AIX-MARSEILLE UNIVERSITÉ

UNIVERSITÀ DI BOLOGNA

ED184 – MATHÉMATIQUES ET INFORMATIQUE

DOTTORATO DI RICERCA – XXVII CICLO

INSTITUT DE MATHÉMATIQUES DE MARSEILLE (UMR 7373)

DIPARTIMENTO DI INFORMATICA - SCIENZA E INGEGNERIA

Thèse présentée pour obtenir le grade universitaire de Docteur en Mathématiques

Tesi presentata per ottenere il titolo di Dottore di Ricerca in Informatica

On operational properties of quantitative extensions of λ-calculus

Michele ALBERTI

Soutenue le 05/12/2014 devant le jury :

Simona RONCHI DELLA ROCCA	Università di Torino	Présidente
Paul B. LEVY	University of Birmingham	Rapporteur
Michele PAGANI	Université Paris Diderot	Rapporteur
Laurent REGNIER	Aix-Marseille Université	Directeur de thèse
Simone MARTINI	Università di Bologna	Directeur de thèse
Emmanuel BEFFARA	Aix-Marseille Université	Co-directeur (invité)
Lionel VAUX	Aix-Marseille Université	Co-directeur (invité)
Ugo DAL LAGO	Università di Bologna	Co-directeur (invité)

Abstract

In this thesis we deal with the operational behaviours of two quantitative extensions of pure λ-calculus, namely the *algebraic λ-calculus* and the *probabilistic λ-calculus*.

In the first part, we study the β-reduction theory of the algebraic λ-calculus, a calculus allowing formal finite linear combinations of λ-terms to be expressed. Although the system enjoys the Church-Rosser property, reduction collapses in presence of negative coefficients. We exhibit a solution to the consequent loss of the notion of (unique) normal form, allowing the definition of a partial, but consistent, term equivalence. We then introduce a variant of β-reduction defined on canonical terms only, which we show partially characterises the previously established notion of normal form. In the process, we prove a factorisation theorem.

In the second part, we study bisimulation and context equivalence in a λ-calculus endowed with a probabilistic choice. We show a technique for proving congruence of probabilistic applicative bisimilarity. While the technique follows Howe's method, some of the technicalities are quite different, relying on non-trivial "disentangling" properties for sets of real numbers. Finally we show that, while bisimilarity is in general strictly finer than context equivalence, coincidence between the two relations is achieved on pure λ-terms. The resulting equality is that induced by Lévy-Longo trees, generally accepted as the finest extensional equivalence on pure λ-terms under a lazy regime.

Résumé

Cette thèse porte sur les propriétés opérationnelles de deux extensions quantitatives du λ-calcul pur : le *λ-calcul algébrique* et le *λ-calcul probabiliste*.

Dans la première partie, nous étudions la théorie de la β-réduction dans le λ-calcul algébrique. Ce calcul permet la formation de combinaisons linéaires finies de λ-termes. Bien que le système obtenu jouisse de la propriété de Church-Rosser, la relation de réduction devient triviale en présence de coefficients négatifs, ce qui la rend impropre à définir une notion de forme normale. Nous proposons une solution qui permet la définition d'une relation d'équivalence sur les termes, partielle mais cohérente. Nous introduisons une variante de la β-réduction, restreinte aux termes canoniques, dont nous montrons qu'elle caractérise en partie la notion de forme normale précédemment établie, démontrant au passage un théorème de factorisation.

Dans la seconde partie, nous étudions la bisimulation et l'équivalence contextuelle dans un λ-calcul muni d'un choix probabliste. Nous donnons une technique pour établir que la bisimilarité applicative probabiliste est une congruence. Bien que notre méthode soit adaptée de celle de Howe, certains points techniques sont assez différents, et s'appuient sur des propriétés non triviales de « désintrication » sur les ensembles de nombres réels. Nous démontrons finalement que, bien que la bisimilarité soit en général strictement plus fine que l'équivalence contextuelle, elles coïncident sur les λ-termes purs. L'égalité correspondante est celle induite par les arbres de Lévy-Longo, généralement considérés comme l'équivalence extensionnelle la plus fine pour les λ-termes en évaluation paresseuse.

Sommario

Questa tesi ha come oggetto le proprietà operazionali di due estensioni quantitative del λ-calcolo puro: il *λ-calcolo algebrico* e il *λ-calcolo probabilistico*.

Nella prima parte è studiata la teoria della β-riduzione nel λ-calcolo algebrico. Questo calcolo permette la formazione di combinazioni lineari finite di λ-termini. Sebbene il sistema risulti godere della proprietà di Church-Rosser, la relazione di riduzione diviene banale in presenza di coefficienti negativi, rendendola inadatta a definire una nozione di forma normale. Si propone quindi una soluzione a tale problema, che permette la definizione di una relazione di equivalenza sui termini, parziale ma coerente. In seguito, una differente β-riduzione, ristretta ai soli termini canonici, è introdotta, dimostrando che questa caratterizza in parte la nozione di forma normale precedentemente stabilita. Nel fare ciò, si dimostra un teorema di fattorizzazione.

Nella seconda parte sono studiate la bisimulazione e l'equivalenza contestuale in un λ-calcolo munito di un operatore di scelta probabilistica. Una tecnica per stabilire che la bisimilarità applicativa probabilistica è una congruenza viene esibita. Sebbene tale metodo si ispiri a quello di Howe, alcune tecnicità risultano differenti, contando su non banali proprietà di "dipanamento" sugli insiemi di numeri reali. Si dimostra infine che, seppure la bisimilarità sia in generale strettamente più fine dell'equivalenza contestuale, esse coincidono sui λ-termini puri. L'uguaglianza corrispondente è quella indotta dagli alberi di Lévy-Longo, generalmente considerata come la più fine equivalenza estensionale sui λ-termini puri in un contesto di valutazione lazy.

Acknowledgements

Thanks to Laurent Regnier and Simone Martini for having accepted to co-supervise this Ph.D. thesis.

I am particularly grateful to Emmanuel Beffara, Lionel Vaux and Ugo Dal Lago for their invaluable guidance throughout these three years. This thesis is full of their helpful ideas, insights and much more.

I thank Emmanuel and Lionel for many enjoyable moments (often talking nonsense) and interesting scientific discussions. I will always admire their never-ending intellectual curiosity and mathematical soundness.

Ugo has been my first professor in Bologna and it has been a pleasure to end my studies working with him. I thank Ugo for having supported my French scientific adventures, first in Lyon and later in Marseille.

Thanks to Simona Ronchi Della Rocca, Paul B. Levy and Michele Pagani for serving on my Ph.D. committee. I am grateful to Paul B. Levy and Michele Pagani for having accepted to act as referees of my thesis.

Thanks to all the members of *Logique de la Programmation* team. I have enjoyed being part of its always interesting research environment.

Special thanks to Tomasz Miernowski for his help and friendship during my stays in Marseille. I have enjoyed being his teaching assistant and often his first (dumb) student. I appreciate the way he strives for good teaching.

Thanks to all I2M Ph.D. students, and Postdocs, for having bear with my antisocial personality. I have appreciated the effort. In particular, I thank (women first) Anna, Emilie, Eugenia, Francesca, Clément, Michele, Marc, Matteo, Jean-Baptiste, Joël, Paolo, Thomas, Pierre. I hope you will be successful.

Thanks to Marc and JB for having helped me out without hesitate, whether it was dealing with French bureaucracy or mathematics. In particular, I owe Marc the section about factorisation for the algebraic λ-calculus.

Un abbraccio speciale al gruppo degli italiani col quale ho condiviso (seppure raramente per mie colpe) dei bei momenti: Anna, Eugenia, Matteo, Paolo. Ringrazio Matteo per le indimenticabili discussioni sui più svariati argomenti (in ordine: gentil sesso, matematica, accademia, “autismo”) e per le numerose, grasse risate.

Thanks to Marco Solieri for having shared with me his thoughts and doubts about research and more, always with a positive attitude. I envy his willpower.

Thanks to Beniamino Accattoli for the time spent together, especially in San Diego during POPL'14. I admire him for being a terrific researcher as much as for his positive personality. I thank him and Ugo for their discussions (especially, along the way to a restaurant in San Diego) about mathematics, computer science and their experiences as Ph.D. students. Their thoughts have clarified the meaning of being a good researcher, something (I hope) will guide my next choices.

Grazie a Gigio per essersi prestato alla lettura di alcune parti di questa tesi, per il supporto e la sua amicizia.

Un grazie speciale a mamma e babbo per la pazienza e il supporto con i quali hanno affrontato il frenetico periodo necessario alla stesura di questo documento. Grazie anche a Paolo e Linda, è bello essere vostro fratello.

Un abbraccio a zia Gloria (e Andrea) per i caffè offerti, e perché se lo merita.

Infine, grazie ad Alessia. Perché c'è sempre stata e ci sarà.

« I sound my barbaric YAWP over the rooftops of the world. »
(W. Whitman)

Contents

Introduction **1**

I Normal forms for the algebraic λ-calculus **9**

1 Linear combinations of λ-terms **11**

1.1 Monoids, Semirings and R-modules . . . 13
1.2 Algebraic λ-calculus . . . 14
1.2.1 Algebraic terms . . . 14
1.2.2 Induction on terms and Substitution . . . 16
1.3 The module of algebraic λ-terms . . . 18
1.3.1 Raw and Permutative terms . . . 19
1.3.2 Free R-module construction . . . 20
1.3.3 Canonical forms . . . 21
1.4 Reduction relations on algebraic λ-terms . . . 25
1.4.1 β-reduction . . . 27
1.4.2 Parallel β-reduction . . . 30
1.4.3 Confluence . . . 33

2 Normal forms for the algebraic λ-calculus **39**

2.1 Collapse . . . 41
2.1.1 Term equivalence . . . 42
2.2 Algebraic properties and Normalisability . . . 42
2.3 (Unique) Normal forms . . . 45
2.3.1 Preliminaries . . . 45
2.3.2 Strong normalisability . . . 52
2.3.3 Normalisability . . . 56
2.3.4 Unique normal forms for algebraic terms . . . 62
2.3.5 (Consistent) Term equivalence . . . 62
2.4 Compute normal forms . . . 63

2.4.1 Canonical relations from terms to terms 63
2.4.2 Canonical parallel β-reduction 64
2.4.3 Canonical β-reduction 71
2.4.4 Canonical means defectiveness 77

3 Factorisation 83
3.1 Reduction in function/argument position 85
3.1.1 Confluence of function reduction 88
3.2 Decompose + Swap = Factorisation 89
3.3 Head normalisability 98

II Coinductive equivalences in a probabilistic scenario 101

4 Probabilistic applicative bisimulation 103
4.1 Probabilistic operational semantics for λ-calculus 105
4.1.1 Pure, non-deterministic λ-calculus 105
4.1.2 Value distributions and Call-by-name operational semantics . 106
4.2 Probabilistic bisimulation 109
4.3 Probabilistic applicative bisimulation and Howe's method 115
4.3.1 Probabilistic applicative bisimulation 115
4.3.2 $\Lambda_{\oplus}$-relations 120
4.3.3 Howe's construction 124
4.4 Proof of the Key Lemma 133
4.4.1 Disentangling probability assignments 134
4.5 Relating Applicative bisimilarity and Context equivalence 142
4.5.1 Probabilistic context equivalence 142
4.5.2 "Context-free" context equivalence 147
4.5.3 Probabilistic CIU-equivalence 149

5 The discriminating power of probabilistic contexts 159
5.1 Lévy-Longo trees 161
5.2 Pure λ-terms in probabilistic contexts 164
5.2.1 Context equivalence implies Lévy-Longo equality 164
5.2.2 Lévy-Longo equality implies Applicative bisimilarity 169

Conclusions and Perspectives 179

Bibliography 185

Introduction

Since its inception, the λ-calculus has provided a well-suited environment for studying computation. Used to express effective and deterministic computation in terms of higher-order functions, during the last decades it has been endowed with different kinds of operators to accommodate new computational models such as the non-deterministic and, more recently, quantum and probabilistic ones.

In other words, the λ-calculus and its extensions have been successfully used as *programming languages* by means of which diverse forms of computation can be rigorously described and their properties formally established. For this purpose, mathematical tools for reasoning about the *semantics* of *programs* are needed, namely a precise definition of how programs, expressed as λ-terms, compute. Since the '60s and '70s, the two most common approaches to formalising semantics are the *operational* and *denotational* ones. Both have been undoubtedly successful.

Operational semantics establishes the behaviour of a programming language by defining an abstract machine for it. The meaning of a program P is given by the final state (or some other particular configuration) that the machine reaches when instructed by P.

Denotational semantics abstracts from any notion of machine. Here, the behaviour of a programming language is studied in terms of the properties, and the mathematical laws, exhibit by the mathematical objects of some semantic domain. The meaning of a program P is its related object $[\![P]\!]$ in the semantic domain, established by an interpretation function $[\![\cdot]\!]$ mapping terms of the programming language into elements of this latter.

While operational semantics relies on the dynamics of the machine to establish equality between programs, denotational semantics recasts such investigation into the static notion of equality between objects of the semantic domain. This two approaches typically relate to each other by some notion of *full abstraction*, establishing that programs are operationally equivalent if and only if their interpretations in the semantic domain are the same.

This thesis deals with the operational semantics of *quantitative* λ-calculi, namely extensions of the λ-calculus accounting for quantitative information on computation such as the probability of a successful evaluation of a program. In particular, some operational aspects of the *algebraic* and *probabilistic* extensions of the pure λ-calculus are studied. Nonetheless, denotational semantics originates and justifies this work, since the calculi here investigated exhibit well-established semantic counterparts.

Algebraic λ-calculus

The first part of this thesis derives directly from the efforts in denotational semantics that followed Girard's introduction of coherence spaces and, consequently, of linear logic [Gir87]. Although coherence spaces are a *qualitative* model in the tradition of Scott's continuous semantics, Girard's original investigation was about a *quantitative* model of the λ-calculus [Gir88], where the intuition of the meaning of programs as the limit of approximations was related to the fact that analytic functions are given by power series (in fact, Girard used a categorical notion called normal functors).

Recently, this idea has been revised by Ehrhard [Ehr05] who has developed denotational models of the typed λ-calculus where types are interpreted as vector spaces, or modules, and λ-terms as analytic functions defined by power series on these spaces. It turned out that, in these models, all functions are n^{th} differentiable and the Taylor formula holds. This has led to differential linear logic [ER06b] and, by the famous Curry-Howard correspondence, to the syntactic developments of the differential λ-calculus [ER03] and its related idioms [PT09, PR10]. In particular, the introduction of the differential λ-calculus has inspired a decade of (still ongoing) research in logic and computer science, such as:

- The investigation of syntactical differentiation in λ-calculus and related study of its computational content, leading to the syntactic Taylor expansion of λ-terms as a theory of approximation [ER06a] (usually based on Böhm trees).

- The introduction of the resource λ-calculus [PT09], similar to Boudol's λ-calculus with multiplicities [Bou93], from the distinction between reusable and depletable arguments provided by the correspondence of partial derivation with linear substitution. This has also provided correlations with non-determinism and process calculi [EL10b, EL10a].

- The extension of pure λ-calculus, and its β-reduction theory, to an *algebraic setting* provided by the endowing of the set of λ-terms with a structure of vector space or R-module (with R semiring). Here linear combinations of terms naturally arise and, according to the intuition of algebraic linearity as "commutation with sums", application terms are linear in the function but not in the argument. This corresponds with well-known computational characterisations of linearity in λ-calculus, such as the one provided by non-deterministic operational semantics [dP95] and that given by the decomposition of intuitionistic implication in linear logic [DHR96, DR04].

The *algebraic λ-calculus* [Vau07, Vau09] stems from this latter research investigation.

The purpose is the study of the rewriting properties of β-reduction in a R-module of terms, namely the interaction between β-reduction and the algebraic component of the calculus (*i.e.* coefficients in R). This makes the algebraic λ-calculus a valuable setting for studying *fundamental* operational aspects of quantitative higher-order functional programming (*e.g.* the probabilistic [DLZ12, Par03] and quantum paradigms [vT04, Sel04]) in a *comprehensive* and *unifying* way.

The higher-order reduction theory of the algebraic λ-calculus is far from being trivial. Indeed, since his first work on the subject [Vau07], Vaux has shown that β-reduction *collapses* as soon as the semiring R admits negative elements: whenever $-1 \in \mathsf{R}$, then any term is β-reducible to any other term. This is due to the fact that, in this case, the system exhibits both fixpoints and negative coefficients.

The module of terms with non-negative only coefficients is not less delicate to study. For instance, if one considers linear combinations of terms with coefficients in $\mathbb{Q}^+$, strong normalisability holds only in the case of normal terms. Not surprisingly, a non-problematic case is that of the module of terms with coefficients in $\mathbb{N}$.

Contributions. In this thesis we continue the investigation on the β-reduction theory of the algebraic λ-calculus, with a particular focus on normalisability and the related notion of normal form. Our work finds justifications from the aforementioned collapse of β-reduction and the consequent inconsistency of term equivalence. Most notably, we investigate the lack of the crucial notion of normal form, which can be thought of as the basic evidence of a meaningful computation.

Following an insight proposed by Ehrhard and Regnier in their original work on the differential λ-calculus [ER03], we develop a method to identify the set of normalisable terms, hence of normal ones, in the most general case of linear combinations of terms with (potentially) negative coefficients. The solution is based on the intuition that every term can be expressed as an element of the module of terms over the semiring of polynomials with non-negative integer coefficients, namely recasting the study of the behaviour of β-reduction in a sound setting. The normal form here obtained is then the interpretation of the normal form of the original term. This leads us to the definition of a partial, but consistent at last, term equivalence. We detail such a solution in the case of *strong normalisability* and (just) *normalisability*, separately, in order to show the different rewriting techniques put to use. The development of the former case has been published [Alb13], whereas the latter has not yet.

We then deal with the problem of directly computing such normal forms: we consider a natural variant of reduction which is consistent by definition. We show that its parallel version characterises the notion of normal form previously established, although we fail to fulfil the original intent of characterising normal forms by means of a more atomic reduction. In the process we prove a weak formulation of the well-known *factorisation theorem*, by following the guidelines of Takahashi's technique [Tak95] for the pure λ-calculus, and we show that it is the best we can achieve in the algebraic λ-calculus. These latter results are not published either.

Probabilistic λ-calculus

The second part of this thesis investigates on probability in computation. Indeed, probabilistic models are more and more pervasive. Not only is probability the best theory one can rely on when dealing with uncertainty and incomplete information. It sometimes is a *necessity* rather than an option, like in computational cryptography [GM84].

A nice way to deal computationally with probabilistic models is to allow probabilistic choice as a primitive when designing algorithms, this way switching from usual, deterministic computation to a new paradigm, called probabilistic computation. Typically, this latter is made available in the realm of programming languages by endowing any deterministic language with one or more primitives for probabilistic choice, like binary probabilistic choice or primitives for distributions.

One class of languages that cope well with probabilistic computation are functional languages. As a matter of fact, many existing probabilistic programming languages [Pfe01, Goo13] are designed around the λ-calculus or one of its incarnations. This has stimulated foundational research about probabilistic λ-calculi and program equivalence in a probabilistic setting. In particular, some results of adequacy and full-abstraction have recently arise in the realm of denotational (quantitative) models of linear logic [EPT11, EPT14], and also in game semantics [DH02]. The underlying operational theory, which in the λ-calculus is known to be very rich, has remained so far largely unexplored.

This thesis focuses on operational techniques for understanding and reasoning about program equality in higher-order probabilistic languages. Checking computer programs for equivalence is a crucial, but challenging, problem. Equivalence between two programs generally means that the programs should behave "in the same manner" under any context. Specifically, two λ-terms are *context equivalent* if they have the same convergence behaviour (i.e., they do or do not terminate) in any possible context. Finding effective methods for context equivalence proofs is particularly challenging in higher-order languages.

Bisimulation has emerged as a very powerful operational method for proving equivalence of programs in various kinds of languages, due to the associated coinductive proof method. To be useful, the behavioural relation resulting from bisimulation, *i.e. bisimilarity*, should be a *congruence*, and should also be sound with respect to context equivalence. Bisimulation has been transplanted onto λ-calculus by Abramsky [Abr90], under the name of *applicative bisimulation*. In short, two λ-terms M and N are applicative bisimilar when their applications $(M)\,P$ and $(N)\,P$ are applicative bisimilar for any argument term P.

Often, checking a given notion of bisimulation to be a congruence in higher-order languages is non-trivial. In the case of applicative bisimilarity, congruence proofs usually rely on Howe's method [How96]. Other forms of bisimulation have been proposed, such as environmental bisimulation and logical bisimulation [SKS07, SKS11, KLS11], with the goal of relieving the burden of the proof of congruence, and of accommodating language extensions.

In this work, we consider the pure λ-calculus extended with a probabilistic choice operator. In this setting, context equivalence of two terms means that they have the same *probability of convergence* in all contexts. The objective of the second part of this thesis is to understand context equivalence and bisimulation, and how they relate to each other, in this paradigmatic probabilistic higher-order language.

Contributions. We endow the non-deterministic λ-calculus [dP95] with a probabilistic call-by-name operational semantics [DLZ12]. In this setting, we first adapt Abramsky's idea of applicative bisimulation [Abr90] and we later provide a proof of congruence for probabilistic applicative bisimilarity along the lines of Howe's method. Definitionally, we obtain probabilistic applicative bisimulation by setting up a *labelled Markov chain* on top of λ-terms, then adapting to it the coinductive scheme introduced by Larsen and Skou in a first-order setting [LS91]. In the proof of congruence, the construction of Howe's lifting $(\cdot)^H$ closely reflects analogous constructions for non-deterministic extensions of the λ-calculus. The novelties are in the technical details for proving that the resulting relation is a bisimulation: in particular our proof of the so-called Key Lemma, an essential ingredient in Howe's method, relies on non-trivial "disentangling" properties for sets of real numbers, these properties themselves proved by modelling the problem as a flow network and then apply the Max-flow Min-cut theorem.

The congruence of applicative bisimilarity yields soundness with respect to context equivalence as an easy corollary. Completeness, however, fails: applicative bisimilarity is proved to be finer.

We finally show that the presence of higher-order functions and probabilistic choice in contexts gives context equivalence and applicative bisimilarity maximal discriminating power on pure λ-terms. We do so by proving that, on pure λ-terms, both context equivalence and applicative bisimilarity coincide with the *Lévy-Longo tree equality*, which equates terms with the same Lévy-Longo tree, and is generally accepted as the finest extensional equivalence on pure λ-terms under a *lazy* regime. These results have been published [LSA14].

General related work

Research on *quantitative extensions* of λ-calculus has drawn much attention lately, following the arise of novel computational models such as the quantum and probabilistic ones. In this section, we give some pointers to the relevant literature on quantitative λ-calculi, without any hope of being exhaustive.

The algebraic λ-calculus [Vau07, Vau09] stems from quantitative models of linear logic [Ehr02, Ehr05], which have provided a serious grounding to endow the pure λ-calculus with a structure of vector space or module. The purpose of this calculus is the quest for a comprehensive setting where operational semantics of the aforementioned models of computation can be investigated.

The collapse of β-reduction is documented in the original works by Vaux [Vau07, Vau09], whereas normalisability issues were already present in differential λ-calculus [ER03]. There, however, the authors principal objective was the understanding of differentiation in λ-calculus.

Normalisability is a delicate matter in the setting of the algebraic λ-calculus. Vaux [Vau09] has proposed a simple Curry-style type system and, under some conditions on the semiring R of coefficient, typable terms are proved to enjoy strong normalisability. Then, by slightly changing the notion of normal form, those conditions on R are relaxed and a *weak normalisation scheme*, built on top of the same type system, is proposed. This latter was, however, first hinted by Ehrhard and Regnier [ER03]. In this thesis, we fully develop this idea without appealing to any type system, rather we rely on non-trivial rewriting techniques.

Similar problems arise in the setting of the linear-algebraic λ-calculus developed by Arrighi and Dowek [AD08]. Introduced as a candidate λ-calculus for quantum computation, this calculus exhibits many technical dissimilarities in comparison with the algebraic λ-calculus. Terms represent linear operators, hence application is bilinear rather than linear in function position only. Moreover, while the algebraic λ-calculus investigates a notion of (call-by-name) β-reduction on a module of terms, the linear-algebraic λ-calculus is a (call-by-value) calculus based on rewriting rules. The relationship between the two calculi have been studied [DCPTV10].

Concerning the collapse, Arrighi and Dowek's solution follows the tradition of term rewriting: they allow some rewriting rules to take place only on closed terms in normal form. Other approaches are mainly based on typing [BDCJ12, ADC12, ADCV13].

The factorisation theorem [Tak95, Mel97] is a well-established result in pure λ-calculus, commonly used to prove the standardisation theorem and characterise the leftmost strategy as normalising [Bar84]. Pagani and Tranquilli have proved a factorisation theorem in the setting of the resource λ-calculus [PT09]. Our result, however, is different as distinct are the dynamics of this latter and that proper of the algebraic λ-calculus. In particular, the resource λ-calculus exhibits a bilinear form of application.

Various probabilistic λ-calculi have been proposed, starting from the pioneering work by Saheb-Djahromi [SD78], followed by more advanced studies by Jones and Plotkin [JP89]. Both these works are mainly focused on denotational semantics. More recently, there has been a revamp on this line of work, with the introduction of adequate (and sometimes also fully-abstract) denotational models for probabilistic variations of PCF [DH02, EPT11, EPT14]. There is also another thread of research in which various languages derived from the λ-calculus are given types in monadic style, allowing this way to nicely model concrete problems like Bayesian inference and probability models arising in robotics [RP02, PPT08, GAB$^+$13]; these works however, do not study operationally based theories of program equivalence.

Non-deterministic extensions of the λ-calculus have been analysed in typed calculi [AC84, Sie93, Las98] as well as in untyped calculi [JP90, Bou94, Ong93, dP95]. The emphasis in all these works is mainly domain-theoretic. Apart from [Ong93], all cited authors closely follow the testing theory [DH84], in its modalities *may* or *must*, separately or together. Ong's approach [Ong93] inherits both testing and bisimulation elements.

Our definition of applicative bisimulation follows Larsen and Skou's scheme for fully-probabilistic systems [LS91]. Many other forms of probabilistic bisimulation have been introduced in the literature, but their greater complexity is usually due to the presence of *both* non-deterministic and probabilistic behaviors, or to continuous probability distributions. See surveys such as [BDL13, Pan09, Hen12].

Contextual characterisations of Lévy-Longo tree equality include [BL96], in a λ-calculus with multiplicities in which deadlock is observable, and [DCTU99], in a λ-calculus with choice, parallel composition, and both call-by-name and call-by-value applications. See [DCG01] for a survey on observational characterisations of λ-calculus trees.

As a follow up to our work, Crubillé and Dal Lago [CL14] have considered the call-by-value variant of our setting. They have proved that applicative bisimilarity coincides with context equivalence, and that completeness holds only in the symmetric setting (*i.e.* not in the case of applicative similarity).

Plan of the thesis

The thesis is organised in two parts: Chapters from 1 to 3 corcern the algebraic λ-calculus, whereas Chapters from 4 to 5 concern the probabilistic λ-calculus. In particular:

- Chapter 1 introduces the syntax and basic reduction theory of the algebraic λ-calculus. The construction of the free R-module of terms is recalled.

- Chapter 2 starts by recalling the collapse of β-reduction in presence of negative coefficients. Thereafter, a full development of Ehrhard and Regnier's idea of weak normalisation is presented, first in the case of strong normalisability and then in the case of (just) normalisability. The induced partial term equivalence is proved to be consistent. Finally, notions of canonical reduction relations from terms to terms are investigated.

- Chapter 3 provides the factorisation theorem for the algebraic λ-calculus.

- Chapter 4 begins with the introduction of the syntax and call-by-name operational semantics of the probabilistic λ-calculus. Thereafter, the notion of bisimulation on labelled Markov chains is recalled by following Larsen and Skou's influential work. Probabilistic λ-calculus is then presented as a labelled Markov chain, providing the ground on top of which the probabilistic variant of Abramsky's applicative (bi)simulation is defined. By appealing to Howe's method, probabilistic applicative bisimilarity is proved to be a congruence: the proof turns out to be much more difficult than the one for deterministic and non-deterministic cases, as it relies on the Max-flow Min-cut theorem to "disentangle" sets of real numbers. Finally, applicative bisimilarity is shown to be strictly finer than context equivalence by means of probabilistic CIU-equivalence.

- Chapter 5 shows that probabilistic applicative bisimilarity and probabilistic context equivalence collapse if the tested terms are pure, deterministic, λ-terms. In particular, both relations coincide with the Lévy-Longo tree equality, which equates terms with the same Lévy-Longo tree.

Part I

Normal forms for the algebraic λ-calculus

Chapter 1

Linear combinations of λ-terms

We devote this chapter to the introduction of the syntax and basic reduction theory of the *algebraic λ-calculus*. We follow by and large Vaux's original works [Vau07, Vau09], while using Ehrhard and Regnier's [ER03] way of presenting algebraic λ-terms. All in all, no contributions are presented in this chapter, although we pay attention in providing all the technical details the original works may lack of.

Contents

1.1 Monoids, Semirings and R-modules **13**

1.2 Algebraic λ-calculus **14**

1.2.1 Algebraic terms 14

1.2.2 Induction on terms and Substitution 16

1.3 The module of algebraic λ-terms **18**

1.3.1 Raw and Permutative terms 19

1.3.2 Free R-module construction 20

1.3.3 Canonical forms 21

1.4 Reduction relations on algebraic λ-terms **25**

1.4.1 β-reduction 27

1.4.2 Parallel β-reduction 30

1.4.3 Confluence 33

The first section recalls the definition of few basic algebraic structures.

The second section introduces the free R-module of algebraic λ-terms, with R a semiring. Terms are subject to a notion of algebraic linearity which highlights the differential linear logic origin of the calculus. The basic definitions concerning term substitution are given.

The third section summarises Vaux's concrete construction of the free R-module of algebraic λ-terms. This latter introduces equality on linear combinations of λ-terms in a rigorous manner, making explicit the one we consider.

Finally, the fourth section extends the pure λ-calculus notion of β-reduction to the setting of the algebraic λ-calculus. This notion of β-reduction is proved to enjoy the Church-Rosser property by means of the well-known Tait–Martin-Löf technique. In the process, any kind of one-step reduction is shown to intrisically deny strong confluence due to the algebraic nature of the calculus.

1.1 Monoids, Semirings and R-modules

In this preliminary section we quickly review the definition of some basic algebraic structure R, which basically involves a set equipped with one or more operations defined on it. Throughout this first part, we usually call *coefficients* the elements of some precise algebraic structure R.

Definition 1.1.1. *Given a set* R *and binary operation* $\times$ *defined on it, the algebraic structure* $\mathsf{R} = (\mathsf{R}, \times, i)$ *is a* monoid *whenever the following two axioms hold:*

- Associativity*: for all* $a, b, c \in \mathsf{R}$*,*

$$(a \times b) \times c = a \times (b \times c);$$

- Identity element*:* $i \in \mathsf{R}$ *and for all* $a \in \mathsf{R}$*,*

$$i \times a = a \times i = a.$$

A monoid is said to be commutative *whenever, for all* $a, b \in \mathsf{R}$*,* $a \times b = b \times a$ *holds.*

Definition 1.1.2. *Given a set* R *and two binary operation* $+, \times$ *defined on it, the algebraic structure* $\mathsf{R} = (\mathsf{R}, +, 0, \times, 1)$ *is a* semiring *whenever* $(\mathsf{R}, +, 0)$ *is a commutative monoid and* $(\mathsf{R}, \times, 1)$ *is a monoid, moreover satisfying the following axioms:*

- $\times$ *distributes over* $+$*: for all* $a, b, c \in \mathsf{R}$*,*

$$a \times (b + c) = (a \times b) + (a \times c),$$
$$(a + b) \times c = (a \times c) + (b \times c).$$

- 0 *is absorbing for* $\times$*: for all* $a \in \mathsf{R}$*,*

$$0 \times a = a \times 0 = 0.$$

A semiring is said to be positive *if, for all* $a, b \in \mathsf{R}$*,* $a + b = 0$ *implies* $a = b = 0$*.*

A typical example of positive semiring is the set of non-negative integer $\mathbb{N}$, equipped with the usual operations.

Notation. Given a semiring R, we write $\mathsf{R}^{\bullet}$ for $\mathsf{R}\backslash\{0\}$. Moreover, we indicate as P any positive semirings.

In Section 1.2 we introduce an extension of the pure λ-calculus by endowing the set of terms with a structure of *vector space*, where terms can be added together and multiplied by coefficients. In particular, we are interested in the slight generalisation of R-modules, namely vector spaces over a fixed semiring R.

Definition 1.1.3. *Given a semiring* R*, a* R-module *is an algebraic structure* $(\Phi, +, \mathbf{0}, \cdot)$ *such that* $(\Phi, +, \mathbf{0})$ *is a commutative monoid, equipped with an operation* $\cdot : \mathsf{R} \times \Phi \mapsto \Phi$*, verifying the following equations: for all* $a, b \in \mathsf{R}$ *and* $\alpha, \beta \in \Phi$*,*

$$\begin{gathered} 0\alpha = \mathbf{0}, \quad 1\alpha = \alpha, \quad a(\alpha + \beta) = a\alpha + a\beta, \\ (a + b)\alpha = a\alpha + b\alpha, \quad (a \times b)\alpha = a(b\alpha). \end{gathered} \tag{1.1}$$

Given a set Φ, $\mathsf{R}\langle\Phi\rangle$ *is the free* R*-module generated by* Φ*,* i.e. *the set of all formal finite linear combinations of elements of* Φ *with coefficients in* R*.*

Observe that there is a clear injection of Φ into $\mathsf{R}\langle\Phi\rangle$ by means of the multiplicative unit 1 of R.

Almost everywhere null functions. An element S of $\mathsf{R}\langle\Phi\rangle$ can be thought as a R-valued function defined on Φ which vanishes for almost all values of its arguments α: *i.e.* for some finite $\Xi \subset \Phi$, $S(\alpha) = 0$ if and only if $\alpha \notin \Xi$.

In the sections that follow, we call *algebraic terms* such elements S, and we often write them as $\sum_{\alpha\in\Phi} a_\alpha \alpha$, knowing that the latter is actually the finite sum $\sum_{\alpha\in\Xi} a_\alpha \alpha$. This latter is what we call the *canonical form* of S in Section 1.3.

Definition 1.1.4. *Given* $S = \sum_{\alpha\in\Phi} a_\alpha \alpha \in \mathsf{R}\langle\Phi\rangle$*, its* support $\mathsf{Supp}(S)$ *is the set defined as follows:*

$$\mathsf{Supp}(S) = \{\alpha \in \Phi \mid a_\alpha \neq 0\}.$$

1.2 Algebraic λ-calculus

In this section we introduce the algebraic λ-calculus, denoted Λ_Σ. Originally proposed by Vaux [Vau07, Vau09], and later commonly depicted as the "differential λ-calculus without differentiation", Λ_Σ is a *quantitative* extension of the pure λ-calculus obtained by endowing the set of λ-terms with a structure of vector space, or of R-module. The inspiration for Λ_Σ comes from denotational semantics and, in particular, from Ehrhard's works [Ehr02, Ehr05] on quantitative denotational models of linear logic. This already influences the syntax of the calculus.

After having defined the set of algebraic terms, we spend some time extending some common notions of pure λ-calculus to the current setting, namely (structural) induction on terms, free variables, substitution.

1.2.1 Algebraic terms

We now define the set of terms of the algebraic λ-calculus by following the approach proposed by Ehrhard and Regnier in their original work on the *differential* λ-

calculus [ER03]: *i.e.* terms are introduced as an increasing sequence $(\mathsf{R}\langle\Delta_{\mathsf{R}}(k)\rangle)_{k\in\mathbb{N}}$ of free R-modules generated by simple terms of bounded height.

Definition 1.2.1. *Let be given a denumerable set of variables $\mathcal{V} = \{x, y, z, \dots\}$. The set $\Delta_{\mathsf{R}}(k)$ of* simple terms of height (at most) k *is defined by induction on $k \in \mathbb{N}$: let $\Delta_{\mathsf{R}}(0) = \emptyset$; the elements of $\Delta_{\mathsf{R}}(k+1)$ are defined from those of $\Delta_{\mathsf{R}}(k)$ by the following clauses:*

- *if $s \in \Delta_{\mathsf{R}}(k)$, then $s \in \Delta_{\mathsf{R}}(k+1)$;* [Monotonicity]
- *if $x \in \mathcal{V}$, then $x \in \Delta_{\mathsf{R}}(k+1)$;* [Variable]
- *if $x \in \mathcal{V}$ and $s \in \Delta_{\mathsf{R}}(k)$, then $\lambda x.s \in \Delta_{\mathsf{R}}(k+1)$;* [Abstraction]
- *if $s \in \Delta_{\mathsf{R}}(k)$ and $T \in \mathsf{R}\langle\Delta_{\mathsf{R}}(k)\rangle$, then $(s)\,T \in \Delta_{\mathsf{R}}(k+1)$.* [Application]

The set of all simple terms *is defined as $\Delta_{\mathsf{R}} = \bigcup_{k\in\mathbb{N}} \Delta_{\mathsf{R}}(k)$, whereas the set of* terms *is given by $\mathsf{R}\langle\Delta_{\mathsf{R}}\rangle = \bigcup_{k\in\mathbb{N}} \mathsf{R}\langle\Delta_{\mathsf{R}}(k)\rangle$.*

Notation. Simple terms are ranged over by metavariables like s, t, u, whereas terms by S, T, U. Linear combinations of simple terms S are often written as $\sum_{i=1}^{n} a_i s_i$ in place $\sum_{s\in\Delta_{\mathsf{R}}} a_s s$, or by referring to the support set, as $\sum_{s\in\mathsf{Supp}(S)} a_s s$.

Observe that whenever $s \in \Delta_{\mathsf{R}}$ and $T \in \mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$, then $\lambda x.s \in \Delta_{\mathsf{R}}$ and $(s)\,T \in \Delta_{\mathsf{R}}$. However, in view of how $\mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$ is defined, we need to somehow extend the aforementioned syntactic constructs to the cases when an arbitrary term S is in place of s. We do so by the so-called *algebraic linearity*.

Definition 1.2.2. [Abstraction] *and* [Application] *clauses* (Definition 1.2.1) *are extended by* (algebraic) linearity *as follows:*

$$\lambda x.\left(\sum_{s\in\Delta_{\mathsf{R}}} a_s s\right) = \sum_{s\in\Delta_{\mathsf{R}}} a_s \lambda x.s, \tag{1.2}$$

and

$$\left(\sum_{s\in\Delta_{\mathsf{R}}} a_s s\right) T = \sum_{s\in\Delta_{\mathsf{R}}} a_s\,(s)\,T, \tag{1.3}$$

for all terms $\sum_{s\in\Delta_{\mathsf{R}}} a_s s \in \mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$.

Note how the above characterisation of algebraic linearity determines applications to be linear in the function but not in the argument, in accordance with the computational meaning of linearity given by the decomposition of the intuitionistic implication provided by linear logic [Gir87].

1.2.2 Induction on terms and Substitution

Inductive reasoning in Λ_Σ does not always resemble that of pure λ-calculus, in that defining notions or proving properties by *induction on the size of terms* (similarly, by *structural induction on terms*) is not always amenable. This is to be related with the way relations on linear combinations of terms are conceived and the fact that, for all $S \in \mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$, $\mathbf{0} = aS + bS$ whenever $a, b \in \mathsf{R}^\bullet$ with $a + b = 0$.

We then exploit the definition of $\mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$ as the limit of the increasing sequence $(\mathsf{R}\langle\Delta_{\mathsf{R}}(k)\rangle)_{k\in\mathbb{N}}$ and, by a slight abuse of terminology, call *induction on terms* a way of reasoning on $S \in \mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$ by induction on the least k such that $S \in \mathsf{R}\langle\Delta_{\mathsf{R}}(k)\rangle$.

Definition 1.2.3. *The set of* free variables *of a term S, denoted $\mathsf{FV}(S)$, is defined by induction on S as follows:*

- *Whenever S is a simple term:*
  - *if $S = x$, then $\mathsf{FV}(S) = \{x\}$;*
  - *if $S = \lambda x.u$, then $\mathsf{FV}(S) = \mathsf{FV}(u) \setminus \{x\}$;*
  - *if $S = (u)\,V$, then $\mathsf{FV}(S) = \mathsf{FV}(u) \cup \mathsf{FV}(V)$;*
- *Whenver S is a term:*
  - *if $S = \sum_{u\in\Delta_{\mathsf{R}}} a_u u$, then $\mathsf{FV}(S) = \bigcup_{u\in\mathsf{Supp}(S)} \mathsf{FV}(u)$.*

Notice that, in particular, the empty sum $\mathbf{0}$ has no free variable.

From the previous definition of free variables, one can work out a definition of α-equivalence in the spirit of the one for the pure λ-calculus [Kri93]. Hence, we always identify α-equivalent terms.

We now give the fundamental notion of the (capture-avoiding) substitution of T for x in S, which we write $S\,[T/x]$. Again, the definition follows from the standard one [Kri93] by using algebraic linearity when needed.

Definition 1.2.4. *Let $x \in \mathcal{V}$ and $T \in \mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$. The* substitution of T for x in S, *denoted $S\,[T/x]$, is defined by induction on S as follows:*

- *Whenever S is a simple term:*
  - *if $S = y$, then* $S\,[T/x] = \begin{cases} T & \text{if } y = x, \\ y & \text{otherwise;} \end{cases}$
  - *if $S = \lambda y.u$, then $S\,[T/x] = \lambda y.u\,[T/x]$ provided that $y \neq x$ and $y \notin \mathsf{FV}(T)$;*
  - *if $S = (u)\,V$, then $S\,[T/x] = (u\,[T/x])\,V\,[T/x]$;*

- *Whenever S is a term:*
  - *if $S = \sum_{u \in \Delta_{\mathsf{R}}} a_u u$, then $S[T/x] = \sum_{u \in \Delta_{\mathsf{R}}} a_u u [T/x]$.*

In particular, substitution is linear in S but not in T.

Lemma 1.2.5. *For all $S, T, U \in \mathsf{R}\langle \Delta_{\mathsf{R}} \rangle$ and $x, y \in \mathcal{V}$, with $x \neq y$ and $x \notin \mathsf{FV}(U)$, it holds that*

$$S[T/x][U/y] = S[U/y][T[U/y]/x].$$

Proof. By induction on S. We first address the cases in which S is a simple term s. Then one of the following applies:

- $s \in \mathcal{V}$; hence the only interesting cases are $s = x$ or $s = y$. All the other cases trivially follow. In the former case, on the left-hand side $s[T/x][U/y]$ entails $x[T/x][U/y] = T[U/y]$, as well as on the right-hand side $s[U/y][T[U/y]/x]$ entails $x[U/y][T[U/y]/x] = x[T[U/y]/x] = T[U/y]$.

