Séparabilité dans le λ-calcul probabiliste

LEVENTIS Thomas

Introduction

Ce rapport présente le travail effectué lors de mon stage de M2 à l'Institut de Mathématiques de Luminy, sous la direction de Lionel Vaux.

Dans le λ-calcul, il existe deux manières de concevoir des égalités : on peut utiliser des définitions syntaxiques, qui ne portent que sur la structure des λ-termes, ou bien l'on peut donner des constructions sémantiques, qui correspondent à des théories du λ-calcul. Un résultat intéressant est que la notion de séparabilité, qui est une construction purement syntaxique, est équivalente à l'incompatibilité sémantique entre termes.

Il existe de nombreuses extensions du λ-calcul, et pour chacune d'entre elles on peut se demander si ce résultat est conservé. Ici l'on s'intéresse au cas du λ-calcul probabiliste, qui ajoute au calcul de base une structure de somme probabiliste et introduit des choix non déterministes au cours des réductions. Le problème est que si l'on transpose directement la notion de séparabilité, on se rend compte qu'elle n'est pas adaptée à ce calcul, et qu'il faut l'enrichir pour qu'elle soit pertinente. De plus, le λ-calcul probabiliste ne se prête pas facilement à une approche sémantique.

Dans un premier temps, nous allons présenter les résultats disponibles dans le λ-calcul de base. Nous allons ensuite définir ce qu'est le λ-calcul probabiliste et détailler ce que signifie la séparabilité dans ce cadre. Enfin, nous présentons les résultats obtenus pour le λ-calcul probabiliste.

1 La séparabilité dans le λ-calcul

1.1 Définitions de base

On définit tout d'abord l'ensemble des λ-termes Λ sur un ensemble de variables *Var* par :

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

L'*abstraction* $\lambda x.M$ représente la fonction qui à x associe M, et l'*application* MN représente l'application de la fonction M à l'argument N. Concrètement, cette interprétation des λ-termes se traduit par la règle de réduction suivante, appelée β-réduction :

$$(\lambda x.M)N \to_\beta M[N/x]$$

où $M[N/x]$ désigne le terme obtenu en substituant le terme N à toutes les occurrences de la variable x dans M.

Cette règle peut également s'appliquer à l'intérieur d'un terme. On formalise cela en définissant des *contextes* par :

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

et en étendant la réduction à $C[M] \to_\beta C[N]$ si $M \to_\beta N$, où $C[M]$ est le terme obtenu en remplaçant $[\,]$ par M dans $C[\,]$ [1].

On note $\to_\beta^*$ la clôture réflexive transitive de $\to_\beta$, et $=_\beta$ sa clôture réflexive transitive symétrique.

Lorsque l'on veut s'intéresser à la sémantique des λ-termes, on ne considère plus les termes eux-mêmes mais leur classe d'équivalence pour la relation $=_\beta$. Ainsi si l'on se donne une interprétation $\underline{n}$ pour chaque entier $n \in \mathbb{N}$, on considère qu'une fonction $f : \mathbb{N} \to \mathbb{N}$ est représentée par un terme M_f si $\forall n, M_f\underline{n} =_\beta \underline{f(n)}$, et l'on fait abstraction des étapes de β-réduction qui justifient cette égalité.

1. On ne parlera pas ici des règles de manipulation des variables ni de leur comportement par rapport aux diverses substitutions. Pour plus de détails, se référer par exemple à *Lambda-calculus, types and models* de Jean-Louis Krivine

Selon le problème abordé, on souhaite parfois ajouter des égalités entre termes dont les différences ne sont pas pertinentes. De manière générale, une égalité sur les λ-termes, également appelée *λ-théorie*, est une relation θ qui vérifie :

- θ est une relation d'équivalence ;
- si $M \; \theta \; N$ alors pour tout contexte $C[\,]$, $C[M] \; \theta \; C[N]$;
- si $M \rightarrow_\beta N$ alors $M \; \theta \; N$.

Un exemple simple de théorie peut être obtenue en considérant la η-réduction du λ-calcul :

$$\lambda x.Mx \rightarrow_\eta M, x \notin FV(M)$$

La clôture réflexive transitive symétrique et par contexte de $\rightarrow_\beta \cup \rightarrow_\eta$ est une λ-théorie. De manière générale, une théorie θ qui vérifie $M \rightarrow_\eta N \Rightarrow M \; \theta \; N$ est qualifiée d'*extensionnelle*.

Une autre classe de théorie souvent utilisée lors de l'étude de la signification calculatoire des λ-termes est celle des théories dites sensées. En général, on s'intéresse principalement aux termes qui correspondent au résultat d'un calcul achevé, et qui ne peuvent donc plus se réduire. Plus précisément :

- une *forme normale* est un terme qui ne peut plus être β-réduit.
- une *forme normale de tête* est un terme de la forme $\lambda x_1 ... x_n . y M_1 ... M_m$. C'est une forme normale pour la *réduction de tête*, qui n'autorise pas à réduire les sous-termes à droite d'une application.

Un terme β-équivalent à une forme normale est dit *normalisable*. Un terme équivalent à une forme normale de tête est dit *solvable*, car il existe un environnement qui le rend normalisable : il suffit d'instancier la variable de tête y par un terme qui prend m arguments et renvoie une forme normale, par exemple l'identité $\lambda x.x$. Les autres termes sont ceux dont la réduction n'aboutira jamais, ils sont appelés *insolvables*. L'exemple le plus simple de terme insolvable est $\Omega = (\lambda x.xx)(\lambda x.xx)$.

Une théorie est alors dite *sensée* si elle identifie tous les termes insolvables. En général, on utilise le symbole Ω pour désigner n'importe quel terme insolvable, et l'on a, pour tous $x \in \mathit{Var}$ et M, $\lambda x.\Omega = \Omega$, $\Omega M = \Omega$ et $\Omega[M/x] = \Omega$. En revanche, on a a priori $M\Omega \neq \Omega$. Typiquement, une forme normale de tête est par définition différente de Ω sans qu'aucune contrainte ne s'applique aux M_i.

Par la suite, on se place dans un cadre extensionnel sensé, et l'on notera $M = N$ pour dire que M et N sont égaux modulo β-réduction, η-réduction et identification des termes insolvables.

1.2 Séparabilité

On sait donc qu'à l'exception de Ω, tout terme est solvable et par conséquent on peut lui passer des arguments et instancier ses variables libres de manière à l'envoyer sur n'importe quel autre terme. La question qui nous intéresse est de savoir cela peut être effectué conjointement pour plusieurs termes distincts.

Definition 1.1. Des termes $M_1, ..., M_p$ sont dits *séparables* si pour tous termes $F_1, ..., F_p$ il existe un contexte $C[\,]$ tel que $\forall i, C[M_i] = F_i$.

On remarque que l'on peut se contenter de choisir les F_i égaux à des variables, puisque si un contexte $C[\,]$ envoie les M_i sur des variables x_i distinctes, alors $(\lambda x_1...x_p.C[\,])F_1...F_p$ envoie M_i sur le terme F_i.

Il apparait rapidement que tous les termes différents de Ω ne sont pas séparables deux à deux. En effet, considérer Ω comme étant totalement à part n'est pas pertinent, puisqu'il peut apparaître à l'intérieur d'autres termes. Par exemple $\lambda x.x\Omega \neq \Omega$, et pourtant $\lambda x.x\Omega$ et $\lambda x.xy$ ne sont pas séparables. On peut étendre la notion de séparabilité pour prendre en compte les termes insolvables.

Definition 1.2. Deux termes M et N sont dits *semi-séparables* si pour toute variable x, il existe un contexte $C[\,]$ tel que $C[M] = x$ et $C[N] = \Omega$, ou inversement.

Un premier résultat de séparabilité peut être obtenu en comparant la structure en surface des formes normales de tête. Étant donnés deux termes $M = \lambda x_1...x_n.yM_1...M_m$ et $N = \lambda x_1..x_{n'}.zN_1...N_{m'}$, on dit que M et N sont similaires, et on note $M \sim N$, si $y = z$ et $m - n = m' - n'$. C'est une relation d'équivalence, dont on note $cl(y, m-n)$ les classes.

