Probabilistic lambda-theories

Thomas Leventis

AIX-MARSEILLE UNIVERSITE

ECOLE DOCTORALE 184

UFR SCIENCES

INSTITUT DE MATHEMATIQUES DE MARSEILLE/EQUIPE LDP

Thèse présentée pour obtenir le grade universitaire de docteur

Discipline : Mathématiques et Informatique
Spécialité : Mathématiques

Thomas LEVENTIS

Probabilistic λ-theories

Rapporteurs :

Antonino SALIBRA	Université de Venise Ca'Foscari
Sam STATON	Université d'Oxford

Soutenue le 08/12/2016 devant le jury :

Antonino SALIBRA	Université de Venise Ca'Foscari	Rapporteur
Delia KESNER	Université Paris Diderot	Examinateur
Ugo DAL LAGO	Université de Bologne	Examinateur
Michele PAGANI	Université Paris Diderot	Examinateur
Lionel VAUX	Aix-Marseille Université	Directeur de thèse
Laurent REGNIER	Aix-Marseille Université	Co-directeur de thèse

Abstract

The λ-calculus has been invented in 1936 by Alonzo Church as a way to formalize the notion of computation, and can be seen as an ancestor of today's programming languages. Since then it has been widely used as an abstract tool to study computation and programming. In particular many variants has been defined by adding different features to the original calculus. In this thesis we will be interested in some of these variants obtained by adding a non deterministic operator, and we will focus mostly on a probabilistic calculus.

The probabilistic λ-calculus relies on binary operators $+_p$ for $p \in [0;1]$ of probabilistic choice, whose interpretation is that a term $M +_p N$ behaves as M with probability p and as N with probability $1-p$. This calculus has been studied for some time, but the probabilistic behaviour has always been treated as a side effect: the operational semantics of the calculus is defined by saying that a term $M +_p N$ actually reduces into M with probability p and it actually reduces into N with probability $1-p$. This external treatment of the probabilities has some drawbacks, and in particular it requires the use a particular reduction strategy when reducing terms. Our purpose is to give a more equational representation of this calculus, by handling the probabilities inside the reduction rather than as a side effect.

To begin with we give a deterministic and contextual operational semantics for the call-by-name probabilistic λ-calculus. To express the probabilistic behaviour of the sum we quotient this calculus by equations stating the commutativity, associativity and idempotence of the sum, as well as the irrelevance of the behaviours of probability 0. We show that this quotient has little consequence on the calculus: the reduction modulo equivalence of the terms gives (almost) the same relation as the reduction where only the result is considered modulo equivalence. We also prove a standardization result for this calculus.

Then using this operational semantics without side effect we define a notion of equational theories for the probabilistic λ-calculus. We extend some notions of the deterministic λ-theories to this setting, and in particular we give a definition for the sensibility of a theory, stating that all diverging terms are equal. This notion is quite simple in a deterministic setting but becomes more complicated when we have a probabilistic computation, and terms may diverge only with a certain probability.

Finally we prove a generalization of the equality between the observational equivalence, the Böhm tree equality and the maximal coherent sensible λ-theory. To achieve this we give a notion of probabilistic Böhm trees generalizing the deterministic ones, and prove that this forms a model of the probabilistic λ-calculus, i.e. mainly that the equality of Böhm trees is stable under context. Then we prove a separability result stating that two terms with different Böhm trees are separable, i.e. are not observationally equivalent. From there we conclude by proving the correspondence between the probabilistic observational equivalence and the equality of the probabilistic Böhm trees, and that this relation is the maximal consistent sensible probabilistic λ-theory.

Résumé

Le λ-calcul a été inventé en 1936 par Alonzo Church pour formaliser la notion de calcul, et peut être considéré comme un ancêtre des langages de programmation d'aujourd'hui. Depuis, il a été largement utilisé en tant qu'outil pour étudier le calcul et la programmation. En particulier de nombreuses variantes ont vu le jour via l'ajout de nouvelles fonctionnalités à la structure d'origine. Dans cette thèse nous nous intéresserons à certaines de ces variantes obtenues en considérant un nouvel opérateur non déterministe, et nous nous pencherons plus particulièrement sur le cas probabiliste.

Le λ-calcul probabiliste est basé sur des opérateurs binaires $+_p$ pour $p \in [0;1]$ de choix probabiliste, tels qu'un terme $M +_p N$ se comporte soit comme M avec probabilité p, soit comme N avec probabilité $1-p$. Ce calcul a fait l'objet d'études depuis quelques temps, mais toujours en considérant le comportement probabiliste d'un terme comme un effet de bord: la sémantique opérationnelle est définie en disant qu'avec une probabilité p le terme $M +_p N$ se réduit effectivement en M, tandis qu'il se réduit effectivement en N avec une probabilité $1-p$. Cette vision des probabilités comme étant externes au calcul lui-même a des inconvénients, et elle impose entre autres de restreindre la réduction par le choix d'une stratégie particulière. Notre objectif est de présenter ce calcul d'une manière plus équationnelle, en intégrant le comportement probabiliste à la réduction sans le voir comme un effet de bord.

Tout d'abord nous définissons une sémantique opérationnelle déterministe et contextuelle pour le λ-calcul probabiliste en appel par nom. Afin de traduire la signification de la somme comme un choix probabiliste nous quotientons ce calcul par des équations exprimant sa commutativité, son associativité et son idempotence, ainsi que l'absence de pertinence des réductions de probabilité nulle. Nous démontrons que ce quotient ne déforme pas la réduction: considérer toute les règles de calcul modulo équivalence revient (presque) à considérer simplement le résultat du calcul modulo équivalence. Nous prouvons également un résultat de standardisation.

Au moyen de cette sémantique opérationnelle sans effet de bord nous définissons une notion de théorie équationnelle pour le λ-calcul probabiliste. Nous étendons les définitions de certaines notions concernant les λ-théories usuelles, et en particulier celle de bon sens, qui correspond à considérer les termes divergents comme égaux. Cette idée se formalise facilement dans un cadre déterministe mais est bien plus complexe dans le cas probabiliste, où un même terme peut à la fois diverger et converger avec certaines probabilités.

Pour finir nous généralisons un résultat affirmant que l'équivalence observationnelle, l'égalité des arbres de Böhm et la théorie cohérente sensée maximale forment une seule et même relation sur les termes. Pour cela nous définissons une notion d'arbres de Böhm probabilistes et nous prouvons qu'elle forme un modèle, c'est-à-dire essentiellement que l'égalité de ces arbres est stable par contexte. Nous démontrons ensuite un résultat de séparabilité disant que deux termes avec des arbres de Böhm distincts ne sont pas observationnellement équivalents. De là nous concluons en montrant que l'égalité des arbres de Böhms probabilistes correspond à l'équivalence observationnelle probabiliste, et que la théorie ainsi obtenue est maximale parmi toutes les théories probabilistes cohérentes sensées.

Remerciements

Pour commencer je voudrais remercier Mathias, qui m'a fait remarquer qu'il n'existe pas de synonyme convenable au mot "remercier" et grâce à qui je peux insister 16 fois sur des mots partageant la même racine en une seule page l'esprit en paix.

Je remercie ensuite Lionel pour m'avoir encadré durant toute cette thèse et même un peu avant, et pour sa bonne camaraderie. Il a été souvent disponible pour discuter des mes recherches durant la première moitié de mon doctorat, relire mes épreuves lors de la seconde, s'occuper de la paperasse administrative un peu tout le temps et m'aider à m'occuper lorsque je cherchais à procrastiner. Je remercie aussi Laurent, directeur de thèse malgré lui.

Merci aux rapporteurs Antonino Salibra et Sam Staton, pour avoir relu ce manuscrit dans des conditions que je qualifierais de non idéales. Merci à tous les membres du jury qui sont venus à Marseille (ou non, hélas) assister à ma soutenance.

Merci à toute l'équipe LDP, Dimitri, Laurent, Lionel, Myriam, Yves, et Manu le lyonnais grenoblois, ainsi que les autres thésards et invités qui sont restés plus ou moins longtemps à Marseille. J'ai vraiment apprécié mon séjour parmi vous, et je vous quitterai avec regret.

En parlant de bonne ambiance au travail, remercions les organisateurs du séminaire Chocola. Avoir tous les mois l'occasion d'assister à quelques exposés dans notre champ de recherche tout en retrouvant des gens que je verrais peu autrement, et en mangeant du chocolat, est toujours agréable.

Retournant aux origines mes remerciements vont à ma famille qui a indubitablement contribué à développer mon sens scientifique. Ils m'ont en particulier fait rencontrer très jeune Jacques Rouxel, Richard Peyzaret et Raymond Smullyan, qui m'ont longuement préparé à me plonger dans le monde de la logique.

Je remercie mes compagnons lyonnais, en commençant par mes divers colocataires Anaël, Lambin, Robin et Vincent, ainsi que, dans un ordre pseudo-aléatoire, Fred, Elie, Sam, Alvaro, Etienne, Rémi, Adrien que j'aurai aimé voir plus souvent, et tous les autres que j'ai plus ou moins côtoyés mais méritent d'être remerciés ne serait-ce que parce qu'ils sont sympas.

Merci aux (vieux) membres du club jeux, aux participants d'Octogônes, aux organisateurs des InterLudes, parce que travailler étant très stressant il est important de se détendre un peu de temps en temps. Mais juste quelques heures de plus et après on s'arrête, promis.

Un grand merci aux gérants de l'auberge Garibaldi, toujours accueillants et disponibles. C'est à chaque fois un plaisir de venir passer quelques jours dans les collines lyonnaises.

Bref merci à ma famille, à mes amis, à mes collègues, aux vendeurs de bandes dessinées et à tous ceux qui me remercieront dans leur propre thèse.

Contents

Introduction

λ-calculus, along with Turing machines and Gödel's recursive functions, is a way to define precisely what a computation is. In an abstract way, a notion of computation can be thought of as a set of rules which, when applied to some input, yield a result.

$$input \xrightarrow{computation\ rules} output$$

For instance we can define computation rules for the addition of two natural numbers and apply them to 2 and 3 to get

$$2, 3 \xrightarrow{\text{addition}} 5.$$

The most natural way to formalize this idea is to explain how the computation rules can be built. This is how recursive functions are defined: a recursive function necessarily takes as input a finite sequence of natural numbers, if it outputs a result then this result is necessarily a natural number, and there is a clear and precise description of the different ways such a function can be built. Turing machines can also be seen in this way. The input and output are more general than for recursive functions, as they take the form of data written on a tape, but again a machine is mostly defined by its transition table, i.e. the computation rules. In both these models there is (at least) one function or one machine for every computation: one for the addition, one for the multiplication, etc.

Yet in his paper *On computable numbers, with an application to the Entscheidungsproblem*, Turing defines a general notion of automatic machines but he also describes what he calls universal machines. Those are machines that expect as input the description of a machine as well as additional data, and simulates the application of the described machine to the data.

$$\begin{array}{lrcl} \text{If} & input & \xrightarrow{computation\ rules} & output \\ \text{then} & computation\ rules, input & \xrightarrow{\text{universal machine}} & output. \end{array}$$

For instance

$$\text{addition}, 2, 3 \xrightarrow{\text{universal machine}} 5.$$

This idea to describe the computation not in the computation rules but rather in the input is at the core of the λ-calculus. It can be thought of as an ancestor of today's programming languages: the input is a program written using some instructions, and the computation rules are defined once and for all to explain how to interpret these instructions.

Computation and rewriting

To understand how λ-calculus works let us consider a simpler calculus based on the same idea. This calculus will only compute expressions made of multiple sums and products of natural numbers.

We said such a calculus is made of programs, which we call *terms*, and of computation rules explaining how the instructions work. So let us first define the terms.

- For every natural number n we have a term $\underline{n}$.
- If M and N are terms then $M + N$ is a term.
- If M and N are terms then $M \times N$ is a term.

A shorter way to write this is

$$M, N := \underline{n} \mid M + N \mid M \times N.$$

Some examples of terms are $\underline{1}$, $\underline{2} + \underline{3}$ and $(\underline{15} + \underline{4}) \times \underline{6}$.

Next the computation rules should define the meaning of $+$ and $\times$. We write them as a relation $\rightarrow$ of *reduction* on terms.

$$\underline{n} + \underline{m} \rightarrow \underline{n + m}$$
$$\underline{n} \times \underline{m} \rightarrow \underline{n \times m}$$

We have for instance $\underline{2} + \underline{3} \rightarrow \underline{5}$ and $\underline{2} \times \underline{3} \rightarrow \underline{6}$.

What about the term $(\underline{15} + \underline{4}) \times \underline{6}$? It is neither a sum of natural number nor a product of natural numbers, but it contains one. We want to write $(\underline{15}+\underline{4}) \times \underline{6} \rightarrow \underline{19} \times \underline{6}$. In order to do this, we define *contexts* C as "terms with a hole":

$$C := [\,] \mid C + M \mid M + C \mid C \times M \mid M \times C.$$

Contexts are exactly like terms with one occurrence of a hole $[\,]$ in them. For instance $[\,]$, $[\,] + \underline{3}$ or $(\underline{15} + [\,]) \times \underline{6}$ are contexts. We can fill this hole with a term, and we write $C[M]$ the term obtained by filling the hole in C with the term M. For instance if $C = (\underline{15} + [\,]) \times \underline{6}$ then $C[\underline{2} + \underline{3}] = (\underline{15} + (\underline{2} + \underline{3})) \times \underline{6}$.

Now we can say we allow reduction under arbitrary context. This means that we can perform the following reductions for any context C:

$$C[\underline{n} + \underline{m}] \rightarrow C[\underline{n + m}]$$
$$C[\underline{n} \times \underline{m}] \rightarrow C[\underline{n \times m}]$$

Terms of the form $\underline{n} + \underline{m}$ or $\underline{n} \times \underline{m}$ are called *redexes*, for *reducible expressions*, and the reductions $C[\underline{n} + \underline{m}] \rightarrow C[\underline{n + m}]$ and $C[\underline{n} \times \underline{m}] \rightarrow C[\underline{n \times m}]$ are the reductions of the redexes $\underline{n} + \underline{m}$ and $\underline{n} \times \underline{m}$ under the context C. For instance we do have $(\underline{15} + \underline{4}) \times \underline{6} \rightarrow \underline{19} \times \underline{6}$ if we reduce the redex $\underline{15} + \underline{4}$ under the context $C = [\,] \times \underline{6}$.

With this calculus we can compute any expression made of sums and products of natural numbers by performing multiple reductions. We have for instance

$$\begin{aligned}(\underline{15} + \underline{4}) \times \underline{6} &\to \underline{19} \times \underline{6} \\ &\to \underline{114}.\end{aligned}$$

The λ-calculus

Our previous example of calculus is a very simple one, and it is certainly not sufficient to express every possible computation. Just having sums and products as instructions is not powerful enough, and we need to find a better set of instructions.

The λ-calculus is based on a single notion: functions. We consider given an infinite set of variables Var, whose elements will usually be written x, y or z. The set Λ of λ-terms is built as follows.

- Every *variable* $x \in Var$ is a term.
- If M is a term and x is a variable then we can build the function which to x associates M. It corresponds to the notation $x \mapsto M$ in mathematics, but here we write it $\lambda x.M$ and call it the *abstraction* of x in M.
- If M and N are terms then we can build the *application* of M to N, usually noted $M(N)$ in mathematics but written here $M\ N$.

In short:

$$M, N \in \Lambda := x \mid \lambda x.M \mid M\ N.$$

Some examples of terms are $\lambda x.x$, $\lambda x.\lambda y.x$ and $(\lambda x.\lambda y.\lambda z.x\ y\ (x\ z))\ (\lambda u.\lambda v.u)$.

The only reduction rule defines how functions should behave. This reduction is called the β-reduction.

$$(\lambda x.M)\ N \to_\beta M\left[{}^{N}\!/_{x}\right] = \text{"}M \text{ where } x \text{ is replaced by } N\text{"}$$

For instance we have for any term N that

$$(\lambda x.x)\ N \to_\beta N.$$

As in the previous calculus we consider reduction under arbitrary context, where the contexts are defined by

$$C := [\,] \mid \lambda x.C \mid C\ M \mid M\ C.$$

This way we have

$$\begin{aligned}\Big[\,(\lambda x.\lambda y.\lambda z.x\ y\ (x\ z))\ \ (\lambda u.\lambda v.u)\,\Big] &\to_\beta \lambda y.\lambda z.\Big[\,(\lambda u.\lambda v.u)\ \ y\Big]\ ((\lambda u.\lambda v.u)\ \ z)\\ &\to_\beta \lambda y.\lambda z.\Big[\,(\lambda v.y)\ \ ((\lambda u.\lambda v.u)\ \ z)\Big]\\ &\to_\beta \lambda y.\lambda z.y\end{aligned}$$

where at each step we reduce the bracketed redex.

This calculus may not look very appealing, and it is not clear how we can use it to represent every computation. So let us show that we can easily recover sums and products of natural numbers in the λ-calculus.

- Given a natural number n, the n-th *Church numeral* is the term

$$\underline{n} = \lambda f.\lambda x.f\ (f...(f\ x)...)$$

  where the variable f is applied n times to the variable x. We can also write $\underline{n} = \lambda f.\lambda x.f^n\ x$. For instance $\underline{3} = \lambda f.\lambda x.f\ (f\ (f\ x))$.

- The term $\underline{\mathrm{add}}$ is given by

$$\underline{\mathrm{add}} = \lambda m.\lambda n.\lambda f.\lambda x.m\ f\ (n\ f\ x).$$

- The term $\underline{\mathrm{prod}}$ is given by

$$\underline{\mathrm{prod}} = \lambda m.\lambda n.\lambda f.m\ (n\ f).$$

We can check that for any natural numbers m and n, the term $\underline{\mathrm{add}}\ \underline{m}\ \underline{n}$ reduces into $\underline{m+n}$ and the term $\underline{\mathrm{prod}}\ \underline{m}\ \underline{n}$ reduces into $\underline{m \times n}$.

Equational λ-theories

We can define calculi by giving reduction rules on terms, but at some point we would like to extract some results from the computations and forget exactly how we obtained them. For instance we saw that in our first calculus the term $(\underline{15}+\underline{4})\times\underline{6}$ reduces in two steps into $\underline{114}$, but we do not have $(\underline{15}+\underline{4})\times\underline{6} = \underline{114}$. Moreover we also have $\underline{100}+\underline{14} \to \underline{114}$ but there is no relation between $(\underline{15}+\underline{4})\times\underline{6}$ and $\underline{100}+\underline{14}$: neither one reduces into the other.

For that reason we define a notion of equality, or *congruence*, on terms. A congruence is a relation $\simeq$ with the following properties.

- It is reflexive: for all term M we have $M \simeq M$.
- It is transitive: if $M \simeq N$ and $N \simeq P$ then $M \simeq P$.
- It is symmetric: if $M \simeq N$ then $N \simeq M$.

- It is contextual: if $M \simeq N$ then for all context C, $C[M] \simeq C[N]$.

Then from any calculus we can induce a congruence on terms, by replacing the reduction rules by equality rules. In the case of the λ-calculus, this congruence is called the β-equivalence and noted $=_\beta$. It is the least congruence on λ-terms such that for all M, N and x,

$$(\lambda x.M)\ N =_\beta M\left[{}^{N}/_{x}\right].$$

This way we can write $(\lambda x.\lambda y.\lambda z.x\ y\ (x\ z))\ (\lambda u.\lambda v.u) =_\beta (\lambda x.x)\ (\lambda y.\lambda z.y)$ or $\underline{\mathrm{add}}\ \underline{m}\ \underline{n} =_\beta \underline{m+n}$.

An interesting property to have when we deal with such congruences is the *confluence* of the calculus. A reduction $\to$ is said to be confluent if whenever a term M reduces (in any number of steps) into two terms N_1 and N_2, there exists a term P such that both N_1 and N_2 reduce into P.

$$\begin{tikzcd} & M \arrow[dl] \arrow[dr] & \\ N_1 \arrow[dr, dashed] & & N_2 \arrow[dl, dashed] \\ & P & \end{tikzcd}$$

If this holds then the congruence $\simeq$ induced by $\to$ is simple: we have $M \simeq N$ if and only if there is a P such that M and N both reduce into P.

In particular in the λ-calculus the β-reduction is confluent so we have

$$M =_\beta N \text{ if and only if } \exists P : M \to_\beta \ldots \to_\beta P \text{ and } N \to_\beta \ldots \to_\beta P.$$

We defined a particular congruence induced by the reduction rules, but we can also consider a more general notion of "meaningful" congruences. We call *theory* any congruence which respects the reduction rules, but may also equates terms which are not related by the computation. We call λ-theory the theories in the λ-calculus.

An example of a non trivial λ-theory is the $\beta\eta$-equivalence. When we substitute a term N for a variable x in another term M, it may happen that x does not appear in M, in which case $M\left[{}^{N}/_{x}\right] = M$. Now if x does not appear in M, let us consider the term $\lambda x.M\ x$. If we apply it to a term N we get

$$(\lambda x.M\ x)\ N \to_\beta (M\ x)\left[{}^{N}/_{x}\right] = M\ N$$

so this is the same as applying directly M to N. But in general there is no direct relation between $\lambda x.M\ x$ and M, and we have $\lambda x.M\ x \neq_\beta M$. The difference between these terms only disappears once we apply them to another term. But we can decide to add a new rule, called η, saying that if x does not appear in M then

$$\lambda x.M\ x =_\eta M.$$

Then the least congruence that respects both the β and η rules is called the $\beta\eta$-equivalence, and it is a λ-theory.

What about probabilities?

The purpose of this thesis is to see how some properties of the λ-calculus can be recovered if we consider an extension of this calculus with a probabilistic behaviour.

The probabilistic λ-calculus is obtained by adding an instruction of probabilistic choice to the λ-calculus, along with its reduction rules. The set Λ_+ of probabilistic terms is defined by

$$M, N \in \Lambda_+ := x \mid \lambda x.M \mid M\ N \mid M +_p N$$

where p is a probability and ranges over $[0;1]$.

To express the idea of a probabilistic behaviour we do not simply consider the reduction as a relation on terms, but we consider that reductions happen with some probability. The reduction of a β-redex happens with a probability 1, whereas the reduction of a sum $M +_p N$ will return M with probability p and N with probability $1-p$.

$$\begin{array}{rcl} (\lambda x.M)\ N & \xrightarrow{1} & M\left[{}^{N}/_{x}\right] \\ M +_p N & \xrightarrow{p} & M \\ M +_p N & \xrightarrow{1-p} & N \end{array}$$

Dealing with probabilities is usually done outside the calculus. We consider that the probability of a reduction path is equal to the product of the probabilities of its reduction steps: the reduction $M \xrightarrow{p} N \xrightarrow{q} P$ happens with probability $p \times q$. We also consider that the behaviour of a term is described by its reduction paths where we made all the possible choices when reducing sums.

Consider for instance the term $x +_p (y +_q z)$: it has seven maximal reduction paths

$$\begin{array}{ccccc} x +_p (y +_q z) & \xrightarrow{p} & x & & \\ x +_p (y +_q z) & \xrightarrow{1-p} & y +_q z & \xrightarrow{q} & y \\ x +_p (y +_q z) & \xrightarrow{1-p} & y +_q z & \xrightarrow{1-q} & z \\ & & & & \\ x +_p (y +_q z) & \xrightarrow{q} & x +_p q & \xrightarrow{p} & x \\ x +_p (y +_q z) & \xrightarrow{q} & x +_p q & \xrightarrow{1-p} & y \\ x +_p (y +_q z) & \xrightarrow{1-q} & x +_p z & \xrightarrow{p} & x \\ x +_p (y +_q z) & \xrightarrow{1-q} & x +_p z & \xrightarrow{1-p} & z. \end{array}$$

We have two kinds of reductions: either we reduce the outer sum first, or we reduce the inner one first. In the first case we obtain x with probability p, y with probability $(1-p)\times q$ and z with probability $(1-p)\times(1-q)$; in the second case we get x with probability $q\times p+(1-q)\times p=p$, y with probability $q\times(1-p)$ and z with probability $(1-q)\times(1-p)$. We associate the same probability in both cases to each result, and we indeed obtain a probability distribution as $p+(1-p)\times q+(1-p)\times(1-q)=1$. But this information is external to the calculus: there is no object D in the calculus representing this probability distribution, and such that $M\to D$.

But some trouble arise when we mix the β-reduction and the reduction of sums. Let us look at all the maximal reduction paths of the term $(\lambda x.x\,x)\;(y+_{\frac{1}{2}}z)$:

$$
\begin{array}{lclcl}
(\lambda x.x\;x)\;(y+_{\frac{1}{2}}z) & \xrightarrow{\frac{1}{2}} & (\lambda x.x\;x)\;y & \xrightarrow{1} & y\;y \\
(\lambda x.x\;x)\;(y+_{\frac{1}{2}}z) & \xrightarrow{\frac{1}{2}} & (\lambda x.x\;x)\;z & \xrightarrow{1} & z\;z
\end{array}
$$

$$
\begin{array}{lclclcl}
(\lambda x.x\;x)\;(y+_{\frac{1}{2}}z) & \xrightarrow{1} & (y+_{\frac{1}{2}}z)\;(y+_{\frac{1}{2}}z) & \xrightarrow{\frac{1}{2}} & y\;(y+_{\frac{1}{2}}z) & \xrightarrow{\frac{1}{2}} & y\;y \\
(\lambda x.x\;x)\;(y+_{\frac{1}{2}}z) & \xrightarrow{1} & (y+_{\frac{1}{2}}z)\;(y+_{\frac{1}{2}}z) & \xrightarrow{\frac{1}{2}} & y\;(y+_{\frac{1}{2}}z) & \xrightarrow{\frac{1}{2}} & y\;z \\
(\lambda x.x\;x)\;(y+_{\frac{1}{2}}z) & \xrightarrow{1} & (y+_{\frac{1}{2}}z)\;(y+_{\frac{1}{2}}z) & \xrightarrow{\frac{1}{2}} & z\;(y+_{\frac{1}{2}}z) & \xrightarrow{\frac{1}{2}} & z\;y \\
(\lambda x.x\;x)\;(y+_{\frac{1}{2}}z) & \xrightarrow{1} & (y+_{\frac{1}{2}}z)\;(y+_{\frac{1}{2}}z) & \xrightarrow{\frac{1}{2}} & z\;(y+_{\frac{1}{2}}z) & \xrightarrow{\frac{1}{2}} & z\;z
\end{array}
$$

$$
\begin{array}{lclclcl}
(\lambda x.x\;x)\;(y+_{\frac{1}{2}}z) & \xrightarrow{1} & (y+_{\frac{1}{2}}z)\;(y+_{\frac{1}{2}}z) & \xrightarrow{\frac{1}{2}} & (y+_{\frac{1}{2}}z)\;y & \xrightarrow{\frac{1}{2}} & y\;y \\
(\lambda x.x\;x)\;(y+_{\frac{1}{2}}z) & \xrightarrow{1} & (y+_{\frac{1}{2}}z)\;(y+_{\frac{1}{2}}z) & \xrightarrow{\frac{1}{2}} & (y+_{\frac{1}{2}}z)\;y & \xrightarrow{\frac{1}{2}} & z\;y \\
(\lambda x.x\;x)\;(y+_{\frac{1}{2}}z) & \xrightarrow{1} & (y+_{\frac{1}{2}}z)\;(y+_{\frac{1}{2}}z) & \xrightarrow{\frac{1}{2}} & (y+_{\frac{1}{2}}z)\;z & \xrightarrow{\frac{1}{2}} & y\;z \\
(\lambda x.x\;x)\;(y+_{\frac{1}{2}}z) & \xrightarrow{1} & (y+_{\frac{1}{2}}z)\;(y+_{\frac{1}{2}}z) & \xrightarrow{\frac{1}{2}} & (y+_{\frac{1}{2}}z)\;z & \xrightarrow{\frac{1}{2}} & z\;z
\end{array}
$$

To get a probability distribution we need to keep track of the two possible choices whenever we reduce a sum. So here we have three interesting sets of reduction paths:

- in the case where we reduce the sum first and then reduce the β-redex we have two reduction paths and we get $y\;y$ with probability $\frac{1}{2}$ and $z\;z$ with probability $\frac{1}{2}$;
- if we reduce the β-redex, then the leftmost sum and lastly the rightmost sum we have four reduction paths and we get $y\;y$, $y\;z$, $z\;y$ or $z\;z$, each with probability $\frac{1}{4}$;
- if we reduce the β-redex first but reduce the rightmost sum before the leftmost one we also get four reduction paths and we end up with the same probability distribution.

We can see that in the last two cases the distribution probability is the same, but in the first case we obtain a different one. The reason for this is pretty clear: when we perform a β-reduction we may duplicate the argument, and in particular we may duplicate a sum, and making a choice once and for all before duplicating it or actually duplicating the sum and making the choice several times will obviously yield different results.

For that reason if we allow reduction under arbitrary context, the calculus does not make sense. To prevent this we have to impose some *reduction strategy*, i.e. some restriction on the reductions we are allowed to perform. There are mostly two kinds of reduction strategies: those which reduce β-redexes without caring about the shape of the argument (the *call-by-name* strategies) and those which evaluate the argument before reducing a β-redex (the *call-by-value* strategies).

We will be interested in the call-by-name version of the probabilistic λ-calculus. In the literature this calculus is usually restricted to the *head reduction*, i.e. reduction under contexts of the form $\lambda x_1 ... \lambda x_n.[\,]\ P_1\ ...\ P_n$. The reduction is then fully described by the following rules, without any extension to another context.

$$\begin{array}{lcl} \lambda x_1...\lambda x_n.(\lambda x.M)\ N\ P_1\ ...\ P_n & \xrightarrow{1}_h & \lambda x_1...\lambda x_n.M\left[{}^{N}/_{x}\right]\ P_1\ ...\ P_n \\ \lambda x_1...\lambda x_n.(M +_p N)\ P_1\ ...\ P_n & \xrightarrow{p}_h & \lambda x_1...\lambda x_n.M\ P_1\ ...\ P_n \\ \lambda x_1...\lambda x_n.(M +_p N)\ P_1\ ...\ P_n & \xrightarrow{1-p}_h & \lambda x_1...\lambda x_n.N\ P_1\ ...\ P_n \end{array}$$

This reduction never reduces a sum in an argument. Besides we can see that every term is in one of the three following forms:

$$\begin{array}{rccc} & \lambda x_1...\lambda x_n. & y & P_1\ ...\ P_n \\ \text{or} & \lambda x_1...\lambda x_n. & (\lambda x.M)\ N & P_1\ ...\ P_n \\ \text{or} & \lambda x_1...\lambda x_n. & (M +_p N) & P_1\ ...\ P_n. \end{array}$$

In the first case the term does not reduce, in the second case it has one possible β-reduction (with probability 1) and in the last case it has two possible reductions with probabilities p and $1-p$. There are far fewer reductions than when we reduce under arbitrary contexts, where a term could have any number of possible reductions. We can actually associate to every term a tree of reductions. This way it is immediate to associate a probability distribution to a term.

If we look at the possible head reductions paths for our previous example $(\lambda x.x\ x)\ (y +_{\frac{1}{2}} z)$, we get:

$$\begin{array}{lclcl} (\lambda x.x\ x)\ (y +_{\frac{1}{2}} z) & \xrightarrow{1}_h & (y +_{\frac{1}{2}} z)\ (y +_{\frac{1}{2}} z) & \xrightarrow{\frac{1}{2}}_h & y\ (y +_{\frac{1}{2}} z) \\ (\lambda x.x\ x)\ (y +_{\frac{1}{2}} z) & \xrightarrow{1}_h & (y +_{\frac{1}{2}} z)\ (y +_{\frac{1}{2}} z) & \xrightarrow{\frac{1}{2}}_h & z\ (y +_{\frac{1}{2}} z). \end{array}$$

There are only two maximal reduction paths, each with probability $\frac{1}{2}$. This corresponds to the following reduction tree:

$$\begin{array}{c} (\lambda x.x\ x)\ (y +_{\frac{1}{2}} z) \\ \downarrow 1 \\ (y +_{\frac{1}{2}} z)\ (y +_{\frac{1}{2}} z) \\ \tfrac{1}{2}\swarrow \qquad \searrow\tfrac{1}{2} \\ y\ (y +_{\frac{1}{2}} z) \quad z\ (y +_{\frac{1}{2}} z) \end{array}$$

Probabilities and equations

The probabilistic head reduction is quite satisfying from a computational point of a view, but it is much less so from the equational one. To begin with we described the equational theories as purely contextual: if $M \simeq N$ then $C[M] \simeq C[N]$ for any context C. Consequently the appearance of a reduction strategy in the calculus, i.e. a restriction on the contexts, is unwelcome.

But more importantly an important part of the calculus is external to the reduction. The computation of probabilities and probability distributions is done by looking at reductions paths and is not part of the reduction itself. Thus we cannot just define theories as equalities that respect the reduction. We could try to consider a relation $\simeq_p$ of "equality with some probability p", with $M +_p N \simeq_p M$ and $M +_p N \simeq_{1-p} N$. But then if we try to get an equivalent of the transitivity (stating that if $M \simeq N$ and $N \simeq P$ then $M \simeq P$), we have to face the following situation:

$$\begin{array}{rclcrcl} (M +_p N) +_q P & \simeq_q & M +_p N & \text{and} & M +_p N & \simeq_p & M \\ M +_q N & \simeq_q & M & \text{and} & M & \simeq_p & M +_p P \end{array}$$

In the first case the computation gives that $(M +_p N) +_q P$ reduces into M with probability $p \times q$, so we would expect $(M +_p N) +_q P \simeq_{p\times q} M$. In the second case we have that $M +_q N$ reduces to M with probability q and $M +_p P$ reduces to M with probability p, so we should have $M +_q N \simeq_{\min(p,q)} M +_p P$. So what is the "right" transitivity rule?

The situation gets even worse if we consider more than three terms. For instance we have

$$M +_p N \simeq_p M \simeq_q M +_q P \simeq_{1-q} P \simeq_r P +_r Q.$$

Now $M +_p N$ and $P +_r Q$ are completely different terms, so no transitivity can really hold. In this situation we can hardly speak of an equational theory.

Another attempt could be to look at the probability distributions we associate to the terms. We can consider that two terms are equal if they describe the same probability distribution. This is a much more satisfying solution, but this yields a far more complicated definition of equational theories than for the deterministic case. In particular it uses the information given by sets of reduction paths, and is not just an extension of the relation of reduction.

Luckily there is a simple solution to all these problems. All you actually need to do is to change the calculus in order to internalize the probabilistic behaviours. This way not only can we get a calculus with deterministic computation (on terms describing probabilistic information) but we can also recover the full contextuality of the reduction. Then we can use the usual notion of equational theory.

Related works

At first most of the research about non-deterministic and probabilistic calculi revolved around denotational semantics. In 1976, Plotkin [9] described a semantics for simple programming languages with a non-deterministic choice operator *or*, which he later adapted to the probabilistic case [8]. In 1978 Saheb-Djahromi introduced an operational semantics for a probabilistic calculus [10], but he had already observed that such a calculus required the choice of a reduction strategy, so he consider only a weak head reduction and he focuses more on the denotational semantics.

Maybe the first study of an operational semantics for a non-deterministic untyped λ-calculus has been done in 1995 by de'Liguoro and Piperno [4]. Their work is similar to ours: they consider a λ-calculus extended with a choice operator $+$ for which they give some reduction rules, prove a standardization theorem, define some non-deterministic Böhm trees and try to get a separability result. But this last attempt fails, as simple non-determinism does not allow enough quantification. In a probabilistic calculus, the terms $x +_p y$ and $x +_q z$ will return the same result with a probability pq, so if you check whether they are equal n times the answer will always be positive with a probability $(pq)^n$, which converges to 0 unless $p = q = 1$. But in a non-deterministic calculus $x + y$ and $x + z$ may be equal, without further information, providing much less tools to get a general separation result.

Recently more work appeared in this fashion. For instance Dal Lago and Zorzi published some results about two operational semantics for the probabilistic λ-calculus, without using any denotational semantics [2]. But once again the two sets of rules they consider describe a call-by-name and a call-by-value calculi, and they do not allow reduction under any context.

To our knowledge no research has been done on an equational presentation of a non-deterministic calculus. Our work is thus based solely on previous results about the deterministic λ-calculus. In particular our ultimate goal is the generalization of a theorem proven independently by Hyland [7] and Wadsworth [12]: the infinitely extensional Böhm tree equality coincide with the observational equivalence, and this is the maximum for the inclusion of all the consistent sensible theories.

The idea of dealing with sums inside the calculus is actually not new in the field of quantitative λ-calculi. It comes from the differential λ-calculus [6], and can also be found in the algebraic λ-calculus [11]. The reason is that in these cases the sums do not have such a clear computational meaning. For instance the differential of a term is a sum of terms, and it does not make sense to say that this sum can reduce into each of its component. Yet there exists no presentation of a probabilistic calculus in this fashion.

Layout

In the first chapters we will describe an operational semantics for general non-deterministic λ-calculi with labelled sums. In the first chapter we will show how we can present those calculi with deterministic and contextual operational semantics and prove that this presentation actually describes the behaviour we expect from such calculi.

In the second chapter we will establish a standardization result. The introduction of sums in the calculi makes this result more difficult than in the deterministic calculus but we can still recover a useful standardization property. In particular the standardization can be used to simplify the description of the reductions in our calculus.

From the third chapter onwards we will restrict our study to the probabilistic λ-calculus. In this third chapter we will give a definition of probabilistic equational theories. Although this definition is derived in a straightforward way from the operational semantics and can hold for any calculus with labelled sums, we will describe more precisely some specific theories, and in particular we will give a notion of sensible theory which is proper to the probabilistic case.

The fourth chapter will be devoted to the notion of Böhm trees. We will give a definition of probabilistic Böhm trees and infinitely extensional probabilistic Böhm trees, and we will prove that they form a model of the probabilistic λ-calculus, i.e. that the Böhm tree equality is a theory.

Finally we will prove a separability result in the fifth chapter. A reason why the Böhm trees are an interesting structure in the usual deterministic λ-calculus is that they describe exactly the observational equivalence of terms. The way we extend the definition of Böhm trees to the probabilistic calculus is quite natural, so it is satisfying to know that this is indeed the right notion of Böhm trees, in the sense that they also correspond to a natural notion of probabilistic observational equivalence.

Common notations and terminology

Relations. In this thesis we will use some common operations on relations. Given two relations $\mathcal{R}$ and $\mathcal{R}'$:

- we write $\mathcal{R} \cdot \mathcal{R}'$ for the composition of $\mathcal{R}$ and $\mathcal{R}'$: we have $x \mathrel{\mathcal{R} \cdot \mathcal{R}'} y$ if and only if there is z such that $x \mathrel{\mathcal{R}} z$ and $z \mathrel{\mathcal{R}'} y$;

- we write $\mathcal{R}^?$ for the reflexive closure or $\mathcal{R}$: $x \mathcal{R}^? y$ iff $x = y$ or $x \mathcal{R} y$;
- we write $\mathcal{R}^0$ for the equality ($x \mathcal{R}^0 y$ iff $x = y$) and for $n \in \mathbb{N}$ we define $\mathcal{R}^{n+1}$ by $x \mathcal{R}^{n+1} y$ iff $x \mathcal{R} \cdot \mathcal{R}^n y$: for all $n \in \mathbb{N}$ we have $x \mathcal{R}^n y$ iff $x = x_0 \mathcal{R} x_1 \mathcal{R} ... \mathcal{R} x_n = y$ for some x_1,...x_{n-1};
- we write $\mathcal{R}^+$ for the transitive closure of $\mathcal{R}$: $\mathcal{R}^+ = \cup_{n\geq 1} \mathcal{R}^n$;
- we write $\mathcal{R}^*$ for the reflexive and transitive closure of $\mathcal{R}$: $\mathcal{R}^* = \cup_{n\geq 0} \mathcal{R}^n$.

If we consider a reduction $\to$:

- we write $\leftarrow$ for the reverse relation ($x \leftarrow y$ iff $y \to x$);
- we write $\twoheadrightarrow$ for its reflexive and transitive closure $\to^*$;
- we write $\leftrightarrow$ for its symmetric closure: $\leftrightarrow = \to \cup \leftarrow$;
- we write $\twoheadleftarrow\!\!\twoheadrightarrow$ for its reflexive symmetric transitive closure: $\twoheadleftarrow\!\!\twoheadrightarrow = \leftrightarrow^*$.

Reductions. Given a reduction $\to$ on terms, a term M is said to be *normal* or in *normal form* if there is no reduction $M \to N$. It is said to be *normalizing* if it reduces into a normal form, i.e. $M \twoheadrightarrow N$ and there is no reduction $N \to P$. It is said to be *strongly normalizing* if there is no infinite reduction $M \to N_1 \to ...$.

The reduction $\to$ itself is said *normalizing* (resp. *strongly normalizing*) if every term is normalizing (resp. strongly normalizing). Moreover the reduction is said to be *confluent* if any two reduction paths starting from the same term can be made to meet again: if $M \twoheadrightarrow N_1$ and $M \twoheadrightarrow N_2$ then there is P such that $N_1 \twoheadrightarrow P$ and $N_2 \twoheadrightarrow P$. The reduction is weakly confluent if this holds at least for reductions of length 1: whenever $M \to N_1$ and $M \to N_2$ there is P such that $N_1 \twoheadrightarrow P$ and $N_2 \twoheadrightarrow P$. Furthermore a weakly confluent reduction has the *diamond property* if we can use only one-step reductions: whenever $M \to N_1$ and $M \to N_2$ with $N_1 \neq N_2$ there is P such that $N_1 \to P$ and $N_2 \to P$.

Every reduction with the diamond property is confluent, and every confluent reduction is weakly confluent. Besides if a reduction is weakly confluent and strongly normalizing then is it also confluent.

λ-terms. We will study λ-terms with labelled sums:

$$M, N := x \mid \lambda x.M \mid M\ N \mid M +_l N, l \in \mathcal{L}$$

for some set of labels $\mathcal{L}$.

We consider the application as left-associative: the term $M\ N\ P$ is to be read $(M\ N)\ P$. We also consider that the abstraction and the application takes priority over the sum, and the application takes priority over the abstraction: we write $\lambda x.M\ N +_l P\ Q$ for the term $(\lambda x.(M\ N)) +_l (P\ Q)$. Besides when writing multiple abstractions we will omit all but the first λ: we will write $\lambda x_1 ... x_n.M$ for $\lambda x_1 ... \lambda x_n.M$.

To shorten the notations we will sometimes write sequences of abstractions or applications as vectors: the notation $\lambda \overrightarrow{x}_{1...n}.M\ \overrightarrow{P}_{1...m}$ represents the term $\lambda x_1 ... x_n.M\ P_1\ ...\ P_m$.

Free variables. The set $\mathrm{FV}(M)$ of the free variables of a term M is defined by induction on M by:

- $\mathrm{FV}(x) = \{x\}$;
- $\mathrm{FV}(\lambda x.M) = \mathrm{FV}(M) \setminus \{x\}$;
- $\mathrm{FV}(M\ N) = \mathrm{FV}(M) \cup \mathrm{FV}(N)$;
- $\mathrm{FV}(M +_l N) = \mathrm{FV}(M) \cup \mathrm{FV}(N)$.

Substitutions We consider two ways to substitute a term P for a variable x in a term M: the regular substitution $M\left[{}^{P}/_{x}\right]$ and the binding substitution $M\left\{{}^{P}/_{x}\right\}$. Both are defined by induction on M.

The regular substitution is given by:

- $x\left[{}^{P}/_{x}\right] = P$;
- $y\left[{}^{P}/_{x}\right] = y$ if $y \neq x$;
- $(\lambda x.M)\left[{}^{P}/_{x}\right] = \lambda x.M$;
- $(\lambda y.M)\left[{}^{P}/_{x}\right] = \lambda y.\left(M\left[{}^{P}/_{x}\right]\right)$ if $y \neq x$ and $y \notin \mathrm{FV}(P)$;
- $(M\ N)\left[{}^{P}/_{x}\right] = \left(M\left[{}^{P}/_{x}\right]\right)\ \left(N\left[{}^{P}/_{x}\right]\right)$;
- $(M +_l N)\left[{}^{P}/_{x}\right] = \left(M\left[{}^{P}/_{x}\right]\right) +_l \left(N\left[{}^{P}/_{x}\right]\right)$.

The binding substitution is given by:

- $x\left\{{}^{P}/_{x}\right\} = P$;
- $y\left\{{}^{P}/_{x}\right\} = y$ if $y \neq x$;
- $(\lambda x.M)\left\{{}^{P}/_{x}\right\} = \lambda x.M$;
- $(\lambda y.M)\left\{{}^{P}/_{x}\right\} = \lambda y.\left(M\left\{{}^{P}/_{x}\right\}\right)$ if $y \neq x$;
- $(M\ N)\left\{{}^{P}/_{x}\right\} = \left(M\left\{{}^{P}/_{x}\right\}\right)\ \left(N\left\{{}^{P}/_{x}\right\}\right)$;

- $(M +_l N)\left\{{}^{P}/_{x}\right\} = \left(M\left\{{}^{P}/_{x}\right\}\right) +_l \left(N\left\{{}^{P}/_{x}\right\}\right)$.

The difference is that when we substitute P for x in an abstraction $\lambda x.M$, the regular substitution does not allow the free variables of P to be bound by the abstraction λx, whereas the binding substitution does.

A consequence is that while the binding substitution is always define, the regular one is not, and we need to consider the terms modulo α-equivalence:

$$\lambda x.M =_\alpha \lambda y.\left(M\left[{}^{y}/_{x}\right]\right) \text{ if } y \notin \mathrm{FV}(M).$$

If $M =_\alpha M'$ and $M\left[{}^{P}/_{x}\right]$ and $M'\left[{}^{P}/_{x}\right]$ are both defined then we always have $M\left[{}^{P}/_{x}\right] =_\alpha M'\left[{}^{P}/_{x}\right]$, and given any two terms M and P and any variable x we can always find $M' =_\alpha M$ such that $M'\left[{}^{P}/_{x}\right]$ is defined.

A reduction $\to$ is said to be *substitutive* if

$$M \to N \text{ implies } M\left[{}^{P}/_{x}\right] \to N\left[{}^{P}/_{x}\right]$$

for all x and P.

Contexts. Contexts are defined by

$$C := [\,] \mid \lambda x.C \mid C\ M \mid M\ C \mid C +_l M \mid M +_l C$$

and for any term M the term $C[M]$ is defined by induction on C:

- $([\,])[M] = M$;
- $(\lambda x.C)[M] = \lambda x.C[M]$;
- $(C\ N)[M] = C[M]\ N$;
- $(N\ C)[M] = N\ C[M]$;
- $(C +_l N)[M] = C[M] +_l N$;
- $(N +_l C)[M] = N +_l C[M]$.

If we consider $[\,]$ as a variable this definition is equivalent to $C[M] = C\left\{{}^{M}/_{[\,]}\right\}$.

Reductions modulo. We will consider reductions modulo equivalences relations, usually the β-reduction $\to_\beta$. Given an equivalence relation $\equiv$ on terms we write $\to_{\beta/\equiv}$ the β-reduction modulo $\equiv$, such that $M \to_{\beta/\equiv} N$ iff $M \equiv \cdot \to_\beta \cdot \equiv N$.

According to our definitions the relation $\to_{\beta/\equiv}^0$ is the equality on terms. But morally the reduction $\to_{\beta/\equiv}$ is a reduction on classes of terms modulo $\equiv$, so we consider that $M \to_{\beta/\equiv}^0 N$ whenever $M \equiv N$.

1 Non-deterministic λ-calculi

Our first goal is to give a presentation of the probabilistic λ-calculus which internalizes the probabilistic information, in order to obtain a deterministic and confluent calculus. This presentation actually works for a larger set of calculi, including the non-deterministic λ-calculus, with or without multiplicities, and the algebraic λ-calculus.

For a given set of labels $\mathcal{L}$ we consider terms with labelled sums:

$$M, N \in \Lambda_+^{\mathcal{L}} := x \mid \lambda x.M \mid M\ N \mid M +_l N, l \in \mathcal{L}.$$

The β-reduction is defined as usual on these terms by

$$(\lambda x.M)\ N \to_\beta M\left[{}^{N}/_{x}\right]$$

extended to arbitrary context.

To define the interpretation of the sums, rather than giving a reduction making a choice between the two subterms we simply define a redution $\to_+$ by:

$$\begin{aligned} \lambda x.(M +_l N) &\to_+ \lambda x.M +_l \lambda x.N \\ (M +_l N)\ P &\to_+ M\ P +_l N\ P \end{aligned}$$

extended to context.

For the probabilistic calculus, i.e. with $\mathcal{L} = [0;1]$, the usual probabilistic head-reduction is defined by:

$$\begin{aligned} \lambda x_1 ... x_n.(M +_p N)\ P_1\ ...\ P_n &\xrightarrow{p}_h \lambda x_1 ... x_n.M\ P_1\ ...\ P_n \\ \lambda x_1 ... x_n.(M +_p N)\ P_1\ ...\ P_n &\xrightarrow{1-p}_h \lambda x_1 ... x_n.N\ P_1\ ...\ P_n. \end{aligned}$$

Here we have

$$\lambda x_1 ... x_n.(M +_p N)\ P_1\ ...\ P_n \twoheadrightarrow_+ \lambda x_1 ... x_n.M\ P_1\ ...\ P_n +_p \lambda x_1 ... x_n.N\ P_1\ ...\ P_n.$$

Instead of choosing between two terms, we reduce deterministically into their syntactic sum.

As we do not output any information, we need to handle the computation on the probabilities inside the calculus. For instance if the probabilistic reduction gives $M \xrightarrow{p}_h N \xrightarrow{q}_h P$ then we usually consider that M reduces into P with probability pq, so we have to express this in our calculus.

To do so we define a syntactic equivalence on the terms, given by the following rules:

$$
\begin{array}{rcll}
M +_l N & \equiv_{\text{syn}} & N +_{\gamma(l)} M & \gamma : \mathcal{L} \to \mathcal{L} \\
(M +_l N) +_{l'} P & \equiv_{\text{syn}} & M +_{\alpha_1(l,l')} (N +_{\alpha_2(l,l')} P) & \alpha_1, \alpha_2 : \mathcal{L}^2 \to \mathcal{L} \\
M +_l M & \equiv_{\text{syn}} & M & \text{if } l \in \mathcal{I} \subset \mathcal{L} \\
M +_l N & \equiv_{\text{syn}} & M +_l P & \text{if } l \in \mathcal{Z} \subset \mathcal{L}
\end{array}
$$

We assume that the sum is commutative and associative, and the effect on the labels is given by some functions γ, α_1 and α_2. Besides we also consider that it may be idempotent for some particular labels, and that some labels may render the right side of the sum irrelevant.

Note that without any assumption on the functions α_1 and α_2, a term of the form $M +_l (N +_{l'} P)$ can not necessarily be proven equivalent to $(M +_k N) +_{k'} P$ for some $k, k' \in \mathcal{L}$ by the associativity rule. But we always have

$$
\begin{aligned}
M +_l (N +_{l'} P) &\equiv_{\text{syn}} \left(P +_{\gamma(l')} N\right) +_{\gamma(l)} M \\
&\equiv_{\text{syn}} P +_{\alpha_1(\gamma(l'),\gamma(l))} \left(N +_{\alpha_2(\gamma(l'),\gamma(l))} M\right) \\
&\equiv_{\text{syn}} \left(M +_{\gamma(\alpha_2(\gamma(l'),\gamma(l)))} N\right) +_{\gamma(\alpha_1(\gamma(l'),\gamma(l)))} P.
\end{aligned}
$$

It would be natural to expect some conditions on the parameters γ, α, $\mathcal{I}$ and $\mathcal{Z}$. For instance we could expect the associativity rule and the construction above to be mutual inverses, or the function γ to be an involution. But the only property we will actually need on the labels is the one stated in proposition 1.0.0.1.

To have this property we will require the following relations for all $l, l' \in \mathcal{L}$:

$$
\begin{aligned}
\alpha_1(l,l') &= \alpha_1(l',l) \\
\alpha_2(l,l') &= \alpha_1(l',\gamma(\alpha_2(l',l))) \\
l &= \gamma(\alpha_2(l',\gamma(\alpha_2(l',l)))).
\end{aligned}
$$

With these we can prove that, in a sense, sums commute with sums.

Proposition 1.0.0.1. *For any terms M, M', N and N' and for any labels l and l' we have modulo commutativity and associativity:*

$$
(M +_{l'} M') +_l (N +_{l'} N') \equiv_{\text{syn}} (M +_l N) +_{l'} (M' +_l N').
$$

Proof. We have the following equivalences:

$$\begin{aligned}
&(M +_{l'} M') +_l (N +_{l'} N') \\
&\qquad\qquad \equiv_{\text{syn}} M +_{\alpha_1(l',l)} \Big(M' +_{\alpha_2(l',l)} (N +_{l'} N')\Big) \\
&\qquad\qquad \equiv_{\text{syn}} M +_{\alpha_1(l',l)} \Big((N +_{l'} N') +_{\gamma(\alpha_2(l',l))} M'\Big) \\
&\qquad\qquad \equiv_{\text{syn}} M +_{\alpha_1(l',l)} \Big(N +_{\alpha_1(l',\gamma(\alpha_2(l',l)))} \Big(N' +_{\alpha_2(l',\gamma(\alpha_2(l',l)))} M'\Big)\Big) \\
&\qquad\qquad \equiv_{\text{syn}} M +_{\alpha_1(l',l)} \Big(N +_{\alpha_1(l',\gamma(\alpha_2(l',l)))} \Big(M' +_{\gamma(\alpha_2(l',\gamma(\alpha_2(l',l))))} N'\Big)\Big) \\
&(M +_l N) +_{l'} (M' +_l N') \\
&\qquad\qquad \equiv_{\text{syn}} M +_{\alpha_1(l,l')} \Big(N +_{\alpha_2(l,l')} (M' +_l N')\Big)
\end{aligned}$$

□

We do not know if the required relations on the labels carry any particular meaning, so it would be more natural to add this property as an additional axiom on our equivalence. But we will want to study in detail the classes of terms and the β-reduction modulo $\equiv_{\text{syn}}$, so we want to keep its definition as simple as possible. For that reason we do not want to add an additional rule when the existing ones are already sufficient in the cases we are interested in.

In the probabilistic case we have $\mathcal{L} = [0;1]$, and the equivalence is:

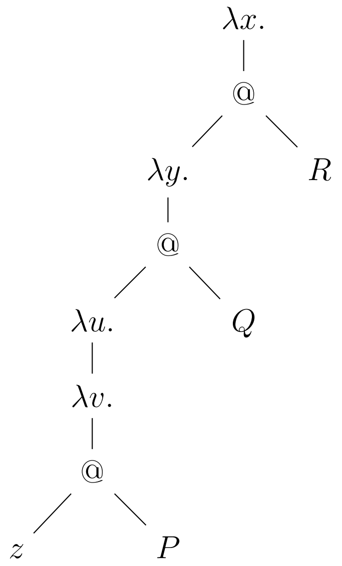
$$\begin{array}{rcll}
M +_p N & \equiv_{\text{syn}} & N +_{1-p} M & \\
(M +_p N) +_q P & \equiv_{\text{syn}} & M +_{pq} (N +_{\frac{(1-p)q}{1-pq}} P) & \text{if } pq \neq 1 \\
(M +_1 N) +_1 P & \equiv_{\text{syn}} & M +_1 (N +_{\frac{1}{2}} P) & \\
M +_p M & \equiv_{\text{syn}} & M & \\
M +_1 N & \equiv_{\text{syn}} & M +_1 P &
\end{array}$$

The purpose of this chapter is to study the calculus obtained by considering the β-reduction and the reduction of sums, both fully contextual, on terms modulo such a syntactic equivalence.

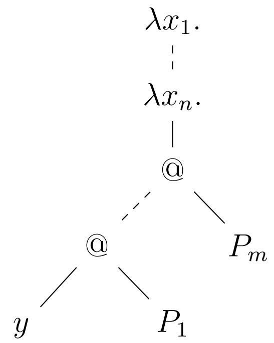
Let us first have a look at the head reduction in the probabilistic case, so that we can compare this calculus with the usual notion of probabilistic head reduction. Note that we try for now to justify the validity of our operational semantics as an alternative to the usual probabilistic one rather than to give precise results. For that reason we will not detail all the definitions and proofs.

In the deterministic λ-calculus, the head reduction can be understood as follows. The main idea is that you never look inside an argument: when a term is an application then the reduction only depends on the shape of its left side term. This way every term can be seen as a sequence of abstractions and applications of an argument over a variable.

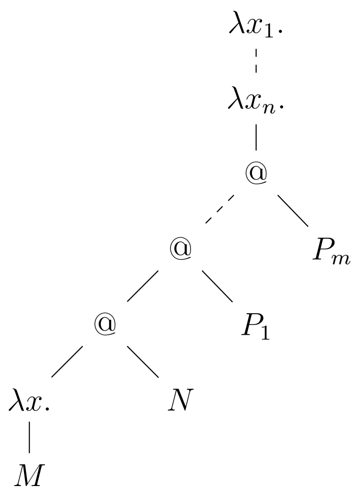
For instance the term $\lambda x.(\lambda y.(\lambda uv.z\ P)\ Q)\ R$ looks like



Now a β-redex is precisely an abstraction on the left of an application. So when you consider a term without looking inside the arguments, there are two possible cases. The first case is that you encounter no β-redex. This means that all abstractions are above the applications, and the term is of the form $\lambda x_1 ... x_n.y\ P_1\ ...\ P_m$.



The second case is that starting from the top we encounter a β-redex, so the term is of the form $\lambda x_1 ... x_n.(\lambda x.M)\ N\ P_1\ ...\ P_m$.



The principle of the head reduction is to only reduce such outer leftmost redexes. This way for every term we have either a unique finite reduction path

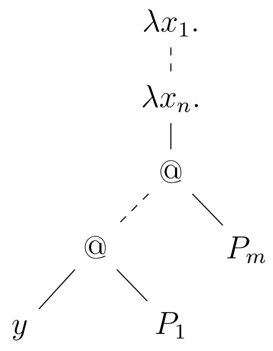
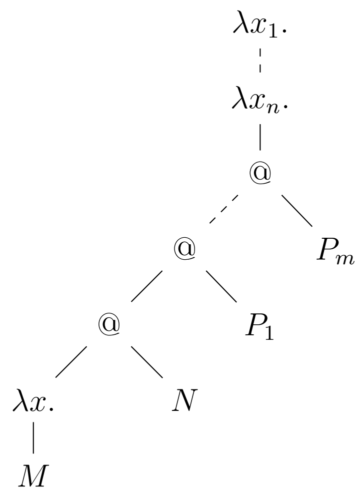
$$\begin{aligned}
\lambda \overrightarrow{x}_{1...n_1}.(\lambda z_1.M_1)\ N_1\ \overrightarrow{P}_{1,1...1,m_1} &\to_h \lambda \overrightarrow{x}_{1...n_2}.(\lambda z_2.M_2)\ N_2\ \overrightarrow{P}_{2,1...2,m_2} \\
&\to_h \cdots \\
&\to_h \lambda \overrightarrow{x}_{1...n_k}.(\lambda z_k.M_k)\ N_k\ \overrightarrow{P}_{k,1...k,m_k} \\
&\to_h \lambda \overrightarrow{x}_{1...n_{k+1}}.y_{k+1}\ \overrightarrow{P}_{k+1,1...k+1,m_{k+1}}
\end{aligned}$$

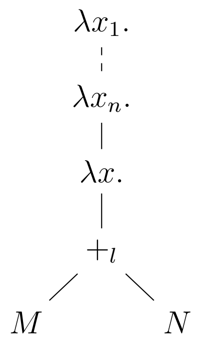
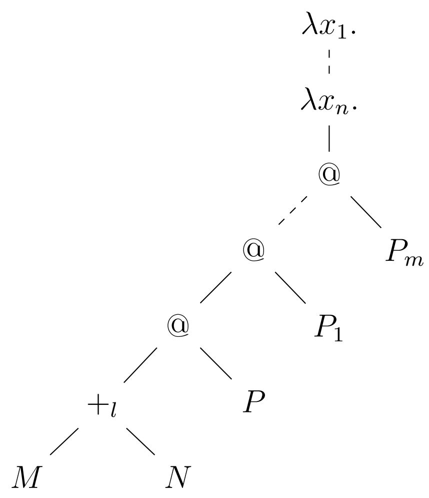
or a unique infinite reduction path

$$\begin{aligned}
\lambda \overrightarrow{x}_{1...n_1}.(\lambda z_1.M_1)\ N_1\ \overrightarrow{P}_{1,1...1,m_1} &\to_h \lambda \overrightarrow{x}_{1...n_2}.(\lambda z_2.M_2)\ N_2\ \overrightarrow{P}_{2,1...2,m_2} \\
&\to_h \cdots \\
&\to_h \lambda \overrightarrow{x}_{1...n_k}.(\lambda z_k.M_k)\ N_k\ \overrightarrow{P}_{k,1...k,m_k} \\
&\to_h \cdots
\end{aligned}$$

Terms of the form $\lambda \overrightarrow{x}_{1...n}.y\ \overrightarrow{P}_{1...m}$, which do not head reduce, are called *head normal forms*, and a term is said to have a head normal form if its head reduction terminates. Terms with a head normal form are also called *solvable* while terms without normal form are called *unsolvable*, as the latter can never be used to produce a result, i.e. a normal form.

Now in our λ-calculi with sums we added two reduction rules, and the new redexes are precisely a sum under an abstraction, and a sum on the left side of an application. So this time every term can be seen as a tree of sums, followed by a sequence of abstractions, then a sequence of applications, and finally either a variable or a redex. In other words every term is either a sum of terms or in one of the following forms:

Thus it is natural to define the head reduction by

$$\frac{M \to_{h+} M'}{M +_l N \to_{h+} M' +_l N} \qquad \frac{N \to_{h+} N'}{M +_l N \to_{h+} M +_l N'}$$

$$\frac{}{\lambda \overrightarrow{x}_{1\ldots n}.(\lambda x.M)\ N\ \overrightarrow{P}_{1\ldots m} \to_{h+} \lambda \overrightarrow{x}_{1\ldots n}.M\left[{}^{N}/_{x}\right]\ \overrightarrow{P}_{1\ldots m}}$$

$$\frac{}{\lambda \overrightarrow{x}_{1\ldots n}.\lambda x.(M +_l N) \to_{h+} \lambda \overrightarrow{x}_{1\ldots n}.(\lambda x.M +_l \lambda x.N)}$$

$$\frac{}{\lambda \overrightarrow{x}_{1\ldots n}.(M +_l N)\ P\ \overrightarrow{P}_{1\ldots m} \to_{h+} \lambda \overrightarrow{x}_{1\ldots n}.(M\ P +_l N\ P)\ \overrightarrow{P}_{1\ldots m}}$$

We can observe that this reduction has the following properties.

- If $\lambda \overrightarrow{x}_{1\ldots n}.(\lambda x.M)\ N\ \overrightarrow{P}_{1\ldots m} \to_{h+}^{+} Q$ then $\lambda \overrightarrow{x}_{1\ldots n}.M\left[{}^{N}/_{x}\right]\ \overrightarrow{P}_{1\ldots m} \twoheadrightarrow_{h+} Q$.
- If $\lambda \overrightarrow{x}_{1\ldots n}.(M +_l N)\ P\ \overrightarrow{P}_{1\ldots m} \to_{h+}^{+} Q$ then $\lambda \overrightarrow{x}_{1\ldots n}.M\ P\ \overrightarrow{P}_{1\ldots m} +_l \lambda \overrightarrow{x}_{1\ldots n}.N\ P\ \overrightarrow{P}_{1\ldots m} \twoheadrightarrow_{h+} Q$.

- $M +_l N \twoheadrightarrow_{h+} Q$ if and only if there are M' and N' such that $Q = M' +_l N'$ and $M \twoheadrightarrow_{h+} M'$, $N \twoheadrightarrow_{h+} N'$.

In the probabilistic case this reduction behaves just as the usual probabilistic head reduction, except that when we should make a choice by reducing a sum we keep both results, and every further reduction step is a reduction in one of these two terms. Another small difference is that we detail more the reduction of the sums in head position: the reduction

$$\lambda \overrightarrow{x}_{1\ldots n}.(M +_l N)\ P\ \overrightarrow{P}_{1\ldots m} \twoheadrightarrow_{h+} \lambda \overrightarrow{x}_{1\ldots n}.M\ P\ \overrightarrow{P}_{1\ldots m} +_l \lambda \overrightarrow{x}_{1\ldots n}.N\ P\ \overrightarrow{P}_{1\ldots m}$$

occurs in $m+n$ steps rather than just one, but these are the only possible $m+n$ first head reduction steps of this term so this decomposition of the reduction does not induce any odd behaviour.

A property this reduction lacks in comparison to the usual deterministic head reduction is the uniqueness of the reduction. It is easy to see that the reduction $\rightarrow_{h+}$ is confluent, and we can also prove that if a term is normalizing then it is strongly normalizing. But if we consider terms which from a probabilistic point of view may reach a head normal form but may also diverge, this head reduction is not satisfactory. For instance if we consider $\Omega = (\lambda x.x\ x)\ (\lambda x.x\ x)$ the well-known term such that $\Omega \rightarrow_\beta \Omega$, we have

$$(\lambda x.x)\ y +_p \Omega \rightarrow_{h+} (\lambda x.x)\ y +_p \Omega \rightarrow_{h+} (\lambda x.x)\ y +_p \Omega \rightarrow_{h+} \ldots$$

but this reduction path never accounts for the probabilistic reduction

$$(\lambda x.x)\ y +_p \Omega \xrightarrow{p} (\lambda x.x)\ y \xrightarrow{1} y.$$

To solve this problem we can define a *complete* head reduction $\rightarrow_{H+}$ which always reduces both sides of the sums.

$$\frac{M \rightarrow_{H+} M' \qquad N \rightarrow_{H+} N'}{M +_l N \rightarrow_{H+} M' +_l N'}$$

$$\frac{}{\lambda \overrightarrow{x}_{1\ldots n}.y\ \overrightarrow{P}_{1\ldots m} \rightarrow_{H+} \lambda \overrightarrow{x}_{1\ldots n}.y\ \overrightarrow{P}_{1\ldots m}}$$

$$\frac{}{\lambda \overrightarrow{x}_{1\ldots n}.(\lambda x.M)\ N\ \overrightarrow{P}_{1\ldots m} \rightarrow_{H+} \lambda \overrightarrow{x}_{1\ldots n}.M\left[{}^{N}/{}_{x}\right]\ \overrightarrow{P}_{1\ldots m}}$$

$$\frac{}{\lambda \overrightarrow{x}_{1\ldots n}.\lambda x.(M +_l N) \rightarrow_{H+} \lambda \overrightarrow{x}_{1\ldots n}.(\lambda x.M +_l \lambda x.N)}$$

$$\frac{}{\lambda \overrightarrow{x}_{1\ldots n}.(M +_l N)\ P\ \overrightarrow{P}_{1\ldots m} \rightarrow_{H+} \lambda \overrightarrow{x}_{1\ldots n}.(M\ P +_l N\ P)\ \overrightarrow{P}_{1\ldots m}}$$

Every term has a unique reduction for $\rightarrow_{H+}$, and this reduction never misses a head normal form: if M has a probabilistic reduction to a head normal form $h = \lambda x_1 \ldots x_n.y\ P_1\ \ldots\ P_m$ then there is a reduction $M \twoheadrightarrow_{H+} N$ such that h appears in N. A minor drawback is that there is strictly speaking no normal form for this reduction, as a head normal form always reduces into itself. Another problem with this idea of a complete reduction which always reduces all head redexes is that when we want to compare two terms, we may have a problem of coordination. Consider for instance the term $\delta(\lambda x.\delta x)$ where $\delta = \lambda x.x\ x$. This term is β-equivalent to Ω and we have

$$\delta(\lambda x.\delta x) \to_{H+} (\lambda x.\delta x)\ (\lambda x.\delta x) \to_{H+} \delta(\lambda x.\delta x).$$

Now if we consider the term $\delta(\lambda x.\delta x) +_l (\lambda x.\delta x)\ (\lambda x.\delta x)$, its reduction for $\to_{H+}$ is

$$\begin{aligned}\delta(\lambda x.\delta x) +_l (\lambda x.\delta x)\ (\lambda x.\delta x) &\to_{H+} (\lambda x.\delta x)\ (\lambda x.\delta x) +_l \delta(\lambda x.\delta x)\\ &\to_{H+} \delta(\lambda x.\delta x) +_l (\lambda x.\delta x)\ (\lambda x.\delta x)\\ &\to_{H+} \cdots\end{aligned}$$

We have a sum of two similar terms, and if the sum is $l \in \mathcal{I}$ (hence $M +_l M \equiv_{\text{syn}} M$ for all M) then we have

$$\begin{aligned}\delta(\lambda x.\delta x) +_l (\lambda x.\delta x)\ (\lambda x.\delta x) &\to_{h+} \delta(\lambda x.\delta x) +_l \delta(\lambda x.\delta x)\\ &\equiv_{\text{syn}} \delta(\lambda x.\delta x)\end{aligned}$$

but the complete head reduction never lets the two sides of the sums reduce into the same term.