  In the latter case, on the left-hand side $s[T/x][U/y]$ equals $y[T/x][U/y] = y[U/y] = U$, as well as on the right-hand side $s[U/y][T[U/y]/x]$ entails, since $x \notin \mathsf{FV}(U)$, $y[U/y][T[U/y]/x] = U[T[U/y]/x] = U$.

- $s = \lambda z.v$ and assume $z \neq x, y$ and $z \notin \mathsf{FV}(T) \cup \mathsf{FV}(U)$; hence, it follows

  $$\begin{aligned} s[T/x][U/y] &= (\lambda z.v)[T/x][U/y] \\ &= \lambda z.v[T/x][U/y] \end{aligned}$$

  and, by the induction hypothesis,

  $$\begin{aligned} \lambda z.v[T/x][U/y] &= \lambda z.v[U/y][T[U/y]/x] \\ &= (\lambda z.v)[U/y][T[U/y]/x] \\ &= s[U/y][T[U/y]/x]. \end{aligned}$$

- $s = (v)\,W$; hence, it follows

  $$\begin{aligned} s[T/x][U/y] &= ((v)\,W)[T/x][U/y] \\ &= (v[T/x][U/y])\,W[T/x][U/y] \end{aligned}$$

  and, by the induction hypothesis,

  $$\begin{aligned} (v[T/x][U/y])\,W[T/x][U/y] &= (v[U/y][T[U/y]/x])\,W[U/y][T[U/y]/x] \\ &= ((v)\,W)[U/y][T[U/y]/x] \\ &= s[U/y][T[U/y]/x]. \end{aligned}$$

Now assume $S = \sum_{v \in \Delta_{\mathsf{R}}} a_v v$. Then, $S\,[T/x]\,[U/y]$ amounts to $\sum_{v \in \Delta_{\mathsf{R}}} a_v v\,[T/x]\,[U/y]$. From what we have just shown in the case of simple terms, the latter implies $v\,[T/x]\,[U/y] = v\,[U/y]\,[T\,[U/y]\,/x]$. Hence it follows

$$\begin{aligned} S\,[T/x]\,[U/y] &= \sum_{v \in \Delta_{\mathsf{R}}} a_v v\,[T/x]\,[U/y] \\ &= \sum_{v \in \Delta_{\mathsf{R}}} a_v v\,[U/y]\,[T\,[U/y]\,/x] = S\,[U/y]\,[T\,[U/y]\,/x]\,, \end{aligned}$$

which concludes the proof. □

***Remark* 1.2.6.** As previously mentioned, the above proof follows a schema of developing the inductive reasoning that we heavily use in the rest of this thesis, especially when dealing with relations on terms. Let us give some details about it.

Suppose $\widetilde{\cdot}$ to be a functional that extends relations on Δ_{R} to relations on $\mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$ in a way that, whenever π is defined as $\bigcup_{k \in \mathbb{N}} \pi_k$, then $\widetilde{\pi}$ results to be $\bigcup_{k \in \mathbb{N}} \widetilde{\pi_k}$ (typically, such functional $\widetilde{\cdot}$ must enjoy some properties such as of being monotone and ω-continous). Then, showing some property P by induction on $\widetilde{\pi}$ boils down to prove the following steps:

1. P holds for π_0;

2. suppose P holds for some $\widetilde{\pi_k}$, then P holds for $\widetilde{\pi_{k+1}}$.

In particular, the second case is carried out by showing that:

- (on simple terms) for all $k > 0$, if P holds for $\widetilde{\pi_k}$ then P holds for π_{k+1};
- (on terms) for all $k > 0$, if P holds for π_k then P holds for $\widetilde{\pi_k}$.

It follows that P holds for all $(\widetilde{\pi_k})_{k \in \mathbb{N}}$, hence P holds under the hypothesis $\widetilde{\pi}$.

1.3 The module of algebraic λ-terms

In this section we recall Vaux's presentation [Vau07, Vau09] of the algebraic λ-calculus in order to shed some light on the structure of $\mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$ as the free R-module generated by Δ_{R}. As a matter of fact, although Definition 1.2.1 appears to be quite simple, the way Ehrhard and Regnier define $\mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$ is rather convoluted. We think the following approach helps in understanding Λ_Σ, not to mention the fact of being an interesting construction of $\mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$ *per se*.

We construct the free R-module of terms by consecutive quotients: we introduce the set of *raw terms* (*i.e.* λ-calculus extended with operators of term summation and coefficient multiplication) which we first refine into *permutative terms* (*i.e.* a monoid of raw terms with respect to the sum operator) and later into *algebraic terms* (*i.e.* a free R-module of terms isomorphic to $\mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$).

This construction makes prominent a notion of *algebraic equality*, implicitly involved as an equality on terms in Section 1.2. It turns out that terms which are identified by the algebraic equality exhibit the same *canonical form*.

Notation. If $\mathcal{R}$ is an equivalence relation on a set $\mathcal{S}$, $\mathcal{S}/\mathcal{R}$ denotes the *quotient* of $\mathcal{S}$ modulo $\mathcal{R}$, namely the set of all equivalence classes of $\mathcal{S}$ modulo $\mathcal{R}$.

1.3.1 Raw and Permutative terms

We begin with the language of raw terms, and we provide a first quotient set resulting from the identification of α-equivalent raw terms. Then, we define the set of permutative terms as the monoid of raw terms with respect to the sum operation.

Definition 1.3.1. *Let $\mathcal{V} = \{x, y, z, \ldots\}$ be a denumerable set of variables. The language Λ^0_{R} of* raw terms *with coefficients in* R *is given by the following grammar:*

$$M, N ::= x \;\mid\; \lambda x.M \;\mid\; (M)\,N \;\mid\; \mathbf{0} \;\mid\; aM \;\mid\; M + N.$$

The usual λ-calculus notion of free occurrences of a variable in a term naturally extends to Λ^0_{R} (*i.e.* λ is the only binder), along with the notion of set of free variables of a raw term M, denoted $\mathsf{FV}(M)$. Finally, one defines α-equivalence and capture-avoiding term substitution following a standard scheme [Kri93].

We can therefore identify α-equivalent raw terms.

Definition 1.3.2. *Let $=_\alpha$ denotes α-equivalence. The set Λ^1_{R} of* raw terms modulo α-equivalence *is defined as the quotient set $\Lambda^0_{\mathsf{R}}/{=_\alpha}$.*

The language of raw terms is pure syntax without any algebraic content. As far as the syntax is concerned, terms like $0M$ and $\mathbf{0}$ are not equivalent. The same holds for terms such as $M + N$ and $N + M$, namely terms which only differ in the order of the summands.

Notation. For all $M_1, \ldots, M_n \in \Lambda^1_{\mathsf{R}}$, we write $M_1 + \cdots + M_n$ or even $\sum_{i=1}^n M_i$ for the term $M_1 + (\ldots + M_n)$. We write $\mathbf{0}$ whenever $n = 0$.

Definition 1.3.3. *A binary relation $\mathcal{R} \subseteq \Lambda^1_{\mathsf{R}} \times \Lambda^1_{\mathsf{R}}$ is said to be* contextual *if it satisfies the following conditions:*

- *$x \;\mathcal{R}\; x$, for all $x \in \mathcal{V}$;*
- *if $M \;\mathcal{R}\; N$, then $\lambda x.M \;\mathcal{R}\; \lambda x.N$;*

- *if $M \mathrel{\mathcal{R}} O$ and $N \mathrel{\mathcal{R}} P$, then $(M)\, N \mathrel{\mathcal{R}} (O)\, P$;*
- $\mathbf{0} \mathrel{\mathcal{R}} \mathbf{0}$*;*
- *if $M \mathrel{\mathcal{R}} N$, then $aM \mathrel{\mathcal{R}} aN$;*
- *if $M \mathrel{\mathcal{R}} O$ and $N \mathrel{\mathcal{R}} P$, then $M + N \mathrel{\mathcal{R}} O + P$.*

This notion of contextual relation is the analogue of the λ-compatible relation in the realm of pure λ-calculus [Kri93].

We extend the equality of terms in order to identify terms modulo associativity, the $\mathbf{0}$ summand and the order of appearance of the summands, so that $(\Lambda^1_{\mathsf{R}}, +, \mathbf{0})$ is a commutative monoid. Since the free variables of a sum do not depend on the order of its summands, they are preserved by the following equality.

Definition 1.3.4. Permutative equality*, denoted $\equiv$, is the least contextual equivalence relation on $\Lambda^1_{\mathsf{R}} \times \Lambda^1_{\mathsf{R}}$ such that the following three identities hold for all $M_1, \ldots, M_n, N, O \in \Lambda^1_{\mathsf{R}}$ and all permutations σ of $\{1, \ldots, n\}$:*

$$(M + N) + O \equiv M + (N + O)\,; \tag{1.4}$$

$$\sum_{i=1}^{n} M_i \equiv \sum_{i=1}^{n} M_{\sigma(i)}; \tag{1.5}$$

$$\mathbf{0} + M \equiv M. \tag{1.6}$$

We write Λ_{R} the quotient set $\Lambda^1_{\mathsf{R}}/\equiv$ and we call permutative terms *the elements of Λ_{R}.*

It is easy to verify that substitution is well defined on Λ_{R}. Except when stated otherwise, we use the same notation for a raw term M and its $\equiv$-class, and use them interchangeably. This is in general harmless, since the properties we consider are all invariant with respect to permutative equality.

1.3.2 Free R-module construction

As in the original work on Λ_Σ [Vau09], we extend permutative equality with the identities listed in (1.1) with the aim of obtaining a language providing linear combinations of terms. We name *algebraic* the resulting equality, which we impose to also fulfil the requirements of linearity characterised by the Identities (1.2) and (1.3).

Definition 1.3.5. Algebraic equality*, denoted $\triangleq$, is the least contextual equivalence relation on $\Lambda^1_{\mathsf{R}} \times \Lambda^1_{\mathsf{R}}$ such that $\equiv \subset \triangleq$, the identities in (1.1) hold, along with the following ones:*

$$\lambda x.\mathbf{0} \triangleq \mathbf{0}; \tag{1.7a}$$

$$\lambda x.(aM) \triangleq a(\lambda x.M); \tag{1.7b}$$

$$\lambda x.(M+N) \triangleq \lambda x.M + \lambda x.N; \tag{1.7c}$$

$$(\mathbf{0})\, M \triangleq \mathbf{0}; \tag{1.7d}$$

$$(aM)\, N \triangleq a((M)\, N); \tag{1.7e}$$

$$(M+N)\, O \triangleq (M)\, O + (N)\, O. \tag{1.7f}$$

Notice that the Identities (1.7a)–(1.7c) axiomatise Identity (1.2), whereas the Identities (1.7d)–(1.7f) axiomatise Identity (1.3). Therefore, the quotient set $\Lambda^1_{\mathsf{R}}/\triangleq$ is a free R-module validating algebraic linearity. Moreover, since algebraic equality already subsumes permutative equality on raw terms, $\triangleq$ is well-defined on Λ_{R} and $(\Lambda^1_{\mathsf{R}}/\triangleq) = (\Lambda_{\mathsf{R}}/\triangleq)$.

Definition 1.3.6. *We call* algebraic λ-terms *the elements of* $\Lambda^1_{\mathsf{R}}/\triangleq$, i.e. *the* $\triangleq$*-classes of raw terms. For every* $M \in \Lambda^1_{\mathsf{R}}$*, we write the corresponding* $\triangleq$*-class as* $\underline{M}$*.*

Intuitively, each element of Λ^1_{R} can be thought as a *writing* of its $\triangleq$-class and, among the set of them all, one would like to distinguish the canonical writing.

To make this meaningful, in Section 1.3.3 we show that each raw term M can be uniquely written as $M \triangleq \sum_{i=1}^{n} a_i s_i$, where the s_i's are pairwise distinct base elements and the a_i's are non-zero. We provide an inductive definition of the syntax of the algebraic λ-calculus by identifying two particular subsets of Λ_{R} (in fact, we need to work modulo permutations of summands nonetheless): the first one is the set of *base terms*, namely the set of terms which are intrinsically not sums, whereas the second one is the set of *canonical terms*, namely the set of linear combinations of base terms. This inductively defined construction is proved to match the quotient set $\Lambda_{\mathsf{R}}/\triangleq$, which we know to be the set $\Lambda^1_{\mathsf{R}}/\triangleq$ of algebraic λ-terms.

1.3.3 Canonical forms

We now introduce canonical forms of raw terms as particular permutative terms, and we show that every class in $\Lambda_{\mathsf{R}}/\triangleq$ contains exactly one of them. We conclude that algebraic terms, as defined in Section 1.2, are just canonical forms of raw terms. In doing so, we provide a summary of the construction proposed by Vaux [Vau09].

Definition 1.3.7. *The set* $\mathsf{C}_{\mathsf{R}} \subset \Lambda_{\mathsf{R}}$ *of* canonical terms *and the set* $\mathsf{B}_{\mathsf{R}} \subset \Lambda_{\mathsf{R}}$ *of* base terms *are defined by mutual induction as follows:*

- $x \in \mathsf{B}_{\mathsf{R}}$*, for all* $x \in \mathcal{V}$*;*
- *if* $s \in \mathsf{B}_{\mathsf{R}}$*, then* $\lambda x.s \in \mathsf{B}_{\mathsf{R}}$*;*

- *if* $s \in \mathsf{B_R}$ *and* $T \in \mathsf{C_R}$, *then* $(s)\, T \in \mathsf{B_R}$;
- *if* $a_1, \ldots, a_n \in \mathsf{R}^\bullet$ *and* $s_1, \ldots, s_n \in \mathsf{B_R}$ *are pairwise distinct, then* $\sum_{i=1}^n a_i s_i \in \mathsf{C_R}$.

The intuition is that each canonical form is the most $\triangleq$ simplified term of an entire $\Lambda_\mathsf{R}/\triangleq$ class. Notice that we can easily inject the set $\mathsf{B_R}$ into $\mathsf{C_R}$ by assigning to every base term s the *singleton linear combination* $1s$.

Notation. Base terms are ranged over by metavariables like s, t, u, whereas canonical terms by S, T, U. Observe that we used this same notation to write the elements of $\mathsf{R}\langle\Delta_\mathsf{R}\rangle$ (Section 1.2). This is not accidental, as we show later in this section (Remark 1.3.13).

Definition 1.3.8. *Let* $M = \sum_{i=1}^n a_i s_i \in \Lambda_\mathsf{R}$, *not necessarily a canonical term. For all* $s \in \mathsf{B_R}$, *the* coefficient of s in M *is the scalar* $\sum_{1 \leq i \leq n,\ s_i = s} a_i$ *(the sum of those* a_i*'s such that* $s_i = s$*), denoted* $M_{(s)}$. *Then* $\mathsf{cansum}\,(M) \in \mathsf{C_R}$ *is defined as*

$$\mathsf{cansum}\,(M) = \sum_{j=1}^{m} M_{(t_j)} t_j$$

where $\{t_1, \ldots, t_m\}$ *is the set of those* s_i*'s with a non-zero coefficient in* M.

Each term of Λ_R can be canonised, as there is a natural way to map each permutative term to its respective canonical form.

Definition 1.3.9. *Canonisation of terms* $\mathsf{can} : \Lambda_\mathsf{R} \to \mathsf{C_R}$ *is given by:*

- $\mathsf{can}(x) = 1x$, *for all* $x \in \mathcal{V}$;
- *if* $\mathsf{can}(M) = \sum_{i=1}^n a_i s_i$, *then* $\mathsf{can}(\lambda x.M) = \sum_{i=1}^n a_i (\lambda x.s_i)$;
- *if* $\mathsf{can}(M) = \sum_{i=1}^n a_i s_i$ *and* $\mathsf{can}(N) = T$, *then* $\mathsf{can}((M)\, N) = \sum_{i=1}^n a_i\, (s_i)\, T$;
- $\mathsf{can}(\mathbf{0}) = \mathbf{0}$;
- *if* $\mathsf{can}(M) = \sum_{i=1}^n a_i s_i$, *then* $\mathsf{can}(aM) = \mathsf{cansum}\,(\sum_{i=1}^n (aa_i) s_i)$;
- *if* $\mathsf{can}(M) = \sum_{i=1}^n a_i s_i$ *and* $\mathsf{can}(N) = \sum_{i=n+1}^{n+m} a_i s_i$, *then* $\mathsf{can}(M + N) = \mathsf{cansum}\,\left(\sum_{i=1}^{n+m} a_i a_i\right)$.

Note that definition of $\mathsf{can}(aM)$ needs cansum to prune all the summands $(aa_i)s_i$ such that $aa_i = 0$. This is important to handle the cases where R might not be an integral domain, namely whenever $ab = 0$ does not imply either a or b to be 0.

Obviously, canonisation does not affect canonical terms.

Lemma 1.3.10. *For all $S \in \mathsf{C}_\mathsf{R}$, $\mathsf{can}(S) = S$.*

Moreover one can prove the following key property:

Theorem 1.3.11. *Algebraic equality is equality of canonical forms: for all $M, N \in \Lambda_\mathsf{R}$, $M \triangleq N$ if and only if $\mathsf{can}(M) = \mathsf{can}(N)$.*

Proof (Sketch). We consider the equivalence relation $\triangleq'$ on Λ_R defined as $M \triangleq' N$ iff $\mathsf{can}(M) = \mathsf{can}(N)$, for all $M, N \in \Lambda_\mathsf{R}$. Then, can makes $\triangleq'$ a contextual relation (Definition 1.4.1) consistent with algebraic linearity (Definition 1.3.5), so that every $\triangleq$-equation holds for $\triangleq'$ as well, *i.e.* $\triangleq \subseteq \triangleq'$. Conversely one verifies that, for all $M \in \Lambda_\mathsf{R}$, $\mathsf{can}(M) \triangleq M$. Hence $M \triangleq' N$ implies $M \triangleq \mathsf{can}(M) = \mathsf{can}(N) \triangleq N$. $\square$

Corollary 1.3.12.

1. *For all $S, T \in \mathsf{C}_\mathsf{R}$, $S \triangleq T$ if and only if $S = T$.*
2. *C_R admits an R-module structure so that can is an isomorphism of R-modules from $\Lambda_\mathsf{R}/\triangleq$ to C_R.*

Proof.

1. This is a direct consequence of Theorem 1.3.11 and Lemma 1.3.10.
2. Consider $\mathsf{C}_\mathsf{R} = (\mathsf{C}_\mathsf{R}, +, \mathbf{0}, \cdot)$ where each operation is defined by means of can. Formally, $\mathbf{0} \in \mathsf{C}_\mathsf{R}$ and the two binary operations $+, \cdot$ are defined as follows:

$$\begin{array}{rrcc} \text{sum:} & (M, N) \in \mathsf{C}_\mathsf{R} \times \mathsf{C}_\mathsf{R} & \mapsto & \mathsf{can}(M + N) \in \mathsf{C}_\mathsf{R} \\ \text{scalar multiplication:} & (a, M) \in \mathsf{R} \times \mathsf{C}_\mathsf{R} & \mapsto & \mathsf{can}(aM) \in \mathsf{C}_\mathsf{R} \end{array}$$

   The isomorphism of R-modules easily holds: by Theorem 1.3.11, can is well defined on $\Lambda_\mathsf{R}/\triangleq$ and it is injective; by Lemma 1.3.10, it is also surjective. Then, the R-module structure on C_R follows from that of $\Lambda_\mathsf{R}/\triangleq$. $\square$

The isomorphism of R-modules between $\Lambda_\mathsf{R}/\triangleq$ and C_R formally confirms that the quotient structure of algebraic terms is subsumed by the mutually inductive structure of base terms and canonical terms. More precisely, if we denote $\underline{\mathsf{C}}$ the set $\{\underline{S} \mid S \in \mathsf{C}\}$ of algebraic terms (*i.e.* $\triangleq$-classes) given a set C of canonical terms, then it follows that $(\Lambda_\mathsf{R}/\triangleq) = \underline{\mathsf{C}_\mathsf{R}}$.

***Remark* 1.3.13.** As already mentioned at the start of this section, $\Lambda_\mathsf{R}/\triangleq$ and $\mathsf{R}\langle\Delta_\mathsf{R}\rangle$ are the same R-module of algebraic λ-terms (to be precise, they are up to isomorphism). Indeed, consider the following notion of height formulated by means of base terms and canonical ones,

Definition 1.3.14. *The* height *of base terms and canonical terms is defined by mutual induction as follows:*

- $\mathsf{h}(x) = 1$, *for all* $x \in \mathcal{V}$;
- $\mathsf{h}(\lambda x.s) = 1 + \mathsf{h}(s)$;
- $\mathsf{h}((s)\,T) = 1 + \max\{\mathsf{h}(s), \mathsf{h}(T)\}$;
- $\mathsf{h}(\sum_{i=1}^{n} a_i s_i) = \max\{\mathsf{h}(s_i) \mid 1 \leq i \leq n\}$ *(which is* 0 *if* $n = 0$*).*

Now, if we define $\mathsf{B}_\mathsf{R}(k)$ and $\mathsf{C}_\mathsf{R}(k)$ to be the sets of base terms and canonical terms of height at most k respectively, then it turns out that $\Delta_\mathsf{R}(k) = \underline{\mathsf{B}_\mathsf{R}(k)}$ and $\mathsf{R}\langle\Delta_\mathsf{R}(k)\rangle = \underline{\mathsf{C}_\mathsf{R}(k)}$. Hence $\Delta_\mathsf{R} = \underline{\mathsf{B}_\mathsf{R}}$ and $\mathsf{R}\langle\Delta_\mathsf{R}\rangle = \underline{\mathsf{C}_\mathsf{R}} = (\Lambda_\mathsf{R}/\triangleq)$.

The correspondence between $\mathsf{R}\langle\Delta_\mathsf{R}\rangle$ and $\Lambda_\mathsf{R}/\triangleq$ reveals the hidden complexity of Definition 1.2.1. In particular, the construction of quotients leading to $\Lambda_\mathsf{R}/\triangleq$ highlights the fact that Definition 1.2.1 involves α-equivalence and the free R-module construction at each height. This especially means that the notion of equality on terms in $\mathsf{R}\langle\Delta_\mathsf{R}\rangle$ implicitly subsumes algebraic equality.

***Example* 1.3.15.** Raw terms Λ^1_R are just syntax. This means that, given $s \in \Lambda^1_\mathsf{R}$,

$$s + \mathbf{0} \neq \mathbf{0} + s \neq s$$

even though, the three are intuitively the same raw term. Permutative terms Λ_R overcome this issue as $s + \mathbf{0} \equiv \mathbf{0} + s \equiv s$. Nonetheless, in the realm of Λ_R,

$$s + s \neq 2s.$$

Then, since $s + s \triangleq 2s$, the two are equivalent algebraic terms because their $\triangleq$-class are: *i.e.* $\underline{s + s} = \underline{2s}$. Observe that $2s$ is the canonical form of $s + s$.

Definition 1.2.1 of $\mathsf{R}\langle\Delta_\mathsf{R}\rangle$ confuses all these different syntactical levels as equality on terms implicitly subsumes algebraic equality (hence, permutative equality as well), insomuch as each of the above examples determines the same term.

The works introducing Λ_Σ [Vau07, Vau09] adopts this last concrete construction of $\mathsf{R}\langle\Delta_\mathsf{R}\rangle$, along with the respective notation for writing terms. This entails different notations to distinguish whenever an algebraic term is written in its canonical form or not. Nonetheless, the way algebraic terms are written, as to remind that they are $\triangleq$-classes, is already burdensome. In those works, this accuracy about notations was somehow mandatory to highlight the diffent quotients.

In this thesis, however, we focus more on the analysis of the dynamics of Λ_Σ. Therefore, we despence with that heavy notation entirely and work instead in the setting *à la* Ehrhard and Regnier [ER03], namely with an implicit way of treating equality on terms, we formerly set up in Section 1.2.

1.4 Reduction relations on algebraic λ-terms

In this section we follow Vaux [Vau09], or even Ehrhard and Regnier [ER03], and provide a dynamics for the algebraic λ-calculus by extending to Λ_Σ the classical notion of β-reduction. The crucial point is the definition of β-reduction in presence of linear combinations of terms: an algebraic term reduces whenever one of its simple terms does. This involves the important notion of contextuality to be revised. We conclude by showing the Church-Rosser property for Λ_Σ by adapting a well-known technique due to Tait and Martin-Löf. In order to do so, we need to define a notion of parallel reduction.

Notation. Given a binary relation $\mathcal{R}$, we denote as $\mathcal{R}^?$ its reflexive closure, as $\mathcal{R}^+$ its transitive closure, whereas $\mathcal{R}^*$ denotes its reflexive and transive closure.

Definition 1.4.1. *A binary relation* $\mathcal{R} \subseteq \mathsf{R}\langle\Delta_\mathsf{R}\rangle \times \mathsf{R}\langle\Delta_\mathsf{R}\rangle$ *is said to be* contextual *if it is reflexive and it satisfies the following clauses:*

1. *if* $S\ \mathcal{R}\ T$, *then* $\lambda x.S\ \mathcal{R}\ \lambda x.T$;
2. *if* $S\ \mathcal{R}\ T$ *and* $U\ \mathcal{R}\ V$, *then* $(S)\,U\ \mathcal{R}\ (T)\,V$;
3. *if* $S\ \mathcal{R}\ T$, *then* $aS\ \mathcal{R}\ aT$;
4. *if* $S\ \mathcal{R}\ T$ *and* $U\ \mathcal{R}\ V$, *then* $S + U\ \mathcal{R}\ T + V$.

Remark **1.4.2.** Contextuality is a crucial property when studying relations on terms. A relation that enjoys such property respects the way terms are constructed, enabling a *compositional* reasoning: a relation between terms can be deduced from those relating the subterms. The induction reasoning is a clear instance.

As every compatible relation of pure λ-calculus [Kri93], this notion of contextual relation admits the following crucial property:

Lemma 1.4.3. *If* $\mathcal{R}$ *is a contextual relation, then* $S\,[T/x]\ \mathcal{R}\ S\,[U/x]$ *as soon as* $T\ \mathcal{R}\ U$.

We identify every relation $\mathcal{R}$ between terms either as a *relation from simple terms to terms*, whenever $\mathcal{R}$ is a subset of $\Delta_\mathsf{R} \times \mathsf{R}\langle\Delta_\mathsf{R}\rangle$, or as a *relation from terms to terms*, whenever $\mathcal{R}$ is a subset of $\mathsf{R}\langle\Delta_\mathsf{R}\rangle \times \mathsf{R}\langle\Delta_\mathsf{R}\rangle$.

Definition 1.4.4. *Given a relation* $\mathcal{R} \subseteq \Delta_{\mathsf{R}} \times \mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$*, the extended relations* $\widetilde{\mathcal{R}}, \overline{\mathcal{R}} \subseteq \mathsf{R}\langle\Delta_{\mathsf{R}}\rangle \times \mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$ *are respectively defined as follows:*

$$S \; \widetilde{\mathcal{R}} \; S' \text{ if } S = au + T \text{ and } S' = aU' + T \text{ where } a \neq 0 \text{ and } u \; \mathcal{R} \; U'; \tag{1.8a}$$

$$S \; \overline{\mathcal{R}} \; S' \text{ if } S = \sum_{i=1}^{n} a_i u_i \text{ and } S' = \sum_{i=1}^{n} a_i U_i' \text{ where, for all } i \in \{1, \ldots, n\}, u_i \; \mathcal{R} \; U_i'. \tag{1.8b}$$

***Remark* 1.4.5.** Observe how Rules (1.8a) and (1.8b) are formulated: the former allows the simple term u to appear in T (*i.e.* u can be part of $\mathsf{Supp}(T)$), whereas the latter does not impose the u_i's to be pairwise distinct simple terms. Using the terminology introduced in Section 1.3, we would say that Rules (1.8a) and (1.8b) do not respectively require $au + T$ nor $\sum_{i=1}^{n} a_i u_i$ to be canonical terms.

Although they might seem convoluted at first sight, the definitions we gave of Rules (1.8a) and (1.8b) turn to be somehow mandatory for obtaining contextual (reduction) relations on terms in the sense of Definition 1.4.1. We recall that contextuality is an important property whenever one wants to reason compositionally or inductively (Remark 1.4.2), crucial principles when studying the reduction theory of a calculus. In Λ_Σ, this is even more important for achieving the intuition we have when thinking about linear combinations of terms as complex objects made of single elements which can behave independently.

A sample of this issue already arises whenever one needs to *write* related terms by starting from related simple terms. For instance, suppose the case of a Rule (1.8b) defined by imposing the u_i's to be pairwise distinct simple terms, and consider $S = u + T$ with $u \; \mathcal{R} \; U'$. Then, since u may appear in T as well, we cannot simply write $S \; \overline{\mathcal{R}} \; S'$ with $S' = U' + T$, in general. On the contrary, the latter is a obvious valid result under the actual formulation of Rule (1.8b).

Nevertheless, Rules (1.8a) and (1.8b) involve non-negligible complications which already appears when defining reduction relations on terms. Clearly $\widetilde{\mathcal{R}} \subseteq \overline{\mathcal{R}}$, and so an obvious idea is to use these constructions in the definition of β-reduction and its parallel version: respectively introduce $\rightarrow$ and $\rightrightarrows$ as relations from simple terms to terms, so that the actual reduction relations on terms are defined as $\widetilde{\rightarrow}$ and $\overline{\rightrightarrows}$. We need to be careful about details though, as already observed in previous works [ER03, Vau09]: generally speaking, reduction notions cannot be simply defined by (structural) induction on terms.

1.4.1 β-reduction

In the pure λ-calculus, (full) β-reduction is informally specified as the least binary relation that satisfies the β-rule

$$(\lambda x.u)\, v \;\mapsto_\beta\; u\,[v/x]\,, \tag{β}$$

under every [Abstraction] and [Application] context.

Notation. A term of the form $(\lambda x.u)\, v$ is called (β-)redex (*i.e. reducible expression*) and the result $u\,[v/x]$ of rewriting a redex according to the β-rule is called (β-)reduct. In the following, we sometimes depict such operation as the *contracting* (or *contraction*) of a redex, or even the *firing* of a redex. In any case, we say that the term $(\lambda x.u)\, v$ reduces to $u\,[v/x]$.

Following Krivine [Kri93], a formal definition of β-reduction $s \to_\beta s'$ would proceed by induction on (the size of) s as follows:

- if $s = \lambda x.u$, then $s' = \lambda x.u'$ whenever $u \to_\beta u'$; [Abstraction]
- if $s = (u)\, v$, then [Application]
  - $s' = (u')\, v$ whenever $u \to_\beta u'$, or
  - $s' = (u)\, v'$ whenever $v \to_\beta v'$;
- if $s = (\lambda x.u)\, v$, then $s' = u\,[v/x]$. [Redex]

A similar definition is not always available in the current setting: whenever S reduces to S' and $a, b \in \mathsf{R}\backslash\{0\}$ such that $a + b = 0$, then $aS + bS = \mathbf{0}$ for all $S \in \mathsf{R}\langle\Delta_\mathsf{R}\rangle$, which implies by Rule (1.8a) that $\mathbf{0}$ can reduce to $aS' + bS$. In a definition by induction like the one above, this would mean having to accept the reduction of $(u)\,\mathbf{0}$ to $(u)\,(aS' + bS)$, suggesting the size of an *arbitrary term* S to be less than the size of $(u)\,\mathbf{0}$.

Accordingly to Ehrhard and Regnier's differential λ-calculus [ER03], we define the reduction notion $\to$ as the relation obtained by taking the union of an increasing sequence of relations on $\Delta_\mathsf{R} \times \mathsf{R}\langle\Delta_\mathsf{R}\rangle$, ordered by the depth of the fired redex.

Definition 1.4.6. *Let $\to_0$ be the empty relation $\varnothing \subseteq \Delta_\mathsf{R} \times \mathsf{R}\langle\Delta_\mathsf{R}\rangle$. Assume $\to_k$ is defined and set $s \to_{k+1} S'$ as soon as one of the following holds:*

- *$s = \lambda x.u$ and $S' = \lambda x.U'$ with $u \to_k U'$;*
- *$s = (u)\, V$ and $S' = (U')\, V$ with $u \to_k U'$, or $S' = (u)\, V'$ with $V \mathrel{\widetilde{\to}_k} V'$;*
- *$s = (\lambda x.u)\, V$ and $S' = u\,[V/x]$.*

Let $\to = \bigcup_{k\in\mathbb{N}} \to_k$. *We call* β-reduction, *or simply* reduction, *the relation* $\widetilde{\to}$.

The operation introduced by Rule (1.8a) has the general property of being monotone and ω-continuous in $\mathcal{R}$, meaning that whenever $\mathcal{R}$ is defined as the union of an increasing sequence of relations $(\mathcal{R}_k)_{k\in\mathbb{N}}$, then $\left(\widetilde{\mathcal{R}_k}\right)_{k\in\mathbb{N}}$ is increasing and its union matches $\widetilde{\mathcal{R}}$. Therefore, one can prove that:

Lemma 1.4.7. $\widetilde{\to} = \bigcup_{k\in\mathbb{N}} \widetilde{\to}_k$.

Definition 1.4.6 can be easily proved to admit the following rephrasing, where Lemma 1.4.7 is needed whenever $\widetilde{\to}$ is involved.

Lemma 1.4.8. *Let* $s \in \Delta_{\mathsf{R}}$ *and* $S' \in \mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$. *It follows that* $s \to S'$ *if and only if one of the following holds:*

- $s = \lambda x.u$ *and* $S' = \lambda x.U'$ *with* $u \to U'$;
- $s = (u)\, V$ *and* $S' = (U')\, V$ *with* $u \to U'$, *or* $S' = (u)\, V'$ *and* $V \widetilde{\to} V'$;
- $s = (\lambda x.u)\, V$ *and* $S' = u\,[V/x]$.

***Remark* 1.4.9.** Observe that the last clause is an instance of the β-rule, although a (β)-redex in Λ_Σ exhibits a term (*i.e.* a linear combination of simple terms) V as argument of the [Application].

We now proceed by proving that β-reduction (actually, for its reflexive and transitive closure $\widetilde{\to}^*$) enjoys the contextual property expressed by Definition 1.4.1.

Lemma 1.4.10. *Let* $S, S', T \in \mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$. *If* $S \widetilde{\to} S'$, *then the following hold:*

1. $\lambda x.S \widetilde{\to} \lambda x.S'$;
2. $(S)\, T \widetilde{\to} (S')\, T$;
3. $(T)\, S \widetilde{\to}^* (T)\, S'$;
4. $aS \widetilde{\to}^? aS'$, *for all* $a \in \mathsf{R}$;
5. $S + T \widetilde{\to} S' + T$.

Proof. By Definition 1.4.6 of $\widetilde{\to}$ and Rule (1.8a), $S \widetilde{\to} S'$ amounts to the following: $S = bu + V$ and $S' = bU' + V$ with $b \neq 0$ and $u \to U'$. Let us prove each statement one at a time:

1. Lemma 1.4.8 on $u \to U'$ entails $\lambda x.u \to \lambda x.U'$. Hence

$$\lambda x.S = b\lambda x.u + \lambda x.V \rightrightarrows b\lambda x.U' + \lambda x.V = \lambda x.S'.$$

2. Lemma 1.4.8 on $u \to U'$ entails $(u)\,T \to (U')\,T$. Hence

$$(S)\,T = b\,(u)\,T + (V)\,T \rightrightarrows b\,\left(U'\right)\,T + (V)\,T = \left(S'\right)\,T.$$

3. Let $T = \sum_{i=1}^{n} a_i t_i$, so that $(T)\,S = \sum_{i=1}^{n} a_i\,(t_i)\,S$ and $(T)\,S' = \sum_{i=1}^{n} a_i\,(t_i)\,S'$. By Lemma 1.4.8 follows $(t_i)\,S \to (t_i)\,S'$, for every $i \in \{1,\ldots,n\}$. Hence

$$(T)\,S = \sum_{i=1}^{n} a_i\,(t_i)\,S \rightrightarrows^* \sum_{i=1}^{n} a_i\,(t_i)\,S' = (T)\,S'.$$

4. For all $a \in \mathsf{R}$, $aS = abu + aV$ and $aS' = abU' + aV$. If $ab = 0$, then $abu = abU' = \mathbf{0}$ which implies $aS = aV = aS'$. Otherwise

$$aS = abu + aV \rightrightarrows abU' + aV = aS'.$$

In any case, $aS \rightrightarrows^? aS'$.

5. Straightforward by Definition 1.4.6 of $\rightrightarrows$ and Rule (1.8a)

$$S + T = bu + V + T \rightrightarrows bU' + V + T = S' + T.$$

This concludes the proof. □

Proposition 1.4.11. *The relation $\rightrightarrows^*$ is contextual.*

Proof. This is a direct consequence of Lemma 1.4.10, using the reflexive and transitive properties of $\rightrightarrows^*$. Let us detail only the 2nd and 4th clauses of Definition 1.4.1 with respect to $\rightrightarrows^*$:

2. Lemma 1.4.10 on $S \rightrightarrows^* T$ and $U \rightrightarrows^* V$, along with the reflexive property of $\rightrightarrows^*$, entails $(S)\,U \rightrightarrows^* (T)\,U$ and $(T)\,U \rightrightarrows^* (T)\,V$. Therefore $(S)\,U \rightrightarrows^* (T)\,V$ follows by the transitive property of $\rightrightarrows^*$.

4. Lemma 1.4.10 on $S \rightrightarrows^* T$ and $U \rightrightarrows^* V$, along with the reflexive property of $\rightrightarrows^*$, entails $S + U \rightrightarrows^* T + U$ and $T + U \rightrightarrows^* T + V$. Therefore $S + U \rightrightarrows^* T + V$ follows by the transitive property of $\rightrightarrows^*$.

The other cases follow by a similar reasoning. □

1.4.2 Parallel β-reduction

Crucial in this work is the notion of parallel reduction by which multiple redexes can be fired simultaneously. Here we introduce the parallel extension of reduction $\widetilde{\to}$, adapting by means of Rule (1.8b) the pure λ-calculus notion of parallel β-reduction. In the classical setting [Bar84], parallel β-reduction $s \rightrightarrows_\beta s'$ would be defined by induction on (the size of) s as follows:

- if $s \in \mathcal{V}$, $s \rightrightarrows_\beta s$; [Variable]
- if $s = \lambda x.u$, then $s' = \lambda x.u'$ whenever $u \rightrightarrows_\beta u'$; [Abstraction]
- if $s = (u)\,v$, then $s' = (u')\,v'$ whenever $u \rightrightarrows_\beta u'$ and $v \rightrightarrows_\beta v'$; [Application]
- if $s = (\lambda x.u)\,v$, then $s' = u'\,[v'/x]$ whenever $u \rightrightarrows_\beta u'$ and $v \rightrightarrows_\beta v'$. [Redex]

The last clause is what we call parallel β-rule. A peculiarity of parallel reduction is the fact that it is reflexive in itself.

The technical problems highlighted in the process of defining β-reduction arise here as well. Thus, we similarly define the reduction notion $\rightrightarrows$ as the relation obtained by taking the union of an increasing sequence of relations on $\Delta_\mathsf{R} \times \mathsf{R}\langle\Delta_\mathsf{R}\rangle$.

Definition 1.4.12. *Let $\rightrightarrows_0$ be the identity relation on $\Delta_\mathsf{R} \times \Delta_\mathsf{R}$, extended as a relation on $\Delta_\mathsf{R} \times \mathsf{R}\langle\Delta_\mathsf{R}\rangle$. Assume $\rightrightarrows_k$ is defined and set $s \rightrightarrows_{k+1} S'$ as soon as one of the following holds:*

- *$s = x$ and $S' = x$, for all $x \in \mathcal{V}$;*
- *$s = \lambda x.u$ and $S' = \lambda x.U'$ with $u \rightrightarrows_k U'$;*
- *$s = (u)\,V$ and $S' = (U')\,V'$ with $u \rightrightarrows_k U'$ and $V \overline{\rightrightarrows_k} V'$;*
- *$s = (\lambda x.u)\,V$ and $S' = U'\,[V'/x]$ with $u \rightrightarrows_k U'$ and $V \overline{\rightrightarrows_k} V'$.*

Let $\rightrightarrows = \bigcup_{k\in\mathbb{N}} \rightrightarrows_k$. We call parallel ($\beta$-)reduction *the relation $\overline{\rightrightarrows}$.*

The following results show that $\overline{\rightrightarrows}$ is indeed the limit of an increasing ω-chain of reduction relations. We use such properties in Lemma 2.3.21 to deduce that a set of such relations admits a greatest one, thus allowing a common inductive reasoning.