On a également que si M est solvable et N est insolvable, alors M et N sont semi-séparables. Par conséquent, deux termes M et N sont soit séparables ou semi-séparables, soit $M \sim N$. Dans ce dernier cas, on voudrait montrer que s'il existe i tel que M_i et N_i sont séparables ou semi-séparables, alors M et N le sont également.

La manière dont on considère ici la structure d'un terme correspond à sa représentation sous forme d'arbre de Böhm.

Definition 1.3. De manière informelle, on définit l'*arbre de Böhm* $BT(M)$ d'un terme M par induction sur sa structure :
- si M est insolvable alors $BT(M) = \Omega$;
- sinon, $M = \lambda x_1 ... x_n.yM_1...M_m$, et $BT(M)$ est l'arbre de racine $\lambda x_1 ... x_n.y$ et de fils $BT(M_1)$,...,$BT(M_m)$.

On utilise des chemins pour désigner les nœuds d'un arbre de Böhm : le chemin vide $\langle\rangle$ désigne la racine et le chemin $\alpha.i$ désigne le i-ème fils du nœud α. Pour tout terme M on note M_α le sous-terme de M correspondant au sous-arbre de $BT(M)$ de racine α.

On peut alors montrer que pour deux arbres de Böhm $BT(M)$ et $BT(N)$, pour tout chemin α tel que $BT(M)$ et $BT(N)$ sont définis en α et sont identiques sur tous les ancêtres de α, alors il existe un contexte $C[\,]$ tel que $C[M] \sim C[N] \Leftrightarrow M_\alpha \sim N_\alpha$ et $C[M]$ (respectivement $C[N]$) est solvable si et seulement si M_α (respectivement N_α) est solvable.

Théorème 1.1 (Hyland)**.** *Deux termes dont les arbres de Böhm sont distincts sont (semi-)séparables.*

Il est important de remarquer que deux termes distincts peuvent avoir le même arbre de Böhm. Cela vient du fait qu'un arbre de Böhm peut être infini. Ainsi pour tout opérateur de point fixe, c'est à dire tous les termes F tels que pour tout M $FM = M(FM)$, l'arbre de Böhm obtenu est toujours le même, à savoir $\lambda f.f(f(...))$. Or il existe des opérateurs de point fixe distincts.

On peut montrer que la réciproque du théorème précédent est vraie : deux termes ayant le même arbre de Böhm ne sont pas semi-séparables. Deux termes distincts dans le cadre extensionnel sensé ne sont donc pas nécessairement semi-séparables.

Martin Hyland a montré dans [1] que la relation $\{(M, N) \mid M$ et N ne sont pas semi-séparables$\}$ définit une λ-théorie, et qu'il s'agit plus précisément de la théorie extensionnelle sensée maximale.

2 Le cas du λ-calcul probabiliste

2.1 Définition du calcul et différences avec le λ-calcul simple

Le λ-calcul probabiliste est construit en ajoutant à la syntaxe du λ-calcul un opérateur de superposition probabiliste de termes :

$$M, N := x \in \mathit{Var} \mid \lambda x.M \mid MN \mid M +_p N, p \in [0;1]$$

La réduction d'un terme n'est plus un processus déterministe, mais elle s'effectue avec une certaine probabilité. Les règles de réduction sont :

$$(\lambda x.M)N \xrightarrow{1}_\beta M[N/x] \qquad M +_p N \xrightarrow{p}_+ M \qquad M +_p N \xrightarrow{1-p}_+ N$$

Une remarque immédiate à propos de ce système est que si l'on regarde les réductions en oubliant les probabilités, on perd la confluence du λ-calcul. En effet, $M + N \to_+ M$ et $M + N \to_+ N$, et il n'y a a priori aucune raison pour que M et N se réduisent en un même terme.

De plus, lorsque l'on fournit une somme comme argument à une fonction, l'ordre dans lequel on effectue les réductions est important. Le résultat pourra être très différent selon que l'on réduise la somme avant ou après avoir effectué la β-réduction. Si l'on étend la notion de réduction par transitivité en notant $M_0 \xrightarrow{\prod p_i}{}^{*} M_n$ si $M_0 \xrightarrow{p_1} M_1 \xrightarrow{p_2} \ldots \xrightarrow{p_n} M_n$, on a que le terme $(\lambda x.xx)(y +_{1/2} z)$ se réduit soit de manière équiprobable en yy ou zz si l'on réduit la somme en premier, soit de manière équiprobable en yy, yz, zy ou zz si l'on réduit d'abord le β-redex.

Il en résulte que le λ-calcul probabiliste se prête beaucoup moins naturellement à une interprétation globale, à une vision modulo réduction que le λ-calcul classique. Alors que dans le cas déterministe, on pouvait décider d'oublier les détails de la réduction d'un terme pour le considérer comme une simple fonction, un terme probabiliste est un programme qui effectue des choix, et dont on doit par conséquent détailler l'exécution.

Pour cette raison, on n'étend pas la réduction à un contexte quelconque, comme pour le λ-calcul classique, mais l'on choisit de n'effectuer que des réductions de tête :

$$\frac{M \xrightarrow{p} M'}{\lambda x.M \xrightarrow{p} \lambda x.M'}$$

$$\frac{M \xrightarrow{p} M'}{MN \xrightarrow{p} M'N}$$

2.2 Théories probabilistes

Cette partie traite de l'extension de la notion de λ-théorie au cas probabiliste. Son objectif est de donner une idée de ce que pourrait être une théorie probabiliste. Puisqu'il ne s'agit pas là d'un travail abouti, les idées présentées ici ne sont pas réutilisées par la suite.

La notion de séparabilité est liée à celle de théorie puisque dans le λ-calcul, deux termes ne sont pas séparables si et seulement si il existe une théorie extensionelle sensée consistante qui les identifie. Même si ce résultat ne se transposait pas au cas probabiliste, il semble naturel que deux termes séparables ne puissent pas être considérés comme égaux.

Or il existe une version non quantifiée du λ-calcul probabiliste, le λ-calcul avec sommes, qui correspond au cas probabiliste dans lequel on ne tient pas compte des valeurs des probabilités :

$$M, N := x \in \mathit{Var} \mid \lambda x.M \mid MN \mid M + N$$

$$(\lambda x.M)N \to_\beta M[N/x] \qquad M + N \to_+ M \qquad M + N \to_+ N$$

On peut alors définir une fonction d'effacement par :
- $|x| = x$
- $|\lambda x.M| = \lambda x.|M|$
- $|MN| = |M|\ |N|$
- $|M +_1 N| = |M|$
- $|M +_0 N| = |N|$
- $|M +_p N| = |M| + |N|$ si $0 < p < 1$

Intuitivement, on serait tenté de dire qu'il s'agit là d'un modèle du λ-calcul probabiliste, c'est-à-dire que $\{(M, N) \mid |M| = |N|\}$ est une théorie. Dans ce cas, le λ-calcul probabiliste ne serait qu'un raffinement du λ-calcul avec sommes, et deux termes ne seraient séparables que s'ils sont séparables dans le cas avec sommes.

Pourtant, la séparabilité correspond à une différence observationnelle entre termes. Il est vrai que, pour tout $\epsilon > 0$, les termes $x +_\epsilon y$ et $x +_{1-\epsilon} y$ ont en commun de se réduire avec probabilité non nulle uniquement en x ou en y, et par conséquent quels que soient les tests effectués sur ces termes, il est possible de n'observer aucune différence. Cependant la probabilité d'avoir un comportement similaire est très faible, et il semble naturel d'avoir une certaine notion de séparabilité qui les distingue.

Une question qui se pose est alors de savoir s'il existe une notion de théorie probabiliste, qui interdirait l'effacement des probabilités. On peut envisager qu'une telle théorie soit la donnée pour chaque terme M, N de la probabilité de réduction de M en N. Au lieu d'être une relation $\theta \subset \Lambda^2$, on aurait des théories de la forme $\theta : \Lambda_+^2 \to [0; 1]$.