Of course this situation only appears with infinite reductions. If two terms can reach the same head normal form then the complete reduction works just fine to compare them. This reduction is actually very useful to associate to a term a probability distribution over head normal forms. This will be detailed in the third chapter.

In the rest of the thesis we will not use exactly this definition of the head reduction. The reason is that we will see the commutation of the sum with the abstraction and the application as a syntactic equivalence rather than as a reduction, and the main reduction rule will be the β-rule.

1.1 λ-terms with sums

For now let us forget about the β-reduction and look at the reduction $\to_+$ along with the equivalence $\equiv_{\mathrm{syn}}$.

1.1.1 Reduction of sums

The reduction $\to_+$ is contextual, so we can associate a congruence $=_+$ to it. This congruence is very easy to describe: indeed we will prove that not only is the reduction $\to_+$ confluent (thus $M =_+ N$ if and only if $M \twoheadrightarrow_+ \cdot \twoheadleftarrow_+ N$) but it is also strongly normalizing.

Proposition 1.1.1.1. $\to_+$ *is weakly confluent.*

Proof. If $M \to_+ N_1$ and $M \to_+ N_2$ with $N_1 \neq N_2$, we reason by induction on M:

- if $\lambda x.M \to_+ \lambda x.N_i$ with $M \to_+ N_i$ for $i \in \{1,2\}$ then we conclude by induction hypothesis;
- similarly if we reduce $M\ M'$ or $M +_l M'$ and the reductions occur both in M or both in M' we conclude by induction hypothesis;
- if $M\ M' \to_+ N_1\ M'$ and $M\ M' \to_+ M\ N_2$ then $N_1\ M' \to_+ N_1\ N_2$ and $M\ N_2 \to_+ N_1\ N_2$;
- similarly if $M +_l M' \to_+ N_1 +_l M'$ and $M +_l M' \to_+ M +_l N_2$ then both reduce in $N_1 +_l N_2$;
- if $\lambda x.(M +_l M') \to_+ \lambda x.M +_l \lambda x.M'$ and $\lambda x.(M +_l M') \to_+ \lambda x.(N +_l M')$ with $M \to_+ N$ then both terms reduce in $\lambda x.N +_l \lambda x.M'$; the case where the second reduction occurs in M' is similar;
- if $(M +_l M')\ P \to_+ M\ P +_l M'\ P$ and we can reduce either M or M' we get a similar result;
- if $(M +_l M')\ P \to_+ M\ P +_l M'\ P$ and $(M +_l M')\ P \to_+ (M +_l M')\ Q$ with $P \to_+ Q$ then $M\ P +_l M'\ P \to_+ M\ Q +_l M'\ P \to_+ M\ Q +_l M'\ Q$ and $(M +_l M')\ Q \to_+ M\ Q +_l M'\ Q$.

$\square$

Proposition 1.1.1.2. $\to_+$ *is strongly normalizing.*

Proof. Is is easy to see that the depth of the sums decrease along the reduction but when we reduce $(M +_l N)\ P \to_+ M\ P +_l N\ P$ we may duplicate sums. We want to define a weight which strictly decreases when we perform a reduction.

We can observe that sums do not reduce when they are on the right side of an application, and that the reduction $\to_+$ only duplicates sums inside an argument. Then it makes sense to define the depth of a subterm as the number of times it is on the right side of an application. This notion of depth is such that the depth of a sum never changes when we perform a reduction, and the reduction of a sum at given depth can only duplicate strictly deeper sums. So a possible weight is the sequence in $\mathbb{N}^{\mathbb{N}}$ whose d-th element is the total number of reductions required to normalize the term at depth d, equipped with the lexicographic order.

Another possible weight function, maybe less intuitive but easier to formalize, is $w : \Lambda_+^{\mathcal{L}} \times \mathbb{N} \to \mathbb{N}$ defined by:

- $w(x, d) = 0$;
- $w(\lambda x.M, d) = w(M, d+1)$;
- $w(M\ N, d) = w(M, d+1) + (w(M, 1) + 1) \times w(N, 0)$;
- $w(M +_l N, d) = w(M, d) + w(N, d) + d$.

The parameter d counts the number of times a sum needs to be reduced. Given a term M the weight $w(M, 1)$ is at least the number of sums at depth 0 in M, so in a term $M\ N$ the argument N will de duplicated at most $w(M, 1)$ times.

The weight of a term is a strictly increasing function of the weight of its subterms in the following sense: for any context C, if $w(M, d) < w(N, d)$ for all $d \in \mathbb{N}$ then $w(C[M], d) < w(C[N], d)$ for all $d \in \mathbb{N}$. We prove this by induction on C:

- if $C = [\ ]$ this is immediate;
- if $C = \lambda x.C'$ then for all d and M, $w(C[M], d) = w(C'[M], d+1)$ so we apply the induction hypothesis to C';
- if $C = C'\ P$ then $w(C[M], d) = w(C'[M], d+1) + (w(C'[M], 1) + 1) \times w(P, 0)$ so again the induction hypothesis applied to C' gives the result;
- if $C = P\ C'$ then $w(C[M], d) = w(P, d+1) + (w(P, 1) + 1) \times w(C[M], 0)$, we have $w(P, 1) + 1 > 0$ so the induction hypothesis allows us to conclude;
- the result is also immediate by induction hypothesis if C is a sum.

If we consider the weight of the redexes of $\to_+$ and their reducts we have

$$\begin{aligned} w(\lambda x.(M +_l N), d) &= w(M, d+1) + w(N, d+1) + d + 1 \\ w(\lambda x.M +_l \lambda x.N) &= w(M, d+1) + w(N, d+1) + d \end{aligned}$$

$$\begin{aligned} w((M +_l N)\ P, d) &= w(M, d+1) + w(N, d+1) \\ &\quad + (w(M, 1) + w(N, 1) + 2) \times w(P, 0) + d + 1 \\ w(M\ P +_l N\ P, d) &= w(M, d+1) + w(N, d+1) \\ &\quad + (w(M, 1) + w(N, 1) + 2) \times w(P, 0) + d \end{aligned}$$

For all $d \in \mathbb{N}$ we have $w(\lambda x.(M +_l N), d) > w(\lambda x.M +_l \lambda x.N)$ and $w((M +_l N)\ P, d) > w(M\ P +_l N\ P, d)$.

Usinge these two facts we get that if $M \to_+ N$ then for all d, $w(M,d) > w(N,d)$, hence the reduction $\to_+$ is strongly normalizing. □

Corollary 1.1.1.3. *$\to_+$ is confluent.*

Corollary 1.1.1.4. *1. Every term has a unique normal form for $\to_+$.*

2. Two terms are equal for $=_+$ if and only if they have the same normal form.

Proof. $\to_+$ is strongly normalizing so every term has a normal form, and the confluence ensures that if a term has two normal forms then they are equal.

The confluence also ensures that $M =_+ N$ if and only if $M \twoheadrightarrow_+ \cdot \twoheadleftarrow_+ N$, and this is equivalent to saying that M and N have the same normal form. □

Every class of terms modulo $=_+$ is represented by a unique normal form. These forms are actually easy to describe, as they have a convenient inductive structure.

Definition 1.1.1.1. The *canonical terms* M, N and *values* v are defined by:

$$\begin{aligned} M, N &:= v \mid M +_l N \\ v &:= x \mid \lambda x.v \mid v\ M. \end{aligned}$$

Proposition 1.1.1.5. *The canonical terms are exactly the normal forms for $\to_+$.*

Proof. First we check by induction that the canonical terms are indeed normal.

- $M +_l N$ is not a redex, and by induction hypothesis M and N are normal.
- x is normal.
- $\lambda x.v$ is a redex only if v is a sum, which is not possible, and by induction hypothesis v is normal.
- $v\ M$ is a redex only if v is a sum, which is not possible, and by induction hypothesis v and M are normal.

Conversely every normal form is a canonical term.

- x is canonical.
- If $\lambda x.M$ is normal then M is normal, hence by induction hypothesis it is canonical, and M is not a sum so it is necessarily a value.
- Similarly if $M\ N$ is normal then M and N are normal, hence canonical, and M is not a sum.
- If $M +_l N$ is normal then M and N are normal and by induction hypothesis they are canonical forms.

□

Definition 1.1.1.2. For all M we write $\mathrm{can}(M)$ the unique canonical (i.e. normal) form of M.

The canonicalizing reduction $\to_{\mathrm{can}}$ is defined by $M \to_{\mathrm{can}} \mathrm{can}(M)$. Note that this reduction is not extended to context: every term has exactly one reduction for $\to_{\mathrm{can}}$.

Since every term has a unique canonical form and those are easily described we are tempted to restrict the calculus to these terms. A problem is that they are not stable by context: if M and N are canonical then $M\ N$ is not necessarily canonical.

Still they are very useful to describe our calculus. We will now see how the canonicalcalization interacts with the other features of the calculus, namely the β-reduction and the syntactic equivalence $\equiv_{\mathrm{syn}}$. For now let us consider this canonicalization modulo $\equiv_{\mathrm{syn}}$.

1.1.2 Terms modulo $\equiv$

We want to prove that the canonicalization preserves the syntactic equivalence $\equiv_{\mathrm{syn}}$: if $M \equiv_{\mathrm{syn}} N$ then $\mathrm{can}(M) \equiv_{\mathrm{syn}} \mathrm{can}(N)$.

To describe precisely what goes on when we consider terms modulo $\equiv_{\mathrm{syn}}$ let us consider the four following reductions, extended to context:

$$\begin{array}{rcll} M +_l N & \to_\gamma & N +_{\gamma(l)} M & \gamma : \mathcal{L} \to \mathcal{L} \\ (M +_l N) +_{l'} P & \to_\alpha & M +_{\alpha_1(l,l')} (N +_{\alpha_2(l,l')} P) & \alpha_1, \alpha_2 : \mathcal{L}^2 \to \mathcal{L} \\ M +_l M & \to_{\mathcal{I}} & M & \text{if } l \in \mathcal{I} \subset \mathcal{L} \\ M +_l N & \to_{\mathcal{Z}} & M +_l P & \text{if } l \in \mathcal{Z} \subset \mathcal{L} \end{array}$$

We will want to prove some commutation between these reductions and $\to_+$. Since some of these reductions duplicate terms we need to use a parallel reduction. Given a reduction $\to$ we define the associated independent parallel reduction $\xrightarrow{i}$ as the reduction of an arbitrary number of disjoint redexes.

$$\frac{}{M \xrightarrow{i} M} \qquad \frac{M \to N}{M \xrightarrow{i} N}$$

$$\frac{M \xrightarrow{i} N}{\lambda x.M \xrightarrow{i} \lambda x.N} \qquad \frac{M \xrightarrow{i} N \qquad M' \xrightarrow{i} N'}{M\ M' \xrightarrow{i} N\ N'} \qquad \frac{M \xrightarrow{i} N \qquad M' \xrightarrow{i} N'}{M +_l M' \xrightarrow{i} N +_l N'}$$

Now we can prove that the syntactic reductions commute with $\to_+$.

Lemma 1.1.2.1. *For $\to$ any of the reductions $\to_\gamma$, $\leftarrow_\gamma$, $\to_\alpha$, $\leftarrow_\alpha$, $\to_{\mathcal{I}}$, $\leftarrow_{\mathcal{I}}$, and $\to_{\mathcal{Z}}$, if $M \xrightarrow{i} N \to_+ N'$ then there is a reduction $N' \twoheadrightarrow_+ N''$ and a term M' such that $M \twoheadrightarrow_+ M' \xrightarrow{i} N''$.*

$$\begin{tikzcd} M \arrow[rr, dashed, two heads, "+"] \arrow[d, "i"'] & & M' \arrow[d, dashed, "i"] \\ N \arrow[r, "+"'] & N' \arrow[r, dashed, two heads, "+"'] & N'' \end{tikzcd}$$

Proof. By induction on the reductions. If both reductions use context rules we easily conclude by induction hypothesis. For instance if $\lambda x.M \xrightarrow{i} \lambda x.N \to_+ \lambda x.N'$ with $M \xrightarrow{i} N \to_+ N'$ then by induction hypothesis we have $N' \twoheadrightarrow_+ N''$ and $M \twoheadrightarrow_+ M' \xrightarrow{i} N''$ so $\lambda x.N' \twoheadrightarrow_+ \lambda x.N''$ and $\lambda x.M \twoheadrightarrow_+ \lambda x.M' \xrightarrow{i} \lambda x.N''$.

We just have to check all the different cases when M or N is a redex. If M is a redex we have seven different cases for the first reduction $M \xrightarrow{i} N$ (i.e. $M \to N$ as M is directly a redex).

- $M_1 +_l M_2 \to_\gamma M_2 +_{\gamma(l)} M_1$;
- $M_2 +_{\gamma(l)} M_1 \leftarrow_\gamma M_1 +_l M_2$;
- $(M_1 +_l M_2) +_{l'} M_3 \to_\alpha M_1 +_{\alpha_1(l,l')} (M_2 +_{\alpha_2(l,l')} M_3)$;
- $M_1 +_{\alpha_1(l,l')} (M_2 +_{\alpha_2(l,l')} M_3) \leftarrow_\alpha (M_1 +_l M_2) +_{l'} M_3$;
- $M_1 +_l M_1 \to_{\mathcal{I}} M_1$;
- $M_1 \leftarrow_{\mathcal{I}} M_1 +_l M_1$;
- $M_1 +_l M_2 \to_{\mathcal{Z}} M_1 +_l M_3$.

Note that no abstraction or application appears here, so the reduction $N \to_+ N'$ necessarily occurs in one of the M_i.

- In the first four cases the reductions obviously commute. For instance if

  $$M_1 +_l M_2 \to_\gamma M_2 +_{\gamma(l)} M_1 \to_+ M_2 +_{\gamma(l)} M_1'$$

  with $M_1 \to_+ M_1'$ then

  $$M_1 +_l M_2 \to_+ M_1' +_l M_2 \to_\gamma M_2 +_{\gamma(l)} M_1'.$$

- If $M_1 +_l M_1 \to_{\mathcal{I}} M_1 \to_+ M_1'$ then we have

  $$M_1 +_l M_1 \to_+ M_1' +_l M_1 \to_+ M_1' +_l M_1' \to_{\mathcal{I}} M_1'.$$

- If $M_1 \leftarrow_{\mathcal{I}} M_1 +_l M_1 \to_+ M_1' +_l M_1$ we have $M_1' +_l M_1 \to_+ M_1' +_l M_1'$ and $M_1 \to_+ M_1' \leftarrow_{\mathcal{I}} M_1' +_l M_1'$. The case is similar if we reduce the term on the right side of the sum.

- If $M_1 +_l M_2 \to_{\mathcal{Z}} M_1 +_l M_3 \to_+ M_1' +_l M_3$ we have

  $$M_1 +_l M_2 \to_+ M_1' +_l M_2 \to_{\mathcal{Z}} M_1' +_l M_3.$$

- If $M_1 +_l M_2 \to_{\mathcal{Z}} M_1 +_l M_3 \to_+ M_1 +_l M_3'$ we can reduce directly

  $$M_1 +_l M_2 \to_{\mathcal{Z}} M_1 +_l M_3'.$$

The remaining case is when M is not directly reduced but N is. This means that the reduction $N \to_+ N'$ is either $\lambda x.(N_1 +_l N_2) \to_+ \lambda x.N_1 +_l \lambda x.N_2$ or $(N_1 +_l N_2)\ Q \to_+ N_1\ Q +_l N_2\ P$. Then necessarily $M = \lambda x.M_0$ or $M = M_0\ P$ with $M_0 \xrightarrow{i} N_1 +_l N_2$ and $P \xrightarrow{i} Q$.

If $M_0 = M_1 +_l M_2$ with $M_j \xrightarrow{i} N_j$ for $j \in \{1;2\}$ then the result is immediate. Otherwise there are seven possible cases, that we only detail for $N = \lambda x.(N_1 +_l N_2)$:

- if $M_0 = N_2 +_{l'} N_1 \to_\gamma N_1 +_{\gamma(l')} N_2$ with $\gamma(l') = l$ then

$$\lambda x.M_0 \to_+ \lambda x.N_2 +_{l'} \lambda x.N_1 \to_\gamma \lambda x.N_1 +_l \lambda x.N_2;$$

- the same goes if $M_0 = N_2 +_{\gamma(l)} N_1 \leftarrow_\gamma N_1 +_l N_2$;
- if

$$M_0 = (N_1 +_{l'} M_1) +_{l''} M_2 \to_\alpha N_1 +_{\alpha_1(l',l'')} (M_1 +_{\alpha_2(l',l'')} M_2)$$

  with $\alpha_1(l',l'') = l$ and $M_1 +_{\alpha_2(l',l'')} M_2 = N_2$ then we have

$$\lambda x.N_1 +_l \lambda x.N_2 \to_+ \lambda x.N_1 +_l (\lambda x.M_1 +_{\alpha_2(l',l'')} \lambda x.M_2)$$

  and

$$\lambda x.M_0 \twoheadrightarrow_+ (\lambda x.N_1 +_{l'} \lambda x.M_1) +_{l''} \lambda x.M_2 \to_\alpha \lambda x.N_1 +_l (\lambda x.M_1 +_{\alpha_2(l',l'')} \lambda x.M_2);$$

- the same works for the reverse reduction $\leftarrow_\alpha$;
- if $M_0 = (N_1 +_l N_2) +_{l'} (N_1 +_l N_2)$ with $l' \in \mathcal{I}$ then

$$\lambda x.M_0 \twoheadrightarrow_+ (\lambda x.N_1 +_l \lambda x.N_2) +_{l'} (\lambda x.N_1 +_l \lambda x.N_2) \to_{\mathcal{I}} \lambda x.N_1 +_l \lambda x.N_2;$$

- if $M_0 = N_1 = N_2$ then we directly have $\lambda x.M_0 \leftarrow_{\mathcal{I}} \lambda x.N_1 +_l \lambda x.N_2$;
- if $M_0 = N_1 +_l M_2 \to_{\mathcal{Z}} N_1 +_l N_2$ then

$$\lambda x.M_0 \to_+ \lambda x.N_1 +_l \lambda x.M_2 \to_{\mathcal{Z}} \lambda x.N_1 +_l \lambda x.N_2.$$

This concludes the proof. □

Corollary 1.1.2.2. *For any reduction $\to$ of the previous lemma, if $M \xrightarrow{i} N$ then* $\mathrm{can}(M) \xrightarrow{i} \mathrm{can}(N)$.

Proof. First we prove that if $M \xrightarrow{i} N$ then there is a term M' such that $M \twoheadrightarrow_+ M' \xrightarrow{i} \mathrm{can}(N)$.

We know that the reduction $\to_+$ is strongly normalizing so for any term N the length of the reductions $N \twoheadrightarrow_+ \mathrm{can}(N)$ is bounded. We reason by induction on this bound. If N is canonical then the result is immediate.

Otherwise if $N \rightarrow_+ N'$ then according to the previous lemma we have $N' \twoheadrightarrow_+ N''$ and $M \twoheadrightarrow_+ M' \xrightarrow{i} N''$, and we can apply the induction hypothesis to N'' to get M'' such that $M' \twoheadrightarrow_+ M'' \xrightarrow{i} \mathrm{can}(N'') = \mathrm{can}(N)$.

Now this holds for the reduction $\rightarrow$ but also for the reverse reduction $\leftarrow$. Hence if $M \xrightarrow{i} N$ we have $M \twoheadrightarrow_+ M' \xrightarrow{i} \mathrm{can}(N)$ but also $\mathrm{can}(N) \twoheadrightarrow_+ N' \xleftarrow{i} \mathrm{can}(M')$. We have $\mathrm{can}(M') = \mathrm{can}(M)$, and necessarily $N' = \mathrm{can}(N)$, so $\mathrm{can}(M) \xrightarrow{i} \mathrm{can}(N)$. $\square$

Proposition 1.1.2.3. *If $M \equiv_{\mathrm{syn}} N$ then $\mathrm{can}(M) \equiv_{\mathrm{syn}} \mathrm{can}(N)$.*

Proof. The equivalence $\equiv_{\mathrm{syn}}$ is the reflexive and transitive closure of the union of the eight reductions $\rightarrow_\gamma$, $\rightarrow_\alpha$, $\rightarrow_{\mathcal{I}}$ and $\rightarrow_{\mathcal{Z}}$ and their symmetric reductions. Thus if $M \equiv_{\mathrm{syn}} N$ we have $M\ (\leftrightarrow_\gamma \cup \leftrightarrow_\alpha \cup \leftrightarrow_{\mathcal{I}} \cup \rightarrow_{\mathcal{Z}})^*\ N$ and

$$\mathrm{can}(M)\ (\leftrightarrow_\gamma \cup \leftrightarrow_\alpha \cup \leftrightarrow_{\mathcal{I}} \cup \rightarrow_{\mathcal{Z}})^*\ \mathrm{can}(N).$$

$\square$

The equality $=_+$ behaves properly with respect to $\equiv_{\mathrm{syn}}$. When we want to consider terms modulo both $=_+$ and $\equiv_{\mathrm{syn}}$ we actually only need to consider canonical terms modulo $\equiv_{\mathrm{syn}}$.

But observe that if M is a canonical term and we have $M \equiv_{\mathrm{syn}} N$ then N is not necessarily canonical. We have $v \equiv_{\mathrm{syn}} v +_l v$ for all canonical value v and $l \in \mathcal{I}$ so we may create redexes for $\rightarrow_+$. For this reason we give an alternate definition of $\equiv_{\mathrm{syn}}$ for canonical terms.

Definition 1.1.2.1. The relations $\equiv^c_{\mathrm{syn}}$ on canonical terms and $\equiv^v_{\mathrm{syn}}$ on values are defined as follows.

$$\frac{M \equiv^c_{\mathrm{syn}} N}{N \equiv^c_{\mathrm{syn}} M} \qquad \frac{M \equiv^c_{\mathrm{syn}} N \qquad N \equiv^c_{\mathrm{syn}} P}{M \equiv^c_{\mathrm{syn}} P}$$

$$\frac{}{M +_l N \equiv^c_{\mathrm{syn}} N +_{\gamma(l)} M} \qquad \frac{}{(M +_l N) +_{l'} P \equiv^c_{\mathrm{syn}} M +_{\alpha_1(l,l')} (N +_{\alpha_2(l,l')} P)}$$

$$\frac{l \in \mathcal{I}}{M +_l M \equiv^c_{\mathrm{syn}} M} \qquad \frac{l \in \mathcal{Z}}{M +_l N \equiv^c_{\mathrm{syn}} M +_l P}$$

$$\frac{M \equiv^c_{\mathrm{syn}} M'}{M +_l N \equiv^c_{\mathrm{syn}} M' +_l N} \qquad \frac{N \equiv^c_{\mathrm{syn}} N'}{M +_l N \equiv^c_{\mathrm{syn}} M +_l N'} \qquad \frac{v \equiv^v_{\mathrm{syn}} v'}{v \equiv^c_{\mathrm{syn}} v'}$$

$$\frac{}{x \equiv^v_{\mathrm{syn}} x} \qquad \frac{v \equiv^v_{\mathrm{syn}} v'}{\lambda x.v \equiv^v_{\mathrm{syn}} \lambda x.v'} \qquad \frac{v \equiv^v_{\mathrm{syn}} v' \qquad M \equiv^c_{\mathrm{syn}} M'}{v\ M \equiv^v_{\mathrm{syn}} v'\ M'}$$

Proposition 1.1.2.4. *If M and N are canonical then $M \equiv_{\mathrm{syn}} N$ if and only if $M \equiv^c_{\mathrm{syn}} N$.*

Proof. If $M \equiv^c_{\text{syn}} N$ then we obviously have $M \equiv_{\text{syn}} N$.

Conversely if $M \to_\gamma N$, $M \to_\alpha N$, $M \to_{\mathcal{I}} N$ or $M \to_{\mathcal{Z}} N$ then we have $M \equiv^c_{\text{syn}} N$. We prove this by induction on the canonical structure of M and N. As $\equiv^c_{\text{syn}}$ is symmetric and transitive this is enough to conclude. □

Corollary 1.1.2.5. *Given any terms M and N we have $M \equiv_{\text{syn}} N$ if and only if* $\text{can}(M) \equiv^c_{\text{syn}} \text{can}(N)$.

Now let us consider the relation built on the rules of $=_+$ and $\equiv_{\text{syn}}$.

Definition 1.1.2.2. The relation $\equiv$ is the least congruence which contains $=_+$ and $\equiv$.

Corollary 1.1.2.6. $M \equiv N$ *if and only if* $\text{can}(M) \equiv^c_{\text{syn}} \text{can}(N)$.

The equivalence $\equiv_{\text{syn}}$ only involves sums, and it was introduced precisely to define the meaning of this constructor. On the other hand canonical forms are terms where the sums are gathered on top of the arguments. So for the different interpretations of the sum we can give a very satisfying characterization of the classes of terms modulo $\equiv$.

We will only detail the case of the probabilistic λ-calculus. We will prove that trees of sums represent probability distributions. More precisely we want to prove the following theorem.

Theorem 1.1.2.7. *The classes of probabilistic terms modulo $\equiv$ are*

$$\begin{aligned} \mathcal{V} \in \Lambda_+^{[0;1]}/_{\equiv} &:= \textit{ finite probability distributions over values } V \\ V &:= x \mid \lambda x.V \mid V\ \mathcal{V}. \end{aligned}$$

According to the previous corollary the classes of terms modulo $\equiv$ are exactly the classes modulo $\equiv^c_{\text{syn}}$ of the canonical terms. So all we need to do is to show that $\equiv^c_{\text{syn}}$ turns sums into probability distributions.

Definition 1.1.2.3. For every canonical term M we define a finite probability distribution $\mathcal{V}_M$ over values (with probability distributions), and for every canonical value v we define a value with probability distributions $\widetilde{v}$.

- $\mathcal{V}_v(\widetilde{v}) = 1$ and $\mathcal{V}_v(V) = 0$ otherwise;
- $\mathcal{V}_{M+_pN} = p \times \mathcal{V}_M + (1-p) \times \mathcal{V}_N$;
- $\widetilde{x} = x$;
- $\widetilde{\lambda x.v} = \lambda x.\widetilde{v}$;
- $\widetilde{v\ M} = \widetilde{v}\ \mathcal{V}_M$.

Proposition 1.1.2.8. *Given M and N two canonical terms, if $M \equiv^c_{\text{syn}} N$ then $\mathcal{V}_M = \mathcal{V}_N$.*

Proof. We prove this as well as the corresponding result for values, i.e. that if $v \equiv^v_{\text{syn}} w$ then $\widetilde{v} = \widetilde{w}$.

We reason by induction on $\equiv^c_{\text{syn}}$ and $\equiv^v_{\text{syn}}$. The basic cases are given by the following equalities:

- $p\mathcal{V}_M + (1-p)\mathcal{V}_N = (1-p)\mathcal{V}_N + (1-(1-p))\mathcal{V}_M$;
- $q\left(p\mathcal{V}_M + (1-p)\mathcal{V}_N\right) + (1-q)\mathcal{V}_P = pq\mathcal{V}_M + (1-pq)\left(\frac{(1-p)q}{1-pq}\mathcal{V}_N + \left(1 - \frac{(1-p)q}{1-pq}\right)\mathcal{V}_P\right)$;
- $(\mathcal{V}_M + 0 \times \mathcal{V}_N) + 0 \times \mathcal{V}_P = \mathcal{V}_M + 0 \times \left(\frac{1}{2} \times \mathcal{V}_N + \frac{1}{2} \times \mathcal{V}_P\right)$;
- $p\mathcal{V}_M + (1-p)\mathcal{V}_M = \mathcal{V}_M$;
- $\mathcal{V}_M + 0 \times \mathcal{V}_N = \mathcal{V}_M + 0 \times \mathcal{V}_P$.

□

We proved that we indeed have a transformation from the classes of terms modulo $\equiv$ to a syntax with probability distributions. Now we need to prove that this transformation is a bijection.

Lemma 1.1.2.9. *Given a canonical term M and a value v we can find canonical terms M_v and M' such that $M \equiv^c_{\text{syn}} M_v +_{\mathcal{V}_M(\widetilde{v})} M'$ and $\mathcal{V}_{M_v} = \mathcal{V}_{\widetilde{v}}$.*

Proof. By induction on M.

- If $M = w$ with $\widetilde{w} = \widetilde{v}$ then $\mathcal{V}_M(\widetilde{v}) = 1$ and $M \equiv^c_{\text{syn}} w +_1 M$.
- If $M = w$ with $\widetilde{w} \neq \widetilde{v}$ then $\mathcal{V}_M(\widetilde{v}) = 0$ and we have $M \equiv^c_{\text{syn}} v +_0 M$.
- If $M = M_1 +_p M_2$ then $\mathcal{V}_M(\widetilde{v}) = p\mathcal{V}_{M_1}(\widetilde{v}) + (1-p)\mathcal{V}_{M_2}(\widetilde{v})$, and by induction hypothesis we have $M_i \equiv^c_{\text{syn}} M_{i,v} +_{\mathcal{V}_{M_i}(\widetilde{v})} M'_i$ for $i \in \{1;2\}$. If $\mathcal{V}_M(\widetilde{v}) = 1$ we have $M \equiv^c_{\text{syn}} M_{1,v} +_p M_{2,v}$, and if $\mathcal{V}_M(\widetilde{v}) = 0$ we have $M \equiv^c_{\text{syn}} M'_1 +_p M'_2$. Otherwise we have

$$\begin{aligned} M &\equiv^c_{\text{syn}} \left(M_{1,v} +_{\mathcal{V}_{M_1}(\widetilde{v})} M'_1\right) +_p \left(M_{2,v} +_{\mathcal{V}_{M_2}(\widetilde{v})} M'_2\right) \\ &\equiv^c_{\text{syn}} \left(M_{1,v} +_{\frac{p\mathcal{V}_{M_1}(\widetilde{v})}{\mathcal{V}_M(\widetilde{v})}} M_{2,v}\right) +_{\mathcal{V}_M(\widetilde{v})} \left(M'_1 +_{\frac{p\left(1-\mathcal{V}_{M_1}(\widetilde{v})\right)}{1-\mathcal{V}_M(\widetilde{v})}} M'_2\right). \end{aligned}$$

□

Proposition 1.1.2.10. *The transformation is injective: for all terms M and N if $\mathcal{V}_M = \mathcal{V}_N$ then $M \equiv^c_{\text{syn}} N$.*

Proof. We prove by induction on $\mathcal{V}$ and V that if $\mathcal{V}_M = \mathcal{V}_N = \mathcal{V}$ then $M \equiv^c_{\text{syn}} N$, and if $\widetilde{v} = \widetilde{w} = V$ then $v \equiv^v_{\text{syn}} w$.

In the cases of values the result is always immediate by induction hypothesis, as values with probability distributions V have exactly the same structure as term values v.

Given a finite probability distribution $\mathcal{V}$, the induction hypothesis states that if $\mathcal{V}(V) \neq 0$ then whenever $\widetilde{v} = \widetilde{w} = V$ we have $v \equiv^v_{\text{syn}} w$.

First remark that for any value v such that $\mathcal{V}(\widetilde{v}) \neq 0$, for any canonical term M such that $\mathcal{V}_M = \mathcal{V}_v$, we have $M \equiv^c_{\text{syn}} v$. This is proven by induction on M:

- if M is a value then by hypothesis $M \equiv^v_{\text{syn}} v$;
- if $M = M_1 +_p M_2$ with $p \in]0;1[$ then $\mathcal{V}_{M_1} = \mathcal{V}_{M_2} = \mathcal{V}_v$ and by induction hypothesis $M \equiv^c_{\text{syn}} v +_p v \equiv^c_{\text{syn}} v$;
- if $M = M_1 +_1 M_2$ or $M = M_2 +_0 M_1$ then $M \equiv^c_{\text{syn}} M_1$ and by induction hypothesis $M_1 \equiv^c_{\text{syn}} v$.

In the general case we reason by induction on the number of values V such that $\mathcal{V}(V) \neq 0$. There is always at least one, and furthermore if there exists M such that $\mathcal{V}_M = \mathcal{V}$ then there exists v such that $\mathcal{V}(\widetilde{v}) \neq 0$.

- If $\mathcal{V} = \mathcal{V}_v$ then we proved that whenever $\mathcal{V}_M = \mathcal{V}$ then $M \equiv^c_{\text{syn}} v$ so the result is immediate.
- Otherwise $0 < \mathcal{V}(\widetilde{v}) < 1$, let

$$\mathcal{V}' : V \mapsto \begin{cases} 0 & \text{if } V = \widetilde{v} \\ \frac{\mathcal{V}(V)}{\mathcal{V}(\widetilde{v})} & \text{otherwise} \end{cases}.$$

  The previous lemma gives that if $\mathcal{V}_M = \mathcal{V}$ then $M \equiv^c_{\text{syn}} M_v +_{\mathcal{V}(\widetilde{v})} M'$ with $\mathcal{V}_{M_v} = \mathcal{V}_v$, hence $M_v \equiv^c_{\text{syn}} v$ and $M \equiv^c_{\text{syn}} v +_{\mathcal{V}(\widetilde{v})} M'$. Besides we have necessarily $\mathcal{V}_{M'} = \mathcal{V}'$, so we can conclude by induction hypothesis on $\mathcal{V}'$.

□

Proposition 1.1.2.11. *The transformation is surjective: for all $\mathcal{V}$ there exists M such that $\mathcal{V} = \mathcal{V}_M$.*

Proof. We reason by induction on $\mathcal{V}$. The cases of values are immediate:

- for a variable x we have $x = \widetilde{x}$;
- for an abstraction $\lambda x.V$ we have by induction hypothesis that $V = \widetilde{v}$ for some canonical value v, and $\lambda x.V = \widetilde{\lambda x.v}$;
- for an application $V\ \mathcal{V}$ we have a canonical value v and a canonical term M such that $\widetilde{v} = V$ and $\widetilde{M} = \mathcal{V}$, hence $\widetilde{v\ M} = V\ \mathcal{V}$.

Now if we have a finite probability distribution $\mathcal{V}$ over values, let $V_0,\ldots,V_n$ be the pairwise distinct values such that $\mathcal{V}(V_i) \neq 0$ for $i \leq n$ and $\mathcal{V}(V) = 0$ otherwise. By induction hypothesis we have $V_i = \widetilde{v_i}$ for some values v_i. Then we can write $\mathcal{V} = \sum_{i=0}^{n} \mathcal{V}(\widetilde{v_i})\mathcal{V}_{v_i}$.

Let us define $M_0 = v_0$ and $M_{i+1} = v_{i+1} +_{p_{i+1}} M_i$ with $p_{i+1} = \frac{\mathcal{V}(\widetilde{v_{i+1}})}{\sum_{j=0}^{i+1}\mathcal{V}(\widetilde{v_j})}$ for $i < n$. We have

$$\mathcal{V}_{M_0} = \mathcal{V}_{v_0}$$
$$\mathcal{V}_{M_{i+1}} = \frac{\mathcal{V}(\widetilde{v_{i+1}})}{\sum_{j=0}^{i+1}\mathcal{V}(\widetilde{v_j})}\mathcal{V}_{v_{i+1}} + \frac{\sum_{j=0}^{i}\mathcal{V}(\widetilde{v_j})}{\sum_{j=0}^{i+1}\mathcal{V}(\widetilde{v_j})}\mathcal{V}_{M_i}$$

so by an easy induction on i we have

$$\mathcal{V}_{M_i} = \sum_{j=0}^{i} \frac{\mathcal{V}(\widetilde{v_j})}{\sum_{k=0}^{i}\mathcal{V}(\widetilde{v_k})}\mathcal{V}_{v_j}.$$

In particular when $i = n$ we have $\sum_{k=0}^{n}\mathcal{V}(\widetilde{v_k}) = 1$ so $\mathcal{V}_{M_n} = \mathcal{V}$. $\square$

We have a correspondence between classes of terms for $\equiv$ and terms with probability distributions, which proves theorem 1.1.2.7.

A consequence of this theorem is that every result we will prove modulo $\equiv$ will still hold if we consider any other calculus also expressing probability distributions. For instance we can consider the following calculus equivalent to ours:

$$M, N := x \mid \lambda x.M \mid M\ N \mid \sum_{i=1}^{n} p_i.M_i, \sum_{i=1}^{n} p_i = 1$$

$$\lambda x.\sum_{i=1}^{n} p_i.M_i \to_+ \sum_{i=1}^{n} p_i.\lambda x.M_i$$
$$\left(\sum_{i=1}^{n} p_i.M_i\right)\ N \to_+ \sum_{i=1}^{n} p_i.M_i\ N$$

with the appropriate syntactic equivalence.

We have similar results for some other calculi, whose proofs we will not detail.

Non-deterministic λ-calculus: $\mathcal{L} = \{*\}$

$$M, N := x \mid \lambda x.M \mid M\ N \mid M + N$$

$$\begin{array}{rcl} M + N & \equiv_{\text{syn}} & N + M \\ (M+N)+P & \equiv_{\text{syn}} & M + (N+P) \\ M + M & \equiv_{\text{syn}} & M \end{array}$$

$$M, N \in \Lambda_+^{\mathcal{L}}/_{\equiv} := \text{ finite sets of values } v$$
$$v := x \mid \lambda x.v \mid v\ M.$$

Non-deterministic λ-calculus with multiplicities: $\mathcal{L} = \{*\}$

$$M, N := x \mid \lambda x.M \mid M\ N \mid M + N$$

$$\begin{array}{rcl} M + N & \equiv_{\text{syn}} & N + M \\ (M+N)+P & \equiv_{\text{syn}} & M + (N+P) \end{array}$$

$$M, N \in \Lambda_+^{\mathcal{L}}/_{\equiv} := \text{ finite multisets of values } v$$
$$v := x \mid \lambda x.v \mid v\ M.$$

Algebraic λ-calculus: $\mathcal{L} = \mathbb{R}^2$

$$M, N := x \mid \lambda x.M \mid M\ N \mid M +_{\alpha,\beta} N$$

$$\begin{array}{rcl} M +_{\alpha,\beta} N & \equiv_{\text{syn}} & N +_{\beta,\alpha} M \\ (M +_{\alpha,\beta} N) +_{\gamma,\delta} P & \equiv_{\text{syn}} & M +_{\alpha\gamma,1} (N +_{\beta\gamma,\delta} P) \\ M +_{\alpha,1-\alpha} M & \equiv_{\text{syn}} & M \\ M +_{\alpha,0} N & \equiv_{\text{syn}} & M +_{\alpha,0} P \\ \\ M +_{\alpha+\gamma,\beta} M & \equiv_{\text{syn}} & M +_{\alpha,\beta+\gamma} M \end{array}$$

$$M, N \in \Lambda_+^{\mathcal{L}}/_{\equiv} := \text{ finite linear combinations of values } v$$
$$v := x \mid \lambda x.v \mid v\ M.$$

In the case of the algebraic λ-calculus we need an additional rule. This equality is necessary for terms to represent linear combinations of values. It is actually derivable from the other rules in most cases.

If $\alpha + \beta \neq 0$ we have :

$$\begin{aligned} M +_{\alpha,\beta} M &\equiv_{\text{syn}} (M +_{\alpha,\beta} M) +_{1,0} (M +_{\alpha,\beta} M) \\ &\equiv_{\text{syn}} (M +_{\alpha,\beta} M) +_{1,0} M \\ &\equiv_{\text{syn}} M +_{\alpha,1} (M +_{\beta,0} M) \\ &\equiv_{\text{syn}} (M +_{\frac{\alpha}{\alpha+\beta},\frac{\beta}{\alpha+\beta}} M) +_{\alpha+\beta,0} M \\ &\equiv_{\text{syn}} M +_{\alpha+\beta,0} M \end{aligned}$$

so $M +_{\alpha,\beta} M \equiv_{\text{syn}} M +_{\gamma,\delta} M$ whenever $\alpha + \beta = \gamma + \delta \neq 0$.

Then if $\alpha \neq 0$ we also have:

$$\begin{aligned} M +_{1,-1} M &\equiv_{\text{syn}} (M +_{\alpha,1-\alpha} M) +_{1,-1} M \\ &\equiv_{\text{syn}} M +_{\alpha,1} (M +_{1-\alpha,-1} M) \\ &\equiv_{\text{syn}} M +_{\alpha,1} (M +_{-\alpha,0} M) \\ &\equiv_{\text{syn}} (M +_{\alpha,-\alpha} M) +_{1,0} M \\ &\equiv_{\text{syn}} (M +_{\alpha,-\alpha} M) +_{1,0} (M +_{\alpha,-\alpha} M) \\ &\equiv_{\text{syn}} M +_{\alpha,-\alpha} M \end{aligned}$$

so $M +_{\alpha,-\alpha} M \equiv_{\text{syn}} M +_{\beta,-\beta} M$ whenever $\alpha \neq 0$ and $\beta \neq 0$.

But as far as we know we can not prove that $M +_{0,0} M \equiv_{\text{syn}} M +_{1,-1} M$ without using this additional rule.

We did not include this rule in our proofs and we will not in the future, as we are mostly interested in the probabilistic calculus, but we believe our results still hold in this case.

1.2 Deterministic β-reduction on non-deterministic terms

Now that we know what our terms modulo $\equiv$ describe, let us look at the most important reduction, i.e. the β-reduction. We will first describe the β-reduction modulo $=_+$, and in a second time we will consider it modulo $\equiv$.

1.2.1 β-reduction modulo $=_+$

An important property we want to prove is the confluence of the reduction. To achieve this we proceed in a standard way and we prove we can define a parallel β-reduction using labelled redexes. Using a parallel reduction is all the more relevant as we may duplicate β-redexes when we reduce a term with $\to_+$, and we want to be able to reduce all the duplicates at once.

The parallel labelled reduction is the reduction of the set of labelled redexes in a term. So first we give a syntax for terms with labelled redexes.

Definition 1.2.1.1. Terms with labelled redexes are:

$$M, N := x \mid \lambda x.M \mid M\ N \mid M +_l N \mid ((\lambda x.M)\ N)_* .$$

The labelled β-reduction is:

$$((\lambda x.M)\ N)_* \to_{\beta_*} M \left[{}^{N}/_{x}\right]$$

extended to arbitrary context.

If we only wanted to prove we can reduce multiple redexes in parallel, we would prove that $\to_{\beta_*}$ is confluent and strongly normalizing, thus every term has a unique normal form which is the reduction of all its labelled redexes. But here we want to consider terms modulo $=_+$, so we also involve the reduction $\to_+$ defined on labelled terms by:

$$\begin{gathered}\lambda x.(M +_l N) \to_+ \lambda x.M +_l \lambda x.N\\ (M +_l N)\ P \to_+ M\ P +_l N\ P\\ ((\lambda x.(M +_l N))\ P)_* \to_+ ((\lambda x.M)\ P)_* +_l ((\lambda x.N)\ P)_*\end{gathered}$$

extended to context.

Proposition 1.2.1.1. *$\to_{\beta_*}$ and $\to_+$ are substitutive: if $M \to_{\beta_*} M'$ (resp. $M \to_+ M'$) then for all x and N we have $M \left[{}^{N}/_{x}\right] \to_{\beta_*} M' \left[{}^{N}/_{x}\right]$ (resp. $M \left[{}^{N}/_{x}\right] \to_+ M' \left[{}^{N}/_{x}\right]$).*

Proof. By a simple induction on the context of the reduction. For instance for the β_*-reduction and the empty context we have

$$((\lambda y.M)\ P)_* \left[{}^{N}/_{x}\right] = \left(\left(\lambda y.M \left[{}^{N}/_{x}\right]\right)\ P \left[{}^{N}/_{x}\right]\right)_*$$

if y is not free in N, which we can assume as we consider terms modulo α-equivalence. Thus $((\lambda y.M)\ P)_* \left[{}^{N}/_{x}\right] \rightarrow_{\beta_*} M\left[{}^{N}/_{x}\right]\left[P\left[{}^{N}/_{x}\right]/_{y}\right]$ and this is known to be the equal to $\left(M\left[{}^{P}/_{y}\right]\right)\left[{}^{N}/_{x}\right]$. □

Proposition 1.2.1.2. *$\rightarrow_{\beta_*}$ and $\rightarrow_{\beta_*} \cup \rightarrow_+$ are weakly confluent.*

Proof. We only deal with $\rightarrow_{\beta_*} \cup \rightarrow_+$, but the same arguments can be used to prove the weak confluence of $\rightarrow_{\beta_*}$. We must prove that if a term has two reductions for $\rightarrow_{\beta_*} \cup \rightarrow_+$ then the reduced terms have a common reduct. We reason by induction on the two contexts of the reductions.

If neither context is empty the result is immediate. Either the two reductions occur in the same subterm (for instance $\lambda x.M$ reduces into $\lambda x.N_1$ and $\lambda x.N_2$ as M reduces into N_1 and N_2) and we conclude by induction hypothesis, or they occur in different subterms (such as $M\ M' \rightarrow N\ M'$ and $M\ M' \rightarrow M\ N'$) and we can immediately compose the reductions ($N\ M'$ and $M\ N'$ both reduce to $N\ N'$).

The remaining cases are when one of the contexts is empty.

- If

$$\begin{aligned} &((\lambda x.M)\ N)_* \rightarrow_{\beta_*} M\left[{}^{N}/_{x}\right] \\ &((\lambda x.M)\ N)_* (\rightarrow_{\beta_*} \cup \rightarrow_+)\, ((\lambda x.M')\ N)_* \end{aligned}$$

  we conclude by substitutivity of $\rightarrow_{\beta_*}$ and $\rightarrow_+$.

- If

$$\begin{aligned} &((\lambda x.M)\ N)_* \rightarrow_{\beta_*} M\left[{}^{N}/_{x}\right] \\ &((\lambda x.M)\ N)_* \rightarrow_{\beta_*} \cup \rightarrow_+ ((\lambda x.M)\ N')_* \end{aligned}$$

  we can prove by an easy induction on M that $M\left[{}^{N}/_{x}\right] \twoheadrightarrow_{\beta_*} \cup \twoheadrightarrow_+ M\left[{}^{N'}/_{x}\right]$.

- If

$$\begin{aligned} &\lambda x.(M +_l N) \rightarrow_+ \lambda x.M +_l \lambda x.N \\ \text{or } &(M +_l N)\ P \rightarrow_+ M\ P +_l N\ P \\ \text{or } &((\lambda x.(M +_l N))\ P)_* \rightarrow_+ ((\lambda x.M)\ P)_* +_l ((\lambda x.N)\ P)_* \end{aligned}$$

  and the second reduction occurs in M, N or P we can easily compose the reductions.

- If

$$\begin{aligned} &((\lambda x.(M +_l N))\ P)_* \rightarrow_{\beta_*} (M +_l N)\left[{}^{P}/_{x}\right] \\ &((\lambda x.(M +_l N))\ P)_* \rightarrow_+ ((\lambda x.M)\ P)_* +_l ((\lambda x.N)\ P)_* \end{aligned}$$

then

$$(M +_l N)\left[{}^{P}\!/\!_{x}\right] = M\left[{}^{P}\!/\!_{x}\right] +_l N\left[{}^{P}\!/\!_{x}\right]$$
$$((\lambda x.M)\ P)_* +_l ((\lambda x.N)\ P)_* \twoheadrightarrow_{\beta_*} M\left[{}^{P}\!/\!_{x}\right] +_l N\left[{}^{P}\!/\!_{x}\right].$$

□

Proposition 1.2.1.3. *$\rightarrow_{\beta_*}$ and $\rightarrow_{\beta_*} \cup \rightarrow_+$ are strongly normalizing.*

Proof. We consider that a weight on a labelled term M is a function w_M which associates a weight $w_M(x^o) \geq 2$ in $\mathbb{N}$ to every variable occurrence x^o in M. For every other subterm N of M we then define $w_M(N)$ by

- $w_M(\lambda x.N) = w_M(N)$;
- $w_M(N\ P) = w_M(N) \times w_M(P)$;
- $w_M(N +_l P) = w_M(N) + w_M(P)$;
- $w_M\left(((\lambda x.N)\ P)_*\right) = w_M(N) \times w_M(P)$.

Remark that to be accurate a subterm of M is given by a term N and a context C such that $M = C[N]$. Here we do not mention the context, which is ambiguous but simplify the notations.

A weight w_M is said to be decreasing if for every subterm $((\lambda x.N)\ P)_*$ of M, for every occurrence x^o of the variable x in N we have $w_M(x^o) \geq w_M(P)$. We want to use decreasing weights to prove that every reduction path of $\rightarrow_{\beta_*} \cup \rightarrow_+$ is finite.

We have by an easy induction on M that for every subterm N of M, $w_M(N) \geq 2$.

Using this we also prove easily that the total weight of a term is a strictly increasing function of the weight of its subterms: given a context C and two weights $w_{C[M]}$ and $w_{C[N]}$ on $C[M]$ and $C[N]$ which agree on the weight of the variable occurrences in C, if $w_{C[M]}(M) > w_{C[N]}(N)$ then $w_{C[M]}(C[M]) > w_{C[N]}(C[N])$.

Now given a β_*-reduction $M = C\left[((\lambda x.N)\ P)_*\right] \rightarrow_{\beta_*} C\left[N\left[{}^{P}\!/\!_{x}\right]\right] = M'$ and a weight w_M on M, we define $w_{M'}$ on M' by

- if y^o is a variable occurrence in C then $w_{M'}(y^o) = w_M(y'^o)$ where y'^o is the corresponding variable occurrence in C in M;
- if y^o is an occurrence of a variable different from x in N then $w_{M'}(y^o) = w_M(y'^o)$ where y'^o is the corresponding variable occurrence in N in M;
- if y^o is a variable occurrence in a copy of P then $w_{M'}(y^o) = w_M(y'^o)$ where y'^o is the corresponding variable occurrence in P in M.

Then $w_M(M) > w_{M'}(M')$ if $w_M\left(((\lambda x.N)\ P)_*\right) > w_{M'}\left(N\left[{}^{P}/_{x}\right]\right)$. If w_M is a decreasing weight then for every variable occurrence x^o of x in N and for every copy of P in M' we have $w_M(x^o) > w_M(P) = w_{M'}(P)$. Thus

$$w_M(N) \geq w_{M'}\left(N\left[{}^{P}/_{x}\right]\right) > 0$$
$$w_M\left(((\lambda x.N)\ P)_*\right) = w_M(N) \times w_M(P) > w_{M'}\left(N\left[{}^{P}/_{x}\right]\right)$$

and $w_M(M) > w_{M'}(M')$.

Furthermore we can check that the weight $w_{M'}$ is decreasing. For every labelled redex $((\lambda x.N')\ P')_*$ in M' we can find a labelled redex $((\lambda x.N)\ P)_*$ in M such that $P = P'$ or $P \to_{\beta_*} P'$, hence $w_M(P) \geq w_{M'}(P')$, and every occurrence x'^o of x in N' inherits its weight from an occurrence x^o of x in N.

We can do the same reasoning with $\to_+$. Given a reduction $M \to_+ M'$, a weight w_M on M induces in an obvious way a weight $w_{M'}$ on M'. For instance if $M = C[(N +_l P)\ Q] \to_+ C[N\ Q +_l P\ Q] = M'$ then:

- the variable occurrences of N in M' have the same weight as the corresponding occurrences of N in M;
- the variable occurrences of P in M' have the same weight as the corresponding occurrences of P in M;
- the variable occurrences of each of the Q's in M' have the same weight as the corresponding occurrences of Q in M.

It is easy to check that if w_M is decreasing then so is $w_{M'}$. Besides the weight of a redex is preserved by reduction. In the example above we have

$$\begin{aligned} w_M((N +_l P)\ Q) &= (w_M(N) + w_M(P)) \times w_M(Q) \\ &= w_{M'}(N) \times w_{M'}(Q) + w_{M'}(P) \times w_{M'}(Q) \\ &= w_{M'}(N\ Q +_l P\ Q). \end{aligned}$$

To sum up, if w_M is a decreasing weight on M, then if $M \to_{\beta_*} N$ there is a decreasing weight w_N on N with $w_M(M) > w_N(N)$, and if $M \to_+ N$ then there is a decreasing weight w_N on N with $w_M(M) = w_N(N)$.

This means that if M has a decreasing weight then every reduction path from M for $\to_{\beta_*} \cup \to_+$ has finitely many β_* steps. In particular $\to_{\beta_*}$ is strongly normalizing. But we also know there is no infinite reduction path for $\to_+$. Then every reduction path for $\to_{\beta_*} \cup \to_+$ is finite.

To conclude we need to prove that we can indeed find a decreasing weight for every term. We easily prove by induction on M that for every family $(\delta_x) \in \mathbb{N}^{Var}$ we can find a decreasing weight on M with $w_M(x^o) \geq \delta_x$ for every variable occurrence x^o of x.

- If $M = x$ we just choose $w_x(x) \geq \max(\delta_x, 2)$.

- If $M = \lambda x.N$ then M and N have the same variable occurrences, and we have by induction hypothesis a decreasing weight w_N on N so we define $w_M(y^o) = w_N(y^o)$ on M.
- Similarly if $M = N\ P$ or $M = N +_l P$ we have weights w_N and w_P and we define $w_M(x^o) = w_N(x^o)$ for every x^o in N and $w_M(x^o) = w_P(x^o)$ for every x^o in P.
- If $M = ((\lambda x.N)\ P)_*$ we have by induction hypothesis a weight w_P, and we have a weight w_N such that $w_N(x^o) > \max(\delta_x, w_P(P))$ for every variable occurrence x^o of x in N. Then we can define w_M as in the previous case.

□

Corollary 1.2.1.4. *1. Every labelled term has a unique normal form for $\to_{\beta_*}$. The normal forms are exactly the terms without labelled redex.*

2. Every labelled term has a unique normal form for $\to_{\beta_*} \cup \to_+$. The normal forms are exactly the canonical terms without labelled redex.

Now to define the parallel β-reduction, we consider the following operations:

- given a term M and a set $\mathcal{F}$ of β-redexes in M, we write $M_{\mathcal{F}}$ the labelled term obtained by labelling the redexes of $\mathcal{F}$ in M;
- given a labelled term M, we write $|M|$ the term obtained by erasing the labels of M.

For every term M we have $|M_{\mathcal{F}}| = M$, and for every labelled term M there exists M' and $\mathcal{F}$ such that $M = M'_{\mathcal{F}}$.

Definition 1.2.1.2. Given a term M and a set $\mathcal{F}$ of β-redexes in M, the reductions $\xrightarrow{\mathcal{F}}_{\beta_{/\!/}}$ and $\xrightarrow{\mathcal{F}}_{\beta^c_{/\!/}}$ of M are defined by

$$\frac{M_{\mathcal{F}} \twoheadrightarrow_{\beta_*} N \qquad N\ \beta_*\text{-normal}}{M \xrightarrow{\mathcal{F}}_{\beta_{/\!/}} N} \qquad \frac{M_{\mathcal{F}}(\to_{\beta_*} \cup \to_+)^* N \qquad N\ \beta_*{+}\text{-normal}}{M \xrightarrow{\mathcal{F}}_{\beta^c_{/\!/}} N}.$$

In general we define the reductions $\to_{\beta_{/\!/}}$ and $\to_{\beta^c_{/\!/}}$ by

$$\frac{\exists \mathcal{F} : M \xrightarrow{\mathcal{F}}_{\beta_{/\!/}} N}{M \to_{\beta_{/\!/}} N} \qquad \frac{\exists \mathcal{F} : M \xrightarrow{\mathcal{F}}_{\beta^c_{/\!/}} N}{M \to_{\beta^c_{/\!/}} N}.$$

The canonical parallel β-reduction is defined on all terms, but we will usually consider it as a reduction between canonical terms, as if $M \to_{\beta^c_{/\!/}} N$ then N is necessarily canonical.

Observe that for every term M and every set of redexes $\mathcal{F}$ there are unique N and N' such that $M \xrightarrow{\mathcal{F}}_{\beta_{/\!/}} N$ and $M \xrightarrow{\mathcal{F}}_{\beta^c_{/\!/}} N'$, and then $N' = \mathrm{can}(N)$.

We can also see that $\to_\beta \subset \to_{\beta_{/\!/}}$ and $\to_{\beta_{/\!/}} \subset \twoheadrightarrow_\beta$ so $\twoheadrightarrow_{\beta_{/\!/}} = \twoheadrightarrow_\beta$. Similarly $\twoheadrightarrow_{\beta^c_{/\!/}} = (\to_\beta \cup \to_+)^*$. Thus the confluences of $\to_\beta$ and $\to_\beta \cup \to_+$ are equivalent to the confluences of $\to_{\beta_{/\!/}}$ and $\to_{\beta^c_{/\!/}}$.

To prove these confluences let us define the notion of residuals of a β-redex.

We can define the β-reduction on labelled terms by $(\lambda x.M)\ N \to_\beta M\left[{}^{N}\!/_{x}\right]$. Then given a reduction on terms $M \to_\beta N$ and a set $\mathcal{F}$ of β-redexes in M, we can reduce $M_\mathcal{F}$ either by β (if the reduced redex is not in $\mathcal{F}$, hence not labelled) or by β_* (if the reduced redex is in $\mathcal{F}$) to get a labelled term N' such that $|N'| = N$. Then there is a set of redexes $\mathcal{G}$ in N such that $N' = N_\mathcal{G}$. We call this set $\mathcal{G}$ the set of residuals of $\mathcal{F}$ for this reduction.

For instance if $M = (\lambda x.x\ x)\ ((\lambda y.P)\ Q) \to_\beta ((\lambda y.P)\ Q)\ ((\lambda y.P)\ Q) = N$, we have:

$$(\lambda x.x\ x)\ ((\lambda y.P)\ Q)_* \to_\beta ((\lambda y.P)\ Q)_*\ ((\lambda y.P)\ Q)_*$$

$$((\lambda x.x\ x)\ ((\lambda y.P)\ Q))_* \to_{\beta_*} ((\lambda y.P)\ Q)\ ((\lambda y.P)\ Q)$$

so for this reduction $((\lambda y.P)\ Q)$ in M has two residuals in N whereas the whole redex $(\lambda x.x\ x)\ ((\lambda y.P)\ Q)$ doesn't have any.

It is well known that when we define the residuals we must say exactly what the reduced redex is. For instance we have the two following reductions on labelled terms:

$$\begin{aligned} (\lambda x.x)\ \ ((\lambda x.x)\ y)_* &\to_\beta ((\lambda x.x)\ y)_* \\ (\lambda x.x)\ \ ((\lambda x.x)\ y)_* &\to_{\beta_*} (\lambda x.x)\ y \end{aligned}$$

There are two ways to obtain the reduction $(\lambda x.x)\ \ ((\lambda x.x)\ y) \to_\beta (\lambda x.x)\ y$: we can reduce the outer redex $(\lambda x.x)\ \ ((\lambda x.x)\ y)$ or the inner redex $(\lambda x.x)\ y$. In the first case the redex $(\lambda x.x)\ y$ has a residual, whereas in the second case it has none.

We define the notion of residuals for $\to_+$ in the same way.

This definition is not standard, as residuals are usually defined for any sub-term, not only redexes, and we would say that N is always the residual of M when $M \to_\beta N$.

The reason we give this definition is that we are interested in the reduction of sets of redexes. In particular we have the following result:

Proposition 1.2.1.5. *Given a term M and a set of redexes $\mathcal{F}$ with $M \xrightarrow{\mathcal{F}}_{\beta_{/\!/}} N$ (resp. $M \xrightarrow{\mathcal{F}}_{\beta^c_{/\!/}} N$), for all $\mathcal{G} \subset \mathcal{F}$ if $M \xrightarrow{\mathcal{G}}_{\beta_{/\!/}} P$ (resp. $M \xrightarrow{\mathcal{G}}_{\beta^c_{/\!/}} P$) and $\mathcal{F}'$ is the set of residuals of $\mathcal{F}$ in P then $P \xrightarrow{\mathcal{F}'}_{\beta_{/\!/}} N$ (resp. $P \xrightarrow{\mathcal{F}'}_{\beta^c_{/\!/}} N$).*

Proof. We have $M_\mathcal{F} \twoheadrightarrow_{\beta_*} P_{\mathcal{F}'}$ (resp. $M_\mathcal{F}\ (\to_{\beta_*} \cup \to_+)^*\ P_{\mathcal{F}'}$), so in either case the normal form of $M_\mathcal{F}$ is the same as the normal form of $P_{\mathcal{F}'}$. □

In the above example if $M = (\lambda x.x\ x)\ ((\lambda y.P)\ Q)$ and $\mathcal{F} = \{(\lambda x.x\ x)\ ((\lambda y.P)\ Q)\}$ the reduction of $\mathcal{F}$ in M gives $((\lambda y.P)\ Q)\ ((\lambda y.P)\ Q)$, but we do not want to say that the redex $((\lambda y.P)\ Q)\ ((\lambda y.P)\ Q)$ is a residual of $\mathcal{F}$.

Proposition 1.2.1.6. *$\to_{\beta_{/\!/}}$ and $\to_{\beta^c_{/\!/}}$ have the diamond property: if $M \to_{\beta_{/\!/}} N_1$ and $M \to_{\beta_{/\!/}} N_2$ (resp. $M \to_{\beta^c_{/\!/}} N_1$ and $M \to_{\beta^c_{/\!/}} N_2$) then there is P such that $N_1 \to_{\beta_{/\!/}} P$ and $N_2 \to_{\beta_{/\!/}} P$ (resp. $N_1 \to_{\beta^c_{/\!/}} P$ and $N_2 \to_{\beta^c_{/\!/}} P$).*

Proof. If we have $M \xrightarrow{\mathcal{F}_1}_{\beta_{/\!/}} N_1$ and $M \xrightarrow{\mathcal{F}_2}_{\beta_{/\!/}} N_2$ then we define P as the reduction of $\mathcal{F}_1 \cup \mathcal{F}_2$ in M. For $i \in \{1; 2\}$, if $\mathcal{G}_i$ is the set of residuals of $\mathcal{F}_1 \cup \mathcal{F}_2$ in N_i then we have $N_i \xrightarrow{\mathcal{G}_i}_{\beta_{/\!/}} P$. □

Corollary 1.2.1.7. *The reductions $\to_\beta$ and $\to_\beta \cup \to_+$ are confluent.*

This is an interesting result but we can deduce even more from the properties of the labelled reduction. When we were looking at the equivalence $\equiv_{\text{syn}}$ modulo $=_+$ we proved that it is entirely characterized by its restriction $\equiv^c_{\text{syn}}$ on canonical terms. Here it does not make sense to restrict the β-reduction to canonical terms. For instance if we consider the reduction

$$(\lambda xy.x)\ (u +_l v) \to_\beta \lambda y.(u +_l v) \to_+ \lambda y.u +_l \lambda y.v$$

there is no way to go from $(\lambda xy.x)\ (u +_l v)$ to $\lambda y.u +_l \lambda y.v$ without using a non canonical form. So we define the following reduction.

Definition 1.2.1.3. The canonical β-reduction $\to_{\beta^c}$ between canonical terms is defined by $\to_{\beta^c} = \to_\beta \cdot \to_{\text{can}}$.

Proposition 1.2.1.8. *If $M \to_\beta N$ then* $\text{can}(M) \twoheadrightarrow_{\beta^c} \text{can}(N)$.

Proof. If Δ is the reduced redex in M we have $M \xrightarrow{\{\Delta\}}_{\beta^c_{/\!/}} \text{can}(N)$. We prove that in general if $P \xrightarrow{\mathcal{F}}_{\beta^c_{/\!/}} Q$ then $\text{can}(P) \twoheadrightarrow_{\beta^c} Q$. We know $\to_{\beta_*} \cup \to_+$ is strongly normalizing so the length of the reductions of P is bounded, we reason by induction on that bound.

If P is not canonical then $P \to_+ P'$ with $\text{can}(P) = \text{can}(P')$ so the result is immediate by induction hypothesis.

If P is canonical then either $\mathcal{F} = \emptyset$ and $P = Q$, or we can reduce a redex of $\mathcal{F}$ to get $P \to_\beta P'$. In this case the induction hypothesis gives $\text{can}(P') \twoheadrightarrow_{\beta^c} Q$ and we have $P \to_{\beta^c} \text{can}(P') \twoheadrightarrow_{\beta^c} Q$. □

Corollary 1.2.1.9. *If $M \twoheadrightarrow_{\beta/=_+} N$ then* $\text{can}(M) \twoheadrightarrow_{\beta^c} \text{can}(N)$.

Proof. If $M \twoheadrightarrow_{\beta/=_+} N$ we have

$$M = M_0 =_+ M'_0 \to_\beta M_1 =_+ M'_1 \to_\beta \ldots \to_\beta M_n =_+ M'_n = N.$$

Using the previous result we get $\text{can}(M'_k) \twoheadrightarrow_{\beta^c} \text{can}(M_{k+1})$ for $k < n$, and since $M_k =_+ M'_k$ we have $\text{can}(M_k) = \text{can}(M'_k)$ for $k \leq n$. Thus

$$\text{can}(M_0) \twoheadrightarrow_{\beta^c} \text{can}(M_1) \twoheadrightarrow_{\beta^c} \ldots \twoheadrightarrow_{\beta^c} \text{can}(M_n).$$

We have $\text{can}(M) \twoheadrightarrow_{\beta^c} \text{can}(N)$. □

To consider the β-reduction modulo $=_+$ is the same as considering the reduction $\to_{\beta^c}$ between canonical terms. There is a difference in the complexity, i.e. the length of the reductions: in canonical terms some redexes are duplicated and need to be reduced several times. But here we are more interested in the equational point of view than the computational one.

Once again canonical forms are very useful to describe the β-reduction modulo $=_+$. But once again they are not stable by context, and the reduction $\to_{\beta^c}$ is not contextual. Actually to study this reduction the most simple way is usually to decompose it into a reduction $\to_\beta \cdot \twoheadrightarrow_+$, so in the future we will not consider only the reduction $\to_{\beta^c}$ but we will go back and forth between the canonical and the non-canonical reductions.

1.2.2 β-reduction modulo $\equiv$

The important computation rule of our calculus is the β-rule, whereas the rules for the sums have a different status: they play an important role in the reduction of terms but it makes as much sense to consider terms modulo $=_+$ than to consider $\to_+$ as an actual reduction. The syntactic rules of $\equiv_{\text{syn}}$, on the other hand, are not supposed to describe a computation on the terms, only on the labels. For that reason we expect the β-reduction modulo $\equiv$ to be similar to the reduction modulo $=_+$.