Lemma 1.4.13. *For all $k \in \mathbb{N}$, with $k \geq 1$, $\overline{\rightrightarrows_{k-1}} \subseteq \overline{\rightrightarrows_k}$ holds.*

Proof. By induction on $k \geq 1$. If $k = 1$, then $\overline{\rightrightarrows_0} \subseteq \overline{\rightrightarrows_1}$ directly follows by Definition 1.4.12 of $\overline{\rightrightarrows}$ and the monotonicity of Rule (1.8b). Suppose the result $\overline{\rightrightarrows_{k-1}} \subseteq \overline{\rightrightarrows_k}$ holds for all $k > 1$, then we prove $\overline{\rightrightarrows_k} \subseteq \overline{\rightrightarrows_{k+1}}$. The latter boils down to show that, for all $S, S' \in \mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$, $S \overline{\rightrightarrows_k} S'$ implies $S \overline{\rightrightarrows_{k+1}} S'$. We proceed by inspecting the possible cases for reduction $S \overline{\rightrightarrows_k} S'$, and we first address the cases in which S is a simple term s and $s \rightrightarrows_k S'$. Then, one of the following applies:

- $s \in \mathcal{V}$; hence the result directly follows by Definition 1.4.12.
- $s = \lambda x.u$ and $S' = \lambda x.U'$ with $u \rightrightarrows_{k-1} U'$; hence, by the induction hypothesis, $u \rightrightarrows_k U'$. By Definition 1.4.12 of $\rightrightarrows$ follows
$$s = \lambda x.u \rightrightarrows_{k+1} \lambda x.U' = S'.$$
- $s = (u)\, V$ and $S' = (U')\, V'$ with $u \rightrightarrows_{k-1} U'$ and $V \overline{\rightrightarrows_{k-1}} V'$; hence, by the induction hypothesis, $u \rightrightarrows_k U'$ and $V \overline{\rightrightarrows_k} V'$. By Definition 1.4.12 of $\rightrightarrows$ follows
$$s = (u)\, V \rightrightarrows_{k+1} \left(U'\right) V' = S'.$$
- $s = (\lambda x.u)\, V$ and $S' = U'\,[V'/x]$ with $u \rightrightarrows_{k-1} U'$ and $V \overline{\rightrightarrows_{k-1}} V'$; hence, by the induction hypothesis, $u \rightrightarrows_k U'$ and $V \overline{\rightrightarrows_k} V'$. By Definition 1.4.12 of $\rightrightarrows$ follows
$$s = (\lambda x.u)\, V \rightrightarrows_{k+1} U' \left[V'/x\right] = S'.$$

Now assume $S \overline{\rightrightarrows_k} S'$. By Rule (1.8b), this amounts to the following: $S = \sum_{i=1}^{n} a_i u_i$ and $S' = \sum_{i=1}^{n} a_i U_i'$ with $u_i \rightrightarrows_k U_i'$, for all $i \in \{1, \dots, n\}$. From what we have just shown in the case of simple terms follows $u_i \rightrightarrows_{k+1} U_i'$, for all $i \in \{1, \dots, n\}$. Hence, by Rule (1.8b) follows
$$S = \sum_{i=1}^{n} a_i u_i \overline{\rightrightarrows_{k+1}} \sum_{i=1}^{n} a_i U_i' = S'.$$

This concludes the proof. $\square$

Lemma 1.4.14. *For all $j, k \in \mathbb{N}$, with $j \leq k$, $\overline{\rightrightarrows_j} \subseteq \overline{\rightrightarrows_k}$ holds.*

Proof. Direct consequence of Lemma 1.4.13 and the fact that $\subseteq$ is a preorder. $\square$

As in the case of β-reduction, the operation introduced by Rule (1.8b) is monotone and ω-continuous in $\mathcal{R}$, therefore implying that:

Lemma 1.4.15. $\overline{\rightrightarrows} = \bigcup_{k \in \mathbb{N}} \overline{\rightrightarrows_k}$.

Similarly to Lemma 1.4.8, the following lemma is just a rephrasing of Definition 1.4.12 using Lemma 1.4.15 whenever $\overline{\rightrightarrows}$ is involved.

Lemma 1.4.16. *Let $s \in \Delta_{\mathsf{R}}$ and $S' \in \mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$. It follows that $s \rightrightarrows S'$ if and only if one of the following holds:*

- *$s = x$ and $S' = x$, for all $x \in \mathcal{V}$;*
- *$s = \lambda x.u$ and $S' = \lambda x.U'$ with $u \rightrightarrows U'$;*
- *$s = (u)\, V$ and $S' = (U')\, V'$ with $u \rightrightarrows U'$ and $V \overline{\rightrightarrows} V'$;*
- *$s = (\lambda x.u)\, V$ and $S' = U'\,[V'/x]$ with $u \rightrightarrows U'$ and $V \overline{\rightrightarrows} V'$.*

We ultimately prove that relation $\tilde{\rightarrow}$ is a strict subrelation of $\overline{\rightrightarrows}$ (Section 1.4.3). Before anything else, this requires $\overline{\rightrightarrows}$ to be a contextual relation, hence reflexive.

Lemma 1.4.17. *The relation $\overline{\rightrightarrows}$ is reflexive.*

Proof. Simple induction on the term S such that $S \overline{\rightrightarrows} S$, using Lemma 1.4.16. □

Proposition 1.4.18. *The relation $\overline{\rightrightarrows}$ is contextual.*

Proof. The reflexive property of $\overline{\rightrightarrows}$ follows by Lemma 1.4.17. Let us prove each clause of Definition 1.4.1 with respect to $\overline{\rightrightarrows}$, one at a time:

1. We need to prove that $\lambda x.S \overline{\rightrightarrows} \lambda x.S'$ whenever $S \overline{\rightrightarrows} S'$. Hence, Rule (1.8b) entails $S = \sum_{i=1}^{n} a_i u_i$ and $S' = \sum_{i=1}^{n} a_i U_i'$ with $u_i \rightrightarrows U_i'$ for every $i \in \{1, \dots, n\}$, and Lemma 1.4.16 implies $\lambda x.u_i \rightrightarrows \lambda x.U_i'$ for every $i \in \{1, \dots, n\}$. By Rule (1.8b) follows
$$\lambda x.S = \sum_{i=1}^{n} a_i \lambda x.u_i \overline{\rightrightarrows} \sum_{i=1}^{n} a_i \lambda x.U_i' = \lambda x.S'.$$

2. We need to prove that $(S)\, T \overline{\rightrightarrows} (S')\, T'$ whenever $S \overline{\rightrightarrows} S'$ and $T \overline{\rightrightarrows} T'$. On the first hypothesis, Rule (1.8b) entails $S = \sum_{i=1}^{n} a_i u_i$ and $S' = \sum_{i=1}^{n} a_i U_i'$ with $u_i \rightrightarrows U_i'$ for every $i \in \{1, \dots, n\}$, and Lemma 1.4.16 implies $(u_i)\, T \rightrightarrows (U_i')\, T'$ for every $i \in \{1, \dots, n\}$. By Rule (1.8b) follows
$$(S)\, T = \sum_{i=1}^{n} a_i\, (u_i)\, T \overline{\rightrightarrows} \sum_{i=1}^{n} a_i\, \left(U_i'\right) T' = \left(S'\right) T'.$$

3. We need to prove that $bS \overline{\rightrightarrows} bS'$ whenever $S \overline{\rightrightarrows} S'$. Hence, Rule (1.8b) entails $S = \sum_{i=1}^{n} a_i u_i$ and $S' = \sum_{i=1}^{n} a_i U_i'$ with $u_i \rightrightarrows U_i'$ for every $i \in \{1, \dots, n\}$. Let $J \subseteq \{1, \dots, n\}$ such that $ba_i = 0$, and write $T = \sum_{j \in J} a_j u_j$ and $T' = \sum_{j \in J} a_j U_j'$ such that $u_j \rightrightarrows U_j'$ for all $j \in J$. It follows $bT = \mathbf{0} = bT'$. By Rule (1.8b) follows
$$bS = \sum_{i=1, i \notin J}^{n} ba_i u_i + bT = \sum_{i=1, i \notin J}^{n} ba_i u_i \overline{\rightrightarrows} \sum_{i=1, i \notin J}^{n} ba_i U_i' = \sum_{i=1, i \notin J}^{n} ba_i U_i' + bT' = bS'.$$

4. We need to prove that $S + T \overline{\rightrightarrows} S' + T'$ whenever $S \overline{\rightrightarrows} S'$ and $T \overline{\rightrightarrows} T'$. This directly follows by Definition 1.4.12 of $\overline{\rightrightarrows}$ and Rule (1.8b).

This concludes the proof. □

Using algebraic linearity (Definition 1.2.2), Lemma 1.4.16 implies the following result, which generalises parallel β-rule to sum terms as well.

Lemma 1.4.19. $(\lambda x.S)\, T \overline{\rightrightarrows} S'\,[T'/x]$ *whenever* $S \overline{\rightrightarrows} S'$ *and* $T \overline{\rightrightarrows} T'$.

This result plays a crucial role in establishing strong confluence for $\overline{\rightrightarrows}$.

Lemma 1.4.20. *Let* $x \in \mathcal{V}$ *and* $S, S', T, T' \in \mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$. *If* $S \overline{\rightrightarrows} S'$ *and* $T \overline{\rightrightarrows} T'$, *then* $S\,[T/x] \overline{\rightrightarrows} S'\,[T'/x]$.

Proof. We refer to Vaux's work [Vau09] for the detailed proof. □

1.4.3 Confluence

We now proceed by showing an important property of reduction in Λ_Σ, namely the fact that reduction relation $\widetilde{\to}$ enjoys *confluence*. This property is also known as the *Church-Rosser property* [Bar84]. In rewriting theory, it establishes as fictional the non-determinism of rewriting and entails the uniqueness of normal forms.

Definition 1.4.21. *A binary relation* $\to$ *on a set* $\mathcal{S}$ *is said to exhibit* confluence *whenever its reflexive and transitive closure* $\to^*$ *enjoys strong confluence.*

In general, a relation $\to$ that exhibits *strong confluence* (also called *diamond property*) directly implies relation $\to^*$ to be strongly confluent, by means of a simple diagram chase. We recall here its definition.

Definition 1.4.22. *A binary relation* $\to$ *on a set* $\mathcal{S}$ *is said to exhibit* strong confluence *(or* diamond property*) whenever, for all* $M, N, O \in \mathcal{S}$ *such that* $M \to N$ *and* $M \to O$, *there exists* $P \in \mathcal{S}$ *such that* $N \to P$ *and* $O \to P$.

However, since Λ_Σ is an extension of pure λ-calculus, reduction $\widetilde{\to}$ does not enjoy strong confluence. The reason is well-known: *duplication*.

A more interesting aspect is how the algebraic nature of Λ_Σ influences confluence. Indeed, it turns out that whenever a relation $\mathcal{R}$ from simple terms to terms is extended to a relation $\widetilde{\mathcal{R}}$ from terms to terms by means of Rule (1.8a), then $\widetilde{\mathcal{R}}$ does not enjoy strong confluence. Observe that $\mathcal{R}$ may be even a strongly confluent relation.

More precisely, the latter depends on the algebraic properties of R: if $a \in \mathsf{R}$ and $a = b + c$ for some $b, c \in \mathsf{R}$, then reduction $\widetilde{\mathcal{R}}$ does not exhibit strong confluence. The following is an example showing such behaviour.

***Example* 1.4.23.** Let us consider the module of terms $\mathbb{N}\langle\Delta_{\mathbb{N}}\rangle$ and $a, b, c \in \mathbb{N}$ such that $a = b + c$. Moreover, consider the simple term $s = ((\lambda xy.x)\, z)\, (\mathbf{I})\, \mathbf{I}$ (where $\mathbf{I} = \lambda x.x$) and verify that $s \to t$ and $s \to u$, with $t = (\lambda y.z)\, (\mathbf{I})\, \mathbf{I}$ and $u = ((\lambda xy.x)\, z)\, \mathbf{I}$. Then, there is $v = (\lambda y.z)\, \mathbf{I}$ such that $t \to v$ and $u \to v$. That is, $\to$ is strongly confluent on s.

Nonetheless, $\widetilde{\to}$ is not strongly confluent on the algebraic term $as \in \mathbb{N}\langle\Delta_{\mathbb{N}}\rangle$: suppose $as \mathrel{\widetilde{\to}} bs + ct$ and $as \mathrel{\widetilde{\to}} au$, then it is simple to verify that the two join with $bu + cv$ as $bs + ct \mathrel{\widetilde{\to}^*} bu + cv$ and $au \mathrel{\widetilde{\to}} bu + cv$. In particular, observe that the second to last reduction is expressed in terms of $\widetilde{\to}^*$.

Generally speaking, whenever an element of R can be expressed by a linear combination of other elements, Rule (1.8a) causes another form of *duplication* of terms (which we may name *algebraic*), hence of redexes. Notice indeed that, in Example 1.4.23, reduction $as \mathrel{\widetilde{\to}} bs + ct$ is due to the fact that as is considered as $bs + cs$ when applying Rule (1.8a). This latter can be considered as a side effect as it is somehow silent with respect to reduction.

Algebraic duplication does not affect relations $\overline{\mathcal{R}}$ from terms to terms defined by means of Rule (1.8b): indeed, this latter allows multiple simple terms of a sum to be reduced simultaneously, no matter how the algebraic properties of the calculus are exploited.

In what follows, we show the Church-Rosser property for reduction $\widetilde{\to}$ by adapting to Λ_Σ a well-known technique in λ-calculus due to Tait and Martin-Löf [Bar84]. This techinique relies on the notion of parallel reduction in order to handle duplication, here algebraic duplication too. The key technical point lies in proving the equivalence between the reflexive and transitive closure of both parallel β-reduction and (one-step) β-reduction. Then, by proving strong confluence for the former, one achieves confluence for the latter as a direct consequence.

Relating reduction to its parallel version

The first step of the Tait–Martin-Löf technique consists in showing that parallel β-reduction can be put in the following relation with β-reduction and this latter reflexive, transitive closure.

Lemma 1.4.24. *It holds that* $\widetilde{\to} \subset \overline{\rightrightarrows} \subset \widetilde{\to}^*$.

Proof. The first inclusion, namely $\widetilde{\to} \subset \overline{\rightrightarrows}$, holds by definitions of $\widetilde{\to}$ (Definition 1.4.6 and Rule (1.8a)) and $\overline{\rightrightarrows}$ (Definition 1.4.12 and Rule (1.8b)), respectively. It is a strict inclusion since $(\mathbf{I})\,(\mathbf{I})\,\mathbf{I} \overline{\rightrightarrows} \mathbf{I}$ whereas $(\mathbf{I})\,(\mathbf{I})\,\mathbf{I} \not\rightrightarrows \mathbf{I}$.

We prove the second inclusion, namely $\overline{\rightrightarrows} \subset \widetilde{\rightarrow}^*$, by induction on k that $\overline{\rightrightarrows_k} \subset \widetilde{\rightarrow}^*$. If $k = 0$, the result follows by the fact that $\widetilde{\rightarrow}^*$ is reflexive. Suppose the result holds for some k so that $S \overline{\rightrightarrows_k} S'$, then we extend it to $k+1$ by inspecting the possible cases for reduction $S \overline{\rightrightarrows_{k+1}} S'$. We first address the cases in which S is a simple term s and $s \rightrightarrows_{k+1} S'$. Then one of the following applies:

- $s \in \mathcal{V}$; hence the result follows by the fact that $\widetilde{\rightarrow}^*$ is reflexive.
- $s = \lambda x.u$ and $S' = \lambda x.U'$ with $u \rightrightarrows_k U'$; hence, by the induction hypothesis, $u \widetilde{\rightarrow}^* U'$. By the contextual property of $\widetilde{\rightarrow}^*$ (Proposition 1.4.11), it follows
$$s = \lambda x.u \widetilde{\rightarrow}^* \lambda x.U' = S'.$$
- $s = (u)\,V$ and $S' = (U')\,V'$ with $u \rightrightarrows_k U'$ and $V \overline{\rightrightarrows_k} V'$; hence, by the induction hypothesis, $u \widetilde{\rightarrow}^* U'$ and $V \widetilde{\rightarrow}^* V'$. By the contextual property of $\widetilde{\rightarrow}^*$ (Proposition 1.4.11), it follows
$$s = (u)\,V \widetilde{\rightarrow}^* \left(U'\right) V' = S'.$$
- $s = (\lambda x.u)\,V$ and $S' = U'\,[V'/x]$ with $u \rightrightarrows_k U'$ and $V \overline{\rightrightarrows_k} V'$; hence, by the induction hypothesis, $u \widetilde{\rightarrow}^* U'$ and $V \widetilde{\rightarrow}^* V'$. By the contextual property of $\widetilde{\rightarrow}^*$ (Proposition 1.4.11) and algebraic linearity (Definition 1.2.2), it follows
$$s = (\lambda x.u)\,V \widetilde{\rightarrow}^* \left(\lambda x.U'\right) V' \widetilde{\rightarrow}^* U' \left[V'/x\right] = S'.$$

Now assume $S \overline{\rightrightarrows_{k+1}} S'$. By Rule (1.8b), this amount to the following: $S = \sum_{i=1}^n a_i u_i$ and $S' = \sum_{i=1}^n a_i U_i'$ with $u_i \rightrightarrows_{k+1} U_i'$ for all $i \in \{1, \dots, n\}$. From what we have just shown in the case of simple terms follows $u_i \widetilde{\rightarrow}^* U_i'$ for all $i \in \{1, \dots, n\}$. By the contextual property of $\widetilde{\rightarrow}^*$ (Proposition 1.4.11), it follows
$$S = \sum_{i=1}^n a_i u_i \widetilde{\rightarrow}^* \sum_{i=1}^n a_i U_i' = S'.$$
Also this second inclusion is strict since $((\mathbf{I})\,\mathbf{I})\,\mathbf{I} \widetilde{\rightarrow}^* \mathbf{I}$ whereas $((\mathbf{I})\,\mathbf{I})\,\mathbf{I} \not\overline{\rightrightarrows} \mathbf{I}$. $\square$

Corollary 1.4.25. *It holds that* $\overline{\rightrightarrows}^* = \widetilde{\rightarrow}^*$.

Proof. Straightforward consequence of Lemma 1.4.24 and the fact that the operator $(\cdot)^*$ is monotone and idempotent (*i.e.* for every binary relation $\mathcal{R}$, $(\mathcal{R}^*)^* = \mathcal{R}^*$). $\square$

By Corollary 1.4.25, Lemma 1.4.20 directly entails the following result.

Lemma 1.4.26. *Let* $x \in \mathcal{V}$ *and* $S, S', T, T' \in \mathsf{R}\langle\Delta_\mathsf{R}\rangle$. *If* $S \widetilde{\rightarrow}^* S'$ *and* $T \widetilde{\rightarrow}^* T'$, *then* $S\,[T/x] \widetilde{\rightarrow}^* S'\,[T'/x]$.

Strong confluence for $\overline{\rightrightarrows}$, confluence for $\widetilde{\rightarrow}$

The second step of the Tait–Martin-Löf technique consists in showing strong confluence for reduction $\overline{\rightrightarrows}$. This is sufficient since Corollary 1.4.25 would then entail strong confluence for $\widetilde{\rightarrow}^*$, and so the Church-Rosser property of $\widetilde{\rightarrow}$.

In order to do so, we characterise parallel reduction as the process of simultaneously firing all the redexes of a term.

Definition 1.4.27. *Let $S^{\circledast}$ be the* complete development of *S inductively defined by:*

$$\begin{aligned}
x^{\circledast} &= x; \\
(\lambda x.u)^{\circledast} &= \lambda x.u^{\circledast}; \\
((u)\,V)^{\circledast} &= \left(u^{\circledast}\right) V^{\circledast}; \ \text{(if s is not a λ-abstraction)} \\
((\lambda x.u)\,V)^{\circledast} &= u^{\circledast}\left[V^{\circledast}/x\right]; \\
\left(\sum_{i=1}^{n} a_i u_i\right)^{\circledast} &= \sum_{i=1}^{n} a_i u_i^{\circledast}.
\end{aligned}$$

In particular, one can prove that every $\overline{\rightrightarrows}$-reduct of a term reduces to its complete development.

Lemma 1.4.28. *For all $S, S' \in \mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$, if $S \,\overline{\rightrightarrows}\, S'$ then $S' \,\overline{\rightrightarrows}\, S^{\circledast}$.*

Proof. Simple induction on k that $S \,\overline{\rightrightarrows_k}\, S'$ implies $S' \,\overline{\rightrightarrows}\, S^{\circledast}$, using the contextual property of $\overline{\rightrightarrows}$ (Proposition 1.4.18) and Lemma 1.4.20 in the redex case. ☐

Theorem 1.4.29. *Relation $\overline{\rightrightarrows}$ is strongly confluent. Hence, relation $\widetilde{\rightarrow}$ enjoys Church-Rosser property.*

Proof. Relation $\overline{\rightrightarrows}$ is strongly confluent as a direct consequence of Lemma 1.4.28. Then, Corollary 1.4.25 implies confluence of relation $\widetilde{\rightarrow}$. ☐

***Remark* 1.4.30.** Notice that Theorem 1.4.29 establishes confluence for reduction $\widetilde{\rightarrow}$ without assuming any particular property about R.

Although such a result suggests that the Church-Rosser property is not affected by the algebraic component of the calculus, we can easily identify a case in which the meaning of Theorem 1.4.29 gets vacuous. Indeed, let us consider the case of a semiring R such that $-1 \in \mathsf{R}$ (*i.e.* $1 + (-1) = 0$), and assume $S \,\widetilde{\rightarrow}^*\, T$ for some $S, T \in \mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$. By the contextual property of $\widetilde{\rightarrow}^*$ (Proposition 1.4.11) follows

$$T = T + (-1)S + S \,\widetilde{\rightarrow}^*\, T + (-1)T + S = S,$$

which proves $\widetilde{\rightarrow}^*$ of being symmetric. Hence Church-Rosser trivially holds.

This previous remark highlights some serious issues about the $\widetilde{\rightarrow}$ rewriting theory of the algebraic λ-calculus, especially as far as normalisation properties are concerned. In Chapter 2, we analyse the details of this problem.

Chapter 2

Normal forms for the algebraic λ-calculus

In this chapter we present the first contributions of this thesis.

We start off by recalling the well-known issues of *normalisability* in the algebraic λ-calculus: in presence of negative coefficients, no term has a normal form [ER03] and term equivalence collapses [Vau09]. We then provide a full development of Ehrhard and Regnier's idea of *weak normalisation* that allows to study the reduction behaviour of terms in a particular module of terms where β-reduction is sound.

Finally, we show that the established notion of (unique) normal form is attained by a parallel variant of β-reduction, defined on canonical terms only.

Contents

2.1 Collapse **41**
2.1.1 Term equivalence 42
2.2 Algebraic properties and Normalisability **42**
2.3 (Unique) Normal forms **45**
2.3.1 Preliminaries 45
2.3.2 Strong normalisability 52
2.3.3 Normalisability 56
2.3.4 Unique normal forms for algebraic terms 62
2.3.5 (Consistent) Term equivalence 62
2.4 Compute normal forms **63**
2.4.1 Canonical relations from terms to terms 63
2.4.2 Canonical parallel β-reduction 64
2.4.3 Canonical β-reduction 71
2.4.4 Canonical means defectiveness 77

The first section reports on the collapse of β-reduction on $\mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$, hence of term equivalence, in presence of negative coefficients in R.

The second section briefly highlights some aspects about the interplay between β-reduction and algebraic rewriting. If it sufficient to consider positive semirings P for guaranteeing term normalisability, strong normalisability needs additional hypothesis on P.

The third section presents the first contribution of the chapter. An untyped weak normalisation scheme is developed in the spirit of Ehrhard and Regnier's idea. The two cases of strongly normalisable and (just) normalisable terms are detailed, as two different rewriting techniques are put into use. Nonetheless, the two share the common idea of studying normalisability properties in the module of terms over the semiring of polynomials with non-negative integer coefficients, namely in a setting where reduction is sound. The induced partial term equivalence is proved to be consistent.

Finally, a notion of canonical reduction relation from terms to terms is investigated in the fourth section. The parallel version is proved to precisely characterise the previously established notion of normal form. However, it is not clear if reduction does. As a matter of fact, it turns out that restricting reduction relations on canonical terms only, causes the lack of some fundamental properties typically needed for developing the reduction theory of a calculus.

2.1 Collapse

Remark 1.4.30 has given a sample of the issues concerning reduction $\widetilde{\rightarrow}$. In this section we report on this matter in a detailed way, discussing especially about the implications on the notion of normal form in Λ_Σ.

We already know that reduction $\widetilde{\rightarrow}^*$ turns into a symmetric relation whenever R is not positive. More strikingly, in such setting there is no *irreducible* term.

Lemma 2.1.1. *Whenever* R *is not positive,* i.e. *there exist* $a, b \in \mathsf{R}^\bullet$ *such that* $a + b = 0$, *then for all* $S \in \mathsf{R}\langle\Delta_\mathsf{R}\rangle$, S *reduces.*

Proof. Consider any $t \in \Delta_\mathsf{R}$ and $T' \in \mathsf{R}\langle\Delta_\mathsf{R}\rangle$ such that $t \to T'$. Then, for every $S \in \mathsf{R}\langle\Delta_\mathsf{R}\rangle$:

$$S = S + at + bt \mathrel{\widetilde{\rightarrow}} S + aT' + bt,$$

which is the thesis. □

Related to the latter, the non-positivity of R causes the *collapse* of $\widetilde{\rightarrow}^*$:

Proposition 2.1.2. *Whenever* R *is not positive,* i.e. *there exist* $a, b \in \mathsf{R}^\bullet$ *such that* $a + b = 0$, *then the reduction* $\mathbf{0} \mathrel{\widetilde{\rightarrow}^*} aS \mathrel{\widetilde{\rightarrow}^*} \mathbf{0}$ *holds.*

Proof. Consider any fixpoint operator Θ of the pure λ-calculus, and define the term $\infty_S = (\Theta)\,\lambda x.(S + x)$. It is simple to verify that ∞_S admits the reduction $\infty_S \mathrel{\widetilde{\rightarrow}^*} S + \infty_S$. Then, it follows

$$\mathbf{0} = a\infty_S + b\infty_S \mathrel{\widetilde{\rightarrow}^*} aS + a\infty_S + b\infty_S = aS$$

and

$$aS = aS + a\infty_S + b\infty_S \mathrel{\widetilde{\rightarrow}^*} aS + a\infty_S + bS + b\infty_S = \mathbf{0},$$

which is the thesis. □

In particular, the fact that the term $\mathbf{0}$ may reduce, and most of all it may reduce to whichever other term, highlights the crucial problem with non-positive semirings: the non-positivity property opens up to the algebraic decomposition of the additive identity of the semiring, which in the end permits to rewrite every term into a different, yet equivalent, term. This obviously leads to inconsistency.

Corollary 2.1.3. *If* R *is such that* 1 *has an opposite,* i.e. $(-1) \in \mathsf{R}$ *with* $1 + (-1) = 0$, *then for all* $S, T \in \mathsf{R}\langle\Delta_\mathsf{R}\rangle$, $S \mathrel{\widetilde{\rightarrow}^*} T$.

Proof. Consider again the term ∞_S, which admits the reduction $\infty_S \widetilde{\to}^* S + \infty_S$. Then, it follows

$$
\begin{aligned}
S &= S + \infty_S - \infty_S + \infty_T - \infty_T \\
&\widetilde{\to} S + \infty_S - S - \infty_S + \infty_T - \infty_T \\
&\widetilde{\to} S + \infty_S - S - \infty_S + T + \infty_T - \infty_T \\
&= S - S + T = T,
\end{aligned}
$$

which is the thesis. □

2.1.1 Term equivalence

Given a reduction notion on terms, it is customary to examine the induced *term equivalence relation*:

Definition 2.1.4. *Let* $\cong \subseteq \mathsf{R}\langle\Delta_{\mathsf{R}}\rangle \times \mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$ *be the contextual equivalence relation defined as the reflexive, symmetric and transitive closure of reduction* $\widetilde{\to}$.

Of particular interest is the behaviour of $\cong$ in the case of non-positive semirings. The prototypical example is the set of $\mathbb{Z}$ of integer equipped the usual operations. It turns out that, whenever the R is not positive, the collapse of reduction $\widetilde{\to}^*$ (Proposition 2.1.2) implies the unsurprising *inconsistency* result that follows.

Corollary 2.1.5. *Let* R *be non-positive. For all* $S, T \in \mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$, $S \cong T$ *holds.*

In other words, $\cong$ identifies terms which have nothing to do with each other.

Again, this inconsistency result is caused only by the possibility of **0** to reduce when expressed as $a\infty_S + b\infty_S$ (for every $S \in \mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$), which ultimately boils down to non-positive R and the definition of β-reduction as relation from terms to terms by means of Rule (1.8a).

2.2 Algebraic properties and Normalisability

Λ_Σ exhibits terms which do not normalise. As far as reduction is concerned, this is not unexpected given that every ordinary λ-term is also a simple term of the algebraic λ-calculus.

On the other hand, we are more interested in understanding how the algebraic component of Λ_Σ, namely the properties of the semiring R from which coefficients are taken, influences its normalisability properties (*w.r.t.* reduction $\widetilde{\to}$), as well as the related notions of normalisable terms and normal forms. This is not a trivial question, in general [Vau09].

Generally speaking, we can rapidly realise that usual pure λ-calculus notions about the normalisability properties of terms do not clearly adapt in the current setting. For instance, we have already established that non-positive semirings cause the loss of the notion of normal forms intended as irreducible terms with respect to a confluent reduction relation. In particular, Lemma 2.1.1 directly implies the following:

Proposition 2.2.1. *Whenever* R *is not positive,* $\mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$ *does not exhibit normal forms.*

Less obvious is the case of positive semirings P: on the one hand, since there is no way 0 can be decomposed as a linear combination of other elements of P, it follows that the notion of normalisable terms recovers the usual meaning. The module $\mathbb{N}\langle\Delta_{\mathbb{N}}\rangle$ is the example *par excellence*:

Proposition 2.2.2. *The module* $\mathbb{N}\langle\Delta_{\mathbb{N}}\rangle$ *exhibits (unique) normal forms and it is conservative with respect to pure* λ*-calculus.*

Proof. This is basically due to the essential property of $\mathbb{N}$ which allows only finetely many ways of writing a non-negative integer as a sum of positive integers. The original paper provide the details [Vau09]. $\square$

On the other hand, *strong normalisability* is a delicate matter in Λ_Σ. In pure λ-calculus, a term enjoys strong normalisability whenever every reduction sequence strarting from it eventually terminates (*i.e.* it is finite).

The question is more brittle in Λ_Σ as strongly normalisable terms cannot be simply defined as the set of linear combinations of strongly normalisable simple terms. Indeed, even if R is positive, it may be the case that the only terms always exhibiting finite sequence of reductions are the normal ones.

***Example* 2.2.3** ([ER03])**.** Assume R to be the positive rig $\mathbb{Q}^+$ of non-negative rational numbers, and let $s, s' \in \Delta_{\mathbb{Q}^+}$ such that $s \to s'$. Then, it rather simple to devise an infinite sequence of reductions from s:

$$\begin{aligned} s = \frac{1}{2}s + \frac{1}{2}s &\mathrel{\widetilde{\rightrightarrows}} \frac{1}{2}s + \frac{1}{2}s' = \frac{1}{4}s + \frac{1}{4}s + \frac{1}{2}s' \\ &\mathrel{\widetilde{\rightrightarrows}} \frac{1}{4}s + \frac{1}{4}s' + \frac{1}{2}s' = \frac{1}{4}s + \frac{3}{4}s' = \dots \\ &\dots \\ &\mathrel{\widetilde{\rightrightarrows}} \frac{1}{2^n}s + \frac{2^n - 1}{2^n}s' = \dots \end{aligned}$$

Notice that such an infinite sequence of reduction is once again due to the definition of $\widetilde{\rightrightarrows}$ on linear combinations of terms, and the possibility of each non-zero rational number to be expressed as an infinite sum of non-zero rational numbers. In other words, the above argument is still valid even when considering s as a strongly normalisable pure λ-term.

We recover the usual notion of strong normalisability as soon as we impose the semiring R to be *finitely splitting* and *integral domain*. Intuitively, the former property ensures that each element of R can only be written as a finite sum of elements of $\mathsf{R}^\bullet$. The latter property guarantees that multiplying elements of R results in the zero element of R only in the case the zero element is itself one of the factors. These two conditions on R efficiently prevents tricky situations which involve coefficients manipulations as in Example 2.2.3.

Definition 2.2.4. *A semiring* R *is said to be* finitely splitting *whenever, for all* $a \in \mathsf{R}$*, the set* $\{(a_1, \ldots, a_n) \in (\mathsf{R}^\bullet)^n \mid a = a_1 + \cdots + a_n\}$ *is finite.*

An obvious example of finitely splitting semiring is $\mathbb{N}$.

Lemma 2.2.5. *If* R *is finitely splitting, then* R *is positive.*

Proof. Assume R finitely splitting and consider the element $0 \in \mathsf{R}$. Since the latter is the neutral element of addition in R, the empty tuple is the only element of the set $\{(a_1, \ldots, a_n) \in (\mathsf{R}^\bullet)^n \mid 0 = a_1 + \cdots + a_n\}$. It follows that R is positive. $\square$

Let us consider the following example to shed light on the reason why the integral domain is a necessary condition to characterise strongly normalisable algebraic terms as linear combinations of strongly normalisable simple terms.

***Example* 2.2.6** ([Vau09]). Consider $\mathsf{R} = \mathbb{N} \times \mathbb{N}$ equipped with operations defined pointwise: *i.e.* $(p, q) + (p', q') = (p + p', q + q')$ and $(p, q)(p', q') = (pp', qq')$. This semiring can be proved to be finitely splitting, hence positive (Lemma 2.2.5).

Now consider the elements $a = (1, 0)$ and $b = (0, 1)$, and check that $a + b = (1, 1)$ and $ab = (0, 0)$, respectively the 1 and 0 of R. Moreover, let $\delta = \lambda x.(x)\, x$ and consider the simple term $(\delta)\, b\delta$. This latter is not normalisable, let alone strongly normalisable. Nonetheless, the term $a\, (\delta)\, b\delta$ always reduces to $\mathbf{0}$, which is normal in this precise setting.

Definition 2.2.7. *A semiring* R *is said to be an* integral domain *whenever, for all* $a, b \in \mathsf{R}$*,* $ab = 0$ *implies either* $a = 0$ *or* $b = 0$*.*

Under these conditions, Vaux [Vau09] has indeed proved that a term $S \in \mathsf{R}\langle \Delta_{\mathsf{R}} \rangle$ is strongly normalisable if and only if every simple term in $\mathsf{Supp}(S)$ is as well. This is sufficient to recover to the classical notion of strong normalisability.

Theorem 2.2.8. *Whenever* R *is finitely splitting and integral domain, the set of strongly normalisable terms is that of the linear combinations of strongly normalisable simple terms.*

The semiring $\mathbb{N}$ is also an integral domain. Indeed, it was the case study of Ehrhard and Regnier when dealing with normalisability properties of the differential λ-calculus [ER03]. In the upcoming section, we exploit the algebraic properties of $\mathbb{N}$ while considering the more interesting semiring $\mathbb{P}$ of polynomials with non-negative integer coefficients over a set of indeterminates.

2.3 (Unique) Normal forms

In the previous section we have set the non-triviality of establishing the set of normalisable terms in the general case of non-positive semirings. In this section we propose a solution inspired by an argument first mentioned by Ehrhard and Regnier [ER03], and later investigated by Vaux [Vau09] under the name of *weak normalisability*. Contrary to the latter where a typed scenario has been considered, our development concerns the untyped version of the algebraic λ-calculus.

We present this contribution by distinguishing two cases: we first consider strong normalisability property, and later we generalise to just normalisability. The former part is published [Alb13], whereas the latter is not yet.

2.3.1 Preliminaries

The purpose of the constructions that follows is to identify in $\mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$ those terms which can be considered as normal when dealing with non-positive semirings R. In view of the issues discussed in Sections 2.1 and 2.2, this obviously needs a slightly different formulation of the notion of normal form and a method to establish the normal form of a term in a system where reduction may be inconsistent.

We now present the idea behind the constructions we provide in Sections 2.3.2 and 2.3.3, along with some preliminary definitions and results, which are based on the intuition that a normal form is a term admitting a redex-free writing.

General idea

The solution we propose exploits the algebraic properties of the semiring $\mathbb{N}[\Xi]$ of polynomials with non-negative integer coefficients over a set Ξ of indeterminates, in order to generalise the usual notions concerning normalisability to whichever module of terms $\mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$.

We highlight, in particular, two properties that $\mathbb{N}[\Xi]$ enjoys:

- As the natural generalisation of $\mathbb{N}$, $\mathbb{N}[\Xi]$ is a finitely splitting and integral domain semiring, so that the module of terms $\mathbb{N}[\Xi]\langle\Delta_{\mathbb{N}[\Xi]}\rangle$ enjoys classical and well-understood notions about normalisability. For instance, it is clear that $\mathbb{N}[\Xi]\langle\Delta_{\mathbb{N}[\Xi]}\rangle$ exhibits unique normal forms;

- As soon as a relation $[\![\cdot]\!]$ between R and Ξ is established, elements of R can be expressed by elements in $\mathbb{N}[\Xi]$.

The latter relation extends to terms of $\mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$ effortlessly, that is every $S \in \mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$ can be related to some term $S_{\Xi} \in \mathbb{N}[\Xi]\langle\Delta_{\mathbb{N}[\Xi]}\rangle$ such that $[\![S_{\Xi}]\!] = S$. Then, the normalisability properties of the latter defines those of the former. Intuitively, this can be accomplished by the following three steps process:

1. A term $S \in \mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$ is turned into $S_{\Xi} \in \mathbb{N}[\Xi]\langle\Delta_{\mathbb{N}[\Xi]}\rangle$ by replacing the coefficients $a_1, \ldots, a_n \in \mathsf{R}$ appearing in it with the corresponding formal indeterminates $\mathsf{X}_1, \ldots, \mathsf{X}_n \in \Xi$;

2. In the module of terms $\mathbb{N}[\Xi]\langle\Delta_{\mathbb{N}[\Xi]}\rangle$, S_{Ξ} normalises to its (unique) normal form $\mathsf{NF}(S_{\Xi})$, if any;

3. By restoring back each indeterminates X_i in $\mathsf{NF}(S_{\Xi})$ to its original value a_i, and by evaluating in R the algebraic expressions therefore obtained, it results in a term of $\mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$ that we determine to be the (unique) normal form of S, hence written $\mathsf{NF}(S)$.

***Example* 2.3.1.** Let us consider the module of terms $\mathbb{Z}\langle\Delta_{\mathbb{Z}}\rangle$. Now suppose $[\![\mathsf{X}]\!] = -1$, $[\![\mathsf{Y}]\!] = 1$, $[\![\mathsf{Z}]\!] = 2$ and consider the terms $-(\mathbf{I})\,2\mathbf{I}, -(\mathbf{I})\,(\mathbf{I}+\mathbf{I}) \in \mathbb{Z}\langle\Delta_{\mathbb{Z}}\rangle$. The two are the same algebraic term in $\mathbb{Z}\langle\Delta_{\mathbb{Z}}\rangle$.

According to the 1st step, their respective terms in $\mathbb{N}[\Xi]\langle\Delta_{\mathbb{N}[\Xi]}\rangle$ are $\mathsf{X}\,(\mathbf{I})\,\mathsf{Z}\mathbf{I}$ and $\mathsf{X}\,(\mathbf{I})\,(\mathsf{Y}\mathbf{I}+\mathsf{Y}\mathbf{I})$, which normalise to $\mathsf{XZ}\mathbf{I}$ and $2\mathsf{XY}\mathbf{I}$ respectively. Then, the 3rd step entails $\mathsf{NF}(\mathsf{X}\,(\mathbf{I})\,\mathsf{Z}\mathbf{I}) = -2\mathbf{I}$ (*i.e.* $[\![\mathsf{XZ}\mathbf{I}]\!]$) and $\mathsf{NF}(\mathsf{X}\,(\mathbf{I})\,(\mathsf{Y}\mathbf{I}+\mathsf{Y}\mathbf{I})) = -2\mathbf{I}$ (*i.e.* $[\![2\mathsf{XY}\mathbf{I}]\!]$). Notice that, although $\mathsf{XZ}\mathbf{I} \neq 2\mathsf{XY}\mathbf{I}$, the normal forms we get at the end are the same term $-2\mathbf{I} \in \mathbb{Z}\langle\Delta_{\mathbb{Z}}\rangle$.

It is straightforward to see that, as soon as a $[\![\cdot]\!]$ is defined, there are infinite different terms of $\mathbb{N}[\Xi]\langle\Delta_{\mathbb{N}[\Xi]}\rangle$ matching the same term of $\mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$ by means of $[\![\cdot]\!]$ (in particular, whenever R is not positive). This pushes us to dispense with the above concrete process entirely, in favour of a more abstract construction.

***Remark* 2.3.2.** Example 2.3.1 highlights a crucial point we must settle in order to claim the soundness of our solution, namely the guarantee that different terms of $\mathbb{N}[\Xi]\langle\Delta_{\mathbb{N}[\Xi]}\rangle$, standing for the same term of $\mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$, provide the same term at the end. On the contrary, we could not set the latter to be a normal form of $\mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$.

In the following we precise how the previous process can be formalised and why the result can be considered a notion of normal form for the algebraic λ-calculus.

Pre-normal terms

In general, the module of terms $\mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$ over non-positive R does not exhibit irreducible terms as we can ultimately exploit the decomposition of the zero term into an equivalent linear combination of terms. In other words, from the point of view of the dynamics of the calculus, we cannot consider any term as an actual redex-free term.

However, we can still define the set of terms that admit a redex-free writing. This is the notion we refer to when talking about normal forms in Λ_{Σ}.

Definition 2.3.3 ([Vau09]). *We define* pre-normal terms *and* pre-neutral terms *by mutual induction as follows:*

- *$S \in \Delta_{\mathsf{R}}$ is a pre-neutral term whenever $S \in \mathcal{V}$, or $S = (u)\,V$ where u is a pre-neutral term and V is a pre-normal term;*
- *$S \in \Delta_{\mathsf{R}}$ is a simple pre-normal term whenever S is a pre-neutral term, or $S = \lambda x.u$ where u is a simple pre-normal term;*
- *S is a pre-normal term whenever, for all $u \in \mathsf{Supp}(S)$, u is a simple pre-normal term.*

Observe that pre-normal terms are those terms whose canonical form is redex-free.

The following proposition justifies our intent of considering pre-normal terms as the normal terms of the algebraic λ-calculus.

Proposition 2.3.4. *If R is positive, then pre-normal terms are the normal terms.*

Proof. Direct consequence of Definition 2.3.3 and the positivity of R. □

A semiring of polynomials: $\mathbb{P}$

The idea previously described prevents the collapse of reduction by somehow forcing non-positive semirings R to exhibit the properties of $\mathbb{N}$. In doing so we do not restrict algebraic manipulations completely, but we tame them to act as in $\mathbb{N}$, namely an harmless setting with respect to $\widetilde{\to}$ normalisability properties.

We now formally introduce the semiring $\mathbb{P}$ of polynomials with non-negative integer coefficients over a set Ξ of indeterminates, and we provide the necessary machinery to relate terms in the module $\mathbb{P}\langle\Delta_{\mathbb{P}}\rangle$ with the corresponding ones in the module $\mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$. In particular, this only needs a morphism between Ξ and R.

Remark **2.3.5.** We consider $\mathbb{P}$ as $\mathbb{N}[\Xi]$ because this latter exhibits algebraic properties that guarantee the classical notions about normalisable terms to hold, even in the strong sense. As a matter of fact, all other examples we are aware of having the same properties are given by variants of $\mathbb{N}[\Xi]$ [Vau09].

Definition 2.3.6. *Let* $\Xi = \{\mathsf{X}, \mathsf{Y}, \mathsf{Z}, \ldots\}$ *be a set of* indeterminates*. We define* $\mathbb{P} = \mathbb{N}[\Xi]$ *the semiring of polynomials with non-negative integer coefficients over indeterminates in* Ξ*.*

Definition 2.3.7. *An* indeterminate assignment *is any total function* $f : \Xi \to \mathsf{R}$.