De ce point de vue, il est intéressant de noter qu'un ensemble $\theta \subset \Lambda^2$ peut également être vu comme une fonction $\Lambda^2 \to \{0; 1\}$. Dans ce cas, une λ-théorie est une fonction θ telle que :

1. $\forall M, \theta(M, M) = 1$
2. $\forall M, \forall N, \theta(M, N) = \theta(N, M)$
3. $\forall M, \forall N, \theta(M, N)\theta(N, Q) \leq \theta(M, Q)$
4. $\forall M, \forall N, \forall x, \theta(M, N) \leq \theta(\lambda x.M, \lambda x.N)$
5. $\forall M, \forall N, \forall Q, \theta(M, N) \leq \theta(MQ, NQ)$
6. $\forall M, \forall N, \forall Q, \theta(M, N) \leq \theta(QM, QN)$

7. $\forall M, \forall N, M \to N \Rightarrow 1 = \theta(M, N)$.

Les conditions 1 à 3 correspondent au fait que θ est une relation d'équivalence, les conditions 4 à 6 correspondent au passage au contexte, et la condition 7 est la compatibilité avec la β réduction.

On remarque que si l'on considère $\theta : \Lambda_+^2 \to [0;1]$, les conditions 1, 3, 4 et 5 sont conformes à l'intuition selon laquelle un terme M se réduit en N avec probabilité $\theta(M, N)$. De plus les conditions 6 et 7 se reécrivent en :

6. $\forall M, \forall N, \forall Q, \inf_n(\theta(M, N))^n \leq \theta(QM, QN)$
7. $\forall M, \forall N, M \xrightarrow{p} N \Rightarrow p \leq \theta(M, N)$.

La justification de cette règle 6 est que si le terme Q utilise n fois son argument, alors on peut considérer les termes $Q(M +_p N)$ et QM comme équivalents avec une probabilité p^n. En pratique, cela revient à dire que $\theta(M, N) = 1 \Rightarrow \theta(QM, QN) = 1$, c'est-à-dire que l'on peut réduire un argument à condition que cette réduction soit déterministe. Cela étend la réduction telle qu'elle a été définie pour le λ-calcul probabiliste, mais reste intuitivement correct.

De la même manière, on peut ajouter une condition à la règle 2 en posant :

2. $\forall M, \forall N, \theta(M, N) = 1 \Rightarrow \theta(N, M) = 1$.

Ainsi, on obtient des conditions satisfaisantes et qui garantissent que la restriction de θ aux λ-termes déterministes est une théorie au sens usuel. De plus, la condition 2 exclut la théorie définie par $\theta(M, N) = 1$ si $|M| \to |N|$, 0 sinon.

La question est ensuite de savoir si ces conditions peuvent être considérées comme une axiomatisation des théories probabilistes. En particulier, il faut que la théorie minimale puisse être considérée comme le pendant probabiliste de la β-équivalence. La plus petite fonction θ vérifiant ces conditions semble être donnée par $\theta(M, N) = \max\{p \mid M \xrightarrow{p}^* N\}$, à la symétrie des réductions déterministes près. Si l'on veut considérer non pas la plus forte probabilité de réduction, mais la somme des probabilités des différentes réductions, il faut rajouter une condition de la forme $\theta(M +_p N, Q) \geq p\theta(M, Q) + (1-p)\theta(N, Q)$.

Si l'on considère que, comme dans le λ-calcul usuel, deux termes sont égaux, c'est-à-dire $\theta(M, N) = 1$, si et seulement s'ils se réduisent de la même manière (par exemple $M +_p N = N +_{1-p} M$), on se trouve confronté à un problème. En effet, dans le cas déterministe, cette propriété est assurée par l'axiome de symétrie. Or cet axiome ne s'étend pas au cas probabiliste. On voudrait obtenir un résultat tel que : *Si M et N possèdent des arbres de réductions dont les feuilles sont identiques et de même poids alors $\theta(M, N) = 1$.* Une telle propriété peut être reformulée sans faire appel à la notion de réduction lorsque l'on dispose de la confluence, puisque tout arbre de réduction peut être étendu en un arbre dont toutes les feuilles sont identiques, mais cela ne fonctionne plus dans le cas probabiliste où certaines réductions provoquent une perte d'information.

En résumé, si l'axiomatisation des théories déterministes semble dans un premier temps s'adapter facilement au cas probabiliste, elle ne permet pas de décrire intégralement le comportement des termes probabilistes. Par conséquent, qu'il soit ou non possible de la compléter en une définition satisfaisante de théorie probabiliste, cette transposition ne permet pas de mieux comprendre sous quelles conditions deux termes sont égalisables, et ne nous aide pas à formuler une définition de la séparabilité probabiliste.

2.3 Choix d'un cadre d'étude

Les contraintes du calcul probabiliste, à savoir la perte de la confluence et la restriction aux réductions de tête, forment un obstacle important lors de l'élaboration de preuves utlisant la structure des termes. Mais d'un autre coté, elles sont dues à la nature particulière de la réduction d'une somme, qui consiste à faire un choix entre deux termes en position de tête. Et cette nature même permet de faire des simplifications du système étudié pour le rendre plus facile à manipuler.

On remarque que d'un point de vue calculatoire, il est possible de faire commuter une somme avec une abstraction ou une application à droite. Par définition de la réduction des sommes, un terme attend que l'on fasse un choix probabiliste si et seulement si il est de la forme $C[M +_p N]$ où $C[\,]$ est un contexte de tête, et l'on a $C[M +_p N] \xrightarrow{p}_+ C[M]$, $C[M +_p N] \xrightarrow{1-p}_+ C[N]$. En d'autres termes, $C[M +_p N]$ se comporte exactement comme $C[M] +_p C[N]$.

Lors de la définition des termes, on peut donc ajouter les égalités syntaxiques suivantes :

$$\lambda x.(M +_p N) \equiv (\lambda x.M) +_p (\lambda x.N) \qquad (M +_p N)Q \equiv (MQ) +_p (NQ)$$

On a alors que tout terme présentant une somme en position de tête est égal à une somme, et plus précisément si $M \xrightarrow{p}_+ M_1$ et $M \xrightarrow{1-p}_+ M_2$ alors $M \equiv M_1 +_p M_2$ ou $M \equiv M_2 +_{1-p} M_1$.

Dans ces conditions, il devient possible de repousser les choix à la fin de la réduction, en transformant toute réduction $M \xrightarrow{p}{}^* N$ en $M \to_\beta^* M' \xrightarrow{p}{}^*_+ N$. Il suffit pour cela d'autoriser les β-réduction sous une somme.

$$\frac{M \to_\beta M'}{M +_p N \to_\beta M' +_p N}$$

$$\frac{N \to_\beta N'}{M +_p N \to_\beta M +_p N'}$$

Comme les réductions $\to_+$ sont des projections par rapport au contructeur $+_p$, il y a une correspondance immédiate entre la représentation d'un terme sous forme d'arbre et son arbre de réduction pour $\to_+$. Par conséquent ces réductions n'apportent aucune information supplémentaire sur les termes. On va donc pouvoir ici les négliger et se concentrer uniquement sur la β-réduction.

L'avantage de se restreindre à la β-réduction est que l'on retrouve les bonnes propriétés du λ-calcul usuel. La restriction aux contextes de tête ayant été imposée en raison du non-déterminisme des réductions, elle n'a plus lieu d'être. De plus, on retrouve la confluence du système, qui découle directement de celle du λ-calcul classique. On peut donc à nouveau raisonner modulo β-équivalence des termes. Le comportement de la somme est quant à lui entièrement décrit par les égalités syntaxiques utilisées.

Pour faire totalement abstraction du détail des réductions, on rajoute également des égalités syntaxiques sur le + :

$$M +_0 N \equiv N \quad M +_1 N \equiv M \quad M +_p M \equiv M \quad M +_p N \equiv N +_{1-p} M$$

$$(M +_p N) +_q Q \equiv M +_{pq} (N +_{\frac{q(1-p)}{1-pq}} Q)$$

Un terme écrit comme un arbre de somme peut donc être vu comme une combinaison linéaire $\sum p_i.M_i$ avec $\sum p_i = 1$.

En résumé, on revient à un système semblable au λ-calcul classique. On se contente d'ajouter des constructeurs $+_p, p \in [0;1]$, auxquels on associe les égalités syntaxiques définies précédemment, sans y associer de règle de réduction. On ne considère que les règles de réduction usuelles, étendues à tous les contextes, qui sont des réductions déterministes. On peut alors définir des égalités comme dans le cas classique.

