_{\mathsf{lab}}(s,\alpha,k+1) = \mathsf{Ev}_{\mathsf{lab}}(t,\gamma,k) \neq \uparrow$. Therefore we have

$$\mathsf{Ev}_{\mathsf{lab}}(s,\alpha,k+1) \neq \uparrow \text{ and } \forall l \in L\ \beta_l = \langle 0\rangle^{\mathsf{Ev}_{\mathsf{lab}}(s,\alpha,k+1)(l)}\,.$$

- The third case we consider is $s = \langle M^l, \pi\rangle$ for some $l \in L$ so that

$$s' = \langle \underline{\mathsf{if}}(\mathsf{lcoin}(l, r_l), M', \Omega^\sigma), \pi'\rangle$$

  where σ is the type of M. Since $\mathsf{Ev}_{\mathsf{lc}}(s',\alpha,\vec{\beta},n) \neq \uparrow$ we must have $n \geq 2$ and $\beta_l = \langle i\rangle\gamma$; indeed, setting $\vec{\beta'} = \vec{\beta}\,[\gamma/l]$,

$$\begin{aligned}\mathsf{Ev}_{\mathsf{lc}}(s',\alpha,\vec{\beta},n) &= \mathsf{Ev}_{\mathsf{lc}}(\langle \mathsf{lcoin}(l,r_l), \mathsf{if}(M',\Omega^\sigma)\cdot\pi'\rangle, \alpha, \vec{\beta}, n-1)\\ &= \mathsf{Ev}_{\mathsf{lc}}(\langle \underline{i}, \mathsf{if}(M',\Omega^\sigma)\cdot\pi'\rangle, \alpha, \vec{\beta'}, n-2)\\ &= \begin{cases}\mathsf{Ev}_{\mathsf{lc}}(\langle M', \pi'\rangle, \alpha, \vec{\beta'}, n-2) & \text{if } i = 0\\ \mathsf{Ev}_{\mathsf{lc}}(\langle \Omega^\sigma, \pi'\rangle, \alpha, \vec{\beta'}, n-2) & \text{if } i = 1\,.\end{cases}\end{aligned}$$

  But whatever is the value of $n \geq 2$ we have $\mathsf{Ev}_{\mathsf{lc}}(\langle \Omega^\sigma, \pi'\rangle, \alpha, \vec{\beta'}, n-2) = \uparrow$ by definition of Ω^σ. It follows that we must have $i = 0$ and $\mathsf{Ev}_{\mathsf{lc}}(\langle M', \pi'\rangle, \alpha, \vec{\beta'}, n-2) \neq \uparrow$. By inductive hypothesis there exists $k \in \mathbb{N}$ such that, setting $t = \langle M, \pi\rangle$

$$\mathsf{Ev}_{\mathsf{lab}}(t,\alpha,k) \neq \uparrow \text{ and } \forall m \in L\ \beta'_m = \langle 0\rangle^{\mathsf{Ev}_{\mathsf{lab}}(t,\alpha,k)(m)}\,.$$

  We have $\mathsf{Ev}_{\mathsf{lab}}(s,\alpha,k+1) = \mathsf{Ev}_{\mathsf{lab}}(s,\alpha,k) + [l] \neq \uparrow$. It follows that if $m \in L \setminus \{l\}$ one has $\beta_m = \beta'_m = \langle 0\rangle^{\mathsf{Ev}_{\mathsf{lab}}(t,\alpha,k)(m)} = \langle 0\rangle^{\mathsf{Ev}_{\mathsf{lab}}(s,\alpha,k+1)(m)}$, and $\beta_l = \langle 0\rangle\beta'_l = \langle 0\rangle\langle 0\rangle^{\mathsf{Ev}_{\mathsf{lab}}(t,\alpha,k)(l)} = \langle 0\rangle^{\mathsf{Ev}_{\mathsf{lab}}(s,\alpha,k+1)(l)}$ proving our contention.

This ends the proof of (2.7), we prove now (2.8), also by induction on n. Assume that the implication holds for all $p < n$ and let us prove it for n so assume that

$$\mathsf{Ev}_{\mathsf{lab}}(s,\alpha,n) \neq \uparrow \text{ and } \forall l \in L\ \beta_l = \langle 0\rangle^{\mathsf{Ev}_{\mathsf{lab}}(s,\alpha,n)(l)}\,.$$

As usual this implies that $n > 0$. We deal with the same three cases as in the proof of (2.7).

- Assume first that $s = \langle \underline{\mathsf{let}}(x, M, N), \pi\rangle$ and let $t = \langle M, \mathsf{let}(x, N)\cdot\pi\rangle$. We have $s' = \langle \underline{\mathsf{let}}(x, M', N'), \pi'\rangle$ and $t' = \langle M', \mathsf{let}(x, N')\cdot\pi'\rangle$. We have $\mathsf{Ev}_{\mathsf{lab}}(t,\alpha,n{-}1) = \mathsf{Ev}_{\mathsf{lab}}(s,\alpha,n) \neq \uparrow$. Also we have, for all $l \in L$, $\beta_l = \langle 0\rangle^{\mathsf{Ev}_{\mathsf{lab}}(s,\alpha,n)(l)} = \langle 0\rangle^{\mathsf{Ev}_{\mathsf{lab}}(t,\alpha,n-1)(l)}$. Hence by ind. hypothesis there is $k \in \mathbb{N}$ such that $\mathsf{Ev}_{\mathsf{lc}}(t',\alpha,\vec{\beta},k) \neq \uparrow$ and hence $\mathsf{Ev}_{\mathsf{lc}}(s',\alpha,\vec{\beta},k+1) = \mathsf{Ev}_{\mathsf{lc}}(t',\alpha,\vec{\beta},k) \neq \uparrow$.
- Assume now that $s = \langle \underline{\mathsf{coin}}(r), \pi\rangle$ so that $s' = \langle \underline{\mathsf{coin}}(r), \pi'\rangle$. Since $\mathsf{Ev}_{\mathsf{lab}}(s,\alpha,n) \neq \uparrow$ we must have $\alpha = \langle i\rangle\gamma$ for some $i \in \{0,1\}$ and we have $\mathsf{Ev}_{\mathsf{lab}}(t,\gamma,n-1) = \mathsf{Ev}_{\mathsf{lab}}(s,\alpha,n) \neq \uparrow$ where $t = \langle \underline{i}, \pi\rangle$. Moreover $\mathsf{Ev}_{\mathsf{lab}}(s,\alpha,n) = \mathsf{Ev}_{\mathsf{lab}}(t,\gamma,n-1)$ and hence for each $l \in L$ we have $\beta_l = \langle 0\rangle^{\mathsf{Ev}_{\mathsf{lab}}(s,\alpha,n)(l)} = \langle 0\rangle^{\mathsf{Ev}_{\mathsf{lab}}(t,\gamma,n-1)(l)}$. By inductive hypothesis there exists k such that $\mathsf{Ev}_{\mathsf{lc}}(t',\gamma,\vec{\beta},k) \neq \uparrow$. We have $\mathsf{Ev}_{\mathsf{lc}}(s',\alpha,\vec{\beta},k+1) = \mathsf{Ev}_{\mathsf{lc}}(t',\gamma,\vec{\beta},k) \neq \uparrow$ as expected.

- Assume last that $s = \langle M^l, \pi\rangle$ for some $l \in L$ so that $s' = \langle \underline{\mathsf{if}}(\mathsf{lcoin}(l, r_l), M', \Omega^\sigma), \pi'\rangle$ where σ is the type of M. Let $t = \langle M, \pi\rangle$. Since $\mathsf{Ev}_{\mathsf{lab}}(s, \alpha, n) \neq \uparrow$ we have $\mathsf{Ev}_{\mathsf{lab}}(t, \alpha, n-1) \neq \uparrow$ and $\mathsf{Ev}_{\mathsf{lab}}(s, \alpha, n) = \mathsf{Ev}_{\mathsf{lab}}(t, \alpha, n-1) + [l]$. We also know that for all $m \in L$
$$\beta_l = \langle 0\rangle^{\mathsf{Ev}_{\mathsf{lab}}(s,\alpha,n)(m)}$$
In particular $\beta_l = \langle 0\rangle\beta'_l$. Setting $\beta'_m = \beta_m$ for $m \neq l$, we have therefore
$$\forall m \in L \quad \beta'_m = \langle 0\rangle^{\mathsf{Ev}_{\mathsf{lab}}(t,\alpha,n)(m)}\,.$$
By inductive hypothesis there is $k \in \mathbb{N}$ such that $\mathsf{Ev}_{\mathsf{lc}}(\mathsf{lc}_{\vec{r}}(t), \alpha, \vec{\beta'}, k) \neq \uparrow$. Since $s' = \mathsf{lc}_{\vec{r}}(s) = \underline{\mathsf{if}}(\mathsf{lcoin}(l, r_l), M', \Omega^\sigma)$, setting $\vec{\beta} = \vec{\beta'}\,[\langle 0\rangle\beta'_l/l]$ we have
$$\begin{aligned}
\mathsf{Ev}_{\mathsf{lc}}(s', \alpha, \vec{\beta}, k+3) &= \mathsf{Ev}_{\mathsf{lc}}(\langle \mathsf{lcoin}(l, r_l), \mathsf{if}(M', \Omega^\sigma)\cdot\pi'\rangle, \alpha, \vec{\beta}, k+2)\\
&= \mathsf{Ev}_{\mathsf{lc}}(\langle \underline{0}, \mathsf{if}(M', \Omega^\sigma)\cdot\pi'\rangle, \alpha, \vec{\beta'}, k+1)\cdot r_l \quad \text{by definition of } \vec{\beta'}\\
&= \mathsf{Ev}_{\mathsf{lc}}(\langle M', \pi'\rangle, \alpha, \vec{\beta'}, k)\cdot r_l\\
&= \mathsf{Ev}_{\mathsf{lc}}(t', \alpha, \vec{\beta'}, k)\cdot r_l\\
&\neq \uparrow
\end{aligned}$$