The following definitions are intended as parametrised over an indeterminate assignment f. Then, given such a function f, we naturally extends it to a morphism evaluating polynomials into elements of R.

Definition 2.3.8. Polynomial evaluation *is the semiring morphism* $[\![\cdot]\!]_f : \mathbb{P} \to \mathsf{R}$ *returning the value in* R *of a given polynomial in* $\mathbb{P}$*:* i.e. *if* $P \in \mathbb{P}$*, then* $[\![P]\!]_f$ *is its* value *calculated in* R *once each indeterminate* X *in* P *has been replaced with its assignment* $f(\mathsf{X})$ *in* R*.*

We extend the above evaluation to the module of terms $\mathbb{P}\langle\Delta_{\mathbb{P}}\rangle$ as the morphism returning for every term in $\mathbb{P}\langle\Delta_{\mathbb{P}}\rangle$ the corresponding term in $\mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$ by replacing each polynomial coefficient with its value.

Definition 2.3.9. Term evaluation *is the module morphism* $[\![\cdot]\!]_f : \mathbb{P}\langle\Delta_{\mathbb{P}}\rangle \to \mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$ *defined by induction on terms as follows:*

$$\begin{aligned}
[\![x]\!]_f &= x; \\
[\![\lambda x.u]\!]_f &= \lambda x.[\![u]\!]_f; \\
[\![(u)\, V]\!]_f &= \left([\![u]\!]_f\right) [\![V]\!]_f; \\
\left[\!\!\left[\sum_{i=1}^{n} P_i u_i\right]\!\!\right]_f &= \sum_{i=1}^{n} [\![P_i]\!]_f\, [\![u_i]\!]_f.
\end{aligned}$$

The above definition induces an obvious equivalence relation on $\mathbb{P}\langle\Delta_{\mathbb{P}}\rangle$:

Definition 2.3.10. *For all* $S, T \in \mathbb{P}\langle\Delta_{\mathbb{P}}\rangle$*,* $S \stackrel{\nabla}{=}_f T$ *whenever* $[\![S]\!]_f = [\![T]\!]_f$*.*

Moreover, along with a term evaluation, it proves useful a way to associate a term in $\mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$ with the subset of terms in $\mathbb{P}\langle\Delta_{\mathbb{P}}\rangle$ evaluating to it.

Definition 2.3.11. *For all* $S \in \mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$*, a* notation for S *is any term* $T \in \mathbb{P}\langle\Delta_{\mathbb{P}}\rangle$ *such that* $[\![T]\!]_f = S$*. The* set of all notations for S *is indicated as* $\langle S\rangle_f$*.*

Notice that, in general, there are more than one notation for a same term in $\mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$. In particular, whenever R is not finitely splitting, then a term in $\mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$ can be associated to an infinite set of notations in $\mathbb{P}\langle\Delta_{\mathbb{P}}\rangle$ evaluating to it. This is not an issue as long as we are able to show that all notations for that term ultimately provide the same normal form.

Definition 2.3.12. *Terms $S, T \in \mathbb{P}\langle\Delta_{\mathbb{P}}\rangle$ such that $S \stackrel{\nabla}{=}_f T$ are called* sibling terms*. For all $S \in \mathbb{P}\langle\Delta_{\mathbb{P}}\rangle$, the* set of sibling terms of *S is defined as follows:*

$$\nabla_f(S) = \{T \in \mathbb{P}\langle\Delta_{\mathbb{P}}\rangle \mid S \stackrel{\nabla}{=}_f T\}.$$

Definition 2.3.13. *A set $\mathcal{S} \subseteq \Delta_{\mathbb{P}}$ is said* uniform *whenever, for all $s, t \in \mathcal{S}$, $s \stackrel{\nabla}{=}_f t$. Two sets $\mathcal{S}, \mathcal{T} \subseteq \Delta_{\mathbb{P}}$ are said* disjoint *whenever, for all $s \in \mathcal{S}$ and $t \in \mathcal{T}$, $s \stackrel{\nabla}{\neq}_f t$.*

In the following we suppose to work with one particular variable assignment. Therefore, we exempt ourselves from precising the variable assignment as subscript in the just introduced symbolism (*e.g.* $[\![\cdot]\!]$ denotes the term evaluation morphism).

The notions introduced by Definition 2.3.11 and Definition 2.3.12 are related in the sense that the siblings of a term $S \in \mathbb{P}\langle\Delta_{\mathbb{P}}\rangle$ are different notations for its term evaluation $[\![S]\!] \in \mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$:

Lemma 2.3.14. *For all $S \in \mathbb{P}\langle\Delta_{\mathbb{P}}\rangle$, it holds that $\nabla(S) = \langle[\![S]\!]\rangle$.*

Proof. Straightforward using Definitions 2.3.9 – 2.3.12. □

Notation. Throughout this section we often write linear combinations of simple terms as $\sum_{u\in\Delta_{\mathsf{R}}}\sum_{i\in I_u} P_i s_i$ suggesting that, if not specified, $s_i \in \langle u\rangle$ for all $i \in I_u$. Note how simple terms $u \in \Delta_{\mathsf{R}}$ are used to index sets I_u of indices of sibling terms. We also write $\sum_{i\in I} P_i s_i + \sum_{j\in J} P_j s_j$ whenever we need to reason on the sets $\{s_i\}_{i\in I}$ and $\{s_j\}_{j\in J}$ of simple terms indexed by elements of I and J, respectively.

Lemma 2.3.15. *Let $S, T \in \mathbb{P}\langle\Delta_{\mathbb{P}}\rangle$ be respectively written as $S = \sum_{u\in\Delta_{\mathsf{R}}}\sum_{i\in I_u} P_i s_i$, with $s_i \in \langle u\rangle$ for all $i \in I_u$, and $T = \sum_{u\in\Delta_{\mathsf{R}}}\sum_{j\in J_u} Q_j t_j$, with $t_j \in \langle u\rangle$ for all $j \in J_u$. Then,*

$$S \stackrel{\nabla}{=} T \text{ if and only if } \left[\!\!\left[\sum_{i\in I_u} P_i\right]\!\!\right] = \left[\!\!\left[\sum_{j\in J_u} Q_j\right]\!\!\right] \text{ for all } u \in \Delta_{\mathsf{R}}.$$

Proof. This is a quite direct consequence of Definitions 2.3.9 – 2.3.12. We provide some details in order to exercise the reasoning technique we intensively use in the following. In particular, the right to left direction goes as follows

$$[\![S]\!] = \sum_{u\in\Delta_{\mathsf{R}}} \left[\!\!\left[\sum_{i\in I_u} P_i s_i\right]\!\!\right] = \sum_{u\in\Delta_{\mathsf{R}}}\sum_{i\in I_u} [\![P_i]\!]\,[\![s_i]\!]$$

$$= \sum_{u \in \Delta_{\mathsf{R}}} \left[\!\!\left[\sum_{i \in I_u} P_i \right]\!\!\right] u$$

$$= \sum_{u \in \Delta_{\mathsf{R}}} \left[\!\!\left[\sum_{j \in J_u} Q_j \right]\!\!\right] u$$

$$= \sum_{u \in \Delta_{\mathsf{R}}} \sum_{j \in J_u} [\![Q_j]\!] \, [\![t_j]\!] = \sum_{u \in \Delta_{\mathsf{R}}} \left[\!\!\left[\sum_{j \in J_u} Q_j t_j \right]\!\!\right] = [\![T]\!] \, .$$

The other way around is similarly provable. $\square$

Lemma 2.3.16. *Let $u \in \Delta_{\mathbb{P}}$ and $S \in \mathbb{P}\langle \Delta_{\mathbb{P}} \rangle$. If $u \overset{\nabla}{=} S$, then $S = \sum_{i \in I} P_i s_i + \sum_{j \in J} P_j s_j$ with:*

- *for all $i \in I$, $s_i \overset{\nabla}{=} u$ and $[\![\sum_{i \in I} P_i]\!] = 1$;*
- *for all $j \in J$, $s_j \overset{\nabla}{\neq} u$ and $[\![\sum_{j \in J} P_j s_j]\!] = \mathbf{0}$.*

Proof. Write $S = \sum_{v \in \Delta_{\mathsf{R}}} \sum_{k \in K_v} P_k s_k$, with $s_k \in \langle v \rangle$ for all $k \in K_v$; since $[\![u]\!] \in \Delta_{\mathsf{R}}$, $S = \sum_{i \in I} P_i s_i + \sum_{v \in \Delta_{\mathsf{R}} \setminus \{[\![u]\!]\}} \sum_{k \in K_v} P_k s_k$ with $s_i \overset{\nabla}{=} u$ for all $i \in I$, and $s_k \overset{\nabla}{\neq} u$ for all $k \in K_v$. Since $u \overset{\nabla}{=} S$, Lemma 2.3.15 implies $[\![\sum_{i \in I} P_i]\!] = 1$ and $\left[\!\!\left[\sum_{k \in K_v} P_k \right]\!\!\right] = 0$ for all $v \in \Delta_{\mathsf{R}} \setminus \{[\![u]\!]\}$. Hence, write $S = \sum_{i \in I} P_i s_i + \sum_{j \in J} P_j s_j$ and verify that the two conditions easily follow. $\square$

We now show that term evaluation is compatible with term substitution.

Lemma 2.3.17. *Let $x \in \mathcal{V}$ and $S, T \in \mathbb{P}\langle \Delta_{\mathbb{P}} \rangle$. Whenever $x \notin \mathsf{FV}(T)$, it holds that*

$$[\![S\,[T/x]]\!] = [\![S]\!] \, [[\![T]\!] \, /x] \, .$$

Proof. By induction on S, using the definition of term evaluation (Definition 2.3.9). We first address the cases in which S is a simple term s. Then one of the following applies:

- $s \in \mathcal{V}$; hence either $s = x$ or $s = y$, for any $y \in \mathcal{V} \setminus \{x\}$. In the former case, it follows

  $$[\![s\,[T/x]]\!] = [\![x\,[T/x]]\!] = [\![T]\!] = x\,[[\![T]\!] \, /x] = [\![x]\!] \, [[\![T]\!] \, /x] = [\![s]\!] \, [[\![T]\!] \, /x] \, ,$$

  whilst, in the latter case

  $$[\![s\,[T/x]]\!] = [\![y\,[T/x]]\!] = [\![y]\!] = y = y\,[[\![T]\!] \, /x] = [\![y]\!] \, [[\![T]\!] \, /x] = [\![s]\!] \, [[\![T]\!] \, /x] \, .$$

- $s = \lambda y.u$ and $y \notin \mathsf{FV}(T)$. The definitions of substitution (Definitions 1.2.4) and term evaluation (Definition 2.3.9) entail

$$\llbracket (\lambda y.u)\,[T/x] \rrbracket = \llbracket \lambda y.u\,[T/x] \rrbracket = \lambda y.\llbracket u\,[T/x] \rrbracket.$$

Hence, by the induction hypothesis, $\llbracket u\,[T/x] \rrbracket = \llbracket u \rrbracket\,[\llbracket T \rrbracket\,/x]$. It follows

$$\begin{aligned}\llbracket s\,[T/x] \rrbracket = \llbracket (\lambda y.u)\,[T/x] \rrbracket &= \lambda y.\llbracket u\,[T/x] \rrbracket \\ &= \lambda y.\llbracket u \rrbracket\,[\llbracket T \rrbracket\,/x] = \llbracket (\lambda y.u) \rrbracket\,[\llbracket T \rrbracket\,/x] = \llbracket s \rrbracket\,[\llbracket T \rrbracket\,/x]\,.\end{aligned}$$

- $s = (u)\,V$. The definitions of substitution (Definitions 1.2.4) and term evaluation (Definition 2.3.9) entail

$$\llbracket ((u)\,V)\,[T/x] \rrbracket = (\llbracket u\,[T/x] \rrbracket)\,\llbracket V\,[T/x] \rrbracket.$$

Hence, by the induction hypothesis, $\llbracket u\,[T/x] \rrbracket = \llbracket u \rrbracket\,[\llbracket T \rrbracket\,/x]$ and $\llbracket V\,[T/x] \rrbracket = \llbracket V \rrbracket\,[\llbracket T \rrbracket\,/x]$. It follows

$$\begin{aligned}\llbracket s\,[T/x] \rrbracket &= \llbracket ((u)\,V)\,[T/x] \rrbracket \\ &= (\llbracket u\,[T/x] \rrbracket)\,\llbracket V\,[T/x] \rrbracket \\ &= (\llbracket u \rrbracket\,[\llbracket T \rrbracket\,/x])\,\llbracket V \rrbracket\,[\llbracket T \rrbracket\,/x] \\ &= \llbracket ((u)\,V) \rrbracket\,[\llbracket T \rrbracket\,/x] = \llbracket s \rrbracket\,[\llbracket T \rrbracket\,/x]\,.\end{aligned}$$

Now assume $S = \sum_{i=1}^{n} P_i u_i$. The definitions of substitution (Definitions 1.2.4) and term evaluation (Definition 2.3.9) entail

$$\left\llbracket \left(\sum_{i=1}^{n} P_i u_i \right)\,[T/x] \right\rrbracket = \sum_{i=1}^{n} \llbracket P_i \rrbracket\,\llbracket u_i\,[T/x] \rrbracket\,.$$

From what we have just shown in the case of simple terms follows $\llbracket u_i\,[T/x] \rrbracket = \llbracket u_i \rrbracket\,[\llbracket T \rrbracket\,/x]$, for all $i \in \{1, \ldots, n\}$. It follows

$$\begin{aligned}\left\llbracket \left(\sum_{i=1}^{n} P_i u_i \right)\,[T/x] \right\rrbracket &= \sum_{i=1}^{n} \llbracket P_i \rrbracket\,\llbracket u_i\,[T/x] \rrbracket = \sum_{i=1}^{n} \llbracket P_i \rrbracket\,(\llbracket u_i \rrbracket\,[\llbracket T \rrbracket\,/x]) \\ &= \left(\sum_{i=1}^{n} \llbracket P_i \rrbracket\,\llbracket u_i \rrbracket \right)\,[\llbracket T \rrbracket\,/x] = \left\llbracket \sum_{i=1}^{n} P_i u_i \right\rrbracket\,[\llbracket T \rrbracket\,/x]\,.\end{aligned}$$

This concludes the proof. □

The following Sections 2.3.2 and 2.3.3 are devoted to show that the property of being sibling terms is stable with respect to normalisation. This is obviously sufficient to resolve the issue we discussed in Remark 2.3.2.

2.3.2 Strong normalisability

The module of terms $\mathbb{P}\langle\Delta_{\mathbb{P}}\rangle$ exhibits normal forms, and enjoy the characterisation of strongly normalisable terms given by Theorem 2.2.8.

Here we exploit this strong normalisability property of $\mathbb{P}\langle\Delta_{\mathbb{P}}\rangle$ in order to prove that sibling, *strongly normalisable* terms entail sibling normal forms. The results here presented have been published [Alb13], although in a slightly different form.

Theorem 2.3.18. *For all strongly normalisable $S, T \in \mathbb{P}\langle\Delta_{\mathbb{P}}\rangle$,*

$$\textit{if } S \overset{\triangledown}{=} T \textit{ then } \mathsf{NF}(S) \overset{\triangledown}{=} \mathsf{NF}(T).$$

The core of the above theorem is the upcoming Lemma 2.3.19, which is the main technical result of the current section.

Lemma 2.3.19. *For all $S, S' \in \mathbb{P}\langle\Delta_{\mathbb{P}}\rangle$ with $S \mathrel{\widetilde{\to}} S'$, there exists $S'' \in \mathbb{P}\langle\Delta_{\mathbb{P}}\rangle$ such that $S' \mathrel{\widetilde{\to}}^* S''$ and, for all $T \in \mathbb{P}\langle\Delta_{\mathbb{P}}\rangle$ with $T \overset{\triangledown}{=} S$, there exists $T'' \in \mathbb{P}\langle\Delta_{\mathbb{P}}\rangle$ such that $T \mathrel{\widetilde{\to}}^* T''$ with $T'' \overset{\triangledown}{=} S''$.*

Proof. By induction on k as $S \mathrel{\widetilde{\to}}_k S'$. If $k = 0$, then the result directly follows. Suppose the result holds for some k, then we extend it to $k+1$ by inspecting the possible cases for reduction $S \mathrel{\widetilde{\to_{k+1}}} S'$. We first address the cases in which S is a simple term s and $s \to_{k+1} S'$. Then one of the following applies:

- $s = \lambda x.u$ and $S' = \lambda x.U'$ with $u \to_k U'$; hence, by the induction hypothesis, there exists $U'' \in \mathbb{P}\langle\Delta_{\mathbb{P}}\rangle$ such that $U' \mathrel{\widetilde{\to}}^* U''$ and, for all $W \in \mathbb{P}\langle\Delta_{\mathbb{P}}\rangle$ with $W \overset{\triangledown}{=} u$, there exists $W'' \in \mathbb{P}\langle\Delta_{\mathbb{P}}\rangle$ such that $W \mathrel{\widetilde{\to}}^* W''$ and $W'' \overset{\triangledown}{=} U''$. Consider $S'' = \lambda x.U''$ and verify that $S' \mathrel{\widetilde{\to}}^* S''$.

  Let us now consider any term $T \in \mathbb{P}\langle\Delta_{\mathbb{P}}\rangle$ with $T \overset{\triangledown}{=} s$. Lemma 2.3.16 implies $T = \sum_{i\in I} Q_i t_i + \sum_{j\in J} Q_j t_j$ with $t_i \overset{\triangledown}{=} s$, $[\![\sum_{i\in I} Q_i]\!] = 1$ and $[\![\sum_{j\in J} Q_j t_j]\!] = \mathbf{0}$. By Definition 2.3.9, $t_i \overset{\triangledown}{=} s = \lambda x.u$ implies $t_i = \lambda x.w_i$ with $w_i \overset{\triangledown}{=} u$ for all $i \in I$; hence, there exists $W_i'' \in \mathbb{P}\langle\Delta_{\mathbb{P}}\rangle$ such that $w_i \mathrel{\widetilde{\to}}^* W_i''$ and $W_i'' \overset{\triangledown}{=} U''$. Consider $T'' = \sum_{i\in I} Q_i \lambda x.W_i'' + \sum_{j\in J} Q_j t_j$ and verify that $T \mathrel{\widetilde{\to}}^* T''$. Moreover, Definition 2.3.9 implies

  $$\begin{aligned} [\![T'']\!] &= \left[\!\!\left[\sum_{i\in I} Q_i \lambda x.W_i''\right]\!\!\right] + \left[\!\!\left[\sum_{j\in J} Q_j t_j\right]\!\!\right] = \sum_{i\in I} [\![Q_i]\!]\, \lambda x.[\![W_i'']\!] + \left[\!\!\left[\sum_{j\in J} Q_j t_j\right]\!\!\right] \\ &= \left[\!\!\left[\sum_{i\in I} Q_i\right]\!\!\right] \lambda x.[\![U'']\!] = \lambda x.[\![U'']\!] = [\![S'']\!], \end{aligned}$$

  i.e. $T'' \overset{\triangledown}{=} S''$.

- $s = (u)\, V$ and $S' = (U')\, V$ with $u \to_k U'$, or $S' = (u)\, V'$ with $V \xrightarrow{\sim}_k V'$.

  In the first case, by the induction hypothesis, there exists $U'' \in \mathbb{P}\langle\Delta_{\mathbb{P}}\rangle$ such that $U' \xrightarrow{\sim}^* U''$ and, for all $W \in \mathbb{P}\langle\Delta_{\mathbb{P}}\rangle$ with $W \stackrel{\nabla}{=} u$, there exists $W'' \in \mathbb{P}\langle\Delta_{\mathbb{P}}\rangle$ such that $W \xrightarrow{\sim}^* W''$ and $W'' \stackrel{\nabla}{=} U''$. Consider $S'' = (U'')\, V$ and verify that $S' \xrightarrow{\sim}^* S''$.

  Let us now consider any term $T \in \mathbb{P}\langle\Delta_{\mathbb{P}}\rangle$ with $T \stackrel{\nabla}{=} s$. Lemma 2.3.16 implies $T = \sum_{i\in I} Q_i t_i + \sum_{j\in J} Q_j t_j$ with $t_i \stackrel{\nabla}{=} s$, $[\![\sum_{i\in I} Q_i]\!] = 1$ and $[\![\sum_{j\in J} Q_j t_j]\!] = \mathbf{0}$. By Definition 2.3.9, $t_i \stackrel{\nabla}{=} s = (u)\, V$ implies $t_i = (w_i)\, Z_i$ with $w_i \stackrel{\nabla}{=} u$ and $Z_i \stackrel{\nabla}{=} V$ for all $i \in I$; hence, there exists $W_i'' \in \mathbb{P}\langle\Delta_{\mathbb{P}}\rangle$ such that $w_i \xrightarrow{\sim}^* W_i''$ and $W_i'' \stackrel{\nabla}{=} U''$. Consider $T'' = \sum_{i\in I} Q_i\, (W_i'')\, Z_i + \sum_{j\in J} Q_j t_j$ and verify that $T \xrightarrow{\sim}^* T''$. Moreover, Definition 2.3.9 implies

  $$\begin{aligned} [\![T'']\!] &= \left[\!\!\left[\sum_{i\in I} Q_i\, (W_i'')\, Z_i\right]\!\!\right] + \left[\!\!\left[\sum_{j\in J} Q_j t_j\right]\!\!\right] = \sum_{i\in I} [\![Q_i]\!]\, ([\![W_i'']\!])\, [\![Z_i]\!] + \left[\!\!\left[\sum_{j\in J} Q_j t_j\right]\!\!\right] \\ &= \left[\!\!\left[\sum_{i\in I} Q_i\right]\!\!\right] ([\![U'']\!])\, [\![V]\!] = ([\![U'']\!])\, [\![V]\!] = [\![S'']\!]\,, \end{aligned}$$

  i.e. $T'' \stackrel{\nabla}{=} S''$.

  In the second case, by the induction hypothesis, there exists $V'' \in \mathbb{P}\langle\Delta_{\mathbb{P}}\rangle$ such that $V' \xrightarrow{\sim}^* V''$ and, for all $W \in \mathbb{P}\langle\Delta_{\mathbb{P}}\rangle$ with $W \stackrel{\nabla}{=} V$, there exists $W'' \in \mathbb{P}\langle\Delta_{\mathbb{P}}\rangle$ such that $W \xrightarrow{\sim}^* W''$ and $W'' \stackrel{\nabla}{=} V''$. Consider $S'' = (u)\, V''$ and verify that $S' \xrightarrow{\sim}^* S''$.

  Let us now consider any term $T \in \mathbb{P}\langle\Delta_{\mathbb{P}}\rangle$ with $T \stackrel{\nabla}{=} s$. Lemma 2.3.16 implies $T = \sum_{i\in I} Q_i t_i + \sum_{j\in J} Q_j t_j$ with $t_i \stackrel{\nabla}{=} s$, $[\![\sum_{i\in I} Q_i]\!] = 1$ and $[\![\sum_{j\in J} Q_j t_j]\!] = \mathbf{0}$. By Definition 2.3.9, $t_i \stackrel{\nabla}{=} s = (u)\, V$ implies $t_i = (w_i)\, Z_i$ with $w_i \stackrel{\nabla}{=} u$ and $Z_i \stackrel{\nabla}{=} V$ for all $i \in I$; hence, there exists $Z_i'' \in \mathbb{P}\langle\Delta_{\mathbb{P}}\rangle$ such that $Z_i \xrightarrow{\sim}^* Z_i''$ and $Z_i'' \stackrel{\nabla}{=} V''$. Consider $T'' = \sum_{i\in I} Q_i\, (w_i)\, Z_i'' + \sum_{j\in J} Q_j t_j$ and verify that $T \xrightarrow{\sim}^* T''$. Moreover, Definition 2.3.9 implies

  $$\begin{aligned} [\![T'']\!] &= \left[\!\!\left[\sum_{i\in I} Q_i\, (w_i)\, Z_i''\right]\!\!\right] + \left[\!\!\left[\sum_{j\in J} Q_j t_j\right]\!\!\right] = \sum_{i\in I} [\![Q_i]\!]\, ([\![w_i]\!])\, [\![Z_i'']\!] + \left[\!\!\left[\sum_{j\in J} Q_j t_j\right]\!\!\right] \\ &= \left[\!\!\left[\sum_{i\in I} Q_i\right]\!\!\right] ([\![u]\!])\, [\![V'']\!] = ([\![u]\!])\, [\![V'']\!] = [\![S'']\!]\,, \end{aligned}$$

  i.e. $T'' \stackrel{\nabla}{=} S''$.

- $s = (\lambda x.u)\, V$ and $S' = u\, [V/x]$. Consider $S'' = S'$.

Let us now consider any term $T \in \mathbb{P}\langle\Delta_{\mathbb{P}}\rangle$ with $T \overset{\nabla}{=} s$. Lemma 2.3.16 implies $T = \sum_{i\in I} Q_i t_i + \sum_{j\in J} Q_j t_j$ with $t_i \overset{\nabla}{=} s$, $[\![\sum_{i\in I} Q_i]\!] = 1$ and $[\![\sum_{j\in J} Q_j t_j]\!] = \mathbf{0}$. By Definition 2.3.9, $t_i \overset{\nabla}{=} s = (\lambda x.u)\, V$ implies $t_i = (\lambda x.w_i)\, Z_i$ with $w_i \overset{\nabla}{=} u$ and $Z_i \overset{\nabla}{=} V$ for all $i \in I$. Consider $T'' = \sum_{i\in I} Q_i w_i\, [Z_i/x] + \sum_{j\in J} Q_j t_j$ and verify that $T \mathrel{\widetilde{\rightarrow}^*} T''$. Moreover, Definition 2.3.9 and Lemma 2.3.17 imply

$$
\begin{aligned}
[\![T'']\!] &= \sum_{i\in I} [\![Q_i]\!]\, [\![w_i\, [Z_i/x]]\!] + \left[\!\!\left[\sum_{j\in J} Q_j t_j\right]\!\!\right] = \sum_{i\in I} [\![Q_i]\!]\, [\![w_i]\!]\, [[\![Z_i]\!]\,/x] \\
&= \left[\!\!\left[\sum_{i\in I} Q_i\right]\!\!\right] [\![u]\!]\, [[\![V]\!]\,/x] = [\![u]\!]\, [[\![V]\!]\,/x] = [\![S'']\!]\,,
\end{aligned}
$$

i.e. $T'' \overset{\nabla}{=} S''$.

Now assume $S \mathrel{\widetilde{\rightarrow_{k+1}}} S'$. By Rule (1.8a), this amounts to the following: $S = Pu + V$ and $S' = PU' + V$ with $u \rightarrow_{k+1} U'$. From what we have just shown in the case of simple terms follows $U'' \in \mathbb{P}\langle\Delta_{\mathbb{P}}\rangle$ such that $U' \mathrel{\widetilde{\rightarrow}^*} U''$ and, for all $W \in \mathbb{P}\langle\Delta_{\mathbb{P}}\rangle$ with $W \overset{\nabla}{=} u$, there exists $W'' \in \mathbb{P}\langle\Delta_{\mathbb{P}}\rangle$ such that $W \mathrel{\widetilde{\rightarrow}^*} W''$ and $W'' \overset{\nabla}{=} U''$. Furthermore

$$
\begin{aligned}
V &= \sum_{s\in\Delta_{\mathsf{R}}} \sum_{k\in K_s} P_k v_k \quad \text{with, for all } k \in K_s,\ v_k \in \langle s\rangle \\
&= \sum_{i\in I} P_i v_i + \sum_{s\in\Delta_{\mathsf{R}}\setminus\{[\![u]\!]\}} \sum_{h\in H_s} P_h v_h,
\end{aligned}
$$

with $v_i \overset{\nabla}{=} u$ for all $i \in I$, and $(v_h)_{h\in H_s} \in \langle s\rangle$ for all $s \in \Delta_{\mathsf{R}}\setminus\{[\![u]\!]\}$ Hence, there exits $V_i'' \in \mathbb{P}\langle\Delta_{\mathbb{P}}\rangle$ such that $v_i \mathrel{\widetilde{\rightarrow}^*} V_i''$ and $V_i'' \overset{\nabla}{=} U''$ for all $i \in I$. Consider $S'' = PU'' + \sum_{i\in I} P_i V_i'' + \sum_{s\in\Delta_{\mathsf{R}}\setminus\{[\![u]\!]\}} \sum_{h\in H_s} P_h v_h$ and verify that $S' \mathrel{\widetilde{\rightarrow}^*} S''$.

Let us now consider any term $T \in \mathbb{P}\langle\Delta_{\mathbb{P}}\rangle$ with $T \overset{\nabla}{=} S$, *i.e.*

$$
\begin{aligned}
T &= \sum_{s\in\Delta_{\mathsf{R}}} \sum_{k\in K_s} Q_k w_k \quad \text{with, for all } k \in K_s,\ w_k \in \langle s\rangle \\
&= \sum_{j\in J} Q_j w_j + \sum_{s\in\Delta_{\mathsf{R}}\setminus\{[\![u]\!]\}} \sum_{k\in K_s} Q_k w_k
\end{aligned}
$$

with $w_j \overset{\nabla}{=} u$ for all $j \in J$, and $(w_k)_{k\in K_s} \in \langle s\rangle$ for all $s \in \mathsf{R}\langle\Delta_{\mathsf{R}}\rangle\setminus\{[\![u]\!]\}$. By Lemma 2.3.15 follows

$$
\left[\!\!\left[P + \sum_{i\in I} P_i\right]\!\!\right] = \left[\!\!\left[\sum_{j\in J} Q_j\right]\!\!\right], \tag{2.1}
$$

and

$$
\left[\!\!\left[\sum_{h\in H_s} P_h\right]\!\!\right] = \left[\!\!\left[\sum_{k\in K_s} Q_k\right]\!\!\right] \quad \text{for all } s \in \mathsf{R}\langle\Delta_{\mathsf{R}}\rangle\setminus\{[\![u]\!]\}. \tag{2.2}
$$

Since $w_j \stackrel{\nabla}{=} u$ for all $j \in J$, there exists $W_j'' \in \mathbb{P}\langle\Delta_{\mathbb{P}}\rangle$ such that $w_j \rightrightarrows^* W_j''$ and $W_j'' \stackrel{\nabla}{=} U''$. Consider $T'' = \sum_{j\in J} Q_j W_j'' + \sum_{s\in\Delta_{\mathsf{R}}\setminus\{[\![u]\!]\}} \sum_{k\in K_s} Q_k w_k$ and verify that $T \rightrightarrows^* T''$. Moreover, Definition 2.3.9 and identities (2.1) and (2.2) imply

$$
\begin{aligned}
[\![T'']\!] &= \left[\!\!\left[\sum_{j\in J} Q_j W_j''\right]\!\!\right] + \left[\!\!\left[\sum_{s\in\Delta_{\mathsf{R}}\setminus\{[\![u]\!]\}} \sum_{k\in K_s} Q_k w_k\right]\!\!\right] \\
&= \sum_{j\in J} [\![Q_j]\!]\,[\![W_j'']\!] + \sum_{s\in\Delta_{\mathsf{R}}\setminus\{[\![u]\!]\}} \sum_{k\in K_s} [\![Q_k]\!]\,[\![w_k]\!] \\
&= \left[\!\!\left[\sum_{j\in J} Q_j\right]\!\!\right] [\![U'']\!] + \sum_{s\in\Delta_{\mathsf{R}}\setminus\{[\![u]\!]\}} \left[\!\!\left[\sum_{k\in K_s} Q_k\right]\!\!\right] s \\
&= \left[\!\!\left[P + \sum_{i\in I} P_i\right]\!\!\right] [\![U'']\!] + \sum_{s\in\Delta_{\mathsf{R}}\setminus\{[\![u]\!]\}} \left[\!\!\left[\sum_{h\in H_s} P_h\right]\!\!\right] s \\
&= [\![P]\!]\,[\![U'']\!] + \sum_{i\in I} [\![P_i]\!]\,[\![V_i'']\!] + \sum_{s\in\Delta_{\mathsf{R}}\setminus\{[\![u]\!]\}} \sum_{h\in H_s} [\![Q_h]\!]\,[\![w_h]\!] \\
&= [\![PU'']\!] + \left[\!\!\left[\sum_{i\in I} P_i V_i''\right]\!\!\right] + \left[\!\!\left[\sum_{s\in\Delta_{\mathsf{R}}\setminus\{[\![u]\!]\}} \sum_{h\in H_s} Q_h w_h\right]\!\!\right] = [\![S'']\!]\,,
\end{aligned}
$$

i.e. $T'' \stackrel{\nabla}{=} S''$, which concludes the case and the proof. $\square$

We now prove Theorem 2.3.18, but we first recall its statement.

Theorem (2.3.18)**.** For all strongly normalisable $S, T \in \mathbb{P}\langle\Delta_{\mathbb{P}}\rangle$,

$$\text{if } S \stackrel{\nabla}{=} T \text{ then } \mathsf{NF}(S) \stackrel{\nabla}{=} \mathsf{NF}(T).$$

Proof. Since both S and T are strongly normalisable terms, let us consider the longest $\rightrightarrows$-reduction sequences to their respective normal form: *i.e.* for some $m, n \in \mathbb{N}$,

1. $S \rightrightarrows S_1 \rightrightarrows \ldots \rightrightarrows S_m = \mathsf{NF}(S)$;
2. $T \rightrightarrows T_1 \rightrightarrows \ldots \rightrightarrows T_n = \mathsf{NF}(T)$.

We prove the result by induction on $m + n$. If $m + n = 0$, then both terms S and T are in normal form and the result directly follows. Otherwise, suppose (at least) S to be a reducible term, that is $m \neq 0$. Then, since $S \stackrel{\nabla}{=} T$ by hypothesis, Lemma 2.3.21 on $S \rightrightarrows S_1$ implies that there exist $S', T' \in \mathbb{P}\langle\Delta_{\mathbb{P}}\rangle$ such that:

- $S_1 \rightrightarrows^* S'$;
- $T \rightrightarrows^* T'$;

- $S' \stackrel{\nabla}{=} T'$.

The strong normalisability hypothesis on S, T implies S' and T' to be strongly normalisable terms as well. Moreover, Church-Rosser property of reduction $\widetilde{\rightarrow}$ (Theorem 1.4.29) assures $\mathsf{NF}(S') = \mathsf{NF}(S)$ and $\mathsf{NF}(T') = \mathsf{NF}(T)$. Therefore, as before, let us consider S', T' longest $\widetilde{\rightarrow}$-reduction sequences to their respective normal forms: *i.e.* for some $p, q \in \mathbb{N}$,

3. $S' \,\widetilde{\rightarrow}\, S'_1 \,\widetilde{\rightarrow}\, \ldots \,\widetilde{\rightarrow}\, S'_p = \mathsf{NF}(S)$;

4. $T' \,\widetilde{\rightarrow}\, T'_1 \,\widetilde{\rightarrow}\, \ldots \,\widetilde{\rightarrow}\, T'_q = \mathsf{NF}(T)$.

As m and n are maximal, with $S \,\widetilde{\rightarrow}^*\, S'$ and $T \,\widetilde{\rightarrow}^*\, T'$, it follows $p + q < m + n$. Hence, the result $\mathsf{NF}(S) \stackrel{\nabla}{=} \mathsf{NF}(T)$ follows by the induction hypothesis. $\square$

2.3.3 Normalisability

We now give a similar result to Theorem 2.3.18, but in the case of just normalisable terms. Then, the following formulation strictly generalises the one presented in Section 2.3.2 as we prove that sibling, *normalisable* sibling terms entail sibling normal forms. This is unpublished work.

Theorem 2.3.20. *For all normalisable $S, T \in \mathbb{P}\langle \Delta_{\mathbb{P}} \rangle$,*

$$\text{if } S \stackrel{\nabla}{=} T \text{ then } \mathsf{NF}(S) \stackrel{\nabla}{=} \mathsf{NF}(T).$$

Generalising to just normalisability obviously comes at a cost, namely a more demanding proof technique. In particular, a similar proof to the one of Theorem 2.3.18 is not available here, due to the fact that there is no way we can appeal to the longest reduction sequence leading to normal form.

The idea then is to express Lemma 2.3.19 by means of parallel reduction, so as to keep control on the reduction length and appeal to the inductive hypothesis. Since parallel reduction may fire multiple redexes at each step, we need a generalisation of Lemma 2.3.19 to multiple terms, allowing the inductive reasoning to go through.

Lemma 2.3.21 is the main technical result of the current section. We deal with multiple $\overline{\rightrightarrows}$-reductions by appealing to Lemma 1.4.14 (hence, Lemma 1.4.13), which ultimately permits a common, single indexed, inductive reasoning.

Lemma 2.3.21. *Let $S_1, \ldots, S_n, S'_1, \ldots, S'_n \in \mathbb{P}\langle \Delta_{\mathbb{P}} \rangle$ such that*

$$\begin{cases} S_i \stackrel{\nabla}{=} S_j, & \text{for all } i, j \in \{1, \ldots, n\} \\ S_i \,\overline{\rightrightarrows}\, S'_i, & \text{for all } i \in \{1, \ldots, n\} \end{cases}.$$

Then there exist $S''_1, \dots, S''_n \in \mathbb{P}\langle\Delta_{\mathbb{P}}\rangle$ *such that* $\begin{cases} S''_i \overset{\nabla}{=} S''_j, & \textit{for all } i, j \in \{1, \dots, n\} \\ S'_i \mathrel{\widetilde{\rightrightarrows}}^* S''_i, & \textit{for all } i \in \{1, \dots, n\} \end{cases}$.

Proof. For all $i \in \{1, \dots, n\}$, Lemma 1.4.15 implies the existence of $k_i \in \mathbb{N}$ such that $S_i \mathrel{\overline{\rightrightarrows}_{k_i}} S'_i$. By Lemma 1.4.14 follows $k = \max{(k_i)}_{i \in \{1,\dots,n\}}$ such that $S_i \mathrel{\overline{\rightrightarrows}_k} S'_i$ for all $i \in \{1, \dots, n\}$: we prove the result by induction on such k.

If $k = 0$, then $S'_i = S_i$ and the result directly follows considering $S''_i = S'_i$ for all $i \in \{1, \dots, n\}$. Suppose the result holds for some k, then we extend it to $k+1$ by inspecting the possible cases for reduction $S_i \mathrel{\overline{\rightrightarrows}_{k+1}} S'_i$. We first address the cases in which every S_i is a simple term s_i and $s_i \rightrightarrows_{k+1} S'_i$. Moreover, since $S_f \overset{\nabla}{=} S_g$ for all $f, g \in \{1, \dots, n\}$, Definitions 2.3.10 and 2.3.9 implies the terms to be simple terms with same structure:

- For all $i \in \{1, \dots, n\}$, $s_i \in \mathcal{V}$; hence $s_i = S'_i$ and the result directly follows;

- For all $i \in \{1, \dots, n\}$, there exist u_i, U'_i such that:

  - $s_i = \lambda x.u_i$;
  - $S'_i = \lambda x.U'_i$;
  - $u_i \rightrightarrows_k U'_i$.

  Moreover, for all $f, g \in \{1, \dots, n\}$, $u_f \overset{\nabla}{=} u_g$; hence, by the induction hypothesis, there exist $U''_1, \dots, U''_n$ such that

  $$\begin{cases} U''_f \overset{\nabla}{=} U''_g, & \text{for all } f, g \in \{1, \dots, n\} \\ U'_i \mathrel{\widetilde{\rightrightarrows}}^* U''_i, & \text{for all } i \in \{1, \dots, n\} \end{cases}.$$

  Then, for all $i \in \{1, \dots, n\}$, consider $S''_i = \lambda x.U''_i$ and easily verify that $S'_i \mathrel{\widetilde{\rightrightarrows}}^* S''_i$. Moreover, for all $f, g \in \{1, \dots, n\}$, it holds

  $$\left[\!\left[S''_f \right]\!\right] = \left[\!\left[\lambda x.U''_f \right]\!\right] = \lambda x.\left[\!\left[U''_f \right]\!\right] = \lambda x.\left[\!\left[U''_g \right]\!\right] = \left[\!\left[\lambda x.U''_g \right]\!\right] = \left[\!\left[S''_g \right]\!\right],$$

  which implies the thesis: *i.e.* $S''_f \overset{\nabla}{=} S''_g$.

- For all $i \in \{1, \dots, n\}$, there exist u_i, U'_i, V_i, V'_i such that:

  - $s_i = (u_i)\, V_i$;
  - $S'_i = (U'_i)\, V'_i$;
  - $u_i \rightrightarrows_k U'_i$;
  - $V_i \mathrel{\overline{\rightrightarrows}_k} V'_i$.

Moreover, for all $f, g \in \{1, \ldots, n\}$, $u_f \overset{\nabla}{=} u_g$ and $V_f \overset{\nabla}{=} V_g$; hence, by the induction hypothesis:

– there exist $U''_1, \ldots, U''_n$ such that $\begin{cases} U''_f \overset{\nabla}{=} U''_g, & \text{for all } f, g \in \{1, \ldots, n\} \\ U'_i \mathrel{\overline{\rightrightarrows}}^* U''_i, & \text{for all } i \in \{1, \ldots, n\} \end{cases}$;

– there exist $V''_1, \ldots, V''_n$ such that $\begin{cases} V''_f \overset{\nabla}{=} V''_g, & \text{for all } f, g \in \{1, \ldots, n\} \\ V'_i \mathrel{\overline{\rightrightarrows}}^* V''_i, & \text{for all } i \in \{1, \ldots, n\} \end{cases}$.

Then, for all $i \in \{1, \ldots, n\}$, consider $S''_i = (U''_i)\, V''_i$ and easily verify that $S'_i \mathrel{\overline{\rightrightarrows}}^* S''_i$. Moreover, for all $f, g \in \{1, \ldots, n\}$, it holds

$$\llbracket S''_f \rrbracket = \llbracket \left(U''_f\right) V''_f \rrbracket = \left(\llbracket U''_f \rrbracket\right) \llbracket V''_f \rrbracket = \left(\llbracket U''_g \rrbracket\right) \llbracket V''_g \rrbracket = \llbracket \left(U''_g\right) V''_g \rrbracket = \llbracket S''_g \rrbracket,$$

which implies the thesis: *i.e.* $S''_f \overset{\nabla}{=} S''_g$.