It seems fairly obvious that the commutativity and associativity rules do not influence the reduction. This is less trivial for the rule $M +_l N \equiv_{\text{syn}} M +_l P$ if $l \in \mathcal{Z} \subset \mathcal{L}$: we can replace a subterm by any other one, without restriction. But the meaning of this rule is that everything on the right side of a sum $+_l$ with $l \in \mathcal{Z}$ is irrelevant and should not influence the reduction.

The rule with the most consequences is the idempotence of the sum. If we have a reduction $M \to_\beta N$ then for any $l \in \mathcal{I}$ we have $M \to_{\beta/\equiv} M +_l M$. We will prove that the reduction is confluent so even if we keep reducing $M \twoheadrightarrow_{\beta/\equiv} M_1$ and $M \twoheadrightarrow_{\beta/\equiv} M_2$ there will be N such that M_1 and M_2 both reduce to N, and we will be able to contract the sum again: $M_1 +_l N_2 \twoheadrightarrow_{\beta/\equiv} N +_l N \equiv N$. In particular this rule will not influence the normalization of a term. But if we want to characterize precisely the β-reduction modulo $\equiv$ the splitting of terms will play an important role.

1.2.2.1 Splitting terms

In section 1.1.2 we defined the reductions $\to_\gamma$, $\to_\alpha$, $\to_\mathcal{I}$ and $\to_\mathcal{Z}$ associated to the rules of $\equiv_{\text{syn}}$. We will again use these reductions, but we will isolate the splitting of terms.

Definition 1.2.2.1. 1. The reduction $\to_{\text{split}}$ is defined by

$$M \to_{\text{split}} M +_l M \text{ if } l \in \mathcal{I}$$

extended to context.

2. The reduction $\rightarrow_{\text{syn}}$ is $\rightarrow_{\mathcal{I}} \cup \leftrightarrow_{\gamma} \cup \leftrightarrow_{\alpha} \cup \rightarrow_{\mathcal{Z}}$.

We showed that sometimes we need to split a term before reducing it: if $M \rightarrow_{\text{split}} M +_l M \twoheadrightarrow_{\beta/=_+} M +_l N$ then there is no way to perform the β-reduction before the splitting. On the other hand we will prove that we can always do all the splitting first, and then use the other reductions: if $M \twoheadrightarrow_{\beta/\equiv} N$ then $M \twoheadrightarrow_{\text{split}} \cdot (\rightarrow_{\beta/=_+} \cup \rightarrow_{\text{syn}})^* N$.

First let us show we can reason on the reduction $\twoheadrightarrow_{\text{split}}$ by structural induction.

Proposition 1.2.2.1. *The reduction $\twoheadrightarrow_{\text{split}}$ is characterized inductively by:*

$$\frac{}{x \twoheadrightarrow_{\text{split}} x} \qquad \frac{M \twoheadrightarrow_{\text{split}} N}{\lambda x.M \twoheadrightarrow_{\text{split}} \lambda x.N}$$

$$\frac{M \twoheadrightarrow_{\text{split}} N \qquad M' \twoheadrightarrow_{\text{split}} N'}{M\ M' \twoheadrightarrow_{\text{split}} N\ N'} \qquad \frac{M_1 \twoheadrightarrow_{\text{split}} N_1 \qquad M_2 \twoheadrightarrow_{\text{split}} N_2}{M_1 +_l M_2 \twoheadrightarrow_{\text{split}} N_1 +_l N_2}$$

$$\frac{M \twoheadrightarrow_{\text{split}} N_1 \qquad M \twoheadrightarrow_{\text{split}} N_2 \qquad l \in \mathcal{I}}{M \twoheadrightarrow_{\text{split}} N_1 +_l N_2}$$

Proof. Let us write $\twoheadrightarrow$ the reduction defined by these rules. We prove easily that these rules hold for $\twoheadrightarrow_{\text{split}}$ so $\twoheadrightarrow \subset \twoheadrightarrow_{\text{split}}$. We prove that $\twoheadrightarrow_{\text{split}} \subset \twoheadrightarrow$ by induction on the length of the reduction $\twoheadrightarrow_{\text{split}}$.

For a reduction of length 0 we prove easily that $\twoheadrightarrow$ is reflexive. Otherwise we have $M \twoheadrightarrow_{\text{split}} N \rightarrow_{\text{split}} P$ and by induction hypothesis on the reduction $M \twoheadrightarrow_{\text{split}} N$ we have $M \twoheadrightarrow N$. We then reason by induction on the context C of the reduction $N \rightarrow_{\text{split}} P$ and on $M \twoheadrightarrow N$.

- If $C \neq [\,]$ and $M \twoheadrightarrow N$ is given by a context rule the result is immediate by induction hypothesis. For instance if we have $N = \lambda x.N' \rightarrow_{\text{split}} \lambda x.P'$ and $M = \lambda x.M' \twoheadrightarrow \lambda x.N'$ then by induction hypothesis $M' \twoheadrightarrow P'$, hence $M \twoheadrightarrow \lambda x.P'$.

- If $C \neq [\,]$ and $M \twoheadrightarrow N_1 +_l N_2$ with $M \twoheadrightarrow N_1$ and $M \twoheadrightarrow N_2$, then we have $P = P_1 +_l P_1$ with $N_i \rightarrow^?_{\text{split}} P_i$ for $i \in \{1; 2\}$ so by hypothesis or by induction hypothesis we have $M \twoheadrightarrow P_i$. Thus $M \twoheadrightarrow P_1 +_l P_2$.

- If $C = [\,]$ then $P = N +_l N$ so using the last rule of $\twoheadrightarrow$ we have $M \twoheadrightarrow N +_l N$.

□

Now let us prove that $\rightarrow_{\text{split}}$ commutes with the other reductions.

Proposition 1.2.2.2. *If $M \rightarrow_{\mathcal{I}} N \twoheadrightarrow_{\text{split}} P$ then there is a term N' such that $M \twoheadrightarrow_{\text{split}} N' \twoheadrightarrow_{\mathcal{I}} P$.*

Proof. By induction on $N \twoheadrightarrow_{\text{split}} P$ and the context C of the reduction $M \rightarrow_{\mathcal{I}} N$.

- If $N \twoheadrightarrow_{\text{split}} P_1 +_l P_2$ with $N \twoheadrightarrow_{\text{split}} P_i$ for $i \in \{1; 2\}$ then by induction hypothesis there are N'_1 and N'_2 such that $M \twoheadrightarrow_{\text{split}} N'_i \twoheadrightarrow_{\mathcal{I}} P_i$ for $i \in \{1; 2\}$, and $M \twoheadrightarrow_{\text{split}} N'_1 +_l N'_2 \twoheadrightarrow_{\mathcal{I}} P_1 +_l P_2$.
- If $C = [\,]$ then $M = N +_l N$ and we have $M \twoheadrightarrow_{\text{split}} P +_l P \rightarrow_{\mathcal{I}} P$.
- If $N \twoheadrightarrow_{\text{split}} P$ does not use the last rule and $C \neq [\,]$ then the result is immediate by induction hypothesis.

□

Corollary 1.2.2.3. $\rightarrow_{\text{split}}$ *is confluent.*

Proof. We just proved that if $M \twoheadrightarrow_{\text{split}} N_1$ and $M \rightarrow_{\text{split}} N_2$ then there is P such that $N_i \twoheadrightarrow_{\text{split}} P$ for $i \in \{1; 2\}$. We prove the confluence of $\rightarrow_{\text{split}}$ by induction on the length of one of the reductions. □

Proposition 1.2.2.4. *1. If $M \rightarrow_\gamma N \twoheadrightarrow_{\text{split}} P$ then there is N' such that $M \twoheadrightarrow_{\text{split}} N' \twoheadrightarrow_\gamma P$.*

2. If $M \leftarrow_\gamma N \twoheadrightarrow_{\text{split}} P$ then there is N' such that $M \twoheadrightarrow_{\text{split}} N' \twoheadleftarrow_\gamma P$.

3. If $M \rightarrow_{\mathcal{Z}} N \twoheadrightarrow_{\text{split}} P$ then there is N' such that $M \twoheadrightarrow_{\text{split}} N' \twoheadrightarrow_{\mathcal{Z}} P$.

Proof. As in the proof of the previous proposition we have the result by induction hypothesis if both reductions are given by context rules, and the same method works if $N \rightarrow_{\text{split}} P_1 +_l P_2$ with $N \rightarrow_{\text{split}} P_i$ for $i \in \{1; 2\}$. We only need to deal with the case when $N \twoheadrightarrow_{\text{split}} P$ is given by a structural rule and the context of the first reduction is empty.

- If $M_1 +_l M_2 \rightarrow_\gamma M_2 +_{\gamma(l)} M_1 \twoheadrightarrow_{\text{split}} P_2 +_{\gamma(l)} P_1$ with $M_i \twoheadrightarrow_{\text{split}} P_i$ for $i \in \{1; 2\}$ then $M_1 +_l M_2 \twoheadrightarrow_{\text{split}} P_1 +_l P_2 \rightarrow_\gamma P_2 +_{\gamma(l)} P_1$.
- If $M_1 +_{\gamma(l)} M_2 \leftarrow_\gamma M_2 +_l M_1$ we proceed in the same way.
- If $M +_l M' \rightarrow_{\mathcal{Z}} M +_l N' \twoheadrightarrow_{\text{split}} P +_l P'$ with $l \in \mathcal{Z}$, $M \twoheadrightarrow_{\text{split}} P$ and $N' \twoheadrightarrow_{\text{split}} P'$ then we have $M +_l M' \twoheadrightarrow_{\text{split}} P +_l M' \rightarrow_{\mathcal{Z}} P +_l P'$.

□

We can not reason in the same way for $\rightarrow_\alpha$. The associativity is the only rules which involves two sums. So if we proceed as for the others we need to deal with the case

$$(M_1 +_l M_2) +_{l'} M_3 \rightarrow_\alpha M_1 +_{\alpha_1(l,l')} (M_2 +_{\alpha_2(l,l')} M_3) \twoheadrightarrow_{\text{split}} P$$

where we know that the reduction $\twoheadrightarrow_{\text{split}}$ is given by a structural rule, hence $P = P_1 +_{\alpha_1(l,l')} Q$ with $M_1 \twoheadrightarrow_{\text{split}} P_1$ and $M_2 +_{\alpha_2(l,l')} M_3 \twoheadrightarrow_{\text{split}} Q$, but we dot not know if Q is of the form $P_2 +_{\alpha_2(l,l')} P_3$. Actually this reduction could use an arbitrary number of times the last rule of $\twoheadrightarrow_{\text{split}}$, and Q would be an arbitrary large tree of sums whose leaves are of the form $P_2 +_{\alpha_2(l,l')} P_3$.

For that reason we first prove that we can push down the splitting of a term.

Lemma 1.2.2.5. *If $M +_l N \twoheadrightarrow_{\text{split}} P$ then there are terms M' and N' such that $M \twoheadrightarrow_{\text{split}} M'$, $N \twoheadrightarrow_{\text{split}} N'$ and $M' +_l N' \ (\leftrightarrow_\gamma \cup \leftrightarrow_\alpha)^* \ P$.*

Proof. This is a consequence of proposition 1.0.0.1. We reason by induction on $\twoheadrightarrow_{\text{split}}$.

- If $M +_l N \twoheadrightarrow_{\text{split}} M' +_l N' = P$ with $M \twoheadrightarrow_{\text{split}} M'$ and $N \twoheadrightarrow_{\text{split}} N'$ then the result is immediate.

- If $M +_l N \twoheadrightarrow_{\text{split}} P_1 +_k P_2 = P$ with $M +_l N \twoheadrightarrow_{\text{split}} P_i$ for $i \in \{1; 2\}$ and $k \in \mathcal{I}$, then by induction hypothesis for $i \in \{1; 2\}$ there are M'_i and N'_i such that $M \twoheadrightarrow_{\text{split}} M'_i$, $N \twoheadrightarrow_{\text{split}} N'_i$ and $M'_i +_l N'_i \ (\leftrightarrow_\gamma \cup \leftrightarrow_\alpha)^* \ P_i$. We have the following reduction:

$$\begin{aligned} M +_l N &\twoheadrightarrow_{\text{split}} (M'_1 +_k M'_2) +_l (N'_1 +_k N'_2) \\ &(\leftrightarrow_\gamma \cup \leftrightarrow_\alpha)^* \ (M'_1 +_l N'_1) +_k (M'_2 +_l N'_2) \\ &(\leftrightarrow_\gamma \cup \leftrightarrow_\alpha)^* \ P_1 +_k P_2. \end{aligned}$$

□

Proposition 1.2.2.6. *1. If $M \rightarrow_\alpha N \twoheadrightarrow_{\text{split}} P$ then there is N' such that $M \twoheadrightarrow_{\text{split}} N' \ (\leftrightarrow_\gamma \cup \leftrightarrow_\alpha)^* \ P$.*

2. If $M \leftarrow_\alpha N \twoheadrightarrow_{\text{split}} P$ then there is N' such that $M \twoheadrightarrow_{\text{split}} N' \ (\leftrightarrow_\gamma \cup \leftrightarrow_\alpha)^* \ P$.

Proof. As before we prove this by induction, and we only need to detail the case where the context of the first reduction is empty and the reduction $\twoheadrightarrow_{\text{split}}$ is given by a contextual rule.

If $(M_1 +_l M_2) +_{l'} M_3 \rightarrow_\alpha M_1 +_{\alpha_1(l,l')} (M_2 +_{\alpha_2(l,l')} M_3) \twoheadrightarrow_{\text{split}} P_1 +_{\alpha_1(l,l')} Q$ with $M_1 \twoheadrightarrow_{\text{split}} P_1$ and $M_2 +_{\alpha_2(l,l')} M_3 \twoheadrightarrow_{\text{split}} Q$ the previous lemma gives P_2 and P_3 such that $M_i \twoheadrightarrow_{\text{split}} P_i$ for $i \in \{2; 3\}$ and $P_2 +_{\alpha_2(l,l')} P_3 \ (\leftrightarrow_\gamma \cup \leftrightarrow_\alpha)^* \ Q$. Then

$$\begin{aligned} (M_1 +_l M_2) +_{l'} M_3 &\twoheadrightarrow_{\text{split}} (P_1 +_l P_2) +_{l'} P_3 \\ &\rightarrow_\alpha P_1 +_{\alpha_1(l,l')} (P_2 +_{\alpha_2(l,l')} P_3) \\ &(\leftrightarrow_\gamma \cup \leftrightarrow_\alpha)^* \ P_1 +_{\alpha_1(l,l')} Q. \end{aligned}$$

The case of $\leftarrow_\alpha$ is similar. □

Corollary 1.2.2.7. *If $M \rightarrow_{\text{syn}} \cdot \twoheadrightarrow_{\text{split}} N$ then $M \twoheadrightarrow_{\text{split}} \cdot \twoheadrightarrow_{\text{syn}} N$.*

We proved that the splitting of a term can be done before using the other syntactic rules.

Now to deal with $\rightarrow_{\text{split}}$ and the reduction of sums, we have again redexes with multiple constructors so we extend the result of the lemma 1.2.2.5.

Lemma 1.2.2.8. *1. If $\lambda x.M \twoheadrightarrow_{\text{split}} P$ then there is M' such that $M \twoheadrightarrow_{\text{split}} M'$ and $\lambda x.M' \twoheadrightarrow_+ P$.*

2. *If $M\ N \twoheadrightarrow_{\text{split}} P$ then there are terms M' and N' such that $M \twoheadrightarrow_{\text{split}} M'$, $N \twoheadrightarrow_{\text{split}} N'$ and $M'\ N'\ (\rightarrow_+ \cup \rightarrow_{\mathcal{I}})^*\ P$.*

Proof. By induction on the reduction $\twoheadrightarrow_{\text{split}}$. If $\twoheadrightarrow_{\text{split}}$ is given by a structural rule (i.e. $\lambda x.M \twoheadrightarrow_{\text{split}} \lambda x.M'$ or $M\ N \twoheadrightarrow_{\text{split}} M'\ N'$) then the result is immediate. Otherwise:

- if $\lambda x.M \twoheadrightarrow_{\text{split}} P_1 +_k P_2$ with $\lambda x.M \twoheadrightarrow_{\text{split}} P_i$ for $i \in \{1;2\}$ and $k \in \mathcal{I}$, then by induction hypothesis for $i \in \{1;2\}$ there is M_i' such that $M \twoheadrightarrow_{\text{split}} M_i'$ and $\lambda x.M_i' \twoheadrightarrow_+ P_i$. We have the following reduction:

$$\begin{aligned}\lambda x.M &\twoheadrightarrow_{\text{split}} \lambda x.(M_1' +_k M_2')\\ &\rightarrow_+ \lambda x.M_1' +_k \lambda x.M_2'\\ &\twoheadrightarrow_+ P_1 +_k P_2;\end{aligned}$$

- if $M\ N \twoheadrightarrow_{\text{split}} P_1 +_k P_2$ with $M\ N \twoheadrightarrow_{\text{split}} P_i$ for $i \in \{1;2\}$ and $k \in \mathcal{I}$, then by induction hypothesis for $i \in \{1;2\}$ there are M_i' and N_i' such that $M \twoheadrightarrow_{\text{split}} M_i'$, $N \twoheadrightarrow_{\text{split}} N_i'$ and $M_i'\ N_i'\ (\rightarrow_+ \cup \rightarrow_{\mathcal{I}})^*\ P_i$. By the confluence of $\rightarrow_{\text{split}}$ there is Q such that $N_i' \twoheadrightarrow_{\text{split}} Q$, hence $Q \twoheadrightarrow_{\mathcal{I}} N_i'$, for $i \in \{1;2\}$. We have the following reduction:

$$\begin{aligned}M\ N &\twoheadrightarrow_{\text{split}} (M_1' +_k M_2')\ Q\\ &\rightarrow_+ M_1'\ Q +_k M_2'\ Q\\ &\twoheadrightarrow_{\mathcal{I}} M_1'\ N_1' +_k M_2'\ N_2'\\ &(\rightarrow_+ \cup \rightarrow_{\mathcal{I}})^*\ P_1 +_k P_2.\end{aligned}$$

□

Now we can prove that the splitting commutes with $\rightarrow_+$.

Proposition 1.2.2.9. *If $M \rightarrow_+ N \twoheadrightarrow_{\text{split}} P$ then there is a term N' such that $M \twoheadrightarrow_{\text{split}} N'\ (\rightarrow_+ \cup \rightarrow_{\mathcal{I}})^*\ P$.*

Proof. We reason by the same induction as usual.

If $\lambda x.(M_1 +_l M_2) \rightarrow_+ \lambda x.M_1 +_l \lambda x.M_2 \twoheadrightarrow_{\text{split}} P_1 +_l P_2$ then for $i \in \{1;2\}$ there is M_i' such that $M_i \twoheadrightarrow_{\text{split}} M_i'$ and $\lambda x.M_i' \twoheadrightarrow_+ P_i$. We have

$$\begin{aligned}\lambda x.(M_1 +_l M_2) &\twoheadrightarrow_{\text{split}} \lambda x.(M_1' +_l M_2')\\ &\twoheadrightarrow_+ \lambda x.M_1' +_l \lambda x.M_2'\\ &\twoheadrightarrow_+ P_1 +_l P_2.\end{aligned}$$

If $(M_1 +_l M_2)\ N \twoheadrightarrow_+ M_1\ N +_l M_2\ N \twoheadrightarrow_{\text{split}} P_1 +_l P_2$ then for $i \in \{1;2\}$ there are M_i' and Q_i such that $M_i \twoheadrightarrow_{\text{split}} M_i'$, $N \twoheadrightarrow_{\text{split}} Q_i$ and $M_i'\ Q_i\ (\rightarrow_+ \cup \rightarrow_{\mathcal{I}})^*\ P_i$. By the confluence of $\rightarrow_{\text{split}}$ there is Q such that $Q_i \twoheadrightarrow_{\text{split}} Q$ for $i \in \{1;2\}$, and we have

$$\begin{aligned}(M_1 +_l M_2)\ N \twoheadrightarrow_{\text{split}} & \ (M_1' +_l M_2')\ Q \\ & \to_+ M_1'\ Q +_l M_2'\ Q \\ & \twoheadrightarrow_{\mathcal{I}} M_1'\ Q_1 +_l M_2'\ Q_2 \\ & (\to_+ \cup \to_{\mathcal{I}})^*\ P_1 +_l P_2.\end{aligned}$$

□

The β-reduction is different in the sense that it involves a substitution which not only duplicates subterms but is also moves them inside the term.

Proposition 1.2.2.10. *If $M \to_\beta N \twoheadrightarrow_{\text{split}} P$ then there is a term N' such that $M \twoheadrightarrow_{\text{split}} N' \ (\to_\beta \cup \to_+ \cup \to_{\text{syn}})^*\ P$.*

Proof. If $(\lambda x.M)\ N \to_\beta M\left[{}^{N}/_{x}\right] \twoheadrightarrow_{\text{split}} P$ then we prove by induction on M that there are terms M' and N' such that $M \twoheadrightarrow_{\text{split}} M'$, $N \twoheadrightarrow_{\text{split}} N'$ and $M'\left[{}^{N'}/_{x}\right]\ (\to_+ \cup \to_{\text{syn}})^*\ P$.

- If $M = x$ then by hypothesis $N \twoheadrightarrow_{\text{split}} P$.
- If $M = \lambda y.M_1$ then $\lambda y.\left(M_1\left[{}^{N}/_{x}\right]\right) \twoheadrightarrow_{\text{split}} P$ and we know there is a term Q_1 such that $M_1\left[{}^{N}/_{x}\right] \twoheadrightarrow_{\text{split}} Q_1$ and $\lambda y.Q_1 \twoheadrightarrow_+ P$. By induction hypothesis there are terms M_1' and N' such that $M_1 \twoheadrightarrow_{\text{split}} M_1'$, $N \twoheadrightarrow_{\text{split}} N'$ and $M_1'\left[{}^{N'}/_{x}\right]\ (\to_+ \cup \to_{\text{syn}})^*\ Q_1$. Then $M \twoheadrightarrow_{\text{split}} \lambda y.M_1'$ and

$$\begin{aligned}(\lambda y.M_1')\left[{}^{N'}/_{x}\right]\ (\to_+ \cup \to_{\text{syn}})^* & \ \lambda y.Q_1 \\ & \twoheadrightarrow_+ P.\end{aligned}$$

- If $M = M_1\ M_2$ then $\left(M_1\left[{}^{N}/_{x}\right]\right)\ \left(M_2\left[{}^{N}/_{x}\right]\right) \twoheadrightarrow_{\text{split}} P$ and there are terms Q_1 and Q_2 such that $M_i\left[{}^{N}/_{x}\right] \twoheadrightarrow_{\text{split}} Q_i$ for all $i \in \{1;2\}$ and $Q_1\ Q_2\ (\to_+ \cup \to_{\mathcal{I}})^*\ P$. By induction hypothesis there are M_i' and N_i' such that $M_i \twoheadrightarrow_{\text{split}} M_i'$, $N \twoheadrightarrow_{\text{split}} N_i'$ and $M_i'\left[{}^{N_i'}/_{x}\right]\ (\to_+ \cup \to_{\text{syn}})^*\ Q_i$ for $i \in \{1;2\}$. Then there is N' such that $N_i' \twoheadrightarrow_{\text{split}} N'$ for $i \in \{1;2\}$ and we have $M \twoheadrightarrow_{\text{split}} M_1'\ M_2'$ and

$$\begin{aligned}(M_1'\ M_2')\left[{}^{N'}/_{x}\right] \twoheadrightarrow_{\mathcal{I}} & \left(M_1'\left[{}^{N_1'}/_{x}\right]\right)\ \left(M_2'\left[{}^{N_2'}/_{x}\right]\right) \\ & (\to_+ \cup \to_{\text{syn}})^*\ Q_1\ Q_2 \\ & (\to_+ \cup \to_{\mathcal{I}})^*\ P.\end{aligned}$$

- If $M = M_1 +_l M_2$ then $\left(M_1\left[{}^{N}/_{x}\right]\right) +_l \left(M_2\left[{}^{N}/_{x}\right]\right) \twoheadrightarrow_{\text{split}} P$ and there are terms Q_1 and Q_2 such that $M_i\left[{}^{N}/_{x}\right] \twoheadrightarrow_{\text{split}} Q_i$ for all $i \in \{1;2\}$ and

$Q_1 +_l Q_2$ $(\leftrightarrow_\gamma \cup \leftrightarrow_\alpha)^*$ P. By induction hypothesis there are M'_i and N'_i such that $M_i \twoheadrightarrow_{\text{split}} M'_i$, $N \twoheadrightarrow_{\text{split}} N'_i$ and $M'_i \left[{}^{N'_i}/_{x}\right]$ $(\rightarrow_+ \cup \rightarrow_{\text{syn}})^*$ Q_i for $i \in \{1; 2\}$. Then there is N' such that $N'_i \twoheadrightarrow_{\text{split}} N'$ for all $i \in \{1; 2\}$ and we have $M \twoheadrightarrow_{\text{split}} M'_1 +_l M'_2$ and

$$\begin{aligned}(M'_1 +_l M'_2)\left[{}^{N'}/_{x}\right] &\twoheadrightarrow_{\mathcal{I}} \left(M'_1\left[{}^{N'_1}/_{x}\right]\right) +_l \left(M'_2\left[{}^{N'_2}/_{x}\right]\right)\\ &(\rightarrow_+ \cup \rightarrow_{\text{syn}})^*\ Q_1 +_l Q_2\\ &(\leftrightarrow_\gamma \cup \leftrightarrow_\alpha)^*\ P.\end{aligned}$$

This conclude the proof of the intermediary result. Now if

$$(\lambda x.M)\ N \rightarrow_\beta M\left[{}^{N}/_{x}\right] \twoheadrightarrow_{\text{split}} P$$

then we have

$$\begin{aligned}(\lambda x.M)\ N &\twoheadrightarrow_{\text{split}} (\lambda x.M')\ N'\\ &\rightarrow_\beta M'\left[{}^{N'}/_{x}\right]\\ &(\rightarrow_+ \cup \rightarrow_{\text{syn}})^*\ P.\end{aligned}$$

□

We showed that the splitting of terms commutes with all the other reductions.

Proposition 1.2.2.11. *If $M \twoheadrightarrow_{\beta/\equiv} N$ then*

$$M \twoheadrightarrow_{\text{split}} \cdot (\rightarrow_\beta \cup \rightarrow_+ \cup \rightarrow_{\text{syn}})^*\ \mathrm{can}(N).$$

Proof. $M \twoheadrightarrow_{\beta/\equiv} N$ if and only if there is a reduction

$$M = M_0 \equiv_{\text{syn}} M'_0 \twoheadrightarrow_{\beta/=_+} M_1 \equiv_{\text{syn}} M'_1 \twoheadrightarrow_{\beta/=_+} \ldots \twoheadrightarrow_{\beta/=_+} M_n \equiv_{\text{syn}} M'_n = N.$$

But according to the proposition 1.1.2.3 if $M_i \equiv_{\text{syn}} M'_i$ then $\mathrm{can}(M_i) \equiv_{\text{syn}} \mathrm{can}(M'_i)$, and if $M'_i \twoheadrightarrow_{\beta/=_+} M_{i+1}$ then $\mathrm{can}(M'_i) \twoheadrightarrow_{\beta^c} \mathrm{can}(M_{i+1})$. Thus

$$\begin{array}{lll} M \twoheadrightarrow_+ & & \mathrm{can}(M_0)\ (\rightarrow_{\text{split}} \cup \rightarrow_{\text{syn}})^*\ \mathrm{can}(M'_0)\\ & (\rightarrow_\beta \cup \rightarrow_+)^* & \mathrm{can}(M_1)\ (\rightarrow_{\text{split}} \cup \rightarrow_{\text{syn}})^*\ \mathrm{can}(M'_1)\\ & (\rightarrow_\beta \cup \rightarrow_+)^* & \ldots\\ & (\rightarrow_\beta \cup \rightarrow_+)^* & \mathrm{can}(M_n)\ (\rightarrow_{\text{split}} \cup \rightarrow_{\text{syn}})^*\ \mathrm{can}(M'_n) \quad =_+ N. \end{array}$$

In short if $M \twoheadrightarrow_{\beta/\equiv} N$ then M $(\rightarrow_\beta \cup \rightarrow_+ \cup \rightarrow_{\text{split}} \cup \rightarrow_{\text{syn}})^*$ $\mathrm{can}(N)$. But we proved that the reductions $\rightarrow_{\text{split}}$ commutes with $\rightarrow_\beta$, $\rightarrow_+$ and $\rightarrow_{\text{syn}}$ so this implies that $M \twoheadrightarrow_{\text{split}} \cdot (\rightarrow_\beta \cup \rightarrow_+ \cup \rightarrow_{\text{syn}})^*\ \mathrm{can}(N)$. □

1.2.2.2 Getting rid of $\equiv_{\text{syn}}$

We dealt with the splitting of terms but our characterization of the β-reduction modulo $\equiv$ still involves syntactic equivalences along the reduction. So we will prove that every use of a syntactic rule other than the splitting can be pushed at the end of the reduction.

Proposition 1.2.2.12. *If $M \to_{\text{syn}} N$ then $\operatorname{can}(M) \twoheadrightarrow_{\text{syn}} \operatorname{can}(N)$.*

Proof. The corollary 1.1.2.2 gives that if $M(\leftrightarrow_\gamma \cup \leftrightarrow_\alpha \cup \to_{\mathcal{I}} \cup \to_{\mathcal{Z}})N$ then $\operatorname{can}(M)\ (\leftrightarrow_\gamma \cup \leftrightarrow_\alpha \cup \to_{\mathcal{I}} \cup \to_{\mathcal{Z}})^*\ \operatorname{can}(N)$. We conclude by induction on the length of the reduction $\to_{\text{syn}}$. □

To prove the commutation of $\to_{\text{syn}}$ with the β-reduction we use the notion of independent parallel reduction $\xrightarrow{i}$ defined in section 1.1.2.

Proposition 1.2.2.13. *If M is canonical and $M \to_{\text{syn}} N \xrightarrow{i}_\beta P$ then there is N' such that $M \xrightarrow{i}_\beta N' \twoheadrightarrow_{\text{syn}} P$.*

Proof. The reason we require M to be canonical is that the reduction $\to_{\mathcal{I}}$ erases a sum and can create a β-redex: with non canonical terms we may have $(\lambda x.M +_l \lambda x.M)\ N \to_{\mathcal{I}} (\lambda x.M)\ N \to_\beta M\left[{}^{N}/{}_{x}\right]$, and here we can not commute the reductions.

We reason by induction on the contexts of the reductions.

- If neither context is empty the result is immediate by induction hypothesis.
- If the context of the reduction $\to_{\text{syn}}$ is empty we have six cases:
  - if $M_1 +_l M_2 \to_\gamma M_2 +_{\gamma(l)} M_1 \xrightarrow{i}_\beta P$ then $P = P_2 +_{\gamma(l)} P_1$ with $M_i \xrightarrow{i}_\beta P_i$ for $i \in \{1;2\}$, and

    $$M_1 +_l M_2 \xrightarrow{i}_\beta P_1 +_l P_2 \to_\gamma P_2 +_{\gamma(l)} P_1;$$

  - if

    $$(M_1 +_l M_2) +_{l'} M_3 \to_\alpha M_1 +_{\alpha_1(l,l')} (M_2 +_{\alpha_2(l,l')} M_3) \xrightarrow{i}_\beta P$$

    then $P = P_1 +_{\alpha_1(l,l')} (P_2 +_{\alpha_2(l,l')} P_3)$ with $M_i \xrightarrow{i}_\beta P_i$ for $i \in \{1;2;3\}$ and

    $$(M_1 +_l M_2) +_{l'} M_3 \xrightarrow{i}_\beta (P_1 +_l P_2) +_{l'} P_3 \to_\alpha P_1 +_{\alpha_1(l,l')} (P_2 +_{\alpha_2(l,l')} P_3);$$

  - the cases of $\leftarrow_\gamma$ and $\leftarrow_\alpha$ are similar;
  - if $M +_l M \to_{\mathcal{I}} M \xrightarrow{i}_\beta P$ then

    $$M +_l M \xrightarrow{i}_\beta P +_l P \to_{\mathcal{I}} P;$$

– if $M_1 +_l M_2 \to_{\mathcal{Z}} M_1 +_l M_3 \xrightarrow{i}_\beta P$ then $P = P_1 +_l P_3$ with $M_i \xrightarrow{i}_\beta P_i$ for $i \in \{1;3\}$ and

$$M_1 +_l M_2 \xrightarrow{i}_\beta P_1 +_l M_2 \to_{\mathcal{Z}} P_1 +_l P_3.$$

- If the context of $\to_{\mathrm{syn}}$ is not empty but the context of the β-reduction is, we have $M = M'\ M_2$, $N = (\lambda x.N_1)\ N_2$ and $P = N_1 \left[{}^{N_2}/_{x}\right]$, and by hypothesis M is canonical so M' is not a sum. Then necessarily $M' = \lambda x.M_1$ and we have $M_i \to^?_{\mathrm{syn}} N_i$ for $i \in \{1;2\}$. We get

$$(\lambda x.M_1)\ M_2 \to_\beta M_1 \left[{}^{M_2}/_{x}\right] \twoheadrightarrow_{\mathrm{syn}} N_1 \left[{}^{N_2}/_{x}\right].$$

□

With these two results we can conclude.

Proposition 1.2.2.14. *If* $M\ (\to_\beta \cup \to_+ \cup \to_{\mathrm{syn}})^*\ N$ *then*

$$\mathrm{can}(M) \twoheadrightarrow_{\beta^c} P \twoheadrightarrow_{\mathrm{syn}} \mathrm{can}(N).$$

Proof. If $M\ (\to_\beta \cup \to_+ \cup \to_{\mathrm{syn}})^*\ N$ then we have

$$\mathrm{can}(M) = M_0 \twoheadrightarrow_{\mathrm{syn}} M_0' \to_{\beta^c} M_1 \twoheadrightarrow_{\mathrm{syn}} M_1' \to_{\beta^c} \ldots \to_{\beta^c} M_n \twoheadrightarrow_{\mathrm{syn}} M_n' = \mathrm{can}(N)$$

where the terms M_i and M_i' are all canonical.

But if $M_i \twoheadrightarrow_{\mathrm{syn}} M_i' \to_\beta P_i \twoheadrightarrow_+ M_{i+1}$ we know there is a term Q_i such that $M_i \xrightarrow{i}_\beta Q_i \twoheadrightarrow_{\mathrm{syn}} P_i \twoheadrightarrow_+ M_{i+1}$. Then we also have $\mathrm{can}(Q_i) \twoheadrightarrow_{\mathrm{syn}} \mathrm{can}(P_i)$ and $\mathrm{can}(P_i) = M_{i+1}$. Hence $M_i \twoheadrightarrow_{\beta^c} \mathrm{can}(Q_i) \twoheadrightarrow_{\mathrm{syn}} M_{i+1}$.

We conclude by induction on the length n of the reduction. □

We can sum up all the results we have so far about the β-reduction modulo $\equiv$.

Theorem 1.2.2.15. *If* $M \twoheadrightarrow_{\beta/\equiv} N$ *then*

$$\mathrm{can}(M) \twoheadrightarrow_{\mathrm{split}} P \twoheadrightarrow_{\beta^c} Q \twoheadrightarrow_{\mathrm{syn}} \mathrm{can}(N).$$

Proof. If $M \twoheadrightarrow_{\beta/\equiv} N$ then $M \twoheadrightarrow_{\mathrm{split}} P\ (\to_\beta \cup \to_+ \cup \to_{\mathrm{syn}})^*\ \mathrm{can}(N)$. According to the corollary 1.1.2.2 we have $\mathrm{can}(M) \twoheadrightarrow_{\mathrm{split}} \mathrm{can}(P)$, and according to the previous result we have $\mathrm{can}(P) \twoheadrightarrow_{\beta^c} Q \twoheadrightarrow_{\mathrm{syn}} \mathrm{can}(N)$. □

Using this result we can prove the confluence of the reduction modulo $\equiv$.

Corollary 1.2.2.16. $\twoheadrightarrow_{\beta/\equiv}$ *is confluent.*

Proof. Let us assume we have two reductions $M \twoheadrightarrow_{\mathrm{split}} M_i \twoheadrightarrow_{\beta^c} N_i \twoheadrightarrow_{\mathrm{syn}} P_i$ for $i \in \{1;2\}$, where all terms are canonical. We know that $\twoheadrightarrow_{\mathrm{split}}$ is confluent so there is M' such that $M_i \twoheadrightarrow_{\mathrm{split}} M'$, or equivalently $M' \twoheadrightarrow_{\mathcal{I}} M_i$, for $i \in \{1;2\}$. But then

$M' \twoheadrightarrow_{\mathcal{I}} M_i \twoheadrightarrow_{\beta^c} N_i \twoheadrightarrow_{\mathrm{syn}} P_i$, hence $M' \left(\to_\beta \cup \to_+ \cup \to_{\mathrm{syn}}\right)^* P_i$, and we know there is M_i' canonical such that $M' \twoheadrightarrow_{\beta^c} M_i' \twoheadrightarrow_{\mathrm{syn}} P_i$.

Now we have $M' \twoheadrightarrow_{\beta^c} M_i'$ for $i \in \{1; 2\}$ and we know that $\to_{\beta^c}$ is also confluent so there is Q such that $M_i' \twoheadrightarrow_{\beta^c} Q$. But then we have both $M_i' \twoheadrightarrow_{\beta^c} Q$ and $M_i' \twoheadrightarrow_{\mathrm{syn}} P_i$ so $P_i \twoheadrightarrow_{\beta/\equiv} Q$.

$$\begin{tikzcd}
 & & & M \arrow[dl, two heads, "\mathrm{split}"'] \arrow[dr, two heads, "\mathrm{split}"] & & & \\
 & & M_1 \arrow[dl, two heads, "\beta^c"'] & & M_2 \arrow[dr, two heads, "\beta^c"] & & \\
 & N_1 \arrow[dl, two heads, "\mathrm{syn}"'] & & M' \arrow[ul, dashed, two heads, "\mathcal{I}"] \arrow[ur, dashed, two heads, "\mathcal{I}"'] \arrow[dl, dashed, two heads, "\beta^c"'] \arrow[dr, dashed, two heads, "\beta^c"] & & N_2 \arrow[dr, two heads, "\mathrm{syn}"] & \\
P_1 & & M_1' \arrow[ll, dashed, two heads, "\mathrm{syn}"] \arrow[dr, dashed, two heads, "\beta^c"'] & & M_2' \arrow[rr, dashed, two heads, "\mathrm{syn}"'] \arrow[dl, dashed, two heads, "\beta^c"] & & P_2 \\
 & & & Q & & &
\end{tikzcd}$$

$\square$

2 Standardization

A very useful property of the usual deterministic β-reduction is the standardization: every β-reduction can be turned into a reduction where the redexes are reduced from the left to the right. More precisely:

- the reduction

  $$\lambda x_1...x_n.(\lambda y.M)\ P\ Q_1\ ...\ Q_m \rightarrow_\beta \lambda x_1...x_n.M\left[{}^{P}/_{y}\right]\ Q_1\ ...\ Q_m \twoheadrightarrow_\beta N$$

  is standard if the reduction $\lambda x_1...x_n.M\left[{}^{P}/_{y}\right]\ Q_1\ ...\ Q_m \twoheadrightarrow_\beta N$ is standard;

- the reduction

  $$\begin{aligned} \lambda x_1...x_n.(\lambda y.M)\ P_1\ P_2\ ...\ P_m &\twoheadrightarrow_\beta \lambda x_1...x_n.(\lambda y.N)\ P_1\ P_2\ ...\ P_m \\ &\twoheadrightarrow_\beta \lambda x_1...x_n.(\lambda y.N)\ Q_1\ P_2\ ...\ P_m \\ &\twoheadrightarrow_\beta ... \\ &\twoheadrightarrow_\beta \lambda x_1...x_n.(\lambda y.N)\ Q_1\ Q_2\ ...\ Q_m \end{aligned}$$

  is standard if the reductions $M \twoheadrightarrow_\beta N$ and $P_i \twoheadrightarrow_\beta Q_i$ for $i \leq m$ are standard;

- the reduction

  $$\begin{aligned} \lambda x_1...x_n.y\ P_1\ P_2\ ...\ P_m &\twoheadrightarrow_\beta \lambda x_1...x_n.y\ Q_1\ P_2\ ...\ P_m \\ &\twoheadrightarrow_\beta ... \\ &\twoheadrightarrow_\beta \lambda x_1...x_n.y\ Q_1\ Q_2\ ...\ Q_m \end{aligned}$$

  is standard if the reductions $P_i \twoheadrightarrow_\beta Q_i$ for $i \leq m$ are standard.

We can also characterize the existence of a standard reduction between terms by:

- there is a standard reduction between $\lambda x_1...x_n.(\lambda y.M)\ P\ Q_1\ ...\ Q_m$ and N if there is a standard reduction between $\lambda x_1...x_n.M\left[{}^{P}/_{y}\right]\ Q_1\ ...\ Q_m$ and N;

- there is a standard reduction between $\lambda x_1...x_n.(\lambda y.M)\ P_1\ ...\ P_m$ and $\lambda x_1...x_n.(\lambda y.N)\ Q_1\ ...\ Q_m$ if there are standard reductions between M and N and between P_i and Q_i for $i \leq m$;

- there is a standard reduction between $\lambda x_1 ... x_n.y\ P_1\ ...\ P_m$ and $\lambda x_1 ... x_n.y\ Q_1\ ...\ Q_m$ if there is a standard reduction between P_i and Q_i for all $i \leq m$.

In the usual λ-calculus there is a standard reduction from M to N if and only if there is a reduction $M \twoheadrightarrow_\beta N$.

Given the structure of canonical terms it seems fairly simple to extend the notion of standard reduction to our calculus. However the fact that canonical terms are not stable by β-reduction implies that we have a choice to make.

Consider the term $(\lambda x.(\lambda y.y)\ (u +_l v))\ M$. If there are reductions $M \rightarrow_\beta M_1$ and $M \rightarrow_\beta M_2$ (necessarily standard as they are of length 1), we have:

$$\begin{aligned}(\lambda x.(\lambda y.y)\ (u +_l v))\ M &\rightarrow_\beta (\lambda x.(u +_l v))\ M \\ &\twoheadrightarrow_+ (\lambda x.u)\ M +_l (\lambda x.v)\ M \\ &\rightarrow_\beta (\lambda x.u)\ M_1 +_l (\lambda x.v)\ M \\ &\rightarrow_\beta (\lambda x.u)\ M_1 +_l (\lambda x.v)\ M_2.\end{aligned}$$

Is this reduction standard? If we look at the shape of the reduction itself, it is indeed from left to right and it seems natural to call it standard. But if we look at the inductive characterization of the existence of a standard reduction we would rather say that a standard reduction of $(\lambda x.(\lambda y.y)\ (u +_l v))\ M$ which does not reduce the head redex is given by a standard reductions for $(\lambda y.y)\ (u +_l v)$ and another for M, and this is not the case here.

So given a term $\lambda x_1 ... x_n.(\lambda y.M)\ P_1\ ...\ P_m$ we must decide whether we allow the canonicalization following the reduction of M to duplicate the P_i's before we reduce them, or if we must decide of all the reduction in parallel.

Before dealing with this matter let us make another remark: the standardization fails in our calculus. Let us consider the following reduction:

$$\begin{aligned}(\lambda x.(\lambda y.y)\ (u +_l v))\ M &\rightarrow_\beta (\lambda x.(u +_l v))\ M \\ &\twoheadrightarrow_+ (\lambda x.u)\ M +_l (\lambda x.v)\ M \\ &\rightarrow_\beta u +_l (\lambda x.v)\ M.\end{aligned}$$

In a standard reduction the head redex $(\lambda x.(\lambda y.y)\ (u +_l v))\ M$ is either reduced first or it is never reduced. If we do reduce it we get

$$(\lambda x.(\lambda y.y)\ (u +_l v))\ M \rightarrow_\beta (\lambda y.y)\ (u +_l v)$$

and this term can not reduce to $u +_l (\lambda x.v)\ M$. If on the other hand we do not reduce it then the only other choice is

$$(\lambda x.(\lambda y.y)\ (u +_l v))\ M \rightarrow_\beta (\lambda x.(u +_l v))\ M$$

and we are not allowed to reduce the remaining redex.

Still we immediately see that we have a reduction

$$(\lambda x.(\lambda y.y)\ (u +_l v))\ M \to_\beta (\lambda y.y)\ (u +_l v) \to_\beta u +_l v$$

as well as

$$u +_l (\lambda x.v)\ M \to_\beta u +_l v.$$

The standardization fails because the canonicalization can duplicate redexes and we may reduce only some copies, not all of them. Then we can build a standardizable reduction by reducing the remaining copies of those redexes.

We will prove that every reduction can be completed into a standardizable one. Now to get back to our other problem, we had to decide whether we allowed $(\lambda x.(\lambda y.y)\ (u +_l v))\ M$ to have a standard reduction to $(\lambda x.u)\ M_1 +_l (\lambda x.v)\ M_2$. But the reduction being confluent we know there is N such that $M_1 \twoheadrightarrow_\beta N$ and $M_2 \twoheadrightarrow_\beta N$. So we can extend this reduction by

$$(\lambda x.u)\ M_1 +_l (\lambda x.v)\ M_2 \twoheadrightarrow_\beta (\lambda x.u)\ N +_l (\lambda x.v)\ N$$

and we do have a standard reduction in the stronger sense from $(\lambda x.(\lambda y.y)\ (u +_l v))\ M$ to $(\lambda x.u)\ N +_l (\lambda x.v)\ N$.

Since we can not standardize every reduction but we expect to be able extend them into standardizable ones, and since every standard reduction in the weaker sense can be extended into a standard one in the stronger sense, we do not lose anything if we use the stronger definition of standardization.

Definition 2.0.2.1. We define the relation $M \twoheadrightarrow_S N$ between canonical terms saying that there is a standard relation from M to N by:

$$\frac{M_1 \twoheadrightarrow_S N_1 \qquad M_2 \twoheadrightarrow_S N_2}{M_1 +_l M_2 \twoheadrightarrow_S N_1 +_l N_2} \qquad \frac{\operatorname{can}\left(\lambda x_1 ... x_n.v\left[{}^{P}/_{y}\right]\ Q_1\ ...\ Q_m\right) \twoheadrightarrow_S N}{\lambda x_1 ... x_n.(\lambda y.v)\ P\ Q_1\ ...\ Q_m \twoheadrightarrow_S N}$$

$$\frac{v \twoheadrightarrow_S N \qquad \forall i \leq m, P_i \twoheadrightarrow_S Q_i \qquad m > 0}{\lambda x_1 ... x_n.(\lambda y.v)\ P_1\ ...\ P_m \twoheadrightarrow_S \operatorname{can}(\lambda x_1 ... x_n.(\lambda y.N)\ Q_1\ ...\ Q_m)}$$

$$\frac{\forall i \leq m, P_i \twoheadrightarrow_S Q_i}{\lambda x_1 ... x_n.y\ P_1\ ...\ P_m \twoheadrightarrow_S \lambda x_1 ... x_n.y\ Q_1\ ...\ Q_m}$$

Proposition 2.0.2.1. *If $M \twoheadrightarrow_S N$ then $M \twoheadrightarrow_{\beta^c} N$.*

Proof. By a simple induction on $\twoheadrightarrow_S$. □

Before dealing with the general case we will consider a restriction of the reduction $\to_\beta \cup \to_+$ for which the standardization holds directly. We will then show that every reduction can be extended into a standardizable one.

2.1 Standardization results

When we defined our operational semantics we wanted be able to simulate the usual probabilistic head reduction of probabilistic terms. We wanted to have for every head context C that $C[M +_l N] \twoheadrightarrow_+ C[M] +_l C[N]$. But we could adopt a different point of view on the reduction of sums: the reason we need to reduce them in the first place is that sums may prevent us from performing β-reduction. For instance if we consider the term $(\lambda x.M +_l N)\ P$, we can not β-reduce it, but we have the reduction

$$(\lambda x.M +_l N)\ P \to_+ (\lambda x.M)\ P +_l N\ P \to_\beta M\left[{}^{P}/_{x}\right] +_l N\ P.$$

So the only necessary reduction rule is

$$(M +_l N)\ P \to_{+^{\mathrm{w}}} M\ P +_l N\ P.$$

While the reduction $\to_+$ corresponds to a head reduction strategy on sums, the reduction $\to_{+^{\mathrm{w}}}$ is closer to a weak head reduction strategy.

An interesting consequence to this restriction is that every weak reduction $(\to_\beta \cup \to_{+^{\mathrm{w}}})^*$ is standardizable.

2.1.1 Weak reduction of sums

First we can see that restricting $\to_+$ to $\to_{+^{\mathrm{w}}}$ does not really change the structure of our calculus.

Proposition 2.1.1.1. *$\to_{+^{\mathrm{w}}}$ is confluent and strongly normalizing.*

Proof. The strong normalization of $\to_+$ (proposition 1.1.1.2) immediately gives the strong normalization of $\to_{+^{\mathrm{w}}}$, and the proof of proposition 1.1.1.1 giving the weak confluence of $\to_+$ can be used to prove the weak confluence of $\to_{+^{\mathrm{w}}}$. □

Definition 2.1.1.1. We call *weak canonical form* of a term M its unique normal form $\mathrm{can}^{\mathrm{w}}(M)$ for $\to_{+^{\mathrm{w}}}$. We note $\to_{\mathrm{can}^{\mathrm{w}}}$ the corresponding canonicalizing reduction $M \to_{\mathrm{can}^{\mathrm{w}}} \mathrm{can}^{\mathrm{w}}(M)$.

Proposition 2.1.1.2. *The weakly canonical terms are given by:*

$$\begin{aligned} M, N &:= v \mid M +_l N \\ v &:= x \mid \lambda x.M \mid v\ M. \end{aligned}$$

Proof. Every such term is weakly canonical, and we prove the converse following the proof of proposition 1.1.1.5. □

We can also define the weakly canonical parallel β-reduction following the results of the section 1.2, using the labelled terms along with the reduction

$$(M +_l N)\ P \to_{+^{\mathrm{w}}} M\ P +_l N\ P$$

extended to context.

Proposition 2.1.1.3. *$\to_{\beta_*} \cup \to_{+^{\mathrm{w}}}$ is confluent and strongly normalizing.*

Proof. Again the strong normalization is given by the strong normalization of $\to_{\beta_*} \cup \to_{+}$ from proposition 1.2.1.3, and the weak confluence is derived from the proof of proposition 1.2.1.2. □

Definition 2.1.1.2. Given a weakly canonical term M and a set $\mathcal{F}$ of β-redexes in M we write $M \xrightarrow{\mathcal{F}}_{\beta^{\mathrm{w}}_{/\!/}} N$ if N is the unique normal form of $M_{\mathcal{F}}$ for $\to_{\beta_*} \cup \to_{+^{\mathrm{w}}}$.

We write $M \to_{\beta^{\mathrm{w}}_{/\!/}} N$ whenever there exists $\mathcal{F}$ such that $M \xrightarrow{\mathcal{F}}_{\beta^{\mathrm{w}}_{/\!/}} N$.

Proposition 2.1.1.4. *$\to_{\beta^{\mathrm{w}}_{/\!/}}$ has the diamond property.*

Proof. If $M \xrightarrow{\mathcal{F}_1}_{\beta^{\mathrm{w}}_{/\!/}} N_1$ and $M \xrightarrow{\mathcal{F}_2}_{\beta^{\mathrm{w}}_{/\!/}} N_2$ then let P such that $M \xrightarrow{\mathcal{F}_1 \cup \mathcal{F}_2}_{\beta^{\mathrm{w}}_{/\!/}} P$, we have $N_i \to_{\beta^{\mathrm{w}}_{/\!/}} P$ for $i \in \{1; 2\}$. □

Corollary 2.1.1.5. *$\to_{\beta} \cup \to_{+^{\mathrm{w}}}$ is confluent.*

Definition 2.1.1.3. The weakly canonical β-reduction $\to_{\beta^{\mathrm{w}}}$ between weakly canonical terms is $\to_{\beta} \cdot \to_{\mathrm{can}^{\mathrm{w}}}$.

Proposition 2.1.1.6. *If $M \to_{\beta} N$ then $\mathrm{can}^{\mathrm{w}}(M) \twoheadrightarrow_{\beta^{\mathrm{w}}} \mathrm{can}^{\mathrm{w}}(N)$.*

Proof. We proceed as for the proposition 1.2.1.8. We prove that if $P \xrightarrow{\mathcal{F}}_{\beta^{\mathrm{w}}_{/\!/}} Q$ then $\mathrm{can}^{\mathrm{w}}(P) \twoheadrightarrow_{\beta^{\mathrm{w}}} Q$ by induction on the bound on the length of the reductions $P\ (\to_{\beta} \cup \to_{+^{\mathrm{w}}})^* \ Q$. □

Corollary 2.1.1.7. *If $M\ (\to_{\beta} \cup \to_{+^{\mathrm{w}}})^*\ N$ then $\mathrm{can}^{\mathrm{w}}(M) \twoheadrightarrow_{\beta^{\mathrm{w}}} \mathrm{can}^{\mathrm{w}}(N)$.*

We can prove a standardization theorem by following the same idea as in the deterministic case. The key result is that every reduction can be turned into head reductions followed by internal reductions. This fails in general but is works when we consider only the weak reduction of sums, as in this case head redexes are never duplicated.

Definition 2.1.1.4. We define the *head β-redexes* and *internal β-redexes* of a weakly canonical term as follows:

- a redex in $M +_l N$ is a head redex (resp. an internal redex) if it is a head redex of M or a head redex of N (resp. an internal redex of M or an internal redex of N);
- a redex in $\lambda x.M$ is a head redex (resp. an internal redex) if it is a head redex of M (resp. an internal redex of M);
- the redex $(\lambda x.M)\ N$ is a head redex in $(\lambda x.M)\ N\ P_1\ \ldots\ P_m$;

- a redex is internal in $(\lambda x.M)\ P_1\ ...\ P_m$ with $m > 0$ or in $y\ P_1\ ...\ P_m$ if it is a redex in M or a P_i for some $i \leq m$.

Definition 2.1.1.5. 1. We write $M \to_h N$ if $M \to_\beta N$ by the reduction of a head redex.

2. We write $M \to_i N$ if $M \to_\beta N$ by the reduction of an internal redex.

Proposition 2.1.1.8. *1. If M is weakly canonical and $M \to_h N$ then the residuals of a head redex in M are head redexes in* $\mathrm{can}^{\mathrm{w}}(N)$.

2. If M is weakly canonical and $M \to_i N$ then the residuals of an internal redex in M are internal redexes in $\mathrm{can}^{\mathrm{w}}(N)$.

Proof. By induction on M.

- If $M = M_1 +_l M_2$ and $M \to_h N$ (resp. $M \to_i N$) then $N = N_1 +_l N_2$ and $\mathrm{can}^{\mathrm{w}}(N) = \mathrm{can}^{\mathrm{w}}(N_1) +_l \mathrm{can}^{\mathrm{w}}(N_2)$ with $M_i \to_h^? N_i$ (resp. $M_i \to_i^? N_i$) for $i \in \{1; 2\}$ so the result is immediate by induction hypothesis.

- If $M = \lambda x.M_0$ and $M \to_h N$ (resp. $M \to_i N$) then $N = \lambda x.N_0$ and $\mathrm{can}^{\mathrm{w}}(N) = \lambda x.\mathrm{can}^{\mathrm{w}}(N_0)$ with $M_0 \to_h N_0$ (resp. $M_0 \to_i N_0$) so the result is immediate by induction hypothesis.

- If $M = y\ P_1\ ...\ P_m$ then M has no head redex, and if $M \to_i N$ then we have $\mathrm{can}^{\mathrm{w}}(N) = y\ \mathrm{can}^{\mathrm{w}}(Q_1)\ ...\ \mathrm{can}^{\mathrm{w}}(Q_m)$ with $P_i \to_\beta^? Q_i$ for $i \leq m$, hence every redex in $\mathrm{can}^{\mathrm{w}}(N)$ is internal.

- If $M = (\lambda x.M_0)\ P\ Q_1\ ...\ Q_m$ and $M \to_h N$ then $N = M_0 \left[{}^{P}/_{x}\right]\ Q_1\ ...\ Q_m$ and the unique head redex $(\lambda x.M_0)\ P$ has no residual in $\mathrm{can}^{\mathrm{w}}(N)$.

- If $M = (\lambda x.M_0)\ P_1\ ...\ P_m$ and $M \to_i N$ we have

$$\mathrm{can}^{\mathrm{w}}(N) = (\lambda x.\mathrm{can}^{\mathrm{w}}(N_0))\ \mathrm{can}^{\mathrm{w}}(Q_1)\ ...\ \mathrm{can}^{\mathrm{w}}(Q_m)$$

  with $M_0 \to_\beta^? N_0$ and $P_i \to_\beta^? Q_i$ for $i \leq m$. Then it is easy to see that the residuals of an internal redex of M are internal in $\mathrm{can}^{\mathrm{w}}(N)$.

□

As head and internal redexes are preserved respectively by head and internal reductions we have that if $M \xrightarrow{\mathcal{F}}_{\beta^{\mathrm{w}}_{/\!/}} N$ and $\mathcal{F}$ is a set of head (resp. internal) redexes then $M\ (\to_h \cup \to_{+^{\mathrm{w}}})^*\ N$ (resp. $M\ (\to_i \cup \to_{+^{\mathrm{w}}})^*\ N$). Thus it makes sense to define head and internal parallel reductions.

Definition 2.1.1.6. Given a parallel reduction $M \xrightarrow{\mathcal{F}}_{\beta^{\mathrm{w}}_{/\!/}} N$ we write $M \xrightarrow{\mathcal{F}}_{h^{\mathrm{w}}_{/\!/}} N$ if $\mathcal{F}$ is a set of head redexes in M, and we write $M \xrightarrow{\mathcal{F}}_{i^{\mathrm{w}}_{/\!/}} N$ if $\mathcal{F}$ is a set of internal redexes.

Again we write $M \to_{h^{\mathrm{w}}_{/\!/}} N$ (resp. $M \to_{i^{\mathrm{w}}_{/\!/}} N$) if there is $\mathcal{F}$ such that $M \xrightarrow{\mathcal{F}}_{h^{\mathrm{w}}_{/\!/}} N$ (resp. $M \xrightarrow{\mathcal{F}}_{i^{\mathrm{w}}_{/\!/}} N$).

Every reduction $M \twoheadrightarrow_{\beta^{\mathrm{w}}} N$ between weakly canonical terms can be seen as a parallel reduction $M \twoheadrightarrow_{\beta^{\mathrm{w}}_{/\!/}} N$. Then we want to prove that there is always a reduction $M \twoheadrightarrow_{h^{\mathrm{w}}_{/\!/}} \cdot \twoheadrightarrow_{i^{\mathrm{w}}_{/\!/}} N$.

Proposition 2.1.1.9. *If $M \to_{\beta^{\mathrm{w}}_{/\!/}} N$ then $M \twoheadrightarrow_{h^{\mathrm{w}}_{/\!/}} \cdot \to_{i^{\mathrm{w}}_{/\!/}} N$.*

Proof. Let $\mathcal{F}$ be such that $M \xrightarrow{\mathcal{F}}_{\beta^{\mathrm{w}}_{/\!/}} N$. We reason by induction on the bound on the length of the reduction of $\mathcal{F}$.

If $\mathcal{F}$ is a set of internal redexes of M then we already have $M \to_{i^{\mathrm{w}}_{/\!/}} N$. Otherwise let $\mathcal{F}' \subset \mathcal{F}$ be the (non empty) set of head redexes in $\mathcal{F}$, we have $M \xrightarrow{\mathcal{F}'}_{h^{\mathrm{w}}_{/\!/}} M'$ and $M' \to_{\beta^{\mathrm{w}}_{/\!/}} N$. Then by induction hypothesis $M' \twoheadrightarrow_{h^{\mathrm{w}}_{/\!/}} \cdot \to_{i^{\mathrm{w}}_{/\!/}} N$. □

Proposition 2.1.1.10. *If $M \to_{i^{\mathrm{w}}_{/\!/}} N$ then every head redex in N is the unique residual of a head redex in M.*

Proof. We reason by induction on M.

- If M is a sum or an abstraction then the result is immediate by induction hypothesis.
- If $M = y\ P_1\ ...\ P_m$ then $N = y\ Q_1\ ...\ Q_m$ with $P_i \to_{\beta^{\mathrm{w}}_{/\!/}} Q_i$ for $i \leq m$ so N has no head redex.
- If $M = (\lambda x.M_0)\ P_1\ ...\ P_m$ then $N = (\lambda x.N_0)\ Q_1\ ...\ Q_m$ with $M_0 \to_{\beta^{\mathrm{w}}_{/\!/}} N_0$ and $P_i \to_{\beta^{\mathrm{w}}_{/\!/}} Q_i$ for $i \leq m$ so the unique head redex of N is the residual of the unique head redex of M.

□

Proposition 2.1.1.11. *If $M \to_{i^{\mathrm{w}}_{/\!/}} \cdot \to_{h^{\mathrm{w}}_{/\!/}} N$ then $M \to_{\beta^{\mathrm{w}}_{/\!/}} N$ and $M \twoheadrightarrow_{h^{\mathrm{w}}_{/\!/}} \cdot \to_{i^{\mathrm{w}}_{/\!/}} N$.*

Proof. If $M \xrightarrow{\mathcal{F}}_{i^{\mathrm{w}}_{/\!/}} P \xrightarrow{\mathcal{G}}_{h^{\mathrm{w}}_{/\!/}} N$ then every element of $\mathcal{G}$ is the unique residual of a head redex in M, so there is a set $\mathcal{G}'$ of head redexes in M such that $M_{\mathcal{F}\cup\mathcal{G}'}(\to_{\beta_*} \cup \to_{+^{\mathrm{w}}})^* P_{\mathcal{G}}$, hence $M \xrightarrow{\mathcal{F}\cup\mathcal{G}'}_{\beta^{\mathrm{c}}_{/\!/}} N$. □

Proposition 2.1.1.12. *If $M \to_{i^{\mathrm{w}}_{/\!/}} \cdot \twoheadrightarrow_{h^{\mathrm{w}}_{/\!/}} N$ then $M \twoheadrightarrow_{h^{\mathrm{w}}_{/\!/}} \cdot \to_{i^{\mathrm{w}}_{/\!/}} N$.*

Proof. By induction on the number of head reductions. If there is none the result is immediate. Otherwise $M \to_{i^{\mathrm{w}}_{/\!/}} \cdot \to_{h^{\mathrm{w}}_{/\!/}} \cdot \to^{n}_{h^{\mathrm{w}}_{/\!/}} N$ so according to the previous result $M \twoheadrightarrow_{h^{\mathrm{w}}_{/\!/}} \cdot \to_{i^{\mathrm{w}}_{/\!/}} \cdot \to^{n}_{h^{\mathrm{w}}_{/\!/}} N$, and by induction hypothesis $M \twoheadrightarrow_{h^{\mathrm{w}}_{/\!/}} \cdot \twoheadrightarrow_{h^{\mathrm{w}}_{/\!/}} \cdot \to_{i^{\mathrm{w}}_{/\!/}} N$. □

Proposition 2.1.1.13. *If $M\ (\to_{h^{\mathrm{w}}_{/\!/}} \cup \to_{i^{\mathrm{w}}_{/\!/}})^*\ N$ with n internal reduction steps then $M \twoheadrightarrow_{h^{\mathrm{w}}_{/\!/}} \cdot \to^{n}_{i^{\mathrm{w}}_{/\!/}} N$.*

Proof. By induction on n. If $n = 0$ the result is immediate. Otherwise

$$M \ (\to_{h^{\mathrm{w}}_{/\!/}} \cup \to_{i^{\mathrm{w}}_{/\!/}})^* \cdot \to_{i^{\mathrm{w}}_{/\!/}} \cdot \twoheadrightarrow_{h^{\mathrm{w}}_{/\!/}} N$$

with $n+1$ internal reductions so according to the previous result

$$M \ (\to_{h^{\mathrm{w}}_{/\!/}} \cup \to_{i^{\mathrm{w}}_{/\!/}})^* \cdot \twoheadrightarrow_{h^{\mathrm{w}}_{/\!/}} \cdot \to_{i^{\mathrm{w}}_{/\!/}} N$$

and by induction hypothesis $M \twoheadrightarrow_{h^{\mathrm{w}}_{/\!/}} \cdot \to^n_{i^{\mathrm{w}}_{/\!/}} \cdot \to_{i^{\mathrm{w}}_{/\!/}} N$. □

2.1.2 Strong and weak standardization theorems

As we claimed before the β-reduction with weak reduction of sums enjoys a standardization property. Let us first adapt the definition of the relation $\twoheadrightarrow_S$ to the weak case.

Definition 2.1.2.1. We define the relation $M \twoheadrightarrow_{S^{\mathrm{w}}} N$ between weakly canonical terms by:

$$\frac{M_1 \twoheadrightarrow_{S^{\mathrm{w}}} N_1 \qquad M_2 \twoheadrightarrow_{S^{\mathrm{w}}} N_2}{M_1 +_l M_2 \twoheadrightarrow_{S^{\mathrm{w}}} N_1 +_l N_2} \qquad \frac{M \twoheadrightarrow_{S^{\mathrm{w}}} N}{\lambda x.M \twoheadrightarrow_{S^{\mathrm{w}}} \lambda x.N}$$

$$\frac{\mathrm{can}^{\mathrm{w}}\left(M\left[{}^{P}/_{y}\right]\ Q_1\ ...\ Q_m\right) \twoheadrightarrow_{S^{\mathrm{w}}} N}{(\lambda y.M)\ P\ Q_1\ ...\ Q_m \twoheadrightarrow_{S^{\mathrm{w}}} N}$$

$$\frac{M \twoheadrightarrow_{S^{\mathrm{w}}} N \qquad \forall i \leq m, P_i \twoheadrightarrow_{S^{\mathrm{w}}} Q_i \qquad m > 0}{(\lambda y.M)\ P_1\ ...\ P_m \twoheadrightarrow_{S^{\mathrm{w}}} \mathrm{can}^{\mathrm{w}}\left((\lambda y.N)\ Q_1\ ...\ Q_m\right)}$$

$$\frac{\forall i \leq m, P_i \twoheadrightarrow_{S^{\mathrm{w}}} Q_i}{y\ P_1\ ...\ P_m \twoheadrightarrow_{S^{\mathrm{w}}} y\ Q_1\ ...\ Q_m}$$

Proposition 2.1.2.1. *If $M \twoheadrightarrow_{S^{\mathrm{w}}} N$ then $M \twoheadrightarrow_{\beta^{\mathrm{w}}} N$.*

Proof. By induction on $\twoheadrightarrow_{S^{\mathrm{w}}}$. □

Now let us prove the standardization theorem.

Proposition 2.1.2.2. *If $M \twoheadrightarrow_{h^{\mathrm{w}}_{/\!/}} \cdot \twoheadrightarrow_{S^{\mathrm{w}}} N$ then $M \twoheadrightarrow_{S^{\mathrm{w}}} N$.*

Proof. By induction on the number of head reductions. If there is none the result is immediate. Otherwise $M \to_{h^{\mathrm{w}}_{/\!/}} \cdot \twoheadrightarrow_{h^{\mathrm{w}}_{/\!/}} \cdot \twoheadrightarrow_{S^{\mathrm{w}}} N$ and by induction hypothesis $M \to_{h^{\mathrm{w}}_{/\!/}} \cdot \twoheadrightarrow_{S^{\mathrm{w}}} N$. We reason by induction on M. Remark that if $M \xrightarrow{\emptyset}_{h^{\mathrm{w}}_{/\!/}} \cdot \twoheadrightarrow_{S^{\mathrm{w}}} N$ then $M \twoheadrightarrow_{S^{\mathrm{w}}} N$.

