Metric Reasoning About λ-Terms: The Affine Case (Long Version)

Raphaëlle Crubillé Ugo Dal Lago

October 22, 2018

Abstract

Terms of Church's λ-calculus can be considered *equivalent* along many different definitions, but context equivalence is certainly the most direct and universally accepted one. If the underlying calculus becomes probabilistic, however, equivalence is too discriminating: terms which have totally unrelated behaviours are treated the same as terms which behave very similarly. We study the problem of evaluating the *distance* between affine λ-terms. The most natural definition for it, namely a natural generalisation of context equivalence, is shown to be characterised by a notion of *trace* distance, and to be bounded from above by a coinductively defined distance based on the Kantorovich metric on distributions. A different, again fully-abstract, tuple-based notion of trace distance is shown to be able to handle nontrivial examples.

1 Introduction

Probabilistic models are formidable tools when abstracting the behaviour of complicated, intractable systems by simpler ones, at the price of introducing uncertainty. But there is more: randomness can be seen as *a way to compute*; in modern cryptography, as an example, having access to a source of uniform randomness is essential to achieve security in an asymmetric setting [14]. Other domains where probabilistic models play a key role include machine learning [24], robotics [27], and linguistics [21].

Probabilistic models of computation have been studied not only directly, but also through concrete or abstract programming languages, which are most often extensions of their deterministic siblings. Among the many ways probabilistic choice can be captured in programming, the simplest one consists in endowing the language of programs with an operator modelling the flipping of a fair coin. This renders program evaluation a probabilistic process, and under mild assumptions the language becomes universal for probabilistic computation. Particularly fruitful in this sense has been the line of work on the functional paradigm, both at a theoretical [17, 26, 23] and at a more practical level [15].

In presence of higher-order functions, program equivalence can be captured by so-called *context equivalence*: two programs M and N are considered equivalent if they behave the same no matter how the environment interacts with them: for every context $\mathcal{C}$, it holds that $\mathrm{Obs}(\mathcal{C}[M]) = \mathrm{Obs}(\mathcal{C}[N])$. However, this definition has the drawback of being based on an universal quantification over *all* contexts: showing that two programs are equivalent, requires considering their interaction with every possible context. The problem of giving handier characterisations of context equivalence can be approached in many different ways. As an example, coinductive methodologies for program equivalence have been studied thoroughly in deterministic [1, 25] and non-deterministic [19] computation, with new and exciting results appearing recently also for probabilistic languages: applicative bisimilarity, a coinductively defined notion of equivalence for functional programs, has been shown to be sound, and sometime even fully abstract, for probabilistic λ-calculi [6, 4].

In a probabilistic setting, however, equivalences are too strong if defined as above. Indeed, two programs are equivalent if their probabilistic behaviour is *exactly* the same (in every context). The actual value of probabilities in a probabilistic model often comes from statistical measurements, and should be considered more as an approximation to the actual probability law. Consequently, we would like to compare programs by appropriately reflecting small variations in them. Another scenario in which a richer, more informative way of comparing programs is needed is cryptography, where a central notion of equivalence, called *computational indistinguishability* [13] is indeed based on statistical distance rather than equality: the adversary *can* win the game, but with a *small* probability. Summing up, equivalences should be refined into metrics, and this is the path we will follow in this paper.

In probabilistic λ-calculi, the notion of observation $\mathrm{Obs}(\cdot)$ is quantitative: it is either the *probability* of convergence to a certain observable base value (e.g. the empty string), or the probability of convergence *tout court*. One can then easily define a notion of *context distance* as the maximal distance contexts can achieve when separating two terms:

$$\delta^{\mathrm{ctx}}(M, N) = \sup_{\mathcal{C}} \ |\mathrm{Obs}(\mathcal{C}[M]) - \mathrm{Obs}(\mathcal{C}[N])|.$$

This looks very close to computational indistinguishability, except for the absence of a security parameter: a scheme is secure if the advantage of any adversary in a given game (e.g., consisting in distinguishing between the case where the scheme is used, and the case where it is replaced by a truly random process) is "small" (e.g., negligible). Again, however, we find ourselves in front of a definition which risks to be useless in proofs, given that all contexts must be taken into account. But how difficult is evaluating the distance between concrete higher-order terms? Are there ways to alleviate the burden of dealing with all contexts, like for equivalences? These are the questions we address in this paper, and which have to the authors' knowledge not been investigated before.

As we will discuss in Section 2 below, finding handier characterisations of the context distance poses challenges which are simply different (and often harder) than the ones encountered in context equivalence. In particular, the context distance tends to *trivialise* and, perhaps worse, naively applying the natural generalisation of techniques known for equivalence is bound to lead to unsound methodologies. Indeed, one immediately realises that the number of times contexts access their argument is a crucial parameter, which must necessarily be dealt with. This is the reason why we work with an affine λ-calculus in this paper: this is a necessary first step, but also points to the right way to tame the general, non-linear case.

An extended version of this paper with more details is available [5].

Contributions

We introduce in this paper *three* distinct notions of distance for terms in an untyped, probabilistic, and affine λ-calculus. The first one is a notion of *trace distance*, in which terms are faced with *linear* tests, i.e. sequences of arguments. The distance between two terms is then defined as the greatest separation any linear test achieves. The first results of this paper are the non-expansiveness of the trace distance, which implies (given that any linear test can easily be implemented by an affine context) that the trace and context distances coincide. This is the topic of Section 4 below.

Section 5, instead, focuses on another notion of distance, which is coinductively defined following the well-known Kantorovich metric [18] for distributions of states in any labelled Markov chain (LMC in the following), and that we dub the *bisimulation distance*. This second notion of a distance is not only smaller than the trace distance, which is well expected, but non-expansive itself. This is proved by a variation on the Howe's method [16], a well-known technique for proving that bisimilarity is a congruence in an higher-order setting, and which has never been used for metrics before. On the other hand, the bisimulation distance *does not* coincide with the context distance, a fact that we do not only prove by giving a counterexample, but that we justify by relying on a test-based characterisation of the bisimulation distance known from the literature.

For the sake of simplicity, the trace and bisimulation distances are analysed on a purely applicative λ-calculus, keeping in mind that pairs could be very easily handled, and can even be encoded

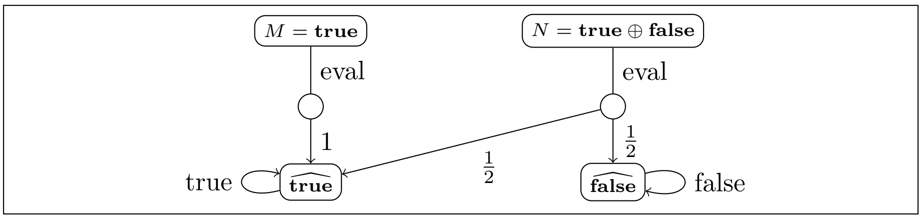

Figure 1: M and N as states of a LMC

in the applicative fragment, as discussed in Section 4.4. The presence of pairs, however, allows us to form very interesting examples of distance problems, one of which will drive us throughout the paper but unfortunately turns out hard to handle neither by the trace distance nor by the bisimulation distance. This is the starting point for the third notion of distance introduced in this paper, which is the subject of Section 6, and which we call the *tuple distance*. Our third notion of distance can be proved to coincide with the trace distance, and thus with the context distance. But this is not the end of the story: in the tuple distance, not a single but *many* terms are compared, and this makes the distance between concrete terms much easier to evaluate: interaction is somehow internalised. In particular, our running example can be handled quite easily. The way the tuple distance is defined makes it adaptable to non-affine calculi, a topic which is outside the scope of this paper, but which we briefly discuss in Section 6.3.

Related Work

This is definitely not the first work on metrics for probabilistic systems: notions of coinductively defined metrics for LMCs, as an example, have been extensively studied (e.g. [10, 9, 28]). There has been, to our knowledge, not so many investigations on the meaning of metrics for concrete programming languages [12], and almost nothing on metric for higher-order languages.

If the key property notions of *equivalences* are required to satisfy consists in being *congruences*, the corresponding property for metrics has traditionally been taken as *non-expansiveness*. Indeed, many results from the literature (e.g. [10, 22]) have precisely the form of non-expansiveness results for metrics defined in various forms. The underlying language, however, invariably take the form of a process algebra without any higher-order feature. The work of Gebler, Tini, and co-authors shows that one could go beyond non-expansiveness and towards uniform continuity [12] but, again, higher-order functions remain out of scope.

Notions of *equivalence* for various forms of probabilistic λ-calculi have also been extensively studied, starting from the pioneering work by Plotkin and Jones [17], down to recent results on probabilistic applicative bisimulation [6, 4], logical relations [3], and probabilistic coherent spaces [7, 11]. None of the works above, however, go beyond equivalences and deals with notions of distances between terms.

2 The Anatomy of a Distance

In this section, we describe the difficulties one encounters when trying to characterise the context distance with either bisimulation or trace metrics.

Suppose we have two terms M and N of boolean type written in a probabilistic λ-calculus. As such, M and N do not evaluate to *a* value in the domain of booleans but to a *distribution* over the same domain. M evaluates to the distribution assigning **true** probability 1, while N evaluates to the uniform distribution over booleans, (i.e. the distribution which attributes probability $\frac{1}{2}$ to both **true** and **false**). Figure 1 depicts the relevant fragment of a LMC, whose induced notion of probabilistic bisimilarity has been proved to be sound for context equivalence [4]. M and N are not bisimilar. Indeed, $\widehat{\mathbf{true}}$ and $\widehat{\mathbf{false}}$ are trivially not bisimilar, while M and N go to equivalent states with different probabilities. The two terms are non-equivalent also contextually. But what *should be* the distance between M and N?

For the moment, let us forget about the context distance, and concentrate on the notions of distance for LMCs we mentioned in Section 1. In all cases we are aware of, we obtain that M and N are at distance $\frac{1}{2}$. As an example, if we consider a trace metric, we have to compare the success probability of linear tests, starting from M and N. More precisely, the tests of interest with respect to these two terms are:

$$t := \text{eval}; \qquad s := \text{eval} \cdot \text{true}; \qquad r := \text{eval} \cdot \text{false}.$$

Since neither M nor N has a non-zero divergence probability, they both pass the test t with probability 1. The success probability of the test s corresponds to the probability of evaluating to **true**: it is 1 for M and $\frac{1}{2}$ for N. Similarly, the success probability of r corresponds to the probability to obtain **false** after evaluation: it is 0 for M and $\frac{1}{2}$ for N. So we can see that the maximal separation linear tests can obtain is $\frac{1}{2}$. The situation is quite similar for bisimulation metrics [10], which attribute distance $\frac{1}{2}$ to M and N.

It is easy, however, to find a family of contexts $\{\mathcal{C}_n\}_{n\in\mathbb{N}}$ such that $\mathcal{C}_n[M]$ evaluates to **true** with probability 1, and $\mathcal{C}_n[N]$ evaluates to **false** with probability $1 - \frac{1}{2^n}$: define $\mathcal{C}_n$ as a context that copies its argument n times, returning **false** if *at least one* of the n copies evaluates to **false**, and otherwise returns **true**. As a consequence, the context distance between M and N is 1. In fact, this reasoning can be extended to any pair of programs which are not equivalent but whose probability of convergence is 1: out of a context which separates them of $\varepsilon > 0$, with ε very small, we can construct a context that separates them of 1 performing some statistical reasoning. The situation is more complicated if we take the probability of convergence as an observable: we cannot *always* construct contexts that discriminate terms based on their probability of convergence, although something can be done if the terms' probabilities of convergence are different but close to 1. The context metric, in other words, risks to be not continuous and close to trivial if contexts are too powerful. What the example above shows, however, is something even worse: if contexts are allowed to copy their arguments, then any metric defined upon the usual presentation of probabilistic λ-calculus as an LMC (a fragment of which is depicted in Figure 1) is bound to be *unsound* w.r.t. the context metric.

Whether bisimulation metrics are sound, how close they are to the context distance, and whether they are useful in relieving the burden of evaluating it, are however open and interesting questions even in absence of the copying capability, i.e., when the underlying language is *affine*. This is the main reason why we focus in this work on such a λ-calculus, whose expressive power is limited (although definitely non-trivial [20]) but which is anyway higher-order. We discover this way an elegant and deep theory in which trace and bisimulation metrics are indeed sound, At the end of this paper, some hints will be given about how the case of the untyped λ-calculus can be handled, a problem which we leave for future work.

Evaluating the context distance between affine terms is already an interesting and nontrivial problem. Consider, as an example, a sequence of terms $\{M_n\}_{n\in\mathbb{N}}$ defined inductively as follows (where Ω stands for a term with zero probability of converging):

$$M_0 = \langle \lambda x.\Omega, \lambda x.\Omega \rangle; \qquad M_{n+1} = \langle \lambda x.M_n, \lambda x.\Omega \rangle.$$

M_0 is the pair whose components are both equal to $\lambda x.\Omega$, and M_{n+1} is defined as a pair whose first component is the function which returns M_n whatever its argument is, and the second component is again $\lambda x.\Omega$. We are now going to define another sequence of terms $\{N_n\}_{n\in\mathbb{N}}$, which can be seen as a noisy variation on $\{M_n\}_{n\in\mathbb{N}}$. More precisely, N_0 is the same as M_0, and for each $n \in \mathbb{N}$, N_{n+1} is constructed similarly to M_{n+1}, but adding some negligible noise in *both* components:

$$N_0 = \langle \lambda x.\Omega, \lambda x.\Omega \rangle;$$
$$N_{n+1} = \langle \lambda x.N_n \oplus^{\frac{1}{2^{n+1}}} \Omega, \lambda x.\Omega \oplus^{\frac{1}{2^{n+1}}} I \rangle.$$

(I stands for the identity: $\lambda x.x$, while the term $L \oplus^p K$ has the same behaviour as L with probability $(1-p)$, and the same behaviour as K with probability p.) We would like to study how the distance between M_n and N_n evolves when n tends to infinity: do the little differences we apply at each step n accumulate, and how can we express this accumulation quantitatively?

Intuitively, it is easy for the environment to separate M_n and N_n of $\frac{1}{2^n}$: it is enough to consider a context $\mathcal{C}$ which simply takes the second component of the pair, passes any argument to it, and evaluates it: the convergence probability of $\mathcal{C}[M_n]$ is 0, while the convergence probability of $\mathcal{C}[N_n]$ is $\frac{1}{2^n}$. But the environment can also decide to take the *first* component of the pair, in order to use the fact that M_{n-1} and N_{n-1} can be distinguished: more precisely, let us suppose that we have a context $\mathcal{C}$ which separates M_{n-1} and N_{n-1}. Then we can construct a context $\mathcal{D}$ which takes the first element of the pair, passes any argument to it, tries to evaluate it, and if it succeeds, gives the result as an argument to $\mathcal{C}$. We would like to express the supremum of the separation that such a context can obtain as a function of the distance between M_{n-1} and N_{n-1}. Unfortunately, this is not so simple: if $\mathcal{C}$ is such that the convergence probability of $\mathcal{C}[M_{n-1}]$ is ε and the convergence probability of $\mathcal{C}[N_{n-1}]$ is ι, we can see that the convergence probability of $\mathcal{D}[M_n]$ is ε, whereas the convergence probability of $\mathcal{D}[N_n]$ is $(\iota \cdot (1 - \frac{1}{2^n}))$. But it is not possible to express $|\varepsilon - \iota \cdot (1 - \frac{1}{2^n})|$ as a function of $|\varepsilon - \iota|$ and n: intuitively, the separation that the context $\mathcal{D}$ can achieve depends not only on the separation that the context $\mathcal{C}$ can achieve, but also on *how* $\mathcal{C}$ achieves it. And moreover, the environment may of course decide to use the two components of the pair, and to make them interact in an arbitrary way. Summing up, although the mechanism of construction of these terms seems to be locally easy to measure, it is complicated to have any idea about how the distance between them evolves when n tends to infinity.

3 Preliminaries

In this section, an affine and probabilistic λ-calculus, which is the object of study of this paper, will be introduced formally, together with a notion of context distance for it.

3.1 An Affine, Untyped, Probabilistic λ-Calculus

We endow the λ-calculus with a probabilistic operator $\oplus$, which corresponds to the possibility for the program to choose one between two arguments, each with the same probability. *Terms* are expressions generated by the following grammar:

$$M ::= x \;\Big|\; \lambda x.M \;\Big|\; MM \;\Big|\; M \oplus M \;\Big|\; \Omega,$$

where Ω models divergence[1], and x ranges over a countable set $\mathbf{V}$ of variables.

The class of affine terms, which model functions using their arguments at most once, can be isolated by way of a formal system, whose judgements are in the form $\Gamma \vdash M$ (where Γ is any finite set of variables) and whose rules are the following (where Γ, Δ stands for the union of two disjoints contexts):

$$\frac{}{\Gamma, x \vdash x} \qquad \frac{\Gamma, x \vdash M}{\Gamma \vdash \lambda x.M} \qquad \frac{\Gamma \vdash M \qquad \Delta \vdash N}{\Gamma, \Delta \vdash MN}$$

$$\frac{\Gamma \vdash M \qquad \Gamma \vdash N}{\Gamma \vdash M \oplus N} \qquad \frac{}{\Gamma \vdash \Omega}$$

A *program* is a term such that $\emptyset \vdash M$, and $\mathbf{P}$ is the set of all such terms. We will call them *closed terms*. We say that a program is a *value* if it is of the form $\lambda x.M$, and $\mathcal{V}$ is the set of such programs. The semantics of the just defined calculus is expressed as a binary relation $\Downarrow$ between programs and *value subdistributions* (or simply *value distributions*), i.e. functions from values to real numbers whose sum is *smaller* or equal to 1. The relation $\Downarrow$ is inductively defined by the following rules:

$$\frac{}{\Omega \Downarrow \emptyset} \qquad \frac{V \text{ a value}}{V \Downarrow \{V^1\}} \qquad \frac{M \Downarrow \mathscr{D} \qquad N \Downarrow \mathscr{E}}{M \oplus N \Downarrow \frac{1}{2}\mathscr{D} + \frac{1}{2}\mathscr{E}}$$

[1]since we only consider affine terms, we cannot encode divergence by the usual constructions of λ-calculus

$$\frac{M \Downarrow \mathscr{D} \qquad N \Downarrow \mathscr{E} \qquad \{L\{V/x\} \Downarrow \mathscr{F}_{L,V}\}_{\lambda x.L \in \mathsf{S}(\mathscr{D}), V \in \mathsf{S}(\mathscr{E})}}{MN \Downarrow \sum \mathscr{D}(\lambda x.L) \cdot \mathscr{E}(V) \cdot \mathscr{F}_{L,V}}$$

where $\mathsf{S}(\mathscr{D})$ stands for the support of the distribution $\mathscr{D}$. The divergent program Ω, as expected, evaluates to the *empty* value distribution $\emptyset$ which assigns 0 to any value. The expression $\{V^1\}$ stands for the Dirac's value distribution on V; more generally the expression $\{V_1^{p_1}, \ldots, V_n^{p_n}\}$ indicates the value distribution assigning probability p_i to each V_i (and 0 to any other value).

For every program M, there exists precisely *one* value distribution $\mathscr{D}$ such that $M \Downarrow \mathscr{D}$, that we note $[\![M]\!]$. This holds only because we restrict ourselves to affine terms. Moreover, $[\![M]\!]$ is always a finite distribution. The rule for application expresses the fact that the semantics is call-by-value: the argument is evaluated before being passed to the function. There is no special reason why we adopt call-by-value here, and all we are going to say also holds for (weak) call-by-name evaluation.

In some circumstances, we would need to have a more local view of how the programs behave. For these reason, we define an equivalent notion of small-steps semantics, which allows us to reason about every small execution step. We define first a one-step semantics $\to$ between programs and distribution over programs:

$$\frac{}{\Omega \to \emptyset} \qquad \frac{}{M \oplus N \to \frac{1}{2} \cdot \{M^1\} + \frac{1}{2} \cdot \{N^1\}}$$

$$\frac{}{(\lambda x.M)\, V \to \{M\{V/x\}^1\}}$$

$$\frac{M \to \mathscr{D}}{MN \to \sum \mathscr{D}(L) \cdot \{LN^1\}} \qquad \frac{N \to \mathscr{D}}{VN \to \sum \mathscr{D}(L) \cdot \{VL^1\}}$$

Then we use it to define a small step semantics $\Rightarrow$, which is a relation between programs and value distributions, and corresponds to do as much as possible steps of $\to$. The rules are the following:

$$\frac{}{V \Rightarrow \{V^1\}} \qquad \frac{M \to \mathscr{D} \qquad (N \Rightarrow \mathscr{E}_N)_{N \in \mathsf{S}(\mathscr{D})}}{M \Rightarrow \sum \mathscr{D}(N) \cdot \mathscr{E}_N}$$

Big-step and small-steps semantics are equivalent: for every program M, there exists a unique distribution $\mathscr{D}$ such that $M \Rightarrow \mathscr{D}$, and moreover $\mathscr{D} = [\![M]\!]$.

3.2 Context Distance

We now want to define a notion of observation for programs which somehow measures the convergence probability of a program. We will do that following the previous literature on this subject. For any distribution $\mathscr{D}$ over a set A, its sum $\sum_{a \in A} \mathscr{D}(a)$ is indicated as $\sum \mathscr{D}$ and is said to be the *weight* of $\mathscr{D}$. The *convergence probability* of a term M, that we note $\mathcal{P}^{\mathrm{cv}}(M)$, is simply $\sum [\![M]\!]$, i.e., the weight of its semantics. For instance, the convergence probability of Ω is zero.

The environment, as usual, is modelled by the notion of a *context*, which is nothing more than a term with a single occurrence of the *hole* $[\cdot]$. They are generated by the following grammar:

$$\mathcal{C} ::= [\cdot] \;\Big|\; M \;\Big|\; \lambda x.\mathcal{C} \;\Big|\; \mathcal{C}M \;\Big|\; M\mathcal{C} \;\Big|\; \mathcal{C} \oplus \mathcal{C}.$$

Affine contexts can be identified by a formal system akin to the one for terms. We note as $\mathcal{C}[M]$ the program obtained by replacing $[\cdot]$ by the closed term M in $\mathcal{C}$. The interaction of a program M with a context $\mathcal{C}$ is the execution of the program $\mathcal{C}[M]$.

We now consider three different ways of comparing programs, based on their behaviour when interacting with the environment: a preorder $\leq^{\mathrm{ctx}}$, an equivalence relation $\equiv^{\mathrm{ctx}}$, and a map δ^{ctx}:

Definition 1 (Context Equivalence, Context Distance) *Let M and N be two programs. Then we write that $M \leq^{ctx} N$ if and only if for every context $\mathcal{C}$, it holds that $\mathcal{P}^{cv}(\mathcal{C}[M]) \leq \mathcal{P}^{cv}(\mathcal{C}[N])$. If $M \leq^{ctx} N$ and $N \leq^{ctx} M$, then we say that the two terms are* context equivalent*, and we write $M \equiv^{ctx} N$. With the same hypotheses, we say the* context distance *between M and N is the real number $\delta^{ctx}(M,N)$ defined as $\sup_{\mathcal{C}} |\mathcal{P}^{cv}(\mathcal{C}[M]) - \mathcal{P}^{cv}(\mathcal{C}[N])|$.*

Please observe that, following [8], we only compare programs and not arbitrary terms. This is anyway harmless in an affine setting.

Example 1 *Let I be the identity $\lambda x.x$. I and Ω are as far as two programs can be: $\delta^{ctx}(I, \Omega) = 1$. To prove that, finding a context which always converges for one of the terms, and always diverges for the other one, suffices. We can take $\mathcal{C} = [\cdot]$, and we have that $\mathcal{P}^{cv}(\mathcal{C}[I]) = 1$ and $\mathcal{P}^{cv}(\mathcal{C}[\Omega]) = 0$. Of course, I and Ω are not context equivalent. Throwing in probabilistic choice can complicate matters a bit. Consider the two terms $I \oplus \Omega$ and I. One can easily prove that $\delta^{ctx}(I \oplus \Omega, I) \geq \frac{1}{2}$: just consider $\mathcal{C} = [\cdot]$. However, showing that the above inequality is in fact an equality, requires showing that there cannot exist any context that separates more, which is possible, but definitely harder. This will be shown in the next section, using a trace-based characterisation of context distance.*

3.3 On Pseudometrics

Which properties does the context distance satisfy, and which structure it then gives to the set of programs? This section answers these questions, and prepares the ground for the sequel by fixing some terminology.

Definition 2 (Pseudometrics) *Let S be a set. A* premetric *on S is any function $\mu : S \to S$ such that $0 \leq \mu(s,t) \leq 1$ and $\mu(s,s) = 0$. A* pseudometric *on S is any premetric such that for every $s,t,u \in S$, it holds that $\mu(s,t) = \mu(t,s)$ and $\mu(s,t) \leq \mu(s,u) + \mu(u,t)$. The set of all pseudometrics on S is indicated with $\Delta(S)$.*

Please observe that pseudometrics are not metrics in the usual sense, since $\mu(s,t) = 0$ does not necessarily imply that $s = t$. If we have a pseudometric μ, we can construct an equivalence relation by considering the *kernel* of μ, that is the set of those pairs (s,t) such that $\mu(s,t) = 0$. It is easy to prove that the context distance is indeed a pseudometric, and that its kernel is context equivalence. We would now want to define a preorder $\leq^{\mathrm{metr}}$ on pseudometrics in such a way that if $\mu \leq^{\mathrm{metr}} \rho$, then the kernel of μ is included in the kernel of ρ. The natural choice, then, is to take the following definition, which is the reverse of the pointwise order on $[0,1]$:

Definition 3 (Pseudometric Ordering) *Let S be any set, and let μ and ρ be two metrics in $\Delta(S)$. Then we stipulate that $\mu \leq^{metr} \rho$ if and only if, for every $s,t \in S$ we have that $\rho(s,t) \leq \mu(s,t)$.*

Lemma 1 *For any set S, $(\Delta(S), \cdot \leq^{metr} \cdot)$ is a complete lattice.*

But when, precisely, can a pseudometric on programs be considered as a sound notion of distance? First of all, we would like it to put two programs at least as far as the difference between their convergence probabilities, since this is precisely our notion of observation:

Definition 4 (Adequacy) *Let μ be a pseudometric on the set of programs. Then μ is an* adequate pseudometric *if for any programs M and N, we have that $|\sum [\![M]\!] - \sum [\![N]\!]| \leq \mu(M,N)$.*

Secondly, we are interested in how programs behave when interacting with the environment. Especially, if we have two terms M and N at a given distance ε, and we put them in an environment $\mathcal{C}$, we would like a pseudometric μ to give us some information about the distance between $\mathcal{C}[M]$ and $\mathcal{C}[N]$. This is the idea behind the following, standard, definition:

Definition 5 (Non-Expansiveness) *Let μ be a pseudometric on programs. We say μ is* non-expansive *if for every pair of programs M and N and for every context $\mathcal{C}$, we have that $\mu(\mathcal{C}[M], \mathcal{C}[N]) \leq \mu(M,N)$.*

Non-expansiveness is the natural generalisation of the usual notion of *congruence*: if R is an equivalence relation on program, it is congruent if for every context $\mathcal{C}$, if $M\,R\,N$, then $\mathcal{C}[M]\,R\,\mathcal{C}[N]$. By construction, δ^{ctx} is a non-expansive pseudometric. We can also adapt the notion of soundness to pseudometric; μ is said to be a *sound* pseudometric on programs if $\mu \leq^{\mathrm{metr}} \delta^{\mathrm{ctx}}$. Clearly, any adequate and non-expansive pseudometric is sound. In the rest of this paper, we will only deal with pseudometrics, but for the sake of simplicity we will refer to them simply as metrics.