- For all $i \in \{1, \ldots, n\}$, there exist u_i, U'_i, V_i, V'_i such that:

  – $s_i = (\lambda x. u_i)\, V_i$;

  – either $s'_i = U'_i\, [V'_i / x]$ (at least once) or $S'_i = (\lambda x. U'_i)\, V'_i$;

  – $u_i \rightrightarrows_k U'_i$;

  – $V_i \overline{\rightrightarrows_k} V'_i$.

  Moreover, for all $f, g \in \{1, \ldots, n\}$, $u_f \overset{\nabla}{=} u_g$ and $V_f \overset{\nabla}{=} V_g$; hence, by the induction hypothesis:

  – there exist $U''_1, \ldots, U''_n$ such that $\begin{cases} U''_f \overset{\nabla}{=} U''_g, & \text{for all } f, g \in \{1, \ldots, n\} \\ U'_i \mathrel{\overline{\rightrightarrows}}^* U''_i, & \text{for all } i \in \{1, \ldots, n\} \end{cases}$;

  – there exist $V''_1, \ldots, V''_n$ such that $\begin{cases} V''_f \overset{\nabla}{=} V''_g, & \text{for all } f, g \in \{1, \ldots, n\} \\ V'_i \mathrel{\overline{\rightrightarrows}}^* V''_i, & \text{for all } i \in \{1, \ldots, n\} \end{cases}$.

  Then, for all $i \in \{1, \ldots, n\}$, consider $S''_i = U''_i\, [V''_i / x]$ and easily verify that $S'_i \mathrel{\overline{\rightrightarrows}}^* S''_i$. Moreover, for all $f, g \in \{1, \ldots, n\}$, by Lemma 2.3.17 follows

$$\begin{aligned} \llbracket S''_f \rrbracket = \llbracket U''_f \left[V''_f / x\right] \rrbracket &= \llbracket U''_f \rrbracket \left[\llbracket V''_f \rrbracket / x\right] \\ &= \llbracket U''_g \rrbracket \left[\llbracket V''_g \rrbracket / x\right] = \llbracket U''_g \left[V''_g / x\right] \rrbracket = \llbracket S''_g \rrbracket, \end{aligned}$$

  which implies the thesis: *i.e.* $S''_f \overset{\nabla}{=} S''_g$.

Now assume $S_i \overline{\rightrightarrows_{k+1}} S'_i$ for all $i \in \{1, \dots, n\}$. By Rule (1.8b), this amounts to the following:

- $S_i = \sum_{j=1}^{m_i} P_{i,j} u_{i,j}$;
- $S'_i = \sum_{j=1}^{m_i} P_{i,j} U'_{i,j}$;
- for all $j \in \{1, \dots, m_i\}$, $u_{i,j} \rightrightarrows_{k+1} U'_{i,j}$.

For all $i \in \{1, \dots, n\}$, let us consider $(J_{i,t})_{t \in \Delta_\mathsf{R}} = (\{j \in \{1, \dots, m_i\} \mid u_{i,j} \in \langle t \rangle\})_{t \in \Delta_\mathsf{R}}$ the infinite, almost everywhere null, partition of $\{1, \dots, m_i\}$ such that every S_i can be rewritten as

$$S_i = \sum_{t \in \Delta_\mathsf{R}} \sum_{j \in J_{i,t}} P_{i,j} u_{i,j},$$

with $u_{i,j} \in \langle t \rangle$, for all $j \in J_{i,t}$. Therefore, from the hypothesis that $u_{i,j} \rightrightarrows_{k+1} U'_{i,j}$, respective terms S'_i can be written as

$$S'_i = \sum_{t \in \Delta_\mathsf{R}} \sum_{j \in J_{i,t}} P_{i,j} U'_{i,j}.$$

For all $f, g \in \{1, \dots, n\}$, Lemma 2.3.15 on $S_f \stackrel{\triangledown}{=} S_g$ implies

$$\left[\!\!\left[\sum_{j \in J_{f,t}} P_{f,j} \right]\!\!\right] = \left[\!\!\left[\sum_{j \in J_{g,t}} P_{g,j} \right]\!\!\right], \text{ for all } t \in \Delta_\mathsf{R} \tag{2.3}$$

Now consider the sets of simple terms $\left(\bigcup_{i \in \{1,\dots,n\}} \{u_{i,j} \mid j \in J_{i,t}\} \right)_{t \in \Delta_\mathsf{R}}$, and verify that they are uniform and pairwise disjoint sets (Definition 2.3.13). Thus, for all $t \in \Delta_\mathsf{R}$, the simple terms $(u_{i,j})_{i \in \{1,\dots,n\}, j \in J_{i,t}}$ are sibling terms, with $u_{i,j} \rightrightarrows_{k+1} U'_{i,j}$ by hypothesis; hence, from what we have just shown for simple terms, there exists $\left(U''_{i,j} \right)_{i \in \{1,\dots,n\}, j \in J_{i,t}}$ such that

$$\begin{cases} U''_{f,j} \stackrel{\triangledown}{=} U''_{g,k}, & \text{for all } f, g \in \{1, \dots, n\},\ j \in J_{f,t},\ k \in J_{g,t} \\ U'_{i,j} \rightrightarrows^* U''_{i,j}, & \text{for all } i \in \{1, \dots, n\},\ j \in J_{i,t} \end{cases}. \tag{2.4}$$

Then, for all $i \in \{1, \dots, n\}$, consider S''_i as the term

$$S''_i = \sum_{t \in \Delta_\mathsf{R}} \sum_{j \in J_{i,t}} P_{i,j} U''_{i,j}.$$

One can verify that $S'_i \mathrel{\widetilde{\rightrightarrows}}^* S''_i$, and by the Identities (2.3) and (2.4) conclude that, for all $f, g \in \{1, \ldots, n\}$,

$$
\begin{aligned}
\llbracket S''_f \rrbracket = \left\llbracket \sum_{t \in \Delta_{\mathsf{R}}} \sum_{j \in J_{f,t}} P_{f,j} U''_{f,j} \right\rrbracket &= \sum_{t \in \Delta_{\mathsf{R}}} \left\llbracket \sum_{j \in J_{f,t}} P_{f,j} U''_{f,j} \right\rrbracket \\
&= \sum_{t \in \Delta_{\mathsf{R}}} \left\llbracket \sum_{j \in J_{f,t}} P_{f,j} \right\rrbracket \llbracket U''_{f,j} \rrbracket \\
&= \sum_{t \in \Delta_{\mathsf{R}}} \left\llbracket \sum_{j \in J_{g,t}} P_{g,j} \right\rrbracket \llbracket U''_{g,j} \rrbracket \\
&= \sum_{t \in \Delta_{\mathsf{R}}} \left\llbracket \sum_{j \in J_{g,t}} P_{g,j} U''_{g,j} \right\rrbracket = \left\llbracket \sum_{t \in \Delta_{\mathsf{R}}} \sum_{j \in J_{g,t}} P_{g,j} U''_{g,j} \right\rrbracket = \llbracket S''_g \rrbracket ,
\end{aligned}
$$

i.e. $S''_f \stackrel{\nabla}{=} S''_g$, which concludes the case and the proof. □

We now prove a last lemma needed to prove Theorem 2.3.20.

Lemma 2.3.22. *For every $S, T, U \in \mathbb{P}\langle \Delta_{\mathbb{P}} \rangle$ with $S \mathrel{\widetilde{\rightrightarrows}}^* T$ and $S \mathrel{\overline{\rightrightarrows}} U$, there exists $V \in \mathbb{P}\langle \Delta_{\mathbb{P}} \rangle$ such that $T \mathrel{\overline{\rightrightarrows}} V$ and $U \mathrel{\widetilde{\rightrightarrows}}^* V$.*

Proof. This simply follows by a diagram chase, using the fact that $\overline{\rightrightarrows}^* = \widetilde{\rightrightarrows}^*$ (Corollary 1.4.25). More precisely, the hypothesis $S \mathrel{\widetilde{\rightrightarrows}}^* T$ implies $S \mathrel{\overline{\rightrightarrows}}^* T$: *i.e.* for some $n \in \mathbb{N}$, there exist $S_1, \ldots, S_n$ such that:

- $S \mathrel{\overline{\rightrightarrows}} S_1$;
- $S_i \mathrel{\overline{\rightrightarrows}} S_{i+1}$ for all $i \in \{1, \ldots, n-1\}$;
- $S_n \mathrel{\overline{\rightrightarrows}} T$.

By the property of strong confluence of $\overline{\rightrightarrows}$ (Theorem 1.4.29), there exist $U_1, \ldots, U_n$ such that:

- $U \mathrel{\overline{\rightrightarrows}} U_1$;
- $U_i \mathrel{\overline{\rightrightarrows}} U_{i+1}$ for all $i \in \{1, \ldots, n-1\}$;
- $S_i \mathrel{\overline{\rightrightarrows}} U_i$ for all $i \in \{1, \ldots, n\}$.

Consider then as V the term obtained from $S_n \mathrel{\overline{\rightrightarrows}} T$ and $S_n \mathrel{\overline{\rightrightarrows}} U_n$, again, by Theorem 1.4.29: V is such that $T \mathrel{\overline{\rightrightarrows}} V$ and $U_n \mathrel{\overline{\rightrightarrows}} V$. The latter implies $U \mathrel{\overline{\rightrightarrows}}^* V$, hence $U \mathrel{\widetilde{\rightrightarrows}}^* V$, which concludes the proof. □

We now proceed by providing the proof of Theorem 2.3.20, right after have recalled its statement.

Theorem (2.3.20)**.** For all normalisable $S, T \in \mathbb{P}\langle\Delta_{\mathbb{P}}\rangle$,

$$\text{if } S \stackrel{\triangledown}{=} T \text{ then } \mathsf{NF}(S) \stackrel{\triangledown}{=} \mathsf{NF}(T).$$

Proof. Since both S and T are normalisable terms (*i.e.* weakly normalisable), let us consider two $\overset{\frown}{\rightrightarrows}$-reduction sequences leading to their respective normal form: *i.e.* for some $m, n \in \mathbb{N}$,

- $S \overset{\frown}{\rightrightarrows} S_1 \overset{\frown}{\rightrightarrows} \dots \overset{\frown}{\rightrightarrows} S_m = \mathsf{NF}(S)$;
- $T \overset{\frown}{\rightrightarrows} T_1 \overset{\frown}{\rightrightarrows} \dots \overset{\frown}{\rightrightarrows} T_n = \mathsf{NF}(T)$.

Since $\overset{\frown}{\rightrightarrows} \subset \overline{\rightrightarrows}$ (Lemma 1.4.24), the same holds with $\overline{\rightrightarrows}$ in place of $\overset{\frown}{\rightrightarrows}$, namely

1. $S \overline{\rightrightarrows} S_1 \overline{\rightrightarrows} \dots \overline{\rightrightarrows} S_m = \mathsf{NF}(S)$;
2. $T \overline{\rightrightarrows} T_1 \overline{\rightrightarrows} \dots \overline{\rightrightarrows} T_n = \mathsf{NF}(T)$.

We prove the result by induction on $m + n$. If $m + n = 0$, then both terms S and T are in normal form and the result directly follows. Otherwise, suppose (at least) S to be a reducible term, that is $m \neq 0$. Then, since $S \stackrel{\triangledown}{=} T$ by hypothesis, Lemma 2.3.21 on $S \overline{\rightrightarrows} S_1$ and $T \overline{\rightrightarrows} T$ (indeed, recall that $\overline{\rightrightarrows}$ is reflexive) implies that there exist $S'_1, T' \in \mathbb{P}\langle\Delta_{\mathbb{P}}\rangle$ such that:

- $S_1 \overset{\frown}{\rightrightarrows}^* S'_1$;
- $T \overset{\frown}{\rightrightarrows}^* T'$;
- $S'_1 \stackrel{\triangledown}{=} T'$.

Lemma 2.3.22 implies that there exist $S'_2, \dots, S'_m \in \mathbb{P}\langle\Delta_{\mathbb{P}}\rangle$ such that:

- $S'_i \overline{\rightrightarrows} S'_{i+1}$;
- $S_{i+1} \overset{\frown}{\rightrightarrows}^* S'_{i+1}$;

for all $i \in \{1, \dots, m-1\}$. Moreover, since $S_m = \mathsf{NF}(S)$, reduction $S_m \overset{\frown}{\rightrightarrows}^* S'_m$ is actually an equality, implying $S'_m = \mathsf{NF}(S)$. Similarly, there exist $T'_1, \dots, T'_n \in \mathbb{P}\langle\Delta_{\mathbb{P}}\rangle$ such that:

- $T' \overline{\rightrightarrows} T'_1$;
- $T'_i \overline{\rightrightarrows} T'_{i+1}$;

- $T_i \mathrel{\tilde{\rightrightarrows}}^* T_i'$;

for all $i \in \{1, \ldots, n-1\}$. Reduction $T_n \mathrel{\tilde{\rightrightarrows}}^* T_n'$ also holds and, since $T_n = \mathsf{NF}(T)$, the latter is actually an equality, implying $T_n' = \mathsf{NF}(T)$.

Let us consider the two $\overline{\rightrightarrows}$ reduction sequences therefore obtained:

4. $S_1' \mathrel{\overline{\rightrightarrows}} S_2' \mathrel{\overline{\rightrightarrows}} \ldots \mathrel{\overline{\rightrightarrows}} S_m' = \mathsf{NF}(S)$;
5. $T' \mathrel{\overline{\rightrightarrows}} T_1' \mathrel{\overline{\rightrightarrows}} \ldots \mathrel{\overline{\rightrightarrows}} T_n' = \mathsf{NF}(T)$.

Notice that, whilst reduction sequences (2) and (5) have the same length, reduction sequence (4) is shorter than (1) by one-step of parallel reduction $\overline{\rightrightarrows}$; hence, the result $\mathsf{NF}(S) \overset{\triangledown}{=} \mathsf{NF}(T)$ follows by the induction hypothesis. $\square$

2.3.4 Unique normal forms for algebraic terms

Theorem 2.3.20 supposes nothing on the module $\mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$ in which terms of $\mathbb{P}\langle\Delta_{\mathbb{P}}\rangle$ are evaluated to. Even in the case of non-positive semirings R, such result allows us to give a normal form to a term in $\mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$ by assigning to it the term evaluation of the normal form, if it exists, computed in $\mathbb{P}\langle\Delta_{\mathbb{P}}\rangle$ for its notations.

Definition 2.3.23. *Let $S \in \mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$ and $T \in \mathbb{P}\langle\Delta_{\mathbb{P}}\rangle$, with $T \in \langle S\rangle$. The* normal form of *S is the term evaluation of the normal form of T, if it exists:* i.e. $\mathsf{NF}(S) = [\![\mathsf{NF}(T)]\!]$.

The notion of normal form we have just introduced is well-defined. Indeed, uniqueness follows by the Church-Rosser property of reduction $\tilde{\rightrightarrows}$ (Theorem 1.4.29), hence assuring that every normalisable term in $\mathbb{P}\langle\Delta_{\mathbb{P}}\rangle$ has a unique normal form, and the fact that sibling terms provide sibling normal forms (Theorem 2.3.20).

Notice moreover that Theorem 2.3.20 does not consider at all the actual reduction notion defined on $\mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$. This is reflected in Definition 2.3.23 where, in some sense, we define normal forms as if they are computed in a *big-step* defined operational semantics. In Section 2.4 we provide a way to compute these normal forms directly in the module $\mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$, as in a *small-step* defined operational semantics.

2.3.5 (Consistent) Term equivalence

The previous results permit to define a *partial*, but *consistent* at last, notion of term equivalence on whichever module of terms $\mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$.

Definition 2.3.24. *For all $S, T \in \mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$, $S \doteqdot T$ whenever there exist normalisable terms $U \in \langle S\rangle$ and $V \in \langle T\rangle$ such that* $[\![\mathsf{NF}(U)]\!] = [\![\mathsf{NF}(V)]\!]$.

It is obvious that terms having distinct redex-free writing are not $\doteqdot$-equivalent. Therefore it follows:

Proposition 2.3.25. *Term equivalence $\doteqdot$ is consistent.*

***Remark* 2.3.26.** As a consequence to consistency, there are (significant) terms which are not related by $\doteqdot$-equivalence. For instance, by this partial equivalence $\mathbf{0} \doteqdot \mathbf{0}$ whereas $\mathbf{\Omega} \not\doteqdot \mathbf{\Omega}$. Moreover, $\mathbf{\Omega} \not\doteqdot \mathbf{0}$. Observe that the collapse of $\widetilde{\rightarrow}^*$ (Section 2.1.1) caused the two to be $\cong$-equivalent.

2.4 Compute normal forms

In Section 2.3 we provided a way to *assign* a normal form to algebraic terms by means of the specific module of terms $\mathbb{P}\langle\Delta_{\mathbb{P}}\rangle$ and a term evaluation morphism relating the latter to whichever $\mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$. One might argue that such solution sounds *ad hoc*, and wonder whether there exist a different β-reduction notion for Λ_Σ exhibiting consistency. This section investigates such possibility.

2.4.1 Canonical relations from terms to terms

In virtue of what we have remarked in Section 2.1 in presence of non-positive semirings R, we introduce β-reduction as a relation from terms to terms by a variant of Rule (1.8a), that disallows the reduction of redexes appearing as the result of the algebraic decomposition of the additive identity of R. This has the immediate consequence that $\mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$ exhibits normal forms as Lemma 2.1.1 is no more valid. As in Section 1.4, we also provide a variant of Rule (1.8b) suitable for parallel reduction.

Definition 2.4.1. *Given a relation $\mathcal{R} \subseteq \Delta_{\mathsf{R}} \times \mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$, we define its extended canonical relations $\mathring{\mathcal{R}}, \widehat{\mathcal{R}} \subseteq \mathsf{R}\langle\Delta_{\mathsf{R}}\rangle \times \mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$ respectively as follows:*

$$S \mathrel{\mathring{\mathcal{R}}} S' \text{ if } S = au + T \text{ and } S' = aU' + T \text{ where } a \neq 0,\ u \notin \mathsf{Supp}(T) \text{ and } u \mathrel{\mathcal{R}} U'; \tag{2.5a}$$

$$\begin{aligned} S \mathrel{\widehat{\mathcal{R}}} S' \text{ if } S = \sum_{i=1}^{n} a_i u_i \text{ and } S' = \sum_{i=1}^{n} a_i U_i' \text{ where,}& \\ \text{for all } i, j \in \{1, \ldots, n\}, u_i \neq u_j \text{ and } u_i \mathrel{\mathcal{R}} U_i'.& \end{aligned} \tag{2.5b}$$

***Remark* 2.4.2.** Observe how Rules (2.5a) and (2.5b) are formulated: using the terminology introduced in Section 1.3, they respectively require $au + T$ and $\sum_{i=1}^{n} a_i u_i$ to be canonical terms.

Committing a slight abuse of vocabulary, we often refer to relations defined by means of Definition 2.4.1 as defined on canonical terms only.

The (canonical) extensions provided by Rules (2.5a) and (2.5b) may seem the natural choices for defining reduction notions for Λ_Σ. Indeed, it holds that $\mathring{\mathcal{R}} \subseteq \widehat{\mathcal{R}}$, and so by introducing $\rightarrowtriangle$ and $\Rightarrow$ as relations from simple terms to terms, one defines β-reduction $\mathring{\rightarrowtriangle}$ and its parallel version $\widehat{\Rightarrow}$ as relations from term to terms. However, it turns out that these latter do not enjoy confluence.

In Section 1.4, we detail the differences between these relations and the ones we have considered. However, since neither Rule (2.5a) nor Rule (2.5b) allow $\mathbf{0}$ to reduce as a linear combination of terms, a first evident distinction is that reduction relations $\mathring{\rightarrowtriangle}$ and $\widehat{\Rightarrow}$ can now be defined by induction on terms, as customary in pure λ-calculus.

In the following, we first show that reduction $\widehat{\Rightarrow}$ characterises the normal forms given by Definition 2.3.23. Thereafter, we approach the same problem in terms of the more atomic reduction $\mathring{\rightarrowtriangle}$. We conjecture the same holds for the latter as we are able to achieve the result only in presence of strong normalisability. In doing so, we argue the surprising fact that $\widehat{\Rightarrow}$ cannot be properly considered as the parallel version of $\mathring{\rightarrowtriangle}$, namely results akin to Lemma 1.4.24 and Corollary 1.4.25 are not available here.

2.4.2 Canonical parallel β-reduction

We now introduce a notion of parallel reduction defined on canonical terms only: we first define $\Rightarrow$ as a relation from simple terms to terms by which multiple redexes can be fired simultaneously, so that the actual reduction relation $\widehat{\Rightarrow}$ is obtained by means of Rule (2.5b). Notice that we formulate the following definition by induction on terms: the fact that $\mathbf{0}$ cannot properly reduce garantees the characterisation of a well-founded relation.

Definition 2.4.3. *The relation $\Rightarrow$ on $\Delta_{\mathsf{R}} \times \mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$ is defined by induction on terms by setting $s \Rightarrow S'$ as soon as one of the following holds:*

- $s = x$ *and* $S' = x$, *for all* $x \in \mathcal{V}$;
- $s = \lambda x.u$ *and* $S' = \lambda x.U'$ *with* $u \Rightarrow U'$;
- $s = (u)\, V$ *and* $S' = (U')\, V'$ *with* $u \Rightarrow U'$ *and* $V \mathrel{\widehat{\Rightarrow}} V'$;
- $s = (\lambda x.u)\, V$ *and* $S' = U'\,[V'/x]$ *with* $u \Rightarrow U'$ *and* $V \mathrel{\widehat{\Rightarrow}} V'$.

We call canonical parallel (β-)reduction *the relation* $\widehat{\Rightarrow}$.

Lemma 2.4.4. *The relation $\widehat{\Rightarrow}$ is reflexive.*

Proof. Simple induction on the term S such that $S \mathrel{\widehat{\Rightarrow}} S$. $\square$

Lemma 2.4.5. *The relation $\widehat{\Rrightarrow}$ is not contextual (in the sense of Definition 1.4.1).*

Proof. On the one hand, reduction $\widehat{\Rrightarrow}$ is reflexive (Lemma 2.4.4) and enjoys the first three clauses of Definition 1.4.1. The latter are provable by following the proof of the contextuality property of reduction $\overline{\rightrightarrows}$ (Proposition 1.4.18). On the other hand, it is straightforward to find an example showing that the 4th clause does not hold: *e.g.* consider s such that $s \widehat{\Rrightarrow} S'$ and $s \widehat{\Rrightarrow} S''$, and verify that $s + s \not\widehat{\Rrightarrow} S' + S''$. □

Nonetheless, under the additional hypothesis of an empty intersection for the support sets of the terms added up, $\widehat{\Rrightarrow}$ does exhibit the contextuality property expressed by the 4th clause of Definition 1.4.1.

Proposition 2.4.6. *For all $S, T \in \mathsf{R}\langle \Delta_{\mathsf{R}} \rangle$, with $\mathsf{Supp}(S) \cap \mathsf{Supp}(T) = \emptyset$, it holds that:*

$$S + T \widehat{\Rrightarrow} S' + T' \text{ as soon as } S \widehat{\Rrightarrow} S' \text{ and } T \widehat{\Rrightarrow} T'.$$

Proof. This directly follows from Definition 2.4.3 and Rule (2.5b). □

Lemma 2.4.7. $(\lambda x.S)\, T \widehat{\Rrightarrow} S'\, [T'/x]$ *whenever $S \widehat{\Rrightarrow} S'$ and $T \widehat{\Rrightarrow} T'$.*

Proof. Rule (2.5b) on $S \widehat{\Rrightarrow} S'$ entails $S = \sum_{i=1}^n a_i u_i$ and $S' = \sum_{i=1}^n a_i U_i'$ such that, for every distinct $i, j \in \{1, \ldots, n\}$, $u_i \neq u_j$ and $u_i \Rrightarrow U_i'$. By linearity (Definition 1.2.2) follows $(\lambda x.S)\, T = \sum_{i=1}^n a_i\, (\lambda x.u_i)\, T$ with $(\lambda x.u_i)\, T \neq \left(\lambda x.u_j\right) T$ and $(\lambda x.u_i)\, T \Rrightarrow U_i'\, [T'/x]$ by Definition 2.4.3. By Rule (2.5b) follows

$$(\lambda x.S)\, T = \sum_{i=1}^{n} a_i\, (\lambda x.u_i)\, T \widehat{\Rrightarrow} \sum_{i=1}^{n} a_i S_i' \left[T'/x\right] = S' \left[T'/x\right],$$

which concludes the proof. □

We now proceed in showing that the notion of normal form given by Definition 2.3.23 coincides with the one obtained by means of $\widehat{\Rrightarrow}$-reductions. Intuitively, we provide the latter by *converting* the normalising $\widetilde{\rightarrow}$-reductions in $\mathbb{P}\langle \Delta_{\mathbb{P}} \rangle$ into $\widehat{\Rrightarrow}$-reductions in $\mathsf{R}\langle \Delta_{\mathsf{R}} \rangle$. The following lemma is the cornerstone of this conversion.

Lemma 2.4.8. *For all $S, T \in \mathbb{P}\langle \Delta_{\mathbb{P}} \rangle$ with $S \overline{\rightrightarrows} T$, there exists $U \in \mathbb{P}\langle \Delta_{\mathbb{P}} \rangle$ such that $T \widetilde{\rightarrow}^* U$ and $[\![S]\!] \widehat{\Rrightarrow} [\![U]\!]$.*

We actually prove a stronger result that permits the induction reasoning to go through. Although the proof follows the structure of the one we have provided for Lemma 2.3.21, we give all the details for the sake of completeness.

Lemma 2.4.9. *Let $S_1, \ldots, S_n, S'_1, \ldots, S'_n \in \mathbb{P}\langle \Delta_{\mathbb{P}} \rangle$ such that*

$$\begin{cases} S_i \overset{\nabla}{=} S_j, & \text{for all } i, j \in \{1, \ldots, n\} \\ S_i \overline{\rightrightarrows} S'_i, & \text{for all } i \in \{1, \ldots, n\} \end{cases}.$$

Then there exist $S''_1, \ldots, S''_n \in \mathbb{P}\langle \Delta_{\mathbb{P}} \rangle$ such that $\begin{cases} S''_i \overset{\nabla}{=} S''_j, & \text{for all } i, j \in \{1, \ldots, n\} \\ S'_i \overrightarrow{\rightrightarrows}^* S''_i, & \text{for all } i \in \{1, \ldots, n\} \\ [\![S_i]\!] \overset{=}{\Rightarrow} [\![S''_i]\!], & \text{for all } i \in \{1, \ldots, n\} \end{cases}$.

Proof. For all $i \in \{1, \ldots, n\}$, Lemma 1.4.15 implies the existence of $k_i \in \mathbb{N}$ such that $S_i \overline{\rightrightarrows_{k_i}} S'_i$. By Lemma 1.4.14 follows $k = \max (k_i)_{i \in \{1,\ldots,n\}}$ such that $S_i \overline{\rightrightarrows_k} S'_i$ for all $i \in \{1, \ldots, n\}$: we prove the result by induction on such k.

If $k = 0$, then $S'_i = S_i$ and the result trivially follows considering $S''_i = S'_i$ for all $i \in \{1, \ldots, n\}$. Suppose the result holds for some k, then we extend it to $k+1$ by inspecting the possible cases for reduction $S_i \overline{\rightrightarrows_{k+1}} S'_i$. We first address the cases in which every S_i is actually a simple term s_i and $s_i \rightrightarrows_{k+1} S'_i$. Moreover, since $S_f \overset{\nabla}{=} S_g$ for all $f, g \in \{1, \ldots, n\}$, Definitions 2.3.10 and 2.3.9 implies the terms to be simple terms with same structure:

- For all $i \in \{1, \ldots, n\}$, $s_i \in \mathcal{V}$; hence $s_i = S'_i$ and the result trivially follows;
- For all $i \in \{1, \ldots, n\}$, there exist u_i, U'_i such that:
  - $s_i = \lambda x.u_i$;
  - $S'_i = \lambda x.U'_i$;
  - $u_i \rightrightarrows_k U'_i$.

  Moreover, for all $f, g \in \{1, \ldots, n\}$, $u_f \overset{\nabla}{=} u_g$; hence, by the induction hypothesis, there exist $U''_1, \ldots, U''_n$ such that

  $$\begin{cases} U''_f \overset{\nabla}{=} U''_g, & \text{for all } f, g \in \{1, \ldots, n\} \\ U'_i \overrightarrow{\rightrightarrows}^* U''_i, & \text{for all } i \in \{1, \ldots, n\} \\ [\![u_i]\!] \Rightarrow [\![U''_i]\!], & \text{for all } i \in \{1, \ldots, n\} \end{cases}.$$

  Then, for all $i \in \{1, \ldots, n\}$, consider $S''_i = \lambda x.U''_i$ and easily verify that $S'_i \overrightarrow{\rightrightarrows}^* S''_i$. Moreover, it holds that

  $$[\![s_i]\!] = \lambda x.[\![u_i]\!] \Rightarrow \lambda x.[\![U''_i]\!] = [\![S''_i]\!],$$

  i.e. $[\![s_i]\!] \Rightarrow [\![S''_i]\!]$. Finally, for all $f, g \in \{1, \ldots, n\}$, it holds

  $$[\![S''_f]\!] = [\![\lambda x.U''_f]\!] = \lambda x.[\![U''_f]\!] = \lambda x.[\![U''_g]\!] = [\![\lambda x.U''_g]\!] = [\![S''_g]\!],$$

  which implies the thesis: *i.e.* $S''_f \overset{\nabla}{=} S''_g$.

- For all $i \in \{1, \dots, n\}$, there exist u_i, U'_i, V_i, V'_i such that:
  - $s_i = (u_i)\, V_i$;
  - $S'_i = (U'_i)\, V'_i$;
  - $u_i \rightrightarrows_k U'_i$;
  - $V_i \overline{\rightrightarrows_k} V'_i$.

  Moreover, for all $f, g \in \{1, \dots, n\}$, $u_f \stackrel{\nabla}{=} u_g$ and $V_f \stackrel{\nabla}{=} V_g$; hence, by the induction hypothesis:

  - there exist $U''_1, \dots, U''_n$ such that $\begin{cases} U''_f \stackrel{\nabla}{=} U''_g, & \text{for all } f, g \in \{1, \dots, n\} \\ U'_i \,\widetilde{\rightrightarrows}^*\, U''_i, & \text{for all } i \in \{1, \dots, n\} \\ \llbracket u_i \rrbracket \Rightarrow \llbracket U''_i \rrbracket, & \text{for all } i \in \{1, \dots, n\} \end{cases}$;
  - there exist $V''_1, \dots, V''_n$ such that $\begin{cases} V''_f \stackrel{\nabla}{=} V''_g, & \text{for all } f, g \in \{1, \dots, n\} \\ V'_i \,\widetilde{\rightrightarrows}^*\, V''_i, & \text{for all } i \in \{1, \dots, n\} \\ \llbracket V_i \rrbracket \,\widehat{\Rightarrow}\, \llbracket V''_i \rrbracket, & \text{for all } i \in \{1, \dots, n\} \end{cases}$.

  Then, for all $i \in \{1, \dots, n\}$, consider $S''_i = (U''_i)\, V''_i$ and easily verify that $S'_i \,\widetilde{\rightrightarrows}^*\, S''_i$. Moreover, it holds that

  $$\llbracket s_i \rrbracket = (\llbracket u_i \rrbracket)\, \llbracket V_i \rrbracket \Rightarrow \left(\llbracket U''_i \rrbracket\right) \llbracket V''_i \rrbracket = \llbracket S''_i \rrbracket\,,$$

  i.e. $\llbracket s_i \rrbracket \Rightarrow \llbracket S''_i \rrbracket$. Finally, for all $f, g \in \{1, \dots, n\}$, it holds

  $$\left\llbracket S''_f \right\rrbracket = \left\llbracket \left(U''_f\right) V''_f \right\rrbracket = \left(\left\llbracket U''_f \right\rrbracket\right) \left\llbracket V''_f \right\rrbracket = \left(\left\llbracket U''_g \right\rrbracket\right) \left\llbracket V''_g \right\rrbracket = \left\llbracket \left(U''_g\right) V''_g \right\rrbracket = \left\llbracket S''_g \right\rrbracket,$$

  which implies the thesis: *i.e.* $S''_f \stackrel{\nabla}{=} S''_g$.

- For all $i \in \{1, \dots, n\}$, there exist u_i, U'_i, V_i, V'_i such that:
  - $s_i = (\lambda x.u_i)\, V_i$;
  - either $s'_i = U'_i\, [V'_i/x]$ (at least once) or $S'_i = (\lambda x.U'_i)\, V'_i$;
  - $u_i \rightrightarrows_k U'_i$;
  - $V_i \overline{\rightrightarrows_k} V'_i$.

  Moreover, for all $f, g \in \{1, \dots, n\}$, $u_f \stackrel{\nabla}{=} u_g$ and $V_f \stackrel{\nabla}{=} V_g$; hence, by the induction hypothesis:

– there exist $U''_1, \dots, U''_n$ such that $\begin{cases} U''_f \stackrel{\nabla}{=} U''_g, & \text{for all } f, g \in \{1, \dots, n\} \\ U'_i \rightrightarrows^* U''_i, & \text{for all } i \in \{1, \dots, n\} \\ [\![u_i]\!] \Rightarrow [\![U''_i]\!], & \text{for all } i \in \{1, \dots, n\} \end{cases}$;

– there exist $V''_1, \dots, V''_n$ such that $\begin{cases} V''_f \stackrel{\nabla}{=} V''_g, & \text{for all } f, g \in \{1, \dots, n\} \\ V'_i \rightrightarrows^* V''_i, & \text{for all } i \in \{1, \dots, n\} \\ [\![V_i]\!] \Rrightarrow [\![V''_i]\!], & \text{for all } i \in \{1, \dots, n\} \end{cases}$.

Then, for all $i \in \{1, \dots, n\}$, consider $S''_i = U''_i [V''_i/x]$ and easily verify that $S'_i \rightrightarrows^* S''_i$. Moreover, using Lemma 2.3.17, it holds that

$$\begin{aligned} [\![s_i]\!] = [\![(\lambda x.u_i)\, V]\!] = (\lambda x.[\![u_i]\!])\, [\![V]\!] \\ \Rightarrow [\![U''_i]\!] \left[[\![V''_i]\!] / x \right] = [\![U''_i \left[V''_i/x\right]]\!] = [\![S''_i]\!], \end{aligned}$$

i.e. $[\![s_i]\!] \Rightarrow [\![S''_i]\!]$. Finally, for all $f, g \in \{1, \dots, n\}$, by Lemma 2.3.17 follows

$$\begin{aligned} [\![S''_f]\!] = [\![U''_f \left[V''_f/x\right]]\!] = [\![U''_f]\!] \left[[\![V''_f]\!] / x \right] \\ = [\![U''_g]\!] \left[[\![V''_g]\!] / x \right] = [\![U''_g \left[V''_g/x\right]]\!] = [\![S''_g]\!], \end{aligned}$$

which implies the thesis: *i.e.* $S''_f \stackrel{\nabla}{=} S''_g$.

Now assume $S_i \overline{\rightrightarrows_{k+1}} S'_i$ for all $i \in \{1, \dots, n\}$. By Rule (1.8b), this amounts to the following:

- $S_i = \sum_{j=1}^{m_i} P_{i,j} u_{i,j}$;
- $S'_i = \sum_{j=1}^{m_i} P_{i,j} U'_{i,j}$;
- for all $j \in \{1, \dots, m_i\}$, $u_{i,j} \rightrightarrows_{k+1} U'_{i,j}$.

For all $i \in \{1, \dots, n\}$, let us consider $(J_{i,t})_{t \in \Delta_\mathsf{R}} = (\{j \in \{1, \dots, m_i\} \mid u_{i,j} \in \langle t \rangle\})_{t \in \Delta_\mathsf{R}}$ the infinite, almost everywhere null, partition of $\{1, \dots, m_i\}$ such that every S_i can be rewritten as

$$S_i = \sum_{t \in \Delta_\mathsf{R}} \sum_{j \in J_{i,t}} P_{i,j} u_{i,j},$$

with $u_{i,j} \in \langle t \rangle$, for all $j \in J_{i,t}$. Therefore, from the hypothesis that $u_{i,j} \rightrightarrows_{k+1} U'_{i,j}$, respective terms S'_i can be written as

$$S'_i = \sum_{t \in \Delta_\mathsf{R}} \sum_{j \in J_{i,t}} P_{i,j} U'_{i,j}.$$

For all $f, g \in \{1, \ldots, n\}$, Lemma 2.3.15 on $S_f \overset{\nabla}{=} S_g$ implies

$$\left[\!\!\left[\sum_{j \in J_{f,t}} P_{f,j}\right]\!\!\right] = \left[\!\!\left[\sum_{j \in J_{g,t}} P_{g,j}\right]\!\!\right], \text{ for all } t \in \Delta_{\mathsf{R}} \tag{2.6}$$

Now consider the sets of simple terms $\left(\bigcup_{i \in \{1,\ldots,n\}} \{u_{i,j} \mid j \in J_{i,t}\}\right)_{t \in \Delta_{\mathsf{R}}}$, and verify that they are uniform and pairwise disjoint sets (Definition 2.3.13). Thus, for all $t \in \Delta_{\mathsf{R}}$, the simple terms $(u_{i,j})_{i \in \{1,\ldots,n\}, j \in J_{i,t}}$ are sibling terms, with $u_{i,j} \rightrightarrows_{k+1} U'_{i,j}$ by hypothesis; hence, from what we have just shown for simple terms, there exists $\left(U''_{i,j}\right)_{i \in \{1,\ldots,n\}, j \in J_{i,t}}$ such that

$$\begin{cases} U''_{f,j} \overset{\nabla}{=} U''_{g,k}, & \text{for all } f, g \in \{1, \ldots, n\},\ j \in J_{f,t},\ k \in J_{g,t} \\ U'_{i,j} \overset{\rightrightarrows}{\rightarrow}{}^{*} U''_{i,j}, & \text{for all } i \in \{1, \ldots, n\},\ j \in J_{i,t} \\ [\![u_{i,j}]\!] \Rightarrow \left[\!\!\left[U''_{i,j}\right]\!\!\right], & \text{for all } i \in \{1, \ldots, n\},\ j \in J_{i,t} \end{cases} . \tag{2.7}$$

Then, for all $i \in \{1, \ldots, n\}$, consider S''_i as the term

$$S''_i = \sum_{t \in \Delta_{\mathsf{R}}} \sum_{j \in J_{i,t}} P_{i,j} U''_{i,j}.$$

One can verify that $S'_i \overset{\rightrightarrows}{\rightarrow}{}^{*} S''_i$, and moreover, for all $t \in \Delta_{\mathsf{R}}$, it holds that

$$\left[\!\!\left[\sum_{j \in J_{i,t}} P_{i,j} u_{i,j}\right]\!\!\right] = \left[\!\!\left[\sum_{j \in J_{i,t}} P_{i,j}\right]\!\!\right] [\![u_{i,j}]\!] \Rightarrow \left[\!\!\left[\sum_{j \in J_{i,t}} P_{i,j}\right]\!\!\right] \left[\!\!\left[U''_{i,j}\right]\!\!\right] = \left[\!\!\left[\sum_{j \in J_{i,t}} P_{i,j} U''_{i,j}\right]\!\!\right],$$

which implies, along with Proposition 2.4.6,

$$\begin{aligned} [\![S_i]\!] = \left[\!\!\left[\sum_{t \in \Delta_{\mathsf{R}}} \sum_{j \in J_{i,t}} P_{i,j} u_{i,j}\right]\!\!\right] &= \sum_{t \in \Delta_{\mathsf{R}}} \left[\!\!\left[\sum_{j \in J_{i,t}} P_{i,j} u_{i,j}\right]\!\!\right] \\ &\overset{\Rightarrow}{\Rightarrow} \sum_{t \in \Delta_{\mathsf{R}}} \left[\!\!\left[\sum_{j \in J_{i,t}} P_{i,j} U''_{i,j}\right]\!\!\right] = \left[\!\!\left[\sum_{t \in \Delta_{\mathsf{R}}} \sum_{j \in J_{i,t}} P_{i,j} U''_{i,j}\right]\!\!\right] = [\![S''_i]\!], \end{aligned}$$

i.e. $[\![S_i]\!] \overset{\Rightarrow}{\Rightarrow} [\![S''_i]\!]$. Finally, for all $f, g \in \{1, \ldots, n\}$, the Identities (2.6) and (2.7) entail

$$\begin{aligned} \left[\!\!\left[S''_f\right]\!\!\right] = \left[\!\!\left[\sum_{t \in \Delta_{\mathsf{R}}} \sum_{j \in J_{f,t}} P_{f,j} U''_{f,j}\right]\!\!\right] &= \sum_{t \in \Delta_{\mathsf{R}}} \left[\!\!\left[\sum_{j \in J_{f,t}} P_{f,j} U''_{f,j}\right]\!\!\right] \\ &= \sum_{t \in \Delta_{\mathsf{R}}} \left[\!\!\left[\sum_{j \in J_{f,t}} P_{f,j}\right]\!\!\right] \left[\!\!\left[U''_{f,j}\right]\!\!\right] \end{aligned}$$

$$
\begin{aligned}
&= \sum_{t \in \Delta_{\mathsf{R}}} \left[\!\!\left[\sum_{j \in J_{g,t}} P_{g,j} \right]\!\!\right] \left[\!\!\left[U''_{g,j} \right]\!\!\right] \\
&= \sum_{t \in \Delta_{\mathsf{R}}} \left[\!\!\left[\sum_{j \in J_{g,t}} P_{g,j} U''_{g,j} \right]\!\!\right] = \left[\!\!\left[\sum_{t \in \Delta_{\mathsf{R}}} \sum_{j \in J_{g,t}} P_{g,j} U''_{g,j} \right]\!\!\right] = \left[\!\!\left[S''_g \right]\!\!\right],
\end{aligned}
$$

i.e. $S''_f \overset{\nabla}{=} S''_g$, which concludes the case and the proof. □

Theorem 2.4.10. *For all normalisable* $S \in \mathbb{P}\langle \Delta_{\mathbb{P}} \rangle$, $[\![S]\!] \mathrel{\widehat{\Rrightarrow}}^* [\![\mathsf{NF}(S)]\!]$.