This ends the proof of (2.8). □

To understand the next lemma, remember that $\mathsf{Ev}_{\mathsf{lab}}(s, \alpha) \in \mathcal{M}_{\mathrm{fin}}(L)$ so that $(\vec{r})^{\mathsf{Ev}_{\mathsf{lab}}(s,\alpha)} = \prod_{l\in L} r_l^{\mathsf{Ev}_{\mathsf{lab}}(s,\alpha)(l)}$.

Lemma 2.8. *Let $s \in \mathsf{S}_{\mathsf{lab}}(L)$. Then*
$$\mathsf{Ev}_{\mathsf{lc}}(\mathsf{lc}_{\vec{r}}(s), \alpha, (\langle 0\rangle^{\mathsf{Ev}_{\mathsf{lab}}(s,\alpha)(l)})_{l\in L}) = \mathbb{P}_{\mathsf{strip}(s)}(\{\langle 0\rangle\alpha\})(\vec{r})^{\mathsf{Ev}_{\mathsf{lab}}(s,\alpha)}$$
for all $\alpha \in \mathcal{D}(\mathsf{strip}(s)) = \mathcal{D}_{\mathsf{lab}}(s)$.

Proof. By definition
$$\mathbb{P}_{\mathsf{strip}(s)}(\{\langle 0\rangle\alpha\}) = \mathsf{Ev}(\mathsf{strip}(s), \alpha) \quad \text{and} \quad (\vec{r})^{\mathsf{Ev}_{\mathsf{lab}}(s,\alpha)} = \prod_{l\in L} r_l^{\mathsf{Ev}_{\mathsf{lab}}(s,\alpha)(l)}$$
so we have to prove that
$$\mathsf{Ev}_{\mathsf{lc}}(\mathsf{lc}_{\vec{r}}(s), \alpha, (\langle 0\rangle^{\mathsf{Ev}_{\mathsf{lab}}(s,\alpha)(l)})_{l\in L}) = \mathsf{Ev}(\mathsf{strip}(s), \alpha)\prod_{l\in L} r_l^{\mathsf{Ev}_{\mathsf{lab}}(s,\alpha)(l)}\,.$$
By Lemmas 2.2 and 2.7 we know that these two expressions are defined (that is, all subexpressions are defined) if and only if $\alpha \in \mathcal{D}_{\mathsf{lab}}(s)$.

By induction on n, we prove that if both expressions
$$\mathsf{Ev}_{\mathsf{lc}}(\mathsf{lc}_{\vec{r}}(s), \alpha, (\langle 0\rangle^{\mathsf{Ev}_{\mathsf{lab}}(s,\alpha,n)(l)})_{l\in L}, n) \quad \text{and} \quad \mathsf{Ev}(\mathsf{strip}(s), \alpha, n)\prod_{l\in L} r_l^{\mathsf{Ev}_{\mathsf{lab}}(s,\alpha,n)(l)}$$
are $\neq\uparrow$, then they are equal. Assume that the property holds for all $p < n$ and let us prove it for n, so assume that both
$$\mathsf{Ev}_{\mathsf{lc}}(\mathsf{lc}_{\vec{r}}(s), \alpha, (\langle 0\rangle^{\mathsf{Ev}_{\mathsf{lab}}(s,\alpha,k)(l)})_{l\in L}, k) \quad \text{and} \quad \mathsf{Ev}(\mathsf{strip}(s), \alpha, k)\prod_{l\in L} r_l^{\mathsf{Ev}_{\mathsf{lab}}(s,\alpha,k)(l)}$$
are defined, which implies $n > 0$. We consider the three usual cases as to s.

- Assume first that $s = \langle \underline{\mathsf{let}}(x, M, N), \pi\rangle$ and let $t = \langle M, \mathsf{let}(x, N) \cdot \pi\rangle$. We have, using as usual the notation $u' = \mathsf{lc}_{\vec{r}}(u)$,

$$\begin{aligned}\mathsf{Ev_{lc}}(s', \alpha, (\langle 0\rangle^{\mathsf{Ev_{lab}}(s,\alpha,n)(l)})_{l\in L}, n) &= \mathsf{Ev_{lc}}(t', \alpha, (\langle 0\rangle^{\mathsf{Ev_{lab}}(t,\alpha,n-1)(l)})_{l\in L}, n-1)\\ &= \mathsf{Ev}(\mathsf{strip}(t), \alpha, n-1) \prod_{l\in L} r_l^{\mathsf{Ev_{lab}}(t,\alpha,n-1)(l)} \quad \text{by ind. hyp.}\\ &= \mathsf{Ev}(\mathsf{strip}(s), \alpha, n) \prod_{l\in L} r_l^{\mathsf{Ev_{lab}}(s,\alpha,n)(l)}\,.\end{aligned}$$

- Assume now that $s = \langle \underline{\mathsf{coin}}(r), \pi\rangle$ so that $s' = \langle \underline{\mathsf{coin}}(r), \pi'\rangle$. Since $\mathsf{Ev_{lab}}(s, \alpha, n) \neq \uparrow$ we must have $\alpha = \langle i\rangle\gamma$ for some $i \in \{0, 1\}$, and we have $\mathsf{Ev_{lab}}(t, \gamma, n-1) = \mathsf{Ev_{lab}}(s, \alpha, n) \neq \uparrow$ where $t = \langle \underline{i}, \pi\rangle$. Moreover $\mathsf{Ev_{lab}}(s, \alpha, n) = \mathsf{Ev_{lab}}(t, \gamma, n-1)$ and hence for each $l \in L$ we have $\beta_l = \langle 0\rangle^{\mathsf{Ev_{lab}}(s,\alpha,n)(l)} = \langle 0\rangle^{\mathsf{Ev_{lab}}(t,\gamma,n-1)(l)}$. We have

$$\begin{aligned}&\mathsf{Ev_{lc}}(s', \alpha, (\langle 0\rangle^{\mathsf{Ev_{lab}}(s,\alpha,n)(l)})_{l\in L}, n)\\ &\quad= \mathsf{Ev_{lc}}(t', \gamma, (\langle 0\rangle^{\mathsf{Ev_{lab}}(t,\alpha,n-1)(l)})_{l\in L}, n-1) \cdot \nu_i(r)\\ &\quad= \mathsf{Ev}(\mathsf{strip}(t), \gamma, n-1) \cdot \nu_i(r) \prod_{l\in L} r_l^{\mathsf{Ev_{lab}}(t,\gamma,n-1)(l)} \quad \text{by ind. hyp.}\\ &\quad= \mathsf{Ev}(\mathsf{strip}(s), \alpha, n) \prod_{l\in L} r_l^{\mathsf{Ev_{lab}}(s,\alpha,n)(l)}\,.\end{aligned}$$