On ne conserve pas de notion de réduction de tête, mais l'on appelle toujours forme normale de tête un terme de la forme $\lambda x_1...x_n.yM_1...M_m$. On continue à parler de solvabilité pour les termes qui ne sont pas des sommes : un terme est solvable s'il est β-équivalent à une forme normale de tête, et il est insolvable s'il n'est ni un terme solvable ni une somme. On conserve la définition de la similarité entre formes normales de tête.

On considère à nouveau les termes selon un point de vue extensionnel sensé. On peut remarquer que, selon la définition précédente, si Ω_1 et Ω_2 sont des termes insolvables, alors $\Omega_1 +_p \Omega_2$ n'est pas insolvable, puisqu'il s'agit d'une somme, mais dans un cadre sensé $\Omega_1 +_p \Omega_2 = \Omega +_p \Omega = \Omega$.

On utilisera principalement la notation $\sum p_i.M_i$ pour désigner une somme de termes, mais l'on utilisera parfois le constructeur $+_p$, généralement lorsqu'une somme comporte exactement deux éléments.

La définition de la séparabilité, quant à elle, ne peut pas être transposée directement. Puisqu'il n'existe pas de règle d'élimination du $+_p$, un terme sous forme de somme ne pourra jamais être envoyé sur un terme sans somme. De plus, puisque l'on n'effectue pas les réductions probabilistes, deux termes qui ne se distinguent que sur les indices des sommes ne pourront pas être séparés autrement que par ces indices.

On utilise alors les termes $\mathbb{T} = \lambda xy.x$ et $\mathbb{F} = \lambda xy.y$. Dans le système usuel deux termes sont séparables si et seulement si l'on peut envoyer l'un sur $\mathbb{T}$ et l'autre sur $\mathbb{F}$. Une définition possible de la séparabilité serait alors de dire que M et N sont séparables s'il existe un contexte que envoie M sur $\mathbb{T} +_p \mathbb{F}$ et N sur $\mathbb{T} +_q \mathbb{F}$ avec $p \neq q$.

3 Les résultats de séparabilité

3.1 Lemmes de manipulation des termes

Pour simplifier les notations, on définit pour tout $a \in [0;1]$, $\underline{a} = \mathbb{T} +_a \mathbb{F}$. On peut remarquer que pour tous $a_1, ...a_n, p_1, ..., p_n$ avec $\sum_{i=1}^n p_i = 1$, $\sum_{i=1}^n p_i.\underline{a_i} = \underline{\sum_{i=1}^n p_i a_i}$. De plus pour $a, b \in [0;1]$, $\underline{a}\,\underline{b}\,\mathbb{F} = \underline{ab}$. De là on obtient que pour toute fonction $f : [0;1]^n \to [0;1]$ construite à partir de ces deux opérations il existe un terme T_f tel que pour tous $a_1, ..., a_n$, $T_f \underline{a_1} ... \underline{a_n} = \underline{f(a_1, ..., a_n)}$.

Lemme 3.1. *Pour tous $a_1, ..., a_n$ non nuls deux à deux distincts, pour tous $p_1, ..., p_n$ non nuls, la fonction $S : x \mapsto \sum_{i=1}^n p_i a_i^x$ s'annule en un nombre fini de points.*

Preuve. On raisonne par induction sur n. Si $n = 1$ alors $S(x) = pa^x$ ne s'annule jamais.

Si $S(x) = p_1 a_1^x + ... + p_{n+1} a_{n+1}^x$, comme a_{n+1}^x n'est jamais nul on a $S(x) = 0$ si et seulement si $\frac{S(x)}{a_{n+1}^x} = 0$. En dérivant cette seconde fonction on obtient $\sum_{i=1}^n p_i \ln(\frac{a_i}{a_{n+1}})(\frac{a_i}{a_{n+1}})^x$, avec les $\frac{a_i}{a_{n+1}}$ non nuls deux à deux distincts et les $p_i \ln(\frac{a_i}{a_{n+1}})$ non nuls. Par hypothèse d'induction la dérivée de $\frac{S(x)}{a_{n+1}^x}$ ne s'annule qu'en un nombre fini de points, d'où le résultat. $\square$

Corollaire 3.2. *Si $\mathcal{M}_1, ..., \mathcal{M}_p$ sont des ensembles finis deux à deux distincts de couples $(p, a) \in]0;1]^2$ tels que $(p, a), (p', a) \in \mathcal{M}_i \Rightarrow p = p'$, alors il existe $n \in \mathbb{N}$ tel que $\forall i \neq j$, $\sum_{p,a \in \mathcal{M}_i} pa^n \neq \sum_{p,a \in \mathcal{M}_j} pa^n$.*

Preuve. Pour tous $i \neq j$ la différence $\sum_{p,a \in \mathcal{M}_i} pa^x - \sum_{p,a \in \mathcal{M}_j} pa^x$ ne s'annule qu'en un nombre fini de points. Par conséquent il existe un $n \in \mathbb{N}$ (et il en existe même une infinité) tel que cette différence ne s'annule pour aucun couple (i, j), $i \neq j$. $\square$

L'intérêt de ce corollaire est de pouvoir combiner des résultats de séparabilité. Typiquement, si l'on considère des termes distincts de la forme $\sum p_i.M_i$, et que l'on sait envoyer les M_i vers des $\underline{a_i}$ distincts, alors il existe un n tel que les $\underline{\sum p_i a_i^n}$ sont distincts.

On qualifie de linéaire une opération qui commute avec la somme. En particulier un contexte est linéaire si $C[M +_p N] = C[M] +_p C[N]$. Or l'application à gauche n'est a priori pas linéaire, il ne suffit pas d'appliquer le terme $T_{x \mapsto x^n}$ à une somme pour l'appliquer à chacune de ses composantes. Pour appliquer la méthode ci-dessus afin de séparer des termes, il faut d'abord trouver un moyen de rendre cette application linéaire. Etant donné un terme M quelconque, il n'existe pas toujours de contexte linéaire $C[\,]$ tel $\forall N, C[N] = MN$, mais il est possible de trouver un résultat proche.

Pour tout $h \in \mathbb{N}$, on rappelle que $R_h = \lambda x_1 ... x_{h+1}.x_{h+1} x_1 ... x_h$. Pour tout ensemble fini de variables Y, pour tous $\delta_0, \delta \in \mathbb{N}$, on note $\Sigma_{Y,\delta_0,\delta}$ l'ensemble des substitutions de la forme $\sigma : M \mapsto M[R_{h_1}/y_1]...[R_{h_n}/y_n]$ qui vérifient :
- $Y \subset \{y_1, ..., y_n\}$;
- $\forall i, h_i \geq \delta_0$;
- $\forall i \neq j, |h_i - h_j| > \delta$.

On remarque qu'étant donnés deux systèmes de contraintes Y, δ_0, δ et Y', δ'_0, δ', il est aisé d'en construire l'intersection : $\Sigma_{Y,\delta_0,\delta} \cap \Sigma_{Y',\delta'_0,\delta'} = \Sigma_{Y \cup Y', \max\{\delta_0,\delta'_0\}, \max\{\delta,\delta'\}}$.

On a alors le résultat suivant :

Lemme 3.3. *Soient des formes normales de tête $M_1, ..., M_p$ avec $M_i = \lambda x_1 ... x_{n_i}.y_i N_{i,1} ... N_{i,m_i}$. On peut leur associer des contraintes Y, δ_0, δ telles que :*
- *pour tout $\sigma \in \Sigma_{Y,\delta_0,\delta}$,*
- *pour toute famille de termes $(Q_{y,s})_{y \in Var, s \in \mathbb{Z}}$,*
- *pour toute famille de variables fraîches $(x_j)_{j \geq 1}$,*

il existe une suite finie de termes $\vec{L}(Q_{y,s}, \sigma)$ tels que pour tout $i \leq p$,

$$\sigma(M_i)\vec{L}(Q_{y,s}, \sigma) = Q_{y_i, m_i - n_i} \sigma(N_{i,1}) ... \sigma(N_{i,m_i}) \sigma(x_{n_i+1}) ... \sigma(x_{m-(m_i-n_i)})$$

où $m = \max m_i$.