Proof. Since S is a normalisable term (*i.e.* weakly normalisable), let us consider the $\widetilde{\rightarrow}$-reduction sequence leading to its normal form: *i.e.* for some $m \in \mathbb{N}$,

$$S \mathrel{\widetilde{\rightarrow}} S_1 \mathrel{\widetilde{\rightarrow}} \ldots \mathrel{\widetilde{\rightarrow}} S_m = \mathsf{NF}(S).$$

Since $\widetilde{\rightarrow} \subset \overline{\rightrightarrows}$ (Lemma 1.4.24), the same holds with $\overline{\rightrightarrows}$ in place of $\widetilde{\rightarrow}$, namely

$$S \mathrel{\overline{\rightrightarrows}} S_1 \mathrel{\overline{\rightrightarrows}} \ldots \mathrel{\overline{\rightrightarrows}} S_m = \mathsf{NF}(S). \tag{2.8}$$

We prove the result by induction on m. If $m = 0$, then S is in normal form and the result directly follows. Otherwise, there exists $S_1 \in \mathbb{P}\langle \Delta_{\mathbb{P}} \rangle$ such that:

- $S \mathrel{\overline{\rightrightarrows}} S_1$;
- $S_1 \mathrel{\overline{\rightrightarrows}} \ldots \mathrel{\overline{\rightrightarrows}} S_m = \mathsf{NF}(S)$.

Lemma 2.4.8 implies $T_1 \in \mathbb{P}\langle \Delta_{\mathbb{P}} \rangle$ such that $S_1 \mathrel{\widetilde{\rightarrow}}^* T_1$ and $[\![S]\!] \mathrel{\widehat{\Rrightarrow}} [\![T_1]\!]$. By Lemma 2.3.22 follows $T_2, \ldots, T_m \in \mathbb{P}\langle \Delta_{\mathbb{P}} \rangle$ such that:

- $T_i \mathrel{\overline{\rightrightarrows}} T_{i+1}$;
- $S_{i+1} \mathrel{\widetilde{\rightarrow}}^* T_{i+1}$;

for all $i \in \{1, \ldots, m-1\}$. Moreover, since $S_m = \mathsf{NF}(S)$, reduction $S_m \mathrel{\widetilde{\rightarrow}}^* T_m$ is actually an equality, implying $T_m = \mathsf{NF}(S)$. Consider now the reduction sequence

$$T_1 \mathrel{\overline{\rightrightarrows}} \ldots \mathrel{\overline{\rightrightarrows}} T_m = \mathsf{NF}(S), \tag{2.9}$$

and notice that it is shorter than the original reduction sequence (2.8) by one-step of parallel reduction $\overline{\rightrightarrows}$; hence, by the induction hypothesis, $[\![T_1]\!] \mathrel{\widehat{\Rrightarrow}}^* [\![\mathsf{NF}(S)]\!]$. Then, it follows

$$[\![S]\!] \mathrel{\widehat{\Rrightarrow}} [\![T_1]\!] \mathrel{\widehat{\Rrightarrow}}^* [\![\mathsf{NF}(S)]\!],$$

which implies the thesis. □

2.4.3 Canonical β-reduction

Although it exactly characterises the notion of normal form of Definition 2.3.23, $\widehat{\Rrightarrow}$ is a notion of parallel β-reduction. One might wonder whether a more *atomic* reduction can provide the same result.

We therefore introduce a notion of β-reduction on canonical terms only: we first define $\rightarrowtriangle$ as a relation from simple terms to terms that validates the β-rule, so that the actual reduction relation $\overset{\circ}{\rightarrowtriangle}$ is obtained by means of Rule (1.8a). As it was the case for canonical parallel reduction, we formulate the following definition by induction on terms.

Definition 2.4.11. *The relation $\rightarrowtriangle$ on $\Delta_{\mathsf{R}} \times \mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$ is defined by induction on terms by setting $s \rightarrowtriangle S'$ as soon as one of the following holds:*

- *$s = \lambda x.u$ and $S' = \lambda x.U'$ with $u \rightarrowtriangle U'$;*
- *$s = (u)\,V$ and $S' = (U')\,V$ with $u \rightarrowtriangle U'$, or $S' = (u)\,V'$ with $V \overset{\circ}{\rightarrowtriangle} V'$;*
- *$s = (\lambda x.u)\,V$ and $S' = u\,[V/x]$.*

We call canonical β-reduction *the relation $\overset{\circ}{\rightarrowtriangle}$.*

It turns out that $\overset{\circ}{\rightarrowtriangle}^*$ is not contextual in the sense of Definition 1.4.1. In particular, the 4th case fails by the same argument we considered in Lemma 2.4.5, namely $s \overset{\circ}{\rightarrowtriangle} S'$ and $s \overset{\circ}{\rightarrowtriangle} S''$ do not imply $s + s \overset{\circ}{\rightarrowtriangle}^* S' + S''$ (on the contrary, reductions $s + s \overset{\circ}{\rightarrowtriangle} 2S'$ and $s + s \overset{\circ}{\rightarrowtriangle} 2S''$ are valid).

Contrary to the case of $\widehat{\Rrightarrow}$, considering the additional hypothesis of an empty intersection for the support set of the terms added up is unhelpful:

Proposition 2.4.12. *For all $S, T \in \mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$, with $\mathsf{Supp}(S) \cap \mathsf{Supp}(T) = \emptyset$, it does not hold that $S + T \overset{\circ}{\rightarrowtriangle}^* S' + T'$ as soon as $S \overset{\circ}{\rightarrowtriangle}^* S'$ and $T \overset{\circ}{\rightarrowtriangle}^* T'$.*

Proof. Suppose $s, t, u \in \Delta_{\mathsf{R}}$, with $\mathsf{Supp}(s) \cap \mathsf{Supp}(t) \cap \mathsf{Supp}(u) = \emptyset$, such that $s \rightarrowtriangle t$ and $t \rightarrowtriangle u$. Then it does not always hold that $s + t \overset{\circ}{\rightarrowtriangle}^* t + u$ (it may be the case that the result is $2u$). $\square$

In Section 2.4.4 we see that the lack of such contextuality property entails numerous problems concerning reduction $\overset{\circ}{\rightarrowtriangle}$; among others, there is no such Lemma 2.4.7 for relation $\overset{\circ}{\rightarrowtriangle}^*$ in general.

We now proceed in showing that, under the hypothesis of strong normalisability, the notion of normal form given by Definition 2.3.23 coincides with the one obtained by means of $\overset{\circ}{\rightarrowtriangle}$-reductions. The following lemma continues on the intuitive idea of converting the normalising $\widetilde{\rightrightarrows}$-reductions in $\mathbb{P}\langle\Delta_{\mathbb{P}}\rangle$ into $\overset{\circ}{\rightarrowtriangle}$-reductions in $\mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$.

Observe that, in contrast with Lemma 2.4.8, we consider a one-step of $\widetilde{\rightarrow}$ as reduction to convert in terms of reduction $\overset{\circ}{\rightarrowtriangle}$.

Lemma 2.4.13. *For all $S, S' \in \mathbb{P}\langle\Delta_{\mathbb{P}}\rangle$ with $S \,\widetilde{\rightarrow}\, S'$, there exists $S'' \in \mathbb{P}\langle\Delta_{\mathbb{P}}\rangle$ such that $S' \,\widetilde{\rightarrow}^*\, S''$, $[\![S]\!] \overset{\circ}{\rightarrowtriangle}{}^? [\![S'']\!]$ and, for all $T \in \mathbb{P}\langle\Delta_{\mathbb{P}}\rangle$ with $T \overset{\triangledown}{=} S$, there exists $T'' \in \mathbb{P}\langle\Delta_{\mathbb{P}}\rangle$ such that $T \,\widetilde{\rightarrow}^*\, T''$ with $T'' \overset{\triangledown}{=} S''$.*

Proof. By induction on k as $S \,\widetilde{\rightarrow}_k\, S'$. If $k = 0$, then the result trivially follows. Suppose the result holds for some k, then we extend it to $k+1$ by inspecting the possible cases for reduction $S \,\widetilde{\rightarrow_{k+1}}\, S'$. We first address the cases in which S is actually a simple term s and $s \rightarrow_{k+1} S'$. Then one of the following applies:

- $s = \lambda x.u$ and $S' = \lambda x.U'$ with $u \rightarrow_k U'$; hence, by the induction hypothesis, there exists $U'' \in \mathbb{P}\langle\Delta_{\mathbb{P}}\rangle$ such that $U' \,\widetilde{\rightarrow}^*\, U''$, $[\![u]\!] \rightarrowtriangle^? [\![U'']\!]$ and, for all $W \in \mathbb{P}\langle\Delta_{\mathbb{P}}\rangle$ with $W \overset{\triangledown}{=} u$, there exists $W'' \in \mathbb{P}\langle\Delta_{\mathbb{P}}\rangle$ such that $W \,\widetilde{\rightarrow}^*\, W''$ and $W'' \overset{\triangledown}{=} U''$. Consider $S'' = \lambda x.U''$ and verify that $S' \,\widetilde{\rightarrow}^*\, S''$. Moreover, Definition 2.3.9 implies
$$[\![s]\!] = [\![\lambda x.u]\!] = \lambda x.[\![u]\!] \rightarrowtriangle^? \lambda x.[\![U'']\!] = [\![\lambda x.U'']\!] = [\![S'']\!] .$$
Let us now consider any term $T \in \mathbb{P}\langle\Delta_{\mathbb{P}}\rangle$ with $T \overset{\triangledown}{=} s$. Lemma 2.3.16 implies $T = \sum_{i\in I} Q_i t_i + \sum_{j\in J} Q_j t_j$ with $t_i \overset{\triangledown}{=} s$, $[\![\sum_{i\in I} Q_i]\!] = 1$ and $[\![\sum_{j\in J} Q_j t_j]\!] = \mathbf{0}$. By Definition 2.3.9, $t_i \overset{\triangledown}{=} s = \lambda x.u$ implies $t_i = \lambda x.w_i$ with $w_i \overset{\triangledown}{=} u$ for all $i \in I$; hence, there exists $W_i'' \in \mathbb{P}\langle\Delta_{\mathbb{P}}\rangle$ such that $w_i \,\widetilde{\rightarrow}^*\, W_i''$ and $W_i'' \overset{\triangledown}{=} U''$. Consider $T'' = \sum_{i\in I} Q_i \lambda x.W_i'' + \sum_{j\in J} Q_j t_j$ and verify that $T \,\widetilde{\rightarrow}^*\, T''$. Moreover, Definition 2.3.9 implies
$$\begin{aligned}[\![T'']\!] &= \left[\!\!\left[\sum_{i\in I} Q_i \lambda x.W_i''\right]\!\!\right] + \left[\!\!\left[\sum_{j\in J} Q_j t_j\right]\!\!\right] = \sum_{i\in I} [\![Q_i]\!]\, \lambda x.[\![W_i'']\!] + \left[\!\!\left[\sum_{j\in J} Q_j t_j\right]\!\!\right] \\ &= \left[\!\!\left[\sum_{i\in I} Q_i\right]\!\!\right] \lambda x.[\![U'']\!] = \lambda x.[\![U'']\!] = [\![S'']\!] ,\end{aligned}$$
i.e. $T'' \overset{\triangledown}{=} S''$.

- $s = (u)\, V$ and $S' = (U')\, V$ with $u \rightarrow_k U'$, or $S' = (u)\, V'$ with $V \,\widetilde{\rightarrow}_k\, V'$.

  In the first case, by the induction hypothesis, there exists $U'' \in \mathbb{P}\langle\Delta_{\mathbb{P}}\rangle$ such that $U' \,\widetilde{\rightarrow}^*\, U''$, $[\![u]\!] \rightarrowtriangle^? [\![U'']\!]$ and, for all $W \in \mathbb{P}\langle\Delta_{\mathbb{P}}\rangle$ with $W \overset{\triangledown}{=} u$, there exists $W'' \in \mathbb{P}\langle\Delta_{\mathbb{P}}\rangle$ such that $W \,\widetilde{\rightarrow}^*\, W''$ and $W'' \overset{\triangledown}{=} U''$. Consider $S'' = (U'')\, V$ and verify that $S' \,\widetilde{\rightarrow}^*\, S''$. Moreover, Definition 2.3.9 implies
$$[\![s]\!] = [\![(u)\, V]\!] = ([\![u]\!])\, [\![V]\!] \rightarrowtriangle^? \left([\![U'']\!]\right) [\![V]\!] = [\![(U'')\, V]\!] = [\![U]\!] .$$

Let us now consider any term $T \in \mathbb{P}\langle\Delta_{\mathbb{P}}\rangle$ with $T \stackrel{\nabla}{=} s$. Lemma 2.3.16 implies $T = \sum_{i\in I} Q_i t_i + \sum_{j\in J} Q_j t_j$ with $t_i \stackrel{\nabla}{=} s$, $[\![\sum_{i\in I} Q_i]\!] = 1$ and $[\![\sum_{j\in J} Q_j t_j]\!] = \mathbf{0}$. By Definition 2.3.9, $t_i \stackrel{\nabla}{=} s = (u)\,V$ implies $t_i = (w_i)\,Z_i$ with $w_i \stackrel{\nabla}{=} u$ and $Z_i \stackrel{\nabla}{=} V$ for all $i \in I$; hence, there exists $W_i'' \in \mathbb{P}\langle\Delta_{\mathbb{P}}\rangle$ such that $w_i \mathrel{\widetilde{\rightarrow}}^* W_i''$ and $W_i'' \stackrel{\nabla}{=} U''$. Consider $T'' = \sum_{i\in I} Q_i\,(W_i'')\,Z_i + \sum_{j\in J} Q_j t_j$ and verify that $T \mathrel{\widetilde{\rightarrow}}^* T''$. Moreover, Definition 2.3.9 implies

$$\begin{aligned}
[\![T'']\!] &= \left[\!\!\left[\sum_{i\in I} Q_i\,(W_i'')\,Z_i\right]\!\!\right] + \left[\!\!\left[\sum_{j\in J} Q_j t_j\right]\!\!\right] = \sum_{i\in I} [\![Q_i]\!]\,([\![W_i'']\!])\,[\![Z_i]\!] + \left[\!\!\left[\sum_{j\in J} Q_j t_j\right]\!\!\right] \\
&= \left[\!\!\left[\sum_{i\in I} Q_i\right]\!\!\right]\,([\![U'']\!])\,[\![V]\!] = ([\![U'']\!])\,[\![V]\!] = [\![S'']\!]\,,
\end{aligned}$$

i.e. $T'' \stackrel{\nabla}{=} S''$.

In the second case, by the induction hypothesis, there exists $V'' \in \mathbb{P}\langle\Delta_{\mathbb{P}}\rangle$ such that $V' \mathrel{\widetilde{\rightarrow}}^* V''$, $[\![V]\!] \stackrel{\circ}{\rightarrowtriangle}{}^? [\![V'']\!]$ and, for all $W \in \mathbb{P}\langle\Delta_{\mathbb{P}}\rangle$ with $W \stackrel{\nabla}{=} V$, there exists $W'' \in \mathbb{P}\langle\Delta_{\mathbb{P}}\rangle$ such that $W \mathrel{\widetilde{\rightarrow}}^* W''$ and $W'' \stackrel{\nabla}{=} V''$. Consider $S'' = (u)\,V''$ and verify that $S' \mathrel{\widetilde{\rightarrow}}^* S''$. Moreover, Definition 2.3.9 implies

$$[\![s]\!] = [\![(u)\,V]\!] = ([\![u]\!])\,[\![V]\!] \stackrel{\circ}{\rightarrowtriangle}{}^? ([\![u]\!])\,[\![V'']\!] = [\![(u)\,V'']\!] = [\![V]\!]\,.$$

Let us now consider any term $T \in \mathbb{P}\langle\Delta_{\mathbb{P}}\rangle$ with $T \stackrel{\nabla}{=} s$. Lemma 2.3.16 implies $T = \sum_{i\in I} Q_i t_i + \sum_{j\in J} Q_j t_j$ with $t_i \stackrel{\nabla}{=} s$, $[\![\sum_{i\in I} Q_i]\!] = 1$ and $[\![\sum_{j\in J} Q_j t_j]\!] = \mathbf{0}$. By Definition 2.3.9, $t_i \stackrel{\nabla}{=} s = (u)\,V$ implies $t_i = (w_i)\,Z_i$ with $w_i \stackrel{\nabla}{=} u$ and $Z_i \stackrel{\nabla}{=} V$ for all $i \in I$; hence, there exists $Z_i'' \in \mathbb{P}\langle\Delta_{\mathbb{P}}\rangle$ such that $Z_i \mathrel{\widetilde{\rightarrow}}^* Z_i''$ and $Z_i'' \stackrel{\nabla}{=} V''$. Consider $T'' = \sum_{i\in I} Q_i\,(w_i)\,Z_i'' + \sum_{j\in J} Q_j t_j$ and verify that $T \mathrel{\widetilde{\rightarrow}}^* T''$. Moreover, Definition 2.3.9 implies

$$\begin{aligned}
[\![T'']\!] &= \left[\!\!\left[\sum_{i\in I} Q_i\,(w_i)\,Z_i''\right]\!\!\right] + \left[\!\!\left[\sum_{j\in J} Q_j t_j\right]\!\!\right] = \sum_{i\in I} [\![Q_i]\!]\,([\![w_i]\!])\,[\![Z_i'']\!] + \left[\!\!\left[\sum_{j\in J} Q_j t_j\right]\!\!\right] \\
&= \left[\!\!\left[\sum_{i\in I} Q_i\right]\!\!\right]\,([\![u]\!])\,[\![V'']\!] = ([\![u]\!])\,[\![V'']\!] = [\![S'']\!]\,,
\end{aligned}$$

i.e. $T'' \stackrel{\nabla}{=} S''$.

- $s = (\lambda x.u)\,V$ and $S' = u\,[V/x]$. Consider $S'' = S'$ and verify that Definition 2.3.9 and Lemma 2.3.17 imply

$$[\![s]\!] = [\![(\lambda x.u)\,V]\!] = (\lambda x.[\![u]\!])\,[\![V]\!] \rightarrowtriangle^? [\![u]\!]\,[[\![V]\!]\,/x] = [\![u\,[V/x]]\!] = [\![S'']\!]\,.$$

Let us now consider any term $T \in \mathbb{P}\langle\Delta_{\mathbb{P}}\rangle$ with $T \stackrel{\nabla}{=} s$. Lemma 2.3.16 implies $T = \sum_{i\in I} Q_i t_i + \sum_{j\in J} Q_j t_j$ with $t_i \stackrel{\nabla}{=} s$, $\llbracket \sum_{i\in I} Q_i \rrbracket = 1$ and $\llbracket \sum_{j\in J} Q_j t_j \rrbracket = \mathbf{0}$. By Definition 2.3.9, $t_i \stackrel{\nabla}{=} s = (\lambda x.u)\, V$ implies $t_i = (\lambda x.w_i)\, Z_i$ with $w_i \stackrel{\nabla}{=} u$ and $Z_i \stackrel{\nabla}{=} V$ for all $i \in I$. Consider $T'' = \sum_{i\in I} Q_i w_i\,[Z_i/x] + \sum_{j\in J} Q_j t_j$ and verify that $T \mathrel{\widetilde{\to}}^* T''$. Moreover, Definition 2.3.9 and Lemma 2.3.17 imply

$$\begin{aligned}
\llbracket T'' \rrbracket &= \sum_{i\in I} \llbracket Q_i \rrbracket \, \llbracket w_i\,[Z_i/x] \rrbracket + \left\llbracket \sum_{j\in J} Q_j t_j \right\rrbracket = \sum_{i\in I} \llbracket Q_i \rrbracket \, \llbracket w_i \rrbracket \, [\llbracket Z_i \rrbracket /x] \\
&= \left\llbracket \sum_{i\in I} Q_i \right\rrbracket \llbracket u \rrbracket \, [\llbracket V \rrbracket /x] = \llbracket u \rrbracket \, [\llbracket V \rrbracket /x] = \llbracket S'' \rrbracket\,,
\end{aligned}$$

i.e. $T'' \stackrel{\nabla}{=} S''$.

Now assume $S \mathrel{\widetilde{\to_{k+1}}} S'$. By Rule (1.8a), this amounts to the following: $S = Pu + V$ and $S' = PU' + V$ with $u \to_{k+1} U'$. From what we have just shown in the case of simple terms follows $U'' \in \mathbb{P}\langle\Delta_{\mathbb{P}}\rangle$ such that $U' \mathrel{\widetilde{\to}}^* U''$, $\llbracket u \rrbracket \rightarrowtriangle^{?} \llbracket U'' \rrbracket$ and, for all $W \in \mathbb{P}\langle\Delta_{\mathbb{P}}\rangle$ with $W \stackrel{\nabla}{=} u$, there exists $W'' \in \mathbb{P}\langle\Delta_{\mathbb{P}}\rangle$ such that $W \mathrel{\widetilde{\to}}^* W''$ and $W'' \stackrel{\nabla}{=} U''$. Furthermore

$$\begin{aligned}
V &= \sum_{s\in\Delta_{\mathsf{R}}} \sum_{k\in K_s} P_k v_k \qquad \text{with, for all } k \in K_s,\ v_k \in \langle s \rangle \\
&= \sum_{i\in I} P_i v_i + \sum_{s\in\Delta_{\mathsf{R}}\setminus\{\llbracket u \rrbracket\}} \sum_{h\in H_s} P_h v_h,
\end{aligned}$$

with $v_i \stackrel{\nabla}{=} u$ for all $i \in I$, and $(v_h)_{h\in H_s} \in \langle s \rangle$ for all $s \in \Delta_{\mathsf{R}}\setminus\{\llbracket u \rrbracket\}$. Hence, there exits $V_i'' \in \mathbb{P}\langle\Delta_{\mathbb{P}}\rangle$ such that $v_i \mathrel{\widetilde{\to}}^* V_i''$ and $V_i'' \stackrel{\nabla}{=} U''$ for all $i \in I$. Consider $S'' = PU'' + \sum_{i\in I} P_i V_i'' + \sum_{s\in\Delta_{\mathsf{R}}\setminus\{\llbracket u \rrbracket\}} \sum_{h\in H_s} P_h v_h$ and verify that $S' \mathrel{\widetilde{\to}}^* S''$. Moreover, Definition 2.3.9 implies

$$\begin{aligned}
\llbracket S \rrbracket = \llbracket Pu + V \rrbracket &= \left\llbracket Pu + \sum_{i\in I} P_i v_i + \sum_{s\in\Delta_{\mathsf{R}}\setminus\{\llbracket u \rrbracket\}} \sum_{h\in H_s} P_h v_h \right\rrbracket \\
&= \llbracket P \rrbracket \, \llbracket u \rrbracket + \sum_{i\in I} \llbracket P_i \rrbracket \, \llbracket v_i \rrbracket + \left\llbracket \sum_{s\in\Delta_{\mathsf{R}}\setminus\{\llbracket u \rrbracket\}} \sum_{h\in H_s} P_h v_h \right\rrbracket \\
&= \llbracket P \rrbracket \, \llbracket u \rrbracket + \left\llbracket \sum_{i\in I} P_i \right\rrbracket \llbracket u \rrbracket + \left\llbracket \sum_{s\in\Delta_{\mathsf{R}}\setminus\{\llbracket u \rrbracket\}} \sum_{h\in H_s} P_h v_h \right\rrbracket \\
&= \left\llbracket P + \sum_{i\in I} P_i \right\rrbracket \llbracket u \rrbracket + \left\llbracket \sum_{s\in\Delta_{\mathsf{R}}\setminus\{\llbracket u \rrbracket\}} \sum_{h\in H_s} P_h v_h \right\rrbracket
\end{aligned}$$

$$
\begin{aligned}
&\overset{\circ}{\rightarrowtriangle}{}^{?} \left[\!\!\left[P + \sum_{i\in I} P_i\right]\!\!\right] [\![U'']\!] + \left[\!\!\left[\sum_{s\in\Delta_{\mathsf{R}}\setminus\{[\![u]\!]\}} \sum_{h\in H_s} P_h v_h\right]\!\!\right] \\
&= [\![P]\!]\,[\![U'']\!] + \left[\!\!\left[\sum_{i\in I} P_i\right]\!\!\right] [\![U'']\!] + \left[\!\!\left[\sum_{s\in\Delta_{\mathsf{R}}\setminus\{[\![u]\!]\}} \sum_{h\in H_s} P_h v_h\right]\!\!\right] \\
&= [\![P]\!]\,[\![U'']\!] + \sum_{i\in I} [\![P_i]\!]\,[\![V_i'']\!] + \left[\!\!\left[\sum_{s\in\Delta_{\mathsf{R}}\setminus\{[\![u]\!]\}} \sum_{h\in H_s} P_h v_h\right]\!\!\right] \\
&= \left[\!\!\left[PU'' + \sum_{i\in I} P_i V_i'' + \sum_{s\in\Delta_{\mathsf{R}}\setminus\{[\![u]\!]\}} \sum_{h\in H_s} P_h v_h\right]\!\!\right] = [\![S'']\!]\,.
\end{aligned}
$$

Let us now consider any term $T \in \mathbb{P}\langle\Delta_{\mathbb{P}}\rangle$ with $T \overset{\nabla}{=} S$, *i.e.*

$$
\begin{aligned}
T &= \sum_{s\in\Delta_{\mathsf{R}}} \sum_{k\in K_s} Q_k w_k \quad \text{with, for all } k \in K_s,\ w_k \in \langle s\rangle \\
&= \sum_{j\in J} Q_j w_j + \sum_{s\in\Delta_{\mathsf{R}}\setminus\{[\![u]\!]\}} \sum_{k\in K_s} Q_k w_k
\end{aligned}
$$

with $w_j \overset{\nabla}{=} u$ for all $j \in J$, and $(w_k)_{k\in K_s} \in \langle s\rangle$ for all $s \in \mathsf{R}\langle\Delta_{\mathsf{R}}\rangle\setminus\{[\![u]\!]\}$. By Lemma 2.3.15 follows

$$
\left[\!\!\left[P + \sum_{i\in I} P_i\right]\!\!\right] = \left[\!\!\left[\sum_{j\in J} Q_j\right]\!\!\right], \tag{2.10}
$$

and

$$
\left[\!\!\left[\sum_{h\in H_s} P_h\right]\!\!\right] = \left[\!\!\left[\sum_{k\in K_s} Q_k\right]\!\!\right] \text{ for all } s \in \mathsf{R}\langle\Delta_{\mathsf{R}}\rangle\setminus\{[\![u]\!]\}. \tag{2.11}
$$

Since $w_j \overset{\nabla}{=} u$ for all $j \in J$, there exists $W_j'' \in \mathbb{P}\langle\Delta_{\mathbb{P}}\rangle$ such that $w_j \mathrel{\widetilde{\rightarrow}}{}^* W_j''$ and $W_j'' \overset{\nabla}{=} U''$. Consider $T'' = \sum_{j\in J} Q_j W_j'' + \sum_{s\in\Delta_{\mathsf{R}}\setminus\{[\![u]\!]\}} \sum_{k\in K_s} Q_k w_k$ and verify that $T \mathrel{\widetilde{\rightarrow}}{}^* T''$. Moreover, Definition 2.3.9 and identities (2.10) and (2.11) imply

$$
\begin{aligned}
[\![T'']\!] &= \left[\!\!\left[\sum_{j\in J} Q_j W_j''\right]\!\!\right] + \left[\!\!\left[\sum_{s\in\Delta_{\mathsf{R}}\setminus\{[\![u]\!]\}} \sum_{k\in K_s} Q_k w_k\right]\!\!\right] \\
&= \sum_{j\in J} [\![Q_j]\!]\,[\![W_j'']\!] + \sum_{s\in\Delta_{\mathsf{R}}\setminus\{[\![u]\!]\}} \sum_{k\in K_s} [\![Q_k]\!]\,[\![w_k]\!] \\
&= \left[\!\!\left[\sum_{j\in J} Q_j\right]\!\!\right] [\![U'']\!] + \sum_{s\in\Delta_{\mathsf{R}}\setminus\{[\![u]\!]\}} \left[\!\!\left[\sum_{k\in K_s} Q_k\right]\!\!\right] s
\end{aligned}
$$

$$
\begin{aligned}
&= \left[\!\!\left[P + \sum_{i \in I} P_i \right]\!\!\right] [\![U'']\!] + \sum_{s \in \Delta_{\mathsf{R}} \setminus \{[\![u]\!]\}} \left[\!\!\left[\sum_{h \in H_s} P_h \right]\!\!\right] s \\
&= [\![P]\!]\,[\![U'']\!] + \sum_{i \in I} [\![P_i]\!]\,[\![V_i'']\!] + \sum_{s \in \Delta_{\mathsf{R}} \setminus \{[\![u]\!]\}} \sum_{h \in H_s} [\![Q_h]\!]\,[\![w_h]\!] \\
&= [\![PU'']\!] + \left[\!\!\left[\sum_{i \in I} P_i V_i'' \right]\!\!\right] + \left[\!\!\left[\sum_{s \in \Delta_{\mathsf{R}} \setminus \{[\![u]\!]\}} \sum_{h \in H_s} Q_h w_h \right]\!\!\right] = [\![S'']\!] \,,
\end{aligned}
$$

i.e. $T'' \overset{\nabla}{=} S''$, which concludes the case and the proof. $\square$

By crucially using the strong normalisability hypothesis, Lemma 2.4.13 entails the following result.

Theorem 2.4.14. *For all strongly normalisable $S \in \mathbb{P}\langle \Delta_{\mathbb{P}} \rangle$, $[\![S]\!] \overset{\circ}{\rightarrow}{}^{*} [\![\mathsf{NF}(S)]\!]$.*

Proof. Since S is a strongly normalisable term, let us consider the longest $\widetilde{\rightrightarrows}$-reduction sequence to its normal form: *i.e.* for some $m \in \mathbb{N}$,

$$S \,\widetilde{\rightrightarrows}\, S_1 \,\widetilde{\rightrightarrows}\, \ldots \,\widetilde{\rightrightarrows}\, S_m = \mathsf{NF}(S).$$

We prove the result by induction on m. If $m = 0$, then $S = \mathsf{NF}(S)$ and the result directly follows. Otherwise, Lemma 2.4.13 on $S \,\widetilde{\rightrightarrows}\, S_1$ implies that there exists $T \in \mathbb{P}\langle \Delta_{\mathbb{P}} \rangle$ such that $S_1 \,\widetilde{\rightrightarrows}^{*}\, T$ and $[\![S]\!] \overset{\circ}{\rightarrow}{}^{?} [\![T]\!]$. The strong normalisability hypothesis on S implies T to be a strongly normalisable term as well. Moreover, Church-Rosser property of reduction $\widetilde{\rightrightarrows}$ (Theorem 1.4.29) assures $\mathsf{NF}(T) = \mathsf{NF}(S)$. Therefore, as before, let us consider the longest $\widetilde{\rightrightarrows}$-reduction sequences to its normal form: *i.e.* for some $n \in \mathbb{N}$,

$$T \,\widetilde{\rightrightarrows}\, T_1 \,\widetilde{\rightrightarrows}\, \ldots \,\widetilde{\rightrightarrows}\, T_n = \mathsf{NF}(S).$$

As m is maximal, with $S \,\widetilde{\rightrightarrows}^{+}\, T$, it follows $n < m$. Hence, by the induction hypothesis, $[\![T]\!] \overset{\circ}{\rightarrow}{}^{*} [\![\mathsf{NF}(S)]\!]$. Since $[\![S]\!] \overset{\circ}{\rightarrow}{}^{?} [\![T]\!]$, it follows

$$[\![S]\!] \overset{\circ}{\rightarrow}{}^{?} [\![T]\!] \overset{\circ}{\rightarrow}{}^{*} [\![\mathsf{NF}(S)]\!] \,,$$

i.e. $[\![S]\!] \overset{\circ}{\rightarrow}{}^{*} [\![\mathsf{NF}(S)]\!]$, which concludes the proof. $\square$

The $\overset{\circ}{\rightarrow}$-reduction characterisation of normal forms in the more general case of just normalisability is harder, insomuch as it remains an open problem to this day. We now briefly report about the two directions we explored in order to attack the problem: the first is impractical, whereas the second seems promising albeit far to provide any result.

As first attempt, one can try to persist with the idea of converting reductions from $\mathbb{P}\langle \Delta_{\mathbb{P}} \rangle$ to $\mathsf{R}\langle \Delta_{\mathsf{R}} \rangle$ by rephrasing Lemma 2.4.8, as well as Lemma 2.4.9, in terms of reduction $\overset{\circ}{\rightarrow}$. In particular this would mean replace each occurrence of $\widehat{\Rightarrow}$ with $\overset{\circ}{\rightarrow}^*$, as it easy to realise that $\overset{\circ}{\rightarrow} \subset \widehat{\Rightarrow}$. This way of proceeding turns out to be inadequate as a proof for such result would presume $\overset{\circ}{\rightarrow}^*$ to exhibit contextuality (at least of the kind of Proposition 2.4.6) and enjoy the β-rule. We already observed that the former is false, even in the looser sense (Proposition 2.4.12), like so the latter as we show in Section 2.4.4. To be thorough, we do not have counterexamples showing that this is not possible, but we are skeptical about its feasibility.

The second attempt tries to solve the problem by following the pattern of strong normalisability, as in Theorem 2.4.14. Of course, this latter is no longer valid in the case of normalisable terms as they do not exhibit the longest reduction sequence to the normal form. Analysing the proof of Theorem 2.4.14, this would cause difficulties when we need to consider the longest $\widetilde{\rightarrow}$-reduction sequence from the term T obtained by Lemma 2.4.13. In general, we cannot assert whether the reduction $S_1 \widetilde{\rightarrow}^* T$ is needed (*i.e.* whether T is a term of the original reduction sequence $S \widetilde{\rightarrow}^* \mathsf{NF}(S)$), and so we cannot guarantee on the length of $T \widetilde{\rightarrow}^* \mathsf{NF}(S)$. In Theorem 2.4.14, we appeal to the strong normalisability hypothesis for overcoming this issue.

The idea therefore is to consider a particular $\widetilde{\rightarrow}$-reduction *strategy* (*i.e.* usually described as a deterministic way of firing redexes) to attain normal forms in $\mathbb{P}\langle\Delta_{\mathbb{P}}\rangle$, so that to characterise normalisable terms as strongly normalisable in this sense. This means that a result following Theorem 2.4.14 would be feasible, assuming an updated version of Lemma 2.4.13 in terms of this reduction strategy should not be a problem to obtain. We refer to Chapter 3 for further details.

2.4.4 Canonical means defectiveness

Here we describe the flaws that reductions $\overset{\circ}{\rightarrow}$ and $\widehat{\Rightarrow}$ exhibit (some are surprising). In particular, these defects are common to any reduction notion defined on canonical terms only. We conclude by briefly talking about the term equivalence induced by these latter.

Canonical reduction relations

Reduction relations $\overset{\circ}{\rightarrow}$ and $\widehat{\Rightarrow}$ are consistent, even in presence of non-positive coefficients: it is straightforward to verify that the arguments leading to the collapse of reduction $\widetilde{\rightarrow}^*$ (Section 2.1) apply neither to reduction $\overset{\circ}{\rightarrow}$ nor to reduction $\widehat{\Rightarrow}$. Indeed, these latter do not allow the term $\mathbf{0}$ to properly reduce (to be precise, there is no way $\mathbf{0}$ reduces to $S' - S$, assuming S reduces to S').

This is exclusively due to Definition 2.4.1, which entails $\overset{\circ}{\rightarrow}$ and $\widehat{\Rightarrow}$ to be defined on canonical terms only. Definition 2.4.1 causes nonetheless these latter to exhibit significant defect from the point of view of reduction theory. In the following, we report on them one at a time.

Contextuality. We have already established, respectively in Sections 2.4.3 and 2.4.2, that reductions $\overset{\circ}{\rightarrow}^*$ and $\widehat{\Rrightarrow}$ do not enjoy contextuality in the sense of Definition 1.4.1. In particular, we recall that both fail to verify the 4th clause of the latter. The other issue that follows are almost entirely due to the lack of this contextuality property.

Local confluence. Whenever one needs to establish whether a reduction relation exhibits confluence, one should verify local confluence first, since the latter can be seen as a necessary conditions for the former. We recall here its definition.

Definition 2.4.15. *A binary relation $\rightarrow$ on a set $\mathcal{S}$ is said to exhibit* local confluence *(or,* weak confluence*) whenever, for all $M, N, O \in \mathcal{S}$ such that $M \rightarrow N$ and $M \rightarrow O$, there exists $P \in \mathcal{S}$ such that $N \rightarrow^* P$ and $O \rightarrow^* P$.*

Reduction relations $\overset{\circ}{\rightarrow}$ and $\widehat{\Rrightarrow}$ do not exhibit confluence, since they are not even locally confluent. We now give a counterexample to local confluence for reduction $\overset{\circ}{\rightarrow}$, but the same argument can be reformulated in terms of reduction $\widehat{\Rrightarrow}$.

***Counterexample* 2.4.16.** Let $s, t \in \Delta_{\mathsf{R}}$, with $s \neq t$, such that $s \overset{\circ}{\rightarrow} t$ and $t \overset{\circ}{\rightarrow} s + y$. Then, $s + t \in \mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$ is a counterexample to local confluence for reduction $\overset{\circ}{\rightarrow}$.

Proof. We first prove that the term $s + t$ is indeed a counterexample. Thereafter, we show that s and t are definable simple terms.

By one-step $\overset{\circ}{\rightarrow}$-reduction, the term $s + t$ reduces either to $2t$ or $2s + y$: the former is obtained by reducing s into t, while the latter by reducing t into $s + y$. The two terms have no common term to reduce to by means of reduction $\overset{\circ}{\rightarrow}^*$. Indeed, suppose for the sake of contradiction that there is $U \in \mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$ such that:

- $2t \overset{\circ}{\rightarrow}^* U$;
- $2s + y \overset{\circ}{\rightarrow} 2t + y \overset{\circ}{\rightarrow}^* U$.

It is easy to see that the existence of such U leads to the absurd: in particular, U should admit an even amount of y variables in the first case, whereas such amount should be odd in the second one. This is contradictory.

We now prove that s and t are actual simple terms with the needed $\overset{\circ}{\rightarrow}$ dynamics. Actually, we provide only the term s as t is obtained from it by one-step reduction. Let us write $\propto_y = \lambda x.(\mathbf{I})\,((x)\,x + y)$, and set $s = (\propto_y)\propto_y$. Then, verify that

$$s = (\propto_y)\propto_y \rightarrow (\mathbf{I})\,((\propto_y)\propto_y + y) = t,$$

$$t = (\mathbf{I})\left(\left(\propto_y\right)\propto_y + y\right) \rightarrowtriangle \left(\propto_y\right)\propto_y + y = s + y,$$

which concludes the proof. $\square$

Counterexample 2.4.16 crucially exploits the constraint Rule (2.5a) imposes on reduction $\overset{\circ}{\rightarrowtriangle}$. Indeed, the argument we provide above is no longer valid when considering reduction $\widetilde{\rightarrow}$: in particular, Definition 1.4.6 and Rule (1.8a) allow $U = t + s + y$ as result of the following two $\widetilde{\rightarrow}$-reductions:

$$\begin{aligned} 2t = t + t &\,\widetilde{\rightarrow}\, t + s + y = U, \\ 2s + y = s + s + y &\,\widetilde{\rightarrow}\, t + s + y = U. \end{aligned}$$

Canonical β-reduction is erroneous. Reduction $\overset{\circ}{\rightarrowtriangle}$ fails to comply with a quite obvious extension of the β-rule: *i.e.* it does not holds that

$$(\lambda x.U)\, V \overset{\circ}{\rightarrowtriangle}{}^{*}\, U\,[V/x]\,, \tag{2.12}$$

with $U, V \in \mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$.

At first, reduction (2.12) seems plausible in the current setting. Indeed, if $U = \sum_{i=1}^{n} a_i u_i$, then one may expect (erroneously) that the reduction of each simple term $(\lambda x.u_i)\, V$ into $u_i\,[V/x]$ must imply $(\lambda x.U)\, V \overset{\circ}{\rightarrowtriangle}{}^{*}\, U\,[V/x]$: *i.e.* by using the laws of algebraic linearity, deduce that

$$\left(\lambda x.\sum_{i=1}^{n} a_i u_i\right) V = \sum_{i=1}^{n} a_i\,(\lambda x.u_i)\, V \overset{\circ}{\rightarrowtriangle}{}^{*}\, \sum_{i=1}^{n} a_i u_i\,[V/x] = \left(\sum_{i=1}^{n} a_i u_i\right)[V/x]\,.$$

This turns to be false, as we provide a counterexample to reduction (2.12).

***Counterexample* 2.4.17.** Let $s, t \in \Delta_{\mathsf{R}}$ with $s = (\lambda y.(y)\, y)\, x$ and $t = (x)\, x$. Then, consider $U = s + t$, $V = \lambda x.s$ and verify that $(\lambda x.U)\, V \not\overset{\circ}{\rightarrowtriangle}{}^{*}\, U\,[V/x]$.

Proof. By linearity follows

$$(\lambda x.U)\, V = (\lambda x.s)\, V + (\lambda x.t)\, V.$$

Obviously, we get

$$U\,[V/x] = s\,[V/x] + t\,[V/x] \tag{2.13}$$

only in the case we reduce both the terms $(\lambda x.s)\, V$ and $(\lambda x.t)\, V$. On the one hand, if we reduce the former, it holds (by using α-equivalence, as last operation)

$$(\lambda x.s)\, V = (\lambda x.(\lambda y.(y)\, y)\, x)\, V \rightarrowtriangle (\lambda y.(y)\, y)\, V = (\lambda x.t)\, V, \tag{2.14}$$

that is $s\,[V/x] = (\lambda x.t)\,V$. On the other hand, if we reduce the latter, it holds

$$(\lambda x.t)\,V = (\lambda x.(x)\,x)\,V \rightarrow (V)\,V = (\lambda x.s)\,V, \tag{2.15}$$

that is $t\,[V/x] = (\lambda x.s)\,V$. Then, from the identity (2.13) follows

$$\begin{aligned} U\,[V/x] &= s\,[V/x] + t\,[V/x] \\ &= (\lambda x.t)\,V + (\lambda x.s)\,V. \end{aligned} \tag{2.16}$$

Since $\overset{\circ}{\rightarrow}$ is a one-step reduction relation, by first reducing as in the identity (2.14) results in

$$(\lambda x.U)\,V = (\lambda x.s)\,V + (\lambda x.t)\,V \overset{\circ}{\rightarrow} (\lambda x.t)\,V + (\lambda x.t)\,V,$$

and, as $\overset{\circ}{\rightarrow}$ is a relation on canonical terms only, by reducing as in the identity (2.15) results in

$$(\lambda x.t)\,V + (\lambda x.t)\,V = 2\,(\lambda x.t)\,V \overset{\circ}{\rightarrow} 2\,(\lambda x.s)\,V \neq (\lambda x.t)\,V + (\lambda x.s)\,V,$$

that is $(\lambda x.U)\,V \not\overset{\circ}{\rightarrow}^{*} U\,[V/x]$. The same outcome is obtained even by first reducing as in the identity (2.15) and later as in the identity (2.14):

$$\begin{aligned} (\lambda x.U)\,V &= (\lambda x.s)\,V + (\lambda x.t)\,V \overset{\circ}{\rightarrow} (\lambda x.s)\,V + (\lambda x.s)\,V \\ &= 2\,(\lambda x.s)\,V \overset{\circ}{\rightarrow} 2\,(\lambda x.t)\,V \\ &\neq (\lambda x.t)\,V + (\lambda x.s)\,V, \end{aligned}$$

that is $(\lambda x.U)\,V \not\overset{\circ}{\rightarrow}^{*} U\,[V/x]$. □

Notice that, although not contextual in the sense of Definition 1.4.1, reduction $\widehat{\Rrightarrow}$ is not affected by the above Counterexample 2.4.17: *i.e.* Rule (2.5b) entails

$$(\lambda x.U)\,V = (\lambda x.s)\,V + (\lambda x.t)\,V \widehat{\Rrightarrow} (\lambda x.t)\,V + (\lambda x.s)\,V = U\,[V/x]\,. \tag{2.17}$$

As a matter of fact, reduction (2.12) is valid with respect to $\widehat{\Rrightarrow}$ (Lemma 2.4.7).