- Assume last that $s = \langle M^l, \pi\rangle$ for some $l \in L$ so that $s' = \langle \underline{\mathsf{if}}(\mathsf{lcoin}(l, r_l), M', \Omega^\sigma), \pi'\rangle$ where σ is the type of M. Let $t = \langle M, \pi\rangle$. We have already noticed that $n > 0$; actually, since $\mathsf{Ev_{lc}}(s', \alpha, n) \neq \uparrow$, we have $n \geq 3$, see below. Moreover $\mathsf{Ev_{lab}}(t, \alpha, n-1) \neq \uparrow$ and $\mathsf{Ev_{lab}}(s, \alpha, n) = \mathsf{Ev_{lab}}(t, \alpha, n-1) + [l]$. Let $\vec{\beta} = (\langle 0\rangle^{\mathsf{Ev_{lab}}(s,\alpha,n)(m)})_{m\in L}$, so that $\beta_l = \langle 0\rangle\beta'_l$ and $\beta_m = \beta'_m$ for $m \neq l$, where $\vec{\beta'} = (\langle 0\rangle^{\mathsf{Ev_{lab}}(t,\alpha,n-1)(m)})_{m\in L}$. We have

$$\begin{aligned}\mathsf{Ev_{lc}}(s', \alpha, \vec{\beta}, n) &= \mathsf{Ev_{lc}}(\langle \mathsf{lcoin}(l, r_l), \mathsf{if}(M', \Omega^\sigma) \cdot \pi'\rangle, \alpha, \vec{\beta}, n-1)\\ &= \mathsf{Ev_{lc}}(\langle \underline{0}, \mathsf{if}(M', \Omega^\sigma) \cdot \pi'\rangle, \alpha, \vec{\beta'}, n-2) \cdot r_l\\ &= \mathsf{Ev_{lc}}(t', \alpha, (\langle 0\rangle^{\mathsf{Ev_{lab}}(t,\alpha,n-1)(m)})_{m\in L}, n-3) \cdot r_l\\ &= \mathsf{Ev_{lc}}(t', \alpha, (\langle 0\rangle^{\mathsf{Ev_{lab}}(t,\alpha,n-1)(m)})_{m\in L}, n-1) \cdot r_l\\ &\qquad\qquad \text{by monotonicity of step-indexing}\\ &= \mathsf{Ev}(\mathsf{strip}(t), \alpha, n-1) \cdot r_l \prod_{m\in L} r_m^{\mathsf{Ev_{lab}}(t,\gamma,n-1)(m)} \quad \text{by ind. hyp.}\\ &= \mathsf{Ev}(\mathsf{strip}(s), \alpha, n) \prod_{m\in L} r_m^{\mathsf{Ev_{lab}}(s,\alpha,n)(m)}\end{aligned}$$

  by monotonicity of step-indexing, since $\mathsf{strip}(s) = \mathsf{strip}(t)$.

□

2.6. **The spying translation.** We consider a last translation, from $\mathsf{pPCF_{lab}}(L)$ to pPCF: let $\vec{x}$ be an L-indexed family of pairwise distinct variables (that we identify with the typing context $(x_l : \iota)_{l\in L}$). If $M \in \mathsf{pPCF_{lab}}(L)$ 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 $\mathsf{sp}_{\vec{x}}(M^l) = \underline{\mathsf{if}}(x_l, \mathsf{sp}_{\vec{x}}(M), \Omega^\sigma)$ where σ is the type of M. As another example, we set $\mathsf{sp}_{\vec{x}}(\lambda y^\tau\, M) = \lambda y^\tau\, \mathsf{sp}_{\vec{x}}(M)$ assuming of course that y is distinct from all x_l's.

The following lemma is not technically essential, it is simply an observation useful to understand better what follows.

Lemma 2.9. *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 $\forall l \in L\ \ \mathsf{supp}(\rho_l) \subseteq \{0\}$.*

Let $f : \mathsf{PN}^L \to \mathsf{P}[\![\sigma]\!]$ be the analytic function induced by $[\![\mathsf{sp}_{\vec{x}}(M)]\!]_{\vec{x}}$. This lemma says that $f(\vec{u})$ (where $\vec{u} \in \mathsf{PN}^L$) depends only on $(u(l)_0)_{l\in L} \in [0,1]^L$, that is, on the 0-component of the $u(l)$'s which are probability sub-distributions on $\mathbb{N}$.

Proof. (Sketch) Simple induction on M considering also open terms: we prove that, if $\Gamma \vdash M : \sigma$ with $\Gamma = (y_1 : \tau_1, \dots, y_k : \tau_k)$, so that $\Gamma, \vec{x} \vdash \mathsf{sp}_{\vec{x}}(M) : \sigma$, then given $\mu_i \in \mathcal{M}_{\mathrm{fin}}(|[\![\tau_i]\!]|)$ for $i = 1, \dots, k$, $a \in |[\![\sigma]\!]|$ and $\vec{\rho} \in \mathcal{M}_{\mathrm{fin}}(\mathbb{N})^L$, if $([\![\mathsf{sp}_{\vec{x}}(M)]\!]_{\Gamma,\vec{x}})_{(\vec{\mu},\vec{\rho},a)} \neq 0$ then $\forall l \in L\, \forall n \in \mathbb{N}\ \rho_l(n) \neq 0 \Rightarrow n = 0$. Considering $[\![\mathsf{sp}_{\vec{x}}(M)]\!]_{\Gamma,\vec{x}}$ as a function

$$[\![\mathsf{sp}_{\vec{x}}(M)]\!]_{\Gamma,\vec{x}} : \prod_{i=1}^{k} \mathsf{P}[\![\tau_i]\!] \times \mathsf{PN}^L \to \mathsf{P}[\![\sigma]\!]$$

(see Section 2.1.1) this amounts to showing that given $\vec{v} \in \prod_{i=1}^k \mathsf{P}[\![\tau_i]\!]$ and $\vec{u}, \vec{u'} \in \mathsf{PN}^L$ then

$$(\forall l \in L\ u(l)_0 = u'(l)_0) \Rightarrow [\![\mathsf{sp}_{\vec{x}}(M)]\!]_{\Gamma,\vec{x}}(\vec{v}, \vec{u}) = [\![\mathsf{sp}_{\vec{x}}(M)]\!]_{\Gamma,\vec{x}}(\vec{v}, \vec{u'})\,.$$

In other words the function $[\![\mathsf{sp}_{\vec{x}}(M)]\!]_{\Gamma,\vec{x}}(\vec{v}, \vec{u})$ of $\vec{u}$ depends only on the values taken by the $u(l)$'s (the components of $\vec{u}$) on $0 \in |\mathsf{N}| = \mathbb{N}$. This follows by a straightforward induction on M, the only "interesting" (though obvious) case being when M is of shape N^l: in this case we have

$$[\![\mathsf{sp}_{\vec{x}}(M)]\!]_{\Gamma,\vec{x}}(\vec{v}, \vec{u}) = u(l)_0 [\![\mathsf{sp}_{\vec{x}}(N)]\!]_{\Gamma,\vec{x}}(\vec{v}, \vec{u})$$

since $[\![\Omega^\sigma]\!] = 0$. □

Lemma 2.10. *Let $\vec{r} \in (\mathbb{Q} \cap [0,1])^L$ and $M \in \mathsf{pPCF}_{\mathsf{lab}}(L)$ with $\vdash M : \tau$. Then*

$$[\![\mathsf{sp}_{\vec{x}}(M)]\!]_{\vec{x}}(\vec{r}\,\mathsf{e}_0) = [\![\mathsf{strip}(\mathsf{lc}_{\vec{r}}(M))]\!]\,.$$

Proof. (Sketch) One proves more generally that given M such that $\Gamma \vdash M : \tau$ with $\Gamma = (y_1 : \sigma_1, \dots, y_k : \sigma_k)$, so that $\Gamma, \vec{x} \vdash \mathsf{sp}_{\vec{x}}(M) : \tau$, and $\vec{v} \in \prod_{i=1}^k \mathsf{P}[\![\sigma_i]\!]$, one has

$$[\![\mathsf{sp}_{\vec{x}}(M)]\!]_{\Gamma,\vec{x}}(\vec{u}, \vec{r}\,\mathsf{e}_0) = [\![\mathsf{strip}(\mathsf{lc}_{\vec{r}}(M))]\!]_{\Gamma}(\vec{u})$$

by a simple induction on M. The only interesting case is when $M = N^l$:

$$\begin{aligned}
[\![\mathsf{sp}_{\vec{x}}(M)]\!]_{\Gamma,\vec{x}}(\vec{u}, \vec{r}\,\mathsf{e}_0) &= (\vec{r}\,\mathsf{e}_0)(l)_0 [\![\mathsf{sp}_{\vec{x}}(N)]\!]_{\Gamma,\vec{x}}(\vec{u}, \vec{r}\,\mathsf{e}_0)\\
&= r(l)_0 [\![\mathsf{sp}_{\vec{x}}(N)]\!]_{\Gamma,\vec{x}}(\vec{u}, \vec{r}\,\mathsf{e}_0)\\
&= [\![\mathsf{strip}(\mathsf{lc}_{\vec{r}}(M))]\!]_{\Gamma}(\vec{u})
\end{aligned}$$

where $\vec{r}\,\mathsf{e}_0 = (r_l \mathsf{e}_0)_{l\in L} \in \mathsf{PN}^L$, using twice the fact that $[\![\Omega^\sigma]\!] = 0$. □

With the constructions and observations accumulated so far, we can prove the main result of this section. Let $M \in \mathsf{pPCF}_{\mathsf{lab}}(L)$ and $l \in L$. Remember from Section 2.2 that $\mathsf{Ev}_{\mathsf{lab}}(s)_l : \mathcal{C}_0 \to \mathbb{N}$ is the integer r.v. defined by $\mathsf{Ev}_{\mathsf{lab}}(s)_l(\alpha) = \mathsf{Ev}_{\mathsf{lab}}(s)(\alpha)(l)$, the number of times an l-labeled subterm of M has arrived in head position during the execution of M induced by $\alpha \in \mathcal{C}_0$. So this r.v. allows to evaluate the number of execution steps in this evaluation. For instance if M is obtained by labeling all sub-terms of a given closed term N of pPCF of type ι with the same label $l \in L$, we get an $\mathbb{N}$-valued r.v. which evaluates the number of execution steps in the evaluation of N that is, the number of times a subterm of N arrives in head position during this evaluation (notice that if N is a constant $\underline{n}$, this number is 1).