- If $M = M_1 +_l M_2$ then $N = N_1 +_l N_2$ with $M_i \to_{h^{\mathrm{w}}_{/\!/}} \cdot \twoheadrightarrow_{S^{\mathrm{w}}} N_i$ for $i \in \{1; 2\}$ so we conclude by induction hypothesis.

- If $M = \lambda x.M_0$ then $N = \lambda x.N_0$ with $M_0 \rightarrow_{h^{\mathrm{w}}_{/\!/}} \cdot \twoheadrightarrow_{S^{\mathrm{w}}} N_0$ so we conclude by induction hypothesis.

- If $M = y\ P_1\ ...\ P_m$ then M has no head redex.

- If $M = (\lambda x.M_0)\ P\ Q_1\ ...\ Q_m$ then either $M \xrightarrow{\emptyset}_{h^{\mathrm{w}}_{/\!/}} \cdot \twoheadrightarrow_{S^{\mathrm{w}}} N$ and we have $M \twoheadrightarrow_{S^{\mathrm{w}}} N$, or we have $\mathrm{can}^{\mathrm{w}}\left(M_0\left[{}^{P}/_{x}\right]\ Q_1\ ...\ Q_m\right) \twoheadrightarrow_{S^{\mathrm{w}}} N$. In the second case we have by definition of $\twoheadrightarrow_{S^{\mathrm{w}}}$ that $M \twoheadrightarrow_{S^{\mathrm{w}}} N$.

$\square$

Theorem 2.1.2.3. *Given two weakly canonical terms M and N, if there is a reduction $M\ (\rightarrow_\beta \cup \rightarrow_{+^{\mathrm{w}}})^*\ N$ then $M \twoheadrightarrow_{S^{\mathrm{w}}} N$.*

Proof. If $M\ (\rightarrow_\beta \cup \rightarrow_{+^{\mathrm{w}}})^*\ N$ then $M \twoheadrightarrow_{\beta^{\mathrm{w}}} N$, so $M \twoheadrightarrow_{\beta^{\mathrm{w}}_{/\!/}} N$. We reason by induction on N.

The propositions 2.1.1.9 and 2.1.1.13 give that $M \twoheadrightarrow_{h^{\mathrm{w}}_{/\!/}} M' \twoheadrightarrow_{i^{\mathrm{w}}_{/\!/}} N$, and the previous result gives that if $M' \twoheadrightarrow_{S^{\mathrm{w}}} N$ then $M \twoheadrightarrow_{S^{\mathrm{w}}} N$. So we need to prove that if $M \twoheadrightarrow_{i^{\mathrm{w}}_{/\!/}} N$ then $M \twoheadrightarrow_{S^{\mathrm{w}}} N$.

- If $N = N_1 +_l N_2$ then $M = M_1 +_l M_2$ with $M_i \twoheadrightarrow_{i^{\mathrm{w}}_{/\!/}} N_i$ for $i \in \{1; 2\}$ so we conclude by induction hypothesis.

- If $N = \lambda x.N_0$ then $M = \lambda x.M_0$ with $M_0 \twoheadrightarrow_{i^{\mathrm{w}}_{/\!/}} N_0$ so we conclude by induction hypothesis.

- If $N = y\ Q_1\ ...\ Q_m$ then $M = y\ P_1\ ...\ P_m$ with $P_i \twoheadrightarrow_{\beta^{\mathrm{c}}_{/\!/}} Q_i$ for $i \leq m$ so by induction hypothesis $P_i \twoheadrightarrow_{S^{\mathrm{w}}} Q_i$.

- If $N = (\lambda x.Q_0)\ Q_1\ ...\ Q_m$ then $M = (\lambda x.P_0)\ P_1\ ...\ P_m$ with $P_i \twoheadrightarrow_{\beta^{\mathrm{c}}_{/\!/}} Q_i$ for $i \leq m$ so by induction hypothesis $P_i \twoheadrightarrow_{S^{\mathrm{w}}} Q_i$.

$\square$

From this result we can derive a standardization theorem for the full reduction $\rightarrow_\beta \cup \rightarrow_+$. We will proceed in two steps. First we need to turn any reduction into a weak one, and secondly we will need to derive a standard reduction $\twoheadrightarrow_S$ from a weakly standard one $\twoheadrightarrow_{S^{\mathrm{w}}}$.

Definition 2.1.2.2. We define $\rightarrow_{+^\lambda}$ by

$$\lambda x.(M +_l N) \rightarrow_{+^\lambda} \lambda x.M +_l \lambda x.N$$

extended to context.

Proposition 2.1.2.4. *If M is weakly canonical and $M \rightarrow_{+^\lambda} \cdot \rightarrow_{\mathrm{can}^{\mathrm{w}}} \cdot \rightarrow_{\beta_{/\!/}} N$ then there is N' such that $N \rightarrow_{\beta_{/\!/}} N'$ and $M \rightarrow_{\beta_{/\!/}} \cdot \twoheadrightarrow_{+} N'$.*

$$\begin{tikzcd}
M \arrow[r, "+^\lambda"] \arrow[dd, dashed, "\beta_{/\!/}"'] & {} \arrow[r, "\mathrm{can}^{\mathrm{w}}"] & {} \arrow[d, "\beta_{/\!/}"] \\
 & & N \arrow[d, dashed, "\beta_{/\!/}"] \\
{} \arrow[rr, dashed, two heads, "+"'] & & N'
\end{tikzcd}$$

Proof. By induction on M as a weakly canonical form.

- If $M = M_1 +_l M_2$ then $N = N_1 +_l N_2$ with $M_i(\rightarrow_{+^\lambda} \cdot \rightarrow_{\mathrm{can}^{\mathrm{w}}})^? \cdot \rightarrow_{\beta_{/\!/}} N_i$ for $i \in \{1; 2\}$ so we conclude by induction hypothesis.
- If $M = y\ P_1\ ...\ P_m$ then $N = y\ Q_1\ ...\ Q_m$ with $P_i(\rightarrow_{+^\lambda} \cdot \rightarrow_{\mathrm{can}^{\mathrm{w}}})^? \cdot \rightarrow_{\beta_{/\!/}} Q_i$ for $i \leq m$ so we conclude by induction hypothesis.
- If
$$M = (\lambda y.M_0)\ P_1\ ...\ P_m \rightarrow_{+^\lambda} \cdot \rightarrow_{\mathrm{can}^{\mathrm{w}}} (\lambda y.M_0')\ P_1'\ ...\ P_m'$$
with $M_0(\rightarrow_{+^\lambda} \cdot \rightarrow_{\mathrm{can}^{\mathrm{w}}})^? M_0'$ and $P_i(\rightarrow_{+^\lambda} \cdot \rightarrow_{\mathrm{can}^{\mathrm{w}}})^? P_i'$ for $i \leq m$ then either $N = (\lambda y.N_0)\ Q_1\ ...\ Q_m$ or $N = N_0 \left[Q_1/y\right]\ Q_2\ ...\ Q_m$, with in both cases $M_0' \rightarrow_{\beta_{/\!/}} N_0$ and $P_i' \rightarrow_{\beta_{/\!/}} Q_i$ for $i \leq m$. Then by induction hypothesis we have a term N_0' such that $N_0 \rightarrow_{\beta_{/\!/}} N_0'$ and $M_0 \rightarrow_{\beta_{/\!/}} \cdot \twoheadrightarrow_{+} N_0'$, and terms Q_i' such that $Q_i \rightarrow_{\beta_{/\!/}} Q_i'$ and $P_i \rightarrow_{\beta_{/\!/}} \cdot \twoheadrightarrow_{+} Q_i'$ for $i \leq m$. We choose either $N' = (\lambda y.N_0')\ Q_1'\ ...\ Q_m'$ or $N' = N_0' \left[Q_1'/y\right]\ Q_2'\ ...\ Q_m'$.
- If
$$\begin{aligned} M = (\lambda y.(M_1 +_l M_2))\ P_1\ ...\ P_m &\rightarrow_{+^\lambda} (\lambda y.M_1 +_l \lambda y.M_2)\ P_1\ ...\ P_m \\ &\rightarrow_{\mathrm{can}^{\mathrm{w}}} (\lambda y.M_1)\ P_1 ... P_m +_l (\lambda y.M_2)\ P_1 ... P_m \end{aligned}$$
then
  - either $N = (\lambda y.N_1)\ Q_1^1\ ...\ Q_m^1 +_l (\lambda y.N_2)\ Q_1^2\ ...\ Q_m^2$
  - or $N = N_1 \left[Q_1^1/y\right]\ Q_2^1\ ...\ Q_m^1 +_l (\lambda y.N_2)\ Q_1^2\ ...\ Q_m^2$
  - or $N = (\lambda y.N_1)\ Q_1^1\ ...\ Q_m^1 +_l N_2 \left[Q_1^2/y\right]\ Q_2^2\ ...\ Q_m^2$
  - or $N = N_1 \left[Q_1^1/y\right]\ Q_2^1\ ...\ Q_m^1 +_l N_2 \left[Q_1^2/y\right]\ Q_2^2\ ...\ Q_m^2$

  with in each case $M_i \rightarrow_{\beta_{/\!/}} N_i$ for $i \in \{1; 2\}$ and $P_i \rightarrow_{\beta_{/\!/}} Q_i^j$ for $i \leq m$ and $i \in \{1; 2\}$. Then there are terms Q_i' for $i \leq m$ such that $P_i \rightarrow_{\beta_{/\!/}} Q_i'$ and $Q_i^j \rightarrow_{\beta_{/\!/}} Q_i'$ for $j \in \{1; 2\}$. We choose

– either $N' = (\lambda y.N_1)\ Q'_1\ \dots\ Q'_m +_l (\lambda y.N_2)\ Q'_1\ \dots\ Q'_m$ in the first case

– or $N' = N_1 \left[Q'_1/y\right]\ Q'_2\ \dots\ Q'_m +_l N_2 \left[Q'_1/y\right]\ Q'_2\ \dots\ Q'_m$.

□

Proposition 2.1.2.5. *If M is weakly canonical and $M \to_{\mathrm{can}} \cdot \to_{\beta^c_{/\!/}} N$ then there is N' such that $N \twoheadrightarrow_{\beta^c_{/\!/}} N'$ and $M \to_{\beta_{/\!/}} \cdot \to_{\mathrm{can}} N'$. This implies $M \to_{\beta^{\mathrm{w}}_{/\!/}} \cdot \to_{\mathrm{can}} N'$.*

Proof. If M is weakly canonical then the reduction $M \to_{\mathrm{can}} \mathrm{can}(M)$ can be decomposed in the form $M\ (\to_{+^\lambda} \cdot \to_{\mathrm{can}^{\mathrm{w}}})^*\ \mathrm{can}(M)$. We reason by induction on the number of reductions $\to_{+^\lambda} \cdot \to_{\mathrm{can}^{\mathrm{w}}}$. If there is none then M is canonical and the result is immediate.

Otherwise if $M \to_{+^\lambda} \cdot \to_{\mathrm{can}^{\mathrm{w}}} P\ (\to_{+^\lambda} \cdot \to_{\mathrm{can}^{\mathrm{w}}})^*\ \mathrm{can}(M) \to_{\beta^c_{/\!/}} N$ then by induction hypothesis there is N' such that $N \twoheadrightarrow_{\beta^c_{/\!/}} N'$ and $P \to_{\beta_{/\!/}} Q \to_{\mathrm{can}} N'$. According to the previous result there is Q' such that $Q \to_{\beta_{/\!/}} Q'$ and $M \to_{\beta_{/\!/}} \cdot \twoheadrightarrow_{+} Q'$. But then from $Q \to_{\beta_{/\!/}} Q'$ we can deduce $\mathrm{can}(Q) \twoheadrightarrow_{\beta^c_{/\!/}} \mathrm{can}(Q')$, and we have $\mathrm{can}(Q) = N'$. Thus we have $N \twoheadrightarrow_{\beta^c_{/\!/}} N' \twoheadrightarrow_{\beta^c_{/\!/}} \mathrm{can}(Q')$ and $M \to_{\beta_{/\!/}} \cdot \to_{\mathrm{can}} \mathrm{can}(Q')$.

$$\begin{tikzcd}
M \arrow[r, "+^\lambda \cdot \mathrm{can}^{\mathrm{w}}"] \arrow[ddd, dashed, "\beta_{/\!/}"'] \arrow[dddr, phantom, "(2.1.2.4)"] & P \arrow[r, two heads, "+^\lambda \cdot \mathrm{can}^{\mathrm{w}}"] \arrow[dd, dashed, "\beta_{/\!/}"'] \arrow[ddr, phantom, "(\mathrm{IH})"] & \mathrm{can}(M) \arrow[d, "\beta^c_{/\!/}"] \\
 & & N \arrow[d, dashed, two heads, "\beta^c_{/\!/}"] \\
 & Q \arrow[r, dashed, "\mathrm{can}"'] \arrow[d, dashed, "\beta_{/\!/}"] & N' \arrow[d, dashed, two heads, "\beta^c_{/\!/}"] \\
{} \arrow[r, dashed, two heads, "+"'] & Q' \arrow[r, dashed, "\mathrm{can}"'] & \mathrm{can}(Q')
\end{tikzcd}$$

□

Proposition 2.1.2.6. *If M is weakly canonical and $M \to_{\mathrm{can}} \cdot \twoheadrightarrow_{\beta^c_{/\!/}} N$ then there is N' such that $N \twoheadrightarrow_{\beta^c_{/\!/}} N'$ and $M \twoheadrightarrow_{\beta^{\mathrm{w}}_{/\!/}} \cdot \to_{\mathrm{can}} N'$.*

Proof. By induction on the length of the reduction $\twoheadrightarrow_{\beta^c_{/\!/}}$. If $M \to_{\mathrm{can}} N$ the result is immediate.

Otherwise $\mathrm{can}(M) \to_{\beta^c_{/\!/}} P \to^n_{\beta^c_{/\!/}} N$ and according to the previous result there are M' and P' such that $P \twoheadrightarrow_{\beta^c_{/\!/}} P'$ and $M \to_{\beta^{\mathrm{w}}_{/\!/}} M' \to_{\mathrm{can}} P'$. Then using the diamond property of $\to_{\beta^c_{/\!/}}$ there exists Q such that $N \twoheadrightarrow_{\beta^c_{/\!/}} Q$ and $P' \to^n_{\beta^c_{/\!/}} Q$. By induction hypothesis on the reduction $M' \to_{\mathrm{can}} P' \to^n_{\beta^c_{/\!/}} Q$ there is N' such that $Q \twoheadrightarrow_{\beta^c_{/\!/}} N'$

and $M' \twoheadrightarrow_{\beta^{\mathrm{w}}_{/\!/}} \cdot \rightarrow_{\mathrm{can}} N'$.

$$
\begin{tikzcd}
M \arrow[r, "\mathrm{can}"] \arrow[d, dashed, "\beta^{\mathrm{w}}_{/\!/}"'] \arrow[dr, phantom, "(2.1.2.5)"] & \mathrm{can}(M) \arrow[r, "\beta^c_{/\!/}"] & P \arrow[r, "\beta^c_{/\!/}", "n" near end] \arrow[d, dashed, two heads, "\beta^c_{/\!/}"] & N \arrow[d, dashed, two heads, "\beta^c_{/\!/}"] \\
M' \arrow[rr, dashed, "\mathrm{can}"'] \arrow[d, dashed, two heads, "\beta^{\mathrm{w}}_{/\!/}"'] \arrow[drr, phantom, "\text{(IH)}"] & & P' \arrow[r, dashed, "\beta^c_{/\!/}"', "n" near end] & Q \arrow[d, dashed, two heads, "\beta^c_{/\!/}"] \\
{} \arrow[rrr, dashed, "\mathrm{can}"'] & & & N'
\end{tikzcd}
$$

□

Corollary 2.1.2.7. *If $M\ (\rightarrow_\beta \cup \rightarrow_+)^*\ N$ then there is a term N' such that $N\ (\rightarrow_\beta \cup \rightarrow_+)^*\ N'$ and $M\ (\rightarrow_\beta \cup \rightarrow_{+^{\mathrm{w}}})^* \cdot \twoheadrightarrow_+ N'$.*

Proof. We have $\mathrm{can}^{\mathrm{w}}(M) \rightarrow_{\mathrm{can}} \mathrm{can}(M) \twoheadrightarrow_{\beta^c} \mathrm{can}(N)$ so according to the previous result there is N' such that $\mathrm{can}(N) \twoheadrightarrow_{\beta^c} N'$ and $\mathrm{can}^{\mathrm{w}}(M) \twoheadrightarrow_{\beta^{\mathrm{w}}} \cdot \rightarrow_{\mathrm{can}} N'$. □

Corollary 2.1.2.8. *If $M\ (\rightarrow_\beta \cup \rightarrow_+)^*\ N$ then there is a term N' such that $N\ (\rightarrow_\beta \cup \rightarrow_+)^*\ N'$ and $\mathrm{can}^{\mathrm{w}}(M) \twoheadrightarrow_{S^{\mathrm{w}}} \cdot \twoheadrightarrow_+ N'$.*

The only difference between $\rightarrow_\beta \cup \rightarrow_+$ and $\rightarrow_\beta \cup \rightarrow_{+^{\mathrm{w}}}$ is that allowing the commutation of sums with abstractions make it possible to duplicate more β-redexes with the reductions $(\lambda x.(M_1 +_p M_2))\ N \twoheadrightarrow_+ (\lambda x.M_1)\ N +_p (\lambda x.M_2)\ N$, and then to reduce only some of them. But in the end it does not change the calculus much. We can even deduce from the previous proofs that if we have a reduction $M \twoheadrightarrow_{\beta^c_{/\!/}} N$ of length n then there is N' such that $N \twoheadrightarrow_{\beta^c_{/\!/}} \mathrm{can}(N')$ and there is a reduction $M \twoheadrightarrow_{\beta^{\mathrm{w}}_{/\!/}} N'$ of the same length n.

Now we can use this result to prove that every reduction $\rightarrow_\beta \cup \rightarrow_+$ can be extended into a standard one.

Proposition 2.1.2.9. *The rules of $\twoheadrightarrow_S$ can be extended to non canonical terms:*

- *if* $\mathrm{can}(M_i) \twoheadrightarrow_S \mathrm{can}(N_i)$ *for* $i \in \{1; 2\}$ *then*

$$\mathrm{can}(M_1 +_l M_2) \twoheadrightarrow_S \mathrm{can}(N_1 +_l N_2);$$

- *if* $\mathrm{can}\left(\lambda x_1 ... x_n.M\left[{}^{P}\!/_{y}\right]\ Q_1\ ...\ Q_m\right) \twoheadrightarrow_S \mathrm{can}(N)$ *then*

$$\mathrm{can}(\lambda x_1 ... x_n.(\lambda y.M)\ P\ Q_1\ ...\ Q_m) \twoheadrightarrow_S \mathrm{can}(N);$$

- *if* $\mathrm{can}(M) \twoheadrightarrow_S \mathrm{can}(N)$ *and* $\mathrm{can}(P_i) \twoheadrightarrow_S \mathrm{can}(Q_i)$ *for* $i \leq m$ *then*

$$\mathrm{can}(\lambda x_1 ... x_n.(\lambda y.M)\ P_1\ ...\ P_m) \twoheadrightarrow_S \mathrm{can}\left(\lambda x_1 ... x_n.(\lambda y.N)\ Q_1\ ...\ Q_m\right);$$

- *if* $\mathrm{can}(P_i) \twoheadrightarrow_S \mathrm{can}(Q_i)$ *for* $i \leq m$ *then*

$$\mathrm{can}(\lambda x_1 ... x_n.y\ P_1\ ...\ P_m) \twoheadrightarrow_S \mathrm{can}(\lambda x_1 ... x_n.y\ Q_1\ ...\ Q_m).$$

Proof. The first and last case are immediate. We have

$$\begin{aligned}
\mathrm{can}(M_1 +_l M_2) &= \mathrm{can}(M_1) +_l \mathrm{can}(M_2)\\
&\twoheadrightarrow_S \mathrm{can}(N_1) +_l \mathrm{can}(N_2)\\
&= \mathrm{can}(N_1 +_l N_2)\\
\mathrm{can}(\lambda x_1 ... x_n.y\ R_1\ ...\ R_m) &= \lambda x_1 ... x_n.y\ \mathrm{can}(R_1)\ ...\ \mathrm{can}(R_m)\\
&\twoheadrightarrow_S \lambda x_1 ... x_n.y\ \mathrm{can}(Q_1)\ ...\ \mathrm{can}(Q_m)\\
&= \mathrm{can}(\lambda x_1 ... x_n.y\ Q_1\ ...\ Q_m).
\end{aligned}$$

In the second case observe that we have

$$\begin{aligned}
&\mathrm{can}\left(\lambda x_1 ... x_n.M\left[{}^{P}/_{y}\right]\ Q_1\ ...\ Q_m\right)\\
&\quad = \mathrm{can}\left(\lambda x_1 ... x_n.\mathrm{can}(M)\left[{}^{\mathrm{can}(P)}/_{y}\right]\ \mathrm{can}(Q_1)\ ...\ \mathrm{can}(Q_m)\right)\\
&\mathrm{can}(\lambda x_1 ... x_n.(\lambda y.M)\ P\ Q_1\ ...\ Q_m)\\
&\quad = \mathrm{can}(\lambda x_1 ... x_n.(\lambda y.\mathrm{can}(M))\ \mathrm{can}(P)\ \mathrm{can}(Q_1)\ ...\ \mathrm{can}(Q_m))
\end{aligned}$$

so we can assume w.l.o.g. that M, P, Q_i for $i \leq m$ and N are canonical. Then we reason by induction on M:

- if $M = v$ is a value then

$$\mathrm{can}(\lambda x_1 ... x_n.(\lambda y.v)\ P\ Q_1\ ...\ Q_m) = \lambda x_1 ... x_n.(\lambda y.v)\ P\ Q_1\ ...\ Q_m$$

  and the result is exactly a rule of $\twoheadrightarrow_S$;

- if $M = M_1 +_l M_2$ then

$$\begin{aligned}
&\mathrm{can}(\lambda x_1 ... x_n.(\lambda y.M)\ P\ Q_1\ ...\ Q_m)\\
&\quad = \mathrm{can}(\lambda x_1 ... x_n.(\lambda y.M_1)\ P\ Q_1 ... Q_m) +_l \mathrm{can}(\lambda x_1 ... x_n.(\lambda y.M_2)\ P\ Q_1 ... Q_m)\\
&\mathrm{can}\left(\lambda x_1 ... x_n.M\left[{}^{P}/_{y}\right]\ Q_1\ ...\ Q_m\right)\\
&\quad = \mathrm{can}\left(\lambda x_1 ... x_n.M_1\left[{}^{P}/_{y}\right]\ Q_1 ... Q_m\right) +_l \mathrm{can}\left(\lambda x_1 ... x_n.M_2\left[{}^{P}/_{y}\right]\ Q_1 ... Q_m\right)
\end{aligned}$$

  and necessarily $N = N_1 +_l N_2$ with

$$\mathrm{can}\left(\lambda x_1 ... x_n.M_i\left[{}^{P}/_{y}\right]\ Q_1\ ...\ Q_m\right) \twoheadrightarrow_S N_i$$

  for $i \in \{1; 2\}$ so by induction hypothesis

$$\mathrm{can}(\lambda x_1 ... x_n.(\lambda y.M_1)\ P\ Q_1\ ...\ Q_m) \twoheadrightarrow_S N_i.$$

The third case is similar. Again we can assume that M, N and P_i and Q_i for $i \leq m$ are canonical, and we weason by induction on M. If M is a value then we have exactly a rule of $\twoheadrightarrow_S$, and if $M = M_1 +_l M_2$ then $N = N_1 +_l N_2$ with $M_i \twoheadrightarrow_S N_i$ for $i \in \{1; 2\}$ and the result follows by induction hypothesis. $\square$

Proposition 2.1.2.10. *If M is weakly canonical and $M \twoheadrightarrow_{S^{\mathrm{w}}} N$ then* $\mathrm{can}(M) \twoheadrightarrow_S \mathrm{can}(N)$.

Proof. By induction on $\twoheadrightarrow_{S^{\mathrm{w}}}$.

- If $\lambda x.M \twoheadrightarrow_{S^{\mathrm{w}}} \lambda x.N$ with $M \twoheadrightarrow_{S^{\mathrm{w}}} N$ then by induction hypothesis we have $\mathrm{can}(M) \twoheadrightarrow_S \mathrm{can}(N)$. We prove the result by induction on this reduction:
  - if
    $$\mathrm{can}(M) = M_1 +_l M_2 \twoheadrightarrow_S N_1 +_l N_2 = \mathrm{can}(N)$$
    then $\mathrm{can}(\lambda x.M) = \mathrm{can}(\lambda x.M_1) +_l \mathrm{can}(\lambda x.M_2)$ and by induction hypothesis we have $\mathrm{can}(\lambda x.M_i) \twoheadrightarrow_S \mathrm{can}(\lambda x.N_i)$ for $i \in \{1; 2\}$, hence $\mathrm{can}(\lambda x.M) \twoheadrightarrow_S \mathrm{can}(\lambda x.N)$;
  - if
    $$\mathrm{can}(M) = \lambda x_1 ... x_n.(\lambda y.v)\ P\ Q_1\ ...\ Q_m \twoheadrightarrow_S \mathrm{can}(N)$$
    with $\mathrm{can}\left(\lambda x_1 ... x_n.v\left[{}^{P}/{}_{y}\right]\ Q_1\ ...\ Q_m\right) \twoheadrightarrow_S \mathrm{can}(N)$ then by induction hypothesis $\mathrm{can}\left(\lambda x.\lambda x_1 ... x_n.v\left[{}^{P}/{}_{y}\right]\ Q_1\ ...\ Q_m\right) \twoheadrightarrow_S \mathrm{can}(\lambda x.N)$, hence
    $$\mathrm{can}(\lambda x.M) = \lambda x.\lambda x_1 ... x_n.(\lambda y.v)\ P\ Q_1\ ...\ Q_m \twoheadrightarrow_S \mathrm{can}(\lambda x.N);$$
  - if
    $$\mathrm{can}(M) = \lambda x_1 ... x_n.(\lambda y.v)\ P_1\ ...\ P_m \twoheadrightarrow_S \mathrm{can}\left(\lambda x_1 ... x_n.(\lambda y.N_0)\ Q_1\ ...\ Q_m\right)$$
    with $v \twoheadrightarrow_S N_0$ and $P_i \twoheadrightarrow_S Q_i$ for $i \leq m$ then
    $$\begin{aligned}\mathrm{can}(\lambda x.M) &= \lambda x.\lambda x_1 ... x_n.(\lambda y.v)\ P_1\ ...\ P_m \\ &\twoheadrightarrow_S \mathrm{can}\left(\lambda x.\lambda x_1 ... x_n.(\lambda y.N_0)\ Q_1\ ...\ Q_m\right) = \mathrm{can}(\lambda x.N);\end{aligned}$$
  - if
    $$\mathrm{can}(M) = \lambda x_1 ... x_n.y\ P_1\ ...\ P_m \twoheadrightarrow_S \lambda x_1 ... x_n.y\ Q_1\ ...\ Q_m = \mathrm{can}(N)$$
    with $P_i \twoheadrightarrow_S Q_i$ for $i \leq m$ then
    $$\begin{aligned}\mathrm{can}(\lambda x.M) &= \lambda x.\lambda x_1 ... x_n.y\ P_1\ ...\ P_m \\ &\twoheadrightarrow_S \lambda x.\lambda x_1 ... x_n.y\ Q_1\ ...\ Q_m = \mathrm{can}(\lambda x.N).\end{aligned}$$

The other cases are consequences of the previous result.

- If
$$M_1 +_l M_2 \twoheadrightarrow_{S^{\mathrm{w}}} N_1 +_l N_2$$
with $M_i \twoheadrightarrow_{S^{\mathrm{w}}} N_i$ for $i \in \{1;2\}$ then by induction hypothesis we have $\mathrm{can}(M_i) \twoheadrightarrow_S \mathrm{can}(N_i)$ for $i \in \{1;2\}$ so
$$\mathrm{can}(M_1 +_l M_2) \twoheadrightarrow_S \mathrm{can}(N_1 +_l N_2).$$

- If
$$(\lambda y.M)\ P\ Q_1\ ...\ Q_m \twoheadrightarrow_{S^{\mathrm{w}}} N$$
with $\mathrm{can}^{\mathrm{w}}\left(M\left[{}^{P}/_{y}\right]\ Q_1\ ...\ Q_m\right) \twoheadrightarrow_{S^{\mathrm{w}}} N$ then by induction hypothesis $\mathrm{can}\left(M\left[{}^{P}/_{y}\right]\ Q_1\ ...\ Q_m\right) \twoheadrightarrow_S \mathrm{can}(N)$ so
$$\mathrm{can}((\lambda y.M)\ P\ Q_1\ ...\ Q_m) \twoheadrightarrow_S \mathrm{can}(N).$$

- If
$$(\lambda y.M)\ P_1\ ...\ P_m \twoheadrightarrow_{S^{\mathrm{w}}} (\lambda y.N)\ Q_1\ ...\ Q_m$$
with $M \twoheadrightarrow_{S^{\mathrm{w}}} N$ and $P_i \twoheadrightarrow_{S^{\mathrm{w}}} Q_i$ for $i \leq m$ then by induction hypothesis $\mathrm{can}(M) \twoheadrightarrow_S \mathrm{can}(N)$ and $\mathrm{can}(P_i) \twoheadrightarrow_S \mathrm{can}(Q_i)$ for $i \leq m$ so
$$\mathrm{can}((\lambda y.M)\ P_1\ ...\ P_m) \twoheadrightarrow_S \mathrm{can}((\lambda y.N)\ Q_1\ ...\ Q_m).$$

- If
$$y\ P_1\ ...\ P_m \twoheadrightarrow_{S^{\mathrm{w}}} y\ Q_1\ ...\ Q_m$$
with $P_i \twoheadrightarrow_{S^{\mathrm{w}}} Q_i$ for $i \leq m$ then by induction hypothesis $\mathrm{can}(P_i) \twoheadrightarrow_S \mathrm{can}(Q_i)$ for $i \leq m$ so
$$\mathrm{can}(y\ P_1\ ...\ P_m) \twoheadrightarrow_S \mathrm{can}(y\ S_1\ ...\ S_m).$$

□

Every reduction $\rightarrow_\beta \cup \rightarrow_+$ can be extended into a weak reduction, every weak reduction is weakly standardizable and every weak standard reduction lifts to a standard one, so we have our general standardization theorem.

Theorem 2.1.2.11. *If $M \twoheadrightarrow_{\beta/=_+} N$ then there is N' such that $N \twoheadrightarrow_{\beta/=_+} N'$ and $\mathrm{can}(M) \twoheadrightarrow_S N'$.*

Proof. If $M \twoheadrightarrow_{\beta/=_+} N$ then $\mathrm{can}(M) \twoheadrightarrow_{\beta^{\mathrm{c}}} \mathrm{can}(N)$, hence there is N' such that $\mathrm{can}(N)\ (\rightarrow_\beta \cup \rightarrow_+)^*\ N'$ and $\mathrm{can}(M) \twoheadrightarrow_{S^{\mathrm{w}}} N'$, which gives $\mathrm{can}(M) \twoheadrightarrow_S \mathrm{can}(N')$. □

Corollary 2.1.2.12. *If $M \twoheadrightarrow_{\beta/=_+} N$ where N is a canonical β-normal form then $\mathrm{can}(M) \twoheadrightarrow_S N$.*

2.2 Simplification of the reductions

The standardization is a very useful result to study the reduction. In our case it comes with this requirement to extend the reduction. This is not much of a problem, as we are mostly interested in reductions which end on some sort of normal form.

But it has its drawbacks. For instance if we have $M \twoheadrightarrow_{\beta/=_+} P \twoheadleftarrow_{\beta/=_+} N$ then we can find P' such that $P \twoheadrightarrow_{\beta/=_+} P'$ and $M \twoheadrightarrow_S P' \twoheadleftarrow_{\beta/=_+} N$; then we can also find P'' such that $P' \twoheadrightarrow_{\beta/=_+} P''$ and $N \twoheadrightarrow_S P''$, but there is not guarantee that we still have $M \twoheadrightarrow_S P''$. We actually do not know if we necessarily have $M \twoheadrightarrow_S \cdot \twoheadleftarrow_S N$ or not.

Still this idea of extending a reduction to simplify it is quite convenient, and yields some interesting results.

2.2.1 Non splitting reductions

The first simplification we want to deal with is the case of the splitting of terms:

$$M \to_{\text{split}} M +_l M \text{ if } l \in \mathcal{I}.$$

The opposite reduction $M +_l M \to_{\mathcal{I}} M$ is natural and necessary if we want to have the right semantics for our calculus. And if we want to look at this calculus from an equational point of view then we also have the splitting. Yet this splitting is not natural at all from a computational point of view. We even proved that every reduction modulo $\equiv_{\text{syn}}$ can be turned into a reduction $\twoheadrightarrow_{\beta/=_+}$ if we consider the result modulo $\equiv_{\text{syn}}$, with the exception that we have to perform the splitting first: we have $\twoheadrightarrow_{\beta/\equiv} = \twoheadrightarrow_{\text{split}} \cdot \twoheadrightarrow_{\beta/=_+} \cdot \equiv_{\text{syn}}$.

This annoying behaviour of the splitting can be fixed if we allow the extension of the reduction: every reduction can be extended into a reduction without splitting.

Proposition 2.2.1.1. *If $M \twoheadrightarrow_{\text{split}} \cdot \twoheadrightarrow_S N$ then there is N' such that $N \twoheadrightarrow_{\beta/\equiv} N'$ and $M \twoheadrightarrow_S N'$.*

Proof. Let $M \twoheadrightarrow_{\text{split}} P \twoheadrightarrow_S N$ with P canonical by definition of $\twoheadrightarrow_S$. We reason by induction on $\twoheadrightarrow_S$, using the inductive characterization of $\twoheadrightarrow_{\mathcal{I}}$ given by 1.2.2.1.

If $P = P_1 +_l P_2 \twoheadrightarrow_S N_1 +_l N_2 = N$ with $P_i \twoheadrightarrow_S N_i$ for $i \in \{1; 2\}$, there are two cases for the reduction $M \twoheadrightarrow_{\text{split}} P$.

- If $M \twoheadrightarrow_{\text{split}} P_i$ for $i \in \{1; 2\}$ then by induction hypothesis there are N_1' and N_2' such that $N_i \twoheadrightarrow_{\beta/\equiv} N_i'$ and $M \twoheadrightarrow_S N_i'$ for $i \in \{1; 2\}$. But then we also have $M \twoheadrightarrow_{\beta/=_+} N_i'$ for $i \in \{1; 2\}$, so by confluence of $\twoheadrightarrow_{\beta/=_+}$ there is Q such that $N_i' \twoheadrightarrow_{\beta/=_+} Q$ for $i \in \{1; 2\}$, and there is Q' such that $Q \twoheadrightarrow_{\beta/=_+} Q'$ and $M \twoheadrightarrow_S Q'$. We conclude by observing that $N \twoheadrightarrow_{\beta/=_+} Q' +_l Q' \to_{\mathcal{I}} Q'$.

- If $M = M_1 +_l M_2 \twoheadrightarrow_{\text{split}} P_1 +_l P_2$ with $M_i \twoheadrightarrow_{\text{split}} P_i$ for $i \in \{1;2\}$ then by induction hypothesis we have terms N_i' such that $N_i \twoheadrightarrow_{\beta/\equiv} N_i'$ and $M_i \twoheadrightarrow_S N_i'$ for $i \in \{1;2\}$.

Otherwise P is a value and M has the same shape as P.

- If $P = \lambda x_1...x_n.y\ R_1\ ...\ R_m \twoheadrightarrow_S \lambda x_1...x_n.y\ S_1\ ...\ S_m = N$ with $R_i \twoheadrightarrow_S S_i$ for $i \leq m$ then $M = \lambda x_1...x_n.y\ Q_1\ ...\ Q_m$ with $Q_i \twoheadrightarrow_{\text{split}} R_i$ for $i \leq m$ so the result is immediate by induction hypothesis.

- If $P = \lambda x_1...x_n.(\lambda y.R_0)\ R_1\ ...\ R_m \twoheadrightarrow_S \lambda x_1...x_n.(\lambda y.S_0)\ S_1\ ...\ S_m = N$ with $R_i \twoheadrightarrow_S S_i$ for $i \leq m$ then $M = \lambda x_1...x_n.(\lambda y.Q_0)\ Q_1\ ...\ Q_m$ with $Q_i \twoheadrightarrow_{\text{split}} R_i$ for $i \leq m$, and we conclude by induction hypothesis.

- If $P = \lambda x_1...x_n.(\lambda y.R_0)\ R_1\ ...\ R_m$ with $\lambda x_1...x_n.R_0\left[R_1/y\right]\ R_2\ ...\ R_m \twoheadrightarrow_S N$ then $M = \lambda x_1...x_n.(\lambda y.Q_0)\ Q_1\ ...\ Q_m$ with $Q_i \twoheadrightarrow_{\text{split}} R_i$ for $i \leq m$. We have $\lambda x_1...x_n.Q_0\left[Q_1/y\right]\ Q_2\ ...\ Q_m \twoheadrightarrow_{\text{split}} \lambda x_1...x_n.R_0\left[R_1/y\right]\ R_2\ ...\ R_m$ so by induction hypothesis there is N' such that $N \twoheadrightarrow_{\beta/\equiv} N'$ and

$$\lambda x_1...x_n.Q_0\left[Q_1/y\right]\ Q_2\ ...\ Q_m \twoheadrightarrow_S N'$$

thus $M \twoheadrightarrow_S N'$.

□

Remark that we extend the reduction for two reasons: to standardize and to use the confluence. This means that even with a strong standardization theorem such as for the weak reduction of sums, we do not have $\twoheadrightarrow_{\text{split}} \cdot \twoheadrightarrow_{S^{\text{w}}} = \twoheadrightarrow_{S^{\text{w}}}$.

Theorem 2.2.1.2. *If $M \twoheadrightarrow_{\beta/\equiv} N$ then there is N' such that $N \twoheadrightarrow_{\beta/\equiv} N'$ and $M \twoheadrightarrow_S N'$.*

Proof. We know that if $M \twoheadrightarrow_{\beta/\equiv} N$ then $\text{can}(M) \twoheadrightarrow_{\text{split}} P \twoheadrightarrow_{\beta/=_+} Q \equiv_{\text{syn}} \text{can}(N)$ with P and Q canonical. Then there is Q' such that $Q \twoheadrightarrow_{\beta/=_+} Q'$ and $P \twoheadrightarrow_S Q'$, and according to the previous result there is a term N' such that $Q' \twoheadrightarrow_{\beta/\equiv} N'$ and $M \twoheadrightarrow_S N'$. Besides we can observe that $\text{can}(N) \equiv_{\text{syn}} Q \twoheadrightarrow_{\beta/\equiv} N'$ so we have $N \twoheadrightarrow_{\beta/\equiv} N'$. □

Corollary 2.2.1.3. *If $M \twoheadrightarrow_{\beta/\equiv} N$ and N is a canonical β-normal form then $M \twoheadrightarrow_S \cdot \equiv_{\text{syn}} N$.*

Proof. If N is canonical and β-normal and $N \twoheadrightarrow_{\beta/\equiv} N'$ with N' canonical then $N \equiv_{\text{syn}} N'$. □

2.2.2 Complete reductions

Another question we discussed in the introduction of the first chapter is that of a complete head reduction. A natural definition of a canonical head reduction would be the following.

Definition 2.2.2.1. The *canonical head reduction* $\to_{h^c}$ is defined between canonical terms by:

$$\frac{M_1 \to_{h^c} N_1}{M_1 +_l M_2 \to_{h^c} N_1 +_l M_2} \qquad \frac{M_2 \to_{h^c} N_2}{M_1 +_l M_2 \to_{h^c} M_1 +_l N_2}$$

$$\frac{}{\lambda x_1 ... x_n.(\lambda y.M)\ P\ Q_1\ ...\ Q_m \to_{h^c} \mathrm{can}\left(\lambda x_1 ... x_n.M\left[{}^{P}/_{y}\right]\ Q_1\ ...\ Q_m\right)}$$

Unlike in the deterministic case, a term with sums may have several head redexes. We may have many different head reduction paths which do not necessarily yield the same result. This is quite a loss compared to the usual case, but we can recover a more satisfying notion of head reduction if we always reduce every head redex in parallel.

Definition 2.2.2.2. The *complete head reduction* $\to_H$ is defined between canonical terms by:

$$\frac{M_1 \to_H N_1 \qquad M_2 \to_H N_2}{M_1 +_l M_2 \to_H N_1 +_l N_2} \qquad \frac{}{\lambda x_1 ... x_n.y\ P_1\ ...\ P_m \to_H \lambda x_1 ... x_n.y\ P_1\ ...\ P_m}$$

$$\frac{}{\lambda x_1 ... x_n.(\lambda y.M)\ P\ Q_1\ ...\ Q_m \to_H \mathrm{can}\left(\lambda x_1 ... x_n.M\left[{}^{P}/_{y}\right]\ Q_1\ ...\ Q_m\right)}$$

This reduction has many interesting properties. To begin with we recover the uniqueness of the reduction, and every head reduction can be completed.

Proposition 2.2.2.1. *For any canonical term M there is a unique term, which we note $\mathrm{H}(M)$, such that $M \to_H \mathrm{H}(M)$.*

Proof. By induction on M.

- If $M = M_1 +_l M_2$ then $M \to_H N$ if and only if $N = N_1 +_l N_2$ with $M_i \to_H N_i$ for $i \in \{1; 2\}$, and by induction hypothesis such terms exist and are unique.
- If M is a value then we have either $M = \lambda x_1 ... x_n.y\ P_1\ ...\ P_m$ or $M = \lambda x_1 ... x_n.(\lambda y.M_0)\ P\ Q_1\ ...\ Q_m$ and there is a unique reduction from M.

□

Proposition 2.2.2.2. *If $M \twoheadrightarrow_H N$ then $M \twoheadrightarrow_{h^c} N$.*

Proof. We prove that if $M \to_H N$ then $M \twoheadrightarrow_{h^c} N$, and the result will follow by induction on the length of the reduction. We reason by induction on $\to_H$.

- If $M = M_1 +_l M_2 \rightarrow_H N_1 +_l N_2 = N$ with $M_i \rightarrow_H N_i$ for all $i \in \{1;2\}$ then by induction hypothesis $M_i \twoheadrightarrow_{h^c} N_i$ for $i \in \{1;2\}$, hence $M_1 +_l M_2 \twoheadrightarrow_{h^c} N_1 +_l M_2 \twoheadrightarrow_{h^c} N_1 +_l N_2$.

- If $M = \lambda x_1...x_n.y\ P_1\ ...\ P_m \rightarrow_H \lambda x_1...x_n.y\ P_1\ ...\ P_m = N$ then $M = N$ and $M \twoheadrightarrow_{h^c} N$.

- If $M = \lambda x_1...x_n.(\lambda y.M_0)\ P\ Q_1\ ...\ Q_m \rightarrow_H \text{can}\left(\lambda x_1...x_n.M_0 \left[{}^{P}/_{y}\right]\ Q_1\ ...\ Q_m\right)$ then $M \rightarrow_{h^c} N$.

□

Proposition 2.2.2.3. *If $M \twoheadrightarrow_{h^c} N$ then there is N' such that $N \twoheadrightarrow_{h^c} N'$ and $M \twoheadrightarrow_H N'$.*

Proof. First we prove that if $M \rightarrow_{h^c} N$ then $N \twoheadrightarrow_{h^c} \mathrm{H}(M)$ by induction on the reduction $\rightarrow_{h^c}$.

- If $M = M_1 +_l M_2 \rightarrow_{h^c} N_1 +_l M_2 = N$ with $M_1 \rightarrow_{h^c} N_1$ then by induction hypothesis $N_1 \twoheadrightarrow_{h^c} \mathrm{H}(M_1)$, and $N_1 +_l M_2 \twoheadrightarrow_{h^c} \mathrm{H}(M_1) +_l \mathrm{H}(M_2) = \mathrm{H}(M)$.

- If $M = M_1 +_l M_2 \rightarrow_{h^c} M_1 +_l N_2 = N$ with $M_2 \rightarrow_{h^c} N_2$ the case is similar.

- If $M = \lambda x_1...x_n.(\lambda y.M_0)\ P\ Q_1\ ...\ Q_m \rightarrow_{h^c} \text{can}\left(\lambda x_1...x_n.M_0 \left[{}^{P}/_{y}\right]\ Q_1\ ...\ Q_m\right)$ then $N = \mathrm{H}(M)$.

Next we have that $\rightarrow_{h^c}$ enjoys the diamond property: if we have two different reductions $M \rightarrow_{h^c} N_1$ and $M \rightarrow_{h^c} N_2$ then there is P such that $N_i \rightarrow_{h^c} P$ for $i \in \{1;2\}$. This is very simple to prove by induction on M.

Finally we can conclude by induction on the length of the reduction $M \twoheadrightarrow_{h^c} N$. If $M = N$ the result is immediate. Otherwise we have $M \rightarrow_{h^c}^{n} P \rightarrow_{h^c} N$ so by induction hypothesis there is P' such that $P \twoheadrightarrow_{h^c} P'$ and $M \twoheadrightarrow_H P'$, and using the diamond property of $\rightarrow_{h^c}$ there is Q such that $P' \rightarrow_{h^c}^{?} Q$ and $N \twoheadrightarrow_{h^c} Q$. Then $Q \twoheadrightarrow_{h^c} \mathrm{H}(P')$, so $N \twoheadrightarrow_{h^c} \mathrm{H}(P')$ and $M \twoheadrightarrow_H \mathrm{H}(P')$. □

Corollary 2.2.2.4. *If $M \twoheadrightarrow_{h^c} N$ and N is a head normal form then $M \twoheadrightarrow_H N$.*

Another interesting aspect of the complete head reduction is that it commutes with the canonical reduction and the syntactic equivalence.

Proposition 2.2.2.5. *If $M \rightarrow_{\beta^c} N$ then* $\mathrm{H}(M) \twoheadrightarrow_{\beta^c} \mathrm{H}(N)$.

Proof. By induction on M.

- If $M = M_1 +_l M_2$ then $N = N_1 +_l N_2$ with $M_i \rightarrow_{\beta^c}^{?} N_i$ for $i \in \{1;2\}$ so by induction hypothesis $\mathrm{H}(M_i) \twoheadrightarrow_{\beta^c} \mathrm{H}(N_i)$, hence $\mathrm{H}(M) \twoheadrightarrow_{\beta^c} \mathrm{H}(N)$.

- If $M = \lambda x_1...x_n.y\ P_1\ ...\ P_m$ then $N = \lambda x_1...x_n.y\ Q_1\ ...\ Q_m$ with $P_i \rightarrow_{\beta^c}^{?} Q_i$ for $i \leq m$ so $\mathrm{H}(M) = M$ and $\mathrm{H}(N) = N$.

- If $M = \lambda x_1 ... x_n.(\lambda y.M_0)\ P\ Q_1\ ...\ Q_m$ we note Δ the redex of the reduction $M \to_{\beta^c} N$. If Δ is the head redex of M then $N = \mathrm{H}(M)$ and we obviously have $N \twoheadrightarrow_{\beta^c} \mathrm{H}(N)$. Otherwise we have $M \xrightarrow{\{(\lambda y.M_0)\ P;\Delta\}}_{\beta^c_{/\!/}} \mathrm{H}(N)$ and $M\ (\to_\beta \cup \to_+)^*\ \mathrm{H}(M)\ (\to_\beta \cup \to_+)^*\ \mathrm{H}(N)$.

□

Proposition 2.2.2.6. *If M and N are canonical terms and $M \equiv_{\mathrm{syn}} N$ then $\mathrm{H}(M) \equiv_{\mathrm{syn}} \mathrm{H}(N)$.*

Proof. M, N, $\mathrm{H}(M)$ and $\mathrm{H}(N)$ are canonical terms so we can use the relation $\equiv^c_{\mathrm{syn}}$ defined in the section 1.1.2.

We reason by induction on $\equiv^c_{\mathrm{syn}}$.

- if $M \equiv^c_{\mathrm{syn}} P$ and $P \equiv^c_{\mathrm{syn}} N$ then by induction hypothesis $\mathrm{H}(M) \equiv^c_{\mathrm{syn}} \mathrm{H}(P)$ and $\mathrm{H}(P) \equiv^c_{\mathrm{syn}} \mathrm{H}(N)$, so $\mathrm{H}(M) \equiv^c_{\mathrm{syn}} \mathrm{H}(N)$;
- if $M \equiv^c_{\mathrm{syn}} N$ implies $N \equiv^c_{\mathrm{syn}} M$ then by induction hypothesis we have $\mathrm{H}(M) \equiv^c_{\mathrm{syn}} \mathrm{H}(N)$ so $\mathrm{H}(N) \equiv^c_{\mathrm{syn}} \mathrm{H}(M)$.
- $\mathrm{H}(M +_l N) = \mathrm{H}(M) +_l \mathrm{H}(N) \equiv^c_{\mathrm{syn}} \mathrm{H}(N) +_{\gamma(l)} \mathrm{H}(M) = \mathrm{H}(N +_{\gamma(l)} M)$;
- $$\begin{aligned} \mathrm{H}((M +_l N) +_{l'} P) &= (\mathrm{H}(M) +_l \mathrm{H}(N)) +_{l'} \mathrm{H}(P) \\ &\equiv^c_{\mathrm{syn}} \mathrm{H}(M) +_{\alpha_1(l,l')} (\mathrm{H}(N) +_{\alpha_2(l,l')} \mathrm{H}(P)) \\ &= \mathrm{H}(M +_{\alpha_1(l,l')} (N +_{\alpha_2(l,l')} P)); \end{aligned}$$
- $\mathrm{H}(M +_l M) = \mathrm{H}(M) +_l \mathrm{H}(M) \equiv^c_{\mathrm{syn}} \mathrm{H}(M)$ if $l \in \mathcal{I} \subset \mathcal{L}$;
- $\mathrm{H}(M +_l N) = \mathrm{H}(M) +_l \mathrm{H}(N) \equiv^c_{\mathrm{syn}} \mathrm{H}(M) +_l \mathrm{H}(P) = \mathrm{H}(M +_l P)$ if $l \in \mathcal{Z} \subset \mathcal{L}$.

The remaining cases are those of the contextual rules and they are given by induction hypothesis. □

These properties, and in particular the commutation of the complete head reduction with the syntactic equivalence, make this a very convenient reduction. But once again this is a completion of the basic head reduction, thus it does not work well to compare terms. We gave earlier an example with the term $\delta\ (\lambda x.\delta x)$ such that $\delta\ (\lambda x.\delta x) \to_H (\lambda x.\delta\ x)\ (\lambda x.\delta\ x) \to_H \delta\ (\lambda x.\delta x)$: the terms $\delta\ (\lambda x.\delta x) +_l (\lambda x.\delta\ x)\ (\lambda x.\delta x)$ and $\delta\ (\lambda x.\delta x) +_l \delta\ (\lambda x.\delta x)$ head reduce into a common term but their complete head reductions never meet.

We can define another interesting reduction which extends the complete head reduction by reducing inductively inside the head normal forms.

Definition 2.2.2.3. The *complete left reduction* $\to_L$ is defined between canonical terms by:

$$\frac{M_1 \to_L N_1 \qquad M_2 \to_L N_2}{M_1 +_l M_2 \to_L N_1 +_l N_2} \qquad \frac{\forall i \leq m, P_i \to_L Q_i}{\lambda x_1 ... x_n . y\ P_1\ ...\ P_m \to_L \lambda x_1 ... x_n . y\ Q_1\ ...\ Q_m}$$

$$\frac{}{\lambda x_1 ... x_n . (\lambda y . M)\ P\ Q_1\ ...\ Q_m \to_L \operatorname{can}\left(\lambda x_1 ... x_n . M \left[{}^{P}\!/_{y}\right]\ Q_1\ ...\ Q_m\right)}$$

This reduction enjoys properties similar to those of the complete head reduction.

Proposition 2.2.2.7. *For any canonical term M there is a unique term, which we note $\mathrm{L}(M)$, such that $M \to_L \mathrm{L}(M)$.*

Proof. By induction on M.

- If $M = M_1 +_l M_2$ then $M \to_L N$ if and only if $N = N_1 +_l N_2$ with $M_i \to_L N_i$ for $i \in \{1; 2\}$, and by induction hypothesis such terms exist and are unique.
- If $M = \lambda x_1 ... x_n . y\ P_1 ... P_m$ then $M \to_L N$ if and only if $N = \lambda x_1 ... x_n . y\ Q_1 ... Q_m$ with $P_i \to_L Q_i$ for $i \leq m$, and by induction hypothesis such terms exist and are unique.
- If $M = \lambda x_1 ... x_n . (\lambda y . M)\ P\ Q_1\ ...\ Q_m$ then M has a unique reduction.

□

Proposition 2.2.2.8. *If $M \to_{\beta^c} N$ then $\mathrm{L}(M) \twoheadrightarrow_{\beta^c} \mathrm{L}(N)$.*

Proof. By induction on M. If M is a sum or $M = \lambda x_1 ... x_n . y\ P_1\ ...\ P_m$ the result is immediate by induction hypothesis. If $M = \lambda x_1 ... x_n . (\lambda y . M_0)\ P\ Q_1\ ...\ Q_m$ then $\mathrm{L}(M) = \mathrm{H}(M)$, and we always have $\mathrm{H}(N) \twoheadrightarrow_{\beta^c} \mathrm{L}(N)$ so we can use the corresponding result on $\to_H$. □

Proposition 2.2.2.9. *If $M \equiv_{\mathrm{syn}} N$ then $\mathrm{L}(M) \equiv_{\mathrm{syn}} \mathrm{L}(N)$.*

Proof. Once again this is proved in the same way as the corresponding result for $\to_H$. □

The complete head reduction returns the head normal form of a term: if M has a head normal form N then $\mathrm{H}^n(M) = N$ for all $n \in \mathbb{N}$ big enough. The complete left reduction, on the other hand, can be used to reach any normal form.

Proposition 2.2.2.10. *If $M \twoheadrightarrow_{\beta^c} N$ and N is a canonical β-normal form then $M \twoheadrightarrow_L N$.*

Proof. This is a simple consequence of the standardization result. We know that $M \twoheadrightarrow_S N$ so we reason by induction on this relation.

- If $M = M_1 +_l M_2 \twoheadrightarrow_S N_1 +_l N_2 = N$ with $M_i \twoheadrightarrow_S N_i$ for $i \in \{1;2\}$ then by induction hypothesis we have $M_i \to_L^{n_i} N_i$ for $i \in \{1;2\}$, and as N is β-normal we have $\mathrm{L}(N_i) = N_i$ for $i \in \{1;2\}$. Hence $M_1 +_l M_2 \to_L^{\max(n_1,n_2)} N_1 +_l N_2$.

- If $M = \lambda x_1 ... x_n.(\lambda y.v)\ P\ Q_1\ ...\ Q_m \twoheadrightarrow_S N$ with $\mathrm{can}\left(\lambda x_1 ... x_n.v\left[{}^{P}/_{y}\right]\ Q_1\ ...\ Q_m\right) \twoheadrightarrow_S N$, i.e. $\mathrm{L}(M) \twoheadrightarrow_S N$, we have by induction hypothesis that $\mathrm{L}(M) \twoheadrightarrow_L N$ so $M \twoheadrightarrow_L N$.

- If $M = \lambda x_1 ... x_n.y\ P_1\ ...\ P_m \twoheadrightarrow_S \lambda x_1 ... x_n.y\ Q_1\ ...\ Q_m = N$ with $P_i \twoheadrightarrow_S Q_i$ for $i \leq m$ then by induction hypothesis $P_i \to_L^{n_i} Q_i$ for $i \leq m$ so $M \to_L^{\max_i n_i} N$.

- The case $M = \lambda x_1 ... x_n.(\lambda y.v)\ P_1\ ...\ P_m \twoheadrightarrow_S \mathrm{can}\,(\lambda x_1 ... x_n.(\lambda y.N_0)\ Q_1\ ...\ Q_m)$ with $m > 0$ is impossible as N is β-normal.

$\square$

This result can actually be extended to a more general notion of normal forms, namely the Böhm trees. We will later define probabilistic Böhm trees and show that the tree of a term M is an upper bound in some sense of the terms $\mathrm{L}^n(M)$.

3 Probabilistic equational theories

The intended purpose of our operational semantics was to get a contextual reduction without side effect, in order to easily define a notion of equational theory. This is something we can now do.

Definition 3.0.2.1. An *equational theory* is a congruence $=_{\mathcal{T}}$ on terms such that

$$\begin{array}{rcll} (\lambda x.M)\ N & =_{\mathcal{T}} & M\left[{}^{N}/_{x}\right] & \\ \lambda x.(M +_l N) & =_{\mathcal{T}} & \lambda x.M +_l \lambda x.N & \\ (M +_l N)\ P & =_{\mathcal{T}} & M\ P +_l N\ P & \\ & & & \\ M +_l N & =_{\mathcal{T}} & N +_{\gamma(l)} M & \\ (M +_l N) +_{l'} P & =_{\mathcal{T}} & M +_{\alpha_1(l,l')} (N +_{\alpha_2(l,l')} P) & \\ M +_l M & =_{\mathcal{T}} & M & \text{if } l \in \mathcal{I} \\ M +_l N & =_{\mathcal{T}} & M +_l P & \text{if } l \in \mathcal{Z}. \end{array}$$

The least theory induced by these rules is noted $=_{\beta+}$.

Proposition 3.0.2.1. *Given two canonical terms M and N we have $M =_{\beta+} N$ if and only if*

$$M \twoheadrightarrow_S \cdot \equiv_{\text{syn}} \cdot \twoheadleftarrow_{\beta^c} \cdot \twoheadleftarrow_{\text{split}} N.$$

Proof. The confluence of $\rightarrow_{\beta/\equiv}$ gives that $M =_{\beta+} N$ if and only if M and N reduces into the same term P. Then we know there is P' such that $P \twoheadrightarrow_{\beta/\equiv} P'$ and $M \twoheadrightarrow_S P$, which gives $N \twoheadrightarrow_{\beta/\equiv} P'$ hence $N \twoheadrightarrow_{\text{split}} \cdot \twoheadrightarrow_{\beta^c} \cdot \equiv_{\text{syn}} P'$. $\square$

In the deterministic λ-calculus the two theories defined by the observational equivalence and the infinitely extensional Böhm tree equality corresponds, and this is also the supremum of all sensible consistent theories. We want to prove a similar result in our non-deterministic case. To begin with we can define the consistency of a theory, as well as an other interesting property, namely the extensionality.

Definition 3.0.2.2. A theory $=_{\mathcal{T}}$ is *consistent* if there are terms M and N such that

$$M \neq_{\mathcal{T}} N.$$

Definition 3.0.2.3. A theory is *extensional* if for all term M and all variable $x \notin \mathrm{FV}(M)$ we have

$$\lambda x.M\ x =_{\mathcal{T}} M.$$

Another crucial notion is the sensibility of a theory. A deterministic theory is said to be sensible if the unsolvable terms, i.e. the terms which do not normalize for the head reduction, are all equal. This is a quite simple and straightforward notion. In a non deterministic setting it becomes more complex, as the head reduction of a term is not unique. We gave a notion of complete head reduction to solve this problem, but even so the problem is that such a reduction may require an infinite number of steps to describe the behaviour of a term. Consider for instance the terms

$$\mathrm{N}_k = \Theta\ (\lambda f.\lambda n.(n +_l f\ (\underline{\mathrm{succ}}\ n)))\ \underline{k}$$

where $k \in \mathbb{N}$, $\underline{k}$ is the corresponding Church integer, $\underline{\mathrm{succ}}$ is an encoding of the successor and Θ is a fixed-point operator such that for all M, $\Theta\ M \twoheadrightarrow_{h^c} M\ (\Theta\ M)$. Then we have

$$\mathrm{N}_k \twoheadrightarrow_{h^c} \underline{k} +_l \mathrm{N}_{k+1}$$

and for all k

$$\begin{array}{ccccccc}
 & & +_l & & & & \\
 & \swarrow & & \searrow & & & \\
 & \underline{0} & & +_l & & & \\
 & & \swarrow & & \searrow & & \\
\mathrm{N}_0 \twoheadrightarrow_{h^c} & & \underline{1} & & +_l & & \\
 & & & \swarrow & & \ddots & \\
 & & & \underline{2} & & +_l & \\
 & & & & \swarrow & & \searrow \\
 & & & & \underline{k} & & \mathrm{N}_{k+1}
\end{array}$$

The intuition is that N_0 reduces to the sum of all the Church integers and has one non terminating branch, and in a sensible theory it should be equal to every other such term. For instance if we define N'_k in the same way as N_k but by changing the implementation of the successor or the fixed-point operator, we should have $\mathrm{N}_0 =_{\mathcal{T}} \mathrm{N}'_0$ in every sensible theory $=_{\mathcal{T}}$.

But how do we formalize this intuition? In some cases this seems simple: if you consider for instance the probabilistic λ-calculus and $l = p$ then this infinite reduction occurs with probability $\prod_{k=0}^{\infty}(1-p) = 0$ if $p > 0$, whereas we have $\mathrm{N}_k \twoheadrightarrow_{\beta^c} \underline{k} +_0 \mathrm{N}_{k+1} \equiv_{\mathrm{syn}} \mathrm{N}_{k+1}$ if $p = 0$ so N_0 is unsolvable.

In other cases this is more complicated. If we consider the simple non deterministic calculus (with $\mathcal{L} = \{*\}$) then to prove that $\mathrm{N}_0 =_{\mathcal{T}} \mathrm{N}'_0$ we would like to prove that $\mathrm{N}_1 =_{\mathcal{T}} \mathrm{N}'_1$, but we can not claim that N_1 is "less important" than N_0 as we could in the probabilistic case. The problem is even clearer if we consider the term $\mathrm{N} = \Theta\ (\lambda f.\underline{0} + f)$. Then $\mathrm{N} \twoheadrightarrow_{\beta^c} \underline{0} + \mathrm{N}$, and to prove that if $\mathrm{N}' \twoheadrightarrow_{\beta^c} \underline{0} + \mathrm{N}'$ then $\mathrm{N} =_{\mathcal{T}} \mathrm{N}'$ we would precisely want to prove that $\mathrm{N} =_{\mathcal{T}} \mathrm{N}'$.

There are several ways to solve this problem. One way is to consider a strong sensibility based on the must convergence: a term is unsolvable if it has an infinite head reduction, regardless of whether some of its branches can reach a head normal form or not. In this setting solvable terms are always finite sums of value head normal forms, and we have powerful tools to obtain the result we seek. But this would require some computation on the labels, and we did not study what properties γ, α_1, α_2, $\mathcal{Z}$ and $\mathcal{I}$ should have. In the probabilistic case we are confident we can get the expected result: although we will not prove it, the techniques we will use should be easy to adapt to this notion of sensibility.

Another way is to make explicitly reference to the semantics of the terms. For instance in the simple case where terms describe sets of values we can say that a theory $=_\mathcal{T}$ is sensible if two terms are equal whenever they describe the same set when restricted to head normal values modulo $=_\mathcal{T}$. This corresponds to the intended meaning of the sensibility, but it is not a convenient definition when we study terms from a syntactical point of view.

What we will do will be to restrict our study to the probabilistic λ-calculus, where we have a natural notion of measure which we can use to approximate the infinite behaviours by finite ones. The infinite behaviours are actually always defined as limits of finite ones: for instance in the simple non deterministic calculus we can associate to a term an infinite set of head normal values, which we define as a union of finite sets. But we have no way to quantify how close we are to the result. In the probabilistic case, on the other hand, we associate to a term a subprobability distribution defined as the limit of finite ones, and here we can say that for all $\epsilon > 0$, the infinite distribution is approximated up to ϵ by a finite one.

In the rest of this thesis we will now only consider the probabilistic calculus. We will not consider probabilistic choice with any probability but only with computable ones. The set Λ_+ of terms is:

$$M, N \in \Lambda_+ := x \mid \lambda x.M \mid M\ N \mid M +_p N, p \in [0;1] \text{ computable}.$$

This restriction to computable probabilities is natural if we want to relate our probabilistic λ-calculus to actual programming languages, but in this purely theoretical setting it is less welcome. The only reason we do not consider arbitrary probabilities is to be able to prove the proposition 3.2.2.5, which will justify our definition of sensibility.

We have the following notion of theory.

Definition 3.0.2.4. A *probabilistic λ-theory* is a congruence $=_\mathcal{T}$ on probabilistic terms such that

$$\begin{array}{rcl}
(\lambda x.M)\ N & =_\mathcal{T} & M\left[{}^{N}/_{x}\right] \\
\lambda x.(M +_p N) & =_\mathcal{T} & \lambda x.M +_p \lambda x.N \\
(M +_p N)\ P & =_\mathcal{T} & M\ P +_p N\ P \\
\\
M +_p N & =_\mathcal{T} & N +_{1-p} M \\
(M +_p N) +_q P & =_\mathcal{T} & M +_{pq} (N +_{\frac{(1-p)q}{1-pq}} P) \text{ if } pq \neq 1 \\
M +_p M & =_\mathcal{T} & M \\
M +_1 N & =_\mathcal{T} & M +_1 P.
\end{array}$$

The least theory induced by these rules is noted $=_{\beta+}$.

Remark that the rule $(M +_1 N) +_1 P =_{\mathcal{T}} M +_1 (N +_{\frac{1}{2}} P)$ we gave in the first chapter can be recovered from the other equivalences, and its purpose was only to define total functions α_1 and α_2.

3.1 Probabilistic behaviour and probabilistic observation

3.1.1 Probability measures

The specificity of the probabilistic calculus is that we can associate a probability measure to some events. We consider that an event is given by a set of terms, and we are interested in the probability that such an event occurs if we start with a given term which behaves according to a given relation.

Definition 3.1.1.1. Given a set of terms $\mathcal{M} \subset \Lambda_+$ we define the following sets:

- $\mathcal{M}_+$ is the set of terms which respect $\mathcal{M}$, and is defined as the closure of $\mathcal{M}$ by probabilistic sum:

$$\mathcal{M} \subset \mathcal{M}_+ \text{ and if } M, N \in \mathcal{M}_+ \text{ then } M +_p N \in \mathcal{M}_+$$

- $\mathcal{M}_p$ for $p \in [0;1]$ is the set of terms which respect $\mathcal{M}$ with probability at least p:

$$\mathcal{M}_p = \{M \mid \exists N \in \mathcal{M}_+, \exists P \in \Lambda_+ : M \equiv N +_p P\}$$

Remark that the splitting of terms is quite convenient here, as it gives the following property:

Proposition 3.1.1.1. *If $p \leq q$ then for all $\mathcal{M}$ we have $\mathcal{M}_q \subset \mathcal{M}_p$.*

Proof. If $M \equiv N +_q P$ with $N \in \mathcal{M}_+$, we have

$$N +_q P \equiv (N +_{\frac{p}{q}} N) +_q P \equiv N +_p (N +_{\frac{q-p}{1-p}} P).$$

□

Now we can define the probability of occurrence of an event.