4 The Trace Distance

The first notion of metric we study is based on traces, i.e., linear tests. This is handier than the context distance, since contexts are replaced by objects with a simpler structure.

4.1 Definition

A *trace* s is a sequence in the form $@V_1 \cdot \cdots \cdot @V_n$, where $V_1, \cdots V_n$ are values, and we note $\mathcal{T}r$ the set of traces. In other words, traces are generated by the following grammar:

$$s ::= \epsilon \mid @V \cdot s$$

We define the probability that a program accepts a trace inductively on the length of the trace, as follows:

$$\begin{aligned} Pr(\lambda x.M, \epsilon) &= 1; \\ Pr(\lambda x.M, @V \cdot s) &= Pr(M\{V/x\}, s); \\ Pr(M, s) &= \sum_V [\![M]\!](V) \cdot Pr(V, s) \qquad \text{if } M \notin \mathcal{V}. \end{aligned}$$

Please observe that the probability that a term M accepts a trace $s = @V_1 \cdots @V_n$ is the probability of convergence of $MV_1 \cdots V_n$. We are now going to define a metric, based on the probability that programs accept arbitrary traces:

Definition 6 *Let M, N be two programs. Then we define the* trace distance *between them as* $\delta^{tr}(M, N) = \sup_s |Pr(M, s) - Pr(N, s)|$. *One can then define* trace equivalence *and the* trace preorder*, in the expected way.*

Please observe that δ^{tr} is a pseudometric on programs in the sense of Definition 2, and that it is an adequate one. The kernel of δ^{tr} is nothing more than trace equivalence.

Example 2 $\delta^{tr}(I, \Omega) = 1$*: we have to find a trace that separates them as much. It is enough to consider the empty trace: it holds that* $Pr(\epsilon, I) = 1$*, and* $Pr(\epsilon, \Omega) = 0$*. The trace distance* $\delta^{tr}(I \oplus \Omega, I)$ *between* $I \oplus \Omega$ *and* I *is* $\frac{1}{2}$*. Showing that it is greater than* $\frac{1}{2}$ *is easy: it is sufficient to consider the empty trace. The other inequality, requires evaluating, for any trace* s*, the probability of accepting it. This is however much easier than dealing with all contexts, because we can now control the structure of the overall program we obtain: for any trace* $s = @V_1 \cdots @V_n$*, we can see that:* $Pr(I \oplus \Omega, s) = \frac{1}{2} \cdot \sum [\![V_1 \cdots V_n]\!]$*, and* $Pr(I, s) = \sum [\![V_1 \cdots V_n]\!]$*. The difference (in absolute value) between* $Pr(I \oplus \Omega, s)$ *and* $Pr(I, s)$*, then, cannot be greater than* $\frac{1}{2}$*.*

The trace distance and the context distance indeed *coincide*, as well as the trace and context preorder, and the trace and context equivalence. In the rest of this section, we will give the details of the proof for the pseudometric case, but the proof is similar for $\equiv^{\text{ctx}}$ and $\leq^{\text{ctx}}$. It is easy to realise that the context distance is a lower bound on the trace distance, since any trace $@V_1 \cdots @V_n$ can be seen as the context $[\cdot]V_1 \cdots V_n$.

Lemma 2 $\delta^{ctx} \leq \delta^{tr}$

Proof. For any trace $s = @V_1 \cdots @V_n$ which separates M and N of ε, we can easily construct a context which separates them of the same quantity: just take $\mathcal{C} = [\cdot]V_1 \cdots V_n$. □

4.2 Non-Expansiveness

Are there contexts that can separate strictly more than traces? In order to show that it is not the case, it is enough to show that δ^{tr} is non-expansive:

Theorem 1 *Let be M and N two programs, and let be $\mathcal{C}$ a context. Then* $\delta^{tr}(\mathcal{C}[M], \mathcal{C}[N]) \leq \delta^{tr}(M, N)$.

Since δ^{tr} is adequate, we can conclude that trace metric and context metric actually coincide:

Theorem 2 *$\delta^{ctx} = \delta^{tr}$.*

The rest of this section is devoted to an outline of the proof of Theorem 1. The proof we give here is roughly inspired by the proof of congruence of trace equivalence for a non-deterministic λ-calculus [8]. The overall structure of the proof is the following: we first express the capacity of a program to do a trace by means of a labelled transition system (LTS in the following) $\mathcal{L}^{\mathrm{tr}}$ whose states are distributions over programs. Then we consider *another* LTS $\mathcal{L}^{\mathrm{tr}}_{\mathbf{C}\times\mathbf{P}}$, where the states are distributions over pairs of contexts and programs that intuitively models the execution of $\mathcal{C}[M]$, but keeps the evolution of $\mathcal{C}$ and M apart.

4.2.1 The LTS $\mathcal{L}^{\mathrm{tr}}$

The first LTS, called $\mathcal{L}^{\mathrm{tr}}$, has distributions over programs as states, and traces as actions. We indicate with $\dot{\Rightarrow}$ the transition relation associated to $\mathcal{L}^{\mathrm{tr}}$. We're in fact going to define it on top of an auxilliary labelled relation $\dot{\rightarrow}$. Intuitively, the idea behind $\dot{\rightarrow}$ is to consider a term as a process who can make actions. There are two kinds of possible actions: an internal action τ, which corresponds to the internal reduction of the term, and external actions, which corresponds to the application of an argument V to the term. More precisely, this labelled relation $\dot{\rightarrow}$ is defined as a subset of the set $\mathrm{Distr}\left(\mathbf{P}\right) \times \mathcal{A} \times \mathrm{Distr}\left(\mathbf{P}\right)$, where the set $\mathcal{A}$ is defined as:

Definition 7 *We define the set of actions $\mathcal{A}$ by:*

$$\mathcal{A} = \{\tau\} \bigcup \{@V \mid V \ \mathit{value}\ \}.$$

Intuitively, a τ-step corresponds to an internal computation step for any term in the support of the distribution, while a $@V$-step corresponds to an interaction with the environment, which provides V as an argument.

Definition 8 *We define a labelled transition relation $\dot{\rightarrow} \subseteq \mathit{Distr}\left(\boldsymbol{P}\right) \times \mathcal{A} \times \mathit{Distr}\left(\boldsymbol{P}\right)$, by the rules of Figure 2. (We write $\mathscr{D} \dotplus \mathscr{E}$ for $\mathscr{D} + \mathscr{E}$ when we want to insist on $\mathscr{D}$ and $\mathscr{E}$ to have disjoint supports).*

$$\frac{M \to \mathscr{E}}{\mathscr{D} \dotplus \alpha \cdot \{M^1\} \xrightarrow{\tau} \mathscr{D} + \alpha \cdot \mathscr{E}} \qquad \frac{\mathscr{D} \text{ value distribution}}{\mathscr{D} \xrightarrow{@V} \sum_M \mathscr{D}(\lambda x.M) \cdot \{M\{V/x\}^1\}}$$

Figure 2: One-step Trace Relations on Program Distributions.

The relation $\dot{\Rightarrow}$ is defined as the accumulation of several steps of $\dot{\rightarrow}$. We define now the LTS $\mathcal{L}^{\mathrm{tr}}$.

Definition 9 *We define the LTS $\mathcal{L}^{tr}$ as:*

- *Its set of states is $\mathit{Distr}\left(\boldsymbol{P}\right)$*
- *Its set of labels is the set of traces $\mathcal{T}r$.*
- *Its transition relation $\dot{\Rightarrow}$ is defined by by the rules of Figure 3*

Please observe that these relations are *not* probabilistic. The relation $\xrightarrow{\tau}$ is non-deterministic, since at each step we can decide which term of the distribution we want to reduce. However, $\xrightarrow{\tau}$ is strongly normalising and confluent.

Lemma 3 *The relation: $\xrightarrow{\tau}$ is strongly normalising*

$$\frac{\mathscr{D} \text{ value distribution}}{\mathscr{D} \overset{\epsilon}{\Rightarrow} \mathscr{D}}$$

$$\frac{\mathscr{D} \overset{\tau}{\to} \mathscr{E} \qquad \mathscr{E} \overset{t}{\Rightarrow} \mathscr{F}}{\mathscr{D} \overset{t}{\Rightarrow} \mathscr{F}} \qquad \frac{\mathscr{D} \overset{@V}{\to} \mathscr{E} \qquad \mathscr{E} \overset{t}{\Rightarrow} \mathscr{F}}{\mathscr{D} \overset{@V\cdot t}{\Rightarrow} \mathscr{F}}$$

Figure 3: Small-step Trace Relations on Program Distributions.

Proof. • We show first that it is terminating: for a term M, we define a quantity $\mid M \mid \in \mathbb{N}$ which corresponds to the size of the term:

$$\mid \Omega \mid = 0; \qquad \mid x \mid = 1; \qquad \mid \lambda x.M \mid = 1 + \mid M \mid;$$

$$\mid MN \mid = \mid M \mid + \mid N \mid; \qquad \mid M \oplus N \mid = 1 + \max\{\mid M \mid, \mid N \mid\};$$

Since our λ-calculus is linear, $\mid M \mid$ decreases during the execution for every program M. More precisely: If $M \to \mathscr{D}$, then for every $N \in \mathsf{S}(\mathscr{E})$, $\mid N \mid < \mid M \mid$. (It is easily checked by observing the rules of $\to$).
Moreover, if $M \to \mathscr{D}$, then the cardinal of $\mathsf{S}(\mathscr{D})$ is at most 2. So, if for a distribution $\mathscr{D}$ we note:

$$\mid \mathscr{D} \mid = \sum_{(M) \in \mathsf{S}(\mathscr{D})} 3^{|M|}$$

, we can see that: for every $\mathscr{D}$, if $\mathscr{D} \overset{\tau}{\to} \mathscr{E}$, we have that $\mid \mathscr{E} \mid < \mid \mathscr{D} \mid$.

• Moreover, let $\mathscr{D}$ be a distribution over program, and let be $\mathscr{E}$ such that $\mathscr{D} \overset{\tau^n}{\to} \mathscr{E}$, and $\mathscr{E}$ is a normal form for $\overset{\tau}{\to}$. Then we are going to show by induction over $n \in \mathbb{N}$ that $\mathscr{E} = \sum_{M \in \mathsf{S}(\mathscr{D})} \mathscr{D}(M) \cdot [\![M]\!]$.

- if $n = 0$, then $\mathscr{D} = \mathscr{E}$, and moreover $\mathscr{D}$ is a distribution over values. So the result holds.
- if $n > 0$, it is a consequence of rules of Figure 2.

□

Moreover, we can show that $\overset{\cdot}{\Rightarrow}$ is in fact deterministic. That is, we have the following Lemma:

Lemma 4 *For every trace s, for every $\mathscr{D}$, there exist an only one $\mathscr{E}$ such that $\mathscr{D} \overset{s}{\Rightarrow} \mathscr{E}$.*

The interest of the relation $\overset{\cdot}{\Rightarrow}$ is that it gives an alternate formulation for the probability that a program succeeds in doing a trace:

Lemma 5 *Let be M a program, s a trace, and let be $\mathscr{E}$ the distribution such that $\{M^1\} \overset{s}{\Rightarrow} \mathscr{E}$. Then $Pr(s, M) = \sum \mathscr{E}$.*

In fact, the labelled transition system $\mathcal{L}^{\mathrm{tr}}$ allows us to extend the notion of probability of success for a trace to the case where we start not from a program, but from a probability distribution over program:

$$Pr(s, \mathscr{D}) = \sum \mathscr{E} \text{ when } \mathscr{D} \overset{s}{\Rightarrow} \mathscr{E}$$

In the same way we extend the preorder $\cdot \leq^{\mathrm{tr}} \cdot$, the equivalence relation $\cdot \equiv^{\mathrm{tr}} \cdot$, and the metric δ^{tr} to distributions. We can now use the relation $\overset{\cdot}{\Rightarrow}$ to give an equivalent formulation of Theorem 1: if M and N are such that $\delta^{\mathrm{tr}}(M, N) \leq \varepsilon$, then for every trace s, and context $\mathcal{C}$, if $\{\mathcal{C}[M]^1\} \overset{s}{\Rightarrow} \mathscr{D}$ and $\{\mathcal{C}[N]^1\} \overset{s}{\Rightarrow} \mathscr{E}$, then it holds that $|\sum \mathscr{D} - \sum \mathscr{E}| \leq \varepsilon$. This statement, however, cannot be proved directly, yet, because the way $\mathcal{C}$ and the argument terms interact is lost.

4.2.2 The LTS $\mathcal{L}^{\mathbf{tr}}_{\mathbf{C}\times\mathbf{P}}$

It is then time to introduce our second LTS, called $\mathcal{L}^{\mathrm{tr}}_{\mathbf{C}\times\mathbf{P}}$, which will allow us to relate $\{\mathcal{C}[M]^1\} \stackrel{\cdot}{\Rightarrow} \cdot$ to the behaviour of M: we want to talk about the evolution of a system consisting of the program M and the environment $\mathcal{C}$, while keeping the system and the environment as separate as possible. $\mathbf{C}\times\mathbf{P}$ is the set of pairs of the form $(\mathcal{C}, M)$, where $\mathcal{C}$ is a context and M is a program. The states of $\mathcal{L}^{\mathrm{tr}}_{\mathbf{C}\times\mathbf{P}}$ are distributions over $\mathbf{C}\times\mathbf{P}$, and the labels of $\mathcal{L}^{\mathrm{tr}}_{\mathbf{C}\times\mathbf{P}}$ are traces. The transition relation of $\mathcal{L}^{\mathrm{tr}}_{\mathbf{C}\times\mathbf{P}}$ corresponds to the transition relation of $\mathcal{L}^{\mathrm{tr}}$, where we keep the information about what part of the whole system is the program, and what part is the environment interacting with it. We'll use the following notation, which will be useful in the formal definition of $\mathcal{L}^{\mathrm{tr}}_{\mathbf{C}\times\mathbf{P}}$: If N is a term, $\mathscr{D}$ a distribution over $\mathbf{C}\times\mathbf{P}$, we define $\mathscr{D}\cdot N$ and $N\cdot\mathscr{D}$ as the distributions over $\mathbf{C}\times\mathbf{P}$ given by:

$$\mathscr{D}\cdot N = \sum_{(\mathcal{C},M)\in\mathsf{S}(\mathscr{D})} \mathscr{D}(\mathcal{C},M)\cdot\{(\mathcal{C}N,M)^1\}$$
$$N\cdot\mathscr{D} = \sum_{(\mathcal{C},M)\in\mathsf{S}(\mathscr{D})} \mathscr{D}(\mathcal{C},M)\cdot\{(N\mathcal{C},M)^1\}$$

And if $\mathcal{C}$ is a context, M a term, and $\mathscr{D}$ a distribution over $\mathbf{P}$, we define $(\mathcal{C}\mathscr{D}, M)$ and $(\mathscr{D}\mathcal{C}, M)$ as the distributions over $\mathbf{C}\times\mathbf{P}$ given by:

$$(\mathcal{C}\mathscr{D}, M) = \sum_{N\in\mathsf{S}(\mathscr{D})} \mathscr{D}(M\{(\mathcal{C}N,M)^1\})$$
$$(\mathscr{D}\mathcal{C}, M) = \sum_{N\in\mathsf{S}(\mathscr{D})} \mathscr{D}(M\{(N\mathcal{C},M)^1\})$$

If $(\mathcal{C}, M)\in\mathbf{C}\times\mathbf{P}$ is such that $\mathcal{C}[M]$ is a value, we say by abuse of notation that $(\mathcal{C}, M)$ is a value. If $\mathscr{D}$ is a distribution over $\mathbf{C}\times\mathbf{P}$ such that every $(\mathcal{C}, M)\in\mathsf{S}(\mathscr{D})$ is a value, we say that $\mathscr{D}$ is a value distribution.

Definition 10 *We define $\mathcal{L}^{tr}_{\mathbf{C}\times\boldsymbol{P}}$ as the labelled transition system such that:*

- *its set of states is the set of probability distributions over $\mathbf{C}\times\boldsymbol{P}$.*
- *its set of labels is the set of traces.*
- *its transition relations $\stackrel{\cdot}{\Rightarrow}_{\mathbf{C}\times\boldsymbol{P}}$ is defined by the rules of Figure 4. The definition uses an auxiliary one-step transition relation $\mathscr{D}\xrightarrow{a}_{\mathbf{C}\times\boldsymbol{P}}\mathscr{E}$, where $a\in\mathcal{A}$, and $\mathscr{D}$, $\mathscr{E}$ are distributions over $\mathbf{C}\times\boldsymbol{P}$.*

Lemma 6 *The relation $\xrightarrow{\tau}$ on distributions over $\mathbf{C}\times\boldsymbol{P}$ is strongly normalising, and normal forms of $\xrightarrow{\tau}$ are value distributions.*

Proof. The proof is exactly the same that for the relation $\xrightarrow{\tau}$ for distribution over programs. We extend the definition of $|\cdot|$ to distribution over $\mathbf{C}\times\mathbf{P}$, by: $|\,\mathscr{D}\,| = \sum_{(\mathcal{C},M)\in\mathsf{S}(\mathscr{D})} 3^{|\mathcal{C}[M]|}$, and we do the same reasoning. $\square$

For $\mathscr{D}$ a distribution over $\mathbf{C}\times\mathbf{P}$, we note $\mathscr{D}^\star$ the normal form of $\mathscr{D}$ for the relation $\xrightarrow{\tau}$. Please observe that Lemma 6 implies that for any distribution $\mathscr{D}$, there exists only one distribution $\mathscr{E}$ such that $\mathscr{D}\stackrel{\epsilon}{\Rightarrow}\mathscr{E}$, and moreover $\mathscr{E}=\mathscr{D}^\star$. The trace semantics for distributions over $\mathbf{C}\times\mathbf{P}$ allows us to extend the notions of trace equivalence, trace preorder and trace metric on distributions over $\mathbf{C}\times\mathbf{P}$ in a natural way.

$$\frac{N \to \mathscr{E}}{\mathscr{D} + p \cdot \{(N, M)^1\} \xrightarrow{\tau}_{\mathbf{C}\times\mathbf{P}} \mathscr{D} + p \cdot (\mathscr{E}, M)} \qquad \frac{}{p \cdot \{(\lambda x.N, M)^1\} \xrightarrow{@V}_{\mathbf{C}\times\mathbf{P}} p \cdot (N\{V/x\}, M)}$$

$$\frac{\{(\mathcal{C}, M)^1\} \xrightarrow{\tau}_{\mathbf{C}\times\mathbf{P}} \mathscr{E}}{\mathscr{D} + p \cdot \{(\mathcal{C}N, M)^1\} \xrightarrow{\tau}_{\mathbf{C}\times\mathbf{P}} \mathscr{D} + p \cdot (\mathscr{E} \cdot N)}$$

$$\frac{\mathcal{C}[M] \text{ is a value} \qquad N \to \mathscr{E}}{\mathscr{D} + p \cdot \{(\mathcal{C}N, M)^1\} \xrightarrow{\tau}_{\mathbf{C}\times\mathbf{P}} \mathscr{D} + p \cdot (\mathcal{C}\mathscr{E}, M)} \qquad \frac{N \to \mathscr{E}}{\mathscr{D} + p \cdot \{(N\mathcal{C}, M)^1\} \xrightarrow{\tau}_{\mathbf{C}\times\mathbf{P}} \mathscr{D} + p \cdot (\mathscr{E}\mathcal{C}, M)}$$

$$\frac{\{(\mathcal{C}, M)^1\} \xrightarrow{\tau}_{\mathbf{C}\times\mathbf{P}} \mathscr{E}}{\mathscr{D} + p \cdot \{(V\mathcal{C}, M)^1\} \xrightarrow{\tau}_{\mathbf{C}\times\mathbf{P}} \mathscr{D} + p \cdot (V\mathscr{E}, M)} \qquad \frac{M \to \mathscr{E}}{\mathscr{D} + p \cdot \{[\cdot], M^1\} \xrightarrow{\tau}_{\mathbf{C}\times\mathbf{P}} \mathscr{D} + p \cdot ([\cdot], \mathscr{E})}$$

$$\frac{}{\mathscr{D} + p \cdot \{[\cdot]V, \lambda x.N^1\} \xrightarrow{\tau}_{\mathbf{C}\times\mathbf{P}} \mathscr{D} + p \cdot \{[\cdot], N\{V/x\}^1\}} \qquad \frac{\mathcal{C}[M] \text{ value}}{\mathscr{D} + p \cdot \{(\lambda x.N)\mathcal{C}, M^1\} \xrightarrow{\tau}_{\mathbf{C}\times\mathbf{P}} \mathscr{D} + p \cdot \{N\{\mathcal{C}/x\}, M^1\}}$$

$$\frac{}{p \cdot \{\lambda x.\mathcal{C}, M^1\} \xrightarrow{@V}_{\mathbf{C}\times\mathbf{P}} p \cdot \{(\mathcal{C}\{V/x\}, M)^1\}} \qquad \frac{}{p \cdot \{([\cdot], \lambda x.M)^1\} \xrightarrow{@V}_{\mathbf{C}\times\mathbf{P}} p \cdot \{([\cdot], M\{V/x\})^1\}}$$

$$\frac{\mathscr{D}_i \xrightarrow{@V}_{\mathbf{C}\times\mathbf{P}} \mathscr{E}_i}{\dot{\sum}_i \mathscr{D}_i \xrightarrow{@V}_{\mathbf{C}\times\mathbf{P}} \sum_i \mathscr{E}_i}$$

$$\frac{\mathscr{D} \text{ value distribution}}{\mathscr{D} \overset{\epsilon}{\Rightarrow}_{\mathbf{C}\times\mathbf{P}} \mathscr{D}} \qquad \frac{\mathscr{D} \xrightarrow{\tau}_{\mathbf{C}\times\mathbf{P}} \mathscr{E} \qquad \mathscr{E} \overset{t}{\Rightarrow}_{\mathbf{C}\times\mathbf{P}} \mathscr{F}}{\mathscr{D} \overset{t}{\Rightarrow}_{\mathbf{C}\times\mathbf{P}} \mathscr{F}} \qquad \frac{\mathscr{D} \xrightarrow{@V}_{\mathbf{C}\times\mathbf{P}} \mathscr{E} \qquad \mathscr{E} \overset{t}{\Rightarrow}_{\mathbf{C}\times\mathbf{P}} \mathscr{F}}{\mathscr{D} \overset{V \cdot t}{\Rightarrow}_{\mathbf{C}\times\mathbf{P}} \mathscr{F}}$$

Figure 4: small-step trace relations on distributions over $\mathbf{C} \times \mathbf{P}$

4.2.3 Relating $\mathcal{L}^{\mathbf{tr}}$ and $\mathcal{L}^{\mathbf{tr}}_{\mathbf{C}\times\mathbf{P}}$

Intuitively, considering a semantics for distributions over $\mathbf{C} \times \mathbf{P}$ allows us to separate the part of the semantics which talks about the program, and the part which talks about the context. We would like to obtain the trace semantics for $\mathcal{C}[M]$, just by looking at the semantics of $(\mathcal{C}, M)$. We are going to express this idea by relating the two trace semantics.

Lemma 7 *Let be M a closed term, $\mathcal{C}$ a context and s a trace. Let be $\mathscr{D}$ and $\mathscr{E}$ such that $\{\mathcal{C}[M]^1\} \overset{s}{\Rightarrow} \mathscr{D}$, and $\{(\mathcal{C}, M)^1\} \overset{s}{\Rightarrow}_{\mathbf{C}\times\mathbf{P}} \mathscr{E}$ Then $\sum \mathscr{D} = \sum \mathscr{E}$.*

Proof. The proof of Lemma 7 is relatively technical, and is based on three auxilliary lemmas : Lemma 8, and Lemma 9. If $\mathscr{D}$ is a distribution over $\mathbf{C} \times \mathbf{P}$, we call $\mathbf{F}(\mathscr{D})$ the distribution obtained by filling each context by its associated term. To express this idea more formally, we define an operator $\mathbf{F}()$ on distributions over $\mathbf{C} \times \mathbf{P}$, which transforms every distribution in its corresponding distribution over terms.

$$\mathbf{F}(\mathscr{D}) = \sum_{\mathcal{C}, M} \mathscr{D}((\mathcal{C}, M)) \cdot \{\mathcal{C}[M]^1\}.$$

We can now express the correspondence between the trace semantics on distributions over programs, and the trace semantics on distributions over $\mathbf{C} \times \mathbf{P}$, by the following lemma.

Lemma 8 *Let $\mathscr{D}, \mathscr{E}$ be distributions over $\mathbf{C} \times \boldsymbol{P}$. If $\mathscr{D} \overset{s}{\Rightarrow}_{\mathbf{C}\times\boldsymbol{P}} \mathscr{E}$, then we have that: $\mathbf{F}(\mathscr{D}) \overset{s}{\Rightarrow} \mathbf{F}(\mathscr{E})$*

But we would like to have some information in the other directions too: if we have the trace semantics of the term $\mathcal{C}[M]$, is it possible to deduce something about the trace semantics of $(\mathcal{C}, M)$? The following lemma give a positive answer:

Lemma 9 *Let $\mathscr{D}$ be a distribution over $\mathbf{C} \times \boldsymbol{P}$ such that $\mathbf{F}(\mathscr{D}) \overset{t}{\Rightarrow} \mathscr{F}$. Let be $\mathscr{G}$ such that $\mathscr{D} \overset{t}{\Rightarrow}_{\mathbf{C}\times\boldsymbol{P}} \mathscr{G}$. Then $\mathscr{F} = \mathbf{F}(\mathscr{G})$.*

Proof. We need first to show an auxiliary lemma, in order to express the correspondence between the one-step relation on distributions over programs, and the one-step relation on distributions over $\mathbf{C} \times \mathbf{P}$.