Preuve. Avant tout, on remarque que l'on utilise la même notation x_j pour les variables liées de tous les M_i, de même que pour la famille de variables fraîches. Ceci dans un but de simplification des notations : ainsi on peut écrire pour tout i, $M_i x_1 ... x_{m+n} = y_i N_{i,1} ... N_{i,m_i} x_{n_i+1} ... x_{m+n}$, puisque de manière générale $(\lambda x.M)x = M[x/x] = M$.

On choisit alors $Y = \cup\{y_i\}$, $\delta_0 = 2m + n$, $\delta = m + n$, et on se donne $\sigma \in \Sigma_{Y,\delta_0,\delta}$. On a $\sigma(y_i) = R_{h_i}$, on pose $h = \max h_i$. Pour tous termes $Q_1, ...Q_{h+1}$, on a :

$$\sigma(M_i)\sigma(x_1)...\sigma(x_{m+n})Q_1...Q_{h+1} = R_{h_i}\sigma(N_{i,1})...\sigma(N_{i,m_i})\sigma(x_{n_i+1})...\sigma(x_{m+n})Q_1...Q_{h+1}$$
$$= Q_{h_i+1-(m+n+m_i-n_i)}\sigma(N_{i,1})...\sigma(N_{i,m_i})\sigma(x_{n_i+1})...\sigma(x_{m+n})Q_1...$$

Mais l'on peut remarquer que la fonction $y_i, m_i - n_i \mapsto h_i + 1 - (m + n + m_i - n_i)$ est injective, puisque si $i \neq j$, alors $|h_i - h - j| > m + n$. Il suffit donc de poser $Q_{h_i+1-(m+n+m_i-n_i)} = \lambda z_1 ... z_{h+m+n+m_i-n_i}.Q_{y_i,m_i-n_i} z_1 ... z_m$. □

Avec ce lemme, il est possible de parcourir la structure d'un terme. Dans le λ-calcul classique, ce résultat suffit à obtenir un théorème de séparabilité, puisque l'on peut dire que deux termes M_1 et M_2 en forme normale de tête sont séparables s'ils ne sont pas similaires ou s'il existe j tel que $N_{1,j}$ et $N_{2,j}$ sont séparables. Les implications sont moins immédiates lorsque l'on travaille avec des sommes de termes.

3.2 Séparabilité des formes normales

Dans un premier temps, nous allons traiter le cas des termes normalisables, c'est-à-dire, puisque l'on considère les termes modulo réduction, des termes en forme normale. La structure des formes normales est donnée par induction par :

$$M, N := \sum_i p_i.(\lambda x_1 ... x_{n_i}.y_i N_{i,1} ... N_{i,m_i})$$

Puisqu'une forme normale est une somme finie de formes normales de tête, dont les sous-termes sont des formes normales, et que l'on sait agir sur n'importe quel ensemble fini de formes normales de tête, on dispose d'une très grande liberté lorsque l'on veut manipuler des formes normales. Pour cette raison, on dispose d'un résultat de séparabilité très fort :

Théorème 3.4. *Soient $M^1, ..., M^p$ des formes normales deux à deux distinctes. Il existe Y, δ_0, δ tels que pour tout $\sigma \in \Sigma_{Y,\delta_0,\delta}$, on puisse trouver une famille de termes $Q_{y,s}$ qui vérifie $\forall i, \sigma(M^i)\vec{L}(Q_{y,s}, \sigma) = \underline{a_i}$ avec $\forall i \neq j, a_i \neq a_j$*

Preuve. On raisonne par induction sur la somme des tailles des termes $M^1, ..., M^p$. On note $M^i = \sum_j p^i_j.(\lambda x_1...x_{n^i_j}.y^i_j N^i_{j,1}...N^i_{j,m^i_j})$, et on pose $m = \max m^i_j$, $n = \max n^i_j$.

On se donne une famille de variables fraîches (x_i), et on pose $\tilde{N}^i_{j,k} = N^i_{j,k}$ si $k \leq m^i_j$ et $\tilde{N}^i_{j,k} = x_{k-(m^i_j - n^i_j)}$ sinon. On observe alors que deux des M^i_j sont égaux si et seulement si et seulement si les termes $y^i_j \tilde{N}^i j, 1...\tilde{N}^i_{j,m}$ sont égaux. C'est ici qu'intervient la nécessité de se placer dans un cadre extensionnel, autrement cette équivalence n'est pas valide.

Le système de contraintes Y, δ_0, δ cherché correspond à l'intersection d'un certain nombre d'autres systèmes de contraintes. En effet, on veut pouvoir effectuer deux choses : d'une part, séparer les M^i_j, et pour cela on doit séparer les $\tilde{N}^i_{j,k}$, et d'autre part, agir linéairement sur les M^i_j pour pouvoir séparer les termes M^i. On applique donc le lemme 3.3 aux M^i_j, et pour chaque $k \leq m$ on applique l'hypothèse d'induction aux $\tilde{N}^i_{j,k}$. On remarque que $\tilde{N}^i_{j,k}$ est soit un sous-terme strict de M^i_j, soit une variable. Par conséquent $\tilde{N}^i_{j,k}$ est strictement plus petit que M^i_j sauf si M^i_j est une variable, or si tous les M^i_j sont des variables, $m = 0$ et l'on n'utilise pas l'hypothèse d'induction.

Soit $\sigma \in \Sigma_{Y,\delta_0,\delta}$, il existe des termes $\vec{L}_k$ tels que $\sigma(\tilde{N}^i_{j,k})\vec{L}_k = \underline{a_{i,j,k}}$, les $a_{i,j,k}$ étant distincts pour des $\tilde{N}^i_{j,k}$ distincts. On peut raffiner cette séparation en disant que les $\frac{k-1+a_{i,j,k}}{m}$ sont distincts pour des $\tilde{N}^i_{j,k}$ distincts ou des k distincts.

On a alors que des termes de la forme $y^i_j \tilde{N}^i j, 1...\tilde{N}^i_{j,m}$ sont distincts si et seulement si les y^i_j sont distincts ou les ensembles $\{\frac{k-1+a_{i,j,k}}{m} \mid k \leq m\}$ sont distincts, et qu'il n'y a pas de redondance dans la définition de ces ensembles. On utilise le corollaire 3.3 pour trouver un entier u qui sépare les sommes $\sum_{k \leq m}(\frac{k-1+a_{i,j,k}}{m})^u$.

Pour totalement séparer les M_j^i, on indexe les variables y_j^i : $\{y_j^i\}_{i,j} = \{y_1, ..., y_q\}$. On note α_j^i l'indice tel que $y_j^i = y_{\alpha_j^i}$, on obtient que des termes M_j^i sont distincts si et seulement si les $\frac{\alpha_j^i - 1}{q} \frac{1}{m} \sum_{k \leq m} (\frac{k-1+a_{i,j,k}}{m})^u$ sont distincts.

Il faut ensuite utiliser la séparation des M_j^i pour séparer les M^i. Pour cela on utilise à nouveau le corollaire 3.3, pour trouver un entier v qui permette de distinguer les $\sum_j p_j [\frac{\alpha_j^i - 1}{q} \frac{1}{m} \sum_{k \leq m} (\frac{k-1+a_{i,j,k}}{m})^u]^v$.