Definition 3.1.1.2. Given a term $M \in \Lambda_+$ and a relation $\mathcal{R} \subset \Lambda_+ \times \Lambda_+$, we define the probability of $\mathcal{M} \subset \Lambda_+$ by

$$\mathcal{P}(M\ \mathcal{R}\ \mathcal{M}) = \sup\{p \in [0;1] \mid \exists N \in \mathcal{M}_p : \mathrm{can}(M)\ \mathcal{R}\ N\}.$$

Remark that this is a generic definition, and for fixed M and $\mathcal{R}$ we may not directly have a probabilistic measure over Λ_+. If for instance $\mathcal{R}$ is the equality, we said that a term M modulo $\equiv$ can be seen as a finite probability distribution over values $M \equiv \sum_i p_i.v_i$, and for any set $\mathcal{M}$ of values modulo $\equiv^v_{\text{syn}}$ we have $\mathcal{P}(M = \mathcal{M}) = \sum_{i \text{ s.t. } v_i \in \mathcal{M}} p_i$, so we get a probability distribution over classes of values. If on the other hand $\mathcal{R} = \twoheadrightarrow_{h^c}$ we get a subprobability distribution over head normal values, but not over the whole set Λ_+.

We will actually only use this definition with sets of head normal values, but when proving some properties of this construction we will try to give the most general results.

First an important property is that under some conditions these probabilities commute with the sums.

Proposition 3.1.1.2. *If $\mathcal{R}$ and $\mathcal{M}$ are such that*

- *if $M \mathrel{\mathcal{R}} M'$ and $N \mathrel{\mathcal{R}} N'$ then $M +_p N \mathrel{\mathcal{R}} M' +_p N'$;*
- *if $M +_p N \mathrel{\mathcal{R}} P$ then $P = M' +_p N'$ with $M \mathrel{\mathcal{R}} M'$ and $N \mathrel{\mathcal{R}} N'$;*
- *$\mathcal{M}$ contains only values*

then

$$\mathcal{P}(M +_p N \mathrel{\mathcal{R}} \mathcal{M}) = p.\mathcal{P}(M \mathrel{\mathcal{R}} \mathcal{M}) + (1-p).\mathcal{P}(N \mathrel{\mathcal{R}} \mathcal{M}).$$

Proof. We assume w.l.o.g. that M and N are canonical.

If $\mathcal{M}$ contains only values then we can view any term M as a probability distribution over values $M \equiv \sum_i p_i.v_i + \sum_j p'_j.w_j$ where $v_i \in \mathcal{M}$ for all i and $w_j \notin \mathcal{M}$ for all j. Then $M \in \mathcal{M}_q$ if and only if $q \leq \sum_i p_i$, and $\mathcal{P}(M = \mathcal{M}) = \sum_i p_i$. From there we get that

$$\mathcal{P}(M +_p N = \mathcal{M}) = p.\mathcal{P}(M = \mathcal{M}) + (1-p).\mathcal{P}(N = \mathcal{M}).$$

We can also remark that for any term M:

$$\begin{aligned}\mathcal{P}(M \mathrel{\mathcal{R}} \mathcal{M}) &= \sup\{p \mid \exists M' \in \mathcal{M}_p : M \mathrel{\mathcal{R}} M'\} \\ &= \sup\{\sup\{p \mid M' \in \mathcal{M}_p\} \mid M \mathrel{\mathcal{R}} M'\} \\ &= \sup\{\mathcal{P}(M' = \mathcal{M}) \mid M \mathrel{\mathcal{R}} M'\}.\end{aligned}$$

Then

$$\mathcal{P}(M +_p N \mathrel{\mathcal{R}} \mathcal{M}) = \sup\{\mathcal{P}(P = \mathcal{M}) \mid M +_p N \mathrel{\mathcal{R}} P\}$$

and

$$\begin{aligned}&p.\mathcal{P}(M \mathrel{\mathcal{R}} \mathcal{M}) + (1-p).\mathcal{P}(N \mathrel{\mathcal{R}} \mathcal{M}) \\ &\qquad = p.\sup\{\mathcal{P}(Q = \mathcal{M}) \mid M \mathrel{\mathcal{R}} Q\} + (1-p).\{\mathcal{P}(R = \mathcal{M}) \mid N \mathrel{\mathcal{R}} R\} \\ &\qquad = \sup\{p.\mathcal{P}(Q = \mathcal{M}) + (1-p).\mathcal{P}(R = \mathcal{M}) \mid M \mathrel{\mathcal{R}} Q, N \mathrel{\mathcal{R}} R\} \\ &\qquad = \sup\{\mathcal{P}(Q +_p R = \mathcal{M}) \mid M \mathrel{\mathcal{R}} Q, N \mathrel{\mathcal{R}} R\}.\end{aligned}$$

But the conditions on $\mathcal{R}$ give precisely that if $M +_p N \mathrel{\mathcal{R}} P$ then $P \equiv Q +_p R$ (hence $\mathcal{P}(P = \mathcal{M}) = \mathcal{P}(Q +_p R = \mathcal{M})$) with $M \mathrel{\mathcal{R}} Q$ and $N \mathrel{\mathcal{R}} R$, and that conversely if $M \mathrel{\mathcal{R}} Q$ and $N \mathrel{\mathcal{R}} R$ then $M +_p N \mathrel{\mathcal{R}} Q +_p R$. Thus the sets $\{\mathcal{P}(P = \mathcal{M}) \mid M +_p N \mathrel{\mathcal{R}} P\}$ and $\{\mathcal{P}(Q +_p R = \mathcal{M}) \mid M \mathrel{\mathcal{R}} Q, N \mathrel{\mathcal{R}} R\}$ are equal. $\square$

The relations we are interested in here are our different reductions and equivalences. Under some conditions on $\mathcal{M}$ we can show that some of these relations express the same behaviour.

Proposition 3.1.1.3. *If $\mathcal{M}_+$ is stable by β-reduction then for all $M \in \Lambda_+$:*

$$\mathcal{P}(M =_{\beta+} \mathcal{M}) = \mathcal{P}(M \twoheadrightarrow_{\beta/\equiv} \mathcal{M}) = \mathcal{P}(M \twoheadrightarrow_S \mathcal{M}).$$

Proof. We know that $\twoheadrightarrow_S \subset \twoheadrightarrow_{\beta/\equiv} \subset =_{\beta+}$ so we have

$$\mathcal{P}(M =_{\beta+} \mathcal{M}) \geq \mathcal{P}(M \twoheadrightarrow_{\beta/\equiv} \mathcal{M}) \geq \mathcal{P}(M \twoheadrightarrow_S \mathcal{M}).$$

We want to prove that if $M =_{\beta+} M' \in \mathcal{M}_p$, i.e. $M =_{\beta+} N +_p P$ with $N \in \mathcal{M}_+$, then $\mathrm{can}(M) \twoheadrightarrow_S \cdot \equiv N' +_p P'$ with $N' \in \mathcal{M}_+$.

If $M =_{\beta+} N +_p P$ then

$$\mathrm{can}(M) \twoheadrightarrow_S \cdot \equiv_{\mathrm{syn}} \cdot \twoheadleftarrow_{\beta^c} \cdot \twoheadleftarrow_{\mathrm{split}} \mathrm{can}(N) +_p \mathrm{can}(P).$$

The reductions $\twoheadrightarrow_{\mathrm{split}}$ and $\twoheadrightarrow_{\beta^c}$ preserve sums, so there are N' and P' such that $\mathrm{can}(N) \twoheadrightarrow_{\mathrm{split}} \cdot \twoheadrightarrow_{\beta^c} N'$, $\mathrm{can}(P) \twoheadrightarrow_{\mathrm{split}} \cdot \twoheadrightarrow_{\beta^c} P'$ and $\mathrm{can}(M) \twoheadrightarrow_S \cdot \equiv_{\mathrm{syn}} N' +_p P'$.

By definition $\mathcal{M}_+$ is stable by $\equiv$ so if it is stable by $\to_\beta$ it is also stable by $\twoheadrightarrow_{\beta/\equiv}$. This implies that $N' \in \mathcal{M}_+$. ☐

Proposition 3.1.1.4. *If $\mathcal{M}_+$ is closed by head reduction then for all $M \in \Lambda_+$:*

$$\mathcal{P}(M \twoheadrightarrow_{h^c} \mathcal{M}) = \mathcal{P}(M \twoheadrightarrow_H \mathcal{M}).$$

Proof. Again $\twoheadrightarrow_H \subset \twoheadrightarrow_{h^c}$ so $\mathcal{P}(M \twoheadrightarrow_{h^c} \mathcal{M}) \geq \mathcal{P}(M \twoheadrightarrow_H \mathcal{M})$, and conversely we want to prove that if $\mathrm{can}(M) \twoheadrightarrow_{h^c} \cdot \equiv N +_p P$ for some $N \in \mathcal{M}_+$ and P then $\mathrm{can}(M) \twoheadrightarrow_H \cdot \equiv N' +_p P'$ with $N' \in \mathcal{M}_+$.

We know there is Q such that $\mathrm{can}(M) \twoheadrightarrow_H Q$ and $N +_p P \equiv \cdot \twoheadrightarrow_{h^c} Q$. Then we have $N +_p P \twoheadrightarrow_{\mathrm{split}} \cdot \twoheadrightarrow_{h^c} \cdot \equiv Q$ so as $\twoheadrightarrow_{\mathrm{split}}$ and $\twoheadrightarrow_{h^c}$ preserve sums there are N' and P' with $N \twoheadrightarrow_{\mathrm{split}} \cdot \twoheadrightarrow_{h^c} N'$, $P \twoheadrightarrow_{\mathrm{split}} \cdot \twoheadrightarrow_{h^c} P'$ and $N' +_p P' \equiv Q$. As $\mathcal{M}_+$ is closed by head reduction and by $\equiv$ we have $N' \in \mathcal{M}_+$. ☐

3.1.2 Observation

In the deterministic λ-calculus we define an observational equivalence where observing a term amounts to checking whether its head reduction terminates. We can remark that for any deterministic term, i.e. any term without sum, we have

$$\mathcal{P}(M \twoheadrightarrow_h \mathcal{H}) = \begin{cases} 1 & \text{if } M \text{ is solvable} \\ 0 & \text{if } M \text{ is unsolvable} \end{cases}$$

where $\mathcal{H}$ is the set of all deterministic head normal forms.

Then extending this notion of observation to the probabilistic case is very natural.

Definition 3.1.2.1. The *convergence probability* of a term M is

$$\mathcal{P}_{\Downarrow}(M) = \mathcal{P}(M \twoheadrightarrow_{h^c} \mathcal{H})$$

where $\mathcal{H} = \{\lambda x_1 ... x_n.y\ P_1\ ...\ P_m\}$ is the set of all head normal values.

Proposition 3.1.2.1. *1. For every set $\mathcal{H}$ of head normal values, every terms $M, N \in \Lambda_+$ and every probability $p \in [0;1]$:*

$$\mathcal{P}(M +_p N \twoheadrightarrow_{h^c} \mathcal{H}) = p.\mathcal{P}(M \twoheadrightarrow_{h^c} \mathcal{H}) + (1-p).\mathcal{P}(N \twoheadrightarrow_{h^c} \mathcal{H}).$$

2. For any set $\mathcal{H}$ of value head normal forms closed by internal β-equivalence and for any term $M \in \Lambda_+$:

$$\mathcal{P}(M \twoheadrightarrow_{h^c} \mathcal{H}) = \mathcal{P}(M \twoheadrightarrow_H \mathcal{H}) = \mathcal{P}(M =_{\beta+} \mathcal{H})$$

Proof. By $\mathcal{H}$ closed by internal equivalence we mean that if $\lambda x_1 ... x_n.y\ P_1\ ...\ P_m \in \mathcal{H}$ and $P_i =_{\beta+} Q_i$ for all $i \leq m$ then $\lambda x_1 ... x_n.y\ Q_1\ ...\ Q_m \in \mathcal{H}$

The first result is a direct application of proposition 3.1.1.2.

For the second one we can apply the two previous results, and we get:

$$\begin{aligned} &\mathcal{P}(M =_{\beta+} \mathcal{H}) = \mathcal{P}(M \twoheadrightarrow_{\beta/\equiv} \mathcal{H}) = \mathcal{P}(M \twoheadrightarrow_S \mathcal{H}) \\ &\mathcal{P}(M \twoheadrightarrow_{h^c} \mathcal{H}) = \mathcal{P}(M \twoheadrightarrow_H \mathcal{H}) \end{aligned}$$

All we have left to prove is that $\mathcal{P}(M \twoheadrightarrow_S \mathcal{H}) = \mathcal{P}(M \twoheadrightarrow_{h^c} \mathcal{H})$. We have $\twoheadrightarrow_{h^c} \subset \twoheadrightarrow_S$ so $\mathcal{P}(M \twoheadrightarrow_S \mathcal{H}) \geq \mathcal{P}(M \twoheadrightarrow_{h^c} \mathcal{H})$.

The converse is given by the fact that if $\mathrm{can}(M) \twoheadrightarrow_S \cdot \equiv N +_p P$ with $N \in \mathcal{H}_+$ then $\mathrm{can}(M) \twoheadrightarrow_{h^c} Q \twoheadrightarrow_{i^c} \cdot \equiv N +_p P$, and since $\mathcal{H}$ is closed by internal β-equivalence we already have $Q \in \mathcal{H}_p$. □

Corollary 3.1.2.2. *1. For every terms $M, N \in \Lambda_+$ and every probability $p \in [0;1]$:*

$$\mathcal{P}_\Downarrow(M +_p N) = p.\mathcal{P}_\Downarrow(M) + (1-p).\mathcal{P}_\Downarrow(N)$$

2. For any term $M \in \Lambda_+$:

$$\mathcal{P}_\Downarrow(M) = \mathcal{P}(M \twoheadrightarrow_H \mathcal{H}) = \mathcal{P}(M =_{\beta+} \mathcal{H})$$

Definition 3.1.2.2. The *observational equivalence* $=_{obs}$ is defined by

$$M =_{obs} N \ \text{ iff } \ \forall C, \mathcal{P}_\Downarrow(C[M]) = \mathcal{P}_\Downarrow(C[N]).$$

Proposition 3.1.2.3. *$=_{obs}$ is a theory.*

Proof. For a fixed context C the relation $\mathcal{P}_\Downarrow(C[M]) = \mathcal{P}_\Downarrow(C[N])$ is trivially an equivalence, thus so is $=_{obs}$.

$=_{obs}$ is also contextual: if $M =_{obs} N$ then for any context C, for any context C' we have $\mathcal{P}_\Downarrow(C'[C[M]]) = \mathcal{P}_\Downarrow(C'[C[N]])$ so $C[M] =_{obs} C[N]$.

We also proved that $\mathcal{P}_\Downarrow(M) = \mathcal{P}(M =_{\beta+} \mathcal{H})$ so $=_{obs}$ is stable by $\to_\beta$, $\to_+$ and $\equiv_{\mathrm{syn}}$. □

It is easy to see that $=_{obs}$ is coherent. It is also extensional but this is much less simple to prove, and it will come as a consequence of our separability result.

We want to prove that $=_{obs}$ is the same as the equality of infinitely extensional Böhm trees, as well as the maximum sensible coherent theory. So first of all we need to define the sensibility of a theory.

3.2 Sensible theories

To define the sensibility of a theory we want to find a congruence $=_\Omega$ which express the equality of unsolvable terms. In a deterministic case such a congruence is simply induced by $M =_\Omega N$ whenever M and N are unsolvable, but as we pointed out before this does not work in the probabilistic case.

We gave a term N_0 whose reduction gives the infinite sum of all Church integers $\underline{k}$:

$$
\mathrm{N}_0 \twoheadrightarrow_{\beta^c}
\begin{array}{ccccccccc}
 & +_p & & & & & & & \\
 / & & \backslash & & & & & & \\
\underline{0} & & & +_p & & & & & \\
 & & / & & \backslash & & & & \\
 & \underline{1} & & & & +_p & & & \\
 & & & & / & & \ddots & & \\
 & & & \underline{2} & & & & +_p & \\
 & & & & & & / & & \backslash \\
 & & & & & \underline{k} & & & \mathrm{N}_{k+1}
\end{array}
$$

Given two such sequence (N_k) and $(\mathrm{N}')_k$ we want to have $\mathrm{N}_0 =_\Omega \mathrm{N}'_0$.

3.2.1 Continuity

This case is actually a particular one, as the terms (N_k) and $(\mathrm{N}')_k$ converge with probability 1. Then we can use another notion in the probabilistic theories, that we call continuity.

Definition 3.2.1.1. A probabilistic theory $=_\mathcal{T}$ is *continuous* if for all M and N,

$$\text{if } \forall \epsilon > 0, \exists P, Q : M +_{1-\epsilon} P =_\mathcal{T} N +_{1-\epsilon} Q \text{ then } M =_\mathcal{T} N.$$

The intuition behind this notion is the same as the one behind the sensibility: two terms which have the same behaviour should be equal, and two behaviours are equal if they are equal up to every $\epsilon > 0$. Given any two terms M and N such that for all $\epsilon > 0$ there are P_ϵ and Q_ϵ such that $M +_{1-\epsilon} P_\epsilon =_{\beta+} N +_{1-\epsilon} Q_\epsilon$, given any set $\mathcal{H}$ of head normal values closed by β, we have

$$\mathcal{P}(M +_{1-\epsilon} P_\epsilon \twoheadrightarrow_{h^c} \mathcal{H}) = \mathcal{P}(N +_{1-\epsilon} Q_\epsilon \twoheadrightarrow_{h^c} \mathcal{H})$$

so

$$\begin{aligned}
&(1-\epsilon).\mathcal{P}(M \twoheadrightarrow_{h^c} \mathcal{H}) + \epsilon.\mathcal{P}(P_\epsilon \twoheadrightarrow_{h^c} \mathcal{H}) \\
=&(1-\epsilon).\mathcal{P}(N \twoheadrightarrow_{h^c} \mathcal{H}) + \epsilon.\mathcal{P}(Q_\epsilon \twoheadrightarrow_{h^c} \mathcal{H})
\end{aligned}$$

and necessarily

$$\mathcal{P}(M \twoheadrightarrow_{h^c} \mathcal{H}) = \mathcal{P}(N \twoheadrightarrow_{h^c} \mathcal{H}).$$

Besides this property has some interesting consequences on the theories.

Proposition 3.2.1.1. *If $=_{\mathcal{T}}$ is a continuous theory then for all terms M, N and P and for all probability $p > 0$:*

$$\textit{if } M +_p P =_{\mathcal{T}} N +_p P \textit{ then } M =_{\mathcal{T}} N.$$

Proof. If $p = 1$ the result is immediate.

Otherwise we have

$$\begin{aligned}
M +_{\frac{p}{1+p}} (M +_p P) &=_{\mathcal{T}} M +_{\frac{p}{1+p}} (N +_p P) \\
&\equiv_{\text{syn}} \frac{p}{1+p}.M + \frac{p}{1+p}.N + \frac{1-p}{1+p}.P \\
&\equiv_{\text{syn}} N +_{\frac{p}{1+p}} (M +_p P) \\
&=_{\mathcal{T}} N +_{\frac{p}{1+p}} (N +_p P)
\end{aligned}$$

thus $M +_{\frac{2p}{1+p}} P =_{\mathcal{T}} N +_{\frac{2p}{1+p}} P$.

Let us define $p_0 = p$ and $p_{n+1} = \frac{2p_n}{1+p_n}$ for $n \in \mathbb{N}$. For all $n \in \mathbb{N}$ we have $M +_{p_n} P =_{\mathcal{T}} N +_{p_n} P$. It is easy to see that the sequence (p_n) is increasing and converges toward 1. Then for all $\epsilon > 0$ there is n such that $p_n > 1 - \epsilon$, and there are Q and R such that $M +_{1-\epsilon} Q \equiv_{\text{syn}} M +_{p_n} P =_{\mathcal{T}} N +_{p_n} P \equiv_{\text{syn}} N +_{1-\epsilon} R$. □

Corollary 3.2.1.2. *If $=_{\mathcal{T}}$ is a continuous theory then for all terms M and N:*

$$\textit{if } \exists p \neq q : M +_p N =_{\mathcal{T}} M +_q N \textit{ then } M =_{\mathcal{T}} N.$$

Proof. Assume w.l.o.g. $p < q$, we have $M +_p N \equiv_{\text{syn}} p.M + (q-p).M + q.N$ and $M +_Q N \equiv_{\text{syn}} p.M + (q-p).N + q.N$ so we can use the previous result twice to simplify $p.M$ and $q.N$. □

With the continuity we can prove that the terms N_0 and N'_0 are equal. If $p = 1$ then they are both equal to $\underline{0}$. Otherwise we have for all k

$$\begin{aligned}
\mathrm{N}_0 &=_{\beta+} \sum_{i=0}^{k} (1-p)^i p.\underline{i} + (1-p)^{k+1} N_{k+1} \\
\mathrm{N}'_0 &=_{\beta+} \sum_{i=0}^{k} (1-p)^i p.\underline{i} + (1-p)^{k+1} N'_{k+1}
\end{aligned}$$

so for all k

$$\frac{1}{1+(1-p)^k}.\mathrm{N}_0 + \frac{(1-p)^k}{1+(1-p)^k}.\mathrm{N}'_k =_{\beta+} \frac{1}{1+(1-p)^k}.\mathrm{N}'_0 + \frac{(1-p)^k}{1+(1-p)^k}.\mathrm{N}_k.$$

$\frac{(1-p)^k}{1+(1-p)^k}$ converges to 0 so we get $\mathrm{N}_0 =_{\mathcal{T}} \mathrm{N}'_0$ if $=_{\mathcal{T}}$ is continuous.

But as we mentioned before the continuity is enough because N_0 and N'_0 converges with probability 1. In this case we need to get rid of an infinite reduction of probability 0, but we do not really need to say that two actual non terminating behaviour are equal.

3.2.2 Weak and strong sensibilities

Let us consider a variant of our term N_0 which diverges with a non zero probability. Given a probability p we define the following terms:

$$\begin{aligned}
\mathbb{T} &= \lambda xy.x \\
\mathbb{F} &= \lambda xy.y \\
\mathrm{P} &= \lambda n.n\ (\lambda x.x +_p \mathbb{F})\ \mathbb{T} \\
\mathrm{N}^{\mathrm{P}}_k &= \Theta\ (\lambda f.\lambda n.\mathrm{P}\ n\ n\ (f\ (\underline{\mathrm{succ}}\ n)))\ \underline{k}.
\end{aligned}$$

We have that

$$\begin{aligned}
\mathrm{P}\ \underline{k} &\twoheadrightarrow_{\beta/\equiv} \mathbb{T} +_{p^k} \mathbb{F} \\
\mathrm{N}^{\mathrm{P}}_k &\twoheadrightarrow_{\beta/\equiv} \underline{k} +_{p^k} \mathrm{N}^{\mathrm{P}}_{k+1}.
\end{aligned}$$

Then

$$\mathrm{N}^{\mathrm{P}}_1 \twoheadrightarrow_{\beta/\equiv} \begin{array}{ccccccc}
 & +_p & & & & & \\
 / & & \backslash & & & & \\
\underline{1} & & & +_{p^2} & & & \\
 & & / & & \backslash & & \\
 & \underline{2} & & & +_{p^3} & & \\
 & & & / & & \ddots & \\
 & & \underline{3} & & & +_{p^k} & \\
 & & & & / & & \backslash \\
 & & & \underline{k} & & & \mathrm{N}^{\mathrm{P}}_{k+1}
\end{array}$$

This term converges with probability $\sum_{k=1}^{\infty} p^k \prod_{i=1}^{k-1}(1-p^i)$, which is strictly less than 1 when $p < 1$.

Here we cannot proceed as before to prove that if some $\mathrm{N}'^{\mathrm{P}}_1$ behaves as $\mathrm{N}^{\mathrm{P}}_k$ then they are equal in every continuous theory, as the probability to reach the subterm $\mathrm{N}^{\mathrm{P}}_k$ does not converge to 0.

We reason once more by approximation up to $\epsilon > 0$. We simply extend the relation $=_\Omega$ which equates unsolvable terms with this idea of probabilistic approximation.

Definition 3.2.2.1. For $\epsilon \geq 0$, the *sensible equality up to* ϵ $=_{\Omega,\epsilon}$ is given by:

$$\frac{\mathcal{P}_{\Downarrow}(M) \leq \epsilon \qquad \mathcal{P}_{\Downarrow}(N) \leq \epsilon}{M =_{\Omega,\epsilon} N} \qquad \frac{}{x =_{\Omega,\epsilon} x} \qquad \frac{M =_{\Omega,\epsilon} N}{\lambda x.M =_{\Omega,\epsilon} \lambda x.N}$$

$$\frac{M =_{\Omega,\epsilon} N \qquad M' =_{\Omega,\epsilon} N'}{M\ M' =_{\Omega,\epsilon} N\ N'} \qquad \frac{M_1 =_{\Omega,\epsilon} N_1 \qquad M_2 =_{\Omega,\epsilon} N_2}{M_1 +_p M_2 =_{\Omega,\epsilon} N_1 +_p N_2}$$

Remark that since some quantification is involved this relation is not transitive. For instance we have $\underline{0} +_\epsilon \Omega =_{\Omega,\epsilon} \Omega$ so

$$\underline{0} +_\epsilon (\underline{0} +_\epsilon \Omega) =_{\Omega,\epsilon} \underline{0} +_\epsilon \Omega =_{\Omega,\epsilon} \Omega$$

but we do not want $\underline{0}+_\epsilon(\underline{0}+_\epsilon\Omega) =_{\Omega,\epsilon} \Omega$. What we have is $\underline{0}+_\epsilon(\underline{0}+_\epsilon\Omega)\Omega =_{\Omega,\epsilon(2-\epsilon)} \Omega$.

Definition 3.2.2.2. A probabilistic theory $=_\mathcal{T}$ is *weakly sensible* if for all M and N,

$$\text{if } \forall\epsilon > 0, M \twoheadrightarrow_{\beta/\equiv} \cdot =_{\Omega,\epsilon} \cdot \twoheadleftarrow_{\beta/\equiv} N \text{ then } M =_\mathcal{T} N.$$

Then for all $\epsilon > 0$ we can find k such that $\mathcal{P}_\Downarrow(\mathrm{N}^\mathrm{P}_k) \leq \epsilon$, so if $\mathrm{N'}^\mathrm{P}_1$ behaves as N^P_k we can prove they are equal in every weakly sensible theory.

The weak sensibility as an interesting consequence. We could expect that for any term M, the diverging part of M is in some sense equal to the unsolvable term Ω in sensible theories. A formalization of this is to say that for every term M there exists a term M_0 such that $M =_\mathcal{T} M_0 +_{\mathcal{P}_\Downarrow(M)} \Omega$.

The only trouble to get such a result is to define M_0. If the behaviour of M is described by a finite term then we easily get this property. For instance if $\mathcal{P}_\Downarrow(M) \neq 0$ and $\mathcal{P}(M \twoheadrightarrow_{h^c} \{h\}) \neq 0$ for finitely many head normal values h, then we can define $M_0 = \sum_h \frac{\mathcal{P}(M\twoheadrightarrow_{h^c}\{h\})}{\mathcal{P}_\Downarrow(M)}.h$ (which is a finite sum) and in every weakly sensible theory we have $M =_\mathcal{T} M_0 +_{\mathcal{P}_\Downarrow(M)} \Omega$. But if we consider for instance our term $M = \mathrm{N}^\mathrm{P}_1$ then morally we want to have $M_0 \twoheadrightarrow_{\beta/\equiv} \sum_{k\in\mathbb{N}} \frac{p^k\prod_{i=1}^{k-1}(1-p^i)}{\mathcal{P}_\Downarrow(M)}.\underline{k}$. As this sum is infinite we can not take this as a definition of M_0.

We are not certain it is always possible to build such a term M_0, but we can approximate it. For any term M we can find computable sequences (h_n) of head normal values and (p_n) of probabilities such that for all head normal value h we have $\mathcal{P}(M \twoheadrightarrow_{h^c} \{h\}) = \sum_{i \text{ s.t. } h_i\equiv h} p_i$. Then given any computable probability C with $C \geq \mathcal{P}_\Downarrow(M)$ we can use these sequences to build a term M_C such that $M =_\mathcal{T} M_C +_C \Omega$.

As we rely on the notion of computability, let us define some notions of representations.

Definition 3.2.2.3.

1. A probability $p = \sum_{k\in\mathbb{N}^*} \frac{p_k}{2^k} \in [0;1]$ with $p_k \in \{0;1\}$ for $k \in \mathbb{N}^*$ is represented by a term P if for all $k \in \mathbb{N}^*$ we have

$$P\ \underline{k} \twoheadrightarrow_\beta \begin{cases} \mathbb{T} & \text{if } p_k = 1 \\ \mathbb{F} & \text{if } p_k = 0 \end{cases}.$$

2. A sequence of probabilities (p_n) is represented by a term S if for all $n \in \mathbb{N}$ the term $S\ \underline{n}$ represents p_n.

3. A sequence of head normal values (h_n) is represented by a term S if for all $n \in \mathbb{N}$ we have $S\ \underline{n} \twoheadrightarrow_\beta h_n$.

Definition 3.2.2.4. The term $\underline{read}$ is given by

$$\underline{read} = \lambda p.\Theta\ (\lambda\varphi.\lambda k.p\ k +_{\frac{1}{2}} \varphi\ (\underline{\text{succ}}\ k))\ \underline{1}.$$

Proposition 3.2.2.1. *If P represents $p \in [0;1]$ then $\mathcal{P}(\underline{read}\ P \twoheadrightarrow_{\beta/\equiv} \mathbb{T}) = p$ and $\mathcal{P}(\underline{read}\ P \twoheadrightarrow_{\beta/\equiv} \mathbb{F}) = 1 - p$.*

Proof. Let $p = \sum_{k\in\mathbb{N}} \frac{p_k}{2^k}$, we have $1 - p = \sum_{k\in\mathbb{N}} \frac{1-p_k}{2^k}$.

Let us write

$$P_k = \Theta\ (\lambda\varphi.\lambda k.P\ k +_{\frac{1}{2}} \varphi\ (\underline{\text{succ}}\ k))\ \underline{k}$$

, we want to prove that P_1 behaves as $\mathbb{T} +_p \mathbb{F}$.

For all $k \geq 1$ we have

$$P_k \twoheadrightarrow_{\beta/\equiv} P\ \underline{k} +_{\frac{1}{2}} P_{k+1}$$

so

$$\begin{aligned} \mathcal{P}(P_k \twoheadrightarrow_{\beta/\equiv} \mathbb{T}) &= \frac{p_k}{2} + \mathcal{P}(P_{k+1} \twoheadrightarrow_{\beta/\equiv} \mathbb{T}) \\ \mathcal{P}(P_k \twoheadrightarrow_{\beta/\equiv} \mathbb{F}) &= \frac{1-p_k}{2} + \mathcal{P}(P_{k+1} \twoheadrightarrow_{\beta/\equiv} \mathbb{F}). \end{aligned}$$

We get $\mathcal{P}(P_1 \twoheadrightarrow_{\beta/\equiv} \mathbb{T}) = \sum_{k\in\mathbb{N}} \frac{p_k}{2^k}$ and $\mathcal{P}(P_1 \twoheadrightarrow_{\beta/\equiv} \mathbb{F}) = \sum_{k\in\mathbb{N}} \frac{1-p_k}{2^k}$. $\square$

Proposition 3.2.2.2. *Given computable sequences (p_n) of probabilities and (h_n) of head normal values represented by terms S_p and S_h, let*

$$M_{(p_n),(h_n)} = \Theta\ (\lambda\varphi.\lambda n.\underline{read}\ (S_p\ n)\ (S_h\ n)\ (\varphi\ (\underline{\text{succ}}\ n)))\ \underline{0}.$$

For all $n \in \mathbb{N}$ and $\epsilon > 0$ there exists terms $M^n_{(p_n),(h_n)}$ and R^n_ϵ such that

$$M_{(p_n),(h_n)} \twoheadrightarrow_{\beta/\equiv} \begin{array}{cccccccccc}
 & & & & +_{1-\epsilon} & & & & & \\
 & & & / & & \backslash & & & & \\
 & & +_{p_0} & & & & R^n_\epsilon & & & \\
 & / & & \backslash & & & & & & \\
h_0 & & & & +_{p_1} & & & & & \\
 & & & / & & \backslash & & & & \\
 & & h_1 & & & & +_{p_2} & & & \\
 & & & & & / & & \ddots & & \\
 & & & & h_2 & & & & +_{p_n} & \\
 & & & & & & & / & & \backslash \\
 & & & & & & h_n & & & M^{n+1}_{(p_n),(h_n)}
\end{array}.$$

Proof. Let

$$M^n_{(p_n),(h_n)} = \Theta\ (\lambda\varphi.\lambda n.\underline{read}\ (S_p\ n)\ (S_h\ n)\ (\varphi\ (\underline{\text{succ}}\ n)))\ \underline{n}.$$

For all $n \in \mathbb{N}$ we have

$$M^n_{(p_n),(h_n)} \twoheadrightarrow_{\beta/\equiv} \underline{read}\ (S_p\ \underline{n})\ (S_h\ \underline{n})\ M^{n+1}_{(p_n),(h_n)}.$$

According to the previous result $\underline{read}\ (S_p\ \underline{n})$ behaves as $\mathbb{T}+_{p_n}\mathbb{F}$, and $S_h\ \underline{n} \twoheadrightarrow_{\beta/\equiv} h_n$ so for all $\epsilon > 0$ we can find Q^n_ϵ such that $M^n_{(p_n),(h_n)} \twoheadrightarrow_{\beta/\equiv} (h_n +_{p_n} M^{n+1}_{(p_n),(h_n)}) +_{1-\epsilon} Q^n_\epsilon$.

We prove the result by induction on n. For $n = -1$ we have that $M_{(p_n),(h_n)} = M^0_{(p_n),(h_n)}$. Otherwise we have by induction hypothesis that for all $\epsilon > 0$

$$M_{(p_n),(h_n)} \twoheadrightarrow_{\beta/\equiv} \left(\sum_{k=0}^{n} p_k \prod_{i=0}^{k-1} (1-p_i).h_k + \prod_{k=0}^{n} (1-p_k).M^{n+1}_{(p_n),(h_n)} \right) +_{1-\epsilon} R^n_\epsilon.$$

We get

$$\begin{aligned}
M_{(p_n),(h_n)} \twoheadrightarrow_{\beta/\equiv} & \sum_{k=0}^{n} (1-\epsilon) p_k \prod_{i=0}^{k-1} (1-p_i).h_k + (1-\epsilon)^2 p_{n+1} \prod_{k=0}^{n} (1-p_k).h_{n+1} \\
& + (1-\epsilon)^2 \prod_{k=0}^{n+1} (1-p_k).M^{n+2}_{(p_n),(h_n)} + (1-\epsilon)\epsilon \prod_{k=0}^{n} (1-p_k).Q^{n+1}_\epsilon + \epsilon.R^n_\epsilon \\
\twoheadrightarrow_{\beta/\equiv} & \sum_{k=0}^{n+1} (1-\epsilon)^2 p_k \prod_{i=0}^{k-1} (1-p_i).h_k + (1-\epsilon)^2 \prod_{k=0}^{n+1} (1-p_k).M^{n+2}_{(p_n),(h_n)} \\
& + \sum_{k=0}^{n} \epsilon(1-\epsilon) p_k \prod_{i=0}^{k-1} (1-p_i).h_k \\
& + (1-\epsilon)\epsilon \prod_{k=0}^{n} (1-p_k).Q^{n+1}_\epsilon + \epsilon.R^n_\epsilon
\end{aligned}$$

We define

$$R^{n+1}_{\epsilon(2-\epsilon)} = \sum_{k=0}^{n} \frac{1-\epsilon}{2-\epsilon} p_k \prod_{i=0}^{k-1} (1-p_i).h_k + \frac{1-\epsilon}{2-\epsilon} \prod_{k=0}^{n} (1-p_k).Q^{n+1}_\epsilon + \frac{1}{2-\epsilon}.R^n_\epsilon$$

which gives the wanted result. □

Given a term M and $C \geq \mathcal{P}_\Downarrow(M)$ we want to build the term M_C using the previous result. We want to find computable sequences (p_n) and (h_n) such that $M =_\mathcal{T} M_{(p_n),(h_n)} +_C \Omega$ in every weakly sensible theory.

Proposition 3.2.2.3. *Let M a term with $0 < \mathcal{P}_\Downarrow(M) < 1$, there are computable* *sequences* (p_n), (h_n) *and* (M^n), *such that* $\lim_{n\to\infty} \mathcal{P}_{\Downarrow}(M^n) = 0$ *and for all* n

$$
M \twoheadrightarrow_{\beta/\equiv}
\begin{array}{ccccccc}
 & & +_{p_0} & & & & \\
 & \diagup & & \diagdown & & & \\
h_0 & & & & +_{p_1} & & \\
 & & & \diagup & & \diagdown & \\
 & & h_1 & & & & +_{p_2} \\
 & & & & & \diagup & \ddots \\
 & & & & h_2 & & & +_{p_n} \\
 & & & & & & \diagup & & \diagdown \\
 & & & & & & h_n & & M^{n+1}
\end{array}
.
$$

Proof. These sequences are given by the following algorithm:

```
Get rid of all the trivial probabilities (0 and 1).
If M contains a head normal value
   rewrite M into h +_p M'
        using commutativity and associativity;
   set p_0 = p, h_0 = h, M^1 = M';
   compute the rest of the sequence from M';
Else apply the algorithm to H(M).
```

where "M contains a head normal value" means M is equal to a head normal value with a non zero probability, i.e. $\mathcal{P}(M = \{\textit{head normal values}\}) > 0$.

First remark that terms without trivial sums $+_0$ and $+_1$ are stable by associativity, commutativity, $\to_\beta$ and $\to_+$. So if we start by erasing all such sums in M using the relation $N +_0 M \equiv_{\text{syn}} M +_1 N \equiv_{\text{syn}} M +_1 M \equiv_{\text{syn}} M$, we never encounter them anymore.

If $\mathcal{P}(M = \{h\}) \neq 0$ with h a head normal value then there is $h' \equiv_{\text{syn}} h$ such that we have either $M = h'$ or $M \equiv_{\text{syn}} h' +_p M'$ modulo associativity and commutativity. But we assumed $\mathcal{P}_{\Downarrow}(M) < 1$ so the first case is impossible. We have necessarily $M \equiv_{\text{syn}} h' +_p M'$ with $\mathcal{P}_{\Downarrow}(M') < 1$. Besides we assumed M reduces into infinitely many head normal values, and necessarily so does M'. In particular $\mathcal{P}_{\Downarrow}(M') > 0$.

Remark also that when applying the algorithm to a term M we always get

$$M \equiv_{\text{syn}} h_0 +_{p_0} (h_1 +_{p_1} \dots (h_n +_{p_n} M^{n+1})\dots)$$

with $\mathcal{P}(M^{n+1} = \{\textit{head normal values}\}) = 0$. Indeed if $M \equiv_{\text{syn}} N$ using only commutativity and associativity then M and N have the same size, so M being of finite size we can not have $M \equiv_{\text{syn}} h_0 +_{p_0} (h_1 +_{p_1} \dots (h_n +_{p_n} M^{n+1})\dots)$ modulo associativity and commutativity for arbitrary large n. Then we apply the algorithm to $\mathrm{H}(M^{n+1})$.

As the h_k are head normal values we have

$$\begin{aligned}\mathrm{H}(M) &\equiv_{\text{syn}} \mathrm{H}(h_0 +_{p_0} (h_1 +_{p_1} \dots (h_n +_{p_n} M^{n+1})\dots)) \\ &= h_0 +_{p_0} (h_1 +_{p_1} \dots (h_n +_{p_n} \mathrm{H}(M^{n+1}))\dots).\end{aligned}$$

Then for all m there is n and a term N^m such that

$$\mathrm{H}^m(M) \equiv_{\mathrm{syn}} h_0 +_{p_0} (h_1 +_{p_1} \dots (h_n +_{p_n} N^m) \dots)$$

with $\mathcal{P}(N^m = \{\mathit{head\ normal\ values}\}) = 0$.

From there we can deduce two things. First the algorithm is productive: indeed for all h such that $\mathcal{P}(M) \twoheadrightarrow_{h^c} \{h\}) \neq 0$ there is m such that $\mathcal{P}(\mathrm{H}^m(M) = \{h\}) \neq 0$, so the algorithm keeps producing p_n, h_n and M^{n+1} provided $\mathcal{P}_{\Downarrow}(M^n) \neq 0$. Secondly we have

$$\mathcal{P}_{\Downarrow}(M) = \sup_m \mathcal{P}(\mathrm{H}^m(M) = \{\mathit{head\ normal\ values}\}) = \sum_{k \in \mathbb{N}} p_k \prod_{i=0}^{k-1} (1 - p_i).$$

Then we have for all $n \in \mathbb{N}$

$$\begin{gathered} M \twoheadrightarrow_{\beta/\equiv} h_0 +_{p_0} (h_1 +_{p_1} \dots (h_n +_{p_n} M^{n+1}) \dots) \\ \mathcal{P}_{\Downarrow}(M) = \sum_{k=0}^{n} p_k \prod_{i=0}^{k-1} (1 - p_i) + \prod_{k=0}^{n} (1 - p_k) \mathcal{P}_{\Downarrow}(M^{n+1}) \end{gathered}$$

This means that $\prod_{k=0}^{n} (1 - p_k) \mathcal{P}_{\Downarrow}(M^{n+1})$ converges to 0. But for all n we have $\prod_{k=0}^{n} (1 - p_k) \geq 1 - \mathcal{P}_{\Downarrow}(M)$ and we assumed $\mathcal{P}_{\Downarrow}(M) < 1$ so necessarily $\mathcal{P}_{\Downarrow}(M^{n+1})$ converges to 0. □

Proposition 3.2.2.4. *Given a sequence of probabilities $(p_n) \in]0; 1[^{\mathbb{N}}$, given a probability $C \in [\sum_{k \in \mathbb{N}} p_k \prod_{i=0}^{k-1} (1 - p_i); 1]$, we define a sequence (q_n) by*

$$q_n = \frac{p_n}{C} \prod_{k=0}^{n-1} \frac{1 - p_k}{1 - q_k}.$$

Then for all $n \in \mathbb{N}$ we have $p_n \leq q_n < 1$.

Proof. Remark that q_n is defined if and only if q_k is defined and different from 1 for all $k < n$.

Let us define $C_0 = C$ and $C_{n+1} = \frac{C_n - p_n}{1 - p_n}$. We will prove that for all n, $q_n = \frac{p_n}{C_n}$, hence the wanted result is $p_n \leq \frac{p_n}{C_n} < 1$, or equivalently $p_n < C_n \leq 1$. So first let us prove by induction on n a more precise result, namely that

$$\sum_{k \geq n} p_k \prod_{i=n}^{k} (1 - p_i) \leq C_n \leq 1.$$

If $n = 0$ this is just the assumption on C. Otherwise we have by induction hypothesis $C_{n+1} = \frac{C_n - p_n}{1 - p_n} \leq \frac{1 - p_n}{1 - p_n} \leq 1$, and

$$C_{n+1} \geq \frac{\sum_{k \geq n} p_k \prod_{i=n}^{k} (1 - p_i) - p_n}{1 - p_n} = \sum_{k \geq n+1} p_k \prod_{i=n+1}^{k} (1 - p_i).$$

Now we prove by induction on n that q_n is well defined, $q_n = \frac{p_n}{C_n}$ and $p_n \leq q_n < 1$. If $n = 0$ then by definition we have $q_0 = \frac{p_0}{C}$ which is well defined, and we have $p_0 < C \leq 1$ so $p_0 \leq q_0 < 1$.

Otherwise we have by induction hypothesis that q_k is well defined and $q_k < 1$ for $k \leq n$ so q_{n+1} is well defined. We can write $q_{n+1} = q_n \frac{p_{n+1}}{p_n} \frac{1-p_n}{1-q_n}$, and by induction hypothesis $q_n = \frac{p_n}{C_n}$ and $q_{n+1} = \frac{p_{n+1}}{C_n} \frac{1-p_n}{1-\frac{p_n}{C_n}} = p_{n+1} \frac{1-p_n}{C_n - p_n} = \frac{p_{n+1}}{C_{n+1}}$. Then from $p_{n+1} < C_{n+1} \leq 1$ we get $p_{n+1} \leq q_{n+1} < 1$. $\square$

We now have all the necessary tools to prove our result.

Proposition 3.2.2.5. *For all M and all computable $C \geq \mathcal{P}_{\Downarrow}(M)$ there is a term M_C such that in all weakly sensible theories*

$$M =_{\mathcal{T}} M_C +_C \Omega.$$

Proof. If $\mathcal{P}_{\Downarrow}(M) = 0$ we have $M =_{\mathcal{T}} \Omega$ in every sensible theory. If $\mathcal{P}_{\Downarrow}(M) = 1$ then necessarily $C = 1$ and $M \equiv_{\text{syn}} M +_1 \Omega$. If there are finitely many head normal values h modulo $\equiv_{\text{syn}}$ such that $\mathcal{P}(M \twoheadrightarrow_{h^c} \{h\}) \neq 0$ then we define M_C such that $M_C \equiv_{\text{syn}} \sum_h \frac{\mathcal{P}(M \twoheadrightarrow_{h^c} \{h\})}{\mathcal{P}_{\Downarrow}(M)}.h$, and we can prove that this term works by using techniques similar to those of the general case.

Otherwise we have computable sequences (p_n) of non trivial probabilities, (h_n) of head normal values and (M^n) of terms such that

$$M \twoheadrightarrow_{\beta/\equiv} h_0 +_{p_0} (h_1 +_{p_1} ... (h_n +_{p_n} M^{n+1})...)$$

for all n and $\mathcal{P}_{\Downarrow}(M^n)$ converges to 0. Then we define $q_n = \frac{p_n}{C} \prod_{k=0}^{n-1} \frac{1-p_k}{1-q_k^C}$, this is a computable sequence and we know that the q_n's are probabilities. We finally define $M_C = M_{(q_n),(h_n)}$, for all n and $\epsilon > 0$ we have

$$M_C \twoheadrightarrow_{\beta/\equiv} \Big(h_0 +_{q_0} (h_1 +_{q_1} ... (h_n +_{q_n} M_C^{n+1})...)\Big) +_{1-\epsilon} R_\epsilon^n.$$

For all n and $\epsilon > 0$ we have

$$\begin{aligned} M_C +_C \Omega \twoheadrightarrow_{\beta/\equiv} & \sum_{k=1}^{n} (1-\epsilon) \left(C q_k \prod_{i=1}^{k-1} (1-q_i) \right) .h_k \\ & + (1-\epsilon) C \prod_{k=1}^{n} (1-q_k).M_C^{n+1} + C\epsilon.R_\epsilon^n + (1-C).\Omega \\ M \twoheadrightarrow_{\beta/\equiv} & \sum_{k=1}^{n} (1-\epsilon) \left(p_k \prod_{i=1}^{k-1} (1-p_i) \right) .h_k \\ & + (1-\epsilon) \prod_{k=1}^{n} (1-p_k).M^{n+1} + \epsilon.M \end{aligned}$$

Using the definition of the sequence (q_n) we see that for all $k \in \mathbb{N}$ we have $C q_k \prod_{i=1}^{k-1} (1-q_i) = p_k \prod_{i=1}^{k-1} (1-p_i)$. Then to prove that for all $\epsilon > 0$ we have

$M \twoheadrightarrow_{\beta/\equiv} \cdot =_{\Omega,\epsilon} \cdot \twoheadleftarrow_{\beta/\equiv} M_C +_C \Omega$ we have to prove that for all $\epsilon > 0$ there are n and $\delta > 0$ such that

$$\mathcal{P}_\Downarrow\left(\frac{(1-\delta)C\prod_{k=1}^n(1-q_k)}{1-c_{\delta,n}}.M_C^{n+1} + \frac{C\delta}{1-c_{\delta,n}}.R_\delta^n + \frac{1-C}{1-c_{\delta,n}}.\Omega\right) \leq \epsilon$$
$$\mathcal{P}_\Downarrow\left(\frac{(1-\delta)\prod_{k=1}^n(1-p_k)}{1-c_{\delta,n}}.M^{n+1} + \frac{\delta}{1-c_{\delta,n}}.M\right) \leq \epsilon$$

where $c_{\delta,n} = (1-\delta)\sum_{k=1}^n p_k\prod_{i=1}^{k-1}(1-p_i)$ and we have $c_{\delta,n} < \mathcal{P}_\Downarrow(M) < 1$.

As $1-c_{\delta,n}$ has a lower bound we just need to prove that $\mathcal{P}_\Downarrow(M_C +_C \Omega) = \sum_{k=1}^\infty Cq_k\prod_{i=1}^{k-1}(1-q_i)$ and $\mathcal{P}_\Downarrow(M) = \sum_{k=1}^\infty p_k\prod_{i=1}^{k-1}(1-p_i)$. Indeed what we want to prove is equivalent to $\forall\epsilon > 0, \exists\delta > 0, \exists n$:

$$(1-c_{\delta,n})\mathcal{P}_\Downarrow\left(\frac{(1-\delta)C\prod_{k=1}^n(1-q_k)}{1-c_{\delta,n}}.M_C^{n+1} + \frac{C\delta}{1-c_{\delta,n}}.R_\delta^n + \frac{1-C}{1-c_{\delta,n}}.\Omega\right) \leq \epsilon$$
$$(1-c_{\delta,n})\mathcal{P}_\Downarrow\left(\frac{(1-\delta)\prod_{k=1}^n(1-p_k)}{1-c_{\delta,n}}.M^{n+1} + \frac{\delta}{1-c_{\delta,n}}.M\right) \leq \epsilon.$$

But this means precisely that $\forall\epsilon > 0, \exists\delta > 0, \exists n$:

$$\mathcal{P}_\Downarrow(M_C +_C \Omega) - c_{\delta,n}\mathcal{P}_\Downarrow\left(\sum_{k=1}^n \frac{(1-\delta)\left(Cq_k\prod_{i=1}^{k-1}(1-q_i)\right)}{c_{\delta,n}}.h_k\right) \leq \epsilon$$
$$\mathcal{P}_\Downarrow(M) - c_{\delta,n}\mathcal{P}_\Downarrow\left(\sum_{k=1}^n \frac{(1-\delta)\left(p_k\prod_{i=1}^{k-1}(1-p_i)\right)}{c_{\delta,n}}.h_k\right) \leq \epsilon$$

i.e. $\forall\epsilon > 0, \exists\delta > 0, \exists n$:

$$\mathcal{P}_\Downarrow(M_C +_C \Omega) - \sum_{k=1}^n (1-\delta)\left(Cq_k\prod_{i=1}^{k-1}(1-q_i)\right) \leq \epsilon$$
$$\mathcal{P}_\Downarrow(M) - \sum_{k=1}^n (1-\delta)\left(p_k\prod_{i=1}^{k-1}(1-p_i)\right) \leq \epsilon$$

We know that $\mathcal{P}_\Downarrow(M) - \sum_{k=1}^\infty p_k\prod_{i=1}^{k-1}(1-p_i)$, and we can check that $\mathcal{P}_\Downarrow(M_C) = \sum_{k=1}^\infty q_k\prod_{i=1}^{k-1}(1-q_i)$. $\square$

The weak sensibility is defined in a quite natural way, as a conjunction of the notions of equality of unsolvable terms and approximations up to ϵ. And this natural definition gives a satisfying result, as given any term we can isolate its diverging part (up to some $\epsilon > 0$) in every weakly sensible theory.

But we call it weak as it is the opposite of the continuity: it deals with the non termination but it can not be used to eliminate infinite branches with probability 0. For instance there is no obvious way to prove that in our first example the terms N_0 and N_0' are equal in every weakly sensible theory.

The reason is that we consider the local convergence probability of the subterms, not their contribution to the convergence probability of the whole term. We can define a strong sensibility which takes this into account.

Definition 3.2.2.5. For $\epsilon, \epsilon' \geq 0$ we define $=^s_{\Omega,\epsilon,\epsilon'}$ by:

$$\frac{\mathcal{P}_{\Downarrow}(M) \leq \epsilon' \qquad \mathcal{P}_{\Downarrow}(N) \leq \epsilon'}{M =^s_{\Omega,\epsilon,\epsilon'} N} \qquad \frac{}{x =^s_{\Omega,\epsilon,\epsilon'} x} \qquad \frac{M =^s_{\Omega,\epsilon,\epsilon'} N}{\lambda x.M =^s_{\Omega,\epsilon,\epsilon'} \lambda x.N}$$

$$\frac{M =^s_{\Omega,\epsilon,\epsilon'} N \qquad M' =^s_{\Omega,\epsilon,\epsilon} N'}{M\ M' =^s_{\Omega,\epsilon,\epsilon'} N\ N'} \qquad \frac{M_1 =^s_{\Omega,\epsilon,\epsilon_1} N_1 \qquad M_2 =^s_{\Omega,\epsilon,\epsilon_2} N_2}{M_1 +_p M_2 =^s_{\Omega,\epsilon,p\epsilon_1+(1-p)\epsilon_2} N_1 +_p N_2}$$

We note $M =^s_{\Omega,\epsilon} N$ when $M =^s_{\Omega,\epsilon,\epsilon} N$.

The intuition is that $M =^s_{\Omega,\epsilon,\epsilon} N$ whenever$N$ is obtained by changing a subterm whose contribution to the convergence of M is less than ϵ in M, and then doing so inductively on the arguments in M. The meaning of this relation may be clearer if we consider it on canonical terms.

Definition 3.2.2.6. For $\epsilon, \epsilon' \geq 0$ we define $=^{s,c}_{\Omega,\epsilon,\epsilon'}$ on canonical terms and $=^{s,v}_{\Omega,\epsilon}$ on values by:

$$\frac{\mathcal{P}_{\Downarrow}(M) \leq \epsilon' \qquad \mathcal{P}_{\Downarrow}(N) \leq \epsilon'}{M =^{s,c}_{\Omega,\epsilon,\epsilon'} N}$$

$$\frac{M_1 =^{s,c}_{\Omega,\epsilon,\epsilon_1} N_1 \qquad M_2 =^{s,c}_{\Omega,\epsilon,\epsilon_2} N_2}{M_1 +_p M_2 =^{s,c}_{\Omega,\epsilon,p\epsilon_1+(1-p)\epsilon_2} N_1 +_p N_2} \qquad \frac{v =^{s,v}_{\Omega,\epsilon} w}{v =^{s,c}_{\Omega,\epsilon,\epsilon'} w}$$

$$\frac{}{x =^{s,v}_{\Omega,\epsilon} x} \qquad \frac{v =^{s,v}_{\Omega,\epsilon} w}{\lambda x.v =^{s,v}_{\Omega,\epsilon} \lambda x.w} \qquad \frac{v =^{s,v}_{\Omega,\epsilon} w \qquad M =^{s,c}_{\Omega,\epsilon,\epsilon} N}{v\ M =^{s,v}_{\Omega,\epsilon} w\ N}$$

We note $M =^{s,c}_{\Omega,\epsilon} N$ when $M =^{s,c}_{\Omega,\epsilon,\epsilon} N$.

Proposition 3.2.2.6. *If M and N are canonical then $M =^s_{\Omega,\epsilon,\epsilon'} N$ if and only if $M =^{s,c}_{\Omega,\epsilon,\epsilon'} N$.*

Proof. It is straightforward to prove by induction on $=^{s,c}_{\Omega,\epsilon,\epsilon'}$ that if $M =^{s,c}_{\Omega,\epsilon,\epsilon'} N$ then $M =^s_{\Omega,\epsilon,\epsilon'} N$ and if $v =^{s,v}_{\Omega,\epsilon} w$ then for all ϵ', $v =^s_{\Omega,\epsilon,\epsilon'} w$.

Conversely we can prove by an easy induction on $=^s_{\Omega,\epsilon,\epsilon'}$ that if $M =^s_{\Omega,\epsilon,\epsilon'} N$ with M canonical then $M =^{s,c}_{\Omega,\epsilon,\epsilon'} N$ and if $v =^s_{\Omega,\epsilon,\epsilon'} w$ then $v =^{s,v}_{\Omega,\epsilon} w$. ☐

Proposition 3.2.2.7. *If $M =^s_{\Omega,\epsilon,\epsilon'} N$ then* $\mathrm{can}(M) =^{s,c}_{\Omega,\epsilon,\epsilon'} \mathrm{can}(N)$.

Proof. We just need to prove that if $M =^s_{\Omega,\epsilon,\epsilon'} N$ and $M \to_+ M'$ then there is N' such that $N \to^?_+ N'$ and $M' =^s_{\Omega,\epsilon,\epsilon'} N'$. Then if $M =^s_{\Omega,\epsilon,\epsilon'} N$ there is N' such that $N \twoheadrightarrow_+ N'$ and $\mathrm{can}(M) =^s_{\Omega,\epsilon,\epsilon'} N'$, by symmetry we get $\mathrm{can}(M) =^s_{\Omega,\epsilon,\epsilon'} \mathrm{can}(N)$, and we can use the previous result to get $\mathrm{can}(M) =^{s,c}_{\Omega,\epsilon,\epsilon'} \mathrm{can}(N)$.

We prove this by induction on $M =^s_{\Omega,\epsilon,\epsilon'} N$ and the context of the reduction $M \to_+ M'$.

- If $\mathcal{P}_{\Downarrow}(M), \mathcal{P}_{\Downarrow}(N) \leq \epsilon'$ then we know that the convergence probability is stable by $\to_+$ so $\mathcal{P}_{\Downarrow}(M') = \mathcal{P}_{\Downarrow}(M)$.

- Otherwise if the context of the reduction is not empty then the result is immediate by induction hypothesis.

- If $M = \lambda x.(M_1 +_p M_2) \to_+ \lambda x.M_1 +_p \lambda x.M_2 = M'$ and we do not have $\mathcal{P}_{\Downarrow}(M), \mathcal{P}_{\Downarrow}(N) \leq \epsilon'$ then necessarily $N = \lambda x.Q$ and $M_1 +_p M_2 =^{s}_{\Omega,\epsilon,\epsilon'} Q$. We have $\mathcal{P}_{\Downarrow}(M_1 +_p M_2) = \mathcal{P}_{\Downarrow}(M)$ and $\mathcal{P}_{\Downarrow}(Q) = \mathcal{P}_{\Downarrow}(N)$ so these are not less than ϵ, and necessarily $Q = N_1 +_p N_2$ with $M_i =^{s}_{\Omega,\epsilon,\epsilon_i} N_i$ for $i \in \{1;2\}$ with $p\epsilon_1 + (1-p)\epsilon_2 = \epsilon$. From there we have $\lambda x.M_i =^{s}_{\Omega,\epsilon,\epsilon_i} \lambda x.N_i$ for $i \in \{1;2\}$, and $M' =^{s}_{\Omega,\epsilon,\epsilon'} \lambda x.N_1 +_p \lambda x.N_2$.

- Similarly if $M = (M_1 +_p M_2)\ P \to_+ M_1\ P +_p M_2\ P = M'$ then $N = Q\ P$ with $\mathcal{P}_{\Downarrow}(M_1 +_p M_2) \leq \mathcal{P}_{\Downarrow}(M)$ and $\mathcal{P}_{\Downarrow}(Q) \leq \mathcal{P}_{\Downarrow}(N)$ so $Q = N_1 +_p N_2$, and $M' =^{s}_{\Omega,\epsilon,\epsilon'} N_1\ P +_p N_2\ P$.

$\square$

The interesting relation on canonical terms is $=^{s,c}_{\Omega,\epsilon}$. If we consider canonical terms modulo $\equiv_{\mathrm{syn}}$ we have

$$\sum_i p_i.v_i + \left(1 - \sum_i p_i\right).M =^{s,c}_{\Omega,\epsilon} \sum_i p_i.w_i + \left(1 - \sum_i p_i\right).N$$

whenever

- $v_i =^{s,v}_{\Omega,\epsilon} w_i$ for all i,
- $(1 - \sum_i p_i)\,\mathcal{P}_{\Downarrow}(M) \leq \epsilon$
- and $(1 - \sum_i p_i)\,\mathcal{P}_{\Downarrow}(N) \leq \epsilon$.

Then if we go back to our previous examples, for all $\epsilon > 0$ we have both $\mathrm{N}_0 \twoheadrightarrow_{\beta/\equiv} \cdot =^{s,c}_{\Omega,\epsilon} \cdot \twoheadleftarrow_{\beta/\equiv} \mathrm{N}_0'$ and $\mathrm{N}_1^{\mathrm{P}} \twoheadrightarrow_{\beta/\equiv} \cdot =^{s,c}_{\Omega,\epsilon} \cdot \twoheadleftarrow_{\beta/\equiv} \mathrm{N'}_1^{\mathrm{P}}$.

Definition 3.2.2.7. A probabilistic theory $=_{\mathcal{T}}$ is *strongly sensible* if for all M and N,

$$\text{if } \forall\epsilon > 0, M \twoheadrightarrow_{\beta/\equiv} \cdot =^{s,c}_{\Omega,\epsilon} \cdot \twoheadleftarrow_{\beta/\equiv} N \text{ then } M =_{\mathcal{T}} N.$$

Proposition 3.2.2.8. *Every strongly sensible theory is weakly sensible.*

Proof. We can prove that if $M =_{\Omega,\epsilon} N$ then $M =^{s}_{\Omega,\epsilon,\epsilon} N$. The only thing we need to remark is that if $M_i =^{s}_{\Omega,\epsilon,\epsilon} N_i$ for $i \in \{1;2\}$ then $M_1 +_p M_2 =^{s}_{\Omega,\epsilon,\epsilon} N_1 +_p N_2$. $\square$

In the following we will not really be interested in the strong sensibility itself. We will consider continuous weakly sensible theories, and we conjecture that such theories are always strongly sensible. So we will simply speak about continuous sensible theories, without specifying that we consider the weak sensibility.

The main reason we mentioned the strong sensibility is that the relations $=^{s}_{\Omega,\epsilon,\epsilon'}$ and $=^{s,c}_{\Omega,\epsilon}$ will be useful to characterize the probabilistic Böhm tree equality.

4 Probabilistic Böhm trees

The Böhm tree of a deterministic term is a generalization of the notion of normal form, obtained by iterating head normalization. Deterministic normal forms are characterized inductively as follows: a term is in normal form if it is a head normal form $\lambda x_1 ... x_n.y\ P_1\ ...\ P_m$ where P_i is normal for $i \leq m$. Böhm trees are obtained by adding a symbol Ω for unsolvable terms, and by extending this definition to infinite forms.

Definition 4.0.2.1. The sets $\mathcal{BT}_d$ for $d \in \mathbb{N}$ of *Böhm trees of depth* d are defined by

$$\mathcal{BT}_0 = \{\Omega\}$$

$$\mathcal{BT}_{d+1} = \left\{ \begin{array}{ccc} & \lambda x_1 ... x_n.y & \\ \swarrow & \downarrow & \searrow \\ T_1 & ... & T_m \end{array} \mid \forall i \leq m, T_i \in \mathcal{BT}_d \right\} \cup \{\Omega\}.$$

In the future we will most often write trees as terms:

$$\mathcal{BT}_{d+1} = \{\lambda x_1 ... x_n.y\ T_1\ ...\ T_m \mid \forall i \leq m, T_i \in \mathcal{BT}_d\} \cup \{\Omega\}.$$

Definition 4.0.2.2. The Böhm tree $BT_d(M)$ of depth $d \in \mathbb{N}$ of a deterministic term M is given by:

$$BT_0(M) = \Omega$$

$$BT_{d+1}(M) = \begin{cases} \lambda x_1 ... x_n.y\ BT_d(P_1)\ ...\ BT_d(P_m) & \text{if } M \twoheadrightarrow_h \lambda x_1 ... x_n.y\ P_1\ ...\ P_m \\ \Omega & \text{otherwise} \end{cases}$$

What we call the Böhm tree of a term M is a possibly infinite tree obtained as the limit of its finite approximations $BT_d(M)$. Such an infinite tree can be defined directly, but here we will simply say that two terms have the same Böhm tree if they have the same finite Böhm trees.

Definition 4.0.2.3. The Böhm tree equality $=_\mathcal{B}$ is defined by

$$M =_\mathcal{B} N \text{ iff } \forall d \in \mathbb{N}, BT_d(M) = BT_d(N).$$

This notion of tree is actually not the one we are interested in. It is fairly simple to see that $=_\mathcal{B}$ does not correspond to the observational equivalence: the terms

x and $\lambda y.x\ y$ have different Böhm trees but they are observationally equivalent. We need to introduce some notion of extensionality in our Böhm tree equality.

As Böhm trees are infinite there are several ways to quotient them by extensionality. We can simply consider the equivalence induced by the rule

$$\lambda x_1...x_n z.y\ T_1\ ...\ T_m\ z = \lambda x_1...x_n.y\ T_1\ ...\ T_m \text{ if } z \neq y \text{ and } z \text{ is not free in the } T_i\text{'s.}$$

But this is not enough as this would mean that two trees are equivalent if we can rewrite one into the other using this rule finitely many times.

Let (x_n) a sequence of pairwise distinct variables, we can define terms X_n for $n \in \mathbb{N}$ such that for all n, $X_n =_\beta \lambda x_{n+1}.x_n\ X_{n+1}$. Then we can prove that X_n is observationally equivalent to x_n for all $n \in \mathbb{N}$. Yet the Böhm trees of the terms X_n have no η-redex. if we consider that the tree of X_{n+1} is equivalent to the tree x_{n+1} then we can perform an η-contraction, but in some sense the first η redex is pushed at an infinite depth.

So instead of considering the η-contraction we rather perform an infinite η-expansion. If $M \twoheadrightarrow_h \lambda x_1...x_n.y\ P_1\ ...\ P_m$ then we will consider

$$BT^\eta_{d+1}(M) = \lambda x_1...x_n x_{n+1}...\ .y\ BT^\eta_d(P_1)\ ...\ BT^\eta_d(P_m)\ BT^\eta_d(x_{n+1})\ ...$$

To make this definition a bit lighter we will consider given a family of pairwise distinct variables $(x_{d,n})_{d,n\in\mathbb{N}}$ (such that $\mathit{Var} \setminus \{x_{d,n} \mid d,n \in \mathbb{N}\}$ is still an infinite set). Then for any $d \in \mathbb{N}$ and for any solvable term M we have a reduction $M \twoheadrightarrow_h \lambda x_{d,1}...x_{d,n}.y\ P_1\ ...\ P_m$, and the infinitely extensional Böhm tree of M at depth $d+1$ is then given by the head variable y and the sequence $(BT^\eta_d(P_1), ..., BT^\eta_d(P_m), BT^\eta_d(x_{n+1}), ...)$.

Definition 4.0.2.4. The sets $\mathcal{BT}^\eta_d$ for $d \in \mathbb{N}$ of *infinitely extensional Böhm trees of depth* d, or *Nakajima trees of depth* d, are defined by

$$\begin{aligned} \mathcal{BT}^\eta_0 &= \{\Omega\} \\ \mathcal{BT}^\eta_{d+1} &= \{(y,(T_n)) \mid \forall n \in \mathbb{N}, T_n \in \mathcal{BT}^\eta_d\} \cup \{\Omega\}. \end{aligned}$$

Definition 4.0.2.5. The infinitely extensional Böhm tree $BT^\eta_d(M)$ of depth $d \in \mathbb{N}$ of a deterministic term M is given by:

$$BT^\eta_0(M) = \Omega$$

$$BT^\eta_{d+1}(M) = \begin{cases} (y,(BT^\eta_d(P_1),..,BT^\eta_d(P_m),BT^\eta_d(x_{d,n+1}),..)) & \text{if } M \twoheadrightarrow_h \lambda x_1...x_n.y\ P_1\ ...\ P_m \\ \Omega & \text{otherwise} \end{cases}$$

Remark that if $BT^\eta_{d+1}(M) = (y,(T_n))$ then there exists $m \in \mathbb{N}$ and $s \in \mathbb{Z}$ such that for all $n > m$ we have $T_n = BT^\eta_d(x_{d,n-s})$. The trees which do not have this property are not useful to describe terms, so they are irrelevant. Thus a possible alternate definition for infinitely extensional Böhm trees would be

$$\begin{aligned}\mathcal{BT}^{\eta}_0 &= \{\Omega\}\\ \mathcal{BT}^{\eta}_{d+1} &= \{(y,(T_n)) \mid \exists m \in \mathbb{N}, \exists s \in \mathbb{Z} : \forall n > m, T_n = BT^{\eta}_d(x_{d,n-s})\} \cup \{\Omega\}.\end{aligned}$$

This restriction will only come useful in the next chapter to obtain a separability result. We will then define classes of trees based on their head variable y and on the integer s, and the trees for which this property does not hold will not belong to any class.