Lemma 10 *Let be $\mathcal{C}$ a context, and N a term. Let be $\mathscr{D}$ such that $\mathcal{C}[N] \to \mathscr{D}$. Then there exists $\mathscr{E}$ such that: $\{(\mathcal{C}, N)^1\} \xrightarrow{\tau}_{\mathbf{C}\times\mathbf{P}} \mathscr{E}$, and $\mathbf{F}(\mathscr{E}) = \mathscr{D}$.*

Using this lemma we are now going to show Lemma 9. The proof is by induction on the derivation of $\mathbf{F}(\mathscr{D}) \overset{t}{\Rightarrow} \mathscr{F}$:

- The basic case is the case where $\mathbf{F}(\mathscr{D})$ is a value distribution (and consequently a normal form for $\cdot \xrightarrow{\tau} \cdot$), and where we are interested in the empty trace. The the derivation tree of $\mathbf{F}(\mathscr{D}) \overset{t}{\Rightarrow} \mathscr{F}$ is of the form:
$$\frac{\mathbf{F}(\mathscr{D}) \text{ value distribution}}{\mathbf{F}(\mathscr{D}) \overset{\epsilon}{\Rightarrow} \mathbf{F}(\mathscr{D})}$$
Then $\mathbf{F}(\mathscr{D})$ is a value distribution. By definition of values for distribution over $\mathbf{C} \times \mathbf{P}$, it means that $\mathscr{D}$ is a value distribution too. And so we can observe that $\mathscr{D} \overset{\epsilon}{\Rightarrow}_{\mathbf{C}\times\mathbf{P}} \mathscr{D}$, and the result holds.

- The first induction case is the case we don't start from a value distribution. Then the derivation tree of $\mathbf{F}(\mathscr{D}) \overset{t}{\Rightarrow} \mathscr{F}$ is of the form:
$$\frac{\mathbf{F}(\mathscr{D}) \xrightarrow{\tau} \mathscr{G} \qquad \mathscr{G} \overset{t}{\Rightarrow} \mathscr{F}}{\mathbf{F}(\mathscr{D}) \overset{t}{\Rightarrow} \mathscr{F}}$$
The only possible way to have obtained: $\mathbf{F}(\mathscr{D}) \xrightarrow{\tau} \mathscr{G}$ is to have used a derivation of the form:
$$\frac{M \to \mathscr{I}}{\mathbf{F}(\mathscr{D}) = \mathscr{H} \dotplus p \cdot \{(M)^1\} \xrightarrow{\tau} \mathscr{G} = \mathscr{H} + p \cdot (\mathscr{I})}$$
Since $\mathbf{F}(\mathscr{D}) = \mathscr{H} \dotplus p \cdot \{(M)^1\}$, $\mathscr{D} = \mathscr{J} \dotplus p \cdot \mathscr{K}$, with $\mathbf{F}(\mathscr{J}) = \mathscr{H}$ and $\mathbf{F}(\mathscr{K}) = \{M^1\}$. So for any $(\mathcal{C}, N) \in \mathsf{S}(\mathscr{K})$, we have that $\mathcal{C}[N] \to \mathscr{I}$. By Lemma 10, there exist $\mathscr{L}_{\mathcal{C},N}$ such that $\mathbf{F}(\mathscr{L}_{\mathcal{C},N}) = \mathscr{I}$, and $\{(\mathcal{C}, N)^1\} \xrightarrow{\tau}_{\mathbf{C}\times\mathbf{P}} \mathscr{L}_{\mathcal{C},N}$. And now we can see by the rules of trace semantics for distributions over $\mathbf{C} \times \mathbf{P}$ that:
$$\mathscr{D} \xrightarrow{\tau}_{\mathbf{C}\times\mathbf{P}} \cdots\cdots \xrightarrow{\tau}_{\mathbf{C}\times\mathbf{P}} \mathscr{J} + p \cdot \sum_{(\mathcal{C},N)\in\mathsf{S}(\mathscr{K})} \mathscr{L}_{(\mathcal{C},N)} \tag{1}$$
Moreover, we can see that $\mathbf{F}(\mathscr{J} + p \cdot \sum_{(\mathcal{C},N)\in\mathsf{S}(\mathscr{K})} \mathscr{L}_{(\mathcal{C},N)}) = \mathscr{H} + p \cdot (\mathscr{I}) = \mathscr{G}$. So now we can apply the induction hypothesis, and we have that there exists a distribution $\mathscr{E}$ such that:
$$\mathscr{J} + p \cdot \sum_{(\mathcal{C},N)\in\mathsf{S}(\mathscr{K})} \mathscr{L}_{(\mathcal{C},N)} \overset{s}{\Rightarrow}_{\mathbf{C}\times\mathbf{P}} \mathscr{E} \tag{2}$$
$$\mathbf{F}(\mathscr{E}) = \mathscr{F}. \tag{3}$$
And now we can conclude (by equations (1) and (2)) that $\mathscr{D} \overset{s}{\Rightarrow}_{\mathbf{C}\times\mathbf{P}} \mathscr{E}$, and so the results holds.

- The second induction case is the case where we start from a value distribution, and we are interested in a non-empty trace. Then the derivation tree of $\mathbf{F}(\mathscr{D}) \overset{t}{\Rightarrow} \mathscr{F}$ is of the form:
$$\frac{\mathbf{F}(\mathscr{D}) \xrightarrow{@V} \mathscr{G} \qquad \mathscr{G} \overset{s}{\Rightarrow} \mathscr{F}}{\mathbf{F}(\mathscr{D}) \overset{t=@V\cdot s}{\Rightarrow} \mathscr{F}}$$
The only possible way to have obtained: $\mathbf{F}(\mathscr{D}) \xrightarrow{@V} \mathscr{G}$ is to have used a derivation of the form:
$$\frac{\mathbf{F}(\mathscr{D}) \text{ value distribution}}{\mathbf{F}(\mathscr{D}) \xrightarrow{@V} \mathscr{G} = \sum \mathbf{F}(\mathscr{D})(\lambda x.M) \cdot \{M\{V/x\}^1\}}$$

For every $(\mathcal{C}, N) \in \mathsf{S}(\mathscr{D})$, let be $M_{(\mathcal{C},N)}$such that $\mathcal{C}[N] = \lambda x.M_{(\mathcal{C},N)}$. Using this notation, we can now express $\mathscr{G}$ as a sum over the support of the distribution $\mathscr{D}$:

$$\mathscr{G} = \sum_{(\mathcal{C},N)\in\mathsf{S}(\mathscr{D})} \mathscr{D}((\mathcal{C}, N)) \cdot \{(M_{(\mathcal{C},N)}\{V/x\})^1\} \tag{4}$$

We are going to define a distribution $\mathscr{H}_{(\mathcal{C},N)}$ over $\mathbf{C} \times \mathbf{P}$ for every $(\mathcal{C}, N)$ in the support of $\mathscr{D}$. We can see that for every $(\mathcal{C}, N) \in \mathsf{S}(\mathscr{D})$, we have two possible cases:

- Or $\mathcal{C} = [\cdot]$, and $N = \lambda x.(M_{(\mathcal{C},N)})$. Then let be $\mathscr{H}_{(\mathcal{C},N)} = \{([\cdot], M_{(\mathcal{C},N)}\{V/x\})^1\}$
- Or $\mathcal{C} = \lambda x.\mathcal{D}$, and $\mathcal{C}[N] = M_{(\mathcal{C},N)}$. Then let be $\mathscr{H}_{(\mathcal{C},N)} = \{(\mathcal{D}\{V/x\}, N)^1\}$. Please observe that, since the calculus is linear, $\mathcal{D}\{V/x\}$ is indeed a context.

Now we can write the equation (4) the following way:

$$\mathscr{G} = \mathbf{F}(\sum_{(\mathcal{C},N)\in\mathsf{S}(\mathscr{D})} \mathscr{D}((\mathcal{C}, N)) \cdot \mathscr{H}_{(\mathcal{C},N)}) \tag{5}$$

Moreover, for every $(\mathcal{C}, N) \in \mathsf{S}(\mathscr{D})$, we have: $\{\mathcal{C}, N^1\} \overset{@V}{\rightarrow}_{\mathbf{C}\times\mathbf{P}} \mathscr{H}_{(\mathcal{C},N)}$, and so the rules of one-step trace semantics for distribution over $\mathbf{C} \times \mathbf{P}$ allow us to say that:

$$\mathscr{D} \overset{@V}{\rightarrow}_{\mathbf{C}\times\mathbf{P}} \sum_{(\mathcal{C},N)\in\mathsf{S}(\mathscr{D})} \mathscr{D}((\mathcal{C}, N)) \cdot \mathscr{H}_{(\mathcal{C},N)} \tag{6}$$

By applying the induction hypothesis to $\mathscr{G} \overset{t}{\Rightarrow} \mathscr{F}$ and using equation (5), we know that there exists $\mathscr{I}$ such that:

$$\sum_{(\mathcal{C},N)\in\mathsf{S}(\mathscr{D})} \mathscr{D}((\mathcal{C}, N)) \cdot \mathscr{H}_{(\mathcal{C},N)} \overset{s}{\Rightarrow}_{\mathbf{C}\times\mathbf{P}} \mathscr{I} \tag{7}$$

$$\text{and } \mathbf{F}(\mathscr{I}) = \mathscr{F} \tag{8}$$

And now we can conclude by using the rules of trace semantics for distributions over $\mathbf{C} \times \mathbf{P}$ that $\mathscr{D} \overset{@V\cdot s}{\Rightarrow}_{\mathbf{C}\times\mathbf{P}} \mathscr{I}$, and since we have equation (8) the result holds.

□

□

4.3 ε-parents distributions

Lemma 7 allows us to give yet another equivalent formulation of Theorem 1: if $\delta^{\mathrm{tr}}(M, N) \leq \varepsilon$, then if $\{(\mathcal{C}, M)^1\} \overset{s}{\Rightarrow}_{\mathbf{C}\times\mathbf{P}} \mathscr{D}$ and $\{(\mathcal{C}, N)^1\} \overset{s}{\Rightarrow}_{\mathbf{C}\times\mathbf{P}} \mathscr{E}$, it holds that $|\sum \mathscr{D} - \sum \mathscr{E}| \leq \varepsilon$. We are in fact going to show a stronger result, which uses the notion of ε-related distributions:

Definition 11 *We say that two distributions $\mathscr{D}$ and $\mathscr{D}'$ over $\mathbf{C} \times \boldsymbol{P}$ are ε-related, and we note $\mathscr{D} \equiv_{\varepsilon}^{par} \mathscr{D}'$ if there exist $n \in \mathbb{N}$, and $\mathcal{C}_1, ..., \mathcal{C}_n$ distinct contexts, $p_1, ..., p_n$ positive real numbers with $\sum_i p_i \leq 1$, and $\mathscr{E}_1, ..., \mathscr{E}_n$, and $\mathscr{E}'_1, ..., \mathscr{E}'_n$ distributions over $\boldsymbol{P}$, such that:*

- $\mathscr{D} = \sum_{1\leq i\leq n} p_i \cdot (\mathcal{C}_i, \mathscr{E}_i)$
- $\mathscr{D}' = \sum_{1\leq i\leq n} p_i \cdot (\mathcal{C}_i, \mathscr{E}'_i)$
- $\forall i,\ \delta^{tr}(\mathscr{E}_i, \mathscr{E}'_i) \leq \varepsilon$

Please observe that, if $\delta^{\text{tr}}(M,N) \leq \varepsilon$, then for every context $\mathcal{C}$, the distributions $\{(\mathcal{C},M)^1\}$ and $\{(\mathcal{C},N)^1\}$ are ε-related. In fact, the notion of ε-relatedness is a way to capture the idea of a pair of distributions over $\mathbf{C} \times \mathbf{P}$ representing *the same* environment, in which we put programs which are close for the trace pseudometric. The following can be seen as a stability result: if we start from ε-related distributions, and we do a trace s, we end up in two distributions which are still ε-related.

Lemma 11 *Let be $\mathscr{D}$, $\mathscr{E}$ distributions over $\mathbf{C} \times \boldsymbol{P}$, and $\varepsilon \in [0,1]$ such that $\mathscr{D} \equiv_{\varepsilon}^{par} \mathscr{E}$. Let be s a trace. Let be $\mathscr{F}$ and $\mathscr{G}$ such that: $\mathscr{D} \stackrel{s}{\Rightarrow}_{\mathbf{C}\times\boldsymbol{P}} \mathscr{F}$, and $\mathscr{E} \stackrel{s}{\Rightarrow}_{\mathbf{C}\times\boldsymbol{P}} \mathscr{G}$. Then $\mathscr{F} \equiv_{\varepsilon}^{par} \mathscr{G}$*

Proof. If $\mathscr{D}$ is a distribution over $\mathbf{P}$, we note $\mathscr{D}^\star$ the distribution such that $\mathscr{D} \stackrel{\epsilon}{\Rightarrow} \mathscr{D}^\star$. Please observe that it is the normal form of $\mathscr{D}$ for the transition relation $\cdot \xrightarrow{\tau} \cdot$. Similarly, if $\mathscr{D}$ is a distribution over $\mathbf{C} \times \mathbf{P}$, $\mathscr{D}^\star$ the distribution such that $\mathscr{D} \stackrel{\epsilon}{\Rightarrow}_{\mathbf{C}\times\mathbf{P}} \mathscr{D}^\star$. We are first going to show two auxiliary lemma:

Lemma 12 *Let be $\mathscr{D}$, $\mathscr{E}$ two distributions over $\mathbf{C} \times \boldsymbol{P}$ such that $\mathscr{D} \equiv_{\iota}^{par} \mathscr{E}$. Then $\mathscr{D}^\star \equiv_{\iota}^{par} \mathscr{E}^\star$.*

Proof. We will use $Kr(a,b)$ as an integer being 1 if a is equal to b, and 0 otherwise. Let be $\mathscr{F}$ any distribution over $\mathbf{C} \times \mathbf{P}$. We note

$$n_{\max}(\mathscr{F}) = \max\{n \mid \mathscr{F} \xrightarrow{\tau^n}_{\mathbf{C}\times\mathbf{P}} \mathscr{F}^\star\}$$

We are going to show the lemma by induction on: $n = \max(n_{\max}(\mathscr{D}), n_{\max}(\mathscr{E}))$.

- If $n = 0$ then $\mathscr{D}^\star = \mathscr{D}$, and $\mathscr{E} = \mathscr{E}^\star$, and the result holds.
- If $n > 0$: Then we have: there exist $p_1,...,p_n$, and $\mathscr{E}_1,...,\mathscr{E}_n$, and $\mathscr{E}_1',...,\mathscr{E}_n'$ such that:

$$\mathscr{D} = \sum_{1\leq i\leq n} p_i \cdot (\mathcal{C}_i, \mathscr{D}_i)$$
$$\mathscr{E} = \sum_{1\leq i\leq n} p_i \cdot (\mathcal{C}_i, \mathscr{E}_i)$$
$$\forall i, \delta^{\text{tr}}(\mathscr{D}_i, \mathscr{E}_i) \leq \varepsilon$$

  Then there exists i such that: there exists $M \in \mathsf{S}(\mathscr{D}_i) \cup \mathsf{S}(\mathscr{E}_i)$, such that $\mathcal{C}_i[M]$ is not an irreducible term. We consider every possible case for the form of $\mathcal{C}_i[M]$:

  - or $\mathcal{C}_i$ is an evaluation context, and there exist $M \in \mathsf{S}(\mathscr{D}_i) \cup \mathsf{S}(\mathscr{E}_i)$, such that M is not an irreducible term. Intuitively, we want to reduce $\mathscr{D}_i$ and $\mathscr{E}_i$ as much as possible, since they are in evaluation position. And the two resulting distributions should be again ε-related distributions. More precisely, for every $M \in \mathsf{S}(\mathscr{D})_i \cup \mathsf{S}(\mathscr{E})_i$, we note $\mathscr{F}_M = \left(\{M^1\}\right)^\star$. Then the rules of $\xrightarrow{\tau}_{\mathbf{C}\times\mathbf{P}}$ allow us to see that there exist k_1 and k_2 such that $k_1 + k_2 > 0$, and:

$$\mathscr{D} \xrightarrow{\tau^{k_1}}_{\mathbf{C}\times\mathbf{P}} \mathscr{D}' = \sum_{j\neq i} p_j \cdot (\mathcal{C}_j, \mathscr{D}_j) + p_i(\mathcal{C}_i, \sum_M \mathscr{D}_i(M) \cdot \mathscr{F}_M),$$

    and

$$\mathscr{E} \xrightarrow{\tau^{k_2}}_{\mathbf{C}\times\mathbf{P}} \mathscr{E}' = \sum_{j\neq i} p_j \cdot (\mathcal{C}_j, \mathscr{E}_j) + p_i(\mathcal{C}_i, \sum_M \mathscr{E}_i(M) \cdot \mathscr{F}_M).$$

    We can easily show that $\mathscr{D}' \equiv_{\varepsilon}^{\text{par}} \mathscr{E}'$, and moreover, $\max(n_{\max}(\mathscr{D}'), n_{\max}(\mathscr{E}')) < \max(n_{\max}(\mathscr{D}), n_{\max}(\mathscr{E}))$. So we can apply the induction hypothesis, and we have that $\mathscr{D}^\star \equiv_{\varepsilon}^{\text{par}} \mathscr{E}^\star$.

– If $\mathcal{C}_i$ is such that the reduction depends only of the $\mathcal{C}_i$, that is if there exist $q_1, .., q_m$,..., such that for every term N,

$$\mathcal{C}_i[N] \to q_1 \cdot \mathcal{D}_1[N] + ... + q_m \cdot \mathcal{D}_m[N].$$

Then the rules of $\xrightarrow{\tau}_{\mathbf{C}\times\mathbf{P}}$ allows us to show that there exist k_1 and k_2, such that $k_1 + k_2 > 0$:

$$\mathscr{D} \xrightarrow{\tau^{k_1}}_{\mathbf{C}\times\mathbf{P}} \mathscr{D}' = \sum_{j\neq i} p_j \cdot (\mathcal{C}_j, \mathscr{D}_j) + p_i \cdot \sum_{1\leq k\leq m} q_k \cdot (\mathcal{D}_k, \mathscr{D}_i)$$

and that

$$\mathscr{E} \xrightarrow{\tau^{k_2}}_{\mathbf{C}\times\mathbf{P}} \mathscr{E}' = \sum_{j\neq i} p_j \cdot (\mathcal{C}_j, \mathscr{E}_j) + p_i \cdot \sum_{1\leq k\leq m} q_k \cdot (\mathcal{D}_k, \mathscr{E}_i).$$

In the definition of ε-related distribution, we consider contexts $(\mathcal{C}_j)_j$ disjoints. So since we want to show that the new distributions we have obtained are still ε-related, we have to regroup the identical contexts (for instance, it can be the case that: $\mathcal{D}_k = \mathcal{C}_j$): We note $\mathbf{C} = \{(\mathcal{C}_j)_{j\neq i} \cup (\mathcal{D}_k)_{1\leq k\leq m}\}$ the set of all contexts that can have been obtained at this step. For $\mathcal{C} \in \mathbf{C}$, we take $p'_\mathcal{C}$ his total probability: $p'_\mathcal{C} = \sum_{j\neq i} Kr(\mathcal{C}, \mathcal{C}_j) \cdot p_j + \sum_{1\leq k\leq m} Kr(\mathcal{C}, \mathcal{D}_k) \cdot p_i \cdot q_k$, and similarly:

$$\mathscr{D}'_\mathcal{C} = \sum_{j\neq i} Kr(\mathcal{C}, \mathcal{C}_j) \cdot \frac{p_j}{p'_\mathcal{C}} \cdot \mathscr{D}_j + \sum_{1\leq k\leq m} Kr(\mathcal{C}, \mathcal{D}_k) \cdot \frac{p_i \cdot q_k}{p'_\mathcal{C}} \cdot \mathscr{D}_i$$

, and

$$\mathscr{E}'_\mathcal{C} = \sum_{j\neq i} Kr(\mathcal{C}, \mathcal{C}_j) \cdot \frac{p_j}{p'_\mathcal{C}} \cdot \mathscr{E}_j + \sum_{1\leq k\leq m} Kr(\mathcal{C}, \mathcal{D}_k) \cdot \frac{p_i \cdot q_k}{p'_\mathcal{C}} \cdot \mathscr{E}_i.$$

And now we have

$$\mathscr{D}' = \sum_{\mathcal{C}\in\mathbf{C}} p'_\mathcal{C}(\mathcal{C}, \mathscr{D}'_\mathcal{C})$$

, and similarly:

$$\mathscr{E}' = \sum_{\mathcal{C}\in\mathbf{C}} p'_\mathcal{C}(\mathcal{C}, \mathscr{E}'_\mathcal{C}),$$

and for every $\mathcal{C} \in \mathbf{C}$, $\delta^{\mathrm{tr}}(\mathscr{D}'_\mathcal{C}, \mathscr{E}'_\mathcal{C}) \leq \varepsilon$.

– The last case is the case where the term and the context really interact: more precisely, $\mathscr{D}_i$ is a value distribution, and moreover:

  * Either $\mathcal{C}_i = \mathcal{D}[[\cdot]V]$, which means that we are in the case where the contexts pass values to the program. Then the following facts are derivable with the rules of $\xrightarrow{\tau}_{\mathbf{C}\times\mathbf{P}}$:

$$\mathscr{D} \xrightarrow{\tau^{k_1}}_{\mathbf{C}\times\mathbf{P}} \mathscr{D}' = \sum_{j\neq i} p_j \cdot (\mathcal{C}_j, \mathscr{D}_j) + p_i \cdot (\mathcal{D}, \sum_{\lambda x.N} \mathscr{D}_i(\lambda x.N) \cdot \{N\{V/x\}^1\})$$

  and

$$\mathscr{E} \xrightarrow{\tau^{k_2}}_{\mathbf{C}\times\mathbf{P}} \mathscr{E}' = \sum_{j\neq i} p_j \cdot (\mathcal{C}_j, \mathscr{E}_j) + p_i \cdot (\mathcal{D}, \sum_{\lambda x.N} \mathscr{E}_i(\lambda x.N) \cdot \{N\{V/x\}^1\})$$

We are now going to show that $\mathscr{D}'$ and $\mathscr{E}'$ are ε-related. We should again regroup the identical contexts (for instance, it can be the case that: $\mathcal{D} = \mathcal{C}_j$): We note $\mathbf{C} = \{(\mathcal{C}_j)_{j \neq i} \cup (\mathcal{D})\}$ the set of all contexts that can have been obtained at this step. For $\mathcal{C} \in \mathbf{C}$, we take $p'_{\mathcal{C}}$ his total probability defined as $p'_{\mathcal{C}} = \sum_{j \neq i} Kr(\mathcal{C}, \mathcal{C}_j) \cdot p_j + Kr(\mathcal{C}, \mathcal{D}) \cdot p_i$, and we obtain:

$$\begin{aligned} \mathscr{D}'_{\mathcal{C}} = & \sum_{j \neq i} Kr(\mathcal{C}, \mathcal{C}_j) \cdot \frac{p_j}{p'_{\mathcal{C}}} \cdot \mathscr{D}_j \\ & + Kr(\mathcal{C}, \mathcal{D}) \cdot \frac{p_i}{p'_{\mathcal{C}}} \cdot \sum_{\lambda x.N} \mathscr{D}_i(\lambda x.N) \cdot \{N\{V/x\}^1\} \end{aligned}$$

and

$$\begin{aligned} \mathscr{E}'_{\mathcal{C}} = & \sum_{j \neq i} Kr(\mathcal{C}, \mathcal{C}_j) \cdot \frac{p_j}{p'_{\mathcal{C}}} \cdot \mathscr{E}_j \\ & + Kr(\mathcal{C}, \mathcal{D}) \cdot \frac{p_i}{p'_{\mathcal{C}}} \cdot \sum_{\lambda x.N} \mathscr{E}_i(\lambda x.N) \cdot \{N\{V/x\}^1\} \end{aligned}$$

We have that: $\forall j \neq i$, $\delta^{\mathrm{tr}}(\mathscr{D}'_j = \mathscr{D}_j, \mathscr{E}'_j = \mathscr{E}_j) \leq \varepsilon$ by hypothesis, and moreover, for every trace s:

$$\begin{aligned} & |Pr(\left(\sum_{\lambda x.N} \mathscr{D}_i(\lambda x.N) \cdot \{N\{V/x\}^1\}\right), s) \\ & \quad - Pr(\left(\sum_{\lambda x.N} \mathscr{E}_i(\lambda x.N) \cdot \{N\{V/x\}^1\}\right), s)| \\ = & | \sum_{\lambda x.N} \mathscr{D}_i(\lambda x.N) \cdot Pr(N\{V/x\}, s) \\ & \quad - \sum_{\lambda x.N} \mathscr{E}_i(\lambda x.N) \cdot Pr(N\{V/x\}, s)| \\ = & \sum_{\lambda x.N} \mathscr{D}_i(\lambda x.N) \cdot Pr(\lambda x.N), @V \cdot s) \\ & \quad - \sum_{\lambda x.N} \mathscr{D}_i(\lambda x.N) \cdot Pr(\lambda x.N), @V \cdot s)| \\ = & |Pr(\mathscr{D}_i, @V \cdot s) \\ & \quad - Pr(\mathscr{E}_i, @V \cdot s)| \\ \leq & \varepsilon \end{aligned}$$

Since the relation $\delta^{\mathrm{tr}}(\cdot, \cdot) \leq \varepsilon$ on terms distribution is stable by convex summations, the result holds.

* Or $\mathcal{C}_i = \mathcal{D}[\lambda x.N[\cdot]]$ Then the rules of $\xrightarrow{\tau}_{\mathbf{C} \times \mathbf{P}}$ allows us to show that there exist k_1 and k_2 with $k_1 + k_2 > 0$, and such that:

$$\mathscr{D} \xrightarrow{\tau^{k_1}}_{\mathbf{C} \times \mathbf{P}} \mathscr{D}' = \sum_{j \neq i} p_j \cdot (\mathcal{C}_j, \mathscr{D}_j) + p_i \cdot (N\{[\cdot]/x\}, \mathscr{D}_i)$$

and:

$$\mathscr{E} \xrightarrow{\tau^{k_2}}_{\mathbf{C} \times \mathbf{P}} \mathscr{E}' = \sum_{j \neq i} p_j \cdot (\mathcal{C}_j, \mathscr{E}_j) + p_i \cdot (N\{[\cdot]/x\}, \mathscr{E}_i)$$

, and we can easily see that $\mathscr{D}' \equiv^{\mathrm{par}}_{\varepsilon} \mathscr{E}'$.