Given $\mu \in \mathcal{M}_{\mathrm{fin}}(L)$, we use $\mu\,[0]$ in the proof of the next result 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.

Theorem 2.11. *Let $M \in \mathsf{pPCF}_{\mathsf{lab}}(L)$ with $\vdash M : \iota$. Then for all $l \in L$*

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

Proof. By Lemma 2.10,

$$[\![\mathsf{strip}(\mathsf{lc}_{\vec{r}}(M))]\!]_0 = \sum_{\mu \in \mathcal{M}_{\mathrm{fin}}(L)} ([\![\mathsf{sp}_{\vec{x}}(M)]\!]_{\vec{x}})_{(\mu\,[0],0)} (\vec{r})^\mu \tag{2.9}$$

By Theorem 2.1, we have

$$\begin{aligned}
[\![\mathsf{strip}(\mathsf{lc}_{\vec{r}}(M))]\!]_0 &= \mathbb{P}_{\mathsf{strip}(\mathsf{lc}_{\vec{r}}(\langle M,\varepsilon\rangle))}(\mathsf{strip}(\mathsf{lc}_{\vec{r}}(\langle M, \varepsilon\rangle)) \downarrow \underline{0}) \\
&= \sum_{(\alpha,\vec{\beta}) \in \mathcal{D}_{\mathsf{lc}}(\mathsf{lc}_{\vec{r}}(\langle M,\varepsilon\rangle))} \mathsf{Ev}_{\mathsf{lc}}(\mathsf{lc}_{\vec{r}}(\langle M, \varepsilon\rangle), \alpha, \vec{\beta}) \quad \text{by Lemma 2.6} \\
&= \sum_{\alpha \in \mathcal{D}(\mathsf{strip}(\langle M,\varepsilon\rangle))} \mathsf{Ev}(\mathsf{strip}(\langle M, \varepsilon\rangle), \alpha) \prod_{l \in L} r_l^{\mathsf{Ev}_{\mathsf{lab}}(\langle M,\varepsilon\rangle,\alpha)(l)} \quad \text{by Lemma 2.8} \\
&= \sum_{\mu \in \mathcal{M}_{\mathrm{fin}}(L)} \left(\sum_{\substack{\alpha \in \langle 0\rangle\mathcal{C}_0 \\ \mathsf{Ev}_{\mathsf{lab}}(\langle M,\varepsilon\rangle)(\alpha) = \mu}} \mathsf{Ev}(\mathsf{strip}(\langle M, \varepsilon\rangle))(\alpha) \right) (\vec{r})^\mu
\end{aligned}$$

and since this holds for all $\vec{r} \in (\mathbb{Q} \cap [0,1])^L$, we must have by Equation (2.9), for all $\mu \in \mathcal{M}_{\mathrm{fin}}(L)$,

$$\begin{aligned}
([\![\mathsf{sp}_{\vec{x}}(M)]\!]_{\vec{x}})_{(\mu\,[0],0)} &= \sum_{\substack{\alpha \in \langle 0\rangle\mathcal{C}_0 \\ \mathsf{Ev}_{\mathsf{lab}}(\langle M,\varepsilon\rangle)(\alpha) = \mu}} \mathsf{Ev}(\mathsf{strip}(\langle M, \varepsilon\rangle))(\alpha) \\
&= \mathbb{P}_{\mathsf{strip}(\langle M,\varepsilon\rangle)}(\mathsf{Ev}_{\mathsf{lab}}(\langle M, \varepsilon\rangle) = \mu)
\end{aligned} \tag{2.10}$$

Let $l \in L$, we have

$$\begin{aligned}
\mathbb{E}(\mathsf{Ev}_{\mathsf{lab}}(\langle M, \varepsilon\rangle)_l) &= \sum_{\mu \in \mathcal{M}_{\mathrm{fin}}(L)} \mu(l) \mathbb{P}_{\langle M,\varepsilon\rangle}(\mathsf{Ev}_{\mathsf{lab}}(\langle M, \varepsilon\rangle) = \mu) \quad \text{by Equation (2.3)} \\
&= \sum_{\mu \in \mathcal{M}_{\mathrm{fin}}(L)} \mu(l) ([\![\mathsf{sp}_{\vec{x}}(M)]\!]_{\vec{x}})_{(\mu\,[0],0)} \quad \text{by Equation (2.10)} \\
&= \frac{\partial [\![\mathsf{sp}_{\vec{x}} M]\!]_{\vec{x}}(\vec{r}\mathsf{e}_0)_0}{\partial r_l}(1, \dots, 1)\,.
\end{aligned}$$

Indeed, given $\vec{r} \in [0,1]^L$ one has

$$[\![\mathsf{sp}_{\vec{x}} M]\!]_{\vec{x}}(\vec{r}\mathsf{e}_0)_0 = \sum_{\mu \in \mathcal{M}_{\mathrm{fin}}(L)} ([\![\mathsf{sp}_{\vec{x}}(M)]\!]_{\vec{x}})_{(\mu\,[0],0)} \vec{r}^\mu$$

and $\frac{\partial \vec{r}^{\mu}}{\partial r_l}(1, ..., 1) = \mu(l)$, whence the last equation. □

Example 2.12. 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 $\vdash M_q : \iota \Rightarrow \iota$:

$$M_q = \underline{\mathsf{fix}}(\lambda f^{\iota \Rightarrow \iota}\, \lambda x^{\iota}\, \underline{\mathsf{if}}(\underline{\mathsf{coin}}(q), \underline{\mathsf{if}}((f)x, \underline{\mathsf{if}}((f)x, \underline{0}, \Omega^{\iota}), \Omega^{\iota}), \underline{\mathsf{if}}(x, \underline{\mathsf{if}}(x, \underline{0}, \Omega^{\iota}), \Omega^{\iota}))),$$

we study $(M_q)\underline{0}^l$ (for a fixed label $l \in \mathcal{L}$). So in this example, "execution time" means "number of uses of the parameter $\underline{0}$". For all $v \in \mathsf{PN}$, we have $[\![M_q]\!](v) = \varphi_q(v_0)\, \mathsf{e}_0$ 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

$$\varphi_q(u) = \begin{cases} \frac{1-\sqrt{1-4q(1-q)u^2}}{2q} & \text{if } q > 0 \\ u^2 & \text{if } q = 0 \end{cases}$$

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 2.1 (for $q \in (0,1]$)

$$\mathbb{P}((M_q)\underline{0} \downarrow \underline{0}) = \varphi_q(1) = \frac{1-|2q-1|}{2q} = \begin{cases} 1 & \text{if } q \leq 1/2 \\ \frac{1-q}{q} & \text{if } q > 1/2 . \end{cases} \tag{2.11}$$

Observe that we have also $\mathbb{P}(M_0 \downarrow \underline{0}) = \varphi_0(1) = 1$ so that Equation (2.11) holds for all $q \in [0,1]$ (the corresponding curve is the second one in Fig. 3).

Then by Theorem 2.11 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

$$\varphi_q'(1)/\varphi_q(1) = \begin{cases} 2(1-q)/(1-2q) & \text{if } q < 1/2 \\ \infty & \text{if } q = 1/2 \\ 2(1-q)/(2q-1) & \text{if } q > 1/2 \end{cases}$$

(using the expression of $\varphi_q(1)$ given by Equation (2.11)), 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 typically twice. For $q = 1/2$ we have almost sure termination with an infinite expected computation time.