Definition 4.0.2.6. The infinitely extensional Böhm tree equality $=_{\mathcal{B}^\eta}$ is defined by

$$M =_{\mathcal{B}^\eta} N \text{ iff } \forall d \in \mathbb{N}, BT^{\eta}_d(M) = BT^{\eta}_d(N).$$

It is known that this equality corresponds to the deterministic observational equivalence.

We want to generalize these definitions to probabilistic terms. Probabilistic normal forms modulo $\equiv$ are finite probability distributions over normal values, and normal values are head normal values $\lambda x_1 ... x_n . y\ P_1\ ...\ P_m$ where P_i is normal for $i \leq m$. Thus it is natural to build probabilistic Böhm trees with two layers: one to represent (potentially infinite) probability distributions and one to represent head normal values.

Definition 4.0.2.7. The sets $\mathcal{PT}_d$ of *probabilistic Böhm trees of depth* d and $\mathcal{VT}_d$ of *probabilistic value Böhm trees of depth* d for $d \in \mathbb{N}$ are defined by

$$\begin{aligned}\mathcal{VT}_0 &= \emptyset\\ \mathcal{VT}_{d+1} &= \{\lambda x_1 ... x_n . y\ T_1\ ...\ T_m \mid \forall i \leq m, T_i \in \mathcal{PT}_d\}\\ \mathcal{PT}_d &= \left\{T : \mathcal{VT}_d \to [0;1] \mid \sum_{t \in \mathcal{VT}_d} T(t) \leq 1\right\}.\end{aligned}$$

Definition 4.0.2.8. The probabilistic Böhm tree $PT_d(M)$ of depth $d \in \mathbb{N}$ of a term M and the probabilistic value Böhm tree $VT_d(h)$ of depth $d \in \mathbb{N}^*$ of a head normal value h are given by:

$$\begin{aligned}VT_{d+1}(\lambda x_1 ... x_n . y\ P_1\ ...\ P_m) = &\quad \lambda x_1 ... x_n . y\ PT_d(P_1)\ ...\ PT_d(P_m)\\ PT_d(M) : &\quad t \mapsto \mathcal{P}(M \twoheadrightarrow_{h^c} \{h \mid VT_d(h) = t\}).\end{aligned}$$

To simplify the notations, given a relation $\mathcal{R}$ and $t \in \mathcal{VT}_d$ we write $\mathcal{P}(M\ \mathcal{R}\ t)$ for $\mathcal{P}(M\ \mathcal{R}\ \{h \mid VT_d(h) = t\})$.

Remark that when defining the Böhm tree of a probabilistic term in such a way it is not convenient to add a label Ω for unsolvable terms. Indeed we can not define $VT_d(v) = \Omega$ when v is unsolvable, as we do not have a satisfying notion of unsolvability. What we do is to associate to every term a subprobability distribution over trees of head normal forms, and the probability of Ω is intuitively given by $PT_d(M)(\Omega) = 1 - \sum_{t \in \mathcal{VT}_d} PT_d(M)(t)$.

Also remark that we defined the Böhm tree at depth 0 of a term M as $PT_0(M) : t \mapsto \mathcal{P}(M =_{\beta+} \{h \mid VT_0(h) = t\})$, even though $VT_0(h)$ is not defined, but one the other hand $\mathcal{VT}_0 = \emptyset$ so $PT_0(M)$ is the unique function from $\emptyset$ to $[0;1]$.

We need to prove that this definition is correct, i.e. that for any term M and any $d \in \mathbb{N}$ we have $\sum_{t \in \mathcal{VT}_d} PT_d(M)(t) \leq 1$.

Lemma 4.0.2.1. *Given two head normal values h and h', if $h =_{\beta+} h'$ then $h = \lambda x_1 ... x_n.y\ P_1\ ...\ P_m$ and $h' = \lambda x_1 ... x_n.y\ P'_1\ ...\ P'_m$ with $P_i =_{\beta+} P'_i$ for $i \leq m$.*

Proof. If $h = \lambda x_1 ... x_n.y\ P_1\ ...\ P_m \twoheadrightarrow_{\text{split}} \cdot \twoheadrightarrow_{\beta^c} M$ then $M \equiv_{\text{syn}} \sum_j p_j.\lambda x_1 ... x_n.y P_{j,1}\ ...\ P_{j,m}$ with $P_i \twoheadrightarrow_{\text{split}} \cdot \twoheadrightarrow_{\beta^c} P_{j,i}$ for all j and $i \leq m$.

Besides if $\sum_j p_j.\lambda x_1 ... x_n.y P_{j,1}\ ...\ P_{j,m} \equiv_{\text{syn}} \sum_j p'_j.\lambda x_1 ... x_{n'}.y' P'_{j,1}\ ...\ P'_{j,m'}$ then necessarily $n = n'$, $y = y'$ and $m = m'$, as the syntactic equivalence can not change the structure of head normal forms. Moreover we can prove by induction on $\equiv_{\text{syn}}$ that for all $i \leq m$ and all j there is j' such that $P_{i,j} \equiv_{\text{syn}} P_{i,j'}$.

Thus if $h = \lambda x_1 ... x_n.y\ P_1\ ...\ P_m$, $h' = \lambda x_1 ... x_{n'}.y'\ P'_1\ ...\ P'_{m'}$ and $h =_{\beta+} h'$, i.e. $h \twoheadrightarrow_{\text{split}} \cdot \twoheadrightarrow_{\beta^c} \cdot \equiv_{\text{syn}} \cdot \twoheadleftarrow_{\beta^c} \cdot \twoheadleftarrow_{\text{split}} h'$, we have $n = n'$, $y = y'$, $m = m'$ and $P_i =_{\beta+} P'_i$ for all $i \leq m$. □

Proposition 4.0.2.2. *For all $d \in \mathbb{N}$:*

1. *if $d > 0$ and $h =_{\beta+} h'$ then $VT_d(h) = VT_d(h')$;*
2. *for $t \in \mathcal{VT}_d$ we have for all M*
$$\mathcal{P}(M \twoheadrightarrow_{h^c} t) = \mathcal{P}(M \twoheadrightarrow_H t) = \mathcal{P}(M =_{\beta+} t) = \mathcal{P}(M \twoheadrightarrow_S t);$$
3. *if $M =_{\beta+} M'$ then $PT_d(M) = PT_d(M')$;*
4. *if $d > 0$, for any term M we have $\sum_{t \in \mathcal{VT}_d} PT_d(M)(t) = \mathcal{P}_{\Downarrow}(M)$.*

In particular for all d and M we have $PT_d(M) \in \mathcal{PT}_d$.

Proof. We reason by induction on d. If $d = 0$ then $\mathcal{VT}_0 = \emptyset$ and there is a unique Böhm tree of depth 0.

Otherwise we have:

- if $h =_{\beta+} h'$ then $h = \lambda x_1 ... x_n.y\ P_1\ ...\ P_m$ and $h' = \lambda x_1 ... x_n.y\ P'_1\ ...\ P'_m$ with $P_i =_{\beta+} P'_i$ for $i \leq m$ so by induction hypothesis $PT_d(P_i) = PT_d(P'_i)$ for $i \leq m$, hence $VT_{d+1}(h) = VT_{d+1}(h')$;
- for $t \in \mathcal{VT}_{d+1}$ the set $\{h \mid VT_{d+1}(h) = t\}$ is a set of head normal values, and by induction hypothesis it is closed by β so we can apply the proposition 3.1.2.1;
- we just proved $PT_{d+1}(M)(t) = \mathcal{P}(M =_{\beta+} t)$ for $t \in \mathcal{VT}_{d+1}$ so if $M =_{\beta+} M'$ we immediately have $PT_{d+1}(M) = PT_{d+1}(M')$;

- we also proved that $PT_{d+1}(M)(t) = \mathcal{P}(M \twoheadrightarrow_H t)$, from which we get that $PT_{d+1}(M)(t) = \sup_{n\in\mathbb{N}} \mathcal{P}(\mathrm{H}^n(M) = t)$, and the sequences $(\mathcal{P}(\mathrm{H}^n(M) = t))_n$ are increasing so

$$\begin{aligned}\sum_{t\in\mathcal{VT}_{d+1}} \sup_{n\in\mathbb{N}} \mathcal{P}(\mathrm{H}^n(M) = t) &= \sup_{n\in\mathbb{N}} \sum_{t\in\mathcal{VT}_{d+1}} \mathcal{P}(\mathrm{H}^n(M) = t) \\ &= \sup_n \mathcal{P}(\mathrm{H}^n(M) = \{h \text{ head normal value}\}) \\ &= \mathcal{P}_{\Downarrow}(M).\end{aligned}$$

□

Definition 4.0.2.9. The probabilistic Böhm tree equality $=_{\mathcal{PB}}$ is defined by

$$M =_{\mathcal{PB}} N \text{ iff } \forall d \in \mathbb{N}, PT_d(M) = PT_d(N).$$

We have a definition for probabilistic Böhm trees, that we can easily extend to infinitely extensional probabilistic Böhm trees.

Definition 4.0.2.10. The sets $\mathcal{PT}_d^\eta$ of *infinitely extensional probabilistic Böhm trees of depth d* and $\mathcal{VT}_d^\eta$ of *infinitely extensional probabilistic value Böhm trees of depth d* for $d \in \mathbb{N}$ are defined by

$$\begin{aligned}\mathcal{VT}_0^\eta &= \emptyset \\ \mathcal{VT}_{d+1}^\eta &= \{(y, (T_n)) \mid \forall n \in \mathbb{N}, T_n \in \mathcal{PT}_d^\eta\} \\ \mathcal{PT}_d^\eta &= \left\{ T : \mathcal{VT}_d^\eta \to [0;1] \mid \sum_{t\in\mathcal{VT}_d^\eta} T(t) \leq 1 \right\}.\end{aligned}$$

Definition 4.0.2.11. The infinitely extensional probabilistic Böhm tree $PT_d^\eta(M)$ of depth $d \in \mathbb{N}$ of a term M and the infinitely extensional probabilistic value Böhm tree $VT_d^\eta(h)$ of depth $d \in \mathbb{N}^*$ of a head normal value h are given by:

$$\begin{aligned}VT_{d+1}^\eta(\lambda x_1 ... x_n.y\ P_1\ ...\ P_m) = &\quad (y, (PT_d^\eta(P_1), ..., PT_d^\eta(P_m), PT_d^\eta(x_{d,n+1}), ...)) \\ PT_d^\eta(M) : &\quad t \mapsto \mathcal{P}(M \twoheadrightarrow_{h^c} \{h \mid VT_d^\eta(h) = t\}).\end{aligned}$$

Proposition 4.0.2.3. *For all $d \in \mathbb{N}$:*

1. *if $d > 0$ and $h =_{\beta+} h'$ then $VT_d^\eta(h) = VT_d^\eta(h')$;*
2. *for all $t \in \mathcal{VT}_d^\eta$ and all M we have*

$$\mathcal{P}(M \twoheadrightarrow_{h^c} t) = \mathcal{P}(M \twoheadrightarrow_H t) = \mathcal{P}(M =_{\beta+} t) = \mathcal{P}(M \twoheadrightarrow_S t);$$

3. *if $M =_{\beta+} M'$ then $PT_d^\eta(M) = PT_d^\eta(M')$;*
4. *if $d > 0$, for any term M we have $\sum_{t\in\mathcal{VT}_d^\eta} PT_d^\eta(M)(t) = \mathcal{P}_{\Downarrow}(M)$.*

In particular for all d and M we have $PT^\eta_d(M) \in \mathcal{PT}^\eta_d$.

Proof. This is the same proof as for the non extensional trees. $\square$

Definition 4.0.2.12. The infinitely extensional Böhm tree equality $=_{\mathcal{PB}^\eta}$ is defined by

$$M =_{\mathcal{PB}^\eta} N \text{ iff } \forall d \in \mathbb{N}, PT^\eta_d(M) = PT^\eta_d(N).$$

We want to prove that $=_{\mathcal{PB}^\eta}$ corresponds to the probabilistic observational equivalence. The first, and most technical, step is to prove that $=_{\mathcal{PB}^\eta}$ is actually a theory. It is obviously an equivalence, and we proved that it is stable by $=_{\beta+}$, so what we need to prove is that it is contextual. Then we will get $=_{\mathcal{PB}^\eta} \subset =_{obs}$: indeed if for all context C we have $C[M] =_{\mathcal{PB}^\eta} C[N]$ then for all C we have

$$\mathcal{P}_\Downarrow(C[M]) = \sum_{t \in \mathcal{VT}^\eta_1} PT^\eta_1(C[M])(t) = \sum_{t \in \mathcal{VT}^\eta_1} PT^\eta_1(C[N])(t) = \mathcal{P}_\Downarrow(C[N]).$$

4.1 Finite approximations

Böhm trees are infinite objects, and thus are not easy to manipulate. It is more convenient to work on finite approximations of these trees. We actually did not directly define Böhm trees but only their approximations at finite depths.

At the end of the section 2.2.2 we claimed that the Böhm tree of a term M is obtained as the limit of its complete left reductions $\mathrm{L}^k(M)$ for $k \in \mathbb{N}$. To formalize this idea, we will associate to every term a (finite) tree, then show that the Böhm tree of a term is indeed the limit of the finite trees associated to its complete left reductions.

For deterministic terms we do not even need to introduce a notion of limit. The restriction to finite depth means that for a term M and a depth d the trees $BT_d(M)$ and $BT_d^\eta(M)$ are given by finitely many complete left reductions on M.

Definition 4.1.0.1. The *local Böhm trees* $\mathrm{bt}_d(M)$ and $\mathrm{bt}_d^\eta(M)$ for a term M and $d \in \mathbb{N}$ are given by

$$\mathrm{bt}_0(M) = \Omega$$

$$\mathrm{bt}_{d+1}(M) = \begin{cases} \lambda x_1 ... x_n . y\ \mathrm{bt}_d(P_1)\ ...\ \mathrm{bt}_d(P_m) & \text{if } M = \lambda x_1 ... x_n . y\ P_1\ ...\ P_m \\ \Omega & \text{otherwise} \end{cases}$$

and

$$\mathrm{bt}_0^\eta(M) = \Omega$$

$$\mathrm{bt}_{d+1}^\eta(M) = \begin{cases} (y, (\mathrm{bt}_d^\eta(P_1), ..., \mathrm{bt}_d^\eta(P_m), \mathrm{bt}_d^\eta(x_{d,n+1}), ...)) & \text{if } M = \lambda x_{d,1} ... x_{d,n} . y\ P_1\ ...\ P_m \\ \Omega & \text{otherwise.} \end{cases}$$

The definition of the local Böhm tree of a term is the same as the definition of Böhm trees, except that we do not check whether a term is solvable, but directly whether it is a head normal form.

Proposition 4.1.0.1. *For every deterministic term M and every $d \in \mathbb{N}$ there exists $k \in \mathbb{N}$ such that*

$$BT_d(M) = \mathrm{bt}_d(\mathrm{L}^k(M)) \text{ and } BT_d^\eta(M) = \mathrm{bt}_d^\eta(\mathrm{L}^k(M)).$$

Proof. First it is easy to prove by induction on d that if $BT_d(M) = \mathrm{bt}_d(M)$ then $BT_d(M) = \mathrm{bt}_d(\mathrm{L}(M))$. The result is immediate if $d = 0$. If M is unsolvable then $\mathrm{L}(M)$ is unsolvable and $BT_{d+1}(M) = \mathrm{bt}_{d+1}(\mathrm{L}(M)) = \Omega$.

Otherwise if $BT_{d+1}(M) = \mathrm{bt}_{d+1}(M)$ and M is solvable then it is necessarily head normal, i.e. $M = \lambda x_1 ... x_n . y\ P_1\ ...\ P_m$, and $\mathrm{L}(M) = \lambda x_1 ... x_n . y\ \mathrm{L}(P_1)\ ...\ \mathrm{L}(P_m)$. By induction hypothesis $BT_d(P_i) = \mathrm{bt}_d(\mathrm{L}(P_i))$ for $i \leq m$ and we immediately get $BT_{d+1}(M) = \mathrm{bt}_{d+1}(\mathrm{L}(M))$.

For the same reason if $BT_d^\eta(M) = \mathrm{bt}_d^\eta(M)$ then $BT_d^\eta(M) = \mathrm{bt}_d^\eta(\mathrm{L}(M))$.

Now we prove the proposition by a simple induction on d. If $d = 0$ then $BT_0(M) = \mathrm{bt}_0(M) = \Omega$ and $BT_0^\eta(M) = \mathrm{bt}_0^\eta(M) = \Omega$. Similarly if M is unsolvable then $BT_{d+1}(M) = \mathrm{bt}_{d+1}(M) = \Omega$ and $BT_{d+1}^\eta(M) = \mathrm{bt}_{d+1}^\eta(M) = \Omega$.

Otherwise if M is solvable then for some k_0 we have that $\mathrm{L}^{k_0}(M)$ is head normal, we write $\mathrm{L}^{k_0}(M) = \lambda x_1 ... x_n.y\ P_1\ ...\ P_m$. By induction hypothesis there are k_1,...,k_m such that $BT_d(P_i) = \mathrm{bt}_d(\mathrm{L}^{k_i}(P_i))$ and $BT_d^\eta(P_i) = \mathrm{bt}_d^\eta(\mathrm{L}^{k_i}(P_i))$ for all $i \leq m$. Let $k = k_0 + \max_{i \leq m} k_i$, we have $BT_d(M) = \mathrm{bt}_d(\mathrm{L}^k(M))$ and $BT_d^\eta(M) = \mathrm{bt}_d^\eta(\mathrm{L}^k(M))$. $\square$

We can define the same kind of local Böhm trees in the probabilistic case.

Definition 4.1.0.2. The *local probabilistic Böhm trees* $\mathrm{pt}_d(M)$ and $\mathrm{pt}_d^\eta(M)$ for a term M and $d \in \mathbb{N}$ are given by

$$\begin{aligned} \mathrm{vt}_{d+1}(\lambda x_1 ... x_n.y\ P_1\ ...\ P_m) = &\quad \lambda x_1 ... x_n.y\ \mathrm{pt}_d(P_1)\ ...\ \mathrm{pt}_d(P_m) \\ \mathrm{pt}_d(M): &\quad t \mapsto \mathcal{P}(M = \{h \mid \mathrm{vt}_d(h) = t\}). \end{aligned}$$

and

$$\begin{aligned} \mathrm{vt}_{d+1}^\eta(\lambda x_1 ... x_n.y\ P_1\ ...\ P_m) = &\quad (y, (\mathrm{pt}_d^\eta(P_1), ..., \mathrm{pt}_d^\eta(P_m), \mathrm{pt}_d^\eta(x_{d,n+1}), ...)) \\ \mathrm{pt}_d^\eta(M): &\quad t \mapsto \mathcal{P}(M = \{h \mid \mathrm{vt}_d^\eta(h) = t\}). \end{aligned}$$

But here we do not have the same result as in the deterministic case. If we consider a term M such that $M \twoheadrightarrow_{\beta^c} \lambda x.x +_p M$ then M has the same Böhm tree as $\lambda x.x$, i.e. for any $d \in \mathbb{N}$ we have $PT_{d+1}(M)(\lambda x.x) = 1$ and $PT_{d+1}(M)(t) = 0$ otherwise. But we can not reduce $M \twoheadrightarrow_{\beta/\equiv} M'$ to get $PT_{d+1}(M) = \mathrm{pt}_{d+1}(M')$.

For that reason we must not only approximate the Böhm trees in depth, but also in width: we can prove that for any term M, any $d \in \mathbb{N}$ and any $\epsilon > 0$ there is k such that $\mathrm{pt}_d(\mathrm{L}^k(M))$ approximates $PT_d(M)$ up to ϵ.

4.1.1 Ordering trees

We will detail the case of the non extensional Böhm trees. The non extensional and infinitely extensional Böhm trees are actually very similar, and they are built in the same way, so it is easy to adapt the following proofs to the infinitely extensional case. The difference between the two notions will arise later, when we will compute with terms knowing only their Böhm trees.

First of all we need to define an order saying that a tree is less defined than another. In the deterministic calculus such an order is simply given by the rule $\Omega \preceq T$ for any tree T. Here we must be a liittle more subtle to take into account the probabilities.

Definition 4.1.1.1. The relations $\preceq_{d,\epsilon}$ on $\mathcal{PT}_d$ and $\preceq_{d,\epsilon}^v$ on $\mathcal{VT}_d$ are defined for $d \in \mathbb{N}$ and $\epsilon \geq 0$ by:

$$\frac{\forall i \leq m, T_i \preceq_{d,\epsilon} T_i'}{\lambda x_1 ... x_n.y\ T_1\ ...\ T_m \preceq_{d+1,\epsilon}^v \lambda x_1 ... x_n.y\ T_1'\ ...\ T_m'}$$

$$\frac{\forall A \subset \mathcal{VT}_d, \sum_{t \in \uparrow A} T(t) \leq \sum_{t \in \uparrow_\epsilon A} T'(t) + \epsilon}{T \preceq_{d,\epsilon} T'}$$

where $\uparrow_\epsilon A = \{t \in \mathcal{VT}_d \mid \exists t' \in A : t' \preceq^v_{d,\epsilon} t\}$ and $\uparrow A = \uparrow_0 A$. We also define $\preceq_d = \preceq_{d,0}$ and $\preceq^v_d = \preceq^v_{d,0}$.

Remark that for $d = 0$, $\preceq^v_{0,\epsilon}$ is a relation on $\mathcal{VT}_0 = \emptyset$ so it is uniquely defined.

We want to prove that $\preceq_d$ is an order on probabilistic Böhm trees, and that the Böhm tree $PT_d(M)$ of a term is the supremum of its approximations $\mathrm{pt}_d(\mathrm{L}^k(M))$. It is easy to prove that the sequence $(\mathrm{pt}_d(\mathrm{L}^k(M)))$ is increasing for $\preceq_d$ and $PT_d(M)$ is an upper bound, but to prove that this is actually its supremum we will also prove that for all $\epsilon > 0$ we have $PT_d(M) \preceq_{d,\epsilon} \mathrm{pt}_d(\mathrm{L}^k(M))$ for k big enough.

For now let us check that $\preceq_{d,\epsilon}$ behaves as an order up to ϵ.

Proposition 4.1.1.1. *For all $d \in \mathbb{N}$ and all $\epsilon \leq \epsilon'$ we have $\preceq_{d,\epsilon} \subset \preceq_{d,\epsilon'}$ and $\preceq^v_{d,\epsilon} \subset \preceq^v_{d,\epsilon'}$.*

Proof. By induction on d. For $d = 0$ we have $\preceq^v_{0,\epsilon} = \emptyset$ and $\Omega \preceq_{0,\epsilon} \Omega$ for all $\epsilon \geq 0$.

Let $\epsilon \leq \epsilon'$, if $t = \preceq^v_{d+1,\epsilon} t'$ then $t = \lambda x_1 ... x_n . y\, T_1 \, ... \, T_m$ and $t' = \lambda x_1 ... x_n . y\, T'_1 \, ... \, T'_m$ with $T_i \preceq_{d,\epsilon} T'_i$ for all $i \leq m$, thus by induction hypothesis $T_i \preceq_{d,\epsilon'} T'_i$ for $i \leq m$ and $t \preceq^v_{d+1,\epsilon'} t'$.

Then for all $A \subset \mathcal{VT}_{d+1}$ we have $\uparrow_\epsilon A \subset \uparrow_{\epsilon'} A$: if $t \in \uparrow_\epsilon A$ then there is $t' \in A$ such that $t' \preceq^v_{d+1,\epsilon} t$, hence $t' \preceq^v_{d+1,\epsilon'} t$. Thus if $T \preceq_{d+1,\epsilon} T'$ we have for all $A \subset \mathcal{VT}_{d+1}$:

$$\sum_{t \in \uparrow A} T(t) \leq \sum_{t \in \uparrow_\epsilon A} T'(t) + \epsilon \leq \sum_{t \in \uparrow_{\epsilon'} A} T'(t) + \epsilon \leq \sum_{t \in \uparrow_{\epsilon'} A} T'(t) + \epsilon'.$$

□

Proposition 4.1.1.2. *For all $d \in \mathbb{N}$ and $\epsilon, \epsilon' \geq 0$ we have $\preceq_{d,\epsilon} \cdot \preceq_{d,\epsilon'} \subset \preceq_{d,\epsilon+\epsilon'}$ and $\preceq^v_{d,\epsilon} \cdot \preceq^v_{d,\epsilon'} \subset \preceq^v_{d,\epsilon+\epsilon'}$.*

Proof. We reason by induction on d, and the result is immediate when $d = 0$ or for value trees.

The result on value trees gives that for all $A \subset \mathcal{VT}_{d+1}$ we have $\uparrow_{\epsilon'} \uparrow_\epsilon A \subset \uparrow_{\epsilon+\epsilon'} A$. Then if $T \preceq_{d+1,\epsilon} T'$ and $T' \preceq_{d+1,\epsilon'} T''$ we have for all $A \subset \mathcal{VT}_{d+1}$:

$$\sum_{t \in \uparrow A} T(t) \leq \sum_{t \in \uparrow_\epsilon A} T'(t) + \epsilon \leq \sum_{t \in \uparrow_{\epsilon'} \uparrow_\epsilon A} T''(t) + \epsilon + \epsilon' \leq \sum_{t \in \uparrow_{\epsilon+\epsilon'} A} T''(t) + \epsilon + \epsilon'.$$

□

Proposition 4.1.1.3. *For all $d \in \mathbb{N}$, $\preceq_d$ and $\preceq^v_d$ are orders.*

Proof. We reason by induction on d. If $d = 0$ the result is immediate, and we easily deduce that $\preceq^v_{d+1}$ is an order from the fact that $\preceq_d$ is an order.

$\preceq_{d+1}$ is trivially reflexive, and we proved that $\preceq_{d+1,0} \cdot \preceq_{d+1,0} \subset \preceq_{d+1,0+0}$, i.e. $\preceq_{d+1}$ is transitive.

If $T \preceq_{d+1} T'$ and $T' \preceq_{d+1} T$ then $\sum_{t\in\uparrow A} T(t) = \sum_{t\in\uparrow A} T'(t)$ for all $A \subset \mathcal{VT}_{d+1}$. In particular given a value tree $t \in \mathcal{VT}_{d+1}$ we have that $\uparrow \{t\} = \{t' \mid t \preceq_{d+1} t'\}$ and $\uparrow \{t' \mid t \preceq_{d+1} t' \text{ and } t \neq t'\} = \{t' \mid t \preceq_{d+1} t' \text{ and } t \neq t'\}$, so

$$\begin{array}{rcrl} \sum_{t\preceq_{d+1}t'} & T(t') = & \sum_{t\preceq_{d+1}t'} & T'(t') \\ \sum_{t\preceq_{d+1}t' \text{ and } t\neq t'} & T(t') = & \sum_{t\preceq_{d+1}t' \text{ and } t\neq t'} & T'(t') \\ & T(t) = & & T'(t) \end{array}$$

$\preceq_{d+1}$ is antisymmetric. □

Lemma 4.1.1.4. *For all $d \in \mathbb{N}$ and $\epsilon \geq 0$ we have $T \preceq_{d,\epsilon} T'$ if and only if for all finite $A \subset_f \mathcal{VT}_d$*

$$\sum_{t\in\uparrow A} T(t) \leq \sum_{t\in\uparrow_\epsilon A} T'(t) + \epsilon.$$

Proof. We just need to observe that for any tree $T \in \mathcal{PT}_d$ all the sums $\sum_{t\in\uparrow_\epsilon A} T(t)$ for $A \subset \mathcal{VT}_d$ can be recovered from the sums $\sum_{t\in\uparrow_\epsilon A} T(t)$ for finite A.

Indeed if $A \subset \mathcal{VT}_d$ then we have $\sum_{t\in\uparrow_\epsilon A} T(t) = \sup_{B\subset_f\uparrow_\epsilon A} \sum_{t\in B} T(t)$, and for all finite $B \subset_f \uparrow_\epsilon A$ there is $B' \subset_f A$ such that $B \subset \uparrow_\epsilon B'$. Thus
$\sum_{t\in\uparrow_\epsilon A} T(t) = \sup_{B'\subset_f A} \sum_{t\in\uparrow_\epsilon B} T(t)$. □

Proposition 4.1.1.5. *Given two trees T and T' in $\mathcal{PT}_d$, if for all $\epsilon > 0$ we have $T \preceq_{d,\epsilon} T'$ then $T \preceq_d T'$.*

Proof. We prove this, as well as the corresponding result for value trees, by induction on d. As usual the result is immediate when $d = 0$ or for value trees.

We know that if $\epsilon \leq \epsilon'$ then for all $A \subset \mathcal{VT}_{d+1}$ we have $\uparrow_\epsilon A \subset \uparrow_{\epsilon'} A$. This means that for a fixed $\epsilon > 0$ we have for all $A \subset \mathcal{VT}_{d+1}$ and for all $\epsilon' > 0$ with $\epsilon' \leq \epsilon$:

$$\sum_{t\in\uparrow A} T(t) \leq \sum_{t\in\uparrow_{\epsilon'} A} T'(t) + \epsilon' \leq \sum_{t\in\uparrow_\epsilon A} T'(t) + \epsilon'$$

hence $\sum_{t\in\uparrow A} T(t) \leq \sum_{t\in\uparrow_\epsilon A} T'(t)$. As $\uparrow_\epsilon A$ decreases with ϵ we get

$$\sum_{t\in\uparrow A} T(t) \leq \sum_{t\in\bigcap_{\epsilon>0}\uparrow_\epsilon A} T'(t).$$

We want to prove that for all $A \subset \mathcal{VT}_{d+1}$ we have $\bigcap_{\epsilon>0} \uparrow_\epsilon A = \uparrow A$. We actually only need to prove this when A is finite. In this case for any $t \in \bigcap_{\epsilon>0} \uparrow_\epsilon A$ there is $t' \in A$ such that $t' \preceq^v_{d+1,\epsilon} t$ for arbitrary small $\epsilon > 0$, hence $t' \preceq^v_{d+1,\epsilon} t$ for all $\epsilon > 0$. Then by induction hypothesis $t' \preceq^v_{d+1} t$ and $t \in \uparrow A$. □

We have the results we need on $\preceq_{d,\epsilon}$ and $\preceq^v_{d,\epsilon}$ to prove that a tree is the supremum of a sequence.

Proposition 4.1.1.6. *Given a sequence (T_n) and a tree T in $\mathcal{PT}_d$, if*

- *for all $n \in \mathbb{N}$, $T_n \preceq_d T$*
- *and for all $\epsilon > 0$ there is n such that $T \preceq_{d,\epsilon} T_n$*

then $T = \sup_n T_n$ *for* $\preceq_d$.

Proof. We have to prove that if $T' \in \mathcal{PT}_d$ is such that $T_n \preceq_d T'$ for all n then $T \preceq_d T'$. Given such a T' we have for all $\epsilon > 0$ that $T \preceq_{d,\epsilon} T_n \preceq_d T'$ for some n, hence $T \preceq_{d,\epsilon} T'$, and according to the previous result $T \preceq_d T'$. □

4.1.2 Böhm trees as a supremum

To use this result with local Böhm trees let us first show that the complete left reduction of a term gives an increasing sequence of local Böhm trees bounded by its Böhm tree.

Proposition 4.1.2.1. *If* M *is a canonical term and* $M \to_{\beta^c} N$ *then for all* $d \in \mathbb{N}$, $\mathrm{pt}_d(M) \preceq_d \mathrm{pt}_d(N)$.

Proof. We prove this, as well as the corresponding result for head normal values, by induction on d. If $d = 0$ the result is immediate.

If $h = \lambda x_1 ... x_n . y\ P_1\ ...\ P_m$ is a head normal value and $h \to_{\beta^c} h'$ then $h' = \lambda x_1 ... x_n . y\ Q_1\ ...\ Q_m$ with $P_i \to^?_{\beta^c} Q_i$ for all $i \leq m$. By induction hypothesis or reflexivity we have $\mathrm{pt}_d(P_i) \preceq_d \mathrm{pt}_d(Q_i)$ for $i \leq m$, and $\mathrm{vt}_{d+1}(h) \preceq^v_{d+1} \mathrm{vt}_{d+1}(h')$.

If M is a canonical term we reason by induction on it.

- If $M = M_1 +_p M_2$ then $\mathrm{pt}_{d+1}(M) = p.\mathrm{pt}_{d+1}(M_1) + (1-p)\mathrm{pt}_{d+1}(M_2)$ and $N = N_1 +_p N_2$ with $M_i \to^?_{\beta^c} N_i$ for $i \in \{1; 2\}$. By induction hypothesis $\mathrm{pt}_{d+1}(M_i) \preceq_{d+1} \mathrm{pt}_{d+1}(N_i)$ for $i \in \{1; 2\}$ so $\mathrm{pt}_{d+1}(M) \preceq_{d+1} \mathrm{pt}_{d+1}(N)$.
- If $M = h$ is a head normal value then $N = h'$ is also a head normal value and we know that $\mathrm{vt}_{d+1}(h) \preceq^v_{d+1} \mathrm{vt}_{d+1}(h')$. Then for all $A \subset \mathcal{VT}_{d+1}$ either $\mathrm{vt}_{d+1}(h) \notin\uparrow A$ and $\sum_{t \in \uparrow A} \mathrm{pt}_{d+1}(M)(t) = 0$, or $\mathrm{vt}_{d+1}(h) \in\uparrow A$ but then necessarily $\mathrm{vt}_{d+1}(h') \in\uparrow A$ and $\sum_{t \in \uparrow A} \mathrm{pt}_{d+1}(M)(t) = \sum_{t \in \uparrow A} \mathrm{pt}_{d+1}(N)(t) = 1$.
- If M is a value but not in head normal form then $\mathrm{pt}_{d+1}(M) = 0$ so necessarily $\mathrm{pt}_{d+1}(M) \preceq_{d+1} \mathrm{pt}_{d+1}(N)$.

□

Proposition 4.1.2.2. *For every canonical term* M *and every* $d \in \mathbb{N}$ *we have* $\mathrm{pt}_d(M) \preceq_d PT_d(M)$.

Proof. We prove this and the corresponding result for value trees by induction on d. The result is immediate if $d = 0$ and for value trees.

Given a canonical term M we reason by induction on M.

- If M is a sum the result is immediate by induction hypothesis.
- If M is a head normal value h then we have $\mathrm{vt}_{d+1}(h) \preceq^v_{d+1} VT_{d+1}(h)$ so $\mathrm{pt}_{d+1}(M) \preceq_{d+1} PT_{d+1}(M)$.
- If M is a value but not head normal then $\mathrm{pt}_{d+1}(M) = 0 \preceq_{d+1} PT_{d+1}(M)$.

□

For any term M the tree $PT_d(M)$ is an upper bound for $(\mathrm{pt}_d(\mathrm{L}^k(M)))$. To prove it is the least one we introduce a notion of normal form up to $\epsilon > 0$.

Definition 4.1.2.1. The sets $\mathrm{NF}_{d,\epsilon}$ of *canonical* d, ϵ*-normal forms* and $\mathrm{NF}^v_{d,\epsilon}$ of d, ϵ*-normal values* are defined for $d \in \mathbb{N}$ and $\epsilon \geq 0$ by

$$\mathrm{NF}^v_{0,\epsilon} = \{v \text{ value}\}$$
$$\mathrm{NF}^v_{d+1,\epsilon} = \{\lambda x_1 ... x_n . y\ P_1\ ...\ P_m \mid \forall i \leq m, P_i \in \mathrm{NF}_{d,\epsilon}\}$$
$$\mathrm{NF}_{d,\epsilon} = \left\{\sum_i p_i.v_i + \left(1 - \sum_i p_i\right).M \mid v_i \in \mathrm{NF}^v_{d,\epsilon} \text{ and } \left(1 - \sum_i p_i\right).\mathcal{P}_{\Downarrow}(M) \leq \epsilon\right\}$$

Proposition 4.1.2.3. *If M is a d, ϵ-normal form then $PT_d(M) \preceq_{d,\epsilon} \mathrm{pt}_d(M)$.*

Proof. We prove this and the corresponding result for values by induction on d. The result is immediate when $d = 0$ and for values.

Given $M \equiv_{\mathrm{syn}} \sum_i p_i.h_i + (1 - \sum_i p_i).P \in \mathrm{NF}_{d+1,\epsilon}$ with $h_i \in \mathrm{NF}^v_{d+1,\epsilon}$ for all i and $(1 - \sum_i p_i).\mathcal{P}_{\Downarrow}(P) \leq \epsilon$ then for all $A \subset \mathcal{VT}_{d+1}$ we have for all i that if $VT_{d+1}(h_i) \in \uparrow A$ then $\mathrm{vt}_{d+1}(h_i) \in \uparrow_\epsilon A$. Hence

$$\sum_{t \in \uparrow A} PT_{d+1}(M)(t) - \sum_{t \in \uparrow_\epsilon A} \mathrm{pt}_{d+1}(M)(t) \leq \left(1 - \sum_i p_i\right) \sum_{t \in \uparrow A} PT_{d+1}(P).$$

We know that $\sum_{t \in \mathcal{VT}_{d+1}} PT_{d+1}(P) = \mathcal{P}_{\Downarrow}(P)$ so $(1 - \sum_i p_i) \sum_{t \in \uparrow A} PT_{d+1}(P) \leq \epsilon$ and

$$\sum_{t \in \uparrow A} PT_{d+1}(M)(t) \leq \sum_{t \in \uparrow_\epsilon A} \mathrm{pt}_{d+1}(M)(t) + \epsilon.$$

□

Proposition 4.1.2.4. *For all $d \in \mathbb{N}$ and all $\epsilon > 0$, for all M there is $k \in \mathbb{N}$ such that $\mathrm{L}^k(M) \in \mathrm{NF}_{d,\epsilon}$.*

Proof. By induction on d, trivial for $d = 0$.

Given any canonical term M there exist $k_0 \in \mathbb{N}$ and head normal values h_i such that $\mathrm{H}^{k_0}(M) \equiv_{\mathrm{syn}} \sum_i p_i.h_i + (1 - \sum_i p_i).P$ with $(1 - \sum_i p_i).\mathcal{P}_{\Downarrow}(P) \leq \epsilon$.

Let $h_i = \lambda x_1 ... x_{n_i}.y_i\ P_{i,1}\ ...\ P_{i,m_i}$, by induction hypothesis there are $k_{i,1}, ..., k_{i,m_i}$ such that $\mathrm{L}^{k_{i,j}}(P_{i,j}) \in \mathrm{NF}_{d,\epsilon}$ for all i and $j \leq m_i$. But $\mathrm{NF}_{d,\epsilon}$ is stable by β-reduction so let $k = k_0 + \max_{i,j \leq m_i} k_{i,j}$, we have $\mathrm{L}^k(M) \in \mathrm{NF}_{d+1,\epsilon}$. □

Corollary 4.1.2.5. *For all $d \in \mathbb{N}$ and $\epsilon > 0$, for all M there is k such that $PT_d(M) \preceq_{d,\epsilon} \mathrm{pt}_d(\mathrm{L}^k(M))$.*

Proof. If $\mathrm{L}^k(M)$ is a d, ϵ-normal form then $PT_d(M) = PT_d(\mathrm{L}^k(M)) \preceq_{d,\epsilon} \mathrm{pt}_d(\mathrm{L}^k(M))$. □

Corollary 4.1.2.6. *For all $d \in \mathbb{N}$ and all term M we have*

$$PT_d(M) = \sup_{k \in \mathbb{N}} \mathrm{pt}_d(\mathrm{L}^k(M)).$$

This result is very useful to prove that two terms have the same Böhm tree, as the Böhm tree at depth d of a term M is entirely defined by the sequence $\mathrm{pt}_d(\mathrm{L}^k(M))$

Proposition 4.1.2.7. *Given any two terms M and N and any $d \in \mathbb{N}$, we have that $PT_d(M) \preceq_d PT_d(N)$ if and only if for all $k \in \mathbb{N}$ and $\epsilon > 0$ there is $k' \in \mathbb{N}$ such that $\mathrm{pt}_d(\mathrm{L}^k(M)) \preceq_{d,\epsilon} \mathrm{pt}_d(\mathrm{L}^{k'}(N))$.*

Proof. If $PT_d(M) \preceq_d PT_d(N)$ then for all k we have $\mathrm{pt}_d(\mathrm{L}^k(M)) \preceq_d PT_d(N)$, and for all $\epsilon > 0$ there is k' such that $PT_d(N) \preceq_{d,\epsilon} \mathrm{pt}_d(\mathrm{L}^{k'}(N))$.

Conversely for all $\epsilon > 0$ there is k such that $PT_d(M) \preceq_{d,\epsilon} \mathrm{pt}_d(\mathrm{L}^k(M))$, so by hypothesis there is k' such that $PT_d(M) \preceq_{d,2\epsilon} \mathrm{pt}_d(\mathrm{L}^{k'}(N))$. We know that $\mathrm{pt}_d(\mathrm{L}^{k'}(N)) \preceq_d PT_d(N)$ so $PT_d(M) \preceq_{d,2\epsilon} PT_d(N)$, hence $PT_d(M) \preceq_d PT_d(N)$. □

To compare the Böhm trees of two terms it is sufficient to compare the local Böhm trees of their complete left reductions.

All these results are for non extensional Böhm trees, but they also extend to infinitely extensional ones. The proofs are exactly the same so we will simply give the definitions of the order and state the final results.

Definition 4.1.2.2. The relation $\preceq^{\eta}_{d,\epsilon}$ on $\mathcal{PT}^{\eta}_d$ and $\preceq^{v\eta}_{d,\epsilon}$ on $\mathcal{VT}^{\eta}_d$ are defined for $d \in \mathbb{N}$ and $\epsilon \geq 0$ by:

$$\frac{\forall n \in \mathbb{N}^*, T_n \preceq^{\eta}_{d,\epsilon} T'_n}{(y,(T_n)) \preceq^{v\eta}_{d+1,\epsilon} (y,(T'_n))} \qquad \frac{\forall A \subset \mathcal{VT}^{\eta}_d, \sum_{t \in \uparrow A} T(t) \leq \sum_{t \in \uparrow_\epsilon A} T'(t) + \epsilon}{T \preceq^{\eta}_{d,\epsilon} T'}$$

where $\uparrow_\epsilon A = \{t \in \mathcal{VT}^{\eta}_d \mid \exists t' \in A : t' \preceq^{v\eta}_{d,\epsilon} t\}$ and $\uparrow A = \uparrow_0 A$. We also define $\preceq^{\eta}_d = \preceq^{\eta}_{d,0}$ and $\preceq^{v\eta}_d = \preceq^{v\eta}_{d,0}$.

Proposition 4.1.2.8. *For all $d \in \mathbb{N}$, $\preceq^{\eta}_d$ and $\preceq^{v\eta}_d$ are orders.*

Proposition 4.1.2.9. *For all $d \in \mathbb{N}$ and all term M we have*

$$PT^{\eta}_d(M) = \sup_{k \in \mathbb{N}} \mathrm{pt}^{\eta}_d(\mathrm{L}^k(M)).$$

Proposition 4.1.2.10. *Given two terms M and N and given $d \in \mathbb{N}$, we have $PT^{\eta}_d(M) \preceq^{\eta}_d PT^{\eta}_d(N)$ if and only if for all $k \in \mathbb{N}$ and $\epsilon > 0$ there is $k' \in \mathbb{N}$ such that $\mathrm{pt}^{\eta}_d(\mathrm{L}^k(M)) \preceq^{\eta}_{d,\epsilon} \mathrm{pt}^{\eta}_d(\mathrm{L}^{k'}(N))$.*

4.2 Böhm trees as a model

4.2.1 Warming up: the non extensional trees

Proving that the Böhm tree equality is contextual is not easy. To understand where the problem lies, and how to solve it, let us consider a similar but more simple example.

Let us consider the equivalence on deterministic term such that two normalizable terms are equivalent if they have the same normal form, and two non normalizable terms are equivalent. It is not immediate that this equivalence is contextual: if M and M' both normalize into N then given a context C we know that $C[M]$ and $C[M']$ both reduce to $C[N]$, but $C[N]$ itself is not necessarily normal. We need to say that if $C[M]$ or $C[M']$ is normalizable then so is $C[N]$, and then both $C[M]$ and $C[M']$ reduce into the normal form of $C[N]$.

The Böhm tree equality is a refinement of this equivalence. Instead of checking whether a term is normalizable we only check whether it is head normalizable, and if two term are head normalizable then we compare inductively their subterms. The problem is that by doing so we leave the realm of terms: while a normal form is still a term, a Böhm tree is an infinite object. Thus if M and M' have the same Böhm tree T then given a context C we can not directly say that the terms $C[M]$ and $C[M']$ are related to some object $C[T]$, which would have the same Böhm tree as $C[M]$ and $C[M']$.

A solution is to define operations of abstraction and application on Böhm trees, so that $C[T]$ is actually defined as a tree. This is how Barendregt proves that Böhm trees form a model [1]. We did not try to extend these definitions to probabilistic trees.

We can also observe that to define our first equivalence on normalizable terms, we actually do not need to consider normal forms. Indeed we know that two normalizable terms M and M' have the same normal form if and only if they both reduce into a same term P, regardless of whether P is normal or not. Then for any context C, if $C[M]$ and $C[M']$ both reduce into $C[P]$ they are equivalent.

We can perform the same trick with Böhm trees. We can give a characterization of the Böhm tree equality which can be easily proven contextual if we forget that Böhm trees are supposed to be normal forms.

The Böhm tree equality is based on two ideas. First this is a sensible equality: all unsolvable terms have the same Böhm tree. Secondly we have a notion of depth and we say that two terms have the same Böhm tree when their differences can be pushed at arbitrary large depth. To these two ideas will correspond two relations on terms. The sensibility is given by the relations $=_\Omega$ on deterministic terms and $=^s_{\Omega,\epsilon}$ on probabilistic terms we define in the section 3.2. We also define a relation of equality up to a finite depth.

Definition 4.2.1.1. The relation $\sim_d$ of *equality up to depth d* for $d \in \mathbb{N}$ is given by

$$\frac{}{M \sim_0 N} \qquad \frac{}{x \sim_{d+1} x} \qquad \frac{M \sim_{d+1} N}{\lambda x.M \sim_{d+1} \lambda x.N}$$

$$\frac{M \sim_{d+1} N \qquad M' \sim_d N'}{M\ M' \sim_{d+1} N\ N'} \qquad\qquad \frac{M_1 \sim_{d+1} N_1 \qquad M_2 \sim_{d+1} N_2}{M_1 +_p M_2 \sim_{d+1} N_1 +_p N_2}$$

In the most simple case, i.e. for non extensional deterministic trees, we can prove that given two deterministic terms M and N we have

$$M =_{\mathcal{B}} N \text{ if and only if } \forall d \in \mathbb{N}, M \twoheadrightarrow_\beta \cdot \sim_d \cdot =_\Omega \cdot \sim_d \cdot \twoheadleftarrow_\beta N.$$

It is easy to prove that if two terms have the same Böhm tree then this relation holds.

Proposition 4.2.1.1. *Given two deterministic terms M and N, if $M =_{\mathcal{B}} N$ then*

$$\forall d \in \mathbb{N}, M \twoheadrightarrow_\beta \cdot \sim_d \cdot =_\Omega \cdot \sim_d \cdot \twoheadleftarrow_\beta N.$$

Proof. For all $d \in \mathbb{N}$ the proposition 4.1.0.1 and its proof give that for $k \in \mathbb{N}$ large enough we have $BT_d(M) = \mathrm{bt}_d(\mathrm{L}^k(M))$, hence for k large enough we have both $BT_d(M) = \mathrm{bt}_d(\mathrm{L}^k(M))$ and $BT_d(N) = \mathrm{bt}_d(\mathrm{L}^k(N))$. Then if $M =_{\mathcal{B}} N$ we have $\mathrm{bt}_d(\mathrm{L}^k(M)) = \mathrm{bt}_d(\mathrm{L}^k(N))$.

From there we can prove by induction on $\mathrm{L}^k(M)$ that $\mathrm{L}^k(M) \sim_d \cdot =_\Omega \mathrm{L}^k(N)$. □

Conversely we want to prove that if this relation holds between two deterministic terms M and N then they have the same Böhm tree. To achieve this we will use the fact that the Böhm tree of a term is the limit of the local Böhm trees of its complete left reductions. So first we prove that this relation is stable by complete left reduction.

Proposition 4.2.1.2. *Given $d \in \mathbb{N}$ and terms M, N, P and Q, if $M \sim_d N$ and $P \sim_d Q$ then $M\left[{}^{P}/_{x}\right] \sim_d N\left[{}^{Q}/_{x}\right]$.*

Proof. By induction on $M \sim_d N$.

- If $d = 0$ then $M\left[{}^{P}/_{x}\right] \sim_0 N\left[{}^{Q}/_{x}\right]$.
- If $M = N = x$ then by hypothesis $P \sim_{d+1} Q$.
- If $M = N = y \neq x$ then $M\left[{}^{P}/_{x}\right] = N\left[{}^{Q}/_{x}\right] = y$.
- If $M = \lambda y.M'$ and $N = \lambda y.N'$ then $M\left[{}^{P}/_{x}\right] = \lambda y.\left(M'\left[{}^{P}/_{x}\right]\right)$ and $N\left[{}^{Q}/_{x}\right] = \lambda y.\left(N'\left[{}^{Q}/_{x}\right]\right)$ so we conclude by induction hypothesis on M' and N'.
- Similarly if M and N are applications or sums the result is immediate by induction hypothesis.

□

Proposition 4.2.1.3. *For any $d \in \mathbb{N}$, if M and N are canonical (or deterministic) terms and $M \sim_{d+1} N$ then $\mathrm{L}(M) \sim_d \mathrm{L}(N)$.*

Proof. The intuition is that when we perform a reduction $(\lambda x.P)\ Q \to_\beta P\left[{}^{Q}/{}_{x}\right]$, we take the argument Q at depth 1 and we substitute it in P, so at most at depth 0. Thus the depth of the subterms decreases of at most 1 when we perform a β-reduction.

Formally we prove the result by induction on M.

- If $M = M_1 +_p M_2$ then $N = N_1 +_p N_2$ with $M_i \sim_{d+1} N_i$ for $i \in \{1; 2\}$ so by induction hypothesis $\mathrm{L}(M_i) \sim_d \mathrm{L}(N_i)$ and $\mathrm{L}(M) \sim_d \mathrm{L}(N)$.

- If $M = \lambda x_1 ... x_n . y\ P_1\ ...\ P_m$ then $N = \lambda x_1 ... x_n . y\ Q_1\ ...\ Q_m$ with $P_i \sim_d Q_i$ for $i \leq m$. Then either $d = 0$ and $\mathrm{L}(M) \sim_0 \mathrm{L}(N)$, or by induction hypothesis we have $\mathrm{L}(P_i) \sim_{d-1} \mathrm{L}(Q_i)$ for $i \leq m$ and $\mathrm{L}(M) \sim_d \mathrm{L}(N)$.

- If $M = \lambda x_1 ... x_n . (\lambda y. M_0)\ P\ R_1\ ...\ R_m$ then $N = \lambda x_1 ... x_n . (\lambda y. N_0)\ Q\ S_1\ ...\ S_m$ with $M_0 \sim_{d+1} N_0$, $P \sim_d Q$ and $R_i \sim_d S_i$ for $i \leq m$. Then $M_0 \sim_d N_0$ and

  $$\mathrm{L}(M) = \lambda x_1 ... x_n . M_0 \left[{}^{P}/{}_{y}\right]\ R_1\ ...\ R_m \sim_d \lambda x_1 ... x_n . N_0 \left[{}^{Q}/{}_{y}\right]\ S_1\ ...\ S_m = \mathrm{L}(N).$$

$\square$

Proposition 4.2.1.4. *Given two deterministic terms M and N, if $M =_\Omega N$ then $\mathrm{L}(M) =_\Omega \mathrm{L}(N)$.*

Proof. First we can prove by an easy induction on a term M that if $M =_\Omega N$ and $P =_\Omega Q$ then $M\left[{}^{P}/{}_{x}\right] =_\Omega N\left[{}^{Q}/{}_{x}\right]$. Indeed if M is unsolvable then $M\left[{}^{P}/{}_{x}\right]$ is unsolvable.

Then the result is given by an easy induction on M. The only non trivial case is when $M = \lambda x_1 ... x_n . (\lambda y. M_0)\ P\ R_1\ ...\ R_m$, in which case either M and N are both unsolvable (and so are $\mathrm{L}(M)$ and $\mathrm{L}(N)$) or $N = \lambda x_1 ... x_n . (\lambda y. N_0)\ Q\ S_1\ ...\ S_m$ with $M_0 =_\Omega N_0$, $P =_\Omega Q$ and $R_i =_\Omega S_i$ for $i \leq m$. $\square$

The proposition 2.2.2.8 gives that the β-reduction commutes with the complete left reduction so we have the following result.

Corollary 4.2.1.5. *If M and N are deterministic terms and for all $d \in \mathbb{N}$ we have*

$$M \twoheadrightarrow_\beta \cdot \sim_d \cdot =_\Omega \cdot \sim_d \cdot \twoheadleftarrow_\beta N$$

then for all $k \in \mathbb{N}$ and $d \in \mathbb{N}$ we have

$$\mathrm{L}^k(M) \twoheadrightarrow_\beta \cdot \sim_d \cdot =_\Omega \cdot \sim_d \cdot \twoheadleftarrow_\beta \mathrm{L}^k(N).$$

Proof. For all d and k, if

$$M \twoheadrightarrow_\beta \cdot \sim_{d+k} \cdot =_\Omega \cdot \sim_{d+k} \cdot \twoheadleftarrow_\beta N$$

then

$$\mathrm{L}^k(M) \twoheadrightarrow_\beta \cdot \sim_d \cdot =_\Omega \cdot \sim_d \cdot \twoheadleftarrow_\beta \mathrm{L}^k(N).$$

$\square$

Next we can see that the relations $\sim_d$ and $=_\Omega$ preserve local Böhm trees.

Proposition 4.2.1.6. *For all $d \in \mathbb{N}$, given two deterministic terms M and N such that $M \sim_d N$ we have*

$$\mathrm{bt}_d(M) = \mathrm{bt}_d(N) \text{ and } \mathrm{bt}^\eta_d(M) = \mathrm{bt}^\eta_d(N).$$

Given two canonical terms M and N such that $M \sim_d N$ we have

$$\mathrm{pt}_d(M) = \mathrm{pt}_d(N) \text{ and } \mathrm{pt}^\eta_d(M) = \mathrm{pt}^\eta_d(N).$$

Proof. By a simple induction on the structure of the terms. For instance for canonical terms such that $M \sim_{d+1} N$ we have

- $M = M_1 +_p M_2$ and $N = N_1 +_p N_2$ with $M_i \sim_{d+1} N_i$ for $i \in \{1; 2\}$ and we get the result by induction hypothesis;
- or $M = \lambda x_1 ... x_n . y \ P_1 \ ... \ P_m$ and $N = \lambda x_1 ... x_n . y \ Q_1 \ ... \ Q_m$ with $P_i \sim_d Q_i$ for $i \leq m$ and we apply the induction hypothesis to the P_i's and Q_i's;
- or $M = \lambda x_1 ... x_n .(\lambda y. M_0) \ P_1 \ ... \ P_m$ and $N = \lambda x_1 ... x_n .(\lambda y. N_0) \ Q_1 \ ... \ Q_m$ with $m > 0$ in which case $\mathrm{pt}_d(M) = \mathrm{pt}_d(N) = 0$ and $\mathrm{pt}^\eta_d(M) = \mathrm{pt}^\eta_d(N) = 0$.

□

Proposition 4.2.1.7. *Given deterministic terms M and N such that $M =_\Omega N$ we have for all $d \in \mathbb{N}$*

$$\mathrm{bt}_d(M) = \mathrm{bt}_d(N) \text{ and } \mathrm{bt}^\eta_d(M) = \mathrm{bt}^\eta_d(N).$$

Proof. By a very simple induction on the terms. □

We have all the results we need to conclude.

Proposition 4.2.1.8. *Given two deterministic terms M and N, if*

$$\forall d \in \mathbb{N}, M \twoheadrightarrow_\beta \cdot \sim_d \cdot =_\Omega \cdot \sim_d \cdot \twoheadleftarrow_\beta N$$

then $M =_{\mathcal{B}} N$.

Proof. We know that given $d \in \mathbb{N}$, for k big enough we have $BT_d(M) = \mathrm{bt}_d(\mathrm{L}^k(M))$ and $BT_d(N) = \mathrm{bt}_d(\mathrm{L}^k(N))$. Then there are terms P and Q such that

$$\mathrm{L}^k(M) \twoheadrightarrow_\beta P \sim_d \cdot =_\Omega \cdot \sim_d Q \twoheadleftarrow_\beta \mathrm{L}^k(N).$$

According to the previous results $\mathrm{bt}_d(P) = \mathrm{bt}_d(Q)$. Besides we have

$$BT_d(M) = \mathrm{bt}_d(\mathrm{L}^k(M)) \preceq_d \mathrm{bt}_d(P) \preceq_d BT_d(P) = BT_d(M)$$

so $\mathrm{bt}_d(P) = BT_d(M)$ and similarly $\mathrm{bt}_d(Q) = BT_d(N)$. □

Corollary 4.2.1.9. *For deterministic terms M and N,*

$$M =_{\mathcal{B}} N \text{ if and only if } \forall d \in \mathbb{N}, M \twoheadrightarrow_{\beta} \cdot \sim_d \cdot =_{\Omega} \cdot \sim_d \cdot \twoheadleftarrow_{\beta} N.$$

Theorem 4.2.1.10. *The relation $=_{\mathcal{B}}$ is a λ-theory.*

Proof. We mentioned before it is an equivalence stable by β, and we want to prove it is contextual.

Given a context C, we can prove by an easy induction on C that if $M \sim_d N$ then $C[M] \sim_d C[N]$. Moreover the β-reduction and the equality $=_{\Omega}$ are both stable by context so if

$$M \twoheadrightarrow_{\beta} \cdot \sim_d \cdot =_{\Omega} \cdot \sim_d \cdot \twoheadleftarrow_{\beta} N$$

then

$$C[M] \twoheadrightarrow_{\beta} \cdot \sim_d \cdot =_{\Omega} \cdot \sim_d \cdot \twoheadleftarrow_{\beta} C[N].$$

□

Generalizing this result to probabilistic trees is very simple. The proof has exactly the same structure, and the only part we need to change is the relation $=_{\Omega}$, as the notion of sensibility is quite different in the probabilistic case. We use instead the relation $=^{s,c}_{\Omega,\epsilon}$.

Proposition 4.2.1.11. *Given M and N such that $PT_d(M) = PT_d(N)$, for all $\epsilon > 0$ we have*

$$M \twoheadrightarrow_{\beta/\equiv} \cdot \sim_d \cdot =^{s,c}_{\Omega,\epsilon} \cdot \sim_d \cdot \twoheadleftarrow_{\beta/\equiv} N.$$

Proof. We reason by induction on d. If $d = 0$ the result is immediate.

Otherwise if $h = \lambda x_1 ... x_n.y\ P_1\ ...\ P_m$ and $h' = \lambda x_1 ... x_n.y\ Q_1\ ...\ Q_m$ are head normal values such that $VT_{d+1}(h) = VT_{d+1}(h')$ then for all $i \leq m$ we have $PT_d(P_i) = PT_d(Q_i)$ so by induction hypothesis there are terms P'_i, P''_i, Q'_i and Q''_i such that

$$P_i \twoheadrightarrow_{\beta/\equiv} P'_i \sim_d P''_i =^{s,c}_{\Omega,\epsilon} Q''_i \sim_d Q'_i \twoheadleftarrow_{\beta/\equiv} Q_i$$

and we have

$$\begin{aligned} h \twoheadrightarrow_{\beta/\equiv} & \lambda \overrightarrow{x}_{1...n}.y\ \overrightarrow{P'}_{1...m} \\ \sim_{d+1} & \lambda \overrightarrow{x}_{1...n}.y\ \overrightarrow{P''}_{1...m} \\ =^{s,v}_{\Omega,\epsilon} & \lambda \overrightarrow{x}_{1...n}.y\ \overrightarrow{Q''}_{1...m} \\ \sim_{d+1} & \lambda \overrightarrow{x}_{1...n}.y\ \overrightarrow{Q'}_{1...m} \twoheadleftarrow_{\beta/\equiv} h'. \end{aligned}$$

In general, given any term M and $\epsilon > 0$ we can find finitely many value trees $t_1, ..., t_k$ such that $\sum_{t \in \mathcal{VT}_{d+1} \setminus \{t_1,...,t_k\}} PT_{d+1}(M)(t) \leq \epsilon$. Then we have a reduction

$$M \twoheadrightarrow_{\beta/\equiv} \sum_{i=1}^{k} \sum_{j} p_{i,j}.h_{i,j} + \left(1 - \sum_{i=1}^{k} \sum_{j} p_{i,j}\right).M'$$

such that the $h_{i,j}$'s are head normal values and for all $i \leq k$ and all j we have $VT_{d+1}(h_{i,j}) = t_i$, and such that for all $i \leq k$, $\sum_j p_{i,j} = (1-\epsilon)PT_{d+1}(M)(t_i)$.

Now if M and N are terms such that $PT_{d+1}(M) = PT_{d+1}(N)$, for all $\epsilon > 0$ we can find such value trees $t_1,\ldots t_k$, and we can get

$$M \twoheadrightarrow_{\beta/\equiv} \sum_{i=1}^{k}\sum_{j} p_{i,j}.h_{i,j} + \left(1 - \sum_{i=1}^{k}\sum_{j} p_{i,j}\right).M'$$
$$N \twoheadrightarrow_{\beta/\equiv} \sum_{i=1}^{k}\sum_{j} p_{i,j}.h'_{i,j} + \left(1 - \sum_{i=1}^{k}\sum_{j} p_{i,j}\right).N'$$

with $VT_{d+1}(h_{i,j}) = VT_{d+1}(h'_{i,j}) = t_i$ for all $i \leq k$ and all j, and with $\sum_j p_{i,j} = (1-\epsilon)PT_{d+1}(M)(t_i)$ for all $i \leq k$.

Then we have

$$\begin{aligned}\left(1 - \sum_{i=1}^{k}\sum_{j} p_{i,j}\right)\mathcal{P}_{\Downarrow}(M') &= \sum_{t \in \mathcal{VT}_{d+1}\setminus\{t_1,\ldots,t_k\}} PT_{d+1}(M)(t) + \epsilon\sum_{i=1}^{k} PT_{d+1}(M)(t_i) \\ &\leq 2\epsilon \\ \left(1 - \sum_{i=1}^{k}\sum_{j} p_{i,j}\right)\mathcal{P}_{\Downarrow}(N') &= \sum_{t \in \mathcal{VT}_{d+1}\setminus\{t_1,\ldots,t_k\}} PT_{d+1}(N)(t) + \epsilon\sum_{i=1}^{k} PT_{d+1}(N)(t_i) \\ &\leq 2\epsilon.\end{aligned}$$

Besides we proved that for all $i \leq k$ and all j we have

$$h_{i,j} \twoheadrightarrow_{\beta/\equiv} \cdot \sim_{d+1} \cdot =^{s,v}_{\Omega,2\epsilon} \cdot \sim_{d+1} \cdot \twoheadleftarrow_{\beta/\equiv} h'_{i,j}.$$

Hence

$$M \twoheadrightarrow_{\beta/\equiv} \cdot \sim_{d+1} \cdot =^{s,c}_{\Omega,2\epsilon} \cdot \sim_{d+1} \cdot \twoheadleftarrow_{\beta/\equiv} N.$$

□

Now as before we check that the relation $=^{s,c}_{\Omega,\epsilon}$ is preserved by complete left reduction.

Proposition 4.2.1.12. *Given $\epsilon, \epsilon', \delta \geq 0$, given probabilistic terms M, N, P and Q such that $M =^{s}_{\Omega,\epsilon,\epsilon'} N$ and $P =^{s}_{\Omega,\delta} Q$, we have $M\left[{}^{P}/_{x}\right] =^{s}_{\Omega,\epsilon+\delta,\epsilon'+\delta} N\left[{}^{Q}/_{x}\right]$.*

Proof. We reason by induction on $=^{s}_{\Omega,\epsilon,\epsilon'}$.

- If $M =^{s}_{\Omega,\epsilon,\epsilon'} N$ with $\mathcal{P}_{\Downarrow}(M) \leq \epsilon'$ and $\mathcal{P}_{\Downarrow}(N) \leq \epsilon'$ then $\mathcal{P}_{\Downarrow}\left(M\left[{}^{P}/_{x}\right]\right) \leq \epsilon'$ and $\mathcal{P}_{\Downarrow}\left(N\left[{}^{Q}/_{x}\right]\right) \leq \epsilon'$.
- If $x =^{s}_{\Omega,\epsilon,\epsilon'} x$ then $P =^{s}_{\Omega,\delta,\delta} Q$ so $P =^{s}_{\Omega,\epsilon+\delta,\epsilon'+\delta} Q$.
- If $y =^{s}_{\Omega,\epsilon,\epsilon'} y$ with $y \neq x$ then $y =^{s}_{\Omega,\epsilon+\delta,\epsilon'+\delta} y$.

- If $\lambda y.M =^{s}_{\Omega,\epsilon,\epsilon'} \lambda y.N$ with $M =^{s}_{\Omega,\epsilon,\epsilon'} N$ then by induction hypothesis $M\left[{}^{P}/_{x}\right] =^{s}_{\Omega,\epsilon+\delta,\epsilon'+\delta} N\left[{}^{Q}/_{x}\right]$ so

$$\lambda y.\left(M\left[{}^{P}/_{x}\right]\right) =^{s}_{\Omega,\epsilon+\delta,\epsilon'+\delta} \lambda y.\left(N\left[{}^{Q}/_{x}\right]\right).$$

- If $M\ M' =^{s}_{\Omega,\epsilon,\epsilon'} N\ N'$ with $M =^{s}_{\Omega,\epsilon,\epsilon'} N$ and $M' =^{s}_{\Omega,\epsilon,\epsilon} N'$ then by induction hypothesis $M\left[{}^{P}/_{x}\right] =^{s}_{\Omega,\epsilon+\delta,\epsilon'+\delta} N\left[{}^{Q}/_{x}\right]$ and $M'\left[{}^{P}/_{x}\right] =^{s}_{\Omega,\epsilon+\delta,\epsilon+\delta} N'\left[{}^{Q}/_{x}\right]$ so

$$\left(M\left[{}^{P}/_{x}\right]\right)\ \left(M'\left[{}^{P}/_{x}\right]\right) =^{s}_{\Omega,\epsilon+\delta,\epsilon'+\delta} \left(N\left[{}^{Q}/_{x}\right]\right)\ \left(N'\left[{}^{Q}/_{x}\right]\right).$$

- If $M_1 +_p M_2 =^{s}_{\Omega,\epsilon,p\epsilon_1+(1-p)\epsilon_2} N_1 +_p N_2$ with $M_i =^{s}_{\Omega,\epsilon,\epsilon_i} N_i$ for $i \in \{1;2\}$ then again the result is immediate by induction hypothesis.