□

Lemma 13 *Let be $\varepsilon > 0$. If $\mathscr{D}$ and $\mathscr{E}$ are two value distributions over $\mathbf{C} \times \boldsymbol{P}$ (and consequently, in normal form for $\cdot \xrightarrow{\tau} \cdot$) with $\mathscr{D} \equiv_{\varepsilon}^{par} \mathscr{E}$, then for every V, there exists $\mathscr{F}$, $\mathscr{G}$ with $\mathscr{F} \equiv_{\varepsilon}^{par} \mathscr{G}$ such that $\mathscr{D} \xrightarrow{@V}_{\mathbf{C}\times\boldsymbol{P}} \mathscr{F}$, and $\mathscr{E} \xrightarrow{@V}_{\mathbf{C}\times\boldsymbol{P}} \mathscr{G}$.*

Proof. By hypothesis, we know that $\mathscr{D} \equiv_{\varepsilon}^{\mathrm{par}} \mathscr{E}$, and so we can write $\mathscr{D}$ and $\mathscr{E}$ as:

$$\mathscr{D} = \sum_i p_i \cdot (\mathcal{C}_i, \mathscr{D}_i) \qquad \text{and } \mathscr{E} = \sum_i p_i \cdot (\mathcal{C}_i, \mathscr{E}_i)$$
$$\text{and } \forall i, \delta^{\mathrm{tr}}(\mathscr{D}_i, \mathscr{E}_i) \leq \varepsilon$$

When $\mathscr{D}_i$ is a term distribution in normal form (i.e with value or non-reducible terms), we note

$$\mathscr{D}_i' = \sum_{\lambda x.M} \mathscr{D}(\lambda x.M) \cdot \{M\{V/x\}^1\}$$
$$\text{and } \mathscr{E}_i' = \sum_{\lambda x.M} \mathscr{E}(\lambda x.M) \cdot \{M\{V/x\}^1\}$$

And we have

$$\mathscr{D} \xrightarrow{@V}_{\mathbf{C}\times\mathbf{P}} \mathscr{F} = \sum_{i \mid \mathcal{C}_i = [\cdot]} p_i \cdot ([\cdot], \mathscr{D}_i') + \sum_{i \mid \mathcal{C}_i = \lambda x.\mathcal{D}_i} p_i \cdot (\mathcal{D}_i\{V/x\}, \mathscr{D}_i)$$

, and similarly:

$$\mathscr{E} \xrightarrow{@V}_{\mathbf{C}\times\mathbf{P}} \mathscr{G} = \sum_{i \mid \mathcal{C}_i = [\cdot]} p_i \cdot ([\cdot], \mathscr{E}_i') + \sum_{i \mid \mathcal{C}_i = \lambda x.\mathcal{D}_i} p_i \cdot (\mathcal{D}_i\{V/x\}, \mathscr{E}_i)$$

, and we can see that $\delta^{\mathrm{tr}}(\mathscr{F}, \mathscr{G}) \leq \varepsilon$. □

We can now use these two auxiliary lemma in order to prove Lemma 11. The proof is by induction on the length of s:

- : if $s = \epsilon$, then we have that $\mathscr{F} = \mathscr{D}^\star$, $\mathscr{G} = \mathscr{E}^\star$ and we have that $\mathscr{F} \equiv_{\varepsilon}^{\mathrm{par}} \mathscr{G}$ by Lemma 12.
- if $s = @V \cdot t$. Let be $\mathscr{H}$, $\mathscr{I}$ such that $\mathscr{D}^\star \xrightarrow{@V}_{\mathbf{C}\times\mathbf{P}} \mathscr{H}$ and $\mathscr{E}^\star \xrightarrow{@V}_{\mathbf{C}\times\mathbf{P}} \mathscr{H}$. We have (since $\xRightarrow{s}$ is confluent):

$$\mathscr{D} \xRightarrow{\epsilon}_{\mathbf{C}\times\mathbf{P}} \mathscr{D}^\star \xrightarrow{@V}_{\mathbf{C}\times\mathbf{P}} \mathscr{H} \xRightarrow{t} \mathscr{F}.$$

  and

$$\mathscr{E} \xRightarrow{\epsilon}_{\mathbf{C}\times\mathbf{P}} \mathscr{E}^\star \xrightarrow{@V}_{\mathbf{C}\times\mathbf{P}} \mathscr{I} \xRightarrow{t} \mathscr{G}.$$

  Then by Lemma 12 we have: $\mathscr{D}^\star \equiv_{\varepsilon}^{\mathrm{par}} \mathscr{E}^\star$. Now we can apply Lemma 13, and we obtain that $\mathscr{H} \equiv_{\varepsilon}^{\mathrm{par}} \mathscr{I}$. And now we apply the induction hypothesis to t, and we obtain that $\mathscr{F} \equiv_{\varepsilon}^{\mathrm{par}} \mathscr{G}$.

□

4.3.1 Proof of Theorem 1.

We can now see that Theorem 1 is a direct consequence of Lemma 11. Indeed, let M and N be two programs at distance at most ε for the trace metric, and let $\mathscr{D}$ and $\mathscr{E}$ be such that $\{\mathcal{C}, M^1\} \xRightarrow{s}_{\mathbf{C}\times\mathbf{P}} \mathscr{D}$, and $\{\mathcal{C}, N^1\} \xRightarrow{s}_{\mathbf{C}\times\mathbf{P}} \mathscr{E}$. Then, as we have already observed, $\{\mathcal{C}, M^1\}$ and $\{\mathcal{C}, N^1\}$ are ε-related. By Lemma 11, we can deduce that $\mathscr{D}$ and $\mathscr{E}$ are ε-related. And it is easy to see that it implies that $|\sum \mathscr{D} - \sum \mathscr{E}| \leq \varepsilon$.

4.4 Adding Pairs to the Calculus

The trace distance and the results we have just presented about it can be extended to an affine λ-calculus *with pairs*, namely a calculus whose language of terms also includes the following two constructs:

$$M ::= \langle M, N\rangle \;\Big|\; \texttt{let}\ \langle x, y\rangle = M\ \texttt{in}\ N.$$

We assume that terms are typed in any linear type system guaranteeing the absence of deadlocks (e.g., simple recursive types), and we add the following rules to the big-step semantics:

$$\frac{}{\langle M, N\rangle \Downarrow \{\langle M, N\rangle^1\}}$$

$$\frac{M \Downarrow \mathscr{D} \qquad (L \Downarrow \mathscr{F}_L,\, K \Downarrow \mathscr{G}_K)_{\langle L,K\rangle \in \mathsf{S}(\mathscr{D})} \qquad N\{V/x\}\{W/y\} \Downarrow \mathscr{E}_{V,W}}{\texttt{let}\ \langle x, y\rangle = M\ \texttt{in}\ N \Downarrow \sum \mathscr{D}(\langle L, K\rangle) \cdot \mathscr{F}_L(V) \cdot \mathscr{G}_K(W) \cdot \mathscr{E}_{V,W}}$$

We would now like to extend the definition of a trace to pairs accordingly: which action should we perform on a term in the form $\langle M, N\rangle$? The naïve solution would be to add projections to the trace language: $s ::= \pi_1 \cdot s \;\Big|\; \pi_2 \cdot s$, with trace interpretation extended in the expected way:

$$\begin{aligned} Pr(\langle M, N\rangle, \pi_1 \cdot t) &= Pr(M, t) \\ Pr(\langle M, N\rangle, \pi_2 \cdot t) &= Pr(N, t) \end{aligned}$$

However, this way the trace distance would not coincide with the context distance, anymore. Indeed, let us consider the following example:

Example 3 *We are going to compare the following terms:*

$$M \;:=\; \langle \lambda z.(I \oplus \Omega), \lambda z.(I \oplus \Omega)\rangle; \qquad N \;:=\; \langle \lambda z.I, \lambda z.I\rangle.$$

These two terms are at context distance at least $\frac{3}{4}$, since we can consider the context $\mathcal{C} := \texttt{let}\ \langle x, y\rangle = [\cdot]\ \texttt{in}\ (xI)(yI)$, and we can see that $\sum [\![\mathcal{C}[M]]\!] = \frac{1}{4}$, while $\sum [\![\mathcal{C}[N]]\!] = 1$. But we cannot find any trace that separates them more than $\frac{1}{2}$. The interesting case is when $s = \pi_i \cdot t$. But then:

$$\begin{aligned} |Pr(M, s) - Pr(N, s)| &= |Pr(\lambda z.(\Omega \oplus I), t) - Pr(\lambda z.I, t)| \\ &\leq \delta^{tr}(\lambda z.(\Omega \oplus I), \lambda z.I). \end{aligned}$$

And it is easy to see that in the calculus with pairs we still have $\delta^{tr}(\lambda z.(\Omega \oplus I), \lambda z.I) = \frac{1}{2}$.

The reason why we cannot recover the context distance by way of projections is that the `let` construct above allows us to access *both* components of a pair, and the distances each of them induce can *add up*. A way out consists in extending the trace language to pairs really following linearity, and considering a new action in the form $\otimes M$ with the following extension of trace interpretation:

$$Pr(\langle M, N\rangle, \otimes L \cdot t) = \sum_{V,W} [\![M]\!](V) \cdot [\![N]\!](W) \cdot Pr(L\{V, W/x, y\}, t)$$

Please observe that we could in fact express the pairs in the original language [2]: let us consider $\Theta : \Lambda_\oplus^{\langle,\rangle} \to$ Terms defined by

$$\begin{aligned} \Theta(\langle M, N\rangle) &:= \lambda x.x\Theta(M)\Theta(N) \\ \Theta(\texttt{let}\ \langle x, y\rangle = M\ \texttt{in}\ N) &:= \Theta(M)\,(\lambda x.(\lambda y.\Theta(N))) \\ \Theta(\lambda x.M) &:= \lambda x.\Theta(M) \cdots \end{aligned}$$

Moreover, we could see that every trace for the language $\Lambda_{\oplus}^{\langle,\rangle}$ can be seen as a trace in the original language: We can extend $\Theta : \mathcal{T}r(\Lambda_{\oplus}^{\langle,\rangle}) \to \mathcal{T}r$, by:

$$\Theta(\epsilon) = \epsilon$$
$$\Theta(@V \cdot s) = @V \cdot \Theta(s)$$
$$\Theta(\otimes M \cdot s) = \lambda x.\lambda y.M \cdot \Theta(s)$$

and we have for every term $M \in \Lambda_{\oplus}^{\langle,\rangle}$, and for every trace $s \in \mathcal{T}r(\Lambda_{\oplus}^{\langle,\rangle})$,

$$Pr(M, s) = Pr(\Theta(M), \Theta(s))$$

This way of handling pairs allows the trace distance and the context distance to coincide, again. However, the trace distance loses its grip with respect to the context distance. Consider, for instance, the terms M and N from Example 3. Showing an upper bound on the distance between M and N is the same thing as showing an upper bound on $\delta^{\mathrm{tr}}(L\{\lambda z.(\Omega \oplus I), \lambda z.(\Omega \oplus I)/x, y\}, L\{\lambda z.I, \lambda z.I/x, y\})$ *for all terms* L such that $x, y \vdash L$, which is in fact not far away from what we should show if we were considering the context distance directly.

5 The Bisimulation Distance

As we realised in the last section, the trace metric can be a way to alleviate the burden of evaluating the context distance between terms but, in particular in presence of pairs, its usefulness can be limited. In this section, we will look at another way to define the distance between programs which is genuinely coinductive, and based on the Kantorovich metric for distributions.

5.1 Definition

A labelled Markov chain (LMC) is a triple $\mathscr{M} = (S, \mathscr{L}, \mathscr{P})$, where S is a countable set of states, $\mathscr{L}$ is a countable set of labels, and $\mathscr{P}$ is a transition probability matrix, that is a function: $\mathscr{P} : S \times \mathscr{L} \to \mathrm{Distr}\,(S)$. Moreover, if the image of $\mathscr{P}$ only consists of distributions with *finite* support, we call $\mathscr{M}$ an *image-finite* LMC. We are now going to define, in a similar way to [10] (but in absence of non-determinism), the metric analog to bisimulation. The idea is to define a metric on the set S of states of the LMC as the greatest fixed point of some monotone operator on metrics. Please recall that $(\Delta(S), \leq^{\mathrm{metr}})$ is a complete lattice, and so any monotone operator has indeed a greatest fixed point.

Lifting Metrics to Distributions

We are going to define a way to turn any premetric over a set S into a metric over finite distribution over S.

Definition 12 *Let μ be a premetric on a set S. We define the* lifting of μ *as the metric on the set of finite distributions over S defined by: for every $\mathscr{D}$, $\mathscr{E}$ finite distributions over S, $\mu(\mathscr{D}, \mathscr{E})$ is the optimum solution to the following linear program:*

$$\begin{array}{ll} \min \sum_{i,j} h_{i,j} \cdot \mu(s_i, s_j) + \sum_i w_i + \sum_j z_j & \\ \textit{subject to} & \sum_i h_{i,j} + z_j = \mathscr{E}(s_j); \\ & \sum_j h_{i,j} + w_i = \mathscr{D}(s_i); \\ & \forall i, j, h_{i,j},\, z_j,\, w_i \geq 0. \end{array}$$

Please observe that this linear program has an optimal solution. We can make use of the notion of duality from linear programming, and obtain an alternative characterisation of lifting:

Theorem 3 *Let μ be a premetric on S and Let $\mathscr{D}$, $\mathscr{E}$ be finite distributions over S. Then:*

$$\begin{aligned} \mu(\mathscr{D},\mathscr{E}) = \max \sum_s a_s \mathscr{D}(s) + b_s \mathscr{E}(s) & \\ \textit{subject to} \quad & \forall s \in S,\, a_s \leq 1; \\ & \forall s \in S,\, b_s \leq 1; \\ & \forall s,\, t \in S,\, a_s + b_t \leq \mu(s,t). \end{aligned}$$

Proof. By strong duality theorem in linear programming. □

We would like to have the lifting of a metric μ behaving coherently with μ itself. If we know the lifting of μ, we should first of all be able to recover μ by considering Dirac distributions:

Lemma 14 *Let μ be a premetric on S, and let $s, t \in S$. Then $\mu(\{s^1\}, \{t^1\}) = \mu(s,t)$.*

Proof. Let be $s, t \in S$, and let be $\mathscr{D} = \{s^1\}$, $\mathscr{E} = \{t^1\}$. Then we can see that:

$$\begin{aligned} \mu(\mathscr{D},\mathscr{E}) &= \max \left\{ \begin{array}{l} \sum_u a_u \cdot \mathscr{D}(u) + b_u \cdot \mathscr{E}(u) \\ \text{with } \forall u \in S, a_u \leq 1 \wedge b_u \leq 1 \\ \text{and } \forall u_1, u_2 \in S,\, a_{u_1} + b_{u_2} \leq \mu(u_1, u_2) \end{array} \right\} \\ &= \max \left\{ \begin{array}{l} a_s + b_t \\ \text{with } a_s \leq 1 \wedge b_t \leq 1 \\ \text{and } a_s + b_t \leq \mu(s,t) \end{array} \right\} \\ &= \mu(s,t) \end{aligned}$$

□

If a premetric on states verifies the triangular inequality, its lifting verifies the triangular inequality too, which is a consequence of the following lemma:

Lemma 15 *Let μ, ρ, ν be three premetrics on S, such that $\forall s, t, u \in S$, $\mu(s,t) \leq \rho(s,u) + \nu(u,t)$. Let $\mathscr{D}$, $\mathscr{E}$, $\mathscr{F}$ be finite distributions over S. Then $\mu(\mathscr{D},\mathscr{F}) \leq \rho(\mathscr{D},\mathscr{E}) + \nu(\mathscr{E},\mathscr{F})$.*

Proof. Let be $\mathscr{D}, \mathscr{E}, \mathscr{F}$ finite distributions over S. We're going to use the minimum-based definition of lifting: Let be ε, ι such that: $\rho(\mathscr{D},\mathscr{E}) = \varepsilon$ and $\nu(\mathscr{E},\mathscr{F}) = \iota$. By assumption, there is a finite number of states which appear in the union of the support of every considered distributions. We numerate these states between 1 and n.
Then let be $(l_{i,j})_{1 \leq i,j \leq n}$, $(x_i)_{1 \leq i \leq n}$, $(y_j)_{1 \leq j \leq n}$ the coefficients for which the minimum of the optimisation problem associated with $\rho(\mathscr{D},\mathscr{E})$ is reached. They verify the following equations:

$$\begin{cases} \varepsilon = \sum_{i,j} l_{i,j} \cdot \rho(s_i, s_j) + \sum_i x_i + \sum_j y_j \\ \forall j,\, \sum_i l_{i,j} + y_j = \mathscr{E}(s_j) \\ \forall i,\, \sum_j l_{i,j} + x_i = \mathscr{D}(s_i) \\ \forall i, j \,:\, l_{i,j}, x_i, y_j \geq 0 \end{cases}$$

Similarly, let be $(h_{i,j})_{1 \leq i,j \leq n}$, $(w_i)_{1 \leq i \leq n}$, $(z_j)_{1 \leq j \leq n}$ the coefficients that reach the minimum for the optimisation problem associated to $\nu(\mathscr{E},\mathscr{F})$. They verify the following equations:

$$\begin{cases} \iota = \sum_{j,k} h_{j,k} \cdot \nu(s_j, s_k) + \sum_j w_j + \sum_k z_k \\ \forall k,\, \sum_j h_{j,k} + z_k = \mathscr{F}(s_k) \\ \forall j,\, \sum_k h_{j,k} + w_j = \mathscr{E}(s_j) \\ \forall j, k \,:\, h_{j,k}, w_j, z_k \geq 0 \end{cases}$$

We want to show that $\mu(\mathscr{D}, \mathscr{F}) \leq \varepsilon + \iota$. In order to do that, we would like to have coefficients $n_{i,k}$, a_i, b_k which verifies the constraints of the optimisation problem associated with $\mu(\mathscr{D}, \mathscr{F})$, and such that the objective function is bounded by $\varepsilon + \iota$. That is, we would like to have:

$$\begin{cases} \sum_i n_{i,k} + b_k = \mathscr{F}(s_k) & (9) \\ \sum_k n_{i,k} + a_i = \mathscr{D}(s_i) & (10) \\ \sum_{i,k} n_{i,k} \cdot \mu(s_i, s_k) + \sum_i a_i + \sum_k b_k \leq \varepsilon + \iota & (11) \end{cases}$$

In order to achieve that, we define the $n_{i,k}$, a_i, b_k on the following way:

$$n_{i,k} = \sum_j \frac{l_{i,j} \cdot hj, k}{\mathscr{E}(s_j)}$$

$$a_i = x_i + \sum_j \frac{l_{i,j} \cdot wj}{\mathscr{E}(s_j)}$$

$$b_k = z_k + \sum_j \frac{h_{j,k} \cdot zj}{\mathscr{E}(s_j)}$$

where we have adopted the following notation: if $\mathscr{E}(s_j) = 0$, then $h_{j,k} = 0 = l_{i,j}$, and then the meaning of $\frac{l_{i,j}}{\mathscr{E}(s_j)}$ is 0.
Now we are going to show that this choice of coefficients gives us what we wanted to have. We first verify that equation (10) holds. Indeed, we have that:

$$\begin{aligned} \sum_k n_{i,k} + a_i &= \sum_k \sum_j \frac{l_{i,j} \cdot hj, k}{\mathscr{E}(s_j)} + (x_i + \sum_j \frac{l_{i,j} \cdot w_j}{\mathscr{E}(s_j)}) \\ &= x_i + \sum_j l_{i,j} \frac{\sum_k h_{j,k} + w_j}{\mathscr{E}(s_j)} \\ &= x_i + \sum_j l_{i,j} \frac{\mathscr{E}(s_j)}{\mathscr{E}(s_j)} = \mathscr{D}(s_i) \end{aligned}$$

We can verify in a very similar way that equation (9) holds, that is: $\sum_i n_{i,k} + b_k = \mathscr{F}(s_k)$. We are now going to verify that equation (11) holds. Indeed, we have that:

$$\begin{aligned} &\sum_{i,k} n_{i,k} \mu(s_i, s_k) + \sum_i a_i + \sum_k b_k \\ &= \left[\begin{array}{l} \sum_{i,k} \sum_j \frac{l_{i,j} \cdot hj,k}{\mathscr{E}(s_j)} \cdot \mu(s_i, s_k) \\ + \sum_i (x_i + \sum_j \frac{l_{i,j} \cdot wj}{\mathscr{E}(s_j)}) \\ + \sum_k (z_k + \sum_j \frac{h_{j,k} \cdot zj}{\mathscr{E}(s_j)}) \end{array} \right. \\ &\leq \left[\begin{array}{l} \sum_{i,k} \sum_j \left(\frac{l_{i,j} \cdot hj,k}{\mathscr{E}(s_j)} \cdot \mu(s_i, s_k) \right) \\ + \sum_i (x_i + \sum_j \frac{l_{i,j} \cdot wj}{\mathscr{E}(s_j)}) \\ + \sum_k (z_k + \sum_j \frac{h_{j,k} \cdot zj}{\mathscr{E}(s_j)}) \end{array} \right. \\ &\leq \left[\begin{array}{l} \sum_{i,k} \sum_j \left(\frac{l_{i,j} \cdot hj,k}{\mathscr{E}(s_j)} \cdot (\rho(s_i, s_j) + \nu(s_i, s_k)) \right) \\ + \sum_i (x_i + \sum_j \frac{l_{i,j} \cdot wj}{\mathscr{E}(s_j)}) \\ + \sum_k (z_k + \sum_j \frac{h_{j,k} \cdot zj}{\mathscr{E}(s_j)}) \end{array} \right. \end{aligned}$$

$$\leq \left[\begin{array}{l} \sum_{i,j} l_{i,j} \cdot \rho(s_i, s_j) \left(\frac{\sum_k hj,k}{\mathscr{E}(s_j)} \right) \\ + \sum_{j,k} hj,k \cdot \nu(s_j, s_k) \left(\frac{\sum_i l_{i,j}}{\mathscr{E}(s_j)} \right) \\ + \sum_i x_i + \sum_j w_j \left(\frac{\sum_i l_{i,j}}{\mathscr{E}(s_j)} \right) \\ + \sum_k z_k + \sum_j zj \cdot \left(\frac{\sum_k h_{j,k}}{\mathscr{E}(s_j)} \right) \end{array} \right.$$

$$\leq \left[\begin{array}{l} \sum_{i,j} l_{i,j} \cdot \rho(s_i, s_j) \\ + \sum_{j,k} hj,k \cdot \nu(s_j, s_k) \\ + \sum_i x_i + \sum_j w_j + \sum_k z_k + \sum_j zj \end{array} \right.$$

$$\leq \varepsilon + \iota$$

□

Metrics as Fixpoints

In a non-probabilistic setting, a relation R is a bisimulation if every pair of states s, t such that $s\,R\,t$ can do the same actions and end up into states which are still bisimilar. More precisely,for every action a, and for every state u such that $s \overset{a}{\rightsquigarrow} u$, there exists v such that $t \overset{a}{\rightsquigarrow} v$, and $u\,R\,v$.

In order to obtain a quantitative counterpart of the scheme above, we define an operator F on the set of metrics over the states of a LMC: intuitively, given a metric μ, we define a new metric $F(\mu)$ which corresponds to the distance obtained by first doing a step of the transition relation, and then applying the lifting of μ to the resulting distributions. More precisely, let be two states s and t: $F(\mu)(s,t)$ is computed in the following way: for every action a, we consider the distance (with respect to μ) between the behaviour obtained from s after doing the action a, and the behaviour obtained from t after doing the same action a, and then we take the maximum over all action a of those quantity.

Definition 13 *Let $\mathscr{M} = (S, \mathscr{L}, \mathscr{P})$ be an image-finite LMC. We define an operator F on $\Delta(S)$ as*

$$F(\mu)(s,t) = \sup\{\mu(\mathscr{P}(s)(a), \mathscr{P}(t)(a)) \mid a \in \mathscr{L}\}.$$

Theorem 4 *For any image-finite LMC $\mathscr{M}$, F has a maximum fixpoint. We call it* the bisimulation metric *for the LMC $\mathscr{M}$, and we note it $\delta^b_{\mathscr{M}}$*

Bisimulation Metric and the Affine λ-Calculus

We are now going to consider a specific LMC $\mathscr{M}^\Lambda$, which captures the interactive behaviour of our calculus.

Definition 14 *We define the LMC $\mathscr{M}^\Lambda = (S^\Lambda, \mathscr{L}^\Lambda, \mathscr{P}^\Lambda)$ where:*

- *The set of states S^Λ is defined as follows:*

$$S^\Lambda = \boldsymbol{P} \uplus \mathcal{V},$$

  A value V in the second component of S^Λ is distinguished from one in the first by using the notation $\hat{V}$.
- *The set of labels $\mathscr{L}^\Lambda$ is taken to be*

$$\mathscr{L}^\Lambda = \{@V \mid V\, a \text{ value}\} \bigcup \{eval\}.$$

- *The transition probability matrix $\mathscr{P}^\Lambda$ is such that: for every $M \in \boldsymbol{P}$, and any value $V \in \mathsf{S}([\![M]\!])$, it holds that $\mathscr{P}^\Lambda(M, eval)(\widehat{V}) = [\![M]\!](V)$, and that for every term M such that $\lambda x.M \in \boldsymbol{P}$, and $V \in \mathcal{V}$, it holds that $\mathscr{P}^\Lambda(\widehat{\lambda x.M}, @V)(M\{x/V\}) = 1$.*

The results we have proved previously in this section apply to $\mathscr{M}^{\Lambda}$. In particular, one can define the bisimulation metric on $\mathscr{M}^{\Lambda}$. The *bisimulation distance* on programs, which we indicate δ^{b}, is defined to be the restriction of $\delta^{\mathrm{b}}_{\mathscr{M}^{\Lambda}}$ to programs.

Definition 15 *We define a metric δ^{b} on the set of closed terms, by: for every M, N,*

$$\delta^{b}(M,N) = \delta^{b}_{\mathscr{M}^{\Lambda}}(M,N)$$

Lemma 16 *For this particular LMC, we have that:*

$$\begin{aligned}
&F(\mu)(\widehat{\lambda x.M}, \widehat{\lambda x.N}) = \sup\{\mu(M\{V/x\}, N\{V/x\}) \mid V \textit{ a value } \}\\
&F(\mu)(M,N) = \mu([\![M]\!],[\![N]\!])\}\\
&F(\mu)(M,\widehat{V}) = 0
\end{aligned}$$

We can see easily that δ^{b} is an adequate metric.

Lemma 17 *δ^{b} is an adequate metric on programs.*

Proof. We have to show: for every M, N:

$$|\sum [\![M]\!] - \sum [\![N]\!]| \leq \delta^{\mathrm{b}}(M,N)$$

□

But there is more, since the bisimulation metric is well-known to be a lower bound on the trace distance: the bisimulation distance is a *sound* metric. In the next section, we anyway show non-expansiveness for it, which is stronger.

5.2 Non-Expansiveness

Proving the non-expansiveness of δ^{b} cannot be done directly, by a plain induction on contexts. Our strategy towards the result is the Howe's technique [16], a way of proving congruence of coinductively-defined equivalences which has been widely used for deterministic and non-deterministic languages, and that we here adapt to metrics.

The idea, then, is to start from δ^{b}, to construct another metric ${\delta^{\mathrm{b}}}^{H}$ on top of δ^{b} (which turns out to be non-expansive by construction), and to show that ${\delta^{\mathrm{b}}}^{H} = \delta^{\mathrm{b}}$. We first need to transform our metric δ^{b} on programs into a metric on (potentially open) terms. Any metric μ on programs can be extended into a metric on open terms, which by abuse of notation we continue to call μ and which is defined as follows

$$\begin{aligned}
\mu(M,N) = \sup_{V_1,\ldots,V_n} &\mu(M\{V_1,\ldots,V_n/x_1,\ldots,x_n\},\\
&N\{V_1,\ldots,V_n/x_1,\ldots,x_n\}),
\end{aligned}$$

where $x_1,\ldots,x_n$ are the variables occurring free in either M or N.