On pose $f_\alpha : a_1, ...a_m \mapsto [\frac{\alpha-1}{q} \frac{1}{m} \sum_{k \leq m} (\frac{k-1+a_k}{m})^u]^v$, ce fonctions sont représentées par des termes T_{f_α}. La famille $(Q_{y,s})$ recherchée est alors donnée par $Q_{y_\alpha, s} = \lambda z_1 ... z_m . T_{f_\alpha}(z_1 \vec{L}_1)...(z_m \vec{L}_m)$, ce qui donne :

$$\begin{aligned}
\sigma(M^i)\vec{L}(Q_{y,s}, \sigma) &= \sum_j p_j . \sigma(M_j^i)\vec{L}(Q_{y,s}, \sigma) \\
&= \sum_j p_j . Q_{y_j^i, m_j^i - n_j^i} \sigma(\tilde{N}_{j,1}^i)...\sigma(\tilde{N}_{j,m}^i) \\
&= \sum_j p_j . T_{f_{\alpha_j^i}} (\sigma(\tilde{N}_{j,1}^i)\vec{L}_1)...(\sigma(\tilde{N}_{j,m}^i)\vec{L}_m) \\
&= \sum_j p_j . T_{f_{\alpha_j^i}} \underline{a_{i,j,1}} ... \underline{a_{i,j,m}} \\
&= \underline{\sum_j p_j f_{\alpha_j^i}(a_{i,j,1}, ..., a_{i,j,m})}
\end{aligned}$$

□

Corollaire 3.5. *Si M_1 et M_2 sont des formes normales distinctes alors pour tout $\epsilon > 0$ il existe un contexte $C[\,]$ tel que $C[M_i] = \underline{a_i}$ avec $a_1 < \epsilon$ et $a_2 > 1 - \epsilon$.*

Preuve. Il existe une substitution σ et des termes $\vec{L}$ tels que $\sigma(M_i)\vec{L} = \underline{a_i}$ avec $a_1 \neq a_2$. On peut supposer $a_1 < a_2$, puisque pour tout a, $\underline{a}\mathbb{F}\mathbb{T} = \underline{1-a}$.

Par définition de la β-réduction, toute substitution peut être effectuée par un contexte, il existe donc $C_1[\,]$ tel que $C_1[M_i] = \underline{a_i}$.

On pose $S_{a_0} = \lambda x.x(x +_{a_0} \mathbb{T})(xx\mathbb{F})$, on peut remarquer que pour tout $a_0, a \in [0;1]$, $S_{a_0}\underline{a} = \underline{a - a(1-a)(a_0 - a)} = \underline{P(a)}$. Or si $0 < a < a_0$ alors $P(a) < a$, d'où $\lim P^n(a) = 0$, et si $a_0 < a < 1$ alors $a < P(a)$, d'où $\lim P^n(a) = 1$. Le contexte recherché est donc de la forme $C[\,] = S_{a_0}...S_{a_0}C_1[\,]$ avec $a_1 < a_0 < a_2$. $\square$

3.3 Formes « presque sûrement normales »

De manière générale, un terme est normalisable s'il peut se réduire en forme normale en un nombre fini d'étapes. Ici, puisque l'on travaille avec des probabilités, il existe également des termes qui se réduisent en forme normale en un temps fini avec probabilité 1, mais ne sont pas normalisables pour autant. Par exemple si Θ désigne un combinateur de point fixe, $\Theta(\lambda y.y +_{\frac{1}{2}} x) = \Theta(\lambda y.y +_{\frac{1}{2}} x) +_{2^{-n}} x$ pour tout $n \in \mathbb{N}$, mais $\Theta(\lambda y.y +_{\frac{1}{2}} x) \neq x$.

Les termes comme $\Theta(\lambda y.y +_{\frac{1}{2}} x)$, que l'on peut considérer comme « égaux avec probabilité 1 », ou « presque sûrement égaux » à une forme normale peuvent être traité comme des formes normales. En revanche, il existe des termes qui peuvent être considérés comme des sommes infinies de formes normales. Par exemple si l'on considère les entiers de Church $\underline{n} = \lambda fx.f^n x$ et la fonction successeur $\underline{succ} = \lambda nfx.nf(fx)$, on a pour tout n :

$$N = \Theta(\lambda n.\underline{succ}\, n +_{\frac{1}{2}} \underline{0}) = \sum_{i=0}^{n} 2^{-(i+1)}.f^i x + 2^{-(n+1)}.\Theta(\lambda n.\underline{succ}\, n +_{\frac{1}{2}} \underline{0})$$

N est la superposition de tous les entiers. Pour autant, on ne peux pas écrire $N = \sum 2^{-(n+1)}\underline{n}$, puisque notre calcul ne comporte pas de terme infini.

On parlera donc de terme *presque sûrement normal* pour désigner un terme s tel que pour tout $\epsilon > 0$, il existe un terme s', une forme normale M et $p > 1 - \epsilon$ tels que $s = M +_p s'$. Puisque $M = M +_p M$ pour tout p, cela revient à dire que pour tout $\epsilon > 0$ il existe M et s' tels que $s = M +_{1-\epsilon} s'$.

Cette manière de voir les termes, en regardant la limite de leur réduction, permet de définir une nouvelle égalité. Deux termes s et t ont le même comportement à l'infini si pour tout $\epsilon > 0$ on a $s = u +_{1-\epsilon} s'$ et $t = u +_{1-\epsilon} t'$ ou, de manière équivalente, si pour tout $\epsilon > 0$ on a $s +_{1-\epsilon} t' = t +_{1-\epsilon} s'$. On note

$s \approx_\epsilon t$ s'il existe s' et t' tels que $s +_{1-\epsilon} t' = t +_{1-\epsilon} s'$, et $s \approx t$ si pour tout $\epsilon > 0$ on a $s \approx_\epsilon t$.

Théorème 3.6. *Soient s_1 et s_2 deux termes presque sûrement normaux tels que $s_1 \not\approx s_2$, alors pour tout $\epsilon > 0$ suffisamment petit il existe un contexte $C[\,]$ tel que $C[s_i] = p_i.\mathbb{T} + q_i.\mathbb{F} + \epsilon.s'_i$ avec $|p_1 - p_2| > \epsilon$.*

Preuve. s_1 et s_2 sont presque sûrement normaux, donc pour tout $\epsilon > 0$ il existe des formes normales M_1 et M_2 et des termes s'_1 et s'_2 tels que $s_i = M_i +_{1-\epsilon} s'_i$. Comme M_1 et M_2 sont des formes normales, il est possible de les séparer s'ils sont différents. Cependant, le résultat du théorème 3.4 est trop faible pour servir ici : il ne donne aucune borne inférieure sur la différence entre les $\underline{a_i}$ obtenus, or l'on veut que cette différence soit supérieure à $\frac{\epsilon}{1-\epsilon}$, pour avoir les $\underline{a_i} +_{1-\epsilon} s'_i$ de la forme voulue.

Il est cependant possible d'obtenir un résultat plus précis. On peut écrire $M_i = \sum_j p_{i,j}.N_j$, quitte à avoir des p^i_j nuls. Si l'on regarde la preuve du théorème 3.4, on voit qu'il est possible d'envoyer les N_j sur des termes $\underline{b_j}$ distincts. La dernière étape de la preuve consiste à dire qu'il existe v tel que $\sum_j p^1_j b^v_j \neq \sum_j p^2_j b^v_j$. Mais ici, on ne veut séparer que deux termes, on peut donc appliquer une autre méthode.

Pour tout $b \in [0;1]$ avec $\forall j, b \neq b_j$, pour tout $\epsilon' > 0$, on sait trouver un terme S tel que $S\underline{b_j} = \underline{b'_j}$ avec $b'_j < \epsilon'$ si $b_j < b$ et $b'_j > 1 - \epsilon'$ si $b_j > b$. On peut alors trouver un contexte linéaire qui envoie M_i sur un terme aussi proche que l'on veut de $\underline{\sum_{b<b_j} p^i_j}$. Pour tout $\epsilon' > 0$, en choisissant correctement b on peut envoyer les M_i sur des termes $\underline{a_i}$ avec $|a_1 - a_2| \geq \frac{1}{2}\max_j |p^1_j - p^2_j| - \epsilon'$.