□

Corollary 4.2.1.13. *Given values v and w and canonical terms P and Q, if $v =^{s,v}_{\Omega,\epsilon} w$ and $P =^{s,c}_{\Omega,\epsilon} Q$ then $v\left[{}^{P}/_{x}\right] =^{s}_{\Omega,2\epsilon} w\left[{}^{Q}/_{x}\right]$.*

Proposition 4.2.1.14. *If $M =^{s,c}_{\Omega,\epsilon,\epsilon'} N$ then $\mathrm{L}(M) =^{s,c}_{\Omega,2\epsilon,\epsilon+\epsilon'} \mathrm{L}(N)$.*

Proof. We prove this result and that if $v =^{s,v}_{\Omega,\epsilon} w$ then $\mathrm{L}(v) =^{s,c}_{\Omega,2\epsilon,2\epsilon} \mathrm{L}(w)$ by induction on $M =^{s,c}_{\Omega,\epsilon,\epsilon'} N$.

- If $\mathcal{P}_{\Downarrow}(M), \mathcal{P}_{\Downarrow}(N) \leq \epsilon'$ then $\mathcal{P}_{\Downarrow}(\mathrm{L}(M)), \mathcal{P}_{\Downarrow}(\mathrm{L}(N)) \leq \epsilon'$.

- If $M = M_1 +_p M_2$ and $N = N_1 +_p N_2$ with $M_i =^{s,c}_{\Omega,\epsilon,\epsilon_i} N_i$ for $i \in \{1;2\}$ and $p\epsilon_1 + (1-p)\epsilon_2 = \epsilon'$ then by induction hypothesis $\mathrm{L}(M_i) =^{s,c}_{\Omega,2\epsilon,\epsilon+\epsilon_i} \mathrm{L}(N_i)$ for $i \in \{1;2\}$ and $\mathrm{L}(M) =^{s,c}_{\Omega,2\epsilon,\epsilon+p\epsilon_1+(1-p)\epsilon_2} \mathrm{L}(N_i)$.

- If M and N are values such that $M =^{s,v}_{\Omega,\epsilon} N$ then by induction hypothesis $\mathrm{L}(M) =^{s,v}_{\Omega,2\epsilon} \mathrm{L}(N)$ and $\mathrm{L}(M) =^{s,c}_{\Omega,2\epsilon,\epsilon+\epsilon'} \mathrm{L}(N)$.

- If

$$v = \lambda x_1...x_n.y\ P_1\ ...\ P_m =^{s,v}_{\Omega,\epsilon} \lambda x_1...x_n.y\ Q_1\ ...\ Q_m = w$$

with $P_i =^{s,c}_{\Omega,\epsilon} Q_i$ for $i \leq m$ then by induction hypothesis $\mathrm{L}(P_i) =^{s,c}_{\Omega,2\epsilon,2\epsilon} \mathrm{L}(Q_i)$ for $i \leq m$ and $\mathrm{L}(v) =^{s,v}_{\Omega,2\epsilon} \mathrm{L}(w)$.

- If

$$v = \lambda x_1...x_n.(\lambda y.v_0)\ P\ R_1\ ...\ R_m =^{s,v}_{\Omega,\epsilon} \lambda x_1...x_n.(\lambda y.w_0)\ Q\ S_1\ ...\ S_m = w$$

with $v_0 =^{s,v}_{\Omega,\epsilon} w_0$, $P =^{s,c}_{\Omega,\epsilon Q}$ and $R_i =^{s,c}_{\Omega,\epsilon} S_i$ for $i \leq m$ then

$$\mathrm{L}(v) = \mathrm{can}\left(\lambda x_1...x_n.v_0\left[{}^{P}/_{y}\right]\ R_1\ ...\ R_m\right)$$
$$\mathrm{L}(w) = \left(\lambda x_1...x_n.w_0\left[{}^{Q}/_{y}\right]\ S_1\ ...\ S_m\right).$$

The previous corollary gives $v' \left[{}^{P_1}/_{y}\right] =^{s}_{\Omega,2\epsilon} w' \left[{}^{Q_1}/_{y}\right]$, thus

$$\lambda x_1 \ldots x_n . v' \left[{}^{P_1}/_{y}\right] \; P_2 \; \ldots \; P_m =^{s}_{\Omega,2\epsilon} \lambda x_1 \ldots x_n . w' \left[{}^{Q_1}/_{y}\right] \; Q_2 \; \ldots \; Q_m .$$

Then the proposition 3.2.2.7 gives $\mathrm{L}(v) =^{s,c}_{\Omega,2\epsilon} \mathrm{L}(w)$.

□

Corollary 4.2.1.15. *If $M =^{s,c}_{\Omega,\epsilon} N$ then $\mathrm{L}(M) =^{s,c}_{\Omega,2\epsilon} \mathrm{L}(N)$.*

Corollary 4.2.1.16. *Given two terms M and N, if for all $d \in \mathbb{N}$ and $\epsilon > 0$ we have*

$$M \twoheadrightarrow_{\beta/\equiv} \cdot \sim_d \cdot =^{s,c}_{\Omega,\epsilon} \cdot \sim_d \cdot \twoheadleftarrow_{\beta/\equiv} N$$

then for all $k \in \mathbb{N}$, $d \in \mathbb{N}$ and $\epsilon > 0$ we have

$$\mathrm{L}^k(M) \twoheadrightarrow_{\beta/\equiv} \cdot \sim_d \cdot =^{s,c}_{\Omega,\epsilon} \cdot \sim_d \cdot \twoheadleftarrow_{\beta/\equiv} \mathrm{L}^k(N).$$

The relation $=^{s,c}_{\Omega,\epsilon}$ does not preserve local Böhm trees, but we can relate it to the approximate ordering $\preceq_{d,\epsilon}$.

Proposition 4.2.1.17. *Given canonical terms M and N and $\epsilon \geq 0$, if $M =^{s,c}_{\Omega,\epsilon} N$ then for all $d \in \mathbb{N}$, $\mathrm{pt}_d(M) \preceq_{d,\epsilon} \mathrm{pt}_d(N)$.*

Proof. We reason by induction on d, immediate when $d = 0$.

Given two values v and w with $v =^{s,v}_{\Omega,\epsilon} w$, if v or w is head normal then they are both head normal, so we can write $v = \lambda x_1 \ldots x_n . y \; P_1 \; \ldots \; P_m$ and $w = \lambda x_1 \ldots x_n . y \; Q_1 \; \ldots \; Q_m$ with $P_i =^{s,c}_{\Omega,\epsilon} Q_i$ for all $i \leq m$. Then by induction hypothesis $\mathrm{pt}_d(P_i) \preceq_{d,\epsilon} \mathrm{pt}_d(Q_i)$ so $\mathrm{vt}_{d+1}(v) \preceq^{v}_{d+1,\epsilon} \mathrm{vt}_{d+1}(w)$.

Now if $M =^{s,c}_{\Omega,\epsilon} N$ we have

$$M \equiv_{\mathrm{syn}} \sum_i p_i . v_i + \left(1 - \sum_i p_i\right) . M'$$
$$N \equiv_{\mathrm{syn}} \sum_i p_i . w_i + \left(1 - \sum_i p_i\right) . N'$$

with $(1 - \sum_i p_i) \mathcal{P}_{\Downarrow}(M'), (1 - \sum_i p_i) \mathcal{P}_{\Downarrow}(N') \leq \epsilon$ and $v_i =^{s,v}_{\Omega,\epsilon} w_i$ for all i.

Then for all $A \subset \mathcal{VT}_{d+1}$, if v_i is a head normal form and $\mathrm{vt}_{d+1}(v_i) \in\uparrow A$ then from $\mathrm{vt}_{d+1}(v_i) \preceq^{v}_{d,\epsilon} \mathrm{vt}_{d+1}(w_i)$ we deduce $\mathrm{vt}_{d+1}(w_i) \in\uparrow_\epsilon A$. Thus

$$\begin{aligned} \sum_{t \in \uparrow A} \mathrm{pt}_{d+1}(M)(t) &\leq \sum_{i \text{ s.t.} \mathrm{vt}_{d+1}(v_i) \in \uparrow A} p_i + \epsilon \\ &\leq \sum_{i \text{ s.t.} \mathrm{vt}_{d+1}(w_i) \in \uparrow_\epsilon A} p_i + \epsilon \\ &\leq \sum_{t \in \uparrow_\epsilon A} \mathrm{pt}_{d+1}(N)(t) + \epsilon. \end{aligned}$$

□

With these results we can prove that our relation does describe the probabilistic Böhm tree equality.

Proposition 4.2.1.18. *If for all $d \in \mathbb{N}$ and $\epsilon > 0$*

$$M \twoheadrightarrow_{\beta/\equiv} \cdot \sim_d \cdot =^{s,c}_{\Omega,\epsilon} \cdot \sim_d \cdot \twoheadleftarrow_{\beta/\equiv} N$$

then $M =_{\mathcal{PB}} N$.

Proof. We know that for all $d \in \mathbb{N}$ and all $\epsilon > 0$ there is $k \in \mathbb{N}$ such that $PT_d(M) \preceq_{d,\epsilon} \mathrm{pt}_d(\mathrm{L}^k(M))$. Then we know there are terms Q, Q', R and R' such that

$$\mathrm{L}^k(M) \twoheadrightarrow_{\beta/\equiv} Q \sim_d R =^{s,c}_{\Omega,\epsilon} R' \sim_d Q' \twoheadleftarrow_{\beta/\equiv} \mathrm{L}^k(N).$$

We proved that $\mathrm{pt}_d(Q) = \mathrm{pt}_d(R)$ and $\mathrm{pt}_d(R') = \mathrm{pt}_d(Q')$, and the previous result gives that $\mathrm{pt}_d(R) \preceq_{d,\epsilon} \mathrm{pt}_d(R')$. Besides we know that $\mathrm{pt}_d(\mathrm{L}^k(M)) \preceq_d \mathrm{pt}_d(Q)$ and $\mathrm{pt}_d(Q') \preceq_d PT_d(Q') = PT_d(N)$. To sum up:

$$\begin{aligned} PT_d(M) \preceq_{d,\epsilon} &\mathrm{pt}_d(\mathrm{L}^k(M)) \preceq_d \mathrm{pt}_d(Q) = \mathrm{pt}_d(R) \\ &\preceq_{d,\epsilon} \mathrm{pt}_d(R') = \mathrm{pt}_d(Q') \preceq_d PT_d(Q') = PT_d(N). \end{aligned}$$

For all $d \in \mathbb{N}$ and all $\epsilon > 0$, we have $PT_d(M) \preceq_{d,2\epsilon} PT_d(N)$, so we have $PT_d(M) \preceq_d PT_d(N)$. By symmetry $PT_d(N) \preceq_d PT_d(M)$, and by antisymmetry $PT_d(M) = PT_d(N)$. □

Theorem 4.2.1.19. *The relation $=_{\mathcal{PB}}$ is a λ-theory.*

Proof. We proved that $M =_{\mathcal{PB}} N$ if and only if

$$\forall d \in \mathbb{N}, \forall \epsilon > 0, M \twoheadrightarrow_{\beta/\equiv} \cdot \sim_d \cdot =^{s,c}_{\Omega,\epsilon} \cdot \sim_d \cdot \twoheadleftarrow_{\beta/\equiv} N$$

and this relation is contextual. □

4.2.2 Introducing infinite extensionality

The technique we used to prove the contextuality of $=_{\mathcal{B}}$ and $=_{\mathcal{PB}}$ can be adapted to the infinitely extensional cases, but we need to use more complex and less straightforward relations. For that reason we will not deal with the deterministic case first, but we will directly consider the probabilistic trees.

To compare extensional trees it may be natural to use the η-expansion: two Böhm trees of finite depth are equal up to infinite extensionality if they have a common η-expansion. But we do not necessarily have $\mathrm{pt}^\eta_d(M) \preceq^\eta_d \mathrm{pt}^\eta_d(\lambda x.M\ x)$, so this relation does not fit well in our technique. The η-contraction is better behaved, as we do have $\mathrm{pt}^\eta_d(\lambda x.M\ x) \preceq^\eta_d \mathrm{pt}^\eta_d(M)$.

We actually use a reduction which mixes the β-reduction modulo $\equiv$, the relation $\sim_d$ and the η-contraction.

Definition 4.2.2.1. The relation $\to_{\eta_d}$ is defined inductively for $d \in \mathbb{N}$ by

$$\frac{}{M \to_{\eta_0} N} \qquad \frac{\forall \epsilon > 0, M \twoheadrightarrow_{\beta/\equiv} M_0 +_{1-\epsilon} M' \text{ with } M_0 \to_{\eta_{d+1}} N}{M \to_{\eta_{d+1}} N}$$

$$\frac{M_1 \to_{\eta_{d+1}} N \qquad M_2 \to_{\eta_{d+1}} N}{M_1 +_p M_2 \to_{\eta_{d+1}} N}$$

$$\frac{M \to_{\eta_{d+1}} N \qquad \forall i, z_i \notin \mathrm{FV}(M) \cup \mathrm{FV}(N) \cup \{z_j \mid j \neq i\} \qquad \forall i, PT^\eta_d(Z_i) = z_i}{\lambda \overrightarrow{z}_{1\ldots k}.M\ \overrightarrow{Z}_{1\ldots k} \to_{\eta_{d+1}} N}$$

$$\frac{}{x \to_{\eta_{d+1}} x} \qquad \frac{M \to_{\eta_{d+1}} N}{\lambda x.M \to_{\eta_{d+1}} \lambda x.N}$$

$$\frac{M \to_{\eta_{d+1}} N \qquad M' \to_{\eta_d} N'}{M\ M' \to_{\eta_{d+1}} N\ N'} \qquad \frac{M_1 \to_{\eta_{d+1}} N_1 \qquad M_2 \to_{\eta_{d+1}} N_2}{M_1 +_p M_2 \to_{\eta_{d+1}} N_1 +_p N_2}$$

Proposition 4.2.2.1. *If $M \to_{\eta_{d+1}} N$ then $M \to_{\eta_d} N$.*

Proof. Immediate by induction on $M \to_{\eta_d} N$. □

Proposition 4.2.2.2. *If $M \twoheadrightarrow_{\beta/\equiv} \cdot \to_{\eta_d} N$ then $M \to_{\eta_d} N$.*

Proof. If $M \twoheadrightarrow_{\beta/\equiv} P$ with $P \to_{\eta_d} N$ then for all $\epsilon > 0$, $M \twoheadrightarrow_{\beta/\equiv} P +_{1-\epsilon} P$. □

Proposition 4.2.2.3. *If $PT^\eta_d(X) = x$ then $X \to_{\eta_d} x$.*

Proof. If $d = 0$ then $X \to_{\eta_0} x$.

Otherwise if h is a head normal value with $VT^\eta_{d+1}(h) = x$ then $h = \lambda \overrightarrow{z}_{1\ldots k}.x\ \overrightarrow{Z}_{1\ldots k}$ where the variables z_i are pairwise distinct and distinct from x, and $PT^\eta_d(Z_i) = z_i$ for all $i \leq k$. Hence $h \to_{\eta_{d+1}} x$.

Now if $PT^\eta_{d+1}(X) = x$ then for all $\epsilon > 0$ we have $X \twoheadrightarrow_{\beta/\equiv} \sum_i p_i.h_i +_{1-\epsilon} X'$ with $VT^\eta_{d+1}(h_i) = x$ for all i, hence $\sum_i p_i.h_i \to_{\eta_{d+1}} x$. □

To characterize the infinitely extensional probabilistic Böhm tree equality it appears that the usual η-contraction is not enough. Indeed the η-contraction involves the notion of free variables. Given a term and a variable, we can ask whether the variable is free in the term, but we can also ask whether it is free in its Böhm tree, and the two answers do not necessarily coincide. A very simple example is $M = (\lambda y.z)\ x$: the variable x is free in M but $M \to_\beta z$ and x is not free in the Böhm tree of M. A more subtle example is $M = \Theta\ (\lambda f.\lambda y.z\ (f\ y))\ x$: we have $M \twoheadrightarrow_\beta z\ M$ and x is free in every reduct of M, but it is not free in its Böhm tree. What happens is that when we reduce M the variable x is being pushed down, so it is always free but it appears at infinitely increasing depth. With probabilities a third case arises, where the variable is not pushed down but appears with a decreasing probability: if $M = \Theta\ (\lambda f.\lambda y.\underline{0} +_{\frac{1}{2}} f\ y)\ x$ then $M \twoheadrightarrow_{\beta^c} \underline{0} +_{\frac{1}{2}} M$ and x is always free at a constant depth in the reducts of M but not in its Böhm tree.

Definition 4.2.2.2. The sets $\mathrm{FV}_{\mathcal{B}}(T)$ of free variables in a tree $T \in \mathcal{PT}^{\eta}_{d}$ and $\mathrm{FV}^{v}_{\mathcal{B}}(t)$ of free variables in $t \in \mathcal{VT}^{\eta}_{d+1}$ are given by

$$\mathrm{FV}_{\mathcal{B}}(T) = \bigcup_{t \text{ s.t. } T(t)\neq 0} \mathrm{FV}^{v}_{\mathcal{B}}(t)$$

$$\mathrm{FV}^{v}_{\mathcal{B}}(y,(T_n)) = \left(\{y\} \cup \bigcup_{n\geq 1} \mathrm{FV}_{\mathcal{B}}(T_n)\right) \setminus \{x_{d,1}, x_{d,2}, ...\}.$$

The set of Böhm free variables at depth d of a term M is $\mathrm{FV}^{d}_{\mathcal{B}}(M) = \mathrm{FV}_{\mathcal{B}}(PT^{\eta}_{d}(M))$, and we write $\mathrm{FV}_{\mathcal{B}}(M) = \bigcup_{d\in\mathbb{N}} \mathrm{FV}^{d}_{\mathcal{B}}(M)$.

This notion of free variables has good properties. We can prove some results which support the intuition that variables which are not free in a term are irrelevant.

Proposition 4.2.2.4. *If M is canonical and $x \notin \mathrm{FV}_{\mathcal{B}}(M)$ then for all P we have*

$$\mathrm{L}\left(\mathrm{can}\left(M\left[{}^{P}/_{x}\right]\right)\right) =_{+} \mathrm{L}(M)\left[{}^{P}/_{x}\right].$$

Proof. By a simple induction on M. □

Proposition 4.2.2.5. *For all $d \in \mathbb{N}$, if $x \notin \mathrm{FV}^{d}_{\mathcal{B}}(M)$ then for any term P,*

$$\mathrm{pt}^{\eta}_{d}(M) = \mathrm{pt}^{\eta}_{d}\left(\mathrm{can}\left(M\left[{}^{P}/_{x}\right]\right)\right).$$

Proof. To simplify we write $\mathrm{pt}^{\eta}_{d}\left(M\left[{}^{P}/_{x}\right]\right)$ for $\mathrm{pt}^{\eta}_{d}\left(\mathrm{can}\left(M\left[{}^{P}/_{x}\right]\right)\right)$.

We reason by induction on d. If $d = 0$ the result is immediate.

Otherwise if M is a value but it is not in head normal form then we have $\mathrm{pt}^{\eta}_{d+1}(M) = \mathrm{pt}^{\eta}_{d+1}\left(M\left[{}^{P}/_{x}\right]\right) = 0$.

If M is a head normal value $h = \lambda x_1 ... x_n.y\ Q_1\ ...\ Q_m$ then $y \neq x$ and $x \notin \mathrm{FV}^{d}_{\mathcal{B}}(Q_i)$ for $i \leq m$, so by induction hypothesis $\mathrm{pt}^{\eta}_{d}(Q_i) = \mathrm{pt}^{\eta}_{d}\left(Q_i\left[{}^{P}/_{x}\right]\right)$ and $\mathrm{can}\left(h\left[{}^{P}/_{x}\right]\right)$ is a head normal value with $\mathrm{vt}^{\eta}_{d+1}(h) = \mathrm{vt}^{\eta}_{d+1}\left(h\left[{}^{P}/_{x}\right]\right)$, hence $\mathrm{pt}^{\eta}_{d+1}(M) = \mathrm{pt}^{\eta}_{d+1}\left(M\left[{}^{P}/_{x}\right]\right)$.

Finally if $M \equiv \sum_i p_i.v_i$ then

$$\mathrm{pt}^{\eta}_{d+1}(M) = \sum_i p_i.\mathrm{pt}^{\eta}_{d+1}(v_i) = \sum_i p_i.\mathrm{pt}^{\eta}_{d+1}\left(v_i\left[{}^{P}/_{x}\right]\right) = \mathrm{pt}^{\eta}_{d+1}\left(M\left[{}^{P}/_{x}\right]\right).$$

□

Corollary 4.2.2.6. *For all $d \in \mathbb{N}$, if $x \notin \mathrm{FV}^{d}_{\mathcal{B}}(M)$ then for any term P,*

$$PT^{\eta}_{d}(M) = PT^{\eta}_{d}\left(M\left[{}^{P}/_{x}\right]\right)$$

.

Proof. If $x \notin \mathrm{FV}^d_{\mathcal{B}}(M)$ then $\mathrm{L}\left(M\left[{}^{P}/_{x}\right]\right) = \mathrm{L}(M)\left[{}^{P}/_{x}\right]$. But we know that $PT^\eta_d(\mathrm{L}(M)) = PT^\eta_d(M)$ so $x \notin \mathrm{FV}^d_{\mathcal{B}}(\mathrm{L}(M))$, and we get by induction that for all $k \in \mathbb{N}$, $\mathrm{L}^k\left(M\left[{}^{P}/_{x}\right]\right) = \mathrm{L}^k(M)\left[{}^{P}/_{x}\right]$. Then $\mathrm{pt}^\eta_d(\mathrm{L}^k(M)) = \mathrm{pt}^\eta_d\left(\mathrm{L}^k\left(M\left[{}^{P}/_{x}\right]\right)\right)$, hence $PT^\eta_d(M) = PT^\eta_d\left(M\left[{}^{P}/_{x}\right]\right)$. $\square$

Proposition 4.2.2.7. *If $x \notin \mathrm{FV}_{\mathcal{B}}(M)$ then for any context C which does not contain x, we have $x \notin \mathrm{FV}_{\mathcal{B}}(C[M])$.*

Proof. We based the definition of the Böhm free variables of a term on its infinitely extensional Böhm tree, but we can observe that we can give an equivalent definition on non extensional Böhm trees. In particular we have that if $M =_{\mathcal{PB}} N$ then $\mathrm{FV}_{\mathcal{B}}(M) = \mathrm{FV}_{\mathcal{B}}(N)$, and if $x \notin \mathrm{FV}_{\mathcal{B}}(M)$ then for all P, $M =_{\mathcal{PB}} M\left[{}^{P}/_{x}\right]$.

Then we can use the contextuality of the non extensional Böhm trees. If $x \notin \mathrm{FV}_{\mathcal{B}}(M)$ then there is M' with $M =_{\mathcal{PB}} M'$ and $x \notin \mathrm{FV}(M')$ (take for instance $M' = M\left[{}^{\lambda x.x}/_{x}\right]$), and then we have both $x \notin \mathrm{FV}(C[M'])$ and $C[M'] =_{\mathcal{PB}} C[M]$, hence $x \notin \mathrm{FV}_{\mathcal{B}}(C[M'])$. $\square$

We call the variables which are free in a term but not in its Böhm tree dummy variables. We just proved that we can substitute any term to a dummy variable without influencing the Böhm tree of a term.

Definition 4.2.2.3. The relation $\equiv_{dum}$ is given by

$$M\left\{{}^{P}/_{x}\right\} \equiv_{dum} M\left\{{}^{Q}/_{x}\right\} \text{ if } x \notin \mathrm{FV}_{\mathcal{B}}(M)$$

closed by transitivity, where $M\left\{{}^{P}/_{x}\right\}$ is the substitution of P for x in M with binding of variables.

Proposition 4.2.2.8. *$\equiv_{dum}$ is contextual.*

Proof. If $M\left\{{}^{P}/_{x}\right\} \equiv_{dum} M\left\{{}^{Q}/_{x}\right\}$ with $x \notin \mathrm{FV}_{\mathcal{B}}(M)$, let C be a context, y a fresh variable, we have $C\left[M\left[{}^{y}/_{x}\right]\right]\left\{{}^{P}/_{y}\right\} = C\left[M\left\{{}^{P}/_{x}\right\}\right]$ and the proposition 4.2.2.7 gives $y \notin \mathrm{FV}_{\mathcal{B}}\left(C\left[M\left[{}^{y}/_{x}\right]\right]\right)$. Then we have

$$C\left[M\left\{{}^{P}/_{x}\right\}\right] = C\left[M\left[{}^{y}/_{x}\right]\right]\left\{{}^{P}/_{y}\right\} \equiv_{dum} C\left[M\left[{}^{y}/_{x}\right]\right]\left\{{}^{Q}/_{y}\right\} = C\left[M\left\{{}^{Q}/_{x}\right\}\right].$$

$\square$

To use the reduction $\to_{\eta_d}$ in conjunction with $\equiv_{dum}$ it is useful to prove that it is stable by substitution.

Proposition 4.2.2.9. *If $M \to_{\eta_d} N$ and $P \to_{\eta_d} Q$ then $M\left[{}^{P}/_{x}\right] \to_{\eta_d} N\left[{}^{Q}/_{x}\right]$.*

Proof. By induction on $M \to_{\eta_d} N$.

- If $M \to_{\eta_0} N$ then $M\left[{}^{P}/_{x}\right] \to_{\eta_0} N\left[{}^{Q}/_{x}\right]$.

- If for all $\epsilon > 0$, $M \twoheadrightarrow_{\beta/\equiv} M_0 +_{1-\epsilon} M'$ with $M_0 \to_{\eta_{d+1}} N$ then the substitutivity of the β-reduction gives $M\left[{}^{P}/_{x}\right] \twoheadrightarrow_{\beta/\equiv} M_0\left[{}^{P}/_{x}\right] +_{1-\epsilon} M'\left[{}^{P}/_{x}\right]$ and by induction hypothesis $M_0\left[{}^{P}/_{x}\right] \to_{\eta_{d+1}} N\left[{}^{Q}/_{x}\right]$. This holds for all $\epsilon > 0$ so $M\left[{}^{P}/_{x}\right] \to_{\eta_{d+1}} N\left[{}^{Q}/_{x}\right]$.
- If $M_1 +_p M_2 \to_{\eta_{d+1}} N$ with $M_i \to_{\eta_{d+1}} N$ for $i \in \{1; 2\}$ then by induction hypothesis $M_i\left[{}^{P}/_{x}\right] \to_{\eta_{d+1}} N\left[{}^{Q}/_{x}\right]$ and $(M_1 +_p M_2)\left[{}^{P}/_{x}\right] \to_{\eta_{d+1}} N\left[{}^{Q}/_{x}\right]$.
- If $\lambda \overrightarrow{z}_{1\ldots k}.M\ \overrightarrow{Z}_{1\ldots k} \to_{\eta_{d+1}} N$ with $M \to_{\eta_{d+1}} N$ then by induction hypothesis $M\left[{}^{P}/_{x}\right] \to_{\eta_{d+1}} N\left[{}^{P}/_{x}\right]$. Besides for all $i \leq k$ we have $z_i \neq x$ so $x \notin \mathrm{FV}^d_{\mathcal{B}}(Z_i)$ and $PT^\eta_d\left(Z_i\left[{}^{P}/_{x}\right]\right) = PT^\eta_d(Z_i) = z_i$.
- If $x \to_{\eta_{d+1}} x$ then $P \to_{\eta_{d+1}} Q$.
- If $y \to_{\eta_{d+1}} y$ with $y \neq x$ then $y\left[{}^{P}/_{x}\right] = y\left[{}^{Q}/_{x}\right] = y$.
- The other cases are immediate by induction hypothesis.

□

We can use these relations to relate terms with the same infinitely extensional probabilistic Böhm tree.

Proposition 4.2.2.10. *If $M =_{\mathcal{PB}^\eta} N$ then for all $d \in \mathbb{N}$ and $\epsilon > 0$ we have*

$$M \twoheadrightarrow_{\beta/\equiv} \cdot \equiv_{dum} \cdot \to_{\eta_d} \cdot =^{s,c}_{\Omega,\epsilon} \cdot \leftarrow_{\eta_d} \cdot \equiv_{dum} \cdot \twoheadleftarrow_{\beta/\equiv} N.$$

Proof. We prove that if M and N are d,ϵ-normal forms and $M =_{\mathcal{PB}^\eta} N$ then

$$M \equiv \cdot \equiv_{dum} \cdot \to_{\eta_d} \cdot =^{s,c}_{\Omega,2\epsilon} \cdot \leftarrow_{\eta_d} \cdot \equiv_{dum} \cdot \equiv N$$

from which we can deduce the proposition.

We reason by induction on d. If $d = 0$ the result is immediate.

Otherwise we can write

$$M \equiv \sum_i p_i.h_i + \left(1 - \sum_i p_i\right).M'$$
$$N \equiv \sum_i p_i.h'_i + \left(1 - \sum_i p_i\right).N'$$

where h_i and h'_i are $d+1,\epsilon$-normal values with $VT^\eta(h_i) = VT^\eta(h'_i)$ for all i, and $(1 - \sum_i p_i)\,\mathcal{P}_\Downarrow(M') \leq 2\epsilon$, $(1 - \sum_i p_i)\,\mathcal{P}_\Downarrow(N') \leq 2\epsilon$.

If h and h' are $d+1,\epsilon$-normal values with $VT^\eta(h) = VT^\eta(h')$ then we have w.l.o.g. $h = \lambda \overrightarrow{x}_{1\ldots n}.y\ \overrightarrow{P}_{1\ldots m}$ and $h' = \lambda \overrightarrow{x}_{1\ldots n+k}.y\ \overrightarrow{Q}_{1\ldots m}\ \overrightarrow{X}_{1\ldots k}$ with

$PT^\eta(P_i) = PT^\eta(Q_i)$ for all $i \leq m$ and $PT^\eta(X_i) = x_{n+i}$ for all $i \leq k$. By induction hypothesis we have

$$P_i \equiv \cdot \equiv_{dum} P'_i \to_{\eta_d} P''_i =^{s,c}_{\Omega,2\epsilon} Q''_i \leftarrow_{\eta_d} Q'_i \equiv_{dum} \cdot \equiv Q_i$$

for all $i \leq m$. Besides we have $x_{n+j} \notin \mathrm{FV}(P_i)$ for all $j \leq k$ and all $i \leq m$, so $x_{n+j} \notin \mathrm{FV}_{\mathcal{B}}(P_i) = \mathrm{FV}_{\mathcal{B}}(Q_i)$. Let $I = \lambda x.x$ (or any closed term), we have $Q'_i \equiv_{dum} Q'_i \left[{}^{I}/x_{n+1}\right] \ldots \left[{}^{I}/x_{n+k}\right]$, the proposition 4.2.2.9 gives

$$Q'_i \left[{}^{I}/x_{n+1}\right] \ldots \left[{}^{I}/x_{n+k}\right] \to_{\eta_d} Q''_i \left[{}^{I}/x_{n+1}\right] \ldots \left[{}^{I}/x_{n+k}\right]$$

and

$$P'_i \left[{}^{I}/x_{n+1}\right] \ldots \left[{}^{I}/x_{n+k}\right] \to_{\eta_d} P''_i \left[{}^{I}/x_{n+1}\right] \ldots \left[{}^{I}/x_{n+k}\right]$$

with according to the proposition 4.2.1.12

$$P''_i \left[{}^{I}/x_{n+1}\right] \ldots \left[{}^{I}/x_{n+k}\right] =^{s,c}_{\Omega,2\epsilon} Q''_i \left[{}^{I}/x_{n+1}\right] \ldots \left[{}^{I}/x_{n+k}\right].$$

This means that for all $j \leq k$ and $i \leq m$ we can assume that x_{n+j} is not free in P'_i, P''_i, Q''_i and Q'_i, hence

$$h \equiv \cdot \equiv_{dum} \cdot \to_{\eta_{d+1}} \lambda \overrightarrow{x}_{1\ldots n}.y\ \overrightarrow{P''}_{1\ldots m}$$
$$h' \equiv \cdot \equiv_{dum} \cdot \to_{\eta_{d+1}} \lambda \overrightarrow{x}_{1\ldots n}.y\ \overrightarrow{Q''}_{1\ldots m}$$

with

$$\lambda \overrightarrow{x}_{1\ldots n}.y\ \overrightarrow{P''}_{1\ldots m} =^{s,v}_{\Omega,2\epsilon} \lambda \overrightarrow{x}_{1\ldots n}.y\ \overrightarrow{Q''}_{1\ldots m}.$$

From there we have

$$M \equiv \cdot \equiv_{dum} \cdot \to_{\eta_d} \cdot =^{s,c}_{\Omega,2\epsilon} \cdot \leftarrow_{\eta_d} \cdot \equiv_{dum} \cdot \equiv N.$$

□

To prove the converse we prove that the relations $\equiv_{dum}$ and $\to_{\eta_d}$ are stable by complete left reduction, and we study their influence on local Böhm trees. First we can check that $\equiv_{dum}$ has absolutely no influence on Böhm trees.

Proposition 4.2.2.11. *If $M \equiv_{dum} N$ then $\mathrm{L}(M) \equiv_{dum} \mathrm{L}(N)$ and for all $d \in \mathbb{N}$, $PT^\eta_d(M) = PT^\eta_d(N)$ and $\mathrm{pt}^\eta_d(M) = \mathrm{pt}^\eta_d(N)$.*

Proof. This is given by the propositions 4.2.2.4 and 4.2.2.5 and the corollary 4.2.2.6. □

To deal with the reduction $\to_{\eta_d}$ we begin by checking that, as most of our relations, it is preserved by canonicalization.

Lemma 4.2.2.12. *If $M \to_{\eta_{d+1}} N_1 +_p N_2$ then for all $\epsilon > 0$ we have*

$$M \twoheadrightarrow_{\beta/\equiv} (M_1 +_p M_2) +_{1-\epsilon} M'$$

with $M_i \to_{\eta_{d+1}} N_i$ *for* $i \in \{1; 2\}$.

Proof. We reason by induction on $M \to_{\eta_{d+1}} N_1 +_p N_2$.

- If for all $\epsilon > 0$ we have $M \twoheadrightarrow_{\beta/\equiv} M_0 +_{1-\epsilon} M'$ with $M_0 \to_{\eta_{d+1}} N_1 +_p N_2$ then by induction hypothesis $M_0 \twoheadrightarrow_{\beta/\equiv} (M_1 +_p M_2) +_{1-\epsilon} M''$ with $M_i \to_{\eta_{d+1}} N_i$ for $i \in \{1; 2\}$, thus

$$M \twoheadrightarrow_{\beta/\equiv} (M_1 +_p M_2) +_{(1-\epsilon)^2} (M' +_{\frac{1}{2-\epsilon}} M'').$$

- If $M_1 +_q M_2 \to_{\eta_{d+1}} N_1 +_p N_2$ with $M_i \to_{\eta_{d+1}} N_1 +_p N_2$ for $i \in \{1; 2\}$ then by induction hypothesis for all $\epsilon > 0$ we have $M_i \twoheadrightarrow_{\beta/\equiv} (M_{i,1} +_p M_{i,2}) +_{1-\epsilon} M'_i$ with $M_{i,j} \to_{\eta_{d+1}} N_j$ for $j \in \{1; 2\}$. Then

$$M_1 +_p M_2 \twoheadrightarrow_{\beta/\equiv} ((M_{1,1} +_q M_{2,1}) +_p (M_{1,2} +_q M_{2,2})) +_{1-\epsilon} (M'_1 +_q M'_2).$$

- If $M_1 +_p M_2 \to_{\eta_{d+1}} N_1 +_p N_2$ with $M_i \to_{\eta_{d+1}} N_i$ for $i \in \{1; 2\}$ then the result is immediate.

□

Proposition 4.2.2.13. *If* $M \to_{\eta_d} N$ *then* $M \to_{\eta_d} \mathrm{can}(N)$.

Proof. We prove that if $M \to_{\eta_d} N \to_+ N'$ then $M \to_{\eta_d} N'$, and the result follows. We reason by induction on $M \to_{\eta_d} N$.

- If $M \to_{\eta_0} N \to_+ N'$ then $M \to_{\eta_0} N'$.
- If for all $\epsilon > 0$ we have $M \twoheadrightarrow_{\beta/\equiv} M_0 +_{1-\epsilon} M'$ with $M_0 \to_{\eta_{d+1}} N$ and $N \to_+ N'$ then by induction hypothesis $M_0 \to_{\eta_{d+1}} N'$, hence $M \to_{\eta_{+1}} N'$.
- If $M_1 +_p M_2 \to_{\eta_{d+1}} N \to_+ N'$ with $M_i \to_{\eta_{d+1}} N_i$ for $i \in \{1; 2\}$ then the result is immediate by induction hypothesis.
- If $\lambda \overrightarrow{z}_{1\ldots k}.M\ \overrightarrow{Z}_{1\ldots k} \to_{\eta_{d+1}} N \to_+ N'$ with $M \to_{\eta_{d+1}} N$ then the result is immediate by induction hypothesis.
- If $x \to_{\eta_{d+1}} x$ then we cannot reduce with $\to_+$.
- If $\lambda x.M \to_{\eta_{d+1}} \lambda x.(N_1 +_p N_2)$ with $M \to_{\eta_{d+1}} N_1 +_p N_2$ then the previous lemma gives $M \twoheadrightarrow_{\beta/\equiv} (M_1 +_p M_2) +_{1-\epsilon} M'$ with $M_i \to_{\eta_{d+1}} N_i$ for $i \in \{1; 2\}$ for all $\epsilon > 0$, thus

$$\lambda x.M \twoheadrightarrow_{\beta/\equiv} (\lambda x.M_1 +_p \lambda x.M_2) +_{1-\epsilon} \lambda x.M'$$

  and $\lambda x.M \to_{\eta_{d+1}} \lambda x.N_1 +_p \lambda x.N_2$.
- Similarly if $M\ M' \to_{\eta_{d+1}} (N_1 +_p N_2)\ N'$ then we use the previous result to get $M\ M' \to_{\eta_{d+1}} N_1\ N' +_p N_2\ N'$.

- The other cases are simple context rules and the result is immediate by induction hypothesis.

$\square$

We can use the structure of canonical terms to define a simplified version of $\to_{\eta_d}$ on those. At the same time we will show that we can restrict the use of the rule $\lambda \overrightarrow{z}_{1\dots k}.M\ \overrightarrow{Z}_{1\dots k} \to_{\eta_{d+1}} N$.

Definition 4.2.2.4. The relations $\to^c_{\eta_d}$ from terms to canonical terms and $\to^v_{\eta_d}$ between values are defined inductively by

$$\frac{}{M \to^c_{\eta_0} N} \qquad \frac{\forall \epsilon > 0, M \twoheadrightarrow_{\beta/\equiv} M_0 +_{1-\epsilon} M' \text{ with } M_0 \to^c_{\eta_{d+1}} N}{M \to^c_{\eta_{d+1}} N}$$

$$\frac{M_1 \to^c_{\eta_{d+1}} N \qquad M_2 \to^c_{\eta_{d+1}} N}{M_1 +_p M_2 \to^c_{\eta_{d+1}} N} \qquad \frac{M_1 \to^c_{\eta_{d+1}} N_1 \qquad M_2 \to^c_{\eta_{d+1}} N_2}{M_1 +_p M_2 \to^c_{\eta_{d+1}} N_1 +_p N_2}$$

$$\frac{v \to^v_{\eta_{d+1}} w \qquad \forall i, z_i \notin \mathrm{FV}(v) \cup \mathrm{FV}(w) \cup \{z_j \mid j \neq i\} \qquad \forall i, PT^\eta_d(Z_i) = z_i}{\lambda \overrightarrow{z}_{1\dots k}.v\ \overrightarrow{Z}_{1\dots k} \to^c_{\eta_{d+1}} w}$$

$$\frac{}{x \to^v_{\eta_{d+1}} x} \qquad \frac{v \to^c_{\eta_{d+1}} w}{\lambda x.v \to^v_{\eta_{d+1}} \lambda x.w} \qquad \frac{v \to^v_{\eta_{d+1}} w \qquad M \to^c_{\eta_d} N}{v\ M \to^v_{\eta_{d+1}} w\ N}$$

Remark that this definition seems better suited for weakly canonical terms: to reduce a value $v\ M$ we treat v as a value, but to reduce $\lambda x.v$ we treat v as a general term. The reason is that we do not want to η-contract a variable which also forms a β-redex: the term $(\lambda z.M\ z)\ N$ with $z \notin \mathrm{FV}(M)$ both β-reduces and η-reduces into $M\ N$, and we want to see this as a β-reduction. But we do not mind η-reducing under an abstraction.

Proposition 4.2.2.14. *If $M \to^c_{\eta_d} N$ then $M \to_{\eta_d} N$.*

Proof. Immediate by induction on $\to^c_{\eta_d}$. $\square$

Lemma 4.2.2.15. *If $M \twoheadrightarrow_{\beta/\equiv} \lambda \overrightarrow{y}_{1\dots l}.M_0\ \overrightarrow{Y}_{1\dots l}$ where the variables y_i are pairwise distinct, $y_i \notin \mathrm{FV}(M_0)$ and $PT^\eta_d(Y_i) = y_i$ for $i \leq l$, then for any pairwise distinct variables z_i and any terms Z_i with $PT^\eta_d(Z_i) = z_i$ for $i \leq k$ we have*

$$\lambda \overrightarrow{z}_{1\dots k}.M\ \overrightarrow{Z}_{1\dots k} \twoheadrightarrow_{\beta/\equiv} \lambda \overrightarrow{x}_{1\dots m}.M_0\ \overrightarrow{X}_{1\dots m}$$

where $m = \max(k, l)$, $x_i = z_i$ for $i \leq k$, $x_{k+i} = y_i$ for $i \leq l - k$ and $PT^\eta_d(X_i) = x_i$ for $i \leq m$.

Proof. We prove this lemma by induction on d.

First observe that if $l \leq k$ then we have

$$\begin{aligned}\lambda \overrightarrow{z}_{1\dots k}.M\ \overrightarrow{Z}_{1\dots k} &\twoheadrightarrow_{\beta/\equiv} \lambda \overrightarrow{z}_{1\dots k}.(\lambda \overrightarrow{y}_{1\dots l}.M_0\ \overrightarrow{Y}_{1\dots l})\ \overrightarrow{Z}_{1\dots k} \\ &\twoheadrightarrow_{\beta/\equiv} \lambda \overrightarrow{z}_{1\dots k}.M_0\ \overrightarrow{Y}_{1\dots l}\left[{}^{Z_1}/{}_{y_1}\right] \dots \left[{}^{Z_l}/{}_{y_l}\right]\ \overrightarrow{Z}_{l+1\dots k}.\end{aligned}$$

If $l \geq k$ then

$$\begin{aligned}\lambda \overrightarrow{z}_{1...k}.M \ \overrightarrow{Z}_{1...k} &\twoheadrightarrow_{\beta/\equiv} \lambda \overrightarrow{z}_{1...k}.(\lambda \overrightarrow{y}_{1...l}.M_0 \ \overrightarrow{Y}_{1...l}) \ \overrightarrow{Z}_{1...k} \\ &\twoheadrightarrow_{\beta/\equiv} \lambda \overrightarrow{z}_{1...k} \overrightarrow{y}_{k+1...l}.M_0 \ \overrightarrow{Y}_{1...l} \left[{}^{Z_1}/{}_{y_1}\right] ... \left[{}^{Z_k}/{}_{y_k}\right].\end{aligned}$$

So we want to prove that in general if $PT_d^\eta(X) = x$ and $PT_d^\eta(Y) = y$ then $PT_d^\eta\left(X\left[{}^{Y}/{}_{x}\right]\right) = y$. If $d = 0$ the result is immediate.

For head normal values, if we have $h = \lambda \overrightarrow{z}_{1...k}.x \ \overrightarrow{Z}_{1...k}$ with $PT_d^\eta(Z_i) = z_i$ and $h' = \lambda \overrightarrow{u}_{1...l}.y \ \overrightarrow{U}_{1...l}$ with $PT_d^\eta(U_i) = u_i$ then by induction hypothesis

$$\lambda \overrightarrow{z}_{1...k}.h' \ \overrightarrow{Z}_{1...k} \twoheadrightarrow_{\beta/\equiv} \lambda \overrightarrow{x}_{1...m}.y \ \overrightarrow{X}_{1...m}$$

with $PT_d^\eta(X_i) = x_i$, hence $PT_{d+1}^\eta\left(\lambda \overrightarrow{z}_{1...k}.h' \ \overrightarrow{Z}_{1...k}\right) = y$.

In general if $PT_{d+1}^\eta(X) = x$ and $PT_{d+1}^\eta(Y) = y$ then for all $\epsilon > 0$ we have

$$\begin{aligned}X &\twoheadrightarrow_{\beta/\equiv} \sum_i p_i.h_i +_{1-\epsilon} X' \\ Y &\twoheadrightarrow_{\beta/\equiv} \sum_j q_j.h'_j +_{1-\epsilon} Y'\end{aligned}$$

where the terms h_i and h'_j are head normal values.

Let us write $h_i = \lambda \overrightarrow{z}_{1...k_i}.x \ \overrightarrow{Z}_{i,1...i,k_i}$, we get

$$X\left[{}^{Y}/{}_{x}\right] \twoheadrightarrow_{\beta/\equiv} \sum_i \sum_j p_i q_j.\lambda \overrightarrow{z}_{1...k_i}.h'_j \ \overrightarrow{Z'}_{i,1...i,k_i} +_{(1-\epsilon)^2} Y''$$

with $Z'_{i,j} = Z_{i,j}\left[{}^{Y}/{}_{x}\right]$ and $PT_d^\eta(Z'_{i,j}) = z_j$ for all i and $j \leq k_i$.

From there, $PT_{d+1}^\eta\left(X\left[{}^{Y}/{}_{x}\right]\right) = y$. □

Proposition 4.2.2.16. *If $M \to_{\eta_d}^c N$ and the variables z_i are pairwise distinct, $z_i \notin \mathrm{FV}(M) \cup \mathrm{FV}(N)$ and $PT_d^\eta(Z_i) = z_i$ for all $i \leq k$ then*

$$\lambda \overrightarrow{z}_{1...k}.M \ \overrightarrow{Z}_{1...k} \to_{\eta_d}^c N.$$

Proof. By induction on $M \to_{\eta_d}^c N$.

- If $M \to_{\eta_0}^c N$ then $\lambda \overrightarrow{z}_{1...k}.M \ \overrightarrow{Z}_{1...k} \to_{\eta_0}^c N$.
- If for all $\epsilon > 0$, $M \twoheadrightarrow_{\beta/\equiv} M_0 +_{1-\epsilon} M'$ with $M_0 \to_{\eta_{d+1}}^c N$ then

$$\lambda \overrightarrow{z}_{1...k}.M \ \overrightarrow{Z}_{1...k} \twoheadrightarrow_{\beta/\equiv} \lambda \overrightarrow{z}_{1...k}.M_0 \ \overrightarrow{Z}_{1...k} +_{1-\epsilon} \lambda \overrightarrow{z}_{1...k}.M' \ \overrightarrow{Z}_{1...k}$$

  with by induction hypothesis $\lambda \overrightarrow{z}_{1...k}.M_0 \ \overrightarrow{Z}_{1...k} \to_{\eta_{d+1}}^c N$.

- If $M_1 +_p M_2 \to^c_{\eta_{d+1}} N$ with $M_i \to^c_{\eta_{d+1}} N$ for $i \in \{1;2\}$ then

$$\lambda \overrightarrow{z}_{1\dots k}.(M_1 +_p M_2)\ \overrightarrow{Z}_{1\dots k} \equiv \lambda \overrightarrow{z}_{1\dots k}.M_1\ \overrightarrow{Z}_{1\dots k} +_p \lambda \overrightarrow{z}_{1\dots k}.M_2\ \overrightarrow{Z}_{1\dots k}$$

  and by induction hypothesis $\lambda \overrightarrow{z}_{1\dots k}.M_i\ \overrightarrow{Z}_{1\dots k} \to^c_{\eta_{d+1}} N$ for $i \in \{1;2\}$.

- If $M_1 +_p M_2 \to^c_{\eta_{d+1}} N_1 +_p N_2$ with $M_i \to^c_{\eta_{d+1}} N_i$ for $i \in \{1;2\}$ then

$$\lambda \overrightarrow{z}_{1\dots k}.(M_1 +_p M_2)\ \overrightarrow{Z}_{1\dots k} \equiv \lambda \overrightarrow{z}_{1\dots k}.M_1\ \overrightarrow{Z}_{1\dots k} +_p \lambda \overrightarrow{z}_{1\dots k}.M_2\ \overrightarrow{Z}_{1\dots k}.$$

  with by induction hypothesis $\lambda \overrightarrow{z}_{1\dots k}.M_i\ \overrightarrow{Z}_{1\dots k} \to^c_{\eta_{d+1}} N_i$ for $i \in \{1;2\}$.

- If $M = \lambda \overrightarrow{y}_{1\dots l}.v\ \overrightarrow{Y}_{1\dots l} \to^c_{\eta_{d+1}} w$ with $v \to^v_{\eta_{d+1}} w$ then the previous lemma gives

$$\lambda \overrightarrow{z}_{1\dots k}.M\ \overrightarrow{Z}_{1\dots k} \twoheadrightarrow_{\beta/\equiv} \lambda \overrightarrow{x}_{1\dots m}.v\ \overrightarrow{X}_{1\dots m}$$

  and we have $\lambda \overrightarrow{x}_{1\dots m}.v\ \overrightarrow{X}_{1\dots m} \to^c_{\eta_{d+1}} w$.

□

Proposition 4.2.2.17. *If N is canonical and $M \to_{\eta_d} N$ then $M \to^c_{\eta_d} N$.*

Proof. We reason by induction on N and $M \to_{\eta_d} N$.

- If $M \to_{\eta_0} N$ then $M \to^c_{\eta_0} N$.
- If for all $\epsilon > 0$ we have $M \twoheadrightarrow_{\beta/\equiv} M_0 +_{1-\epsilon} M'$ with $M_0 \to_{\eta_{d+1}} N$ then by induction hypothesis $M_0 \to^c_{\eta_{d+1}} N$ hence $M \to^c_{\eta_{d+1}} N$.
- If $M_1 +_p M_2 \to_{\eta_{d+1}} N$ with $M_i \to_{\eta_{d+1}} N$ for $i \in \{1;2\}$ then by induction hypothesis $M_i \to^c_{\eta_{d+1}} N$ and $M_1 +_p M_2 \to^c_{\eta_{d+1}} N$.
- If $\lambda \overrightarrow{z}_{1\dots k}.M\ \overrightarrow{Z}_{1\dots k} \to_{\eta_{d+1}} N$ with $M \to_{\eta_{d+1}} N$ then by induction hypothesis $M \to^c_{\eta_{d+1}} N$ and the previous proposition gives $\lambda \overrightarrow{z}_{1\dots k}.M\ \overrightarrow{Z}_{1\dots k} \to^c_{\eta_{d+1}} N$.
- If $x \to_{\eta_{d+1}} x$ then $x \to^c_{\eta_{d+1}} x$.
- If $\lambda x.M \to_{\eta_{d+1}} \lambda x.w$ with $M \to_{\eta_{d+1}} w$ then by induction hypothesis we have $M \to^c_{\eta_{d+1}} w$, and $\lambda x.M \to^c_{\eta_{d+1}} \lambda x.w$.
- If $M\ M' \to_{\eta_{d+1}} w\ N'$ with $M \to_{\eta_{d+1}} w$ and $M' \to_{\eta_d} N'$ then by induction hypothesis $M \to^c_{\eta_{d+1}} w$, we reason by induction on this reduction.
  - If for all $\epsilon > 0$, $M \twoheadrightarrow_{\beta/\equiv} M_0 +_{1-\epsilon} M''$ with $M_0 \to^c_{\eta_{d+1}} w$ then by induction hypothesis $M_0\ M' \to^c_{\eta_{d+1}} w\ N'$, and $M\ M' \twoheadrightarrow_{\beta/\equiv} M_0\ M' +_{1-\epsilon} M''\ M'$ so $M\ M' \to^c_{\eta_{d+1}} w\ N'$.
  - If $M = M_1 +_p M_2$ with $M_i \to^c_{\eta_{d+1}} N$ for $i \in \{1;2\}$ then we have $M\ M' \equiv M_1\ M' +_p M_2\ M'$ and we conclude by induction hypothesis.

– If $M = v$ with $v \to^v_{\eta_{d+1}} w$ then the induction hypothesis on N gives that $M' \to^c_{\eta_d} N'$ and $v\ M' \to^v_{\eta_{d+1}} w\ N'$.

– If $M = \lambda \overrightarrow{z}_{1 \ldots k+1}.v\ \overrightarrow{Z}_{1 \ldots k+1}$ with $v \to^v_{\eta_{d+1}} w$ then

$$M\ M' \to_\beta \lambda \overrightarrow{z}_{2 \ldots k+1}.v\ \overrightarrow{Z}_{1 \ldots k+1} \left[{}^{M'}/_{z_1}\right]$$

and the proposition 4.2.2.9 gives $Z_1 \left[{}^{M'}/_{z_1}\right] \to_{\eta_d} N'$. Using the outer induction hypothesis on N we have that $Z_1 \left[{}^{M'}/_{z_1}\right] \to^c_{\eta_d} N'$, and by induction hypothesis $v\ Z_1 \left[{}^{M'}/_{z_1}\right] \to^c_{\eta_{d+1}} w\ N'$, hence

$$\lambda \overrightarrow{z}_{2 \ldots k+1}.v\ Z_1 \left[{}^{M'}/_{z_1}\right]\ \overrightarrow{Z}_{2 \ldots k+1} \left[{}^{M'}/_{z_1}\right] \to^c_{\eta_{d+1}} w\ N'.$$

- If $M_1 +_p M_2 \to_{\eta_{d+1}} N_1 +_p N_2$ with $M_i \to_{\eta_{d+1}} N_i$ for $i \in \{1; 2\}$ then by induction hypothesis $M_i \to^c_{\eta_{d+1}} N_i$ and $M_1 +_p M_2 \to^c_{\eta_{d+1}} N_1 +_p N_2$.

□

Now we can prove that the reduction $\to^c_{\eta_d}$ behaves properly with respect to the complete left reduction.

Proposition 4.2.2.18. *If $M \to^c_{\eta_{d+1}} N$ then $\mathrm{L}(M) \to^c_{\eta_d} \mathrm{L}(N)$.*

Proof. Again we prove that $\mathrm{L}(M) \to_{\eta_d} N'$ where N' does not depend on M and $\mathrm{can}(N') = \mathrm{L}(N)$.

We reason by induction on N and $M \to^c_{\eta_{d+1}} N$. The result is immediate if $d = 0$.

- If for all $\epsilon > 0$, $M \twoheadrightarrow_{\beta/\equiv} M_0 +_{1-\epsilon} M'$ with $M_0 \to^c_{\eta_{d+2}} N$ then we have $\mathrm{L}(M) \twoheadrightarrow_{\beta/\equiv} \mathrm{L}(M_0) +_{1-\epsilon} \mathrm{L}(M')$ and by induction hypothesis $\mathrm{L}(M_0) \to_{\eta_{d+1}} N'$, hence $M \to_{\eta_{d+1}} N'$.

- If $M_1 +_p M_2 \to^c_{\eta_{d+2}} N$ with $M_i \to^c_{\eta_{d+2}} N_i$ for $i \in \{1; 2\}$ then by induction hypothesis $\mathrm{L}(M_i) \to_{\eta_{d+1}} N'$ and $\mathrm{L}(M_1 +_p M_2) = \mathrm{L}(M_1) +_p \mathrm{L}(M_2) \to_{\eta_{d+1}} N'$.

- If $M_1 +_p M_2 \to^c_{\eta_{d+2}} N_1 +_p N_2$ with $M_i \to^c_{\eta_{d+2}} N_i$ for $i \in \{1; 2\}$ then by induction hypothesis $\mathrm{L}(M_i) \to_{\eta_{d+1}} N'_i$ and $\mathrm{L}(M_1 +_p M_2) \to_{\eta_{d+1}} N'_1 +_p N'_2$.

- If $v \to^c_{\eta_{d+2}} w$ with $v \to^v_{\eta_{d+2}} w$ then by induction hypothesis $\mathrm{L}(v) \to_{\eta_{d+1}} w'$.

- If $\lambda \overrightarrow{z}_{1 \ldots k+1}.v\ \overrightarrow{Z}_{1 \ldots k} \to^c_{\eta_{d+2}} w$ with $v \to^v_{\eta_{d+2}} w$ and v is not an abstraction then

$$\mathrm{L}(\lambda \overrightarrow{z}_{1 \ldots k+1}.v\ \overrightarrow{Z}_{1 \ldots k}) =_+ \lambda \overrightarrow{z}_{1 \ldots k+1}.\mathrm{L}(v)\ \overrightarrow{Z}_{1 \ldots k}$$

or

$$\mathrm{L}(\lambda \overrightarrow{z}_{1 \ldots k+1}.v\ \overrightarrow{Z}_{1 \ldots k}) =_+ \lambda \overrightarrow{z}_{1 \ldots k+1}.\mathrm{L}(v)\ \mathrm{L}(\overrightarrow{Z}_{1 \ldots k})$$

and by induction hypothesis $\mathrm{L}(v) \to_{\eta_{d+1}} w'$, hence $\mathrm{L}(\lambda \overrightarrow{z}_{1 \ldots k+1}.v\ \overrightarrow{Z}_{1 \ldots k}) \to_{\eta_{d+1}} w'$.

- If $\lambda \overrightarrow{z}_{1\ldots k+1}.(\lambda x.v)\ \overrightarrow{Z}_{1\ldots k} \to^{c}_{\eta_{d+2}} \lambda x.w$ with $v \to^{v}_{\eta_{d+2}} w$ then by induction hypothesis $\mathrm{L}(v) \to_{\eta_{d+1}} w'$, and $Z_1 \to_{\eta_{d+1}} z_1$ so the proposition 4.2.2.9 gives $\mathrm{L}(v) \left[{}^{Z_1}/_{x}\right] \to_{\eta_{d+1}} w' \left[{}^{z_1}/_{x}\right]$. We have

$$\begin{aligned}\mathrm{L}(\lambda \overrightarrow{z}_{1\ldots k+1}.(\lambda x.v)\ \overrightarrow{Z}_{1\ldots k}) &=_{+} \lambda \overrightarrow{z}_{1\ldots k+1}.v \left[{}^{Z_1}/_{x}\right]\ \overrightarrow{Z}_{2\ldots k}\\ &\twoheadrightarrow_{\beta/\equiv} \lambda \overrightarrow{z}_{1\ldots k+1}.\mathrm{L}(v) \left[{}^{Z_1}/_{x}\right]\ \overrightarrow{Z}_{2\ldots k}\\ &\to_{\eta_{d+1}} \lambda z_1.w' \left[{}^{z_1}/_{x}\right]\\ &= \lambda x.w'.\end{aligned}$$

- If $\lambda x.v \to^{v}_{\eta_{d+2}} \lambda x.w$ with $v \to^{c}_{\eta_{d+2}} w$ then by induction hypothesis we have $\mathrm{L}(v) \to_{\eta_{d+1}} w'$ and $\mathrm{L}(\lambda x.v) =_{+} \lambda x.\mathrm{L}(v) \to_{\eta_{d+1}} \lambda x.w'$.

- If $x\ \overrightarrow{P}_{1\ldots m} \to^{v}_{\eta_{d+2}} x\ \overrightarrow{Q}_{1\ldots m}$ with $P_i \to^{c}_{\eta_{d+1}} Q_i$ for $i \leq m$ then by induction hypothesis $\mathrm{L}(P_i) \to_{\eta_d} Q_i'$ for $i \leq m$ and $x\ \mathrm{L}(\overrightarrow{P}_{1\ldots m}) \to_{\eta_{d+1}} x\ \overrightarrow{Q'}_{1\ldots m}$.

- If $(\lambda x.v)\ P\ \overrightarrow{R}_{1\ldots m} \to^{v}_{\eta_{d+2}} (\lambda x.w)\ Q\ \overrightarrow{S}_{1\ldots m}$ with $v \to^{c}_{\eta_{d+2}} w$, $P \to^{c}_{\eta_{d+1}} Q$ and $R_i \to^{c}_{\eta_{d+1}} S_i$ for all $i \leq m$ then the proposition 4.2.2.9 gives that $v \left[{}^{P}/_{x}\right] \to_{\eta_{d+1}} w \left[{}^{Q}/_{x}\right]$, hence

$$\begin{aligned}\mathrm{L}((\lambda x.v)\ P\ \overrightarrow{R}_{1\ldots m}) &=_{+} v \left[{}^{P}/_{x}\right]\ \overrightarrow{R}_{1\ldots m}\\ &\to_{\eta_{d+1}} w \left[{}^{Q}/_{x}\right]\ \overrightarrow{S}_{1\ldots m}.\end{aligned}$$

□

We obtain the following result.

Proposition 4.2.2.19. *If for all $d \in \mathbb{N}$ and all $\epsilon > 0$ we have*

$$M \twoheadrightarrow_{\beta/\equiv} \cdot \equiv_{dum} \cdot \to^{c}_{\eta_d} \cdot =^{s,c}_{\Omega,\epsilon} \cdot \leftarrow^{c}_{\eta_d} \cdot \equiv_{dum} \cdot \twoheadleftarrow_{\beta/\equiv} N$$

then for all $k \in \mathbb{N}$, $d \in \mathbb{N}$ and all $\epsilon > 0$ we have

$$\mathrm{L}^k(M) \twoheadrightarrow_{\beta/\equiv} \cdot \equiv_{dum} \cdot \to^{c}_{\eta_d} \cdot =^{s,c}_{\Omega,\epsilon} \cdot \leftarrow^{c}_{\eta_d} \cdot \equiv_{dum} \cdot \twoheadleftarrow_{\beta/\equiv} \mathrm{L}^k(N).$$

The relation of the previous proposition holds between terms with the same infinitely extensional Böhm tree and it is stable by complete left reduction, all we have left to do is to prove that if two d, ϵ-normal forms are related then they have almost the same Böhm trees of depth d. First it is easy to see that the local Böhm trees increase with the reduction $\to^{c}_{\eta_d}$.

Proposition 4.2.2.20. *If M is canonical and $M \to^{c}_{\eta_d} N$ then $\mathrm{pt}^{\eta}_{d}(M) \preceq^{\eta}_{d} \mathrm{pt}^{\eta}_{d}(N)$.*

Proof. We reason by induction on $M \to^{c}_{\eta_d} N$.

- If $M \to^c_{\eta_0} N$ then the result is immediate.

- If for all $\epsilon > 0$ we have $M \twoheadrightarrow_{\beta/\equiv} M_0 +_{1-\epsilon} M'$ with $M_0 \to^c_{\eta_{d+1}} N$ then by induction hypothesis $\mathrm{pt}^\eta_{d+1}(M_0) \preceq^\eta_{d+1} \mathrm{pt}^\eta_{d+1}(N)$. From there we have $\mathrm{pt}^\eta_{d+1}(M) \preceq^\eta_{d+1,\epsilon} \mathrm{pt}^\eta_{d+1}(N)$ and $\mathrm{pt}^\eta_{d+1}(M) \preceq^\eta_{d+1} \mathrm{pt}^\eta_{d+1}(N)$.

- If $M_1 +_p M_2 \to^c_{\eta_{d+1}} N$ with $M_i \to^c_{\eta_{d+1}} N$ for $i \in \{1;2\}$ then by induction hypothesis $\mathrm{pt}^\eta_{d+1}(M_i) \preceq^\eta_{d+1} \mathrm{pt}^\eta_{d+1}(N)$ thus
$$p\mathrm{pt}^\eta_{d+1}(M_1) + (1-p)\mathrm{pt}^\eta_{d+1}(M_2) \preceq^\eta_{d+1} \mathrm{pt}^\eta_{d+1}(N).$$

- If $M_1 +_p M_2 \to^c_{\eta_{d+1}} N_1 +_p N_2$ with $M_i \to^c_{\eta_{d+1}} N_i$ for $i \in \{1;2\}$ then by induction hypothesis $\mathrm{pt}^\eta_{d+1}(M_i) \preceq^\eta_{d+1} \mathrm{pt}^\eta_{d+1}(N_i)$ and
$$p\mathrm{pt}^\eta_{d+1}(M_1) + (1-p)\mathrm{pt}^\eta_{d+1}(M_2) \preceq^\eta_{d+1} p\mathrm{pt}^\eta_{d+1}(N_1) + (1-p)\mathrm{pt}^\eta_{d+1}(N_2).$$

- If $v \to^c_{\eta_{d+1}} w$ with $v \to^v_{\eta_{d+1}} w$ then by induction hypothesis we have $\mathrm{pt}^\eta_{d+1}(v) \preceq^\eta_{d+1} \mathrm{pt}^\eta_{d+1}(w)$.

- If $\lambda \overrightarrow{z}_{1\ldots k+1}.x\ \overrightarrow{P}_{1\ldots m}\ \overrightarrow{Z}_{1\ldots k} \to^c_{\eta_{d+1}} x\ \overrightarrow{Q}_{1\ldots m}$ with $P_i \to^c_{\eta_d} Q_i$ for $i \leq m$ then
$$\mathrm{pt}^\eta_{d+1}(\lambda \overrightarrow{z}_{1\ldots k+1}.x\ \overrightarrow{P}_{1\ldots m}\ \overrightarrow{Z}_{1\ldots k}) = \mathrm{pt}^\eta_{d+1}(x\ \overrightarrow{P}_{1\ldots m})$$
and by induction hypothesis $\mathrm{pt}^\eta_d(P_i) \preceq^\eta_d \mathrm{pt}^\eta_d(Q_i)$ for $i \leq m$ so
$$\mathrm{pt}^\eta_{d+1}(x\ \overrightarrow{P}_{1\ldots m}) \preceq^\eta_{d+1} \mathrm{pt}^\eta_{d+1}(x\ \overrightarrow{Q}_{1\ldots m}).$$

- If $\lambda \overrightarrow{z}_{1\ldots k+1}.(\lambda x.v)\ \overrightarrow{P}_{1\ldots m}\ \overrightarrow{Z}_{1\ldots k} \to^c_{\eta_{d+1}} (\lambda x.w)\ \overrightarrow{Q}_{1\ldots m}$ then
$$\mathrm{pt}^\eta_{d+1}(\lambda \overrightarrow{z}_{1\ldots k+1}.(\lambda x.v)\ \overrightarrow{P}_{1\ldots m}\ \overrightarrow{Z}_{1\ldots k}) = 0.$$

- If $\lambda x.v \to^v_{\eta_{d+1}} \lambda x.w$ with $v \to^c_{\eta_{d+1}} w$ then by induction hypothesis we have $\mathrm{pt}^\eta_{d+1}(v) \preceq^\eta_{d+1} \mathrm{pt}^\eta_{d+1}(w)$ and $\mathrm{pt}^\eta_{d+1}(\lambda x.v) \preceq^\eta_{d+1} \mathrm{pt}^\eta_{d+1}(\lambda x.w)$.

- If $x\ \overrightarrow{P}_{1\ldots m} \to^v_{\eta_{d+1}} x\ \overrightarrow{Q}_{1\ldots m}$ with $P_i \to^c_{\eta_d} Q_i$ for $i \leq m$ then we get the result with the induction hypothesis.

- If $(\lambda x.v)\ \overrightarrow{P}_{1\ldots m+1} \to^v_{\eta_{d+1}} (\lambda x.w)\ \overrightarrow{Q}_{1\ldots m+1}$ then $\mathrm{pt}^\eta_{d+1}((\lambda x.v)\ \overrightarrow{P}_{1\ldots m+1}) = 0$.

$\square$

It is more complicated to see that if we reduce a d,ϵ-normal form with $\to^c_{\eta_d}$ then the former tree is greater than the latter up to ϵ.