Definition 16 *Let be μ a metric on terms. An* Howe judgement *is a element of the form $(\Gamma,(M,N),\varepsilon)$, where Γ is a typing context, M and N are two terms, and $\varepsilon \in [0,1]$. We say that an Howe judgement is* valid*, and we note $\Gamma \vdash \mu^{H}(M,N) \leq \varepsilon$, if it can be derived by the rules of figure 5.*

Please observe that, potentially, there are several different ε such that $\Gamma \vdash \mu^{H}(M,N) \leq \varepsilon$.

We are finally in a position to define the Howe's lifting of μ:

Definition 17 *Let be μ a metric on terms. We define a premetric μ^{H} on terms by:*

$$\mu^{H}(M,N) = \inf\left(\{\varepsilon \mid \exists\Gamma, \Gamma \vdash \mu^{H}(M,N) \leq \varepsilon\} \bigcup \{1\}\right).$$

$$\frac{\mu(x,M)\leq\varepsilon \qquad x,\Gamma\vdash M}{x,\Gamma\vdash\mu^H(x,M)\leq\varepsilon}$$

$$\frac{\Gamma\vdash\mu^H(M,K)\leq\varepsilon \qquad \Delta\vdash\mu^H(N,T)\leq\gamma \qquad \mu(KT,L)\leq\iota \qquad \Gamma,\Delta\vdash L}{\Gamma,\Delta\vdash\mu^H(MN,L)\leq\varepsilon+\gamma+\iota}$$

$$\frac{\Gamma\vdash\mu^H(M,K)\leq\varepsilon \qquad \Gamma\vdash\mu^H(N,T)\leq\iota \qquad \mu(K\oplus T,L)\leq\gamma \qquad \Gamma\vdash L}{\Gamma\vdash\mu^H(M\oplus N,L)\leq\frac{\varepsilon+\iota}{2}+\gamma}$$

$$\frac{x,\Gamma\vdash\mu^H(M,K)\leq\varepsilon \qquad \mu(\lambda x.K,L)\leq\iota \qquad \Gamma\vdash L}{\Gamma\vdash\mu^H(\lambda x.M,L)\leq\varepsilon+\iota}$$

Figure 5: Rule for Howe's constructor on metrics

The following lemma says that the optimum value of ε can be reached with any typing context Γ which contains the free variables of M and N.

Lemma 18 *For every terms M, N, for every typing contexts Γ, and every real ε such that $\Gamma\vdash\mu^H(M,N)\leq\varepsilon$, we have that: $FV(M)\cup FV(N)\subseteq\Gamma$. Moreover, for any context Δ such that $\{\iota\mid\Delta\vdash\mu^H(M,N)\leq\iota\}\neq\emptyset$, then $\inf\{\iota\mid\Delta\vdash\mu^H(M,N)\leq\varepsilon\}\leq\varepsilon$.*

We can see that ${\delta^{\mathrm{b}}}^H$ is a premetric on open terms. Please observe that it is not necessarily a metric, since its construction entails neither symmetry nor the triangular inequality.

Lemma 19 *If μ is any premetric on closed terms, then μ^H is a premetric on (potentially open) terms.*

Lemma 20 *For every terms M, N:*

$$\mu^H(M,N)\leq\mu(M,N)$$

The construction of Howe's lifting allows us to have the two following properties:

Lemma 21 (Pseudo-Transitivity) *Let be μ a metric on terms. For every terms M, N, L:*

$$\mu^H(M,N)\leq\mu^H(M,L)+\mu(L,N)$$

Proof. Let be ε such that $\Gamma\vdash\mu^H(M,L)\leq\varepsilon$ is a valid judgement. It is enough to show that

$$\Gamma\vdash\mu^H(M,N)\leq\varepsilon+\mu(L,N)$$

is a valid judgement.The proof is by induction on the rules of the construction of valid judgements. □

Lemma 22 (Pseudo-substitutivity) *If μ verifies that, for every terms M,N, for every values V: $\mu(M\{V/x\},N\{V/x\})\leq\mu(M,N)$. Then for every terms M,N, for every values V, W:*

$$\mu^H(M\{V/x\},N\{W/x\})\leq\mu^H(M,N)+\mu^H(V,W)$$

Please observe that, the open extension of a metric on closed term verifies the hypothesis.

Proof. Let be ε such that: $\Gamma\vdash\mu^H(M,N)\leq\varepsilon$. The proof is by induction on the structure of the derivation of $\Gamma\vdash\mu^H(M,N)\leq\varepsilon$.

- If the derivation is: $\dfrac{\mu(x,N)\leq\varepsilon}{\Gamma,x\vdash\mu^H(x,N)\leq\varepsilon}$: Then $M\{V/x\}=V$. Then, since μ is pseudo substitutive: $\mu(W,N\{W/x\})\leq\mu(M,N)\leq\varepsilon$. Now by pseudo-transitivity of μ^H, we have that: $\mu^H(V,N\{W/x\})\leq\mu^H(V,W)+\mu(W,N\{W/x\})\leq\mu^H(V,W)+\varepsilon$.

- If the derivation is:

$$\frac{\Gamma \vdash \mu^H(T,K) \le \varepsilon \qquad \Delta \vdash \mu^H(U,P) \le \gamma \qquad \Gamma,\Delta \vdash L \qquad \mu(KP,L) \le \iota}{\Gamma,\Delta \vdash \mu^H(TU,L) \le \varepsilon + \iota + \gamma}$$

We know that x cannot appear both in Γ and in Δ. Suppose for example that x doesn't appear in Δ . Then (by Lemma 18) x doesn't appear in $FV(U) \cup FV(P)$. Then: We apply the induction hypothesis to: $\mu^H(T,K) \le \varepsilon$. We have:

$$\mu^H(T\{V/x\}, K\{W/x\}) \le \mu^H(T,K) + \mu^H(V,W) \le \varepsilon + \mu^H(V,W).$$

Moreover, since $\mu(KP,L) \le \iota$, we have that (since μ is value substitutive):

$$\mu((KP)\{W/x\}, L\{W/x\}) \le \iota$$

So now, we have that:

$$\frac{\begin{array}{c}\Gamma \setminus x \vdash \mu^H(T\{V/x\}, K\{W/x\}) \le \varepsilon + \mu^H(V,W) \\ \Delta \vdash \mu^H(U,P) \le \gamma\end{array} \qquad \mu(K\{W/x\}P, L\{W/x\}) \le \iota \qquad \Gamma \setminus x, \Delta \vdash L\{W/x\}}{\Gamma \setminus x, \Delta \vdash \mu^H((TU)\{V/x\}, L) \le \varepsilon + \iota + \gamma + \mu^H(V,W)}$$

- Other cases are similar.

□

The interest of this construction is that the metric ${\delta^{\mathrm{b}}}^H$ is (more or less by construction) non-expansive:

Lemma 23 (Non-expansiveness of ${\delta^{\mathrm{b}}}^H$) *For every context $\mathcal{C}$ and for every terms M, N it holds that ${\delta^b}^H(\mathcal{C}[M], \mathcal{C}[N]) \le {\delta^b}^H(M,N)$.*

Proof. The proof is by induction on the structure of the context $\mathcal{C}$. □

The goal now is to show that ${\delta^{\mathrm{b}}}^H \le^{\mathrm{metr}} \delta^{\mathrm{b}}$. Since δ^{b} is the greatest fixed point of F for our LMC $\mathscr{M}^\Lambda$, we are going to show that ${\delta^{\mathrm{b}}}^H$ can be extended into a metric on the states of $\mathscr{M}^\Lambda$, obtaining a fixed point for the operator F. First we extend ${\delta^{\mathrm{b}}}^H$ to a premetric on S^Λ:

Definition 18 *We define the extension of ${\delta^b}^H$ to S^Λ (that we note still ${\delta^b}^H$ by abuse of notation), by:*

$$\begin{aligned}
{\delta^b}^H(M,N) &= {\delta^b}^H(M,N);\\
{\delta^b}^H(\widehat{V},\widehat{W}) &= {\delta^b}^H(V,W);\\
{\delta^b}^H(M,\widehat{W}) &= 1.
\end{aligned}$$

Since ${\delta^{\mathrm{b}}}^H$ isn't guaranteed to be a metric, we are forced to further refine it, by adding rules corresponding to symmetry and to the triangular inequality: we define ${\delta^{\mathrm{b}}}^H_\triangle$ over S^Λ by the rules of Figure 6.

Definition 19 *We define a valid ${\delta^b}^H_\triangle$-judgement $\vdash {\delta^b}^H_\triangle(s,t) \le \varepsilon$, where $s,t \in S^\Lambda$, $\varepsilon \in [0,1]$, as the judgements which have a finite proof-tree by using the rules of Figure 6.*

We define ${\delta^b}^H_\triangle$ a metric over S^Λ by:

$${\delta^b}^H_\triangle(s,t) = \inf\{\varepsilon \mid \vdash {\delta^b}^H_\triangle(s,t) \le \varepsilon\}$$

Lemma 24 ${\delta^b}^H_\triangle$ *is a metric.*

$$\frac{\delta^{\mathrm{b}^H}(s,t) \leq \varepsilon}{\vdash \delta^{\mathrm{b}^H}_{\triangle}(s,t) \leq \varepsilon} \qquad \frac{\vdash \delta^{\mathrm{b}^H}_{\triangle}(s,t) \leq \varepsilon \qquad \vdash \delta^{\mathrm{b}^H}_{\triangle}(t,s) \leq \iota}{\vdash \delta^{\mathrm{b}^H}_{\triangle}(s,t) \leq \min(\varepsilon,\iota)}$$

$$\frac{\vdash \delta^{\mathrm{b}^H}_{\triangle}(s,t) \leq \varepsilon \qquad \vdash \delta^{\mathrm{b}^H}_{\triangle}(t,u) \leq \iota}{\vdash \delta^{\mathrm{b}^H}_{\triangle}(s,u) \leq \varepsilon + \iota}$$

Figure 6: Metric Closure of δ^{b^H}

We can see easily that $\delta^{\mathrm{b}} \leq^{\mathrm{metr}} \delta^{\mathrm{b}^H} \leq^{\mathrm{metr}} \delta^{\mathrm{b}^H}_{\triangle}$ with respect to the preorder on terms. We want to show that $\delta^{\mathrm{b}^H} = \delta^{\mathrm{b}}$. In order to have that, we will show that $\delta^{\mathrm{b}^H}_{\triangle} \leq^{\mathrm{metr}} \delta^{\mathrm{b}}$. That is a direct consequence of the following theorem:

Theorem 5 *$\delta^{\mathrm{b}^H}_{\triangle}$ is a pre-fixpoint of F.*

Proof. We need to show that $\delta^{\mathrm{b}^H}_{\triangle} \leq^{\mathrm{metr}} F(\delta^{\mathrm{b}^H}_{\triangle})$. Please remember that the preorder on metrics corresponds to the reverse of the point-wise preorder for states. So if we read this inequality on metrics as an inequality on the states of $\mathscr{M}^{\Lambda}$, we see that it is equivalent to: for every $s,t \in S^{\Lambda}$, $F(\delta^{\mathrm{b}^H}_{\triangle})(s,t) \leq \delta^{\mathrm{b}^H}_{\triangle}(s,t)$. If we unfold the definition of the operator F on metrics, we can see that it means that for every $a \in \mathscr{L}^{\Lambda}$, $\delta^{\mathrm{b}^H}_{\triangle}(\mathscr{P}^{\Lambda}(s,a), \mathscr{P}^{\Lambda}(t,a)) \leq \delta^{\mathrm{b}^H}_{\triangle}(s,t)$. Please remember that there are two kinds of actions in our LMC: the action eval of evaluating a program to obtain a value distribution, and the action $@V$, which corresponds to passing the value V to a distinguished value. If we consider separately each of these actions, we see that the result we want to have is equivalent to:

- Let be M, N closed terms. Then $\delta^{\mathrm{b}^H}_{\triangle}(\widehat{[\![M]\!]}, \widehat{[\![N]\!]}) \leq \delta^{\mathrm{b}^H}_{\triangle}(M,N)$
- Let be M, N such that $x \vdash M$ and $x \vdash N$, and let V be a value. Then it holds that: $\delta^{\mathrm{b}^H}_{\triangle}(M\{V/x\}, N\{V/x\}) \leq \delta^{\mathrm{b}^H}_{\triangle}(\widehat{\lambda x.M}, \widehat{\lambda x.N})$

We are first going to show these two result to the original premetric on terms δ^{b^H}, and we will extend them later to $\delta^{\mathrm{b}^H}_{\triangle}$.

Lemma 25 (Key-Lemma) *Let be M and N two closed terms. Then:*

$$\delta^{b^H}([\![M]\!],[\![N]\!]) \leq \delta^{b^H}(M,N)$$

Proof. We show in fact that, for every ε such that the judgement $\vdash \delta^{\mathrm{b}^H}(M,N) \leq \varepsilon$ is a valid one, it holds that $\delta^{\mathrm{b}^H}([\![M]\!],[\![N]\!]) \leq \varepsilon$. We show that by induction on the structure of the derivation of: $M \Downarrow [\![M]\!]$

- If M is a value: then $M = \lambda x.K$, and the derivation of $M \Downarrow [\![M]\!]$ is of the following form:

$$\frac{}{\lambda x.K \Downarrow \{\lambda x.K^1\}}$$

Then the proof tree allowing to certify the validity of $\vdash \delta^{\mathrm{b}^H}(M,N) \leq \varepsilon$ should be of the form:

$$\frac{x \vdash \delta^{\mathrm{b}^H}(T,K) \leq \gamma \qquad \delta^{\mathrm{b}}(\lambda x.K, N) \leq \iota}{\vdash \delta^{\mathrm{b}^H}(\lambda x.T, N) \leq \varepsilon = \gamma + \iota}$$

Since δ^{b} is a fixpoint of F, we have that: $\delta^{\mathrm{b}}([\![\lambda x.K]\!],[\![N]\!]) \leq \delta^{\mathrm{b}}(\lambda x.K, N) \leq \iota$. And so:

$$\begin{aligned} \delta^{\mathrm{b}^H}([\![\lambda x.T]\!],[\![N]\!]) &\leq \delta^{\mathrm{b}^H}([\![\lambda x.T]\!],[\![\lambda x.K]\!]) + \delta^{\mathrm{b}}([\![\lambda x.K]\!],[\![N]\!]) && \text{by lemma 21} \\ &\leq \delta^{\mathrm{b}^H}([\![\lambda x.T]\!],[\![\lambda x.K]\!]) + \delta \end{aligned}$$

Moreover, since $\lambda x.T$ and $\lambda x.K$ are values, we know that: $[\![\lambda x.T]\!] = \{\lambda x.T^1\}$, and $[\![\lambda x.K]\!] = \{\lambda x.K^1\}$. By Lemma 14, we can see that: ${\delta^{\mathrm{b}}}^H([\![\lambda x.T]\!], [\![\lambda x.K]\!]) = {\delta^{\mathrm{b}}}^H(\lambda x.T, \lambda x.K)$. It follows that: ${\delta^{\mathrm{b}}}^H([\![\lambda x.T]\!], [\![N]\!]) \leq {\delta^{\mathrm{b}}}^H(\lambda x.T, \lambda x.K) + \delta$ and since we have the following proof tree, it allows us to conclude.

$$\frac{x \vdash {\delta^{\mathrm{b}}}^H(T, K) \leq \gamma \qquad \delta^{\mathrm{b}}(\lambda x.K, \lambda x.K) \leq 0}{\vdash {\delta^{\mathrm{b}}}^H(\lambda x.T, \lambda x.K) \leq \gamma}$$

- If $M = UL$. Then the derivation of $M \Downarrow [\![M]\!]$ is the following:

$$\frac{\begin{array}{c} U \Downarrow [\![U]\!] \qquad L \Downarrow [\![L]\!] \\ \{P\{V/x\} \Downarrow [\![P\{V/x\}]\!]\}_{\lambda x.P \in \mathsf{S}([\![U]\!]), V \in \mathsf{S}([\![L]\!])} \end{array}}{MN \Downarrow \sum [\![U]\!](\lambda x.L) \cdot [\![L]\!](V) \cdot [\![P\{V/x\}]\!]}$$

And the proof tree corresponding to the validity of $\vdash {\delta^{\mathrm{b}}}^H(M, N) \leq \varepsilon$ has the following form:

$$\frac{\vdash {\delta^{\mathrm{b}}}^H(U, K) \leq \beta \qquad \vdash {\delta^{\mathrm{b}}}^H(L, T) \leq \gamma \qquad \delta^{\mathrm{b}}(KT, N) \leq \iota}{\vdash {\delta^{\mathrm{b}}}^H(UL, N) \leq \varepsilon = \beta + \gamma + \iota}$$

We have:

$$\begin{aligned} {\delta^{\mathrm{b}}}^H([\![UL]\!], [\![N]\!]) &\leq {\delta^{\mathrm{b}}}^H([\![UL]\!], [\![KT]\!]) + \delta^{\mathrm{b}}([\![KT]\!], N) \\ &\leq {\delta^{\mathrm{b}}}^H([\![UL]\!], [\![KT]\!]) + \iota \end{aligned}$$

So it is enough to show that: ${\delta^{\mathrm{b}}}^H([\![UL]\!], [\![KT]\!]) \leq \varepsilon + \gamma$.

So we have that:

$$\begin{aligned} {\delta^{\mathrm{b}}}^H([\![UL]\!], [\![KT]\!]) &= \max\left\{ \sum_s a_s \cdot [\![UL]\!](s) + b_s [\![KT]\!](s) \mid a_s \leq 1, b_s \leq 1 \wedge a_s + b_t \leq {\delta^{\mathrm{b}}}^H(s, t) \right\} \\ &= \max\{ \sum_s (a_s \cdot \sum_P \sum_V [\![U]\!](\lambda x.P) \cdot [\![L]\!](V) \cdot [\![P\{V/x\}]\!](s) + \\ &\qquad b_s \cdot \sum_Q \sum_W [\![K]\!](\lambda x.Q) \cdot [\![T]\!](W) \cdot [\![Q\{W/x\}]\!](s)) \\ &\quad \mid a_s \leq 1, b_s \leq 1 \wedge a_s + b_t \leq {\delta^{\mathrm{b}}}^H(s, t)\} \end{aligned}$$

We are now going to use the dual characterisation of the lifting of a metric to a distribution:
We know that: ${\delta^{\mathrm{b}}}^H(U, K) \leq \varepsilon$.
So there exist $(l_{P,Q})_{\lambda P. \in \mathsf{S}([\![U]\!]), \lambda Q. \in \mathsf{S}([\![K]\!])}$, and $(x_P)_{\lambda x.P \in \mathsf{S}([\![U]\!])}$, and $(y_Q)_{\lambda x.Q \in \mathsf{S}([\![K]\!])}$, such that:

$$\sum_{P,Q} l_{P,Q} \cdot {\delta^{\mathrm{b}}}^H(\lambda x.P, \lambda x.Q) + \sum_P x_P + \sum_Q y_Q = {\delta^{\mathrm{b}}}^H(U, K) \tag{12}$$

$$\sum_P l_{P,Q} + y_Q = [\![K]\!](\lambda x.Q) \tag{13}$$

$$\sum_Q l_{P,Q} + x_P = [\![U]\!](\lambda x.P) \tag{14}$$

Please observe that the equation (12) implies in particular that: $\sum_{P,Q} l_{P,Q} \leq 1$. Similarly, the equations (13) and (14) implies that $\sum_P x_{P,} \leq 1$ and $\sum_Q y_Q \leq 1$

and

$$\begin{aligned}
{\delta^{\mathrm{b}}}^{H}([\![UL]\!],[\![KT]\!]) = \max\{&\sum_s (a_s \cdot \sum_P \sum_V \left(\sum_Q l_{P,Q} + x_P\right) \cdot [\![L]\!](V) \cdot [\![P\{V/x\}]\!](s) \\
&+ b_s \sum_Q \sum_W \left(\sum_P l_{P,Q} + y_Q\right) \cdot [\![T]\!](W) \cdot [\![Q\{W/x\}]\!](s)) \\
&\mid a_s \leq 1 \wedge b_s \leq 1 \wedge a_s + b_t \leq {\delta^{\mathrm{b}}}^{H}(s,t)\}
\end{aligned}$$

$$\begin{aligned}
= \max\{&\sum_s \sum_{P,Q} (l_{P,Q} \sum_V a_s \cdot [\![L]\!](V) \cdot [\![P\{V/x\}]\!](s) + b_s \cdot [\![T]\!](W) \cdot [\![Q\{W/x\}]\!](s) \\
&+ \sum_s a_s \cdot \sum_P x_P \sum_V \cdot [\![L]\!](V) \cdot [\![P\{V/x\}]\!](s) \\
&+ \sum_s b_s \sum_Q y_Q \sum_W \cdot [\![T]\!](W) \cdot [\![Q\{W/x\}]\!](s) \\
&\mid a_s \leq 1 \wedge b_s \leq 1 \wedge a_s + b_t \leq {\delta^{\mathrm{b}}}^{H}(s,t)\}
\end{aligned}$$

$$\begin{aligned}
\leq \max\{&\sum_s \sum_{P,Q} (l_{P,Q} \sum_V a_s \cdot [\![L]\!](V) \cdot [\![P\{V/x\}]\!](s) + b_s \cdot [\![T]\!](W) \cdot [\![Q\{W/x\}]\!](s) \\
&\mid a_s \leq 1 \wedge b_s \leq 1 \wedge a_s + b_t \leq {\delta^{\mathrm{b}}}^{H}(s,t)\} \\
+ \max\{&\sum_s a_s \cdot \sum_P x_P \sum_V \cdot [\![L]\!](V) \cdot [\![P\{V/x\}]\!](s) \\
&\mid a_s \leq 1, b_s \leq 1 \wedge a_s + b_t \leq {\delta^{\mathrm{b}}}^{H}(s,t)\} \\
+ \max\{&\sum_s \sum_Q y_Q \sum_W b_s \cdot [\![T]\!](W) \cdot [\![Q\{W/x\}]\!](s) \\
&\mid a_s \leq 1, b_s \leq 1 \wedge a_s + b_t \leq {\delta^{\mathrm{b}}}^{H}(s,t)\}
\end{aligned}$$

$$\begin{aligned}
\leq \max\{&\sum_s \sum_{P,Q} (l_{P,Q} \sum_V a_s \cdot [\![L]\!](V) \cdot [\![P\{V/x\}]\!](s) + b_s \cdot [\![T]\!](W) \cdot [\![Q\{W/x\}]\!](s) \\
&\mid a_s \leq 1 \wedge b_s \leq 1 \wedge a_s + b_t \leq {\delta^{\mathrm{b}}}^{H}(s,t)\} \\
+ &\sum_P x_P + \sum_Q y_Q
\end{aligned}$$

We can now apply the induction hypothesis to ${\delta^{\mathrm{b}}}^{H}(L,T) \leq \gamma$. We obtain that: ${\delta^{\mathrm{b}}}^{H}([\![L]\!],[\![T]\!]) \leq \gamma$. So there exist $(h_{V,W})_{V \in \mathsf{S}([\![L]\!]), W \in \mathsf{S}([\![T]\!])}$, and $(w_V)_{V \in \mathsf{S}([\![L]\!])}$, and $(z_W)_{W \in \mathsf{S}([\![T]\!])}$, such that:

$$\sum_{V,W} h_{V,W} \cdot {\delta^{\mathrm{b}}}^{H}(V,W) + \sum_V w_V + \sum_W z_W = {\delta^{\mathrm{b}}}^{H}(L,T) \leq \gamma \tag{15}$$

$$\sum_V h_{V,W} + z_W = [\![T]\!](W) \tag{16}$$

$$\sum_W h_{V,W} + w_V = [\![L]\!](V) \tag{17}$$

And now we have:

${\delta^{\mathrm{b}}}^{H}([\![UL]\!],[\![KT]\!])$

$$
\begin{aligned}
&\leq \max\{\sum_s \sum_{P,Q} l_{P,Q}(\sum_V (\sum_W h_{V,W} + w_V) \cdot [\![P\{V/x\}]\!](s) \cdot a_s \\
&\qquad\qquad + \sum_W (\sum_V h_{V,W} + z_W) \cdot [\![Q\{W/x\}]\!](s) \cdot b_s) \\
&\qquad \mid a_s \leq 1, b_s \leq 1 \wedge a_s + b_t \leq {\delta^{\mathrm{b}}}^H(s,t)\} \\
&\quad + \sum_P x_P + \sum_Q y_Q
\end{aligned}
$$

$$
\begin{aligned}
&\leq \max\{\sum_{P,Q} (l_{P,Q}(\sum_{V,W} h_{V,W} \sum_s [\![P\{V/x\}]\!](s) \cdot a_s + [\![Q\{W/x\}]\!](s) \cdot b_s \\
&\qquad + \sum_V w_V \cdot \sum_s [\![P\{V/x\}]\!](s) \cdot a_s \\
&\qquad + \sum_W z_W \cdot \sum_s [\![Q\{W/x\}]\!](s)) \cdot b_s \\
&\qquad \mid a_s \leq 1 \wedge b_s \leq 1 \wedge a_s + b_t \leq {\delta^{\mathrm{b}}}^H(s,t)\} \\
&+ \sum_P x_P + \sum_Q y_Q
\end{aligned}
$$

$$
\begin{aligned}
&\leq \sum_{P,Q} l_{P,Q}(\sum_{V,W} h_{V,W} \max\{\sum_s [\![P\{V/x\}]\!](s) \cdot a_s + [\![Q\{W/x\}]\!](s) \cdot b_s \mid a_s \leq 1 \wedge b_s \leq 1 \wedge a_s + b_t \leq {\delta^{\mathrm{b}}}^H(s,t)\} \\
&\qquad + \sum_V w_V \cdot \max\{\sum_s [\![P\{V/x\}]\!](s) \cdot a_s \mid a_s \leq 1 \wedge b_s \leq 1 \wedge a_s + b_t \leq {\delta^{\mathrm{b}}}^H(s,t)\} \\
&\qquad + \sum_W z_W \cdot \max\{\sum_s [\![Q\{W/x\}]\!](s) \cdot b_s \mid a_s \leq 1 \wedge b_s \leq 1 \wedge a_s + b_t \leq {\delta^{\mathrm{b}}}^H(s,t)\}) \\
&+ \sum_P x_P + \sum_Q y_Q
\end{aligned}
$$

Now, we can use equation (15), and the fact that the sum of a distribution is always lesser or equal to 1:

$$
\begin{aligned}
{\delta^{\mathrm{b}}}^H([\![UL]\!], [\![KT]\!]) \leq &\sum_{P,Q} l_{P,Q} \left(\sum_{V,W} h_{V,W} {\delta^{\mathrm{b}}}^H(P\{V/x\}, Q\{W/x\}) + \sum_V w_V + \sum_W z_W \right) \\
&+ \sum_P x_P + \sum_Q y_Q
\end{aligned}
$$