Of course such explicit computations are not always possible. For instance, using more occurrences of $(f)x$ we can modify the definition of M_q in such a way that computing $\varphi_q(u)$ would require solving a quintic. Or even we could set

$$M_q = \underline{\mathsf{fix}}(\lambda f^{\iota \Rightarrow \iota}\, \lambda x^{\iota}\, \underline{\mathsf{if}}(\underline{\mathsf{coin}}(q), \underline{\mathsf{if}}((f)x, \underline{\mathsf{if}}((f)(f)x, \underline{0}, \Omega^{\iota}), \Omega^{\iota}), \underline{\mathsf{if}}(x, \underline{\mathsf{if}}(x, \underline{0}, \Omega^{\iota}), \Omega^{\iota}))),$$

and then our solution function φ_q satisfies $\varphi_q(u) = (1-q)\, u^2 + q\, \varphi_q(u)\varphi_q(\varphi_q(u))$; in such a case we cannot expect to have an explicit expression for $\varphi_q(u)$. Approximating the value of $\varphi_q'(1)$ from below is possible by performing a finite number of iterations of the fixpoint and

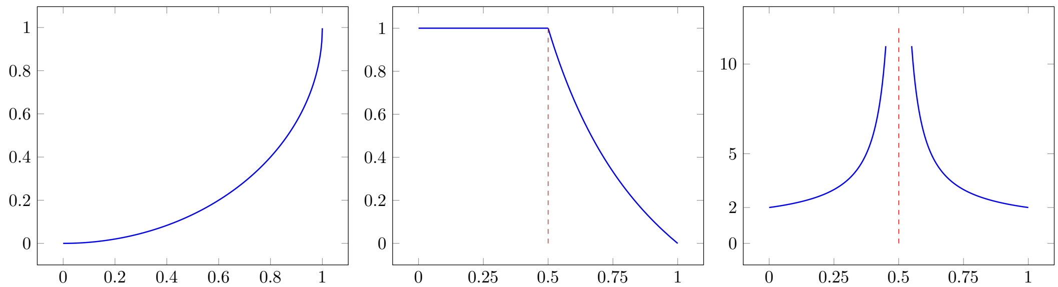

Figure 3: Plot of $\varphi_{0.5}(u)$ with u on the x-axis (vertical slope at $u = 1$). Plots of $\varphi_q(1)$ and $\mathbb{E}(\mathsf{Ev}_{\mathsf{lab}}(\langle(M_q)\underline{0}^l, \varepsilon\rangle)_l \mid \langle(M_q)\underline{0}, \varepsilon\rangle \downarrow \underline{0})$ with q on the x-axis. See Example 2.12.

approximating it from above is a more subtle problem. We could also expect to use more efficient approaches typically based on Newton's method.

Remark 2.13. (Connection with relational and coherence semantics.) It is possible to interpret terms of pPCF in **Rel**, the relational model of Linear Logic (see for instance [BE01]). In this model each type σ is interpreted as a set $[\![\sigma]\!]^{\mathbf{Rel}}$:

$$[\![\mathsf{N}]\!]^{\mathbf{Rel}} = \mathbb{N}$$
$$[\![\sigma \Rightarrow \tau]\!]^{\mathbf{Rel}} = \mathcal{M}_{\mathrm{fin}}([\![\sigma]\!]^{\mathbf{Rel}}) \times [\![\tau]\!]^{\mathbf{Rel}}$$

so that for each type σ we have $[\![\sigma]\!]^{\mathbf{Rel}} = |[\![\sigma]\!]|$. If $\Gamma \vdash M : \tau$ with $\Gamma = (x_1 : \sigma_1, \dots, x_k : \sigma_k)$ then

$$[\![M]\!]_\Gamma \in \mathbf{Rel}\left(\prod_{i=1}^{k} \mathcal{M}_{\mathrm{fin}}([\![\sigma_i]\!]^{\mathbf{Rel}}), [\![\tau]\!]^{\mathbf{Rel}}\right) = \mathcal{P}\left(\prod_{i=1}^{k} \mathcal{M}_{\mathrm{fin}}([\![\sigma_i]\!]^{\mathbf{Rel}}) \times [\![\tau]\!]^{\mathbf{Rel}}\right).$$

This semantics is "qualitative" in the sense that a point can only belong or not belong to the interpretation of a term whereas the **Pcoh** semantics is quantitative in the sense that the interpretation of the same term also provides a coefficient $\in \mathbb{R}_{\geq 0}$ for this point. We explain shortly the connection between the two models. To this end we describe first the relational model. One of the shortest ways to do so is by means of the "intersection typing system" given in Fig. 4 where we use the following conventions:

- $\Phi, \Phi_0 \cdots$ are *semantic contexts* of shape $\Phi = (x_1 : \mu_1 : \sigma_1, \dots, x_k : \mu_k : \sigma_k)$ where the x_i's are pairwise distinct variables and $\mu_i \in \mathcal{M}_{\mathrm{fin}}([\![\sigma_i]\!]^{\mathbf{Rel}})$;
- if $\Phi = (x_1 : \mu_1 : \sigma_1, \dots, x_k : \mu_k : \sigma_k)$ is such a semantic context then $\underline{\Phi} = (x_1 : \sigma_1, \dots, x_k : \sigma_k)$ is the underlying typing context;
- given a typing context $\Gamma = (x_1 : \sigma_1, \dots, x_k : \sigma_k)$, 0_Γ stands for the semantic context $0_\Gamma = (x_1 : [] : \sigma_1, \dots, x_k : [] : \sigma_k)$;
- given semantic contexts $\Phi_j = (x_1 : \mu_1^j : \sigma_1, \dots, x_k : \mu_k^j : \sigma_k)$ for $j = 1, \dots, n$ which have all the same underlying typing context $\Gamma = (x_1 : \sigma_1, \dots, x_k : \sigma_k)$, $\sum_{j=1}^n \Phi_j$ stands for the semantic context $(x_1 : \sum_{j=1}^n \mu_1^j : \sigma_1, \dots, x_k : \sum_{j=1}^n \mu_k^j : \sigma_k)$ whose underlying typing context is Γ (0_Γ can be seen as the case $n = 0$ of this construct with the slight problem that Γ cannot be derived from the Φ_j's in that case since there are none, whence the special construct 0_Γ).

Then, assuming that $(x_i : \sigma_i)_{i=1}^k \vdash M : \tau$, this typing system is such that, given $\vec{\mu} \in \prod_{i=1}^k \mathcal{M}_{\mathrm{fin}}(\llbracket \sigma_i \rrbracket^{\mathbf{Rel}})$ and $a \in \llbracket \tau \rrbracket^{\mathbf{Rel}}$, one has $(\vec{\mu}, a) \in \llbracket M \rrbracket_\Gamma^{\mathbf{Rel}}$ iff the judgment

$$(x_i : \mu_i : \sigma_i)_{i=1}^k \vdash M : a : \tau$$

is derivable. The interpretation of a term in **Rel** is simply the support of its interpretation in **Pcoh**:

$$(\vec{\mu}, a) \in \llbracket M \rrbracket_\Gamma^{\mathbf{Rel}} \Leftrightarrow (\llbracket M \rrbracket_\Gamma)_{\vec{\mu}, a} \neq 0$$

as soon as all occurrences of $\underline{\mathsf{coin}}(r)$ in M are such that $r \notin \{0, 1\}$ (occurrences of $\underline{\mathsf{coin}}(0)$ and $\underline{\mathsf{coin}}(1)$ can be replaced by $\underline{1}$ and $\underline{0}$ respectively without changing the semantics of M). This is easy to prove by a simple induction on M.

Since [BE01, Bou11] we know that Girard's coherence space semantics can be modified as follows: a *non-uniform coherence space* is a triple $X = (|X|, \frown_X, \smile_X)$ where $|X|$ is an at most countable set (the web of X) and $\frown_X$, $\smile_X$ are disjoint binary symmetric relations on $|X|$ called *strict coherence* and *strict incoherence* but contrarily to ordinary coherence spaces we can have $a \frown_X a$ or $a \smile_X a$ for some $a \in |X|$. These objects can be organized into a categorical model of classical linear logic **nCoh** whose associated Kleisli cartesian closed category is a model of PCF (that is, pPCF without the $\underline{\mathsf{coin}}(r)$ construct). Contrarily to what happens with Girard's coherence spaces[10], we have

$$|\llbracket \sigma \rrbracket^{\mathbf{nCoh}}| = \llbracket \sigma \rrbracket^{\mathbf{Rel}} = |\llbracket \sigma \rrbracket|\,.$$

Moreover given a PCF term M such that $\vdash M : \tau$, the set $\llbracket M \rrbracket^{\mathbf{nCoh}}$, which is a clique of the non-uniform coherence space $\llbracket \tau \rrbracket^{\mathbf{nCoh}}$ — meaning that $\forall a, a' \in \llbracket M \rrbracket\ \neg(a \smile_{\llbracket \tau \rrbracket^{\mathbf{nCoh}}} a')$ —, satisfies

$$\llbracket M \rrbracket^{\mathbf{nCoh}} = \llbracket M \rrbracket^{\mathbf{Rel}} = \{a \in |\llbracket \tau \rrbracket| \mid \llbracket M \rrbracket_a \neq 0\}\,.$$

In other words, the interpretation of a PCF term in **nCoh** is *exactly the same* as its interpretation in the basic model **Rel**. So what is the point of the model **nCoh**? It teaches us something we couldn't see in **Rel**: $\llbracket M \rrbracket^{\mathbf{Rel}}$ is a clique in the non-uniform coherence space associated with its type in **nCoh**.