Relating reductions $\overset{\circ}{\rightarrow}$ and $\widehat{\Rrightarrow}$. In this work, we make use of a common pattern when studying the reduction theory of λ-calculus, namely characterise properties of β-reduction in terms of parallel β-reduction. For example, in the setting of the algebraic λ-calculus, this becomes useful for proving the Church-Rosser property (Section 1.4.3), to provide a notion of normal form (Section 2.3) or to show a factorisation theorem (Chapter 3).

This methodology relies on the fact that reduction and its parallel version exhibit the crucial property of the kind stated by Lemma 1.4.24 and Corollary 1.4.25: in our current case, this would mean $\overset{\circ}{\rightarrow} \subset \widehat{\Rrightarrow} \subset \overset{\circ}{\rightarrow}^{*}$ and $\widehat{\Rrightarrow}^{*} = \overset{\circ}{\rightarrow}^{*}$ respectively.

This seems a clear fact, as Rule (2.5b) is undoubtedly the most obvious generalisation of Rule (2.5a) and Definition 2.4.3 follows the commen pattern in providing a parallel version of Definition 2.4.11. However, it is not hard to realise that, in general, reductions $\overset{\circ}{\rightharpoonup}$ and $\widehat{\Rrightarrow}$ do not enjoy such properties: of course $\overset{\circ}{\rightharpoonup} \subset \widehat{\Rrightarrow}$, but Counterexample 2.4.17 and reduction (2.17) show $\widehat{\Rrightarrow} \not\subset \overset{\circ}{\rightharpoonup}^*$ (in particular, the two are not comparable). Hence, $\widehat{\Rrightarrow}^* \neq \overset{\circ}{\rightharpoonup}^*$.

Nevertheless, we need to precise that $\widehat{\Rrightarrow}^* = \overset{\circ}{\rightharpoonup}^*$ is not valid in the *general case*. Notice indeed that the terms we provide in Counterexample 2.4.17 exhibit a circular dynamics, as both reduce to the other by one-step of $\overset{\circ}{\rightharpoonup}$-reduction. That is, those terms are not normalisable (*w.r.t.* any of the reduction relations we have considered in this thesis). It remains to understand if the aforementioned equivalence holds in the case of normalisable terms, given that we already know it is valid on strongly normalisable terms (Theorem 2.4.14).

***Remark* 2.4.18.** Observe that these problematic characteristics are not limited to reductions $\overset{\circ}{\rightharpoonup}$ and $\widehat{\Rrightarrow}$. It is straightforward to verify that every reduction relation $\mathcal{R}$ from simple terms to terms extended to a relation on terms either as $\mathring{\mathcal{R}}$ by Rule (2.5a) or as $\widehat{\mathcal{R}}$ by Rule (2.5b) exhibits those same properties. Indeed, we only need to exploit the fact that these relations are defined on canonical terms only.

Term equivalence [Vau09]

Rules (2.5a) and (2.5b) prevent the collapse of reduction relations by construction: every reduction $\to$ defined on terms by means of Definition 2.4.1 does not exhibit inconsistency (Corollary 2.1.3). Nonetheless, its induced term equivalences $\sim$ remains inconsistent.

Lemma 2.4.19. *Let* R *be non-positive. For all* $S, T \in \mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$*,* $S \sim T$ *holds.*

Proof. First of all notice that, for all $S \in \mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$, the term $\infty_S = (\Theta)\,\lambda x.(S + x)$ admits the reduction $\infty_S \to^* S + \infty_S$. Recall that whenever R is not positive, then there are $a, b \in \mathsf{R}^\bullet$ such that $a + b = 0$. Therefore, both the following two reduction sequences are valid: for all $U \in \mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$,

- $a\infty_U + b\,(\mathbf{I})\,\infty_U \to^* aU + a\infty_U + b\,(\mathbf{I})\,\infty_U \to aU + a\infty_U + b\infty_U = aU$;
- $a\infty_U + b\,(\mathbf{I})\,\infty_U \to a\infty_U + b\infty_U = \mathbf{0}$.

Then, it follows

$$aS \sim a\infty_U + b\,(\mathbf{I})\,\infty_U \sim \mathbf{0} \sim a\infty_U + b\,(\mathbf{I})\,\infty_U \sim aT,$$

i.e. $S \sim T$, which concludes the proof. $\square$

Chapter 3

Factorisation

In this chapter we prove the *factorisation theorem* for the algebraic λ-calculus.

The reason for considering the factorisation theorem is the quest for a deterministic and normalising way of contracting redexes (*i.e.* a normalising strategy) that we discussed at the end of Section 2.4.3. Indeed, it is well-known [Bar84] that in pure λ-calculus the factorisation theorem can be used to characterise the *leftmost strategy* as normalising, namely the strategy that repeatedly reduces the redex whose λ is the furthest to the left (hence, the leftmost redex).

The factorisation theorem asserts that any β-reduction sequence $\to_\beta^*$ can be decomposed into a sequence of *head reduction* and *internal ones* as $(\to_{\mathtt{h}} \cup \to_{\mathtt{i}})^*$, and then reorganised into two sequences $\to_{\mathtt{h}}^*\to_{\mathtt{i}}^*$, that is by first reducing on the head and then everywhere else [Tak95]. As a consequence, head reductions result as the *essential* part of a computation.

Here we prove a weaker formulation that uses *function* and *argument* decomposition of β-reduction, nonetheless sufficient to characterise head normalisability. In particular, function reduction is non-deterministic, hence not a proper strategy. We show that this is the best one can achieve in the setting of the algebraic λ-calculus, which denies the usual technique to prove that the leftmost strategy is normalising. The question on whether or not this latter holds remains unanswered.

A similar result has been obtained by Pagani and Tranquilli for the resource λ-calculus [PT09]. Although both calculi are derived from the differential λ-calculus, their dynamics are quite different.

Contents

3.1 Reduction in function/argument position **85**

3.1.1 Confluence of function reduction 88

3.2 Decompose + Swap = Factorisation **89**

3.3 Head normalisability **98**

The first section introduces the decomposition of β-reduction into function and argument reductions. The former is also proved to enjoy confluence.

The second section is devoted to prove the factorisation theorem on the basis of Takahashi's technique [Tak95] for the pure λ-calculus. The method is adapted to the algebraic λ-calculus, obtaining a factorisation property with respect to function reduction. The section concludes by showing that a stronger formulation of theorem, resembling the classical one, is out of reach.

Finally, the third section applies the factorisation theorem in order to show that function reduction is sufficient to characterise head normalisability.

3.1 Reduction in function/argument position

In this section we provide a decomposition of $\widetilde{\rightarrow}$-reduction into two, which we call *function reduction* and *argument reduction* respectively. Generally speaking, we name function reduction every reduction contracting a redex whose λ-abstraction is never on the right side of an application. As the dual one, we name argument reduction every reduction which is not a function one. In pure λ-calculus, a formal definition of function reduction $s \rightarrow_{\mathtt{f}} s'$ would proceed by induction on (the size of) s as follows:

- if $s = \lambda x.u$, then $s' = \lambda x.u'$ whenever $u \rightarrow_{\mathtt{f}} u'$; [Abstraction]
- if $s = (u)\, v$, then $s' = (u')\, v$ whenever $u \rightarrow_{\mathtt{f}} u'$; [Application]
- if $s = (\lambda x.u)\, v$, then $s' = u\, [v/x]$. [Redex]

Notice how funtion reduction is indeed a special case of β-reduction, namely $\rightarrow_{\mathtt{f}} \subset \rightarrow_{\beta}$: the [Application] case of the former provides only the first clause of the two exhibited by the latter (Section 1.4.1), as reduction takes place only on the left subterm (*i.e.* function position) of an application term. One then defines application reduction $\rightarrow_{\mathtt{a}}$ as $\rightarrow_{\beta} \setminus \rightarrow_{\mathtt{f}}$.

Definition 3.1.1. *A binary relation $\mathcal{R} \subseteq \mathsf{R}\langle\Delta_{\mathsf{R}}\rangle \times \mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$ is said to be* function contextual *if it is contextual, but for the 2nd clause which is replaced by:*

- *if $S\ \mathcal{R}\ T$, then $(S)\, U\ \mathcal{R}\ \ (T)\, U$.*

Contrary to Definition 1.4.6, where the reduction $\rightarrow$ needs to be defined as the union of an increasing sequence of relations due to the use of Rule (1.8a), function reduction can be defined as a relation from simple terms to terms by a common inductive definition.

Definition 3.1.2. *The relation $\rightarrow_{\mathtt{f}}$ on $\Delta_{\mathsf{R}} \times \mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$ is defined by induction on terms. In particular, set $s \rightarrow_{\mathtt{f}} S'$ as soon as one of the following holds:*

- *$s = \lambda x.u$ and $S' = \lambda x.U'$ with $u \rightarrow_{\mathtt{f}} U'$;*
- *$s = (u)\, V$ and $S' = (U')\, V$ with $u \rightarrow_{\mathtt{f}} U'$;*
- *$s = (\lambda x.u)\, V$ and $S' = u\, [V/x]$.*

We call function β-reduction*, or simply* function reduction*, the relation $\widetilde{\rightarrow}_{\mathtt{f}}$.*

Definition of $\to_{\mathtt{f}}$ corresponds to the outline of function reduction we gave at the very beginning of this section: since no clause is defined in terms of function reduction on linear combination of terms, $\to_{\mathtt{f}}$ is unrelated with the algebraic extension we endowed the pure λ-calculus.

Function reduction $\widetilde{\to_{\mathtt{f}}}$ is non-deterministic (indeed, this is already the case in λ-calculus), hence not a strategy.

***Remark* 3.1.3.** Observe that function reductions take place in linear position, namely where algebraic linearity applies. As a matter of fact, this decomposition of β-reduction is already known in the literature of pure λ-calculus after the works on *σ-equivalence* [Reg94] and *linear head reduction* [MP94, DR04], influenced by the fine-grained notion of computation of linear logic proof-nets. These notions have recently been revised and put to use in a series of works [ADL14, ABM14].

We now proceed to prove a couple of properties concerning function reduction. In particular, we show $\widetilde{\to_{\mathtt{f}}}^*$ to be function contextual.

Lemma 3.1.4. *Let $S, S', T \in \mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$. If $S \widetilde{\to_{\mathtt{f}}} S'$, then the following hold:*

1. $\lambda x.S \widetilde{\to_{\mathtt{f}}} \lambda x.S'$;
2. $(S)\,T \widetilde{\to_{\mathtt{f}}} (S')\,T$;
3. $aS \widetilde{\to_{\mathtt{f}}}^{?} aS'$, *for all* $a \in \mathsf{R}$;
4. $S + T \widetilde{\to_{\mathtt{f}}} S' + T$.

Proof. By Definition 3.1.2 and Rule (1.8a), as in Lemma 1.4.10. □

Lemma 3.1.5. *The relation $\widetilde{\to_{\mathtt{f}}}^*$ is function contextual.*

Proof. This is a direct consequence of Lemma 3.1.4, using the reflexive and transitive properties of $\widetilde{\to_{\mathtt{f}}}^*$. □

Lemma 3.1.6. *Let $x \in \mathcal{V}$ and $S, S', T \in \mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$. If $S \widetilde{\to_{\mathtt{f}}} S'$ then $S\,[T/x] \ \widetilde{\to_{\mathtt{f}}}^* S'\,[T/x]$. In particular, the same holds for $S \widetilde{\to_{\mathtt{f}}}^* S'$.*

Proof. This is an easy induction on $S \widetilde{\to_{\mathtt{f}}} S'$. Notice that, in contrast with what it would happen in pure λ-calculus, the conclusion is expressed in terms of $\widetilde{\to_{\mathtt{f}}}^*$ instead of $\widetilde{\to_{\mathtt{f}}}$. Indeed, consider $S = (\lambda y.x)\,z$ which reduces to $S' = x$ with one-step reduction, and $T = \sum_{i=1}^{n} a_i t_i$. The result follows by n-steps of function reduction:

$$S\,[T/x] = \left(\lambda y.\sum_{i=1}^{n} a_i t_i\right) z = \sum_{i=1}^{n} a_i\,(\lambda y.t_i)\,z \widetilde{\to_{\mathtt{f}}}^* \ \sum_{i=1}^{n} a_i t_i = x\,[T/x] = S'\,[T/x]\,.$$

The general case follows from this latter by induction on the length of $S \mathrel{\widetilde{\to_{\mathtt{f}}}^{*}} S'$. $\square$

We define argument reduction as the dual one of function reduction. Observe that, since argument reduction fires redexes on the right subterm (*i.e.* argument position) of an application, its direct characterisation is a relation from simple terms to simple terms. Moreover, it is given by a simple inductive definition as the crucial clause of reduction (*i.e.* the last one) is based on reduction $\widetilde{\to}$.

Definition 3.1.7. *The relation $\to_{\mathtt{a}}$ on $\Delta_{\mathsf{R}} \times \Delta_{\mathsf{R}}$, extended as a relation on $\Delta_{\mathsf{R}} \times \mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$, is defined as follows:*

$$\to_{\mathtt{a}} = \to \setminus \to_{\mathtt{f}} . \tag{3.1}$$

Relation $\to_{\mathtt{a}}$ can be directly given by induction on terms. In particular, set $s \to_{\mathtt{a}} s'$ as soon as one of the following holds:

- *$s = \lambda x.u$ and $s' = \lambda x.u'$ with $u \to_{\mathtt{a}} u'$;*
- *$s = (u)\, V$ and $s' = (u')\, V$ with $u \to_{\mathtt{a}} u'$;*
- *$s = (u)\, V$ and $s' = (u)\, V'$ with $V \mathrel{\widetilde{\to}} V'$;*

We call argument β-reduction, *or simply* argument reduction, *the relation $\widetilde{\to_{\mathtt{a}}}$.*

The technique we develop to devise the factorisation theorem requires a parallel version of argument reduction, which we denote $\overline{\rightrightarrows_{\mathtt{a}}}$. Even in this case, we only need to write down a standard inductive definition.

Definition 3.1.8. *The relation $\rightrightarrows_{\mathtt{a}}$ on $\Delta_{\mathsf{R}} \times \Delta_{\mathsf{R}}$ is defined by induction on terms. In particular, set $s \rightrightarrows_{\mathtt{a}} s'$ as soon as one of the following holds:*

- *$s = x$ and $s' = x$, for all $x \in \mathcal{V}$;*
- *$s = \lambda x.u$ and $s' = \lambda x.u'$ with $u \rightrightarrows_{\mathtt{a}} u'$;*
- *$s = (u)\, T$ and $s' = (u')\, T'$ with $u \rightrightarrows_{\mathtt{a}} u'$ and $T \mathrel{\overline{\rightrightarrows}} T'$;*

We call parallel argument (β-)reduction *the relation $\overline{\rightrightarrows_{\mathtt{a}}}$.*

Lemma 3.1.9. *It holds that:*

1. *The relations $\widetilde{\to_{\mathtt{a}}}^{*}$ and $\overline{\rightrightarrows_{\mathtt{a}}}$ are contextual;*
2. *$\widetilde{\to_{\mathtt{a}}} \subset \overline{\rightrightarrows_{\mathtt{a}}} \subset \widetilde{\to_{\mathtt{a}}}^{*}$. In particular, $\overline{\rightrightarrows_{\mathtt{a}}}^{*} = \widetilde{\to_{\mathtt{a}}}^{*}$.*

Proof. The first result follows from the contextual properties of reductions $\widetilde{\to}^{*}$ and $\overline{\rightrightarrows}$ (respectively, Propositions 1.4.11 and 1.4.18). The second result follows from how reductions $\widetilde{\to}$ and $\overline{\rightrightarrows}$ are related (Lemma 1.4.24 and Corollary 1.4.25). $\square$

3.1.1 Confluence of function reduction

In Section 3.2 we provide a factorisation theorem for Λ_Σ with respect to function and argument reductions. Thereafter, in Section 3.3, we show that function reduction characterises head normalisability. Since function reduction is non-deterministic, one might wonder whether such results are universal as the classical ones for pure λ-calculus. We now settle this doubt by showing confluence property of function reduction, which ultimately implies the soundness of the results obtained.

It is rather easy to see that relation $\to_{\mathtt{f}}$ enjoys the diamond property. This is already well-known in the realm of pure λ-calculus [Reg94, DR04]. However reduction $\widetilde{\to_{\mathtt{f}}}$ does not, due to the reasons we detailed in Section 1.4.3.

We appeal again to the Tait–Martin-Löf technique, and so we introduce a parallel version of function reduction which is proved to be strongly confluent.

Definition 3.1.10. *The relation $\rightrightarrows_{\mathtt{f}}$ on $\Delta_{\mathsf{R}} \times \mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$ is defined by induction on terms. In particular, set $s \rightrightarrows_{\mathtt{f}} S'$ as soon as one of the following holds:*

- *$s = x$ and $S' = x$, for all $x \in \mathcal{V}$;*
- *$s = \lambda x.u$ and $S' = \lambda x.U'$ with $u \rightrightarrows_{\mathtt{f}} U'$;*
- *$s = (u)\,T$ and $S' = (U')\,T$ with $u \rightrightarrows_{\mathtt{f}} U$;*
- *$s = (\lambda x.u)\,T$ and $S' = U'\,[T/x]$ with $u \rightrightarrows_{\mathtt{f}} U'$.*

We call parallel function (β-)reduction *the relation $\overline{\rightrightarrows_{\mathtt{f}}}$.*

We now assert some properties of parallel function reduction, without proving them (by now, their proof are standard). In particular, contrary to function reduction, its parallel version is stable under substitution as the extension provided by Rule (1.8b) permits to handle the situation we described in the proof of Lemma 3.1.6.

Lemma 3.1.11. *It holds that:*

1. *The relation $\overline{\rightrightarrows_{\mathtt{f}}}$ is function contextual;*
2. *$\widetilde{\to_{\mathtt{f}}} \subset \overline{\rightrightarrows_{\mathtt{f}}} \subset \widetilde{\to_{\mathtt{f}}}^*$. In particular, $\overline{\rightrightarrows_{\mathtt{f}}}^* = \widetilde{\to_{\mathtt{f}}}^*$.*

Lemma 3.1.12. *Let $x \in \mathcal{V}$ and $S, S', T \in \mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$. If $S \;\overline{\rightrightarrows_{\mathtt{f}}}\; S'$, then $S\,[T/x] \;\overline{\rightrightarrows_{\mathtt{f}}}\; S'\,[T/x]$.*

Proof. This is an easy induction on $S \;\overline{\rightrightarrows_{\mathtt{f}}}\; S'$. Notice that, in contrast with what happen in Lemma 3.1.6, the conclusion is expressed in terms of $\overline{\rightrightarrows_{\mathtt{f}}}$. Indeed, if we consider once again $S = (\lambda y.x)\,z$ which reduces to $S' = x$ by one-step of $\widetilde{\to_{\mathtt{f}}}$-reduction (hence by a step of $\overline{\rightrightarrows_{\mathtt{f}}}$ as well), and $T = \sum_{i=1}^{n} a_i t_i$, then the result follows by just one parallel function reduction:

$$S[T/x] = \left(\lambda y.\sum_{i=1}^{n} a_i t_i\right) z = \sum_{i=1}^{n} a_i (\lambda y.t_i) z \overline{\rightrightarrows_{\mathtt{f}}} \sum_{i=1}^{n} a_i t_i = x[T/x] = S'[T/x].$$

This property is crucial for proving Lemma 3.2.4. □

Following what we did in Section 1.4.3, we characterise parallel function reduction as the process of simultanuosly firing all the redexes, this time in function position only, appearing in a term.

Definition 3.1.13. *Let $S^{\odot}$ be the* complete function development of S *inductively defined by:*

$$\begin{aligned}
x^{\odot} &= x; \\
(\lambda x.u)^{\odot} &= \lambda x.u^{\odot}; \\
((u)\,V)^{\odot} &= \left(u^{\odot}\right) V; \text{ (if } s \text{ is not a } \lambda\text{-abstraction)} \\
((\lambda x.u)\,V)^{\odot} &= u^{\odot}[V/x]; \\
\left(\sum_{i=1}^{n} a_i u_i\right)^{\odot} &= \sum_{i=1}^{n} a_i u_i^{\odot}.
\end{aligned}$$

In particular, one can prove that every $\overline{\rightrightarrows_{\mathtt{f}}}$-reduct reduces to its complete function development. We do not provide a proof as it would replicate that of Lemma 1.4.28.

Lemma 3.1.14. *For all $S, S' \in \mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$, if $S \overline{\rightrightarrows_{\mathtt{f}}} S'$ then $S' \overline{\rightrightarrows_{\mathtt{f}}} S^{\odot}$.*

Theorem 3.1.15. *Relation $\overline{\rightrightarrows_{\mathtt{f}}}$ is strongly confluent. Hence, relation $\widetilde{\rightarrow_{\mathtt{f}}}$ enjoys the Church-Rosser property.*

Proof. Relation $\overline{\rightrightarrows_{\mathtt{f}}}$ is strongly confluent as a direct consequence of Lemma 3.1.14. Then, Lemma 3.1.11-2 implies confluence of relation $\widetilde{\rightarrow_{\mathtt{f}}}$. □

3.2 Decompose + Swap = Factorisation

Takahashi's technique [Tak95] for pure λ-calculus provides *factorisation* by first decomposing a parallel reduction sequence into sequences of head reductions punctuated by internal parallel reductions (*decomposition* phase), which are later reorganized into two (sub)sequences that first reduce on the head and then everywhere else (*swap* phase). In this section we transport the factorisation theorem in the setting of the algebraic λ-calculus, by proving these two phases (Lemma 3.2.7 and 3.2.8) with respect to function and argument reductions. After that, we discuss the mismatch between the two versions of the theorem, and we explain why Theorem 3.2.9 is the best we can achieve in Λ_Σ.

Definition 3.2.1. *For all $S, S' \in \mathsf{R}\langle\Delta_\mathsf{R}\rangle$, write $S \Rrightarrow S'$ whenever there exist $n \geq 0$ and $S_0, \ldots, S_n \in \mathsf{R}\langle\Delta_\mathsf{R}\rangle$ such that*

$$S = S_0 \ \overline{\rightrightarrows_\mathtt{f}}\ S_1 \ \overline{\rightrightarrows_\mathtt{f}}\ \ldots \ \overline{\rightrightarrows_\mathtt{f}}\ S_n \ \overline{\rightrightarrows_\mathtt{a}}\ S',$$

and $S_i \ \overline{\rightrightarrows}\ S'$ for all $i \in \{0, \ldots, n\}$.

Lemma 3.2.2. *Let $S, S', T, T' \in \mathsf{R}\langle\Delta_\mathsf{R}\rangle$ with $S \Rrightarrow S'$ and $T \ \overline{\rightrightarrows}\ T'$. Then, $(S)\, T \Rrightarrow (S')\, T'$.*

Proof. By Definition 3.2.1, $S \Rrightarrow S'$ implies $n \geq 0$ and $S_0, \ldots, S_n \in \mathsf{R}\langle\Delta_\mathsf{R}\rangle$ such that

$$S = S_0 \ \overline{\rightrightarrows_\mathtt{f}}\ S_1 \ \overline{\rightrightarrows_\mathtt{f}}\ \ldots \ \overline{\rightrightarrows_\mathtt{f}}\ S_n \ \overline{\rightrightarrows_\mathtt{a}}\ S',$$

and $S_i \ \overline{\rightrightarrows}\ S'$ for all $i \in \{0, \ldots, n\}$. The result follows by the function contextual property of $\overline{\rightrightarrows_\mathtt{f}}$ (Lemma 3.1.11-1), the contextual property of $\overline{\rightrightarrows_\mathtt{a}}$ and $\overline{\rightrightarrows}$ (respectively, Lemma 3.1.9-2 and Proposition 1.4.18), using in the large the definition of $\overline{\rightrightarrows_\mathtt{a}}$ (Definition 3.1.8). In particular, it holds that:

- for all $i \in \{0, \ldots, n-1\}$, $S_i \ \overline{\rightrightarrows_\mathtt{f}}\ S_{i+1}$ implies $(S_i)\, T \ \overline{\rightrightarrows_\mathtt{f}}\ (S_{i+1})\, T$;
- $S_n \ \overline{\rightrightarrows_\mathtt{a}}\ S'$ and $T \ \overline{\rightrightarrows}\ T'$ imply $(S_n)\, T \ \overline{\rightrightarrows_\mathtt{a}}\ (S')\, T'$;
- for all $i \in \{0, \ldots, n\}$, $S_i \ \overline{\rightrightarrows}\ S'$ and $T \ \overline{\rightrightarrows}\ T'$ imply $(S_i)\, T \ \overline{\rightrightarrows}\ (S')\, T'$;

which together imply the thesis, that is $(S)\, T \Rrightarrow (S')\, T'$. $\square$

As usual, we proceed by showing a contextuality result for the new relation $\Rrightarrow$.

Proposition 3.2.3. *The relation $\Rrightarrow$ is contextual.*

Proof. Relation $\Rrightarrow$ is obviously reflexive as relations $\overline{\rightrightarrows_\mathtt{f}}^*$ and $\overline{\rightrightarrows_\mathtt{a}}$ already are. Then, we prove the clauses of Definition 1.4.1 with respect to relation $\Rrightarrow$ as follows:

1. By Definition 3.2.1, $S \Rrightarrow S'$ implies $n \geq 0$ and $S_0, \ldots, S_n \in \mathsf{R}\langle\Delta_\mathsf{R}\rangle$ such that

   $$S = S_0 \ \overline{\rightrightarrows_\mathtt{f}}\ S_1 \ \overline{\rightrightarrows_\mathtt{f}}\ \ldots \ \overline{\rightrightarrows_\mathtt{f}}\ S_n \ \overline{\rightrightarrows_\mathtt{a}}\ S'$$

   and $S_i \ \overline{\rightrightarrows}\ S'$ for all $i \in \{0, \ldots, n\}$. The result follows by the function contextual property of $\overline{\rightrightarrows_\mathtt{f}}$ (Lemma 3.1.11-1) and the contextual property of $\overline{\rightrightarrows_\mathtt{a}}$ and $\overline{\rightrightarrows}$ (respectively, Lemma 3.1.9-2 and Proposition 1.4.18), using in the large the definition of $\overline{\rightrightarrows_\mathtt{a}}$ (Definition 3.1.8). In particular, it holds that:

   - for all $i \in \{0, \ldots, n-1\}$, $S_i \ \overline{\rightrightarrows_\mathtt{f}}\ S_{i+1}$ implies $\lambda x.S_i \ \overline{\rightrightarrows_\mathtt{f}}\ \lambda x.S_{i+1}$;

- $S_n \overline{\rightrightarrows_{\mathtt{a}}} S'$ implies $\lambda x.S_n \overline{\rightrightarrows_{\mathtt{a}}} \lambda x.S'$;
- for all $i \in \{0, \ldots, n\}$, $S_i \overline{\rightrightarrows} S'$ implies $\lambda x.S_i \overline{\rightrightarrows} \lambda x.S'$;

  which together imply the thesis, that is $\lambda x.S \Rrightarrow \lambda x.S'$.

2. By Definition 3.2.1, $T \Rrightarrow T'$ implies $T \overline{\rightrightarrows} T'$ and the thesis directly follows by Lemma 3.2.2.

3. By Definition 3.2.1, $S \Rrightarrow S'$ implies $n \geq 0$ and $S_0, \ldots, S_n \in \mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$ such that

$$S = S_0 \overline{\rightrightarrows_{\mathtt{f}}} S_1 \overline{\rightrightarrows_{\mathtt{f}}} \ldots \overline{\rightrightarrows_{\mathtt{f}}} S_n \overline{\rightrightarrows_{\mathtt{a}}} S'$$

   and $S_i \overline{\rightrightarrows} S'$ for all $i \in \{0, \ldots, n\}$. The result follows by the same reasons of the first case. In particular, it holds that:

   - for all $i \in \{0, \ldots, n-1\}$, $S_i \overline{\rightrightarrows_{\mathtt{f}}} S_{i+1}$ implies $aS_i \overline{\rightrightarrows_{\mathtt{f}}} aS_{i+1}$;
   - $S_n \overline{\rightrightarrows_{\mathtt{a}}} S'$ implies $aS_n \overline{\rightrightarrows_{\mathtt{a}}} aS'$;
   - for all $i \in \{0, \ldots, n\}$, $S_i \overline{\rightrightarrows} S'$ implies $aS_i \overline{\rightrightarrows} aS'$;

   which together imply the thesis, that is $aS \Rrightarrow aS'$.

4. By Definition 3.2.1, $S \Rrightarrow S'$ and $T \Rrightarrow$ T' respectively imply $m, n \geq 0$ and

   - $S_0, \ldots, S_m \in \mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$ such that $S = S_0 \overline{\rightrightarrows_{\mathtt{f}}} S_1 \overline{\rightrightarrows_{\mathtt{f}}} \ldots \overline{\rightrightarrows_{\mathtt{f}}} S_m \overline{\rightrightarrows_{\mathtt{a}}} S'$ and $S_i \overline{\rightrightarrows} S'$ for all $i \in \{0, \ldots, m\}$;
   - $T_0, \ldots, T_n \in \mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$ such that $T = T_0 \overline{\rightrightarrows_{\mathtt{f}}} T_1 \overline{\rightrightarrows_{\mathtt{f}}} \ldots \overline{\rightrightarrows_{\mathtt{f}}} T_n \overline{\rightrightarrows_{\mathtt{a}}} T'$ and $T_j \overline{\rightrightarrows} T'$ for all $j \in \{0, \ldots, n\}$.

   The result follows by the same reasons of the first case. In particular, it holds that:

   - for all $i \in \{0, \ldots, m-1\}$ and all $j \in \{0, \ldots, n-1\}$, $S_i \overline{\rightrightarrows_{\mathtt{f}}} S_{i+1}$ and $T_i \overline{\rightrightarrows_{\mathtt{f}}} T_{j+1}$ imply $S_i + T_0 \overline{\rightrightarrows_{\mathtt{f}}} S_{i+1} + T_0$ and $S_m + T_j \overline{\rightrightarrows_{\mathtt{f}}} S_m + T_{j+1}$, hence $S_0 + T_0 \overline{\rightrightarrows_{\mathtt{f}}}^* S_m + T_n$;
   - $S_m \overline{\rightrightarrows_{\mathtt{a}}} S'$ and $T_n \overline{\rightrightarrows_{\mathtt{a}}} T'$ imply $S_m + T_n \overline{\rightrightarrows_{\mathtt{a}}} S' + T'$;
   - for all $i \in \{0, \ldots, m\}$ and $j \in \{0, \ldots, n\}$, $S_i \overline{\rightrightarrows} S'$ and $T_j \overline{\rightrightarrows} T'$ imply $S_i + T_j \overline{\rightrightarrows} S' + T'$;

   which together imply the thesis, that is $S + T \Rrightarrow S' + T'$.

This concludes the proof. $\square$

Lemma 3.2.4. *Let $x \in \mathcal{V}$ and $S, S', T, T' \in \mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$. If $S \Rrightarrow S'$ and $T \Rrightarrow T'$, then $S\,[T/x] \Rrightarrow S'\,[T'/x]$.*

We show the above lemma as a direct consequence of the following ones.

Lemma 3.2.5. *Let $x \in \mathcal{V}$ and $S, T, T' \in \mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$. If $T \Rrightarrow T'$, then $S\,[T/x] \Rrightarrow S\,[T'/x]$.*

Proof. Direct consequence of Lemma 1.4.3 and the contextual property of relation $\Rrightarrow$ (Proposition 3.2.3). □

Lemma 3.2.6. *Let $x \in \mathcal{V}$ and $S, S', T, T' \in \mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$. If $S \overline{\rightrightarrows_{\mathsf{a}}} S'$ and $T \Rrightarrow T'$, then $S\,[T/x] \Rrightarrow S'\,[T'/x]$.*

Proof. By induction on $S \overline{\rightrightarrows_{\mathsf{a}}} S'$ as we inspect the possible cases for such reduction. If $S = S'$, then the result follows by Lemma 3.2.5. Otherwise, we first address the cases in which S is a simple term s (and so S' is s', by Definition 3.1.8) and $s \rightrightarrows_{\mathsf{a}} s'$. Then, one of the following applies:

- $s = x$ and $s' = x$; hence the result directly follows. The result vacuously follows whenever $s = y = s'$, for all $y \in \mathcal{V}$ and $y \neq x$.

- $s = \lambda y.u$ and $s' = \lambda y.u'$ with $u \rightrightarrows_{\mathsf{a}} u'$; hence, by the induction hypothesis, $u\,[T/y] \Rrightarrow u'\,[T'/y]$ and the contextual property of $\Rrightarrow$ (Proposition 3.2.3) implies

$$s\,[T/x] = (\lambda y.u)\,[T/x] = \lambda y.u\,[T/x] \\ \Rrightarrow \lambda y.u'\,\left[T'/x\right] = (\lambda y.u')\,\left[T'/x\right] = s'\,\left[T'/x\right].$$

- $s = (u)\,V$ and $s' = (u')\,V'$, with $u \rightrightarrows_{\mathsf{a}} u'$ and $V \overline{\rightrightarrows} V'$; hence, by the induction hypothesis, $u\,[T/x] \Rrightarrow u'\,[T'/x]$. Moreover, Definition 3.2.1 on $T \Rrightarrow T'$ entails $T \overline{\rightrightarrows} T'$ and so, by Lemma 1.4.20, $V\,[T/x] \overline{\rightrightarrows} V'\,[T'/x]$. Lemma 3.2.2 implies

$$\begin{aligned} s\,[T/x] = ((u)\,V)\,[T/x] &= (u\,[T/x])\,V\,[T/x] \\ &\Rrightarrow \left(u'\,\left[T'/x\right]\right)V'\,\left[T'/x\right] \\ &= \left(\left(u'\right)V'\right)\left[T'/x\right] = s'\,\left[T'/x\right]. \end{aligned}$$

Now let S be a term and $S \overline{\rightrightarrows_{\mathsf{a}}} S'$. By definition, this amount to the following: $S = \sum_{i=1}^{n} a_i u_i$ and $S' = \sum_{i=1}^{n} a_i u_i'$, with $u_i \rightrightarrows_{\mathsf{a}} u_i'$ for all $i \in \{1, \ldots, n\}$. From what we have just shown in the case of simple terms, the latter implies $u_i\,[T/x] \Rrightarrow u_i'\,[T'/x]$

for all $i \in \{1, \ldots, n\}$. The contextual property of $\Rrightarrow$ (Proposition 3.2.3) entails

$$S\,[T/x] = \left(\sum_{i=1}^{n} a_i u_i\right)[T/x] = \sum_{i=1}^{n} a_i u_i\,[T/x]$$
$$\Rrightarrow \sum_{i=1}^{n} a_i u_i'\,[T'/x] = \left(\sum_{i=1}^{n} a_i u_i'\right)[T/x] = S'\,[T'/y]\,.$$

This concludes the proof. $\square$

We now proceed to prove Lemma 3.2.4.

Proof (Lemma 3.2.4). By Definition 3.2.1, $S \Rrightarrow S'$ implies $n \geq 0$ and $S_0, \ldots, S_n \in \mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$ such that

$$S = S_0 \overline{\rightrightarrows_{\mathtt{f}}}\, S_1 \overline{\rightrightarrows_{\mathtt{f}}} \ldots \overline{\rightrightarrows_{\mathtt{f}}}\, S_n \overline{\rightrightarrows_{\mathtt{a}}}\, S'$$

and $S_i \overline{\rightrightarrows} S'$ for all $i \in \{0, \ldots, n\}$. Then, Lemma 3.1.12 entails

$$S_0\,[T/x] \overline{\rightrightarrows_{\mathtt{f}}}\, S_1\,[T/x] \overline{\rightrightarrows_{\mathtt{f}}} \ldots \overline{\rightrightarrows_{\mathtt{f}}}\, S_n\,[T/x]$$

and Lemma 3.2.6 on $S_n \overline{\rightrightarrows_{\mathtt{a}}}\, S'$ and $T \Rrightarrow T'$ entails $S_n\,[T/x] \Rrightarrow S'\,[T'/x]$. That is:

$$S\,[T/x] = S_0\,[T/x] \overline{\rightrightarrows_{\mathtt{f}}}\, S_1\,[T/x] \overline{\rightrightarrows_{\mathtt{f}}} \ldots \overline{\rightrightarrows_{\mathtt{f}}}\, S_n\,[T/x] \Rrightarrow S'\,[T'/x]$$

Moreover $T \Rrightarrow T'$ implies $T \overline{\rightrightarrows} T'$, so that Lemma 1.4.20 entails $S_i\,[T/x] \overline{\rightrightarrows} S'\,[T'/x]$ for all $i \in \{1, \ldots, n\}$. It follows $S\,[T/x] \Rrightarrow S'\,[T'/x]$ (Definition 3.2.1 of $\Rrightarrow$). $\square$

We now prove the crucial two phases of Takahashi's technique. We begin with the phase called *decomposition*, by which a parallel reduction step is decomposed into a sequence of parallel function reductions and a step of parallel argument one.

Lemma 3.2.7. *For all $S, S' \in \mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$ with $S \overline{\rightrightarrows} S'$, there exists $S'' \in \mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$ such that $S \overline{\rightrightarrows_{\mathtt{f}}}^{*} S'' \overline{\rightrightarrows_{\mathtt{a}}}\, S'$.*

Proof. By induction on k that $S \overline{\rightrightarrows_k} S'$ implies $S \Rrightarrow S'$. If $k = 0$, then $S' = S$ and the result directly follows by reflexivity of relation $\Rrightarrow$ (Proposition 3.2.3). Suppose the result holds for some k, then we extend it to $k+1$ by inspecting the possible cases for reducing $S \overline{\rightrightarrows_{k+1}} S'$. We first address the cases in which S is a simple term s and $s \rightrightarrows_{k+1} S'$. Then, one of the following applies:

- $s \in \mathcal{V}$; hence the result follows by reflexivity of $\Rrightarrow$ (Proposition 3.2.3).
- $s = \lambda x.u$ and $S' = \lambda x.U'$ with $u \rightrightarrows_k U'$; hence, by the induction hypothesis, $u \Rrightarrow U'$. By the contextual property of $\Rrightarrow$ (Proposition 3.2.3) follows

$$s = \lambda x.u \Rrightarrow \lambda x.U' = S'.$$

- $s = (u)\,T$ and $S' = (U')\,T'$ with $u \rightrightarrows_k U'$ and $T \overline{\rightrightarrows_k} T'$; hence, by a double induction hypothesis, $u \Rrightarrow U'$ and $T \Rrightarrow T'$. By the contextual property of $\Rrightarrow$ (Proposition 3.2.3) follows

$$s = (u)\,T \Rrightarrow \left(U'\right) T' = S'.$$

- $s = (\lambda x.u)\,T$ and $S' = U'\,[T'/x]$ with $u \rightrightarrows_k U'$ and $T \overline{\rightrightarrows_k} T'$; hence, by a double induction hypothesis, $u \Rrightarrow U'$ and $T \Rrightarrow T'$. Lemma 3.2.4 on these latter entails $u\,[T/x] \Rrightarrow U'\,[T'/x]$. Then, along with the fact that $(\lambda x.u)\,T \widetilde{\rightarrow_{\mathtt{f}}} u\,[T/x]$, hence $(\lambda x.u)\,T \overline{\rightrightarrows_{\mathtt{f}}} u\,[T/x]$ (Lemma 3.1.11-2), it follows

$$s = (\lambda x.u)\,T \Rrightarrow U'\left[T'/x\right] = S'.$$

Now assume $S \overline{\rightrightarrows_{k+1}} S'$. By definition, this amount to the following: $S = \sum_{i=1}^{n} a_i u_i$ and $S' = \sum_{i=1}^{n} a_i U_i'$, with $u_i \rightrightarrows_{k+1} U_i'$ for all $i \in \{1, \ldots, n\}$. From what we have just shown in the case of simple terms, the latter implies $u_i \Rrightarrow U_i'$ for all $i \in \{1, \ldots, n\}$. By the contextual property of $\Rrightarrow$ (Proposition 3.2.3) follows

$$S = \sum_{i=1}^{n} a_i u_i \Rrightarrow \sum_{i=1}^{n} a_i U_i' = S'.$$

This concludes the proof. $\square$

We proceed with the phase called *swap*, by which parallel argument reductions can be postponed in favor of parallel function reductions.

Lemma 3.2.8. *For all $S, S', T \in \mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$ with $S \overline{\rightrightarrows_{\mathtt{a}}} S' \overline{\rightrightarrows_{\mathtt{f}}} T$, there exists $U \in \mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$ such that $S \overline{\rightrightarrows_{\mathtt{f}}}^{+} U \overline{\rightrightarrows_{\mathtt{a}}} T$.*

Proof. We prove that $S \overline{\rightrightarrows_{\mathtt{a}}} S' \overline{\rightrightarrows_{\mathtt{f}}} T$ implies $S \overline{\rightrightarrows_{\mathtt{f}}} V \Rrightarrow T$, for some $V \in \mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$, by induction on $S \overline{\rightrightarrows_{\mathtt{a}}} S'$, and then by cases on $S' \overline{\rightrightarrows_{\mathtt{f}}} T$. If $S' = S$, then the result follows by considering $V = T$. Otherwise, we first address the cases in which S is a simple term s (and so S' is s', by Definition 3.1.8) with $s \rightrightarrows s'$, and so $s' \overline{\rightrightarrows_{\mathtt{f}}} T$. Then, one of the following applies:

- $s \in \mathcal{V}$; hence the result directly follows.