We can here use lemma 22, which states that ${\delta^{\mathrm{b}}}^H$ is pseudo-substitutive:

$$
\begin{aligned}
{\delta^{\mathrm{b}}}^H([\![UL]\!], [\![KT]\!]) \leq &\sum_{P,Q} l_{P,Q} \left(\sum_{V,W} h_{V,W} ({\delta^{\mathrm{b}}}^H(P,Q) + {\delta^{\mathrm{b}}}^H(V,W)) + \sum_V w_V + \sum_W z_W \right) \\
&+ \sum_P x_P + \sum_Q y_Q
\end{aligned}
$$

$$\leq \sum_{P,Q} l_{P,Q} (\sum_{V,W} h_{V,W}) \cdot \delta^{\mathrm{b}^H}(P,Q) + \sum_P x_P + \sum_Q y_Q$$
$$+ \sum_{P,Q} l_{P,Q} \sum_{V,W} h_{V,W} \cdot \delta^{\mathrm{b}^H}(V,W) + \sum_V w_V + \sum_W z_W)$$

and, since $\sum_{V,W} h_{V,W} \leq 1$, and similarly $\sum_{P,Q} l_{P,Q} \leq 1$, we have that:

$$\delta^{\mathrm{b}^H}(\llbracket UL \rrbracket, \llbracket KT \rrbracket) \leq \sum_{P,Q} l_{P,Q} \cdot \delta^{\mathrm{b}^H}(P,Q) + \sum_P x_P + \sum_Q y_Q$$
$$+ \sum_{V,W} h_{V,W} \cdot \delta^{\mathrm{b}^H}(V,W) + \sum_V w_V + \sum_W z_W$$

We can now use equations (12) and (15):

$$\delta^{\mathrm{b}^H}(\llbracket UL \rrbracket, \llbracket KT \rrbracket) \leq \delta^{\mathrm{b}^H}(\llbracket U \rrbracket, \llbracket K \rrbracket) + \delta^{\mathrm{b}^H}(\llbracket L \rrbracket, \llbracket T \rrbracket) \leq \varepsilon + \gamma$$

$\square$

Lemma 26

$$\delta^{b^H}(M\{V/x\}, N\{V/x\}) \leq \delta^{b^H}(\lambda x.M, \lambda x.N)$$

Proof. Let be ε such that: $\vdash \delta^{\mathrm{b}^H}(\lambda x.M, \lambda x.N) \leq \varepsilon$

The only rule that can have been applied is:

$$\frac{x \vdash \delta^{\mathrm{b}^H}(M,K) \leq \gamma \qquad \delta^{\mathrm{b}}(\lambda x.K, \lambda x.N) \leq \iota \qquad \vdash \lambda x.N}{\vdash \delta^{\mathrm{b}^H}(\lambda x.M, \lambda x.N) \leq \varepsilon = \gamma + \iota}$$

We can now apply Lemma 22 to $x \vdash \delta^{\mathrm{b}^H}(M,K) \leq \varepsilon$, and we see that: $\delta^{\mathrm{b}^H}(M\{V/x\}, K\{V/x\}) \leq \delta^{\mathrm{b}^H}(M,K) \leq \varepsilon$. Moreover, we know that $\delta^{\mathrm{b}}(\lambda x.K, \lambda x.N) \leq \iota$. Since δ^{b} is a fixpoint for F, we can see that:

$$\gamma \geq \delta^{\mathrm{b}}(\lambda x.K, \lambda x.N) = \delta^{\mathrm{b}}(\widehat{\lambda x.K}, \widehat{\lambda x.N}) \geq \delta^{\mathrm{b}}(K\{V/x\}, N\{V/x\})$$

and now we can conclude by Lemma 21 that: $\delta^{\mathrm{b}^H}(M\{V/x\}, N\{V/x\}) \leq \iota + \gamma$ $\square$

Now we extend these two lemmas to $\delta^{\mathrm{b}^H}_{\triangle}$:

Lemma 27 *Let be M, N two terms. Then*

$$\delta^{\mathrm{b}^H}_{\triangle}(\widehat{\llbracket M \rrbracket}, \widehat{\llbracket N \rrbracket}) \leq \delta^{\mathrm{b}^H}_{\triangle}(M,N)$$

Proof. Let be ε such that the judgement $\vdash \delta^{\mathrm{b}^H}_{\triangle}(M,N) \leq \varepsilon$ is valid by the rules of figure 6. We are going to show by induction on the structure of its derivation that: $\delta^{\mathrm{b}^H}_{\triangle}(\widehat{\llbracket M \rrbracket}, \widehat{\llbracket N \rrbracket}) \leq \varepsilon$. We consider different cases depending of the structure of the proof tree used to derive the validity of $\vdash \delta^{\mathrm{b}^H}_{\triangle}(M,N) \leq \varepsilon$:

- If the proof tree is:
$$\frac{\delta^{\mathrm{b}^H}(M,N) \leq \varepsilon}{\vdash \delta^{\mathrm{b}^H}_{\triangle}(M,N) \leq \varepsilon}$$

We can use Lemma 25, and we obtain that ${\delta^{\mathrm{b}}}^H(\llbracket M\rrbracket, \llbracket N\rrbracket) \leq \varepsilon$. Now we can see that:

$$\begin{aligned}
{\delta^{\mathrm{b}}}^H_{\triangle}(\widehat{\llbracket M\rrbracket}, \widehat{\llbracket N\rrbracket}) &\leq {\delta^{\mathrm{b}}}^H(\widehat{\llbracket M\rrbracket}, \widehat{\llbracket N\rrbracket}) && \text{since } {\delta^{\mathrm{b}}}^H \leq^{\mathrm{metr}} {\delta^{\mathrm{b}}}^H_{\triangle} \\
&= {\delta^{\mathrm{b}}}^H(\llbracket M\rrbracket, \llbracket N\rrbracket) && \text{by construction of the extension of } {\delta^{\mathrm{b}}}^H \text{ to } S^{\Lambda} \\
&\leq \varepsilon
\end{aligned}$$

- If the proof tree is of the form:

$$\frac{\vdash {\delta^{\mathrm{b}}}^H_{\triangle}(M,N) \leq \gamma \qquad \vdash {\delta^{\mathrm{b}}}^H_{\triangle}(N,M) \leq \iota}{\vdash {\delta^{\mathrm{b}}}^H_{\triangle}(M,N) \leq \varepsilon = \min(\gamma,\iota)}$$

  We can apply the induction hypothesis to $\vdash {\delta^{\mathrm{b}}}^H_{\triangle}(M,N) \leq \gamma$ and $\vdash {\delta^{\mathrm{b}}}^H_{\triangle}(N,M) \leq \iota$. We obtain that ${\delta^{\mathrm{b}}}^H_{\triangle}(\widehat{\llbracket M\rrbracket}, \widehat{\llbracket N\rrbracket}) \leq \gamma$ and that ${\delta^{\mathrm{b}}}^H_{\triangle}(\widehat{\llbracket N\rrbracket}, \widehat{\llbracket M\rrbracket}) \leq \iota$. Since ${\delta^{\mathrm{b}}}^H_{\triangle}$ is symmetric, it means that: ${\delta^{\mathrm{b}}}^H_{\triangle}(\widehat{\llbracket M\rrbracket}, \widehat{\llbracket N\rrbracket}) \leq \iota$. And so we have the result.

- If the proof tree is of the form:

$$\frac{\vdash {\delta^{\mathrm{b}}}^H_{\triangle}(M,s) \leq \gamma \qquad \vdash {\delta^{\mathrm{b}}}^H_{\triangle}(s,N) \leq \leq \iota}{\vdash {\delta^{\mathrm{b}}}^H_{\triangle}(M,N) \leq \varepsilon = \gamma + \iota}$$

  If $\varepsilon = 1$, the result holds. Otherwise, please observe that s cannot be a distinguished value. So there exist a closed term L such that $s = L$. By induction hypothesis: ${\delta^{\mathrm{b}}}^H_{\triangle}(\widehat{\llbracket M\rrbracket}, \widehat{\llbracket L\rrbracket}) \leq \gamma$, and ${\delta^{\mathrm{b}}}^H_{\triangle}(\widehat{\llbracket L\rrbracket}, \widehat{\llbracket N\rrbracket}) \leq \iota$. So by Lemma 15, and since ${\delta^{\mathrm{b}}}^H_{\triangle}$ verifies the triangular inequality, we have: ${\delta^{\mathrm{b}}}^H_{\triangle}(\widehat{\llbracket M\rrbracket}, \widehat{\llbracket N\rrbracket}) \leq \gamma + \iota$.

$\square$

Lemma 28 *For every M, N:*

$${\delta^{b}}^H_{\triangle}(M\{V/x\}, N\{V/x\}) \leq {\delta^{b}}^H_{\triangle}(\widehat{\lambda x.M}, \widehat{\lambda x.N})$$

Proof. Let be ε such that the judgement $\vdash {\delta^{\mathrm{b}}}^H_{\triangle}(\widehat{\lambda x.M}, \widehat{\lambda x.N}) \leq \varepsilon$ is valid. As for the previous lemma, the proof is by induction on the structure of the proof tree for this judgement. $\square$

$\square$

Since ${\delta^{\mathrm{b}}}^H$ is non-expansive by construction, we now have the result we were aiming for:

Theorem 6 *δ^b is non-expansive.*

Proof. As a consequence of Theorem **??**, $\delta^{\mathrm{b}} = {\delta^{\mathrm{b}}}^H$. Since ${\delta^{\mathrm{b}}}^H$ is non-expansive, the result holds. $\square$

5.3 On Full-Abstraction and Pairs

The bisimulation distance is a sound approximation of the context distance. But how about full-abstraction? Is there any hope to prove that the two coincide? The answer is negative: there are terms whose distance is *strictly* higher in the bisimulation metric than in the context (or trace) metric.

Example 4 *Consider the following terms: M corresponds to the program that takes an argument, and then returns I with probability $\frac{1}{2}$, and diverges with probability $\frac{1}{2}$. N corresponds to the program which chooses first between the function which return I whenever it is called, and the function which diverges whenever called. Formally:*

$$M := \lambda x.(I \oplus \Omega); \qquad N := (\lambda x.I) \oplus (\lambda x.\Omega).$$

These two terms are at distance 0 for the context distance: since the calculus is linear, the step where the choice is done is irrelevant. However, $\delta^b(M, N) = \frac{1}{2}$: the proof, use the characterisation of bisimulation distance by testing from [9], in which not only linear tests, but also more complicated tests (like threshold tests*) are available.*

But how about pairs? Indeed, for the sake of simplicity, we have presented the metatheory of the bisimulation metric for a purely applicative λ-calculus. Following the lines of our discussion in Section 4.4, however, the LMC $\mathscr{M}^\Lambda$ can be extended into one handling pairs in a relatively simple way. The difficulties we encountered when trying to evaluate the (trace, or context) distance between pairs of terms unfortunately remain: it is not clear whether coinduction could provide any additional *advantage* over contextual distance. As for the trace metric in the previous section, we would like to extend the bisimulation metric to a language with pairs. In order to do that, we add the action $\otimes K$ to the LMC $\mathscr{M}^\Lambda$. We transform the definition to the probability matrix $\mathscr{P}^\Lambda$ by adding:

$$\mathscr{P}^\Lambda(\langle M, N\rangle)(\otimes L) = \sum_{V,W} [\![M]\!](V) \cdot [\![N]\!](W) \cdot \{L\{V, W/x, y\}^1\}$$

We now have to transform the definition of validity for Howe's judgement in order to consider the case of pairs:

$$\frac{\begin{array}{cc}\Gamma \vdash \mu^H(M, K) \leq \varepsilon & \Gamma, \Delta \vdash L \\ \Delta \vdash \mu^H(N, T) \leq \gamma & \mu(\langle K, T\rangle, L) \leq \iota\end{array}}{\Gamma, \Delta \vdash \mu^H(\langle M, N\rangle, L) \leq \varepsilon + \gamma + \iota}$$

6 The Tuple Distance

The two metrics we have just defined have been shown to be non-expansive, even if the calculus is extended with pairs. In that case, however, they do not represent so much of an improvement with respect to the context distance. Please recall *where* the problem comes from: we would like to define actions starting from $\langle M, N\rangle$, and respecting the affine paradigm. We have seen that taking *projections* as actions lead to an unsound metric, and we have circumvented the problem by considering an action $\otimes L$, following [8]. Intuitively the action $\otimes L$ corresponds to replacing the free variables of L (which are supposed to be included in $\{x, y\}$) by the components of the pair: if for instance V and W are values, we have that $\langle V, W\rangle \overset{\otimes L}{\rightsquigarrow} \{L\{V, W/x, y\}^1\}$. But what can any environment L do if we give it V and W as two values to interact with? Let us suppose that both V and W are functions, and remember that we are in an affine setting. The environment can (probabilistically) pass some arguments to V, and independently some other arguments to W, and then possibly pass to one of the two programs an argument that contains the other one. The idea behind the construction we present in this section, then, is to *keep* the information about the two components of the pairs in the states until they really interact with each other.

Our idea can be made concrete by introducing another LMC, whose states are not closed terms anymore, but *tuples* in the form $[V_1, \cdots, V_n]$, where $V_1, \cdots V_n$ are values. The possible actions the environment can perform on a tuple $[V_1, \cdots, V_n]$ correspond to the choice of an index $i \in \{1, \cdots, n\}$ and of an action to apply to the value V_i. If V_i is a pair, the only possible action is to split it into two components. We call this action unfold^i. If V_i is a function, the environment can pass it an argument, which can possibly be constructed using other V_j's. More precisely, the argument is built by way of an open term $\mathcal{C}$, and a typing context Γ, such that $\Gamma \vdash \mathcal{C}$, and Γ is a subset of $\{x_j \mid j \neq i\}$: the free variables of Γ represent the places where other values V_j, with

$$S^{\Lambda}_{\text{mul}} = \{[V_1, \cdots, V_n] \mid V_1, \cdots, V_n \text{ closed values }\}$$
$$\mathscr{A}^{\Lambda}_{\text{mul}} = \{\text{unfold}^i \mid i \in \mathbb{N}\} \cup \{@(\Gamma, \mathcal{C})^i \mid i \in \mathbb{N}, (\Gamma, \mathcal{C}) \text{ a } (n,i)\text{-open-value }\}.$$

$$\mathscr{P}^{\Lambda}_{\text{mul}}([s_1, \cdots, \langle N, L\rangle, \cdots, s_n], \text{unfold}^i)([s_1, \cdots s_{i-1}, V, W, s_{i+1} \cdots, s_n]) = [\![N]\!](V) \cdot [\![L]\!](W)$$
$$\mathscr{P}^{\Lambda}_{\text{mul}}([s_1, \cdots, \lambda y.N, \cdots, s_n], @(\Gamma, \mathcal{C})^i)([s_{h_1}, \cdots, W, \cdots, s_{h_m}]) = [\![N\{\mathcal{C}\{s_{j_l}/x_{j_l}\}/y\}]\!](W)$$
$$\text{with } \{1, \cdots, n\} = i \cup \{j_1, ..., j_k\} \cup \{h_1, ..., h_m\} \text{ (disjoint union) and } \Gamma = x_{j_1}, \cdots x_{j_k}$$

Figure 7: The Tuple LMC

$j \neq i$, are used. Moreover, we ask that for any value $W_1, \cdots W_n$, the term obtained in substituting x_j by W_j is a value: it means that $\mathcal{C}$ is one of the x_j, or of the form $\lambda y.\mathcal{D}$. We call a pair $(\Gamma, \mathcal{C})$ which verifies these conditions a (n,i)*-open value*. Formally, the LMC $\mathscr{M}^{\Lambda}_{\text{mul}} = (S^{\Lambda}_{\text{mul}}, \mathscr{A}^{\Lambda}_{\text{mul}}, \mathscr{P}^{\Lambda}_{\text{mul}})$ is defined in Figure 7.

6.1 The Metric

We are going to define a metric on closed terms which corresponds to linear tests in $\mathscr{M}^{\Lambda}_{\text{mul}}$. First, we define *tuple traces* simply as words over $\mathscr{A}^{\Lambda}_{\text{mul}}$. The probability to succeed in doing a trace s starting from a tuple $K \in S^{\Lambda}_{\text{mul}}$ can be naturally defined, and paves the way to defining a metric on tuples of values:

$$Pr^{\text{mul}}(K, \varepsilon) = 1;$$
$$Pr^{\text{mul}}(K, a \cdot s) = \sum_H \mathscr{P}(K, a)(H) \cdot Pr^{\text{mul}}(H, s);$$
$$\delta^{\text{mul}}(K, H) = \sup_s |Pr^{\text{mul}}(K, s) - Pr^{\text{mul}}(H, s)|.$$

What we need, however, is a metric on *programs*. Please remember that states of the LMC $\mathscr{M}^{\Lambda}_{\text{mul}}$ are tuples of *values*. Any program M, however, can be viewed as the distribution of values obtained by evaluating it, i.e. its semantics $[\![M]\!]$:

$$\delta^{\text{mul}}(M, N) = \sup_s |\sum [\![M]\!](V) \cdot Pr^{\text{mul}}([V], s) - \sum [\![N]\!](W) \cdot Pr^{\text{mul}}([W], s)|.$$

The just introduced metric should at least be put in relation to the context metric for it to be useful. We know from Section 4 that the context metric coincides with the trace metric. The following theorem relates the trace metric δ^{tr} and the metric δ^{mul}:

Theorem 7 *Let I be any finite set of variables, and $\{V_x\}_{x \in I}$ and $\{W_x\}_{x \in I}$ any two collections of values. For any open term $\mathcal{C}$ such that $I \vdash \mathcal{C}$, it holds that:*

$$\delta^{tr}(\mathcal{C}\{V_x/x\}_{(x \in I)}, \mathcal{C}\{W_x/x\}_{(x \in I)}) \leq \delta^{mul}([V_x]_{(x \in I)}, [W_x]_{(x \in I)})$$

Proof. The proof of Theorem 7 is similar to the proof of non-expansiveness for the trace metric : first we define a small step semantics, which corresponds to the transition relation in the Markov Chain $\mathscr{M}^{\Lambda}_{\text{mul}}$, then we define another small step semantics, which corresponds to keep separated the context, which is now seen as a term with several holes, and the tuple used to fill it, and we end the proof by defining a notion of ε-parentality for disributions over pairs of contexts and tuple, and showing a stability result for ε-parents distributions. These steps are displayed in more details below.

6.1.1 Trace Semantics Big Steps for Tuples

We're going to be interested in the labelled transition system on finite distributions over S^{Λ}_{mul} induced by the Markov Chain $\mathscr{M}^{\Lambda}_{\text{mul}}$.

Definition 20 *We'll note $\Delta(S^{\Lambda}_{mul})$ the set of finite distributions over S^{Λ}_{mul}. We define a reduction relation $\mathscr{D} \xrightarrow{a}_{\Delta(S^{\Lambda}_{mul})} \mathscr{E}$, where $\mathscr{D}, \mathscr{E} \in \Delta(S^{\Lambda}_{mul})$, and $a \in \mathscr{A}^{\Lambda}_{mul}$, by :*

$$\mathscr{D} \xrightarrow{a}_{\Delta(S^{\Lambda}_{mul})} \sum_{K} \mathscr{D}(K) \cdot \mathscr{P}^{\Lambda}_{mul}(K)(a)$$

Now, we define the success probability of a trace for a distribution as :

Definition 21 *If $s = a_1 \cdots a_n$,*

$$Pr^{bs}_{mul}(\mathscr{D}, s) = \sum \mathscr{E} \textit{ with } \mathscr{D} \xrightarrow{a_1}_{\Delta(S^{\Lambda}_{mul})} \cdots \xrightarrow{a_n}_{\Delta(S^{\Lambda}_{mul})} \mathscr{E}$$

The relation between this deterministic labelled transition system and the Markov Chain $\mathscr{M}^{\Lambda}_{\text{mul}}$ can be expressed by the following lemma :

Lemma 29 *Let be $K \in S^{\Lambda}_{mul}$, and s a trace. Then $Pr^{bs}_{mul}(\{K^1\}, s) = Pr^{mul}(K, a)$.*

6.1.2 Trace Semantics Small Steps for Tuples

We would like now to have a notion of small-step semantics for tuples corresponding to the trace semantics of the Markov Chain. Since we are now small steps, we should consider not only tuples of values, but tuples of terms as well. Moreover, during the execution, we should remember which term of the tuple is being reduced. For this reason, we must add intermediate states, where there is explicit focus on terms being evaluated.

Definition 22

- *We define a set $\mathcal{TV}$ consisting in closed terms of $\Lambda^{\langle,\rangle}_{\oplus}$, and distinguished values of $\Lambda^{\langle,\rangle}_{\oplus}$: $\mathcal{TV} = \{M \mid M \text{ closed term }\} \cup \{\widehat{V} \mid V \text{ closed value }\}$. Then we define the corresponding set of tuples $S = \{[s_1, \cdots, s_n] \mid s_1, \cdots, s_n \in \mathcal{TV}\}$*
- *$\mathcal{TV}^{focus} = \{M \mid M \text{ closed term }\} \cup \{focus^i(M) \mid M \text{ closed term}, i \in \mathbb{N}\} \cup \{\widehat{V} \mid V \text{ closed value }\}$, and $\mathcal{S}^{focus} = \left\{[s_1, \cdots, s_n] \mid s_1, \cdots, s_n \in \mathcal{TV}^{focus} \text{ and the focus integer are all distincts }\right\}$*

The term which should be reduced first is the term which has the smaller focus index. That's the sense of the following definition.

Definition 23 *For any $K \in \mathcal{S}^{focus}$, we note $f(K)$ defined as:*

- *$f(K) = \infty$ if K has no element with focus.*
- *$f([s_1, \cdots, s_n]) = j$ such that $s_i = focus^j(M)$ and j is the smaller focus in K*

We now define a small step probabilistic labelled reduction relation, where the actions can be divided in two kinds :

Definition 24 *We define a labelled reduction relation $K \xrightarrow{a}_{[]} \mathscr{D}$ where $K \in \mathcal{S}^{focus}$, $\mathscr{D}$ a distribution over $\mathcal{S}^{focus}$, and where $a \in Act^{ss} = \{\tau\} \cup \{eval^i \mid i \in \mathbb{N}\} \cup \{unfold^i \mid i \in \mathbb{N}\} \cup \{@(\Gamma, \mathcal{C})^i \mid i \in \mathbb{N}, (\Gamma, \mathcal{C}) \text{ a } (n,i)\text{open-value for a } n \in \mathbb{N}\}$. The rules are the one given in figure 8.*

τ is called an internal action, and corresponds to the internal reduction terms under focus in the tuple. The other actions are called external actions, and correspond to interactions with the environment. The definition given in Figure 8 use the small step semantics for term $\rightarrow$.

We want to formalize the probability of doing a trace for a distribution. First we lift the trace semantics to a reduction (non probabilistic) to distributions. We'll note $\Delta(\mathcal{S}^{\text{focus}})$ the set of finite distributions over $\mathcal{S}^{\text{focus}}$.

$$\frac{f(\big[s_1, \cdots, \mathrm{focus}^i(M), \cdots, s_n\big]) = i \qquad M \to \mathscr{D}}{\big[s_1, \cdots, \mathrm{focus}^i(M), \cdots, s_n\big] \xrightarrow{\tau}_{[]} \sum \mathscr{D}(N) \cdot \{\big[s_1, \cdots, \mathrm{focus}^i(N), \cdots, s_n\big]^1\}}$$

$$\frac{f(\big[s_1, \cdots, \mathrm{focus}^i(M), \cdots, s_n\big]) = i \qquad M \not\to}{\big[s_1, \cdots, \mathrm{focus}^i(M), \cdots, s_n\big] \xrightarrow{\tau}_{[]} \emptyset}$$

$$\frac{f(\big[s_1, \cdots, \mathrm{focus}^i(V), \cdots, s_n\big]) = i \qquad V \text{ is a value}}{\big[s_1, \cdots, \mathrm{focus}^i(V), \cdots, s_n\big] \xrightarrow{\tau}_{[]} \{\Big[s_1, \cdots, \widehat{V}, \cdots, s_n\Big]^1\}}$$

$$\frac{f([s_1, \cdots, s_n]) = \infty}{[s_1, \cdots, s_{i-1}, M, s_{i+1}, \cdots, s_n] \xrightarrow{\mathrm{eval}^i}_{[]} \{\big[s_1, \cdots, s_{i-1}, \mathrm{focus}^1(M), s_{i+1}, \cdots, s_n\big]^1\}}$$

$$\frac{f([s_1, \cdots, s_n]) = \infty}{\Big[s_1, \cdots, s_{i-1}, \widehat{\langle M, N \rangle}, s_{i+1}, \cdots, s_n\Big] \xrightarrow{\mathrm{unfold}^i}_{[]} \{\big[s_1, \cdots, s_{i-1}, \mathrm{focus}^1(M), \mathrm{focus}^2(N), s_{i+1}, \cdots, s_n\big]^1\}}$$

$$\frac{f([s_1, \cdots, s_n]) = \infty}{\Big[s_1, \cdots, s_{i-1}, \widehat{\lambda y.M}, s_{i+1}, \cdots, s_{j-1}, \widehat{V}, s_{j+1}, \cdots, s_n\Big] \xrightarrow{@x_j{}^i}_{[]} \{\big[s_1, \cdots, s_{i-1}, \mathrm{focus}^1(M\{V/y\}), s_{i+1}, \cdots, s_{j-1}, s_{j+1}, \cdots, s_n\big]^1\}}$$

$$\frac{f([s_1, \cdots, s_n]) = \infty \qquad (x_{j_k})_{1 \le k \le l} \vdash \mathcal{D} \qquad \{1, \cdots, n\} = \{j_k \mid 1 \le k \le n\} \sqcup \{h_k \mid 1 \le k \le m\}}{\Big[s_1, \cdots, s_{i-1}, \widehat{\lambda y.M}, s_{i+1}, \cdots, s_n\Big] \xrightarrow{@\lambda z.\mathcal{D}^i}_{[]} \{\Big[s_{h_1}, \cdots, \mathrm{focus}^1(M\{\lambda z.\mathcal{D}\{s_{j_k}/x_{j_k}\}_{1 \le k \le l}/y\}), \cdots, s_{h_m}\Big]^1\}}$$

Figure 8: small-step trace semantics for tuples

$$\frac{K \xrightarrow{\tau}_{[]} \mathscr{E}}{\mathscr{D} \dot{+} p \cdot \{K^1\} \xrightarrow{\epsilon}_{\Delta([])} \mathscr{D} + p \cdot \mathscr{E}}$$

$$\frac{K \xrightarrow{a}_{[]} \mathscr{E}_K \qquad \mathscr{D} \text{ in normal form}}{\mathscr{D} \xrightarrow{a}_{\Delta([])} \sum_{K \text{s.t } f(K)=\infty} \mathscr{D}(K) \cdot \mathscr{E}_K}$$

$$\frac{\mathscr{D} \xrightarrow{s}_{\Delta([])} \mathscr{E} \qquad \mathscr{E} \xRightarrow{t}_{\Delta([])} \mathscr{F}}{\mathscr{D} \xRightarrow{s \cdot t}_{\Delta([])} \mathscr{F}}$$