Indeed if $M \equiv \sum_i p_i.h_i + (1-\sum_i p_i).M'$ with $(1-\sum_i p_i).\mathcal{P}_{\Downarrow}(M') \leq \epsilon$ and if $M \to^c_{\eta_d} N$ then $N \equiv \sum_i p_i.h'_i + (1-\sum_i p_i).N'$ with $h_i \to^c_{\eta_d} h'_i$ and $M' \to^c_{\eta_d} N'$, but for now we do not know anything about the convergence $\mathcal{P}_{\Downarrow}(N')$.

Proposition 4.2.2.21. *If $M \to^c_{\eta_{d+1}} h$ with h a head normal value then $\mathcal{P}_{\Downarrow}(M) = 1$.*

Proof. We reason by induction on $M \to^c_{\eta_{d+1}} h$.

- If for all $\epsilon > 0$ we have $M \twoheadrightarrow_{\beta/\equiv} M_0 +_{1-\epsilon} M'$ with $M_0 \to^c_{\eta_{d+1}} h$ then by induction hypothesis $\mathcal{P}_{\Downarrow}(M_0) = 1$, hence $\mathcal{P}_{\Downarrow}(M) \geq 1 - \epsilon$.
- If $M_1 +_p M_2 \to^c_{\eta_{d+1}} h$ then by induction hypothesis $\mathcal{P}_{\Downarrow}(M_i) = 1$ for $i \in \{1;2\}$ and $\mathcal{P}_{\Downarrow}(M) = 1$.
- If $v \to^c_{\eta_{d+1}} h$ with $v \to^v_{\eta_{d+1}} h$ then by induction hypothesis $\mathcal{P}_{\Downarrow}(v) = 1$.
- If $\lambda \overrightarrow{z}_{1\ldots k+1}.v\ \overrightarrow{Z}_{1\ldots k} \to^c_{\eta_{d+1}} h$ with $v \to^v_{\eta_{d+1}} h$ then by induction hypothesis $\mathcal{P}_{\Downarrow}(v) = 1$. Then for all $\epsilon > 0$ we have $v \twoheadrightarrow_{\beta/\equiv} \sum_i p_i.h_i +_{1-\epsilon} v'$ where the terms h_i are head normal values, and

  $$\lambda \overrightarrow{z}_{1\ldots k+1}.v\ \overrightarrow{Z}_{1\ldots k} \twoheadrightarrow_{\beta/\equiv} \sum_i p_i.\lambda \overrightarrow{z}_{1\ldots k+1}.h_i\ \overrightarrow{Z}_{1\ldots k} +_{1-\epsilon} \lambda \overrightarrow{z}_{1\ldots k+1}.v'\ \overrightarrow{Z}_{1\ldots k}.$$

  We can then check that for all i we have $\mathcal{P}_{\Downarrow}(\lambda \overrightarrow{z}_{1\ldots k+1}.h_i\ \overrightarrow{Z}_{1\ldots k}) = 1$, hence $\mathcal{P}_{\Downarrow}(\lambda \overrightarrow{z}_{1\ldots k+1}.v\ \overrightarrow{Z}_{1\ldots k}) \geq 1 - \epsilon$ and $\mathcal{P}_{\Downarrow}(\lambda \overrightarrow{z}_{1\ldots k+1}.v\ \overrightarrow{Z}_{1\ldots k}) = 1$.
- If $\lambda x.v \to^v_{\eta_{d+1}} \lambda x.h$ with $v \to^c_{\eta_{d+1}} h$ then by induction hypothesis $\mathcal{P}_{\Downarrow}(v) = 1$ and $\mathcal{P}_{\Downarrow}(\lambda x.v) = 1$.
- If $x\ \overrightarrow{P}_{1\ldots m} \to^v_{\eta_{d+1}} x\ \overrightarrow{Q}_{1\ldots m}$ with $P_i \to^c_{\eta_d} Q_i$ for $i \leq m$ then $\mathcal{P}_{\Downarrow}(x\ \overrightarrow{P}_{1\ldots m}) = 1$.

□

Proposition 4.2.2.22. *For all $d \in \mathbb{N}$ and $\epsilon > 0$, if M is a d, ϵ-normal form and $M \to^c_{\eta_d} N$ then $\mathrm{pt}^\eta_d(N) \preceq^\eta_{d,\epsilon} PT^\eta_d(M)$.*

Proof. We reason by induction on d, the result is immediate when $d = 0$.

Otherwise using the lemma 4.2.2.12 we can write $N \equiv \sum_i p_i.v_i$ and get for all $\epsilon' > 0$ that $M \twoheadrightarrow_{\beta/\equiv} \sum_i p_i.M_i +_{1-\epsilon'} M'$ with $M_i \to^c_{\eta_{d+1}} v_i$ for all i. The previous result gives that if v_i is a head normal value then $\mathcal{P}_{\Downarrow}(M_i) = 1$, and the fact that M is a $d+1, \epsilon$-normal form (hence $\sum_i p_i.M_i +_{1-\epsilon'} M'$ is a $d+1, \epsilon$-normal form) gives

$$(1-\epsilon') \sum_{i \text{ s.t. } M_i \text{ is not head normal}} p_i.\mathcal{P}_{\Downarrow}(M_i) \leq \epsilon.$$

Together these two results give

$$(1-\epsilon') \sum_{i \text{ s.t. } M_i \text{ is not head normal}, v_i \text{ is head normal}} p_i \leq \epsilon.$$

But if M_i is head normal and $M_i \to^c_{\eta_{d+1}} v_i$ then v_i is head normal, and we prove by induction on $M_i \to^c_{\eta_{d+1}} v_i$ that $\mathrm{pt}^\eta_d(v_i) \preceq^\eta_{d,\epsilon} PT^\eta_d(M_i)$.

- If for all $\epsilon'' > 0$ we have $M_i \twoheadrightarrow_{\beta/\equiv} M_{i,0} +_{1-\epsilon''} M_i'$ with $M_{i,0} \to^c_{\eta_{d+1}} v_i$ then by induction hypothesis $\mathrm{pt}^\eta_d(v_i) \preceq^\eta_{d,\epsilon} PT^\eta_d(M_{i,0})$, hence $\mathrm{pt}^\eta_d(v_i) \preceq^\eta_{d,\epsilon+\epsilon''} PT^\eta_d(M_i)$. This holds for all $\epsilon'' > 0$ so we have $\mathrm{pt}^\eta_d(v_i) \preceq^\eta_{d,\epsilon} PT^\eta_d(M_i)$.

- If $M_i = M_{i,1} +_p M_{i,2}$ with $M_{i,j} \to^c_{\eta_{d+1}} v_i$ then the result is immediate by induction hypothesis.

- If $M_i = \lambda \overrightarrow{x}_{1\dots n} \overrightarrow{z}_{1\dots k}.y\ \overrightarrow{P}_{1\dots m}\ \overrightarrow{Z}_{1\dots k}$ and $v_i = \lambda \overrightarrow{x}_{1\dots n}.y\ \overrightarrow{Q}_{1\dots m}$ with $P_i \to^c_{\eta_d} Q_i$ for all $i \leq m$ then the induction hypothesis on d gives $\mathrm{pt}^\eta_d(Q_i) \preceq^\eta_{d,\epsilon} PT^\eta_d(P_i)$ and we have

$$\mathrm{pt}^\eta_d(\lambda \overrightarrow{x}_{1\dots n}.y\ \overrightarrow{Q}_{1\dots m}) \preceq^\eta_{d,\epsilon} PT^\eta_d(\lambda \overrightarrow{x}_{1\dots n} \overrightarrow{z}_{1\dots k}.y\ \overrightarrow{P}_{1\dots m}\ \overrightarrow{Z}_{1\dots k}).$$

Then we have

$$\begin{aligned} PT^\eta_{d+1}(M) = \quad & (1-\epsilon') \sum_{i \text{ s.t. } M_i \text{ hnf}} p_i PT^\eta_{d+1}(M_i) \\ & +(1-\epsilon') \sum_{i \text{ s.t. } M_i \text{ not hnf}} p_i PT^\eta_{d+1}(M_i) \\ & +\epsilon' PT^\eta_{d+1}(M') \end{aligned}$$

and

$$\begin{aligned} \mathrm{pt}^\eta_{d+1}(N) = \quad & (1-\epsilon') \sum_{i \text{ s.t. } M_i \text{ hnf}} p_i \mathrm{pt}^\eta_{d+1}(v_i) \\ & +(1-\epsilon') \sum_{i \text{ s.t. } M_i \text{ not hnf}} p_i \mathrm{pt}^\eta_{d+1}(v_i) \\ & +\epsilon' \sum_i p_i \mathrm{pt}^\eta_{d+1}(v_i) \end{aligned}$$

hence

$$\mathrm{pt}^\eta_d(N) \preceq^\eta_{d,\epsilon+\epsilon'} PT^\eta_d(M).$$

This holds for all $\epsilon' > 0$ so $\mathrm{pt}^\eta_d(N) \preceq^\eta_{d,\epsilon} PT^\eta_d(M)$. □

Proposition 4.2.2.23. *If for all $d \in \mathbb{N}$ and all $\epsilon > 0$ we have*

$$M \twoheadrightarrow_{\beta/\equiv} \cdot \equiv_{dum} \cdot \to^c_{\eta_d} \cdot =^{s,c}_{\Omega,\epsilon} \cdot \leftarrow^c_{\eta_d} \cdot \equiv_{dum} \cdot \twoheadleftarrow_{\beta/\equiv} N$$

then $M =_{\mathcal{PB}^\eta} N$.

Proof. Given $d \in \mathbb{N}$ and $\epsilon > 0$, for $k \in \mathbb{N}$ large enough the terms $\mathrm{L}^k(M)$ and $\mathrm{L}^k(N)$ are d,ϵ-normal forms, and we have

$$\mathrm{L}^k(M) \twoheadrightarrow_{\beta/\equiv} M_1 \equiv_{dum} M_2 \to^c_{\eta_d} M_3 =^{s,c}_{\Omega,\epsilon} N_3 \leftarrow^c_{\eta_d} N_2 \equiv_{dum} N_1 \twoheadleftarrow_{\beta/\equiv} \mathrm{L}^k(N).$$

Then we have

$$\begin{aligned}\mathrm{pt}^\eta_d(\mathrm{L}^k(M)) &\preceq^\eta_d \mathrm{pt}^\eta_d(M_1) = \mathrm{pt}^\eta_d(M_2) \preceq^\eta_d \mathrm{pt}^\eta_d(M_3)\\ &\preceq^\eta_{d,\epsilon} \mathrm{pt}^\eta_d(N_3) \preceq^\eta_{d,\epsilon} PT^\eta_d(N_2) = PT^\eta_d(N_1) = PT^\eta_d(N).\end{aligned}$$

For all $d \in \mathbb{N}$ and $k \in \mathbb{N}$, for all $\epsilon > 0$ we have $\mathrm{pt}^\eta_d(\mathrm{L}^k(M)) \preceq^\eta_{d,2\epsilon} PT^\eta_d(N)$, hence $\mathrm{pt}^\eta_d(\mathrm{L}^k(M)) \preceq^\eta_d PT^\eta_d(N)$ and $PT^\eta_d(M) \preceq^\eta_d PT^\eta_d(N)$. By symmetry we have $PT^\eta_d(M) = PT^\eta_d(N)$, and $M =_{\mathcal{PB}^\eta} N$. □

Theorem 4.2.2.24. *If $M =_{\mathcal{PB}^\eta} N$ then for all context C, $C[M] =_{\mathcal{PB}^\eta} C[N]$.*

Proof. The relation

$$\bigcap_{d\in\mathbb{N}} \bigcap_{\epsilon>0} \twoheadrightarrow_{\beta/\equiv} \cdot \equiv_{dum} \cdot \rightarrow^c_{\eta_d} \cdot =^{s,c}_{\Omega,\epsilon} \cdot \leftarrow^c_{\eta_d} \cdot \equiv_{dum} \cdot \twoheadleftarrow_{\beta/\equiv}$$

is stable by context. □

Theorem 4.2.2.25. *The relation $=_{\mathcal{PB}^\eta}$ is a theory.*

5 Separability

The infinitely extensional probabilistic Böhm tree equality forms a theory, and it is contained in the observational equivalence. To prove that these two theories are the same we need to show that if two terms have different Böhm trees then they are not observationally equivalent. In other words we need a separability result.

In the deterministic calculus there are two notions of separation. Two terms M and N are usually said to be separable if given any pair of terms P and Q there is a context C such that $C[M] \twoheadrightarrow_\beta P$ and $C[N] \twoheadrightarrow_\beta Q$. This is a strong property, and although it is useful to know that two terms are separable, it is much less informative to know that they are not. If we consider normalizable terms then two terms are β-equivalent if and only if they are not separable. But if we consider the typical diverging term Ω, any context C such that $C[\Omega] \twoheadrightarrow_\beta \lambda x.x$ does not use its argument: for any term M we also have $C[M] \twoheadrightarrow_\beta \lambda x.x$. This means that Ω is never separable from any other term.

The second notion of separation, called the weak separation, is obtained by weakening the observation made on terms. Two terms M and N are weakly separable if there is a context C such that one of the terms $C[M]$ and $C[N]$ is solvable, while the other is not. This corresponds exactly to the definition of the observational equivalence: two terms are weakly separable if and only if they are not observationally equivalent.

The situation is the same in the probabilistic calculus. We can define many notions of separability, and each will be useful on some particular terms. For instance we can prove that if $M \neq_{\beta+} N$ are probabilistic terms which normalize for the reduction $\to_\beta \cup \to_+$ then for any terms P and Q, for any $\epsilon > 0$ there is a context C such that $C[M] \twoheadrightarrow_{\beta/\equiv} P +_{1-\epsilon} Q$ and $C[N] \twoheadrightarrow_{\beta/\equiv} P +_\epsilon Q$.

Here we are interested in the observational equivalence, so we will use the corresponding notion of separability.

Definition 5.0.2.1. Two terms M and N are *separable* if there is a context C such that

$$\mathcal{P}_\Downarrow(C[M]) \neq \mathcal{P}_\Downarrow(C[N]).$$

We want to prove that two terms with different infinitely extensional Böhm trees are separable. This result is not much more difficult to obtain than in the deterministic case, but we encounter the same problem as for the contextuality of Böhm trees. Given a deterministic term M, for any $d \in \mathbb{N}$ there is $k \in \mathbb{N}$ such that $BT^\eta_d(M) = \mathrm{bt}^\eta_d(\mathrm{L}^k(M))$, so it is easy to deduce the behaviour of a term from its Böhm tree. With probabilistic terms, on the other hand, we always need to work up to a small probability $\epsilon > 0$.

For that reason we split the separability proof in two parts. Given finitely many trees $T_1,\ldots,T_n$ we will describe a way to get probabilities $\alpha_1,\ldots,\alpha_n$ such that $\alpha_i = \alpha_j$ if and only if $T_i = T_j$, and then we will show that given terms $M_1,\ldots,M_n$ with Böhm trees $T_1,\ldots,T_n$, for all $\epsilon > 0$ we can find a context C such that $|\mathcal{P}_{\Downarrow}(C[M_i]) - \alpha_i| \leq \epsilon$ for all $i \leq n$.

5.1 Separating trees

Our goal is to find a way to read the structure of a tree and to encode it in a single probability.

As we mentioned when we first defined infinitely extensional Böhm trees, the trees obtained from terms have a particular structure. For probabilistic trees, we get that for every head normal value h the value tree $VT^{\eta}_{d+1}(h)$ is of the form $(y, (T_1, ..., T_m, x_{d,n+1}, x_{d,n+2}, ...))$, and for every term M the tree $PT^{\eta}_{d+1}(M)$ is a subprobability distribution over such value trees.

Definition 5.1.0.1. For $d \in \mathbb{N}$, given any variable y and any $s \in \mathbb{Z}$ we define the sets of value Böhm tree of depth $d+1$ in the class (y, s) by

$$\mathcal{VT}^{\eta}_{d+1,(y,s)} = \{(y, (T_n)) \mid \exists m \in \mathbb{N}, \exists s \in \mathbb{Z} : \forall n > m, T_n = x_{d,n-s}\}.$$

It is easy to see that if $(y, s) \neq (y', s')$ then $\mathcal{VT}^{\eta}_{d+1,(y,s)} \cap \mathcal{VT}^{\eta}_{d+1,(y',s')} = \emptyset$. On the other hand we do not have $\mathcal{VT}^{\eta}_{d+1} = \bigcup_{y,s} \mathcal{VT}^{\eta}_{d+1,(y,s)}$, but for every head normal value h there is a (unique) class (y, s) such that $VT^{\eta}_{d+1}(h) \in \mathcal{VT}^{\eta}_{d+1,(y,s)}$.

We call *term Böhm tree* the trees T such that $T = PT^{\eta}_{d}(M)$ for some term M. To have a separation result on terms we only need to separate term Böhm trees.

Using standard Böhm out techniques we know that two head normal values whose Böhm trees are in different classes are separable. To separate two different value Böhm trees in the same class we use functions $\varphi : [0; 1]^k \to [0; 1]$ to encode the difference of the subtrees into a single probability.

Definition 5.1.0.2. The *evaluation functions* from $[0; 1]^k$ to $[0; 1]$ are

- the constant functions $\alpha_1, ..., \alpha_k \mapsto \alpha$ for $\alpha \in [0; 1]$;
- the projections $\alpha_1, ..., \alpha_k \mapsto \alpha_i$ for $i \leq k$;
- the products $\alpha_1, ..., \alpha_k \mapsto \varphi(\alpha_1, ..., \alpha_k) \times \psi(\alpha_1, ..., \alpha_k)$ where φ and ψ are evaluation functions;
- the probabilistic sums $\alpha_1, ..., \alpha_k \mapsto p \times \varphi(\alpha_1, ..., \alpha_k) + (1 - p) \times \psi(\alpha_1, ..., \alpha_k)$ where $p \in [0; 1]$ and φ and ψ are evaluation functions.

Definition 5.1.0.3. For $d \in \mathbb{N}$, and evaluation structure $\mathcal{S}$ of depth $d+1$ is given by:

- a finite set of classes $\mathcal{C}^{\mathcal{S}} \subset_f Var \times \mathbb{Z}$;
- for all $c \in \mathcal{C}^{\mathcal{S}}$ an arity $m^{\mathcal{S}}_c \in \mathbb{N}$ and an evaluation function $\varphi^{\mathcal{S}}_c : [0; 1]^{m^{\mathcal{S}}_c} \to [0; 1]$;
- for all $c \in \mathcal{C}^{\mathcal{S}}$ and all $i \leq m^{\mathcal{S}}_c$ an evaluation structure $\mathcal{S}_{c,i}$ of depth d.

There is a unique evaluation structure of depth 0, with $\mathcal{C} = \emptyset$.

Definition 5.1.0.4. Given a tree T of depth d and an evaluation structure $\mathcal{S}$ of depth d, the evaluation $\mathcal{S}(T)$ of T by $\mathcal{S}$ is given by

$$\mathcal{S}(T) = \sum_{t \in \mathcal{VT}_d^\eta} T(t) \times \mathcal{S}^v(t)$$

$$\mathcal{S}^v(y,(T_n)) = \begin{cases} \varphi_c^{\mathcal{S}}\left(\mathcal{S}_{c,1}(T_1), ..., \mathcal{S}_{c,m_c^{\mathcal{S}}}(T_{m_c^{\mathcal{S}}})\right) & \text{if } (y,(T_n)) \in \mathcal{VT}_{d+1,c}^\eta \text{ and } c \in \mathcal{C}^{\mathcal{S}} \\ 0 & \text{otherwise} \end{cases}.$$

What we want to prove is that given finitely many trees T_1,...,T_k there is an evaluation structure $\mathcal{S}$ such that for all $i, j \leq k$, $\mathcal{S}(T_i) = \mathcal{S}(T_j)$ if and only if $T_i = T_j$.

The evaluation functions are some particular polynomials, so the evaluation of a tree is a possibly infinite sum of polynomials functions of the evaluation of its subtrees. For that reason the following result will be very useful.

Proposition 5.1.0.1. *Given pairwise distinct probabilities $\alpha_i \in [0;1]$ for $i \in I$ with $I \subset \mathbb{N}$ and given two probability subdistributions $(p_i)_{i\in I}$ and $(p'_i)_{i\in I}$ we have*

$$\left(\forall n \in \mathbb{N}, \sum_{i\in I} p_i \alpha_i^n = \sum_{i \in I} p'_i \alpha_i^n\right) \Rightarrow \forall i \in I, p_i = p'_i$$

Proof. The following proof is thanks to Franck Boyer.

If for all n we have $\sum_{i\in I} p_i\alpha_i^n = \sum_{i\in I} p'_i\alpha_i^n$ then for all polynomial P we have $\sum_{i\in I} p_i P(\alpha_i) = \sum_{i\in I} p'_i P(\alpha_i)$.

If $f : [0;1] \to \mathbb{R}$ is continuous let $\|f\| = \max_{x\in[0;1]}|f(x)|$. We know by the Stone-Weierstrass theorem that for every continuous f and for all $\epsilon > 0$ there is a polynomial P such that $\|f-P\| \leq \epsilon$. Thus for all continuous function $f : [0;1] \to \mathbb{R}$ we have $\sum_{i\in I} p_i f(\alpha_i) = \sum_{i\in I} p'_i f(\alpha_i)$.

Now let $i \in I$, we want to show $p_i = p'_i$. Let $\varphi : \mathbb{R} \to \mathbb{R}$ be a continuous function with compact support such that $\varphi(0) = 1$, $\|\varphi\|= 1$ and $\varphi(x) = 0$ when $|x| > 1$. We can choose for instance $\varphi : x \mapsto \max(0, 1 - |x|)$. Then for $\epsilon > 0$ let $f_\epsilon : x \mapsto \varphi\left(\frac{x-\alpha_i}{\epsilon}\right)$, we have $f_\epsilon(\alpha_i) = 1$, $\|f_\epsilon\|= 1$ and $f_\epsilon(x) = 0$ when $|x| > \epsilon$.

For all $\epsilon > 0$ there exists a finite subset $J \subset_f I$ with $i \in J$ such that $\sum_{j\in I\setminus J} p_j \leq \epsilon$ and $\sum_{j\in I\setminus J} p'_j \leq \epsilon$. Then there exists $\delta > 0$ such that for all $j \in J\setminus\{i\}$, $|\alpha_i - \alpha_j| > \delta$, hence $f_\delta(\alpha_j) = 0$. We get

$$\sum_{j\in I} p_j f_\delta(\alpha_j) = \sum_{j\in I} p'_j f_\delta(\alpha_j)$$
$$p_i + \sum_{j\in I\setminus J} p_j f_\delta(\alpha_j) = p'_i + \sum_{j\in I\setminus J} p'_j f_\delta(\alpha_j)$$

We have $|p_i - p'_i| \leq \epsilon$ for all $\epsilon > 0$ so $p_i = p'_i$. □

Another useful result is that we can combine the evaluations functions to separate elements of $[0;1]^m$.

Proposition 5.1.0.2. *Given finitely many vectors $\vec{\alpha}_i \in \mathbb{N}^m$, given finitely many evaluation functions $\varphi_1, ..., \varphi_k : [0;1]^m \to [0;1]$ with $k \geq 1$, there is an evaluation function $\psi : [0;1]^m \to [0;1]$ such that for all i, i'*

$$\psi(\vec{\alpha}_i) = \psi(\vec{\alpha}_{i'}) \text{ iff } \forall j \leq k, \varphi_j(\vec{\alpha}_i) = \varphi_j(\vec{\alpha}_{i'}).$$

Proof. We reason by induction on k. If $k = 1$ then $\psi = \varphi_1$.

Otherwise if for all i, i' and $j \leq k$ we have $\varphi_j(\vec{\alpha}_i) = \varphi_j(\vec{\alpha}_{i'})$ then $\psi = \varphi_{k+1}$. If it is not the case by induction hypothesis there is ψ such that the $\psi(\vec{\alpha}_i)$ are equal if and only if the $\varphi_j(\vec{\alpha}_i)$ are equal for all $j \leq k$. Then let $\delta = \min\{|\psi(\vec{\alpha}_i) - \psi(\vec{\alpha}_{i'})| \neq 0\}$, we have

$$\begin{array}{ll} & \psi(\vec{\alpha}_i) + \frac{\delta}{2}\varphi_{k+1}(\vec{\alpha}_i) = \psi(\vec{\alpha}_{i'}) + \frac{\delta}{2}\varphi_{k+1}(\vec{\alpha}_{i'}) \\ \text{iff} & \psi(\vec{\alpha}_i) = \psi(\vec{\alpha}_{i'}) \text{ and } \varphi_{k+1}(\vec{\alpha}_i) = \varphi_{k+1}(\vec{\alpha}_{i'}) \\ \text{iff} & \qquad \forall j \leq k+1, \varphi_j(\vec{\alpha}_i) = \varphi_j(\vec{\alpha}_{i'}). \end{array}$$

The evaluation function $\vec{\alpha} \mapsto \frac{2}{2+\delta}\psi(\vec{\alpha}) + \frac{\delta}{2+\delta}\varphi_{k+1}(\vec{\alpha})$ has the required property. ☐

With these two results we can directly prove that evaluation structures can separate trees.

Proposition 5.1.0.3. *Given term Böhm trees T_1,...,T_k of depth d there is an evaluation structure $\mathcal{S}$ such that for all $i, j \leq k$, $\mathcal{S}(T_i) = \mathcal{S}(T_j)$ if and only if $T_i = T_j$.*

Proof. To simplify the proof we assume the T_i's pairwise distinct.

We reason by induction on d. If $d = 0$ there are a unique tree and a unique evaluation structure of depth 0.

Otherwise we first cut the trees in width. Observe that for any tree T and any value tree $t = (y, (U_n)) \in \mathcal{VT}^{\eta}_{d+1,c}$ of depth $d+1$, $T(t)$ is the limit when $m \to \infty$ of the sequence

$$\sum_{t'=(y,(U'_n))\in\mathcal{VT}^{\eta}_{d+1,c} \text{ s.t.} \forall n \leq m, U_n = U'_n} T(t').$$

So given two distinct trees $T \neq T'$ of depth $d+1$ there is $t = (y, (U_n))$ of class c such that $T(t) \neq T'(t)$, and there is $m \in \mathbb{N}$ such that

$$\sum_{t'=(y,(U'_n))\in\mathcal{VT}^{\eta}_{d+1,c} \text{ s.t.} \forall n \leq m, U_n = U'_n} T(t') \neq \sum_{t'=(y,(U'_n))\in\mathcal{VT}^{\eta}_{d+1,c} \text{ s.t.} \forall n \leq m, U_n = U'_n} T'(t').$$

More generally given finitely many pairwise distinct trees T_1,...,T_k of depth $d+1$ there is $m \in \mathbb{N}$ such that for all $i \neq j$ there are a class $c_{i,j}$ and trees $U_{i,j,1}$,...,$U_{i,j,m}$ of depths d with

$$\begin{array}{ll} & \sum_{t'=(y,(U'_n))\in\mathcal{VT}^{\eta}_{d+1,c_{i,j}} \text{ s.t.} \forall n \leq m, U_{i,j,n} = U'_n} T_i(t') \\ \neq & \sum_{t'=(y,(U'_n))\in\mathcal{VT}^{\eta}_{d+1,c_{i,j}} \text{ s.t.} \forall n \leq m, U_{i,j,n} = U'_n} T_j(t'). \end{array}$$

For all $i \leq k$ we define

$$\tau_i^1 : c, (U_1, ..., U_m) \mapsto \sum_{t'=(y,(U'_n)) \in \mathcal{VT}^\eta_{d+1,c} \text{ s.t.} \forall n \leq m, U_n = U'_n} T_i(t').$$

The τ_i^1's are pairwise distinct subprobability distributions over $Var \times \mathbb{Z} \times (\mathcal{PT}^\eta_d)^m$.

We can use the induction hypothesis to associate to every τ_i^1 a subprobability distribution τ_i^2 over $Var \times \mathbb{Z} \times [0;1]^m$. We can find a finite set $\mathcal{C}$ of labels and finite sets $\mathcal{U}_{c,n}$ of trees of depth d for $c \in \mathcal{C}$ and $n \leq m$ such that

- for all $i \neq j$, $c_{i,j} \in \mathcal{C}$ and for all $n \leq m$, $U_{i,j,n} \in \mathcal{U}_{c,n}$;
- and for all $i \neq j$,

  $$\sum_{t \in \mathcal{R}} T_i(t) < |\tau_i^1(y_{i,j}, s_{i,j}, (U_{i,j,1}, ..., U_{i,j,m})) - \tau_j^1(y_{i,j}, s_{i,j}, (U_{i,j,1}, ..., U_{i,j,m}))|$$

  where $\mathcal{R} = \{(y, (U_n)) \in \mathcal{VT}^\eta_{d+1,c} \mid c \notin \mathcal{C} \text{ or } \exists n \leq m : U_n \notin \mathcal{U}_{c,n}\}$.

Applying the induction hypothesis to the sets $\mathcal{U}_{c,n}$ gives evaluation structures $\mathcal{S}_{c,n}$. Then for $i \leq k$ we define

$$\tau_i^2 : c, (\alpha_1, ..., \alpha_m) \mapsto \sum_{\mathcal{S}_{c,1}(U_1)=\alpha_1} ... \sum_{\mathcal{S}_{c,m}(U_m)=\alpha_m} \tau_i^1(c, (U_1, ..., U_m)).$$

For all $c \in \mathcal{C}$ and $U_1, ..., U_m \in \mathcal{U}_{c,1} \times ... \times \mathcal{U}_{c,m}$ we have for all $i \leq k$

$$\tau_i^1(c, (U_1, ..., U_m)) \leq \tau_i^2(c, (\mathcal{S}_{c,1}(U_1), ..., \mathcal{S}_{c,m}(U_m))) \leq \tau_i^1(c, (U_1, ..., U_m)) + \sum_{t \in \mathcal{R}} T_i(t).$$

In particular for $i \neq j$ we have

$$\tau_i^2(c_{i,j}, (\mathcal{S}_{c_{i,j},1}(U_{i,j,1}), ..., \mathcal{S}_{c_{i,j},m}(U_{i,j,m}))) \neq \tau_j^2(c_{i,j}, (\mathcal{S}_{c_{i,j},1}(U_{i,j,1}), ..., \mathcal{S}_{c_{i,j},m}(U_{i,j,m}))).$$

Next we want to build subprobability distributions τ_i^3 over $Var \times \mathbb{Z} \times [0;1]$. For each $c \in \mathcal{C}$ we simply use the previous result with the vectors $(\mathcal{S}_{c,1}(U_1), ..., \mathcal{S}_{c,m}(U_m))$ for $(U_1, ..., U_m) \in \mathcal{U}_{c,1} \times ... \times \mathcal{U}_{c,m}$ and the projections $\alpha_1, ..., \alpha_m \mapsto \alpha_n$ for $n \leq m$. This gives evaluation functions ψ_c such that the $\psi_c(\mathcal{S}_{c,1}(U_1), ..., \mathcal{S}_{c,m}(U_m))$ for $(U_1, ..., U_m) \in \mathcal{U}_{c,1} \times ... \times \mathcal{U}_{c,m}$ are pairwise distinct. Then for $i \leq k$ we define

$$\tau_i^3 : c, \alpha \mapsto \sum_{\psi_c(\alpha_1, ..., \alpha_m)=\alpha} \tau_i^2(c, (\alpha_1, ..., \alpha_m)).$$

Again we have

$$\tau_i^1(c, (U_1, ..., U_m)) \leq \tau_i^3(c, \psi_c(\mathcal{S}_{c,1}(U_1), ..., \mathcal{S}_{c,m}(U_m))) \leq \tau_i^1(c, (U_1, ..., U_m)) + \sum_{t \in \mathcal{R}} T_i(t)$$

for $c \in \mathcal{C}$ and $U_1, ..., U_m \in \mathcal{U}_{c,1} \times ... \times \mathcal{U}_{c,m}$ so the τ_i^3 are pairwise distinct.

The next step is to define τ_i^4 over $[0;1]$. Let us write $\mathcal{C} = \{c^0, ..., c^{l-1}\}$, for $i \leq k$ we define

$$\tau_i^4 : \alpha \mapsto \begin{cases} \tau_i^3(c^a, \beta) & \text{if } \alpha = \frac{a}{l} + \frac{\beta}{2l} \\ 0 & \text{otherwise} \end{cases} .$$

For $i \neq j$, if $\tau_i^3(c^a, \beta) \neq \tau_j^3(c^a, \beta)$ then $\tau_i^4\left(\frac{a}{l} + \frac{\beta}{2l}\right) \neq \tau_j^4\left(\frac{a}{l} + \frac{\beta}{2l}\right)$ so the τ_i^4 are pairwise distinct.

Then for all $i \neq j$, according to proposition 5.1.0.1 there is $n_{i,j} \in \mathbb{N}$ such that $\sum_{\alpha \in [0;1]} \tau_i^4(\alpha)\alpha^{n_{i,j}} \neq \sum_{\alpha \in [0;1]} \tau_j^4(\alpha)\alpha^{n_{i,j}}$.

Finally for all $c^a \in \mathcal{C}$ we can define an evaluation function φ_{c^a} as a linear combination of the evaluation functions

$$\alpha_1, ...\alpha_m \mapsto \left(\frac{a}{l} + \frac{1}{2l}\psi_{c^a}(\alpha_1, ..., \alpha_m)\right)^{n_{i,j}}$$

such that the sums

$$\sum_{(y,s) \in \mathcal{C}} \sum_{U_1, ..., U_m} \tau_i^1(c, (U_1, ..., U_m))\varphi_c(\mathcal{S}_{c,1}(U_1), ..., \mathcal{S}_{c,m}(U_m))$$

for $i \leq k$ are pairwise distinct. Then the evaluation structure given by $\mathcal{C}$, the evaluation function φ_c and the substructure $\mathcal{S}_{c,n}$ separates the trees T_1,...,T_k. $\square$

5.2 Evaluating terms

To obtain a separabilty result all we have left to do is to show that we can simulate at the level of terms the evaluation of a tree by an evaluation structure. Given finitely many terms M_1,...,M_k and an evaluation structure $\mathcal{S}$ of depth d we want to find a context C such that for all $i \leq k$, the term $C[M_i]$ converges with probability $\mathcal{S}(PT_d^\eta(M_i))$. As usual we can not obtain exactly this result, so we will need to work up to a small probability $\epsilon > 0$.

Definition 5.2.0.1. For $\alpha \in [0;1]$ and $\epsilon \geq 0$ we say that a term M *represents* α *up to* ϵ if

$$\alpha - \epsilon \leq \mathcal{P}(M =_{\beta+} I) \leq \mathcal{P}_\Downarrow(M) \leq \alpha + \epsilon$$

with $I = \lambda x.x$.

First it is easy to show that we can represent the evaluation functions by terms, and that this representation is uniformly continuous.

Proposition 5.2.0.1. *Given an evaluation function $\varphi : [0;1]^m \to [0;1]$ there is a term $\underline{\varphi}$ such that for all $\epsilon > 0$ there exists $\delta > 0$ such that whenever M_1,...,M_m represent α_1,...,α_m up to δ then $\underline{\varphi}\ M_1\ ...\ M_m$ represents $\varphi(\alpha_1, ..., \alpha_m)$ up to ϵ.*

Proof. We reason by induction on the definition of evaluation function.

- The term $\lambda a_1...a_m.(I +_\alpha \Omega)$ represents the constant function $\alpha_1, ..., \alpha_m \mapsto \alpha$.
- If M_1,...,M_m represent α_1,...,α_m up to ϵ then for all $i \leq m$ the term $\lambda a_1...a_m.a_i$ represents the projection $\alpha_1, ..., \alpha_m \mapsto \alpha_i$.
- If $\underline{\varphi}$ and $\underline{\psi}$ represent φ and ψ then for all $\epsilon \geq 0$ there is δ such that if M_1,...,M_m represent α_1,...,α_m up to δ then $\underline{\varphi}\ M_1\ ...\ M_m$ and $\underline{\psi}\ M_1\ ...\ M_m$ represent $\varphi(\alpha_1, ..., \alpha_m)$ and $\psi(\alpha_1, ..., \alpha_m)$ up to ϵ. Then

$$\begin{aligned}
&\mathcal{P}((\underline{\varphi}\ M_1\ ...\ M_m)\ (\underline{\psi}\ M_1\ ...\ M_m) =_{\beta+} I)\\
&\geq \mathcal{P}(\underline{\varphi}\ M_1\ ...\ M_m =_{\beta+} I) \times \mathcal{P}(\underline{\psi}\ M_1\ ...\ M_m =_{\beta+} I)\\
&\geq (\varphi(\alpha_1, ..., \alpha_m) - \epsilon)(\psi(\alpha_1, ..., \alpha_m) - \epsilon)\\
&\geq \varphi(\alpha_1, ..., \alpha_m) \times \psi(\alpha_1, ..., \alpha_m) - 2\epsilon
\end{aligned}$$

  and

$$\begin{aligned}
&\mathcal{P}_\Downarrow((\underline{\varphi}\ M_1\ ...\ M_m)\ (\underline{\psi}\ M_1\ ...\ M_m))\\
&\leq \mathcal{P}(\underline{\varphi}\ M_1\ ...\ M_m =_{\beta+} I) \times \mathcal{P}_\Downarrow(\underline{\psi}\ M_1\ ...\ M_m) + 2\epsilon\\
&\leq (\varphi(\alpha_1, ..., \alpha_m) + \epsilon)(\psi(\alpha_1, ..., \alpha_m) + \epsilon) + 2\epsilon\\
&\leq \varphi(\alpha_1, ..., \alpha_m) \times \psi(\alpha_1, ..., \alpha_m) + 4\epsilon + \epsilon^2.
\end{aligned}$$

  The term $\lambda a_1...a_m.(\underline{\varphi}\ a_1\ ...\ a_m)\ (\underline{\psi}\ a_1\ ...\ a_m)$ is a uniformly continuous representation of the function $\alpha_1, ..., \alpha_m \mapsto \varphi(\alpha_1, ..., \alpha_m) \times \psi(\alpha_1, ..., \alpha_m)$.

- Finally if $\underline{\varphi}$ and $\underline{\psi}$ represent φ and ψ then for all $\epsilon \geq 0$ there is δ such that if M_1,...,M_m represent α_1,...,α_m up to δ then $\underline{\varphi}\ M_1\ ...\ M_m$ and $\underline{\psi}\ M_1\ ...\ M_m$ represent $\varphi(\alpha_1, ..., \alpha_m)$ and $\psi(\alpha_1, ..., \alpha_m)$ up to ϵ. Then

$$\begin{aligned}
&\mathcal{P}((\underline{\varphi} +_p \underline{\psi})\ M_1\ ...\ M_m =_{\beta+} I) \\
&= p\mathcal{P}(\underline{\varphi}\ M_1\ ...\ M_m =_{\beta+} I) + (1-p)\mathcal{P}(\underline{\psi}\ M_1\ ...\ M_m =_{\beta+} I) \\
&\geq p(\varphi(\alpha_1, ..., \alpha_m) - \epsilon) + (1-p)(\psi(\alpha_1, ..., \alpha_m) - \epsilon) \\
&= p\varphi(\alpha_1, ..., \alpha_m) + (1-p)\psi(\alpha_1, ..., \alpha_m) - \epsilon
\end{aligned}$$

and

$$\begin{aligned}
&\mathcal{P}_{\Downarrow}((\underline{\varphi} +_p \underline{\psi})\ M_1\ ...\ M_m) \\
&= p\mathcal{P}_{\Downarrow}(\underline{\varphi}\ M_1\ ...\ M_m) + (1-p)\mathcal{P}_{\Downarrow}(\underline{\psi}\ M_1\ ...\ M_m) \\
&\leq p(\varphi(\alpha_1, ..., \alpha_m) + \epsilon) + (1-p)(\psi(\alpha_1, ..., \alpha_m) + \epsilon) \\
&= p\varphi(\alpha_1, ..., \alpha_m) + (1-p)\psi(\alpha_1, ..., \alpha_m) + \epsilon.
\end{aligned}$$

The term $\underline{\varphi} +_p \underline{\psi}$ represents the function $\alpha_1, ..., \alpha_m \mapsto p\varphi(\alpha_1, ..., \alpha_m) + (1 - p)\psi(\alpha_1, ..., \alpha_m)$.

$\square$

Next we want to mimic the action of an evaluation structure on terms. For any term M and $\epsilon > 0$ we have

$$M \twoheadrightarrow_{\beta/\equiv} \sum_i p_i.h_i + \left(1 - \sum_i p_i\right).M'$$

where the h_i's are head normal values and $(1 - \sum_i p_i)\,\mathcal{P}_{\Downarrow}(M') \leq \epsilon$. So for any evaluation structure $\mathcal{S}$ of depth $d+1$, if we have a context C such that for all i, $C[h_i]$ represents $\mathcal{S}^v(VT^{\eta}_{d+1}(h_i))$ up to ϵ and such that C commutes with sums, then $C[M]$ represents $\mathcal{S}(PT^{\eta}_{d+1}(M))$ up to 2ϵ.

Besides given a head normal value $h = \lambda x_1 ... x_n.y\ P_1\ ...\ P_m$ we have

$$\mathcal{S}^v(VT^{\eta}_{d+1}(h)) = \varphi^{\mathcal{S}}_{y,m-n}(\mathcal{S}_{y,m-n,1}(PT^{\eta}_d(P'_1)), ..., \mathcal{S}_{y,m-n,m^{\mathcal{S}}_{y,m-n}}(PT^{\eta}_d(P'_{m^{\mathcal{S}}_{y,m-n}})))$$

where $P'_i = \begin{cases} P_i & \text{if } i \leq m \\ x_{d,i-m+n} & \text{if } i > m \end{cases}$. Hence there is $\delta > 0$ such that if $C_i[P'_i]$ represents $\mathcal{S}_{y,m-n,i}(PT^{\eta}_d(P'_i))$ up to δ for $i \leq m^{\mathcal{S}}_{y,m-n}$ then

$$\underline{\varphi^{\mathcal{S}}_{y,m-n}}\ C_1[P'_1]\ ...\ C_{m^{\mathcal{S}}_{y,m-n}}[P'_{m^{\mathcal{S}}_{y,m-n}}]$$

represents $\mathcal{S}^v(VT^{\eta}_{d+1}(h))$ up to ϵ.

We can see that what we need to do is very similar to the usual separability techniques for the deterministic λ-calculus.

Definition 5.2.0.2. We describe the parallel substitutions by functions between terms $\sigma : \Lambda_+ \to \Lambda_+$ such that

- $\sigma(x) \neq x$ for finitely many variables x;
- $\sigma(\lambda x.M) = \lambda x.\sigma(M)$ if $\sigma(x) = x$;
- $\sigma(M\ N) = \sigma(M)\ \sigma(N)$;
- $\sigma(M +_p N) = \sigma(M) +_p \sigma(N)$.

We note Σ the set of all such functions.

Proposition 5.2.0.2. *For all $\sigma \in \Sigma$ there is a context C such that for all M, $C[M] \twoheadrightarrow_\beta \sigma(M)$.*

Proof. Let x_1,...,x_n be the variables such that $\sigma(x) \neq x$, then we can choose $C = (\lambda x_1...x_n.[\ ])\ \sigma(x_1)\ ...\ \sigma(x_n)$. ☐

Definition 5.2.0.3. For $n \in \mathbb{N}$ we define $R_n = \lambda x_1...x_{n+1}.x_{n+1}\ x_1\ ...\ x_{n+1}$.

Definition 5.2.0.4. Given a finite set of variables $Y \subset_f Var$ and $\delta \in \mathbb{N}$, the set $\Sigma_{Y,\delta} \subset \Sigma$ of substitutions respecting the conditions Y and δ is defined by $\sigma \in \Sigma_{Y,\delta}$ iff

- $\forall x \in Var, \sigma(x) \neq x \Rightarrow \exists n \in \mathbb{N} : \sigma(x) = R_n$
- $\forall y \in Y, \sigma(y) \neq y$
- $\forall x \in Var, \sigma(x) = R_n \Rightarrow n > \delta$
- $\forall x \neq x' \in Var, \sigma(x) = R_n$ and $\sigma(x') = R_{n'} \Rightarrow |n - n'| > \delta$.

Proposition 5.2.0.3. *Given finite sets of variables Y and Y' and integers δ and δ' we have $\Sigma_{Y,\delta} \cap \Sigma_{Y',\delta'} = \Sigma_{Y\cup Y',\max(\delta,\delta')}$.*

Proof. Immediate by definition. ☐

Proposition 5.2.0.4. *Given $d \in \mathbb{N}$ and finitely many head normal forms h_1,...,h_k*

- *for every family of integers $(m_{y,s})$*
- *there exists $Y \subset_f Var$ and $\delta \in \mathbb{N}$ such that*
- *for every $\sigma \in \Sigma_{Y,\delta}$ and every family $(F_{y,s})$ of terms*
- *there exist terms L_1,...,L_l*

such that for all $i \leq k$, if $h_i = \lambda x_1...x_n.y\ P_1\ ...\ P_m$ then

$$\sigma(h_i)\ L_1\ ...\ L_l =_{\beta+} F_{y,m-n}\ \sigma(P'_1)\ ...\ \sigma(P'_{m_{y,m-n}})$$

with $P'_i = \begin{cases} P_i & \text{if } i \leq m \\ x_{d,i-m+n} & \text{if } i > m \end{cases}$.

Proof. To simplify the notations we consider $d \in \mathbb{N}$ given and we simply write x_i for $x_{d,i}$.

Let us write $h_i = \lambda x_1 ... x_{n_i}.y_i\ P_{i,1}\ ...\ P_{i,m_i}$ for $i \leq k$.

Let $n \geq \max_{i \leq k} n_i$, we have for all $i \leq k$ that

$$h_i\ x_1\ ...\ x_n \twoheadrightarrow_\beta y_i\ P_{i,1}\ ...\ P_{i,m_i}\ x_{n_i+1}\ ...\ x_n.$$

Then we also have

$$\begin{aligned}\sigma(h_i)\ \sigma(x_1)\ ...\ \sigma(x_n) &= \sigma(h_i\ x_1\ ...\ x_n)\\ &\twoheadrightarrow_\beta \sigma(y_i\ P_{i,1}\ ...\ P_{i,m_i}\ x_{n_i+1}\ ...\ x_n)\\ &= \sigma(y_i)\ \sigma(P_{i,1})\ ...\ \sigma(P_{i,m_i})\ \sigma(x_{n_i+1})\ ...\ \sigma(x_n)\\ &= \sigma(y_i)\ \sigma(P'_{i,1})\ ...\ \sigma(P'_{i,n+m_i-n_i}).\end{aligned}$$

Given any $Y \subset_f$ *Var* such that $y_i \in Y$ for all $i \leq k$ and any $\delta \in \mathbb{N}$ such that $\delta \geq n + m_i - n_i$, for all $l \in \mathbb{N}$ large enough and all terms $F_1,...,F_l$, we have for all $i \leq k$ with $\sigma(y_i) = R_a$

$$\sigma(h_i)\ \sigma(x_1)\ ...\ \sigma(x_n)\ F_1\ ...\ F_l \twoheadrightarrow_\beta F_{a-(n+m_i-n_i)}\ \sigma(P'_{i,1})\ ...\ \sigma(P'_{i,n+m_i-n_i})\ F_1\ ...\ F_l.$$

Besides if $\delta \geq |(m_i - n_i) - (m_j - n_j)|$ for all $i \neq j$, we have for all $i, j \leq k$ that if $\sigma(y_i) = R_a$, $\sigma(y_j) = R_b$ and $a - (n + m_i - n_i) = b - (n + m_j - n_j)$ then necessarily $y_i = y_j$ and $m_i - n_i = m_j - n_j$.

So to sum up:

- consider given $d \in \mathbb{N}$, the head normal values $h_1,...,h_k$ and a family of integers $(m_{y,s})$;
- let n such that $n \geq n_i$ and $n + m_i - n_i \geq m_{y_i,m_i-n_i}$ for all $i \leq k$;
- let $Y = \{y_1; ...; y_k\}$ and let us choose δ such that $\delta \geq n + m_i - n_i$ for all $i \leq k$ and $\delta \geq |(m_i - n_i) - (m_j - n_j)|$ for all $i \neq j$;
- then consider given $\sigma \in \Sigma_{Y,\delta}$ and a family of terms $(F_{y,s})$;
- for $i \leq k$ let $a_i \in \mathbb{N}$ the integer such that $\sigma(y_i) = R_{a_i}$ and let $l \in \mathbb{N}$ such that $l \geq a_i - (n + m_i - n_i)$ for all $i \leq k$;
- for $i \leq k$ let $F_{a_i-(n+m_i-n_i)} = \lambda z_1 ... z_{n+m_i-n_i+l}.F_{y_i,m_i-n_i}\ z_1\ ...\ z_{m_{y_i,m_i-n_i}}$, and let us choose any term for F_q if $q \leq l$ is not of this form.

Then for all $i \leq k$ we have

$$\sigma(h_i)\ \sigma(x_1)\ ...\ \sigma(x_n)\ F_1\ ...\ F_l \twoheadrightarrow_\beta F_{y_i,m_i-n_i}\ \sigma(P'_{i,1})\ ...\ \sigma(P'_{i,m_{y_i,m_i-n_i}}).$$

$\square$

With these results we can simulate on the terms the evaluation of some trees.

Proposition 5.2.0.5.

- *Given an evaluation structure $\mathcal{S}$ of depth d, finitely many terms M_1,...,M_k and $\epsilon > 0$,*
- *there are a finite set of variables Y and an integer δ such that*
- *for all $\sigma \in \Sigma_{Y,\delta}$,*
- *there are terms L_1,...,L_l*

such that for all $i \leq k$, $\sigma(M_i)\ L_1\ ...\ L_l$ represents $\mathcal{S}(PT^\eta_d(M_i))$ up to ϵ.

Proof. We already sketched the proof of this result.

We reason by induction on d. If $d = 0$ then for all $i \leq k$, $\mathcal{S}(PT^\eta_d(M_i)) = 0$ and we can find Y and δ such that for all $\sigma \in \Sigma_{Y,\delta}$, the sequence of terms $\Omega, ..., \Omega$ works.

Otherwise for all $i \leq k$ we have

$$M_i \twoheadrightarrow_{\beta/\equiv} \sum_j p_{i,j}.h_{i,j} + \left(1 - \sum_j p_{i,j}\right).M'_i$$

where the $h_{i,j}$ are head normal values and $\left(1 - \sum_j p_{i,j}\right)\mathcal{P}_{\Downarrow}(M'_i) \leq \epsilon$.

The previous result applied with d, the head normal values $h_{i,j}$ and the family $\left(m_{y,s} = \begin{cases} m^{\mathcal{S}}_{y,s} & \text{if } (y,s) \in \mathcal{C}^{\mathcal{S}} \\ 0 & \text{otherwise} \end{cases}\right)$ gives some $Y_0 \subset_f Var$ and $\delta_0 \in \mathbb{N}$.

For all $(y, s) \in \mathcal{C}^{\mathcal{S}}$ there is a term $\underline{\varphi^{\mathcal{S}}_{y,s}}$ which uniformly continuously represents $\varphi^{\mathcal{S}}_{y,s}$, and the uniform continuity applied to ϵ gives some ϵ'.

Now we write $h_{i,j} = \lambda x_{d,1}...x_{d,n_{i,j}}.y_{i,j}\ P_{i,j,1}\ ...\ P_{i,j,m_{i,j}}$, we define

$$P'_{i,j,q} = \begin{cases} P_{i,j,q} & \text{if } q \leq m_{i,j} \\ x_{d,q-m_{i,j}+n_{i,j}} & \text{if } q > m_{i,j} \end{cases}.$$

For all $(y, s) \in \mathcal{C}^{\mathcal{S}}$ and for all $q \leq m^{\mathcal{S}}_{y,s}$, we apply the induction hypothesis with $\mathcal{S}_{y,s,q}$, the terms $P'_{i,j,q}$ for i and j such that $(y_{i,j}, m_{i,j} - n_{i,j}) = (y, s)$ and ϵ' to get $Y_{y,s,q}$ and $\delta_{y,s,q}$.

Let

$$Y = Y_0 \cup \bigcup_{(y,s)\in\mathcal{C}^{\mathcal{S}}} \bigcup_{q\leq m_{y,s}} Y_{y,s,q}$$
$$\delta = \max\{\delta_0\} \cup \{\delta_{y,s,q} \mid (y,s) \in \mathcal{C}^{\mathcal{S}}, q \leq m_{y,s}\}.$$

We have

$$\Sigma_{Y,\delta} = \Sigma_{Y_0,\delta_0} \cap \bigcap_{(y,s)\in\mathcal{C}^{\mathcal{S}}} \bigcap_{q\leq m_{y,s}} \Sigma_{Y_{y,s,q},\delta_{y,s,q}}.$$

Let $\sigma \in \Sigma_{Y,\delta}$, for all $(y, s) \in \mathcal{C}^{\mathcal{S}}$ and $q \leq m_{y,s}$ we have $\sigma \in \Sigma_{Y_{y,s,q},\delta_{y,s,q}}$ so there are terms $L_{y,s,q,1}$,...,$L_{y,s,q,l_{y,s,q}}$ such that for all i and j such that $(y_{i,j}, m_{i,j} - n_{i,j}) = (y, s)$ we have that $\sigma(P'_{i,j,q})\ L_{y,s,q,1}\ ...,\ L_{y,s,q,l_{y,s,q}}$ represents $\mathcal{S}_{y,s,q}(PT^\eta_d(P'_{i,j,q}))$ up to ϵ'.

For $(y, s) \in \mathcal{X}^{\mathcal{S}}$ we define

$$F_{y,s} = \lambda z_1 ... z_{m_{y,s}}.\underline{\varphi^{\mathcal{S}}_{y,s}} \; (z_1 \; L_{y,s,1,1} \; .. \; L_{y,s,1,l_{y,s,1}}) \; .. \; (z_{m_{y,s}} \; L_{y,s,m_{y,s},1} \; .. \; L_{y,s,m_{y,s},l_{y,s,m_{y,s}}}).$$

For all i, j such that $(y_{i,j}, m_{i,j} - n_{i,j}) = (y, s)$ the term $F_{y,s} \; \sigma(P'_{i,j,1}) \; ... \; \sigma(P'_{i,j,m_{y,s}})$ represents $\varphi^{\mathcal{S}}_{y,s}(\mathcal{S}_{y,s,1}(PT^{\eta}_{d}(P'_{i,j,1})), ..., \mathcal{S}_{y,s,m^{\mathcal{S}}_{y,s}}(PT^{\eta}_{d}(P'_{i,j,m^{\mathcal{S}}_{y,s}})))$ up to ϵ, i.e. it represents $\mathcal{S}^{v}(VT^{\eta}_{d+1}(h_{i,j}))$.

We define $F_{y,s} = \Omega$ if $(y, s) \notin \mathcal{C}^{\mathcal{S}}$, as $\sigma \in \Sigma_{Y_0,\delta_0}$ there are terms L_1,...,L_l such that for all i, j,

$$\sigma(h_{i,j}) \; L_1 \; ... \; L_l =_{\beta+} F_{y_{i,j},m_{i,j}-n_{i,j}} \; \sigma(P'_{i,j,1}) \; ... \; \sigma(P'_{i,j,m_{y_{i,j},m_{i,j}-n_{i,j}}}).$$

Finally for all $i \leq k$

$$\sigma(M_i) \; L_1 \; ... \; L_l =_{\beta+} \sum_j p_{i,j}.\sigma(h_{i,j}) \; L_1 \; ... \; L_l + \left(1 - \sum_j p_{i,j}\right).M''_i$$

so $\sigma(M_i) \; L_1 \; ... \; L_l$ represents $\mathcal{S}(PT^{\eta}_{d+1}(M_i))$ up to 2ϵ. □

Corollary 5.2.0.6. *Given an evaluation structure $\mathcal{S}$ of depth d and finitely many terms M_1,...,M_k, for all $\epsilon > 0$ there exists a context C such that for all $i \leq k$, $C[M_i]$ represents $\mathcal{S}(PT^{\eta}_{d}(M_i))$ up to ϵ.*

Theorem 5.2.0.7. *Given finitely many terms M_1,...,M_k, there exists a context C such that for all $i, j \leq k$,*

$$\mathcal{P}_{\Downarrow}(C[M_i) = \mathcal{P}_{\Downarrow}(C[M_j]) \textit{ if and only if } M_i =_{\mathcal{PB}^{\eta}} M_j.$$

Proof. The contextuality of the Böhm trees gives that for all $i, j \leq k$ if $M_i =_{\mathcal{PB}^{\eta}} M_j$ then for all C, $\mathcal{P}_{\Downarrow}(C[M_i]) = \mathcal{P}_{\Downarrow}(C[M_j])$.

Conversely there exists $d \in \mathbb{N}$ such that for all $i, j \leq k$, $M_i =_{\mathcal{PB}^{\eta}} M_j$ if and only if $PT^{\eta}_{d}(M_i) = PT^{\eta}_{d}(M_j)$. Then the proposition 5.1.0.3 applied to the trees $PT^{\eta}_{d}(M_1)$,...,$PT^{\eta}_{d}(M_k)$ gives an evaluations structure $\mathcal{S}$ such that for all $i, j \leq k$, $M_i =_{\mathcal{PB}^{\eta}} M_j$ if and only if $\mathcal{S}(PT^{\eta}_{d}(M_i)) = \mathcal{S}(PT^{\eta}_{d}(M_j))$.

Let $\epsilon > 0$ such that $2\epsilon < |\mathcal{S}(PT^{\eta}_{d}(M_i)) - \mathcal{S}(PT^{\eta}_{d}(M_j))|$ whenever $M_i \neq_{\mathcal{PB}^{\eta}} M_j$. There is a context C such that $C[M_i]$ represents $\mathcal{S}(PT^{\eta}_{d}(M_i))$ up to ϵ for all $i \leq k$. In particular we have for all $i \leq k$

$$\mathcal{S}(PT^{\eta}_{d}(M_i)) - \epsilon \leq \mathcal{P}_{\Downarrow}(C[M_i]) \leq \mathcal{S}(PT^{\eta}_{d}(M_i)) + \epsilon.$$

For all $i, j \leq k$ we have $\mathcal{P}_{\Downarrow}(C[M_i]) = \mathcal{P}_{\Downarrow}(C[M_j])$ if and only if $\mathcal{S}(PT^{\eta}_{d}(M_i)) = \mathcal{S}(PT^{\eta}_{d}(M_j))$. □

Theorem 5.2.0.8. *For any terms M and N,*

$$M =_{\mathcal{PB}^{\eta}} N \textit{ if and only if } M =_{obs} N.$$

Conclusion

We have described a probabilistic λ-calculus which resembles the usual deterministic one. To begin with, the λ-calculus is meant to be very simple and contains the bare minimum to get a complete model of computation. So to introduce non-determinism we preferred considering a syntax with binary labeled sums, rather than directly using distributions. And to avoid using side-effects we gave deterministic reductions rules for these sums, rather than implementing the notion of non-deterministic choice in the operational semantics. We then proved that allowing such a reduction under arbitrary context yields a confluent calculus, which enjoys a standardization property.

To carry the interpretation of the sums as non-deterministic choices inside the calculus we gave syntactic equivalences, such that for instance in the probabilistic cases sums modulo equivalence actually define probability distributions. An interesting result is that this equivalence has actually little influence on the reduction: reducing a term modulo equivalence is basically the same as reducing it and considering only the result modulo equivalence. The only exception to this is when we introduce irrelevant choices, saying that $M \to_{\text{split}} M +_l M$, but this is not something we want to do. We may decide not to allow such a thing, turning the equivalence into reduction rules, or we know that we can get rid of these splittings by extending the β-reductions.

Once we have a contextual and deterministic operational semantics we can define theories in a natural way. In the case of a positive quantification (such as with the simple non-deterministic calculus or the probabilistic calculus, but not with the algebraic calculus), there is also a straightforward notion of Böhm trees. In the probabilistic case we proved that the Böhm trees form a model of the calculus, and we established a separation result which implies that the infinitely extensional Böhm trees describe the observational equivalence on the terms.

To reach the goal we set, i.e. to prove that this particular theory is the maximum sensible theory, we also had to define a notion of sensibility. This one is unfortunately much less natural than the others. We still managed to find a definition which allows us to state that in every sensible theory the diverging part of a term is equal to the diverging term Ω up to an arbitrary probability $\epsilon > 0$. This is sufficient for us to conclude.

Final theorem. *The infinitely extensional Böhm tree equality $=_{\mathcal{PB}^\eta}$ (and the observational equivalence $=_{obs}$) is the maximum of all probabilistic sensible continuous consistent λ-theories.*

Proof. The theorem 4.2.2.25 states that the relation $=_{\mathcal{PB}^\eta}$ is a theory, and moreover the theorem 5.2.0.8 states that it corresponds to the observational equivalence. Besides $=_{\mathcal{PB}^\eta}$ is clearly continuous, strongly sensible and consistent. So we want that if M and N have different infinitely extensional Böhm trees, or equivalently if they are not observationally equivalent, then every sensible continuous theory $=_{\mathcal{T}}$ such that $M =_{\mathcal{T}} N$ is inconsistent.

The proof of the separation theorem 5.2.0.7 gives probabilities $\alpha \neq \beta$ such that for all small $\epsilon > 0$ there is a context C such that

$$\alpha - \epsilon \leq \mathcal{P}(M =_{\beta+} I) \leq \mathcal{P}_{\Downarrow}(M) \leq \alpha + \epsilon$$
$$\beta - \epsilon \leq \mathcal{P}(N =_{\beta+} I) \leq \mathcal{P}_{\Downarrow}(N) \leq \beta + \epsilon.$$

In a sensible theory $=_{\mathcal{T}}$ we can use the proposition 3.2.2.5 to get

$$M =_{\mathcal{T}} (\alpha - 2\epsilon).I + (1 - \alpha - 2\epsilon).\Omega + 4\epsilon.M'$$
$$N =_{\mathcal{T}} (\beta - 2\epsilon).I + (1 - \beta - 2\epsilon).\Omega + 4\epsilon.N'$$

for some terms M' and N'.

In continuous theories we can simplify sums (proposition 3.2.1.1), so let us assume w.l.o.g. $\alpha < \beta$, we have

$$\frac{\beta - \alpha}{\beta - \alpha + 4\epsilon}.\Omega + \frac{4\epsilon}{\beta - \alpha + 4\epsilon}.M' =_{\mathcal{T}} \frac{\beta - \alpha}{\beta - \alpha + 4\epsilon}.I + \frac{4\epsilon}{\beta - \alpha + 4\epsilon}.N'.$$

Then for all $\epsilon > 0$ there are terms M' and N' such that

$$\Omega +_{1-\epsilon} M' =_{\mathcal{T}} I +_{1-\epsilon} N'.$$

By continuity this implies

$$\Omega =_{\mathcal{T}} I.$$

From there the proof of inconsistency is the same as for deterministic theories. Using a fixed point operator Θ, the term $\Theta\ (\lambda xy.x)$ is unsolvable so

$$\Theta\ (\lambda xy.x) =_{\mathcal{T}} I$$

and $\Theta\ (\lambda xy.x) \twoheadrightarrow_\beta \lambda y.\,(\Theta\ (\lambda xy.x))$ so

$$\Theta\ (\lambda xy.x) =_{\mathcal{T}} \lambda y.\,(\Theta\ (\lambda xy.x))$$

hence

$$I =_{\mathcal{T}} \lambda y.I.$$

Then for any term P we have

$$P =_{\mathcal{T}} I\ P =_{\mathcal{T}} (\lambda y.I)\ P =_{\mathcal{T}} I.$$

$=_{\mathcal{T}}$ is inconsistent. $\square$

Future work

We claimed that all the definitions involved in our final theorem are natural extensions of the definitions in the deterministic λ-calculus except for the sensibility of a theory. This definition is not without consequences: in all our work we assumed that our probabilities were computable, but the only reason we do so is to be able to prove the proposition 3.2.2.5 stating that the diverging part of a term is actually equal to Ω up to an arbitrary $\epsilon > 0$.

Restricting the calculus to computable probabilities is fine in practice, and besides we know how to prove that any continuous consistent theory in which the proposition 3.2.2.5 holds is included in the Böhm tree equality even if we allow arbitrary probabilities. But is would still be interesting to find a better definition for the sensibility which suits the general case.

Another question which arises is the relation between our theories and existing models of the probabilistic λ-calculus. For instance we know of a model in the probabilistic coherence spaces [3] [5], and even though it is presented as a model of a calculus with a probabilistic head reduction it is easy to see that it fits our notion of theory. In particular it is entirely contextual: if M and N have the same denotational semantics then so do $C[M]$ and $C[N]$ for any context C. Then we can wonder whether this model is fully abstract, i.e. if two terms have the same denotational semantics if and only if they are observationally equivalent.

We can also try to extend our results to other quantitative calculi. In the first two chapters of this thesis we tried be be as general as possible when describing our operational semantics, which gives a good basis to adapt our results. Unfortunately in the simple cases of non deterministic calculi with a unique, non indexed sum, the separation seems to fail.

And in the algebraic λ-calculus our results are hardly relevant. Indeed we know that given any algebraic term M we can build a term ω_M representing an infinite sum of M, such that $\omega_M \twoheadrightarrow_\beta M + \omega_M$. Then

$$0 \equiv \omega_M - \omega_M \twoheadrightarrow_\beta (M + \omega_M) - \omega_M \equiv M.$$

This way given any two terms M and N we can find a reduction $M \twoheadrightarrow_{\beta/\equiv} N$.

It is clear why this should not happen. This is like saying that if we consider integers with an infinite element ω such that $1 + \omega = \omega$ then

$$0 = \omega - \omega = 1 + \omega - \omega = 1.$$

The problem here is that the difference $\omega - \omega$ is not well defined, and so in the algebraic λ-calculus the difference $\omega_M - \omega_M$ should not be equal to 0.

So our results may be useful in the study of the algebraic λ-calculus, but one has to be very careful about the syntactic equivalence, and it is very likely that even our general framework would need to be adapted to get a meaningful calculus.

Bibliography

[1] H.P. Barendregt. *The Lambda Calculus, its Syntax and Semantics.* Stud. Log. F. Math., Vol. 103. North-Holland, 1984.

[2] Ugo Dal Lago and Margherita Zorzi. Probabilistic operational semantics for the lambda calculus. *RAIRO-Theoretical Informatics and Applications*, 46(3):413–450, 2012.

[3] Vincent Danos and Thomas Ehrhard. Probabilistic coherence spaces as a model of higher-order probabilistic computation. *Information and Computation*, 209(6):966–991, 2011.

[4] Ugo Deliguoro and Adolfo Piperno. Nondeterministic extensions of untyped λ-calculus. *Information and Computation*, 122(2):149–177, 1995.

[5] Thomas Ehrhard, Michele Pagani, and Christine Tasson. The computational meaning of probabilistic coherence spaces. In *Logic in Computer Science (LICS), 2011 26th Annual IEEE Symposium on*, pages 87–96. IEEE, 2011.

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

[7] J Martin E Hyland. A syntactic characterization of the equality in some models for the lambda calculus. *J. London Math. Soc*, 12(2):361–370, 1976.

[8] Claire Jones and Gordon D Plotkin. A probabilistic powerdomain of evaluations. In *Logic in Computer Science, 1989. LICS'89, Proceedings., Fourth Annual Symposium on*, pages 186–195. IEEE, 1989.

[9] Gordon D Plotkin. A powerdomain construction. *SIAM Journal on Computing*, 5(3):452–487, 1976.

[10] Nasser Saheb-Djahromi. Probabilistic lcf. In *International Symposium on Mathematical Foundations of Computer Science*, pages 442–451. Springer, 1978.

[11] Lionel Vaux. The algebraic lambda calculus. *Mathematical Structures in Computer Science*, 19(05):1029–1059, 2009.

[12] Christopher P Wadsworth. The relation between computational and denotational properties for scott's d_∞-models of the lambda-calculus. *SIAM Journal on Computing*, 5(3):488–521, 1976.