In the model **nCoh**, the interpretation of the object of integers $\mathsf{N}^{\mathbf{nCoh}} = \llbracket \iota \rrbracket^{\mathbf{nCoh}}$ satisfies $|\mathsf{N}^{\mathbf{nCoh}}| = \mathbb{N}$, $n \smile n'$ if $n \neq n'$ and $\neg(n \frown n) \wedge \neg(n \smile n)$ for all $n \in \mathbb{N}$. In the model **nCoh** at least two exponentials are available; the free one is characterized in [Bou11]. With this exponential, $!\mathsf{N}^{\mathbf{nCoh}}$ has $\mathcal{M}_{\mathrm{fin}}(\mathbb{N})$ as web and

- $\mu \smile \mu'$ if $\exists n \in \mathsf{supp}(\mu)$, $n' \in \mathsf{supp}(\mu')$ $n \neq n'$
- $\mu \frown \mu'$ if $\mathsf{supp}(\mu) \cup \mathsf{supp}(\mu')$ has at most one element and $\mu \neq \mu'$.

Notice in particular that $[0, 1] \smile [0, 1]$ in $!\mathsf{N}^{\mathbf{nCoh}}$.

Let $M \in \mathsf{PCF}_{\mathsf{lab}}(L)$ (that is $M \in \mathsf{pPCF}_{\mathsf{lab}}(L)$ and M contains no instances of $\underline{\mathsf{coin}}(r)$) such that $\vdash M : \iota$ so that $\mathsf{sp}_{\vec{x}}(M) \in \mathsf{PCF}$ satisfies $(x_l : \iota)_{l \in L} \vdash \mathsf{sp}_{\vec{x}}(M) : \iota$ where $\vec{x} = (x_l : \iota)_{l \in L}$ is a list of pairwise distinct variables. Then $\llbracket \mathsf{sp}_{\vec{x}}(M) \rrbracket_{\vec{x}}^{\mathbf{nCoh}}$ is a clique of the non-uniform coherence space $X = !\mathsf{N}^{\mathbf{nCoh}} \otimes \cdots \otimes !\mathsf{N}^{\mathbf{nCoh}} \multimap \mathsf{N}^{\mathbf{nCoh}}$ (one occurrence of $!\mathsf{N}^{\mathbf{nCoh}}$ for each element of L). If $(\vec{\mu}, n) \in \llbracket \mathsf{sp}_{\vec{x}}(M) \rrbracket_{\vec{x}}^{\mathbf{nCoh}}$ then we know that each $\mu(l) \in \mathcal{M}_{\mathrm{fin}}(\mathbb{N})$ satisfies $\mathsf{supp}(\mu(l)) \subseteq \{0\}$ (see Lemma 2.9). And since the set $\llbracket \mathsf{sp}_{\vec{x}}(M) \rrbracket_{\vec{x}}^{\mathbf{nCoh}}$ is a clique in

[10]Indeed in Girard's coherence spaces, the web of $!X$ is the set of all finite cliques, or all finite multicliques of elements of $|X|$ (there are two versions of this exponential), hence this web *depends on the coherence relation* $\frown_X$. This is no more the case with non-uniform coherence spaces and this is the most important difference between the two models.

$$\frac{\mu_j = \begin{cases}[a] & \text{if } j = i\\ [\,] & \text{otherwise}\end{cases}}{(x_j : \mu_j : \sigma_j)_{j=1}^{k} \vdash x_i : a : A} \qquad \frac{}{0_\Gamma \vdash \underline{n} : n : \iota}$$

$$\frac{\Phi \vdash M : n : \iota}{\Phi \vdash \underline{\mathsf{succ}}(M) : n+1 : \iota} \qquad \frac{\Phi \vdash M : 0 : \iota}{\Phi \vdash \underline{\mathsf{pred}}(M) : 0 : \iota} \qquad \frac{\Phi \vdash M : n+1 : \iota}{\Phi \vdash \underline{\mathsf{pred}}(M) : n : \iota}$$

$$\frac{r \in (0,1] \cap \mathbb{Q}}{0_\Gamma \vdash \underline{\mathsf{coin}}(r) : 0 : \iota} \qquad \frac{r \in [0,1) \cap \mathbb{Q}}{0_\Gamma \vdash \underline{\mathsf{coin}}(r) : 1 : \iota}$$

$$\frac{\Phi \vdash M : n : \iota}{\Phi \vdash \underline{\mathsf{succ}}(M) : n+1 : \iota} \qquad \frac{\Phi_0 \vdash M : 0 : \iota \qquad \Phi_1 \vdash P : a : A \qquad \Gamma \vdash Q : A \text{ where } \Gamma = \underline{\Phi_0} = \underline{\Phi_1}}{\Phi_0 + \Phi_1 \vdash \underline{\mathsf{if}}(M, P, Q) : a : A}$$

$$\frac{\Phi_0 \vdash M : n : \iota \qquad \Phi_1, x : [n, \dots, n] : \iota \vdash N : a : \sigma \qquad \underline{\Phi_0} = \underline{\Phi_1}}{\Phi_0 + \Phi_1 \vdash \underline{\mathsf{let}}(x, M, N) : a : \sigma}$$

$$\frac{\Phi_0 \vdash M : n+1 : \iota \qquad \Phi_2 \vdash Q : a : A \qquad \Gamma \vdash P : A \text{ where } \Gamma = \underline{\Phi_0} = \underline{\Phi_2}}{\Phi_0 + \Phi_2 \vdash \underline{\mathsf{if}}(M, P, Q) : a : A}$$

$$\frac{\Phi, x : \mu : A \vdash M : b : B}{\Phi \vdash \lambda x^A\, M : (\mu, b) : A \Rightarrow B}$$

$$\frac{\Phi_0 \vdash M : ([a_1, \dots, a_n], b) : A \Rightarrow B \qquad \Phi_i \vdash P : a_i : A \text{ and } \underline{\Phi_i} = \underline{\Phi_0} \text{ for } i = 1, \dots, n}{\sum_{i=0}^{n} \Phi_i \vdash (M)P : b : B}$$

$$\frac{\Phi_0 \vdash M : ([a_1, \dots, a_n], a) : A \Rightarrow A \qquad \Phi_i \vdash \underline{\mathsf{fix}}(M) : a_i : A \text{ and } \underline{\Phi_i} = \underline{\Phi_0} \text{ for } i = 1, \dots, n}{\sum_{i=0}^{n} \Phi_i \vdash \underline{\mathsf{fix}}(M) : a : A}$$

Figure 4: Relational interpretation of pPCF as an intersection typing system

X and in view of the characterization above of the coherence relation of $!\mathsf{N}^{\mathbf{nCoh}}$, this set contains at most one element. When it is empty, this means that the execution of M does not terminate. When it is a singleton $\{(\vec{\mu}, n)\}$, the execution of M terminates with value $\underline{n}$, using $\mu(l)(0)$ times the l-labeled subterms of M. This can be understood as a version of the denotational characterization of execution time developed in [dC09, dC18], which is based on the model **Rel**.

From the viewpoint of the denotational interpretation in **Pcoh**, non-uniform coherence spaces tell us that, for non-probabilistic labeled closed terms $M \in \mathsf{PCF}_{\mathsf{lab}}(L)$ of type ι, the power series $[\![\mathsf{sp}_{\vec{x}}(M)]\!]$ has at most one monomial whose degree reflects the number of times its labeled subterms are used during its (deterministic) execution. Remember indeed that

$$[\![\mathsf{sp}_{\vec{x}}(M)]\!]_{\vec{x}}^{\mathbf{Rel}} = \{(\vec{\mu}, n) \mid ([\![\mathsf{sp}_{\vec{x}}(M)]\!]_{\vec{x}})_{\vec{\mu}, n} \neq 0\}\,.$$

When $M \in \mathsf{pPCF}_{\mathsf{lab}}(L)$, our Theorem 2.11 is a "smooth" version of this property.

3. Differentials and distances

In this second part of the paper, we also use differentiation in the category **Pcoh**, but contrarily to what we did when relating derivatives with execution time — we used only derivatives of first order functions — we will now consider also the "derivatives" (in that case one rather uses the word "differentials") of higher order functions. This is possible thanks to the fact that, even at higher order, our functions are analytic in some sense.

3.1. **Order theoretic characterization of PCSs.** The following simple lemma will be useful in the sequel. It is proven in [Gir04] in a rather sketchy way, we provide here a detailed proof for further reference. We say that a partially ordered set S is ω-complete if any increasing sequence of elements of S has a least upper bound in S.

Lemma 3.1. *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 [DE11]), we prove the converse, which uses the Hahn-Banach theorem in finite dimension. Notice first that by condition (3) we have $P^\perp, P^{\perp\perp} \subseteq (\mathbb{R}_{\geq 0})^I$.

Let $P \subseteq (\mathbb{R}_{\geq 0})^I$ satisfying the three conditions above and let us prove that $P^{\perp\perp} \subseteq P$, that is, given $y \in (\mathbb{R}_{\geq 0})^I \setminus P$, we must prove that $y \notin P^{\perp\perp}$, that is, we must exhibit a $x' \in P^\perp$ such that $\langle y, x' \rangle > 1$. We first show that we can assume that I is finite.