Figure 9: small-step trace semantics on distributions of tuples

Definition 25 *We define a labelled relation* $\mathscr{D} \Rightarrow^{a}_{\Delta([])} \mathscr{E}$*, where* $\mathscr{D}, \mathscr{E} \in \Delta(\mathcal{S}^{focus})$*, and* $a \in Act^{ss}$*. The rules are the one given in Figure 9.*

Definition 26 $Pr^{ss}_{mul}(\mathscr{D}, s) = \max\{\sum_{f(K)=\infty} \mathscr{E}^\star(K) \mid \mathscr{D} \Rightarrow^{s}_{\Delta([])} \mathscr{E}^\star\}$

Please observe that for any (external or internal) action a, the relations (between tuples and distributions over tuples) $K \xrightarrow{a}_{[]} \mathscr{D}$, and $K \xrightarrow{a}_{\Delta(S^{\Lambda}_{\text{mul}})} \mathscr{D}$ are deterministic. It's not the case anymore when we lift to relations between distributions, but we have the following lemma :

Lemma 30 *The reduction* $\cdot \rightarrow^{\tau}_{\Delta([])} \cdot$ *on distributions over* $\mathcal{S}^{focus}$ *is strongly normalizing.*

Proof. It follows from the fact that the relation $\rightarrow$ over distributions of terms is strongly normalizing. $\square$

We note $\mathscr{D}^\star$ the normal form of $\mathscr{D}$ for the relation $\cdot \rightarrow^{\tau}_{\Delta([])} \cdot$. By abuse of notation, if $s \in \mathcal{TV}^{\text{focus}}$, we note $s^\star$ for $\{s^1\}^\star$. We can in fact be more precised on the shape of the normal form of a distribution :

Definition 27 *Let be* $s \in \mathcal{TV}^{focus}$*. We define* $(s^\star)$ *by :*

- *If* $s = focus^i(M)$*, then* $(s^\star) = \sum_V [\![s]\!](V) \cdot \{\widehat{V}^1\}$
- *otherwise,* $(s^\star) = \{s^1\}$

Lemma 31 *Let be* $K = [s_1, \cdots, s_n] \in \mathcal{S}^{focus}$*. Then* $K^\star = \sum_{t_1,\cdots t_n} \prod_{1 \leq i \leq n}((s_i^\star))(t_i) \cdot \{[t_1, \cdots, t_n]^1\}$

Proof. The proof is by induction on the maximal number of reduction steps from $\mathscr{D}$ to $\mathscr{D}^\star$ (which is well defined since $\cdot \rightarrow^{\tau}_{\Delta([])} \cdot$ is strongly normalizing) $\square$

Now we want to compare the probability to do a trace for the small-step semantics and for the big-step semantics. For doing that, we show first the following lemma :

Lemma 32 *Let be* $a \in \mathscr{A}^{\Lambda}_{mul}$*, and* $\mathscr{D} \in \Delta(S^{\Lambda}_{mul})$*. Then let be* $\mathscr{E}$ *the distribution over* S^{Λ}_{mul} *such that* $\mathscr{D} \xrightarrow{a}_{\Delta(S^{\Lambda}_{mul})} \mathscr{E}$*. Let be* $\mathscr{F}$ *the distribution over* $\mathcal{S}^{focus}$ *such that :* $\widehat{\mathscr{D}} \xrightarrow{a}_{[]} \mathscr{F} \xrightarrow{\tau}_{[]} \cdots \xrightarrow{\tau}_{[]} \mathscr{F}^\star$*. Then :*

$$\widehat{\mathscr{E}} = \mathscr{F}^\star$$

Proof. Let be $a \in \mathscr{A}^{\Lambda}_{\text{mul}}$. We can see that for every $K \in S^{\Lambda}_{\text{mul}}$, there exists an (only one) $H \in \mathcal{S}^{\text{focus}}$ such that : $K \xrightarrow{a}_{[]} \{H^1\}$. It is sufficient to show that : if $\mathscr{D}$ is the distribution over S^{Λ}_{mul} such that $K \xrightarrow{a}_{\Delta(S^{\Lambda}_{\text{mul}})} \mathscr{D}$, we have that $H^\star = \widehat{\mathscr{D}}$. The proof of that is by case analysis on the rules of $\xrightarrow{a}_{[]}$, and using the characterisation given in Lemma 31 of the normal form for $\xrightarrow{\tau}_{[]}$

$\square$

Now we can extend this result to traces :

Lemma 33 *Let be* s *a word over* $\mathscr{A}^{\Lambda}_{mul}$*. Let be* $\mathscr{D}$ *a distribution over* S^{Λ}_{mul}*. Then :* $Pr^{bs}_{mul}(\mathscr{D}, s) = Pr^{ss}_{mul}(\widehat{\mathscr{D}}, s)$

Proof. The proof is by induction on the length of s.

- if $s = \epsilon$: $Pr^{\text{bs}}_{\text{mul}}(\mathscr{D}, \epsilon) = \sum \mathscr{D}$. Since $\widehat{\mathscr{D}}$ is a normal form for $\cdot \xrightarrow{\tau}_{[]} \cdot$, we have that : $Pr^{\text{ss}}_{\text{mul}}(\widehat{\mathscr{D}}, \epsilon) = \sum_{f(K)=\infty} \widehat{\mathscr{D}}(K) = \sum \mathscr{D}$
- if $s = a \cdot t$ then let be $\mathscr{E}$ such that $\mathscr{D} \xrightarrow{a}_{\Delta(S^{\Lambda}_{\text{mul}})} \mathscr{E}$. Then $Pr^{\text{bs}}_{\text{mul}}(\mathscr{D}, s) = Pr^{\text{bs}}_{\text{mul}}(\mathscr{E}, t)$. We apply Lemma 32, and we obtain that : $\widehat{\mathscr{D}} \xrightarrow{a}_{[]} \mathscr{F}$, and $\mathscr{F}^\star = \widehat{\mathscr{E}}$. Moreover, we have that : $Pr^{\text{ss}}_{\text{mul}}(\widehat{\mathscr{D}}, s) = Pr^{\text{ss}}_{\text{mul}}(\mathscr{F}, t) = Pr^{\text{ss}}_{\text{mul}}(\mathscr{F}^\star, t)$

$\square$

6.1.3 Trace semantics for distribution over contexts and tuples

Here we consider the same traces used for defining trace semantics for distribution on closed terms.

We are first going to introduce useful notations :

Definition 28 *We define an operator* $(\phi\,|^n\,\psi)$ *on functions by : If* $\phi : A \to \mathbb{N}$, $\psi : B \to \mathbb{N}$ *such that :*

- $A \cap B = \emptyset$
- $Im(\phi) \subseteq \{1, \cdots, n\}$

Then $(\phi\,|^n\,\psi) : A \cup B \to \mathbb{N}$ *is defined by :*

- $(\phi\,|^n\,\psi)(x) = \phi(x)$ *if* $x \in A$
- $(\phi\,|^n\,\psi)(x) = n + \psi(x)$ *if* $x \in B$

We now want to define a set of pairs of context with several holes, and tuples used for filling these holes. Formally the idea is the following : We first define things for the untyped case (without pairs) :

Definition 29

- *Let be* $\phi : \boldsymbol{V} \to \mathbb{N}$ *a partial injective function,* $\mathcal{C}$ *an (open) term, and* $[s_1, \cdots, s_n]$ *an element of* $\mathcal{S}^{focus}$. *We define the judgment* $x_1, ...x_m \vdash (\mathcal{C}, \phi, [s_1, \cdots, s_n])$ *by :*
  - $\{x_1, \cdots, x_m\} \cap Dom(\phi) = \emptyset$
  - $x_1 \cdots x_n, (y)_{y \in Dom(\phi)} \vdash \mathcal{C}$
  - $Im(\phi) \subseteq \{1 \cdots n\}$

  We define the judgment $x_1, ...x_m \vdash (\mathcal{C}, \phi, [s_1, \cdots, s_n]) :$ ***val*** *by :*
  - $x_1, ...x_m \vdash (\mathcal{C}, \phi, [s_1, \cdots, s_n])$
  - *if* $\mathcal{C} = y$, *then there exists a value* V *such that :* $s_{\phi(y)} = \widehat{V}$
  - *if* $\mathcal{C}$ *is not a variable,* $\mathcal{C}$ *is an abstraction (that is,* $\mathcal{C} = \lambda y.\mathcal{D}$, *where* $\mathcal{D}$ *is an open term), or* $\mathcal{C}$ *is a pair (that is* $\mathcal{C} = \langle \mathcal{D}_1, \mathcal{D}_2 \rangle$, *where* $\mathcal{D}_1$ *and* $\mathcal{D}_2$ *are open terms).*
- *We define the set of pairs of context and tuples which are well formed :*

$$\mathcal{A} = \{(\mathcal{C}, \phi, [s_1, \cdots, s_n]) \text{ such that } \emptyset \vdash (\mathcal{C}, \phi, [s_1, \cdots, s_n])\}$$

- *We define a notion of congruence for elements in* $\mathcal{A}$ *: For every permutation* $\theta : \{1, .., n\} \to \{1, ..., n\}$, $(\mathcal{C}, \phi\,[s_1, \cdots, s_n]) \equiv (\mathcal{C}, \theta^{-1} \circ \phi, ([s_{\theta(1)}, \cdots, s_{\theta(n)}]))$

We should modify the definition if we consider a typed calculus :

Definition 30

- *We define the judgment* $x_1 : \sigma_1, ...x_m : \sigma_m \vdash (\mathcal{C}, \phi, [M_1, \cdots, M_n]) : \tau$ *by :*
  - $\{x_1, \cdots, x_m\} \cap Dom(\phi) = \emptyset$
  - $x_1 : \sigma_1 \cdots x_n : \sigma_n, (y : \gamma_y)_{y \in Dom(\phi)} \vdash \mathcal{C} : \tau$
  - $Im(\phi) \subseteq \{1, \cdots, n\}$, *and* $\vdash M_{\phi(y)} : \gamma_y$ *for every* $y \in Dom(\phi)$
- *We define the set of pairs of context and tuples which are well formed :*

$$\mathcal{A}^\sigma = \{(\mathcal{C}, \phi, [M_1, \cdots, M_n]) \mid \emptyset \vdash (\mathcal{C}, \phi, [M_1, \cdots, M_n]) : \sigma\}$$

- *We define a notion of congruence for elements in* $\mathcal{A}$ *: For every permutation* $\sigma : \{1, .., n\} \to \{1, ..., n\}$, $(\mathcal{C}, \phi\,[M_1, \cdots, M_n]) \equiv (\mathcal{C}, \sigma^{-1} \circ \phi, [M_{\sigma(1)}, \cdots, M_{\sigma(n)}])$

$$\frac{\{(\mathcal{C},\phi,[s_1,\cdots,s_n])^1\}\xrightarrow{\epsilon}\mathscr{E}\qquad \begin{array}{c}\vdash(\mathcal{D},\psi,[t_1,\cdots,t_p])\\ \vdash(\mathcal{C},\phi,[s_1,\cdots,s_n])\end{array}\qquad \mathrm{Dom}(\phi)\cap\mathrm{Dom}(\psi)=\emptyset}{\mathscr{D}\dot{+}p\cdot\{(\mathcal{C}\mathcal{D},(\phi\,|^n\,\psi)\,[s_1,\cdots,s_n,t_1,\cdots t_p])^1\}\xrightarrow{\epsilon}\mathscr{D}+p\cdot\sum\mathscr{E}(\mathcal{E},\nu,[u_1,\cdots,u_q])\cdot(\mathcal{E}\mathcal{D},(\nu\,|^q\,\psi),[u_1,\cdots,u_q,t_1,\cdots t_p])}$$

$$\frac{\{(\mathcal{D},\psi,[t_1,\cdots,t_p])^1\}\xrightarrow{\epsilon}\mathscr{E}\qquad \begin{array}{c}\vdash(\mathcal{C},\phi,[s_1,\cdots,s_n]):\ \mathbf{val}\\ \vdash(\mathcal{D},\psi,[t_1,\cdots,t_p])\\ \mathrm{Dom}(\phi)\cap\mathrm{Dom}(\psi)=\emptyset\end{array}}{\mathscr{D}\dot{+}p\cdot\{(\mathcal{C}\mathcal{D},(\phi\,|^n\,\psi)\,[s_1,\cdots,s_n,t_1,\cdots t_p])^1\}\xrightarrow{\epsilon}\mathscr{D}+p\cdot\sum\mathscr{E}(\mathcal{E},\nu,[u_1,\cdots,u_q])\cdot(\mathcal{C}\mathcal{E},(\phi\,|^n\,\nu),[s_1,\cdots,s_n,u_1,\cdots u_q])}$$

$$\frac{f([M,s_1,\cdots,s_n])=\infty}{\mathscr{D}\dot{+}p\cdot\{x,(\{x\to 1\}\,|^1\,\phi),[M,s_1,\cdots,s_n]^1\}\xrightarrow{\epsilon}\mathscr{D}+p\cdot(x,(\{x\to 1\}\,|^1\,\phi),\left[\mathrm{focus}^1(M),s_2,\cdots,s_n\right])}$$

$$\frac{M\Rightarrow\mathscr{E}\qquad f(\left[\mathrm{focus}^i(M),s_1,\cdots,s_n\right])=i}{\mathscr{D}\dot{+}p\cdot\{x,(\{x\to 1\}\,|^1\,\phi),\left[\mathrm{focus}^i(M),s_1,\cdots,s_n\right]^1\}\xrightarrow{\epsilon}\mathscr{D}+p\cdot(x,(\{x\to 1\}\,|^1\,\phi),\left[\mathrm{focus}^i(\mathscr{E}),s_1,\cdots,s_n\right])}$$

$$\frac{f(\left[\mathrm{focus}^i(V),s_1,\cdots,s_n\right])=i}{\mathscr{D}\dot{+}p\cdot\{x,(\{x\to 1\}\,|^1\,\phi),\left[\mathrm{focus}^i(V),s_1,\cdots,s_n\right]^1\}\xrightarrow{\epsilon}\mathscr{D}+p\cdot(x,(\{x\to 1\}\,|^1\,\phi),\left[\widehat{V},s_1,\cdots,s_n\right])}$$

$$\frac{\vdash(\mathcal{C},\phi,[s_1,\cdots,s_q]):\ \mathbf{val}\qquad V=\mathcal{C}\{M_{\phi(z)}/z\}_{z\in FV(\mathcal{C})}}{\mathscr{D}\dot{+}p\cdot\{x\mathcal{C},(((x\to 1)\,|^1\,\phi)\,|^q\,\psi),\left[\widehat{\lambda y.L},s_1\cdots s_q,t_1\cdots,t_n\right]^1\}\xrightarrow{\epsilon}\mathscr{D}+p\cdot\{x,((x\to 1)\,|^1\,\psi),\left[\mathrm{focus}^1(L\{V/x\}),t_1,\cdots,t_n\right]^1\}}$$

$$\frac{\begin{array}{c}\vdash(\mathcal{C},\phi,[s_1,\cdots,s_n])\\ \vdash(\mathcal{D},\psi,[t_1,\cdots,t_p]):\ \mathbf{val}\\ \mathrm{Dom}(\phi)\cap\mathrm{Dom}(\psi)=\emptyset\end{array}}{\mathscr{D}\dot{+}p\cdot\{(\lambda x.\mathcal{C})\mathcal{D},(\phi\,|^n\,\psi),[s_1,\cdots s_n,t_1,\cdots,t_p]^1\}\xrightarrow{\epsilon}\mathscr{D}+p\cdot\{\mathcal{C}\{\mathcal{D}/x\},(\phi\,|^n\,\psi),[s_1,\cdots s_n,t_1,\cdots,t_p]^1\}}$$

$$\frac{}{p\cdot\{\lambda x.\mathcal{C},\phi,[s_1,\cdots,s_n]^1\}\xrightarrow{@V}p\cdot\{(\mathcal{C}\{V/x\},\phi,[s_1,\cdots,s_n])^1\}}$$

$$\frac{}{p\cdot\{(x,((x\to 1)\,|^1\,\phi),\left[\widehat{\lambda y.N},s_1,\cdots,s_n\right])^1\}\xrightarrow{@V}p\cdot\{(x,((x\to 1)\,|^1\,\phi),\left[\mathrm{focus}^1(N\{V/y\}),s_1,\cdots,s_n\right]^1\}}$$

$$\frac{\mathscr{D}_i\xrightarrow{@V}\mathscr{E}_i\qquad \forall i,\mathscr{D}_i\text{ value distribution}\qquad \mathscr{F}\text{ stopped distribution}}{\dot{\sum}_i\mathscr{D}_i\dot{+}\mathscr{F}\xrightarrow{@V}\sum_i\mathscr{E}_i}$$

$$\frac{}{\mathscr{D}\xRightarrow{\epsilon}\mathscr{D}}\qquad\frac{\mathscr{D}\xrightarrow{s}\mathscr{E}\qquad \mathscr{E}\xRightarrow{t}\mathscr{F}}{\mathscr{D}\xRightarrow{s\cdot t}\mathscr{F}}$$

Figure 10: small-step trace relation on distributions over $\mathcal{A}$ (without pairs)

In the following, we consider equivalence class of $\equiv$. It corresponds to reorder elements of the tuple, and to modify the function ϕ in order to have still the same mapping from the free variables of $\mathcal{C}$.

We define a small-step semantics on elements of $\mathcal{A}$.

Please observe that the rules would be exactly the same for an strictly linear (that is, not affine) calculus. The only thing to change would be the definition of : $\Gamma\vdash(\mathcal{C},\phi,K)$.

We need a definition of $\mathcal{A}$ taking into account possible other free variables We need to add rules specific for the language with pairs :

Please observe that there is two different non-determinism in the rules : the choice of the part of the distribution which is going to be reduced, and the way the tuple is divided (for the affine

$$\frac{\begin{array}{c}\vdash (\mathcal{C},\phi,[M_1,\cdots,M_n]):\sigma\otimes\tau \\ x:\sigma,y:\tau\vdash(\mathcal{D},\psi,[N_1,\cdots,N_p]):\gamma \\ \mathrm{Dom}(\phi)\cap(\mathrm{Dom}(\psi)\cup\{x,y\})=\emptyset\end{array}\qquad \{(\mathcal{C},\phi,[M_1,\cdots,M_n])^1\}\xrightarrow{\epsilon}\mathscr{E}}{\begin{array}{c}\mathscr{D}\dotplus p\cdot\{(\texttt{let }\langle x,y\rangle=\mathcal{C}\texttt{ in }\mathcal{D},(\phi\,|^n\,\psi)\,[M_1,\cdots,M_n,N_1,\cdots N_p])^1\}\xrightarrow{\epsilon} \\ \mathscr{D}+p\cdot\sum\mathscr{E}(\mathcal{E},\nu,[L_1,\cdots,L_q])\cdot(\texttt{let }\langle x,y\rangle=\mathcal{E}\texttt{ in }\mathcal{D},(\nu\,|^q\,\psi),[L_1,\cdots,L_q,N_1,\cdots N_p])\end{array}}$$

$$\frac{\{(\mathcal{C},\phi,[M_1,\cdots,M_n])^1\}\xrightarrow{\epsilon}\mathscr{E}\qquad \begin{array}{c}\mathrm{Dom}(\phi),\mathrm{Dom}(\psi),\{x,y\},\mathrm{Dom}(\nu)\\ \text{disjoints sets}\end{array}\qquad \begin{array}{c}\vdash\mathcal{C},\phi,[M_1,\cdots,M_n]:\sigma\\ \vdash\mathcal{D},\psi,[N_1,\cdots,N_m]:\tau\\ x:\sigma,y:\tau\vdash(\mathcal{E},\nu,[L_1,\cdots,L_l]:\gamma\end{array}}{\begin{array}{c}\mathscr{D}\dotplus p\cdot\{(\texttt{let }\langle x,y\rangle=\langle\mathcal{C},\mathcal{D}\rangle\texttt{ in }\mathcal{E},(\phi\,|^n\,(\psi\,|^m\,\nu)),[M_1,\cdots,M_n,N_1,\cdots N_m,L_1\cdots L_l])^1\}\xrightarrow{\epsilon}\\ \mathscr{D}+p\cdot\sum\mathscr{E}(\mathcal{F},\eta,[K_1,\cdots,K_q])\cdot(\texttt{let }\langle x,y\rangle=\langle\mathcal{F},\mathcal{D}\rangle\texttt{ in }\mathcal{E},(\eta\,|^q\,(\psi\,|^m\,\nu)),[K_1,\cdots,K_q,N_1,\cdots N_m,L_1,\cdots L_l])\end{array}}$$

$$\frac{\{(\mathcal{D},\psi,[N_1,\cdots,N_m])^1\}\xrightarrow{\epsilon}\mathscr{E}\qquad \mathcal{C}\{M_{\phi(x)}/x\}_{x\in\mathrm{Dom}(\phi)}\text{ value}\qquad \begin{array}{c}\vdash\mathcal{C},\phi,[M_1,\cdots,M_n]:\sigma\\ \vdash\mathcal{D},\psi,[N_1,\cdots,N_m]:\tau\\ x:\sigma,y:\tau\vdash(\mathcal{E},\nu,[L_1,\cdots,L_l]:\gamma\\ \mathrm{Dom}(\phi),\mathrm{Dom}(\psi),\{x,y\},\mathrm{Dom}(\nu)\text{ disjoints sets}\end{array}}{\begin{array}{c}\mathscr{D}\dotplus p\cdot\{(\texttt{let }\langle x,y\rangle=\langle\mathcal{C},\mathcal{D}\rangle\texttt{ in }\mathcal{E},(\phi\,|^n\,(\psi\,|^m\,\nu))\,[M_1,\cdots,M_n,N_1,\cdots N_m,L_1\cdots L_l])^1\}\xrightarrow{\epsilon}\\ \mathscr{D}+p\cdot\sum\mathscr{E}(\mathcal{F},\eta,[K_1,\cdots,K_q])\cdot\{(\texttt{let }\langle x,y\rangle=\langle\mathcal{F},\mathcal{D}\rangle\texttt{ in }\mathcal{E},(\phi\,|^n\,(\eta\,|^q\,\nu)),[M_1,\cdots,M_n,K_1,\cdots K_q,L_1,\cdots L_l])^1\}\end{array}}$$

$$\frac{\begin{array}{c}\mathcal{C}\{M_{\phi(z)}/z\}_{z\in\mathrm{Dom}(\phi)}\text{ value}\\ \mathcal{D}\{N_{\psi(z)}/z\}_{z\in\mathrm{Dom}(\psi)}\text{ value}\end{array}\qquad \begin{array}{c}\vdash\mathcal{C},\phi,[M_1,\cdots,M_n]:\sigma\\ \vdash\mathcal{D},\psi,[N_1,\cdots,N_m]:\tau\\ x:\sigma,y:\tau\vdash(\mathcal{E},\nu,[L_1,\cdots,L_l]:\gamma\\ \mathrm{Dom}(\phi),\mathrm{Dom}(\psi),\{x,y\},\mathrm{Dom}(\nu)\text{ disjoints sets}\end{array}}{\begin{array}{c}\mathscr{D}\dotplus p\cdot\{(\texttt{let }\langle x,y\rangle=\langle\mathcal{C},\mathcal{D}\rangle\texttt{ in }\mathcal{E},(\phi\,|^n\,(\psi\,|^m\,\nu))\,[M_1,\cdots,M_n,N_1,\cdots N_m,L_1\cdots L_l])^1\}\xrightarrow{\epsilon}\\ \mathscr{D}+p\cdot\{(\mathcal{E}\{\mathcal{C}/x\}\{\mathcal{D}/y\},(\phi\,|^n\,(\psi\,|^m\,\nu)),[M_1,\cdots,M_n,N_1,\cdots N_m,L_1,\cdots L_l])^1\}\end{array}}$$

$$\frac{M_1=\langle N,L\rangle\qquad \begin{array}{c}x:\sigma,y:\tau\vdash(\mathcal{C},\phi,[M_2,\cdots,M_n]):\gamma\\ z\notin\mathrm{Dom}(\phi)\end{array}}{\begin{array}{c}\mathscr{D}\dotplus p\cdot\{(\texttt{let }\langle x,y\rangle=z\texttt{ in }\mathcal{C},(\{z\to1\}\,|^1\,\phi),[M_1,\cdots,M_n])^1\}\xrightarrow{\epsilon}\\ \mathscr{D}+p\cdot\{(\mathcal{C},(\{x\to1,y\to2\}\,|^2\,\phi),[N,L,M_2,\cdots M_n])^1\}\end{array}}$$

$$\frac{M_1=\langle N,L\rangle}{\begin{array}{c}p\cdot\{(z,(\{z\to1\}\,|^1\,\phi),[M_1,\cdots,M_n])^1\}\xrightarrow{\langle,\rangle\cdot\mathcal{C}}\\ p\cdot\{(\mathcal{C},(\{x\to1,y\to2\}\,|^2\,\phi),[N,L,M_2,\cdots M_n])^1\}\end{array}}$$

$$\frac{M_1=\langle N,L\rangle}{\begin{array}{c}p\cdot\{(\langle\mathcal{D},\mathcal{E}\rangle,\phi,[M_1,\cdots,M_n])^1\}\xrightarrow{\langle,\rangle\cdot\mathcal{C}}\\ p\cdot\{(\mathcal{C}\{\mathcal{D}/x\}\{\mathcal{E}/y\},\phi,[M_1,\cdots,M_n])^1\}\end{array}}$$

Figure 11: small-step trace relation on distributions over $\mathcal{A}$ for pairs

case). The second one is not really meaningful, since we have the following lemma :

Lemma 34 *Suppose that :* $\vdash (\mathcal{C}, \phi, K)$*, and let be* $\mathscr{D}$*,* $\mathscr{E}$ *such that* $\{(\mathcal{C}, \phi, K)^1\} \xrightarrow{a} \mathscr{D}$ *and* $\{(\mathcal{C}, \phi, K)^1\} \xrightarrow{a} \mathscr{E}$*. Then* $\mathscr{D} \equiv \mathscr{E}$*.*

Proof. Let be $\vdash (\mathcal{C}, \phi, K)$.
We are first going to show the following result : Suppose that $K = [M_1, \cdots, M_n, \cdots, M_q]$, et that $\phi(FV(\mathcal{C})) \subseteq \{1, \cdots, n\}$. Then let be $\mathscr{D}$ such that : $(\mathcal{C}, \phi, K) \xrightarrow{a} \mathscr{D}$. Then there exist $\mathscr{E}$ such that : $(\mathcal{C}, (\phi) \restriction_{FV(\mathcal{C})}, [M_1, \cdots, M_n] \xrightarrow{a} \mathscr{E}$, and $\mathscr{D} = \sum \mathscr{E}(\mathcal{D}, \psi, [N_1, \cdots, N_p]) \cdot \{(\mathcal{D}, \nu, [N_1, \cdots N_p, M_{n+1}, \cdots, M_n])^1\}$, with $\nu(x) = \psi(x)$ if $x \in FV(\mathcal{C})$, and $\nu(x) = p - n + \phi(x)$ otherwise. We show that by induction on the derivation of $(\mathcal{C}, \phi, K) \xrightarrow{a} \mathscr{D}$.
Then it is sufficient to remark that, if the free variables of $\mathcal{C}$ correspond exactly to the terms in the tuple, there is only one possible rule that can be applied. $\square$