Pour séparer les s_i, il suffit alors de montrer que pour ϵ suffisamment petit, on peut choisir les termes M_i tels que $\frac{1}{2}\max_j |p^1_j - p^2_j| > \frac{\epsilon}{1-\epsilon}$. Pour cela, on regarde la « forme infinie » des s_i. De manière informelle, on a $s_i = \sum_j q^i_j.N_j$, on note $\delta = \max_j |q^1_j - q^2_j|$. Pour $\epsilon < \frac{1}{3}\delta$, si $M_i = \sum_j p^i_j.N_j$ est une forme normale telle que $s_i = M_i +_{1-\epsilon} s'_i$, on a pour tout j, $q^i_j - \epsilon \leq (1-\epsilon)p^i_j \leq q^i_j$ donc $\max_j |p^1_j - p^2_j| \geq \frac{\delta-\epsilon}{1-\epsilon} > \frac{2}{3}\frac{\delta}{1-\epsilon} > 2\frac{\epsilon}{1-\epsilon}$. $\square$

L'énoncé du théorème précédent est bien plus faible que ce que l'on pourrait souhaiter. Les définitions de normalisation et d'égalité presque sûres telles qu'elles ont été données ne s'étendent pas au contexte. Il est pourtant naturel de considérer que si un terme s converge presque sûrement alors xs converge presque sûrement. Le problème est que dans le cas des formes normales, la transmission des résultats de séparabilité au contexte est possible grâce au fait que les termes de la forme $\underline{a}$ sont faciles à manipuler, et en particulier se prêtent à l'application du corollaire 3.2. Mais ici, l'existence d'un résidu s' nous empêche d'utiliser cette méthode.

Conclusion

La différence entre termes est beaucoup moins nette dans le λ-calcul probabiliste que dans le calcul classique. Deux sommes peuvent être différentes tout en ayant des termes en commun, deux termes peuvent ne se différencier que par des valeurs de probabilité, sans que leurs structures soient fondamentalement différentes. Pour cette raison, la notion de séparabilité obtenue est moins satisfaisante, puisqu'une fois séparés, des termes ont qualitativement le même comportement, et la distinction ne porte que sur les probabilités de réduction.

Cette faiblesse de la séparabilité est cependant parfaitement justifiée, puisque l'on veut séparer un maximum de termes, et donc des termes très proches. Pour obtenir des résultats plus satisfaisants, il est possible de combiner ceux présentés ici avec des théorèmes de séparation du λ-calcul avec somme (voir par exemple [3]), pour pouvoir donner une séparation structurelle pour des termes structurellement différents, et une séparation probabiliste pour des termes qui ne diffèrent que par des probabilités.

Le véritable problème est qu'il est possible d'avoir des constructions infinies dans le λ-calcul probabiliste. Or que ce soit dans le λ-calcul classique ou probabiliste, on ne sait manipuler qu'un nombre fini de termes à la fois. Pour cette raison, dès que l'on considère des termes autres que de formes normales, il devient difficile d'obtenir des résultats.

Références

[1] M. Hyland A syntactic characterisation of the Equality in some Models of the Lambda Calculus *Journal of the London Mathematical Society*, 1976.

[2] H.P. Barendregt The Lambda Calculus, its syntax and semantics *Studies in Logic and the Foudations of Mathematics*, Vol. 103

[3] U. de'Liguoro, A. Piperno. Nondeterministic Extensions of Untyped Lambda Calculus *Information and Computation*,1995

[4] T. Ehrhard, M. Pagani, C. Tasson. The Computational Meaning of Probabilistic Coherence Spaces, February 7, 2009

[5] V. Danos, T. Ehrhard. On Probabilistic Coherence Spaces, May 18, 2008

Lemmes

Pour tous termes $M_1, ..., M_p$ de la forme $M_i = \lambda x_1 ... x_{n_i}.y_i N_{i,1} ... N_{i,m_i}$, on peut par extensionnalité supposer que pour $i \neq j$, si $y_i = y_j$ et $n_i - m_i = n_j - m_j$ alors $n_i = n_j$ et $m_i = m_j$. D'après le lemme 3.3 il existe Y,δ_0 et δ tels que pour tout $\sigma \in \Sigma_{Y,\delta_0,\delta}$ et pour toute famille de termes Q_s indexée par les classes de similitudes, il existe des termes $\vec{L}$ tels que $\sigma(M_i)\ \vec{L} = Q_s \sigma(N_{i,1}) ... \sigma(N_{i,m_i})$.

Toute fonction $f : [0;1]^m \to [0;1]$ obtenue à partir des fonctions constantes de $[0;1]$ et de $x, y, z \mapsto xy + (1-x)z$ peut être représentée par un terme F. Cette représentation est continue, c'est-à-dire que pour tous $a_1, ...a_m$, pour tout $\epsilon > 0$, il existe $\delta > 0$ tel que pour tous $r_1, ...r_m$ il existe r tel que $F(\underline{a_1} +_{1-\delta} r_1)...(\underline{a_m} +_{1-\delta} r_m) = \underline{f(a_1, ...a_m)} +_{1-\epsilon} r$.

Pour tous $a_1, ..., a_p \in [0;1]$, pour tout $i \leq p$ et pour tout $\epsilon > 0$ il existe une fonction représentable f telle que $f(a_i) > 1 - \epsilon$ et pour tout $j \neq i$, $f(a_j) < \epsilon$. On construit d'abord $f_1 : x \mapsto 1 - (x * ux + (1-x)v)$ en choisissant $u, v \in [0;1]$ tels que $v/2u = a_i$, qui nous donne $\forall j \neq j, f_1(a_j) < f_1(a_i)$. De là, on sait construire f_2 telle que $f_2(f_1(a_i)) > 1 - \epsilon$ et $f_2(f_1(a_j)) < \epsilon$ pour $j \neq i$.

Arbres de Böhm

Dans le λ-calcul déterministe, un arbre de Böhm est un arbre dont les nœuds sont étiquetés par des labels de la forme $(\lambda x_1 ... x_n.y, m)$, puisque tout terme est soit insolvable, ce qui correspond à l'arbre vide, soit une forme normale de tête $\lambda x_1 ... x_n.y N_1 ... N_m$. Dans le cas probabiliste, un terme est une somme de telles formes normales de tête. On définit donc par induction mutuelle :

- $PHNBT_0 = \Omega$
- $PHNBT_{n+1} = \{(\lambda x_1 ... x_n.y, m, T_1, ..., T_m \mid T_i \in PBT_n\} \cup \{\Omega\}$
- PBT_n est l'ensemble des sous-distributions de probabilité sur $PHNBT_n$.

Pour simplifier les notations, on assimile les éléments $t \in PHNBT_n$ et les fonctions de PBT_n valant 1 sur t et 0 ailleurs.

Une autre manière de définir les arbres de Böhm probabilistes est de rajouter un nœud étiqueté $+$, pouvant avoir une quantité dénombrable de fils, avec des arêtes pondérées dont la somme est 1. Il faudrait alors considérer les arbres modulo certaines transformations : un nœud $+$ avec un seul fils est semblable à une arête, deux sous-arbres identiques fils d'un même nœud + peuvent être rassemblés, etc.

On définit ensuite les approximants d'un terme M à profondeur n.
- $hn\omega_0(M) = \Omega$
- $hn\omega_{n+1}(\lambda x_1...x_n.yN_1...N_m) = (\lambda x_1...x_n.y, m, \omega_n(N_1), ..., \omega_n(N_m))$
- Si les réductions de tête maximales finies de M sont les $M \xrightarrow{p_i} M_i$, $\omega_n(M) = \sum p_i.hn\omega_n(M_i)$.

On considère les arbres de Böhm modulo η :
- $(\lambda x_1...x_n.y, m, \Omega, ..., \Omega) = (\lambda x_1, ..., x_{n+1}.y, m+1, \Omega, ..., \Omega)$ si $y \neq x_{n+1}$
- $(\lambda x_1...x_n.y, m, T_1, ..., T_m) = (\lambda x_1, ..., x_{n+1}.y, m{+}1, T_1, ..., T_m, 1.(x_{n+1}, 0))$ si $y \neq x_{n+1}$.

On dit que deux termes M et M' ont les mêmes arbres de Böhm et on note $M \approx M'$ si $\forall n, \omega_n(M) = \omega_n(M')$.

Lemme 3.7. *Si $\omega_{n+1}(M) = \omega_{n+1}(M')$ alors $\omega_n(M) = \omega_n(M')$.*

Preuve. Par induction sur n. Si $n = 0$ le résultat est immédiat.

Si $\omega_{n+2}(M) = \omega_{n+2}(M') = \sum p_i.t_i$ où les t_i sont distincts, on a $\omega_{n+2}(M) = \sum q_{i,j}.hn\omega_{n+2}(M_{i,j})$ et $\omega_{n+2}(M') = \sum q'_{i,j}.hn\omega_{n+2}(M_{i,j})$, où $\sum_j q_{i,j} = \sum_j q'_{i,j} = p_i$ et $hn\omega_{n+2}(M_{i,j}) = t_i$.