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 $(y|_{\{i_1,\dots,i_n\}})_{n\in\mathbb{N}}$ (where $i_1, i_2, \dots$ is any enumeration of I) and hence there must be some $n \in \mathbb{N}$ such that $y|_{\{i_1,\dots,i_n\}} \notin P$ by the assumption that P is ω-complete. Therefore it suffices to prove the result for I finite, which we assume. Let $Q = \{x \in \mathbb{R}^I \mid (|x_i|)_{i\in I} \in P\}$ which is a convex subset of $\mathbb{R}^I$ by the assumption that P is convex and downwards-closed.

Let $t_0 = \sup\{t \in \mathbb{R}_{\geq 0} \mid ty \in P\}$. By our assumption that P is ω-complete, we have $t_0 y \in P$ and hence $t_0 < 1$ since $y \notin P$. Let $h : \mathbb{R}y = \{ty \mid t \in \mathbb{R}\} \to \mathbb{R}$ 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[11], there exists a linear $l : \mathbb{R}^I \to \mathbb{R}$ which is upper-bounded by 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.** Given a cone P (see Section 1.2 for the definition) 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-module. 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.

[11] Here is one of the many versions of the Hahn-Banach Theorem: let E be a $\mathbb{R}$-vector space, F a subspace of E, $f : F \to \mathbb{R}$ a linear map, $p : E \to \mathbb{R}_{\geq 0}$ a seminorm such that $|f| \leq p$ on F. Then there is a $g : E \to \mathbb{R}$, linear and extending f and such that $|g| \leq p$ on E.

We specialize this construction to PCSs. 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 3.1 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 $\widehat{t}(x + u) \in \mathsf{P}Y$:

$$\widehat{t}(x+u)_b = \sum_{\mu \in |!X|} t_{\mu,b}(x+u)^\mu = \sum_{\mu \in |!X|} t_{\mu,b} \sum_{\nu \leq \mu} \binom{\mu}{\nu} x^{\mu-\nu} u^\nu \,.$$

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

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

Given $a \in |X_x|$ and $b \in |Y_{\widehat{t}(x)}|$, 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

$$t'(x) \in \mathbf{Pcoh}(X_x, Y_{\widehat{t}(x)}) \,.$$

By definition, this linear morphism $t'(x)$ is the *derivative (or differential, or Jacobian) of t at x*[12]. It is uniquely characterized by the fact that, for all $x \in \mathsf{P}X$ and $u \in \mathsf{P}X_x$, we have

$$\widehat{t}(x+u) = \widehat{t}(x) + t'(x) \cdot u + \widetilde{t}(x, u) \tag{3.1}$$

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

Example 3.2. Consider the case where $Y = {!X}$ and $t = \delta = \mathsf{Id}_{!X} \in \mathbf{Pcoh}_!(X, !X)$, so that $\widehat{\delta}(x) = x^!$. 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 codereliction [Ehr18]. Observe for instance that $\delta'(0)$ satisfies $\delta'(0)_{a,\nu} = \delta_{\nu,[a]}$ and therefore coincides with the ordinary definition of codereliction.

Proposition 3.3 (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 $(t \circ s)'(x) \cdot u = t'(\widehat{s}(x)) \cdot s'(x) \cdot u$.*

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

Proof. It suffices to write

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

by linearity of $t'(\widehat{s}(x))$ which proves our contention by the observation that, in the power series $t'(\widehat{s}(x))\cdot(\widetilde{s}(x,u))+\widetilde{t}(\widehat{s}(x),s'(x)\cdot u+\widetilde{s}(x,u))$, 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 [EPT18, Ehr20]. Given a PCS X, remember that $\|\_\|_X$ denotes the norm $\|\_\|_{\overline{\mathsf{P}}X}$ of the associated cone, see Section 1.2.

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 to prove 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 $\mathsf{d}_X(x,z)\leq\mathsf{d}_X(x,y)+\mathsf{d}_X(y,z)$.

Remark 3.4. In a cone P, glb's do not necessarily exist; we can nevertheless define a distance as follows:

$$\mathsf{d}_P(x,y)=\inf\{\|x-z\|_P+\|y-z\|_P\mid z\in P\text{ and }z\leq x,y\}$$

and it is possible to prove that, equipped with this distance, P is always Cauchy-complete. Of course if $P = \overline{\mathsf{P}}X$ this distance coincides with the distance defined above.

3.4. **A Lipschitz property.** Using the differential of Section 3.2, we prove that all morphisms of $\mathbf{Pcoh}_!$ satisfy a Lipschitz property, with a coefficient which cannot be upper bounded on the whole domain.

First of all, observe that, if $w \in \overline{\mathsf{P}}(X \multimap Y)$ and $x \in \overline{\mathsf{P}}X$, we have

$$\|w \cdot x\|_Y \leq \|w\|_{X \multimap Y} \, \|x\|_X \, .$$

Indeed if $\|w\|_{X \multimap Y} \neq 0$ and $\|x\|_X \neq 0$ we have $\frac{w}{\|w\|_{X \multimap Y}} \in \mathsf{P}(X \multimap Y)$ and $\frac{x}{\|x\|_X} \in \mathsf{P}X$, therefore $\frac{w}{\|w\|_{X \multimap Y}} \cdot \frac{x}{\|x\|_X} \in \mathsf{P}Y$ and our contention follows. And if $\|w\|_{X \multimap Y} = 0$ or $\|x\|_X = 0$ the inequation is obvious since then $w \cdot x = 0$.

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 $\|w \cdot (1-p)u\|_Y \leq 1$ 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 $t'(x) \in \mathsf{P}(X_x \multimap 1_{\widehat{t}(x)}) \subseteq \mathsf{P}(X_x \multimap 1)$. Therefore, if we assume that $\|x\|_X \leq p$, we have

$$(1-p)t'(x) \in \mathsf{P}(X \multimap 1) = \mathsf{P}X^{\perp} \, . \tag{3.2}$$

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

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

which is clearly analytic. 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

$$h'(\theta) = t'(x + \theta(y-x)) \cdot (y-x) \leq \frac{\|y-x\|_X}{1-p}$$

by (3.2), using Proposition 3.3. We have

$$0 \leq \widehat{t}(y) - \widehat{t}(x) = h(1) - h(0) = \int_0^1 h'(\theta)\, d\theta \leq \frac{\|y-x\|_X}{1-p} \, . \tag{3.3}$$

Let now $x, y \in \mathsf{P}X$ be such that $\|x\|_X, \|y\|_X \leq p$ (we don't assume any more that x and y are comparable). We have

$$\begin{aligned} \left|\widehat{t}(x) - \widehat{t}(y)\right| &= \left|\widehat{t}(x) - \widehat{t}(x \wedge y) + \widehat{t}(x \wedge y) - \widehat{t}(y)\right| \\ &\leq \left|\widehat{t}(x) - \widehat{t}(x \wedge y)\right| + \left|\widehat{t}(y) - \widehat{t}(x \wedge y)\right| \\ &\leq \frac{1}{1-p}(\|x - (x \wedge y)\|_X + \|y - (x \wedge y)\|_X) \\ &= \frac{\mathsf{d}_X(x,y)}{1-p} \end{aligned}$$

by (3.3) since $x \wedge y \leq x, y$. So we have proven the following result.

Theorem 3.5. *Let $t \in \mathsf{P}(!X \multimap 1)$. Given $p \in [0,1)$, the function $\widehat{t}$ 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*

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

Remark 3.6. The Lipschitz constant cannot be uniformly upper-bounded on $\mathsf{P}X$, in particular it cannot be upper-bounded by 1, that is t is not always contractive. A typical example is $t = \varphi_{0.5} \in \mathsf{P}(!1 \multimap 1)$ of Example 2.12: the left plot of Fig. 3 shows that the Lipschitz constant goes to ∞ when p goes to 1.

4. Application to the observational distance in pPCF

Given a pPCF 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 [EPT18], so that $\mathbb{P}(M \downarrow \underline{0}) = \mathbb{P}_{\langle M, \varepsilon\rangle}(\langle M, \varepsilon\rangle \downarrow \underline{0})$ with the notations of Section 2. Remember that $\mathbb{P}(M \downarrow \underline{0}) = [\![M]\!]_0$ by the Adequacy Theorem of [EPT18].

Given a type σ and two pPCF terms M, M' such that $\vdash M : \sigma$ and $\vdash M' : \sigma$, we define the *observational distance* $\mathsf{d_{obs}}(M, M')$ between M and M' as the sup of all the

$$\left|\mathbb{P}((C)M \downarrow \underline{0}) - \mathbb{P}((C)M' \downarrow \underline{0})\right|$$

taken over terms C such that $\vdash C : \sigma \Rightarrow \iota$ (testing contexts).