Definition 31 *Let be* $\mathscr{D}$ *a distribution over* $\mathcal{A}$*. We define* $\mathbf{F}(\mathscr{D})$ *a distribution over closed terms by :* $\mathbf{F}(\mathscr{D}) = \sum \mathscr{D}(\mathcal{C}, \phi, [M_1, \cdots, M_n]) \cdot \{\mathcal{C}\{M_{\phi(x)}/x\}_{x \in Dom(\phi)}{}^1\}$

We would like to know that, if a distribution on terms can do a trace, then the correponding distribution where we split contexts and terms filling them can do the same trace. Unfortunately, we need to be more precis on how we split the distribution, and especially on what focus we can have on the components of the tuple. (For example, $\big(x, (x \to 1), \big[M, \text{focus}^1(N)\big]\big) \not\to$, since it is not possible to evaluate M before having evaluated N.) So we define a notion of coherent tuples in $\mathcal{S}^{\text{focus}}$ for a given context, where the idea is : This context could have triggered the evaluation on the terms which are under focus :

Lemma 35 *Let be* $\mathscr{D}$ *a distribution over* $\mathcal{A}$*, and* s *a trace such that :* $\mathbf{F}(\mathscr{D}) \overset{s}{\Rightarrow} \mathscr{E}$*. Then there exists* $\mathscr{F}$ *such that* $\mathscr{D} \overset{s}{\Rightarrow} \mathscr{F}$*, and* $\mathbf{F}(\mathscr{F}) = \mathscr{E}$*.*

The rules of the trace semantics for elements in $\mathcal{A}$ are designed to match the one for trace semantics for terms. More precisely, it means that :

Lemma 36 *Let be* $\vdash (\mathcal{C}, \phi, [M_1, \cdots, M_n])$*, and let be* $\mathscr{D}$ *such that :* $\{(\mathcal{C}, \phi, [M_1, \cdots, M_n])^1\} \xrightarrow{\epsilon} \mathscr{D}$*. Then* $\{\mathcal{C}\{M_{\phi x}/x\}_{x \in FV(\mathcal{C})}{}^1\} \xrightarrow{\epsilon} \mathbf{F}(\mathscr{D})$

Proof. The proof is by case analysis of the derivation of $\{(\mathcal{C}, \phi, [M_1, \cdots, M_n])^1\} \xrightarrow{\epsilon} \mathscr{D}$ $\square$

Lemma 37 *Let be* $\mathscr{D}$ *a distribution over* $\mathcal{A}$*, and* s *a trace. Suppose that* $\mathscr{D} \overset{s}{\Rightarrow} \mathscr{E}$*. Then* $\mathbf{F}(\mathscr{D}) \overset{s}{\Rightarrow} \mathbf{F}(\mathscr{E}^\star)$

Proof. It is in fact sufficient to show :

- If $\mathscr{D} \xrightarrow{\epsilon} \mathscr{E}$, then there exist $\mathscr{F}$, such that $\mathscr{E} \overset{\epsilon}{\Rightarrow} \mathscr{F}$, and $\mathbf{F}(\mathscr{D}) \xrightarrow{\epsilon} \mathbf{F}(\mathscr{F})$. No matter the last rule used in the derivation of $\mathscr{D} \xrightarrow{a} \mathscr{E}$, it is of the form : $\mathscr{D} = \mathscr{G} \dotplus p \cdot (\mathcal{C}, \phi, K) \xrightarrow{\epsilon} \mathscr{G} + p \cdot \mathscr{H}$ with $\{(\mathcal{C}, \phi, K)^1\} \xrightarrow{\epsilon} \mathscr{H}$. Now we have to consider all the possible $(\mathcal{D}, \psi, H) \in \mathsf{S}(\mathscr{G})$ such that $\mathcal{D}\{H_{\psi(x)}/x\}_{x \in FV(\mathcal{D})} = \mathcal{C}\{K_{\psi(x)}/x\}_{x \in FV(\mathcal{C})}$
- and : if $\mathscr{D} \xrightarrow{a} \mathscr{E}$, then $\mathbf{F}(\mathscr{D}) \xrightarrow{a} \mathbf{F}(\mathscr{E})$

$\square$

6.1.4 Link beetween trace semantics on terms and trace semantics on $\mathcal{A}$.

Definition 32 *Let be $\mathscr{D}$ and $\mathscr{E}$ two distributions over $\mathcal{A}$. For $\varepsilon \geq 0$, we say that $\mathscr{D}$ and $\mathscr{E}$ are ε-related if : there exist $p_i, ..., p_m$ positive reals, and $\mathcal{D}_1, ..., \mathcal{D}_d$ distincts contexts, and $\mathscr{F}_1, ..., \mathscr{F}_d$, $\mathscr{G}_1, ..., \mathscr{G}_d$ distributions on tuples such that :*

$$\mathscr{D} = \sum_j p_j \cdot (\mathcal{D}_j, \mathscr{F}_j)$$
$$\mathscr{E} = \sum_j p_j \cdot (\mathcal{D}_j, \mathscr{G}_j)$$
$$\delta^{mul}(\mathscr{F}_j, \mathscr{G}_j) \leq \varepsilon$$

Lemma 38 *The relation $\cdot \overset{\epsilon}{\to} \cdot$ on distributions over $\mathcal{A}$ is strongly normalizing.*

Lemma 39 *Let be $\mathscr{D}$, $\mathscr{E}$ two ε-related distributions. Then $\mathscr{D}^\star$ and $\mathscr{E}^\star$ are related.*

Lemma 40 *Let be $\mathscr{D}$, $\mathscr{E}$ two ε-related distribution. Let be $\mathscr{F}^\star$, and $\mathscr{G}^\star$ in normal form such that : $\mathscr{D} \overset{s}{\Rightarrow} \mathscr{F}^\star$, and $\mathscr{E} \overset{s}{\Rightarrow} \mathscr{G}^\star$. Then $\mathscr{F}$ and $\mathscr{G}$ are ε-related*

Theorem 7 is deduced of Lemma 40 in a similar way as for the trace distance.

□

Theorem 7 can be read as a non-expansiveness result: if we have a system $\mathcal{E}$, playing the role of the environment, and which is prepared to interact with n components, and moreover we have two tuples K and H of length n, then the tuple distance between K and H gives us an upper bound on the trace distance between the system composed of $\mathcal{E}$ interacting with K, and the system composed of $\mathcal{E}$ interacting with H.

We can now see that δ^{mul} coincides with the context metric: one inequality comes from Theorem 7, the other comes from the fact that any trace s over $\mathscr{A}^{\Lambda}_{\mathrm{mul}}$ and designed to start from a single value, can be simulated by a context.

Theorem 8 *On programs, $\delta^{mul} = \delta^{ctx}$*

Proof.
- We apply Theorem 7 to $\lambda x.M$ and $\lambda x.N$, which are values, and the context $\mathcal{C} = [\cdot]$:

$$\begin{aligned}\delta^{\mathrm{tr}}(M, N) = \delta^{\mathrm{tr}}(\lambda x.M, \lambda x.N) & \\ &\leq \delta^{\mathrm{mul}}([\lambda x.M], \lambda x.N) \\ &= \delta^{\mathrm{mul}}(M, N)\end{aligned}$$

- Let be s a trace in the LMC $\mathscr{M}^{\Lambda}_{\mathrm{mul}}$ which starts from a single value. Then we can find a context that simulate this trace.

□

6.2 Examples

The tuple distance, that we have just proved to be fully-abstract, can be seen as yet another presentation of the context distance. But there is much more: it allows to evaluate the distance between concrete programs, even when the latter contains pairs, in a relatively easy way. In this section, we will give two examples.

6.2.1 A Simple Example

Consider the terms M and N defined in Example 3. We can prove that $\delta^{\mathrm{mul}}(M,N) = \frac{3}{4}$. We are first going to show that $\delta^{\mathrm{mul}}(M,N) \geq \frac{3}{4}$. In order to show that, we are going to present a particular trace s such that $|Pr^{\mathrm{mul}}([M], s) - Pr^{\mathrm{mul}}([N], s)| = \frac{3}{4}$. More precisely, we take $s = \mathrm{unfold}^1 \cdot @(\emptyset, I)^1 \cdot @(\emptyset, I)^2$: it corresponds to first separating the two components of the pair, and then passing I as an argument to the first and to the second component. The relevant fragment of $\mathscr{M}^{\Lambda}_{\mathrm{mul}}$ can be found in Figure 12. In particular, we can see that $Pr([M], s) = 1$, and $Pr([N], s) = \frac{1}{4}$. Now we want to show the reverse inequality, namely that $\delta^{\mathrm{mul}}(M,N) \leq \frac{3}{4}$. For

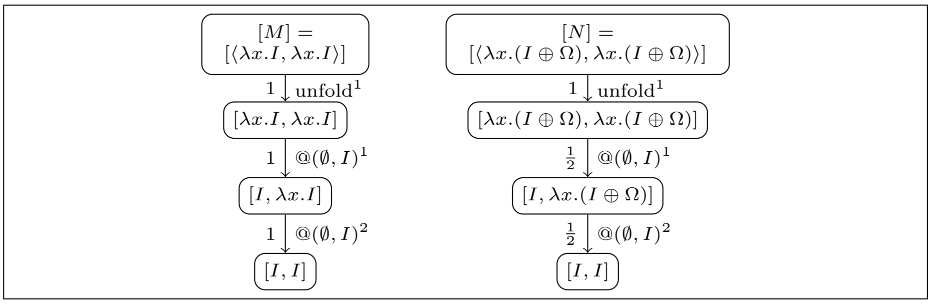

Figure 12: The relevant fragment of the tuple LMC

that, we are going to use the alternative characterisation of trace distance : it is sufficient to find a $\frac{3}{4}$-bisimulation R on the LTS of distributions such that $(\{[M]^1\}), (\{[N]^1\}) \in R$

6.2.2 A More Complicated Example

Please remember the example we presented in Section 2. We note $\{u_n\}_{n\in\mathbb{N}}$ the sequence defined as: $u_n = \prod_{1\leq i\leq n}(1 - \frac{1}{2^i})$. Please observe that the sequence $(u_n)_{n\in\mathbb{N}}$ has a limit strictly between 0 and 1.

Lemma 41 *La suite $(u_n)_{n\in\mathbb{N}}$ has a limit l, and $\frac{1}{2} > l > 0$*

Proof.

- u_n is a decreasing and bounded sequence : it has a limit.
- $l > u_1 = \frac{1}{2}$
- We consider the sequence : $v_n = \log u_n = \sum_{1\leq i\leq n} \log(1 - (\frac{1}{2})^i)$. We pose $w_n = \log 1 - (\frac{1}{2})^i$. Then we consider

$$|\frac{w_n}{w_{n+1}}| = |\frac{\log(1-(\frac{1}{2})^n)}{\log(1-(\frac{1}{2})^{n+1})}| \rightarrow_{n\rightarrow\infty} \frac{1}{2}$$

D'Alembert's theorem for infinite sum implies that the serie is convergent and has a finite limit.

□

Theorem 9 *For every $n \in \mathbb{N}$, $\delta^{mul}(M_n, N_n) = 1 - u_n$.*

Proof. We first show that $\delta^{\mathrm{mul}}(M_n, N_n) \geq 1 - u_n$. As in the previous example, we do that by finding, for each $n \in \mathbb{N}$, a trace s_n such that $|Pr([M_n], s_n) - Pr([N_n], s_n)| = 1 - u_n$. We define the sequence $(s_n)_{n\in\mathbb{N}}$ inductively as follows:

$$s_0 = \epsilon \qquad s_{n+1} = \mathrm{unfold}^1 \cdot @(\emptyset, I)^1 \cdot s_n$$

s_0 is the trace which always succeeds, whatever the starting state is. s_{n+1} corresponds to separating the two components of the pair which is in first position in the tuple, then passing the identity

$$Pr([M_0], s_0) = 1 \qquad Pr([N_0], s_0) = 1$$
$$Pr([M_{n+1}], s_{n+1}) = 1 \cdot Pr([M_n], s_n)$$
$$Pr([N_{n+1}], s_{n+1}) = (1 - \frac{1}{2^{n+1}}) \cdot Pr([N_n], s_n)$$

Figure 13: Recursive equations verified by s_n

as an argument to the first component of this pair, and then executing s_n. For this sequence of traces, the recursive equations of Figure 13 are verified (the proof can be found in [5]). We can see by solving these equations that for every $n \in \mathbb{N}$, $Pr(M_n, s_n) = 1$ and $Pr(N_n, s_n) = u_n$. As a direct consequence, we obtain the result. We want now to show that $\delta^{\mathrm{mul}}(M_n, N_n) \leq 1 - u_n$. To do that, we need to establish that there doesn't exist a trace t such that $|Pr([M_n], t) - Pr([N_n], t)| > 1 - u_n$. We're in fact going to show something stronger: for every $n \in \mathbb{N}$, we're going to define a set A_n of pairs of tuple, which contains the pair $([M]_n, [N_n])$, and such that for every $(K, H) \in A_n$, for every trace t, $|Pr(K, t) - Pr(H, t)| \leq 1 - u_n$. Intuitively, the idea behind the sequence $\{A_n\}_{n \in \mathbb{N}}$ is the following: if we start from $[M_n]$, do a trace of even length, and end up in a tuple K with a non-zero probability, and if when we do *the same trace* starting from $[N_n]$ ending up in the tuple H, then the pair of tuple (K, H) is in one of the A_j, with j smaller than n.

Definition 33 *Let be $n \in \mathbb{N}$. Let A_n be the set of (K, H) such that: there exist $m \in \mathbb{N}$, and $k_i \geq n + 1$ (for $1 \leq i \leq m$), where:*

$$K = [M_n, [\lambda x.\Omega]^m];$$
$$H = \left[N_n, [\lambda x.\Omega \oplus^{\frac{1}{2^{k_i}}} I]_{1 \leq i \leq m}\right].$$

We want now to give an upper bound to the separation between K and H any trace can induce, if $(K, H) \in A_n$.

Lemma 42 *For every $n \in \mathbb{N}$, for every $(K, H) \in A_n$, we can partition the set of traces as:*

$$\mathcal{T}r = \{s \mid Pr(K, s) = 0 \text{ and } Pr(H, s) \leq \frac{1}{2}\}$$
$$\bigcup \{s \mid Pr(K, s) = 1 \text{ and } Pr(H, s) \geq u_n\}.$$

Proof. Let $s \in \mathcal{T}r$. We are going to show by induction on the length of s that for every $n \in \mathbb{N}$, for every $(K, H) \in A_n$, either $Pr(K, s) = 0$ and $Pr(H, s) \leq \frac{1}{2}$, or $Pr(K, s) = 1$ and $Pr(H, s) \geq u_n$.

- If $s = \epsilon$, then for every $n \in \mathbb{N}$ and $(K, H) \in A_n$ $Pr(K, s) = Pr(H, s) = 1$, and we are in the second case.
- If the length of s is $l > 0$. Let be $n \in \mathbb{N}$, and $(K, H) \in A_n$. Then we can write:

$$K = [M_n, [\lambda x.\Omega]^m];$$
$$H = \left[N_n, [\lambda x.\Omega \oplus^{\frac{1}{2^{k_i}}} I]_{1 \leq i \leq m}\right] \qquad \text{with } k_i \geq n + 1.$$

  We are now going to distinguish the cases depending on which element of the tuple is applied the first action of the trace.
  - If the first action is not applied to the first element of the tuple, then $s = @(\Gamma, \mathcal{C})^j \cdot t$, with $j > 1$: Then $Pr(K, s) = 0$, and $Pr(H, s) \leq \frac{1}{2^{k_j}} \leq \frac{1}{2}$: we are in the first case.
  - If the first action is applied to the first element of the tuple : Then we can see that $s = \mathrm{unfold}^1 \cdot t$ (since the first element of the tuple is actually a pair, the only action that can be applied to it is the unfold action).

- First, let's consider the case where $n = 0$. Please remember that by definition we have that $M_0 = N_0 = \langle \lambda x.\Omega, \lambda x.\Omega \rangle$. Observe that:

$$K_1 = \left[[\lambda x.\Omega]^{m+2} \right];$$
$$H_1 = \left[\lambda x.\Omega, \lambda x.\Omega, [\lambda x.\Omega \oplus^{\frac{1}{2^{k_i}}} I]_{1 \leq i \leq m} \right].$$

  With these notations, we can see that $Pr(K, s) = Pr(K_1, t)$, and $Pr(H, s) = Pr(H_1, t)$. If $t = \epsilon$, these two expressions are equal to 1, and we are in the second case. Otherwise, $Pr(K_1, t) = 0$, and $Pr(H_1, t) \leq \frac{1}{2}$, and we are in the first case.
- Now let's consider the case where $n \geq 1$. Please remember that:

$$M_n = \langle \lambda x.M_{n-1}, \lambda x.\Omega \rangle;$$
$$N_n = \langle \lambda x.(N_{n-1} \oplus^{\frac{1}{2^n}} \Omega), \lambda x.(\Omega \oplus^{\frac{1}{2^n}} I) \rangle.$$

  Then we'll note:

$$K_2 = [\lambda x.M_{n-1}, \lambda x.\Omega, [\lambda x.\Omega]^m];$$
$$H_2 = \left[\lambda x.(N_{n-1} \oplus^{\frac{1}{2^n}} \Omega), \lambda x.(\Omega \oplus^{\frac{1}{2^n}} I), [\lambda x.(\Omega \oplus^{\frac{1}{2^{k_i}}} I)]_{1 \leq i \leq m} \right].$$

  With these notations, we can see that $Pr(K, s) = Pr(K_2, t)$, and $Pr(H, s) = Pr(H_2, t)$. Now we have to consider the different possible form of the trace t :
  - if $t = \epsilon$, $Pr(K, s) = Pr(H, t) = 1$.
  - if $t = @(\Gamma, \mathcal{C})^j \cdot u$, with $j > 1$, we have $Pr(K_1, u) = 0$ and $Pr(K_2, u) \leq \frac{1}{2^n} \leq \frac{1}{2}$, and we are in the first case.
  - if $t = @(\Gamma, \mathcal{C})^1 \cdot u$. Please remember the semantics of this action in the Markov Chain : If we start from K_2, with probability 1 we go to a state K_3 of the form :

$$K_3 = \left[M_{n-1}, [\lambda x.\Omega]^l \right]$$

    with $l \leq m$. If we start from H_2, with probability $(1 - \frac{1}{2^n})$ we go in a state H_3 of the form :

$$H_3 = \left[N_{n-1}, [\lambda x.(\Omega \oplus^{\frac{1}{2^{k_i}}} I)]_{1 \leq i \leq l} \right]$$

    with $k_i \geq n$. Now we can see that :

$$Pr(K, s) = Pr(K_3, u);$$
$$Pr(H, s) = (1 - \frac{1}{2^n}) \cdot Pr(H_3, u).$$

    Moreover, please observe that $(K_3, H_3) \in A_{n-1}$, so we can apply the induction hypothesis (since the length of u is strictly smaller that the length of s). Now, there are two possible cases :
    - $Pr(K_3, u) = 0$, and $Pr(H_3, u) \leq \frac{1}{2}$. Then we can see that the result holds, since it implies that : $Pr(K, s) = 0$ and $Pr(H, s) \leq (1 - \frac{1}{2^n}) \cdot \frac{1}{2} \leq \frac{1}{2}$.
    - $Pr(K_3, u) = 1$, and $Pr(H_3, u) \geq u_{n-1}$. Then we can see that the result holds, since it implies that : $Pr(K, s) = 1$ and $Pr(H, s) \geq (1 - \frac{1}{2^n}) \cdot u_{n-1} = u_n$.

□

The result we're seeking to show is a direct consequence of Lemma 42: we can see easily that for any trace s, if $(K, H) \in A_n$, the separation that s can induce is smaller than $1 - u_n$. Indeed, let be $s \in \mathcal{T}r$. Since $([M_n], [N_n]) \in A_n$, we can see that :

- Or the trace s is in the first set of the partition given by Lemma 42, and $|Pr(M_n, s) - Pr(N_n, t)| \leq \frac{1}{2} \leq 1 - u_n$.
- Or the trace s is in the second set of this partition, and then $|Pr(M_n, s) - Pr(N_n, t)| \leq 1 - u_n$.

□

6.3 On Tuples and Copying

The tuple distance naturally suggests a way to handle λ-calculi in which copying is indeed allowed. Although the details are clearly outside the scope of this paper, we anyway want to give some hints about why this is the case.

What makes the trace and behavioural distances unsound in presence of copying is their inability to capture an environment which can access the program at hand *more than once*. In our view, however, the problem does not come from the way those distances are defined in the abstract, but rather in the way *the underlying LMC* reflects the operational semantics of the calculus at hand. In a sense, it is in the responsibility of the LMC to guarantee that the environment can access terms multiple times. The LMC $\mathscr{M}^{\Lambda}$ we introduced in this paper (which is close to the ones from the literature [4, 6, 8]), as an example, is not adequate.

Suppose, however, to extend $\mathscr{M}^{\Lambda}_{\mathrm{mul}}$ to an LMC for a λ-calculus in the style of Wadler's linear λ-calculus [29]: there, the grammar of terms includes a construct $!M$ whose purpose is marking those subterms which can indeed be duplicated. The actions the environment can perform on a term in the form $!M$ simply reflects the above: the environment can create a *new copy* of $!M$, *but also keeps* the possibility to access $!M$ in the future. One immediately realises that tuples are indeed the right way to model the access to both $!M$ and M.

7 Conclusions

We have initiated the study of metrics in higher-order languages, starting with the relatively easy case of affine λ-terms, where copying capabilities are simply not available. We showed that three different notions of distance are sound (and sometime fully-abstract) for the context distance, the natural generalisation of Morris' observational equivalence. One of them, the tuple distance, reflects the inherently monoidal structure of the underlying calculus, this way allowing to solve some nontrivial distance problems.

We are actively working on extending the results described here to the non-affine case, which for various reasons turns out to be more difficult, as discussed in Section 2. We are in particular quite optimistic about the possibility of generalising the tuple distance to a metric reflecting copying. The real challenge, however, consists in handling the case in which copying is indeed available, but the number of copies of a given term the environment can have access to is somehow bounded, maybe polynomially on the value on an security parameter. That would indeed be a way to get closer to computational indistinguishability, a central notion in modern cryptography.

References

[1] S. Abramsky. The Lazy λ-Calculus. In D. Turner, editor, *Research Topics in Functional Programming*, pages 65–117. Addison Wesley, 1990.

[2] A. Asperti and L. Roversi. Intuitionistic light affine logic. *ACM Trans. Comput. Log.*, 3(1):137–175, 2002.

[3] A. Bizjak and L. Birkedal. Step-indexed logical relations for probability. To appear in FoSSaCS 2015.

[4] R. Crubillé and U. Dal Lago. On probabilistic applicative bisimulation and call-by-value λ-calculi. In *ESOP*, pages 209–228, 2014.

[5] R. Crubillé and U. Dal Lago. Metric reasoning about λ-terms: the affine case (long version). Available at `http://eternal.cs.unibo.it/mrltac.pdf`, 2015.

[6] U. Dal Lago, D. Sangiorgi, and M. Alberti. On coinductive equivalences for higher-order probabilistic functional programs. In *POPL*, pages 297–308, 2014.

[7] V. Danos and T. Ehrhard. Probabilistic coherence spaces as a model of higher-order probabilistic computation. *Inf. Comput.*, 209(6):966–991, 2011.

[8] Y. Deng and Y. Zhang. Program equivalence in linear contexts. To appear in Theoretical Computer Science, 2014.

[9] J. Desharnais, V. Gupta, R. Jagadeesan, and P. Panangaden. Metrics for labeled markov systems. In *CONCUR*, 1999.

[10] J. Desharnais, R. Jagadeesan, V. Gupta, and P. Panangaden. The metric analogue of weak bisimulation for probabilistic processes. In *LICS*, pages 413–422, 2002.

[11] T. Ehrhard, C. Tasson, and M. Pagani. Probabilistic coherence spaces are fully abstract for probabilistic PCF. In *POPL*, pages 309–320, 2014.

[12] D. Gebler and S. Tini. Fixed-point characterization of compositionality properties of probabilistic processes combinators. In *EXPRESS-SOS*, pages 63–78, 2014.

[13] O. Goldreich. *Modern Cryptography, Probabilistic Proofs and Pseudorandomness*, volume 17 of *Algorithms and Combinatorics*. Springer, 1998.

[14] S. Goldwasser and S. Micali. Probabilistic encryption. *J. Comput. Syst. Sci.*, 28(2):270–299, 1984.

[15] N. D. Goodman, V. K. Mansinghka, D. M. Roy, K. Bonawitz, and J. B. Tenenbaum. Church: a language for generative models. In *UAI 2008*, pages 220–229, 2008.

[16] D. J. Howe. Proving congruence of bisimulation in functional programming languages. *Inf. Comput.*, 124(2):103–112, 1996.

[17] C. Jones and G. D. Plotkin. A probabilistic powerdomain of evaluations. In *LICS*, pages 186–195, 1989.

[18] L. V. Kantorovich. On the transfer of masses. In *Dokl. Akad. Nauk. SSSR*, volume 37, pages 227–229, 1942.

[19] S. B. Lassen. Relational reasoning about contexts. In *Higher Order Operational Techniques in Semantics, Publications of the Newton Institute*, pages 91–135. Cambridge University Press, 1998.

[20] H. G. Mairson. Linear lambda calculus and ptime-completeness. *J. Funct. Program.*, 14(6):623–633, 2004.

[21] C. D. Manning and H. Schütze. *Foundations of statistical natural language processing*, volume 999. MIT Press, 1999.

[22] J. C. Mitchell, A. Ramanathan, A. Scedrov, and V. Teague. A probabilistic polynomial-time process calculus for the analysis of cryptographic protocols. *Theor. Comput. Sci.*, 353(1-3):118–164, 2006.

[23] S. Park, F. Pfenning, and S. Thrun. A probabilistic language based on sampling functions. *ACM Trans. Program. Lang. Syst.*, 31(1), 2008.

[24] J. Pearl. *Probabilistic reasoning in intelligent systems: networks of plausible inference.* Morgan Kaufmann, 1988.

[25] A. M. Pitts. Operationally-based theories of program equivalence. In *Semantics and Logics of Computation*, pages 241–298. Cambridge University Press, 1997.

[26] N. Ramsey and A. Pfeffer. Stochastic lambda calculus and monads of probability distributions. In *POPL*, pages 154–165, 2002.

[27] S. Thrun. Robotic mapping: A survey. *Exploring artificial intelligence in the new millennium*, pages 1–35, 2002.

[28] F. van Breugel and J. Worrell. A behavioural pseudometric for probabilistic transition systems. *Theor. Comput. Sci.*, 331(1):115–142, 2005.

[29] P. Wadler. A syntax for linear logic. In *MFPS*, number 802 in LNCS, pages 513–529, 1993.