Pour $j \neq j'$, $hn\omega_{n+2}(M_{i,j}) = hn\omega_{n+2}(M_{i,j'})$, on écrit $M_{i,j} = \lambda x_1...x_n.yN_1...N_m$ et $M_{i,j'} = \lambda x_1...x_{n'}.yN'_1...N'_{m'}$, comme les termes et les arbres de Böhm sont tous deux définis modulo η on peut supposer $n = n'$, $m = m'$. Alors nécessairement $\omega_{n+1}(N_k) = \omega_{n+1}(N'_k)$, et par hypothèse d'induction $\omega_n(N_k) = \omega_n(N'_k)$. De là on déduit $hn\omega n + 1(M_{i,j}) = hn\omega n + 1(M_{i,j'})$.

Par conséquent $\omega_{n+1}(M) = \sum q_{i,j}.hn\omega_{n+1}(M_{i,j}) = \sum q'_{i,j}.hn\omega_{n+1}(M_{i,j}) = \omega_{n+1}(M')$. □

Lemme 3.8. *Si M est un terme avec $\omega_{n+1}(M) = \sum p_i.(\lambda x_1...x_{n_i}.y_i, m_i, T_{i,1}..., T_{i,m_i})$, alors pour tout $\epsilon > 0$, pour tout I fini, il existe d et des termes r et $N_{i,j,k}$ pour $i \in I$ et $k \leq m_i + d$ tels que :*

- $M = \sum_{i\in I,j} p_{i,j}.\lambda x_1...x_{n_i+d}.y_i N_{i,j,1}...N_{i,j,m_i+d} + (1 - \sum_{i\in I,j} p_{i,j}).r$
- $\omega_n(N_{i,j,k}) = T_{i,j}$ *si* $k \leq m_i$, *et* $\omega_n(N_{i,j,m_i+k}) = 1.(x_{n_i+k}, 0)$
- $\sum_j p_{i,j} = (1-\epsilon).p_i$.

Definition 3.1. Soit T un arbre de profondeur n, on définit $M \leq_n T$ par :
- $M \leq_0$ si $M = \Omega$
- $M \leq_{n+1} T$ si $T = \sum p_i.\lambda\vec{x}_i.y_i T_{i,1}...T_{i,m_i}$ et $M = \sum_{i\in I} p_i.\lambda\vec{x}_i.y_i N_{i,1}...N_{i,m_i} + (1 - \sum_{i\in I} p_i).\Omega$ avec $N_{i,j} \leq_n T_{i,j}$.

On dit qu'un terme est totalement défini si les sous-distributions de probabilité qui interviennent dans son arbre de Böhm sont des distributions de probabilité. De manière équivalente, un terme n'est pas totalement défini si sa réduction de tête termine avec une probabilité < 1, ou s'il se réduit en une forme normale de tête qui n'est pas totalement définie.

Théorème 3.9. *Si* $M_1, ..., M_p$ *sont des termes totalement définis, alors pour tout* n *il existe* $a_1, ...a_p \in [0;1]$ *tels que* $a_i = a_j$ *si et seulement si* $\omega_n(M_i) = \omega_n(M_j)$, *et pour tout* $\epsilon > 0$ *il existe un contexte linéaire* $C[]$ *et des termes* $r_1, ...r_p$ *tels que pour tout* i, $C[M_i] = \underline{a_i} +_{1-\epsilon} r_i$.

Preuve. On raisonne par induction sur n, en montrant comme dans le cas des formes normales qu'il existe des contraintes telles que pour toutes substitution σ les respectant, il existe des termes $\vec{L}$ tels que $\sigma(M_i)\vec{L}$ donne le résultat attendu.

Si $n = 0$, le résultat est immédiat.

Au rang $n+1$, on a $\omega_{n+1}(M_i) = \sum_j p_{i,j} t_j$, avec par hypothèse $\forall i, \sum_j p_{i,j} = 1$. Si $\omega_{n+1}(M_i) \neq \omega_{n+1}(M_{i'})$ il existe un j tel que $p_{i,j} \neq p_{i',j}$. On a donc des $j_1, ..., j_l$ tels que $\omega_{n+1}(M_i) \neq \omega_{n+1}(M_{i'})$ si et seulement si $p_{i,j_k} \neq p_{i',j_k}$. On définit alors $a_i^1 = 1/2 * p_{i,j_1}$ et $a_i^{k+1} = 1/2 * \min_{u_j^k \neq u_{j'}^k} |u_j^k - u_{j'}^k| * p_{i,j_{k+1}}$, puis $a_i = \sum_{k\leq l} a_i^k$. Les a_i on bien la propriété souhaitée.

Pour prouver le théorème, il ne reste plus qu'a montrer que pour tout $k \leq l$, pour tout $\epsilon > 0$, il existe un contexte C_k tel que $C_k[M_i] = \underline{p_{i,j_k}} +_{1-\epsilon} r_i$. On montre ce résultat plus généralement pour tout j.

Pour tout $\epsilon > 0$ il existe un ensemble fini J tel que $\forall \omega_{n+1}(M_i) \neq \omega_{n+1}(M_{i'}), \sum_{j \in J} p_{i,j}.t_{i,j} \neq \sum_{j \in J} p_{i',j}.t_{i',j}$ et $\forall i, \sum_{j \in J} p_{i,j} > 1 - \epsilon$.

Il existe des termes $N_{i,j} = \lambda x_1 ... x_{n_{i,j}}.y_{i,j} N_{i,j,1} ... N_{i,j,m_{i,j}}$ pour $j \in J$ et des r_i tels que $M_i = (1-\epsilon) \sum_{j \in J} p_{i,j}.N_{i,j} + (1 - (1-\epsilon) \sum_{j \in J} p_{i,j}).r_i$ avec $hn\omega_{n+1}(N_{i,j}) = t_j$.

On peut supposer par extensionnalité que $m_{i,j}$ est constant, on notera simplement m. Alors pour tout $k \leq m$, on applique l'hypothèse d'induction aux $N_{i,j,k}$, pour obtenir des $b_{i,j,k}$. Comme $hn\omega_{n+1}(N_{i,j}) = t_j$, les $b_{i,j,k}$ ne dépendent pas de i, on les note simplement $b_{j,k}$.

On fixe un indice j_0. On a vu qu'il existe des fonctions représentables f_k telles que pour tout $\epsilon' > 0$, $f_k(b_{j_0,k}) > 1 - \epsilon'$ et pour tout j tel que $b_{j,k} \neq b_{j_0,k}$, $f_k(b_{j,k}) < \epsilon'$.

On construit alors de manière usuelle et par continuité des fonctions utilisées un contexte linéaire $C_{j_0}[]$ tel que $C_{j_0}[N_{i,j}] = \underline{0}$ si $N_{i,j}$ n'est pas dans la classe des N_{i,j_0}, et $C_{j_0}[N_{i,j}] = \underline{\prod_k f_k(b_{j,k})} +_{1-\epsilon'} r_{i,j}$ sinon. On note $b_j = \prod_k f_k(b_{j,k})$, on a $b_{j_0} > (1-\epsilon')^m$ et $b_j < \epsilon'$ si $j \neq j_0$.

On a alors

$$
\begin{aligned}
C_{j_0}[M_i] &= (1-\epsilon) \sum_{j \in J} p_{i,j}.C_{j_0}[N_{i,j}] + (1 - (1-\epsilon) \sum_{j \in J} p_{i,j}).C_{j_0}[r_i] \\
&= (1-\epsilon) \sum_{t_j \sim t_{j_0}} p_{i,j}.(\underline{b_j} +_{1-\epsilon'} r_{i,j}) + (1-\epsilon) \sum_{t_j \nsim t_{j_0}} p_{i,j}.\underline{0} + (1 - (1-\epsilon) \sum_{j \in J} p_{i,j}).C_{j_0}[r_i] \\
&= (1-\epsilon)(\sum_{j \in J} p_{i,j}).
\end{aligned}
$$