If $\varepsilon \in [0,1] \cap \mathbb{Q}$ we have $\mathsf{d_{obs}}(\underline{\mathsf{coin}}(0), \underline{\mathsf{coin}}(\varepsilon)) = 1$ as soon as $\varepsilon > 0$. It suffices indeed to consider the context

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

The semantics $[\![C]\!] \in \mathsf{P}(!\mathsf{N} \multimap \mathsf{N})$ is a function $c : \mathsf{PN} \to \mathsf{PN}$ such that

$$\forall u \in \mathsf{PN} \quad c(u) = u_0 c(u) + \Big(\sum_{i=1}^{\infty} u_i\Big)\mathsf{e}_0$$

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

$$c((1-\varepsilon)\mathsf{e}_0 + \varepsilon \mathsf{e}_1) = \begin{cases} 0 & \text{if } \varepsilon = 0 \\ 1 & \text{if } \varepsilon > 0\,. \end{cases}$$

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

$$C^{\langle p\rangle} = \lambda z^\sigma\, (C)\underline{\mathsf{if}}(\underline{\mathsf{coin}}(p), z, \Omega^\sigma)$$

and a tamed version of the observational distance is defined by

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

In other words, we modify our first definition of the observational distance by restricting the universal quantification on contexts to those which are of shape $C^{\langle p\rangle}$.

Theorem 4.1. *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*

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

Proof.

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

by the Adequacy Theorem and by Theorem 3.5. □

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 4.1 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 is that

$$\mathsf{d}_{\mathsf{obs}}^{\langle p\rangle}(M,M') \leq \frac{p}{1-p}\,\mathsf{d}_{\mathsf{lin}}(M,M')\,. \tag{4.1}$$

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

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

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 (4.1) 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 3.5 shows that another and more geometric approach, based on a simple denotational model, is also possible to get Theorem 4.1 which, though weaker than Inequation (4.1), 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 4.2. *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 $M_p = \underline{\mathsf{if}}(\underline{\mathsf{coin}}(p), M, \Omega^\sigma)$. 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 [EPT18], but $[\![M_p]\!] = p[\![M]\!]$ and similarly for M'. Since $p \neq 0$ we get $[\![M]\!] = [\![M']\!]$ and hence $M \sim M'$ by adequacy [EPT18]. $\square$

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 4.1 shows that we can control this distance using the denotational distance. For instance we have $\mathsf{d}_{\mathsf{obs}}^{\langle p \rangle}(\underline{\mathsf{coin}}(0), \underline{\mathsf{coin}}(\varepsilon)) \leq \frac{2p\varepsilon}{1-p}$ so that $\mathsf{d}_{\mathsf{obs}}^{\langle p \rangle}(\underline{\mathsf{coin}}(0), \underline{\mathsf{coin}}(\varepsilon))$ tends to 0 when ε tends to 0.

5. Conclusion

The two results of this paper are related: both use derivatives wrt. probabilities to evaluate the number of times arguments are used. The derivatives used in the second part are more general than those of the first part simply because the parameters wrt. which derivatives are taken can be of an arbitrary type whereas in the first part, they are of ground type, but this is essentially the only difference.

Indeed in Section 2 we computed partial derivatives of morphisms

$$t \in \mathbf{Pcoh}(!X, \mathsf{N})$$

where $X = \mathsf{N} \,\&\, \cdots \,\&\, \mathsf{N}$ (k copies). More precisely, the t's we consider in that section are such that if $t_{\vec{\mu},n} \neq 0$ then $n = 0$ and each $\mu_i \in \mathcal{M}_{\mathrm{fin}}(\mathbb{N})$ satisfies $\mathsf{supp}(\mu_i) \subseteq \{0\}$. So actually we can consider such a t as a an element of $\mathbf{Pcoh}(!(1 \,\&\, \cdots \,\&\, 1), 1)$ which induces a function

$$\begin{aligned} \widehat{t} : \mathsf{P}(1 \,\&\, \cdots \,\&\, 1) \simeq [0,1]^k &\to \mathsf{P}1 \simeq [0,1] \\ x &\mapsto \sum_{n_1,\dots,n_k \in \mathbb{N}} t_{n_1,\dots,n_k} \prod_{i=1}^{k} x_i^{n_i} \end{aligned}$$

where $t_{n_1,\dots,n_k} \in \mathbb{R}_{\geq 0}$ for each $n_1, \dots, n_k \in \mathbb{N}$. Let $Y = 1 \,\&\, \cdots \,\&\, 1$. Let $\vec{x} \in \mathsf{PN}^k$ and $x \in \mathsf{P}Y$ so that x can be seen as a tuple $(x_1, \dots, x_k) \in [0,1]$ and $\|x\|_Y = \max_{i=1}^{k} x_i$. It follows that Y_x can be described as follows:

$$|Y_x| = \{i \mid 1 \leq i \leq k \text{ and } x_i < 1\}$$
$$\mathsf{P}(Y_x) = \{u \in (\mathbb{R}_{\geq 0})^{|Y_x|} \mid \forall i \in |Y_x|\ x_i + u_i \leq 1\}\,.$$

and then we have defined the differential of t,

$$t'(x) \in \mathbf{Pcoh}(Y_x, 1)$$

in Section 3.2. This differential relates as follows with the partial derivatives used in Section 2:

$$\forall u \in \mathsf{P}Y_x \quad t'(x) \cdot u = \sum_{i \in |Y_x|} t'_i(x) u_i$$

where $t'_i(x) \in \mathbb{R}_{\geq 0}$ is the ith partial derivative of the function $\widehat{t}$ at x. Notice that in Section 2 we use slightly more general partial derivatives computed also at indices i such that $x_i = 1$ (which are actually left derivatives since the function can be undefined for $x_i > 1$) where they can take infinite values, in accordance with the fact that the expectation of the number of steps can be ∞ like in the example $\varphi_{0.5}$, see Fig. 3. This is not allowed in Section 3.2 where we insist on keeping all derivatives finite for upper-bounding them.

We think that these preliminary results provide motivations for investigating further differential extensions of pPCF and related languages in the spirit of the differential lambda-calculus [ER03].

Acknowledgments

We thank Raphaëlle Crubillé, Paul-André Melliès, Michele Pagani and Christine Tasson for many enlightening discussions on this work. We also thank the referees of the FSCD'19 version of this paper for their precious comments and suggestions. Last but not least we thank warmly the reviewers of this journal version for their in-depth reading and understanding of the paper and for their invaluable help in improving the presentation.

This research was partly funded by the ANR project ANR-19-CE48-0014 *Probabilistic Programming Semantics (PPS)*.

References

[BE01] Antonio Bucciarelli and Thomas Ehrhard. On phase semantics and denotational semantics: the exponentials. *Annals of Pure and Applied Logic*, 109(3):205–241, 2001.

[BLGS16] 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.

[Bou11] Pierre Boudes. Non-uniform (hyper/multi)coherence spaces. *Mathematical Structures in Computer Science*, 21(1):1–40, 2011.

[CCPW18] Simon Castellan, Pierre Clairambault, Hugo Paquet, and Glynn Winskel. The concurrent game semantics of probabilistic PCF. In Dawar and Grädel [DG18], pages 215–224.

[CL17] 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.

[Cru18] Raphaëlle Crubillé. Probabilistic Stable Functions on Discrete Cones are Power Series. In Dawar and Grädel [DG18], pages 275–284.

[dC09] Daniel de Carvalho. Execution time of lambda-terms via denotational semantics and intersection types. *CoRR*, abs/0905.4251, 2009.

[dC18] Daniel de Carvalho. Execution time of λ-terms via denotational semantics and intersection types. *MSCS*, 28(7):1169–1203, 2018.

[DE11] Vincent Danos and Thomas Ehrhard. Probabilistic coherence spaces as a model of higher-order probabilistic computation. *Information and Computation*, 152(1):111–137, 2011.

[DG18] 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*. ACM, 2018.

[DH00] 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.

[Ehr18] Thomas Ehrhard. An introduction to differential linear logic: proof-nets, models and antiderivatives. *Mathematical Structures in Computer Science*, 28(7):995–1060, 2018.

[Ehr20] Thomas Ehrhard. On the linear structure of cones. *CoRR*, abs/2001.04284, 2020. To appear as "Cones as a model of intuitionistic linear logic" in proceedings of LICS'2020.

[EPT18] Thomas Ehrhard, Michele Pagani, and Christine Tasson. Full Abstraction for Probabilistic PCF. *Journal of the ACM*, 65(4):23:1–23:44, 2018.

[ER03] Thomas Ehrhard and Laurent Regnier. The differential lambda-calculus. *Theoretical Computer Science*, 309(1-3):1–41, 2003.

[ET16] Thomas Ehrhard and Christine Tasson. Probabilistic call by push value. Technical report, 2016.

[Gir04] 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.

[KP17] Klaus Keimel and Gordon D. Plotkin. Mixed powerdomains for probability and nondeterminism. *Logical Methods in Computer Science*, 13(1), 2017.

[Sel04] 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.

[VKS19] Matthijs Vákár, Ohad Kammar, and Sam Staton. A domain theory for statistical probabilistic programming. *PACMPL*, 3(POPL):36:1–36:29, 2019.