- $s = \lambda x.u$ and $s' = \lambda x.u'$ with $u \rightrightarrows_{\mathtt{a}} u'$. Definition of $\overline{\rightrightarrows_{\mathtt{f}}}$ (Definition 3.1.10 and Rule (1.8b)) on $s' \overline{\rightrightarrows_{\mathtt{f}}} T$ entails $T = \lambda x.T'$ with $u' \overline{\rightrightarrows_{\mathtt{f}}} T'$: *i.e.* $u \rightrightarrows_{\mathtt{a}} u' \overline{\rightrightarrows_{\mathtt{f}}} T'$. Hence, by the induction hypothesis, there exists $V' \in \mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$ such that $u \overline{\rightrightarrows_{\mathtt{f}}} V' \Rrightarrow T'$. Then, consider $V = \lambda x.V'$. By the function contextual property of $\overline{\rightrightarrows_{\mathtt{f}}}$ (Lemma 3.1.11-1) and the contextual property of $\Rrightarrow$ (Proposition 3.2.3) follows

$$\lambda x.u \;\overline{\rightrightarrows_{\mathtt{f}}}\; \lambda x.V' \Rrightarrow \lambda x.T',$$

i.e. $s \;\overline{\rightrightarrows_{\mathtt{f}}}\; V \Rrightarrow T$.

- $s = (u)\,W$ and $s' = (u')\,W'$ with u not a λ-abstraction, $u \rightrightarrows_{\mathtt{a}} u'$ and $W \;\overline{\rightrightarrows}\; W'$. Definition of $\rightrightarrows_{\mathtt{a}}$ (Definition 3.1.8) entails u' not a λ-abstraction and so definition of $\overline{\rightrightarrows_{\mathtt{f}}}$ (Definition 3.1.10 and Rule (1.8b)) on $s' \;\overline{\rightrightarrows_{\mathtt{f}}}\; T$ implies $T = (T')\,W'$ with $u' \;\overline{\rightrightarrows_{\mathtt{f}}}\; T'$: *i.e.* $u \rightrightarrows_{\mathtt{a}} u' \;\overline{\rightrightarrows_{\mathtt{f}}}\; T'$. Hence, by the induction hypothesis, there exists $V' \in \mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$ such that $u \;\overline{\rightrightarrows_{\mathtt{f}}}\; V' \Rrightarrow T'$. Moreover, Lemma 3.2.7 on $W \;\overline{\rightrightarrows}\; W'$ implies $W \Rrightarrow W'$. Then, consider $V = (V')\,W$. By the function contextual property of $\overline{\rightrightarrows_{\mathtt{f}}}$ (Lemma 3.1.11-1) and the contextual property of $\Rrightarrow$ (Proposition 3.2.3) follows

$$(u)\,W \;\overline{\rightrightarrows_{\mathtt{f}}}\; \left(V'\right) W \Rrightarrow \left(T'\right) W',$$

i.e. $s \;\overline{\rightrightarrows_{\mathtt{f}}}\; V \Rrightarrow T$.

- $s = (\lambda x.u)\,W$ and $s' = (\lambda x.u')\,W'$ with $u \rightrightarrows_{\mathtt{a}} u'$ and $W \;\overline{\rightrightarrows}\; W'$. Definition of $\overline{\rightrightarrows_{\mathtt{f}}}$ (Definition 3.1.10 and Rule (1.8b)) on $s' \;\overline{\rightrightarrows_{\mathtt{f}}}\; T$ entails two subcases:

  - $T = (\lambda x.T')\,W'$ with $u' \;\overline{\rightrightarrows_{\mathtt{f}}}\; T'$: *i.e.* $u \rightrightarrows_{\mathtt{a}} u' \;\overline{\rightrightarrows_{\mathtt{f}}}\; T'$. Hence, by the induction hypothesis, there exists $V' \in \mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$ such that $u \;\overline{\rightrightarrows_{\mathtt{f}}}\; V' \Rrightarrow T'$. Moreover, Lemma 3.2.7 on $W \;\overline{\rightrightarrows}\; W'$ implies $W \Rrightarrow W'$. Then, consider $V = (\lambda x.V')\,W$ and verify that $\lambda x.u \;\overline{\rightrightarrows_{\mathtt{f}}}\; \lambda x.V' \Rrightarrow \lambda x.T'$ (respectively, Lemma 3.1.11-1 and Proposition 3.2.3). By the function contextual property of $\overline{\rightrightarrows_{\mathtt{f}}}$ (Lemma 3.1.11-1) and the contextual property of $\Rrightarrow$ (Proposition 3.2.3) follows

    $$(\lambda x.u)\,W \;\overline{\rightrightarrows_{\mathtt{f}}}\; \left(\lambda x.V'\right) W \Rrightarrow \left(\lambda x.T'\right) W',$$

    i.e. $s \;\overline{\rightrightarrows_{\mathtt{f}}}\; V \Rrightarrow T$.

  - $T = U''\,[W'/x]$ with $u' \;\overline{\rightrightarrows_{\mathtt{f}}}\; U''$: *i.e.* $u \rightrightarrows_{\mathtt{a}} u' \;\overline{\rightrightarrows_{\mathtt{f}}}\; U''$. Hence, by the induction hypothesis, there exists $V' \in \mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$ such that $u \rightrightarrows_{\mathtt{f}} V' \Rrightarrow U''$. Moreover, Lemma 3.2.7 on $W \;\overline{\rightrightarrows}\; W'$ implies $W \Rrightarrow W'$, and so Lemma 3.2.4 on $V' \Rrightarrow U''$ and $W \Rrightarrow W'$ entails $V'\,[W/x] \Rrightarrow U''\,[W'/x]$. Then, consider $V = V'\,[W/x]$. By the definition of $\overline{\rightrightarrows_{\mathtt{f}}}$ (Definition 3.1.10 and Rule (1.8b)) follows

    $$(\lambda x.u)\,W \;\overline{\rightrightarrows_{\mathtt{f}}}\; V'\,[W/x] \Rrightarrow U''\left[W'/x\right],$$

    i.e. $s \;\overline{\rightrightarrows_{\mathtt{f}}}\; V \Rrightarrow T$.

Now let S be a term and $S \overline{\rightrightarrows_{\mathtt{a}}} S' \overline{\rightrightarrows_{\mathtt{f}}} T$. By definition, this amount to the following: $S = \sum_{i=1}^{n} a_i u_i$, $S' = \sum_{i=1}^{n} a_i u_i'$ and $T = \sum_{i=1}^{n} a_i U_i''$ with $u_i \rightrightarrows_{\mathtt{a}} u_i'$ and $u_i' \overline{\rightrightarrows_{\mathtt{f}}} U_i''$ for all $i \in \{1, \ldots, n\}$. That is, for all $i \in \{1, \ldots, n\}$, $u_i \rightrightarrows_{\mathtt{a}} u_i' \overline{\rightrightarrows_{\mathtt{f}}} U_i''$; hence, from what we have just shown in the case of simple terms, there exist $V_i' \in \mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$ such that $u_i \overline{\rightrightarrows_{\mathtt{f}}} V_i' \Rrightarrow U_i''$. Then, consider $V = \sum_{i=1}^{n} a_i V_i'$. By the definition of $\overline{\rightrightarrows_{\mathtt{f}}}$ (Definition 3.1.10 and Rule (1.8b)) and the contextual property of $\Rrightarrow$ (Proposition 3.2.3) follows

$$\sum_{i=1}^{n} a_i u_i \overline{\rightrightarrows_{\mathtt{f}}} \sum_{i=1}^{n} a_i V_i' \Rrightarrow \sum_{i=1}^{n} a_i U_i'',$$

i.e. $S \overline{\rightrightarrows_{\mathtt{f}}} V \Rrightarrow T$. □

Decomposition (Lemma 3.2.7) and swap (Lemma 3.2.8) together imply the *factorisation theorem* for the algebraic λ-calculus.

Theorem 3.2.9. *For all $S, T \in \mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$ with $S \widetilde{\to}^* T$, it holds that $S \widetilde{\to}_{\mathtt{f}}^* \widetilde{\to}_{\mathtt{a}}^* T$.*

Proof. The hypothesis $S \widetilde{\to}^* T$ implies $S_0, \ldots, S_n \in \mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$ such that

$$S = S_0 \widetilde{\to} S_1 \widetilde{\to} \ldots \widetilde{\to} S_{n-1} \widetilde{\to} S_n = T.$$

Since $\widetilde{\to} \subset \overline{\rightrightarrows}$ (Lemma 1.4.24), the same holds with $\overline{\rightrightarrows}$ in place of $\widetilde{\to}$: *i.e.*

$$S = S_0 \overline{\rightrightarrows} S_1 \overline{\rightrightarrows} \ldots \overline{\rightrightarrows} S_{n-1} \overline{\rightrightarrows} S_n = T.$$

For all $i \in \{0, \ldots, n-1\}$, Lemma 3.2.7 on $S_i \overline{\rightrightarrows} S_{i+1}$ entails $S_i \overline{\rightrightarrows_{\mathtt{f}}}^* \overline{\rightrightarrows_{\mathtt{a}}} S_{i+1}$: *i.e.*

$$S = S_0 \overline{\rightrightarrows_{\mathtt{f}}}^* \overline{\rightrightarrows_{\mathtt{a}}} S_1 \overline{\rightrightarrows_{\mathtt{f}}}^* \overline{\rightrightarrows_{\mathtt{a}}} \ldots \overline{\rightrightarrows_{\mathtt{f}}}^* \overline{\rightrightarrows_{\mathtt{a}}} S_{n-1} \overline{\rightrightarrows_{\mathtt{f}}}^* \overline{\rightrightarrows_{\mathtt{a}}} S_n = T.$$

Iterating Lemma 3.2.8 entails $S \overline{\rightrightarrows_{\mathtt{f}}}^* \overline{\rightrightarrows_{\mathtt{a}}}^* T$, and so the result knowing that $\overline{\rightrightarrows_{\mathtt{f}}}^* = \widetilde{\to}_{\mathtt{f}}^*$ and $\overline{\rightrightarrows_{\mathtt{a}}}^* = \widetilde{\to}_{\mathtt{a}}^*$ (respectively, Lemma 3.1.11-2 and Lemma 3.1.9-2). □

The factorisation theorem we achieved (Theorem 3.2.9) is weaker than the well-known one appearing in the literature of the pure λ-calculus [Bar84, Tak95, Mel97]. While the latter is expressed in terms of a proper strategy, which deterministically always reduces the head redex, our result is not. Indeed, function reduction is by definition non-deterministic.

Unfortunately, although head reduction is a function reduction, Theorem 3.2.9 turns out to be the best we can do in the setting of the algebraic λ-calculus: it cannot be strengthened to use only head reductions.

This is disappointing, as in pure λ-calculus the factorisation theorem directly entails two import results: the *standardisation theorem* and the *leftmost reduction* *theorem* [Bar84, Tak95]. More strikingly, a classical formulation of the standardisation theorem is out of reach: since a head redex is always the leftmost one, if it is contracted in the course of a reduction sequence, then it cannot be factorised and the reduction sequence cannot be subject of standardisation.

We highlight the main difficulty in adapting these classical results to Λ_Σ by showing an example of reduction sequence which does not admit a head factorisation.

***Example* 3.2.10.** Consider $s \in \Delta_{\mathsf{R}}$ such that $s \to u + v$, and the following reduction sequence:

$$(\lambda x.s)\, y \,\widetilde{\to}\, (\lambda x.(u+v))\, y = (\lambda x.u)\, y + (\lambda x.v)\, y \,\widetilde{\to}\, u\,[y/x] + (\lambda x.v)\, y. \tag{3.2}$$

Head factorisation, as well as the classical notion of standard reduction, would instead proceed as follows:

$$(\lambda x.s)\, y \,\widetilde{\to}\, s\,[y/x] \;\widetilde{\to}\, (u+v)\,[y/x] = u\,[y/x] + v\,[y/x]\,, \tag{3.3}$$

namely by first firing the head redex $(\lambda x.s)\, y$, and reducing s afterwards.

Since a head reduction causes the loss of the redex $(\lambda x.v)\, y$ as a side effect, it is clear that there is no way a sequence of head reductions can produce the result of reduction (3.2). The same holds for classical standard reduction.

Example 3.2.10 reveals that head reduction does not permute with the other function reductions (typically called *internal*), in general. This obviously prevents the adaptation of those classical results that require reductions to be transformed by permuting their steps according to a precise order. This is the case of factorisation and standardisation theorems in λ-calculus.

***Remark* 3.2.11.** The fact that head reduction does not permute with internal ones marks once again the non-trivial gap between the dynamics of the algebraic λ-calculus and the pure λ-calculus. As a matter of fact, head reduction is a crucial notion in pure λ-calculus and the factorisation theorem emphasises it as the "[...] *efficient part of a computation that can always be separated from its junk*" [Mel97]. Indeed, it is well-known that head redexes cannot be erased nor duplicated.

This is no more true in Λ_Σ: head redexes can be erased or duplicated by exploiting the algebraic component of the calculus. As an example, consider the algebraic term $(\lambda x.(\lambda y.y)\, S)\, z$, and verify the function reduction $(\lambda x.(\lambda y.y)\, S)\, z \,\widetilde{\to}_{\mathtt{f}}\, (\lambda x.S)\, z$. Now, the head redex is erased whenever $S = \mathbf{0}$, whereas "duplicated" whenever $S = \sum_{i=1}^{n} a_i s_i$ with $n \neq 0$. This is already present in the differential λ-calculus [ER03] and, partially, in its derived resource λ-calculus [PT09].

3.3 Head normalisability

In pure λ-calculus, factorisation theorem directly entails the *leftmost reduction theorem* [Bar84] stating that a normal form, if it exists, can always be found by repeatedly contracting the leftmost redex. In fact, Theorem 3.2.9 would be sufficient in that setting as function reduction $\rightarrow_{\mathtt{f}}$ enjoys the diamond property [Reg94].

Our weaker formulation of the factorisation theorem prevents us to directly assert that the leftmost reduction is a normalising strategy in Λ_{Σ}. Although we believe the latter is true anyway, we cannot actualise the idea described at the end of Section 2.4.3 and recalled at the very beginning of this chapter.

Nonetheless, the factorisation theorem (Theorem 3.2.9) is sufficient to characterise head normalisability in terms of function reduction. We first define the free R-module of algebraic terms in head normal form, and then we prove that function reductions are sufficient to attain head normal forms. This intuitively holds given that simple terms in function position are pure λ-terms and Definition 3.1.2 of function reduction precisely matches the one of pure λ-calculus, which indeed characterises head normalisability.

Since terms are subject to algebraic linearity, simple terms in function position share the same structure of pure terms. Therefore, the classical notion of *head normal form* adapts effortlessly to the algebraic λ-calculus as follows:

Definition 3.3.1. *The set* $\mathsf{HNF}_{\mathsf{R}}(k)$ *of* simple head normal forms of height at most k *is defined by induction on k: let* $\mathsf{HNF}_{\mathsf{R}}(0) = \emptyset$*; the elements of* $\mathsf{HNF}_{\mathsf{R}}(k+1)$ *are defined from those of* $\mathsf{HNF}_{\mathsf{R}}(k)$ *by the following clauses:*

- *if* $s \in \mathsf{HNF}_{\mathsf{R}}(k)$*, then* $s \in \mathsf{HNF}_{\mathsf{R}}(k+1)$*;* [Monotonicity]
- *if* $x \in \mathcal{V}$*, then* $x \in \mathsf{HNF}_{\mathsf{R}}(k+1)$*;* [Variable]
- *if* $x \in \mathcal{V}$ *and* $s \in \mathsf{HNF}_{\mathsf{R}}(k)$*, then* $\lambda x.s \in \mathsf{HNF}_{\mathsf{R}}(k+1)$*;* [Abstraction]
- *if* $s \in \mathsf{HNF}_{\mathsf{R}}(k)$ *not a λ-abstraction and* $T \in \mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$*,*
  then $(s)\,T \in \mathsf{HNF}_{\mathsf{R}}(k+1)$*.* [Application]

The set of all simple head normal forms *is defined as* $\mathsf{HNF}_{\mathsf{R}} = \bigcup_{k\in\mathbb{N}} \mathsf{HNF}_{\mathsf{R}}(k)$*, whereas the set of* head normal forms *is given by* $\mathsf{R}\langle\mathsf{HNF}_{\mathsf{R}}\rangle = \bigcup_{k\in\mathbb{N}} \mathsf{R}\langle\mathsf{HNF}_{\mathsf{R}}(k)\rangle$*. Moreover, a term S is said to be* head normalisable *if S is reducible to a term in* $\mathsf{R}\langle\mathsf{HNF}_{\mathsf{R}}\rangle$*, and such term is written as* $\mathsf{HNF}(S)$*.*

Dealing with head normalisability requires the same precautions we detailed in Section 2.1 about the algebraic properties of the semiring of coefficients R. In particular, notice that the argument leading to the collapse of reduction $\widetilde{\rightarrow}$

(Lemma 2.1.1 and Proposition 2.1.2) can be equally reproduced in terms of reduction $\widetilde{\rightarrow}_{\mathtt{f}}$. In view of those problems, Corollary 3.3.2 assumes R to be positive.

Corollary 3.3.2. *For all head normalisable $S \in \mathsf{P}\langle\Delta_{\mathsf{P}}\rangle$, it holds that $S \mathrel{\widetilde{\rightarrow}_{\mathtt{f}}}^{*} \mathsf{HNF}(S)$.*

Proof. Theorem 3.2.9 implies that there exists T such that $S \mathrel{\widetilde{\rightarrow}_{\mathtt{f}}}^{*} T \mathrel{\widetilde{\rightarrow}_{\mathtt{a}}}^{*} \mathsf{HNF}(S)$. Hence the result holds by the fact that, whenever $T \mathrel{\widetilde{\rightarrow}_{\mathtt{a}}}^{*} U$, U is a head normal form if and only if T already is. □

***Remark* 3.3.3.** Generally speaking, the literature concerning λ-calculi extended with (various notions of) "sums" considers different notions of normalisability [dP95, PT09, PR10], named as *may* (or optimistic) and *must* (or pessimistic), respectively. The former is a relaxed version of the latter as the *may* approach requires only one element of the sum to be normalisable. In contrast, the *must* approach requires every element of the sum to be normalisable.

Definition 3.3.1 follows the must approach (as this entire work does, in general), although may-head normalisability is also conceivable. Obviously, Corollary 3.3.2 would be valid in this latter case as well.

Part II

Coinductive equivalences in a probabilistic scenario

Chapter 4

Probabilistic applicative bisimulation

This chapter presents our first contributions concerning probabilistic computation.

We introduce probability in λ-calculus by recalling Dal Lago and Zorzi's [DLZ12] idea of endowing non-deterministic λ-calculus with a probabilistic operational semantics. Here the *lazy* regime is investigated, where closed terms as programs evaluates not to a single value but rather to a probability distribution of values.

On this probabilistic λ-calculus, we study operational techniques for understanding and reasoning about program equality. In particular, we investigate coinductive techniques to characterise context equivalence, which is known to be a challenging matter in higher-order languages.

We first adapt Abramsky's [Abr90] *applicative bisimulation* to this setting, and we show a technique for proving congruence of *probabilistic applicative bisimilarity*. While the technique follows Howe's method [How96], some of the technicalities are quite different, relying on non-trivial "disentangling" properties for sets of real numbers.

Being a congruence, applicative bisimilarity is sound with respect to context equivalence. However, full abstraction fails. To show this, we provide a counterexample in terms of *probabilistic CIU-equivalence*.

Contents

4.1 Probabilistic operational semantics for λ-calculus **105**
4.1.1 Pure, non-deterministic λ-calculus 105
4.1.2 Value distributions and Call-by-name operational semantics 106
4.2 Probabilistic bisimulation **109**
4.3 Probabilistic applicative bisimulation and Howe's method **115**

4.3.1 Probabilistic applicative bisimulation 115
4.3.2 $\Lambda_{\oplus}$-relations 120
4.3.3 Howe's construction 124
4.4 Proof of the Key Lemma **133**
4.4.1 Disentangling probability assignments 134
4.5 Relating Applicative bisimilarity and Context equivalence **142**
4.5.1 Probabilistic context equivalence 142
4.5.2 "Context-free" context equivalence 147
4.5.3 Probabilistic CIU-equivalence 149

The first section introduces Dal Lago and Zorzi's [DLZ12] probabilistic operational semantics on top of non-deterministic λ-calculus, following the so-called call-by-name discipline. The notion of value distribution is defined, along with big-step and small-step inference systems to characterise term convergence behaviour by means of value distributions.

The second section presents bisimulation on labelled Markov chains [DEP02a, Pan11]. Both probabilistic simulation and bisimulation are provided on the basis of Larsen and Skou's influential paper [LS91]. Contrary to what happen in a non-deterministic setting, similarity equivalence is shown to coincide with bisimilarity.

The third section presents the probabilistic λ-calculus, previously introduced, as a labelled Markov chain. This latter provides the ground on top of which the probabilistic variant of Abramsky's [Abr90] applicative (bi)simulation is defined as an instance of probabilistic bisimulation notion. Finally, Howe's method [How96] is set up by following the excellent survey provided by Pitts [Pit11].

The forth section is crucial for proving the congruence property of probabilistic bisimilarity. In particular the so-called Key Lemma, the steppingstone of Howe's method, is shown to hold. The proof turns out to be much more difficult than the one for deterministic and non-deterministic cases. In particular, it relies on the Max-flow Min-cut theorem to "disentangle" sets of real numbers.

Finally, the fifth section shows that probabilistic bisimilarity and context equivalence do not coincide, as the former is proved to be strictly finer than the latter. A counterexample is given, along with the needed technical machinery involving the notion of *CIU-equivalence*.

4.1 Probabilistic operational semantics for λ-calculus

In this section we introduce the set of terms of the probabilistic λ-calculus, denoted $\Lambda_\oplus$, along with its (probabilistic) operational semantics. To be precise, as the denotation $\Lambda_\oplus$ suggests, the underlying calculus is just what in the literature is well-known as the non-deterministic λ-calculus [dP95]. Originally proposed by de'Liguoro and Piperno, $\Lambda_\oplus$ is nothing more than the usual pure λ-calculus extended with a binary operator $\oplus$ representing non-deterministic choice: $M \oplus N$ reduces to either M or N. On top of $\Lambda_\oplus$, Dal Lago and Zorzi [DLZ12] have proposed a probabilistic operational semantics by developing the basic idea of interpreting $\oplus$ as a probabilistic choice. They have investigated call-by-value and call-by-name evaluations to give a probabilistic semantics to $\Lambda_\oplus$, considering both big-step and small-step disciplines, and proved the two resulting semantics equivalent.

4.1.1 Pure, non-deterministic λ-calculus

We now introduce $\Lambda_\oplus$ as the usual pure non-deterministic λ-calculus. Then, we briefly recall its non-deterministic dynamics.

Definition 4.1.1. *Let be given a denumerable set of variables $\mathcal{V} = \{x, y, z, \ldots\}$. The set $\Lambda_\oplus$ of* term expressions *(M, N, L, etc.), or* terms, *is the smallest set such that:*

- *if $x \in \mathcal{V}$, then $x \in \Lambda_\oplus$;* [Variable]
- *if $x \in \mathcal{V}$ and $M \in \Lambda_\oplus$, then $\lambda x.M \in \Lambda_\oplus$;* [Abstraction]
- *if $M, N \in \Lambda_\oplus$, then $(M)\,N \in \Lambda_\oplus$;* [Application]
- *if $M, N \in \Lambda_\oplus$, then $M \oplus N \in \Lambda_\oplus$.* [Choice]

Notation. Terms are ranged over by metavariables like M, N, L. We establish [Choice] to be left-associative in order to let us write $M \oplus N \oplus L$ in place of $(M \oplus N) \oplus L$. Moreover, a sequence of terms $M_1, \ldots, M_n$ is denoted as $\overline{M}$, which we sometimes use in order to concisely write sequences obtained from other sequences and terms: *e.g.* $M \oplus \overline{N}$ denotes the sequence $M \oplus N_1, \ldots, M \oplus N_n$ whenever $\overline{N}$ is $N_1, \ldots, N_n$. We write $\Lambda_\oplus^*$ to denote the set of sequences of terms.

We obviously consider terms modulo renaming of bound variables (*i.e.* α-equivalence). Following the notation of Section 1.2, we indicate as $\mathsf{FV}(M)$ the set of free variables of M and as $M\,[N/x]$ the capture-avoiding substitution of N for the free occurrences of x in M. Both are defined as one expects (in fact, Definitions 1.2.3 and 1.2.4 are intuitively valid here as well).

Throughout this part of the thesis, it is useful to keep track, explicitly, of the free variables we work with: we indicate as $\Lambda_\oplus(\{x_1, \ldots, x_n\})$ the set of terms whose free variables are among the ones in $\{x_1, \ldots, x_n\}$. Therefore, we often write $M \in \Lambda_\oplus(\emptyset)$ to equivalently mean $\mathsf{FV}(M) = \emptyset$, *i.e.* M is a closed term (or *program*).

Among terms, we are particularly interested in the so-called *values*.

Definition 4.1.2. *A term is a* value *whenever it is a closed λ-abstraction (*i.e. *lastly obtained by means of* [Abstraction]*). Values are ranged over by metavariables like V, W, X, and we denote as $\mathsf{V}\Lambda_\oplus$ the set of all values.*

The original work on $\Lambda_\oplus$ [dP95] endows the calculus with a non-deterministic operational semantics by defining β-reduction as the least binary relation enjoying rule $\mapsto_\beta$ (Section 1.4.1) along with the rules $M \oplus N \mapsto M$ and $M \oplus N \mapsto N$, under any possible context. However, we pursue a different approach here.

4.1.2 Value distributions and Call-by-name operational semantics

Following Plotkin's pioneering work [Plo75], we consider close terms as representing *programs*, and as their meaning the *value* they evaluates to in the call-by-name (CbN) reduction strategy. In particular, we focus on the *weak head reduction* (or *lazy reduction*) obtained by preventing reduction of [Abstraction] terms.

As a direct consequence of the interaction between non-deterministic choice and non-terminating computations common in λ-calculus, the meaning of terms may be an infinite set of values. Providing a probabilistic operational semantics to this calculus becomes a delicate matter. As a matter of fact, every inductively defined (hence finite) operational semantics is obviously insufficient here.

We now recall Dal Lago and Zorzi's [DLZ12] probabilistic semantics, which is developed around the notion of *(partial) value distributions* and how to attain them operationally. The idea is that every closed term reduces not to a single value, but rather to a function assigning a probability to every possible value. Full divergence is taken into account by the distribution assigning probability zero to every value.

Definition 4.1.3. *A* value distribution *is a function $\mathscr{D} : \mathsf{V}\Lambda_\oplus \to \mathbb{R}_{[0,1]}$ such that $\sum_{V \in \mathsf{V}\Lambda_\oplus} \mathscr{D}(V) \leq 1$. The set of all value distributions is denoted $\mathcal{P}_{\mathsf{V}\Lambda_\oplus}$. Moreover, given a value distribution $\mathscr{D}$,*

- *its* support $\mathsf{Supp}(\mathscr{D})$ *is the subset of* $\mathsf{V}\Lambda_\oplus$ *whose elements are values to which $\mathscr{D}$ attributes positive probability.*
- *its* sum $\sum \mathscr{D}$ *is* $\sum_{V \in \mathsf{V}\Lambda_\oplus} \mathscr{D}(V)$.

Observe that distributions do not necessarily sum to 1, so to model the possibility of a term to diverge.

Notation. We sometimes need to write $\{V_1^{p_1}, \ldots, V_n^{p_n}\}$ to indicate the value distribution $\mathscr{D}$, with finite support, defined as $\mathscr{D}(V) = \sum_{V_i=V} p_i$. Moreover, notice that $\sum \mathscr{D} = \sum_{i=1}^{n} p_i$.

Value distributions can be ordered point-wise, by lifting the canonical order of $\mathbb{R}$, so that the structure $(\mathcal{P}_{\mathsf{V}\Lambda_\oplus}, \leq)$ turns out forming both a lower semilattice and an ω**CPO** (*i.e.* limits of ω-chains always exist). In particular, $(\mathcal{P}_{\mathsf{V}\Lambda_\oplus}, \leq)$ is not a lattice since the join of two value distributions is not necessary a value distribution: it is rather simple to come up with two value distributions whose join exhibits a sum strictly greater than 1.

Call-by-name operational semantics

We now endow $\Lambda_\oplus$ with a *call-by-name* (CbN) probabilistic operational semantics, in terms of both small-step and big-step disciplines, which assigns a term M a value distribution $[\![M]\!]$. The original work [DLZ12] investigates different formal systems (*i.e.* set of rules for deriving judgements) to model convergence and divergence by interpreting the systems inductively or coinductively. Moreover, the authors prove the semantics for convergence (resp. divergence) equivalent, whereas convergence and divergence are shown to be probabilistically dual with respect to each other. In this work we focus on inductively defined convergence only.

Given a term M, the first step consists in defining a formal system, interpreting it inductively and deriving finite *lower approximations* of the value distribution $[\![M]\!]$. Big-step approximation semantics derives judgements in the form $M \Downarrow \mathscr{D}$, where M is a term and $\mathscr{D}$ is a value distribution with finite support (Figure 4.1). Small-step approximation semantics can be defined similarly, and derives judgements in the form $M \Rightarrow \mathscr{D}$ (Figure 4.2). As a matter of fact, big-step and small-step can simulate each other, that is if $M \Downarrow \mathscr{D}$, then $M \Rightarrow \mathscr{E}$ where $\mathscr{E} \geq \mathscr{D}$, and *vice versa* [DLZ12]. Observe that small-step rule schema requires a notion of (weak) call-by-name reduction, which we define as follows.

Definition 4.1.4. Leftmost (weak) CbN reduction $\mapsto$ *is the least binary relation on* $\Lambda_\oplus \times \Lambda_\oplus^*$ *such that:*

- $(\lambda x.M)\, N \mapsto M\,[N/x]$*;*
- *if* $M \mapsto \overline{L}$*, then* $(M)\, N \mapsto \left(\overline{L}\right) N$*;*
- $M \oplus N \mapsto M, N$*.*

$$\frac{}{M \Downarrow \emptyset}\ \textsf{(be)}$$

$$\frac{}{V \Downarrow \{V^1\}}\ \textsf{(bv)}$$

$$\frac{M \Downarrow \mathscr{D} \quad \{P\,[N/x] \Downarrow \mathscr{E}_{P,N}\}_{\lambda x.P \in \mathsf{Supp}(\mathscr{D})}}{(M)\,N \Downarrow \sum_{\lambda x.P \in \mathsf{Supp}(\mathscr{D})} \mathscr{D}(\lambda x.P) \cdot \mathscr{E}_{P,N}}\ \textsf{(ba)}$$

$$\frac{M \Downarrow \mathscr{D} \quad N \Downarrow \mathscr{E}}{M \oplus N \Downarrow \frac{1}{2}\mathscr{D} + \frac{1}{2}\mathscr{E}}\ \textsf{(bs)}$$

Figure 4.1: Big-step CbN approximation semantics for $\Lambda_\oplus$.

$$\frac{}{M \Rightarrow \emptyset}\ \textsf{(se)}$$

$$\frac{}{V \Rightarrow \{V^1\}}\ \textsf{(sv)}$$

$$\frac{M \mapsto \overline{N} \quad N_i \Rightarrow \mathscr{D}_i}{M \Rightarrow \sum_{i=1}^{n} \frac{1}{n}\mathscr{D}_i}\ \textsf{(st)}$$

Figure 4.2: Small-step CbN approximation semantics for $\Lambda_\oplus$.

Notice that reduction cannot take place under a λ-abstraction. Moreover, as a peculiarity of CbN semantics, it is possible to perform a choice between terms which are not (necessarily) values. In the second step, the value distribution $[\![M]\!]$, called the CbN *semantics* of M, is set as the least upper bound of distributions obtained in either of the two ways.

Definition 4.1.5. *The* CbN semantics *of a term $M \in \Lambda_\oplus$ is the value distribution $[\![M]\!]$ defined as:*

$$[\![M]\!] = \sup_{M \Downarrow \mathscr{D}} \mathscr{D} = \sup_{M \Rightarrow \mathscr{D}} \mathscr{D}.$$

The above $[\![M]\!]$ is well-defined due to the fact that the set of all distributions $\mathscr{D}$ such that $M \Downarrow \mathscr{D}$ is directed, and its least upper bound is a value distribution because of ω-completeness.

The semantics of terms satisfies some useful equations, such as:

Lemma 4.1.6.

- $[\![(\lambda x.M)\, N]\!] = [\![M\, [N/x]]\!]$;
- $[\![M \oplus N]\!] = \frac{1}{2}\, [\![M]\!] + \frac{1}{2}\, [\![N]\!]$.

Proof. We refer to Dal Lago and Zorzi's work [DLZ12] for detailed proofs. □

4.2 Probabilistic bisimulation

In this section we recall the definition and preliminary results concerning *probabilistic (bi)simulation* in the setting of labelled Markov chains (*i.e.* labelled probabilistic transition systems characterised by discrete state space and time), by following Larsen and Skou's influential work [LS91]. In Section 4.3 we adapt the resulting form of bisimilarity to the probabilistic λ-calculus $\Lambda_{\oplus}$ by combining it with Abramsky's notion of applicative bisimilarity [Abr90].

In particular we deal with fully probabilistic systems, namely systems in which transitions (*i.e.* the "internal" non-determinism) are quantified over $\mathbb{R}_{[0,1]}$, whereas the choice of labels (*i.e.* the "external" non-determinism) is unquantified.

Definition 4.2.1. *A* labelled Markov chain *is a triple* $(\mathcal{S}, \mathcal{L}, \mathcal{P})$ *such that:*

- $\mathcal{S}$ *is a countable set of* states*;*
- $\mathcal{L}$ *is set of* labels*;*
- $\mathcal{P}$ *is a* transition probability matrix, *i.e. a function*

$$\mathcal{P} : \mathcal{S} \times \mathcal{L} \times \mathcal{S} \to \mathbb{R}_{[0,1]}$$

such that the following normalisation condition holds:

$$\forall \ell \in \mathcal{L}.\, \forall s \in \mathcal{S}.\, \mathcal{P}(s, \ell, \mathcal{S}) \leq 1$$

where $\mathcal{P}(s, \ell, X)$ *stands for* $\sum_{t \in X} \mathcal{P}(s, \ell, t)$*, whenever* $X \subseteq \mathcal{S}$*.*

Observe that $\mathcal{P}(s, \ell, \mathcal{S})$ is not required to sum to 1, meaning that we rather allow subprobability distributions in which the sum of all probabilities is only bounded by 1. This allows some actions to be rejected, something necessary as long as we distinguish states on the basis of the actions they may, or may not, exhibit.

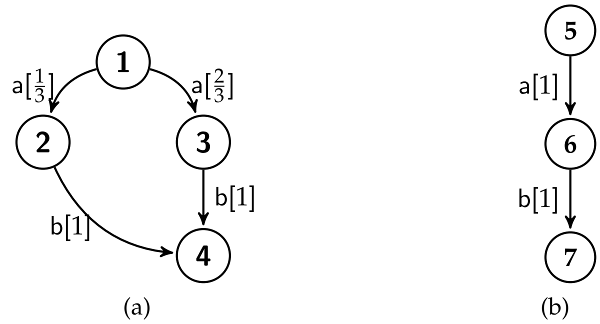


Figure 4.3: Two bisimilar probabilistic systems.

Notation. States are ranged over by metavariables like s, t, v. Moreover, given an equivalence relation $\mathcal{R}$ on $\mathcal{S}$, metavariables like $\mathtt{E}, \mathtt{F}, \mathtt{G}$ indicate the equivalence classes of $\mathcal{S}$ modulo $\mathcal{R}$.

We now give the definition of *probabilistic bisimulation*. The crucial point is that the (non-)deterministic way of establishing the bisimulation game, based on the idea of one-to-one transition matching, is insufficient here. Intuitively, rather than reasoning with respect to states, one must reason with respect to state space partitions and take into account quantitative information. In other words, states are bisimilar whenever they exhibit the same actions, quantified with a same *overall* probability, to state partitions of bisimilar states. This requires transition probabilities to bisimilar states to be added (Example 4.2.3).

Definition 4.2.2. *Given a labelled Markov chain* $(\mathcal{S}, \mathcal{L}, \mathcal{P})$*, a* probabilistic bisimulation *is an equivalence relation* $\mathcal{R}$ *on* $\mathcal{S}$ *such that* $(s, t) \in \mathcal{R}$ *implies that for every* $\ell \in \mathcal{L}$ *and for every* $\mathtt{E} \in \mathcal{S}/\mathcal{R}$*,* $\mathcal{P}(s, \ell, \mathtt{E}) = \mathcal{P}(t, \ell, \mathtt{E})$*.*

Consider the following example to exercise the above definition.

***Example* 4.2.3.** Let ID be the identity relation on the state space $\mathcal{S} = \{1, 2, 3, 4, 5, 6, 7\}$. Consider the two probabilistic systems in Figure 4.3, and the relation $\mathcal{R} = \{(1, 5); (2, 6); (3, 6); (4, 7)\}^* \cup \mathsf{ID}$. It is simple to verify that $\mathcal{R}$ is an equivalence relation and that, moreover, it is a probabilistic bisimulation. Indeed the (significant) equivalence classes of $\mathcal{S}$ modulo $\mathcal{R}$ are $\mathcal{S}/\mathcal{R} = \{\{1, 5\}; \{2, 3, 6\}; \{4, 7\}\}$, and verify $(s, t) \in \mathcal{R}$ implies that, for every $\ell \in \{a, b\}$ and for every $\mathtt{E} \in \mathcal{S}/\mathcal{R}$, $\mathcal{P}(s, \ell, \mathtt{E}) = \mathcal{P}(t, \ell, \mathtt{E})$. We detail only the most significant case: consider the states $(1, 5)$ and verify that on $\ell = a$ and $\mathtt{E} = \{(2, 3, 6)\}$

$$\mathcal{P}(1, a, \mathtt{E}) = \sum_{s \in \mathtt{E}} \mathcal{P}(1, a, s)$$

$$
\begin{aligned}
&= \mathcal{P}(1,a,2) + \mathcal{P}(1,a,3) + \mathcal{P}(1,a,6) \\
&= 1 \\
&= \mathcal{P}(5,a,2) + \mathcal{P}(5,a,3) + \mathcal{P}(5,a,6) \\
&= \sum_{s \in \mathtt{E}} \mathcal{P}(5,a,s) = \mathcal{P}(5,a,\mathtt{E}),
\end{aligned}
$$

whereas on all other cases of $\ell \in \{a,b\}$ and $\mathtt{F} \in \mathcal{S}/\mathcal{R}$ with $\mathtt{F} \neq \mathtt{E}$, it follows $\mathcal{P}(1,\ell,\mathtt{F}) = 0 = \mathcal{P}(5,\ell,\mathtt{F})$.

Notation. We often use the mixfix notation $s\ \mathcal{R}\ t$ instead of $(s,t) \in \mathcal{R}$. Moreover, we write $\mathcal{R}^{op}$ for the *reciprocal relation* of $\mathcal{R}$, namely $\mathcal{R}^{op} = \{(t,s) \mid (s,t) \in \mathcal{R}\}$.

Notice that a probabilistic bisimulation has to be, by definition, an equivalence relation. This means that, in principle, we are not allowed to define probabilistic bisimilarity simply as the union of all probabilistic bisimulations. As a matter of fact, given $\mathcal{R}, \mathcal{T}$ two equivalence relations, $\mathcal{R} \cup \mathcal{T}$ is not necessarily an equivalence relation. In particular, $\mathcal{R} \cup \mathcal{T}$ is not necessarily transitive.

The following is a standard way to overcome the problem.

Lemma 4.2.4. *If $\{\mathcal{R}_i\}_{i \in I}$ is a collection of probabilistic bisimulations, then also their reflexive and transitive closure $(\bigcup_{i \in I} \mathcal{R}_i)^*$ is a probabilistic bisimulation.*

Proof. We set $\mathcal{T} = (\bigcup_{i \in I} \mathcal{R}_i)^*$ and we show $\mathcal{T}$ is an equivalence relation as follows:

- Reflexivity is simple: $\mathcal{T}$ is reflexive by definition.
- Symmetry is a consequence of symmetry of each of the relations in $\{\mathcal{R}_i\}_{i \in I}$: if $s\ \mathcal{T}\ t$, then there are $n \geq 0$ states $v_0, \ldots, v_n$ such that $v_0 = s$, $v_n = t$ and for every $1 \leq i \leq n$ there is $j \in I$ such that $v_{i-1}\ \mathcal{R}_j\ v_i$. The symmetry property of each of the $\mathcal{R}_j$ entails $v_i\ \mathcal{R}_j\ v_{i-1}$. As a consequence, $t\ \mathcal{T}\ s$.
- Transitivity is simple: $\mathcal{T}$ is transitive by definition.

Notice that for every $i \in I$, $\mathcal{R}_i \subseteq \bigcup_{j \in I} \mathcal{R}_j \subseteq \mathcal{T}$, meaning that every equivalence class with respect to $\mathcal{T}$ is the union of equivalence classes with respect to $\mathcal{R}_i$. Suppose $s\ \mathcal{T}\ t$, then there are $n \geq 0$ states $v_0, \ldots, v_n$ such that $v_0 = s$, $v_n = t$ and for every $1 \leq i \leq n$ there is $j \in I$ such that $v_{i-1}\ \mathcal{R}_j\ v_i$. For every $\ell \in \mathcal{L}$ and every $\mathtt{E} \in \mathcal{S}/\mathcal{T}$, it follows

$$\mathcal{P}(s,\ell,\mathtt{E}) = \mathcal{P}(v_0,\ell,\mathtt{E}) = \ldots = \mathcal{P}(v_n,\ell,\mathtt{E}) = \mathcal{P}(t,\ell,\mathtt{E}).$$

This concludes the proof. $\square$

Lemma 4.2.4 allows us to define the largest probabilistic bisimulation, called *probabilistic bisimilarity*.

Definition 4.2.5. Probabilistic bisimilarity *is the relation* $\sim$ *defined as*

$$\sim = \bigcup\{\mathcal{R} \mid \mathcal{R} \textit{ is a probabilistic bisimulation}\}.$$

Lemma 4.2.6. *It holds that* $\sim = (\sim)^*$. *Hence,* $\sim$ *is a probabilistic bisimulation.*

Proof. Notice that, by Lemma 4.2.4, $(\sim)^*$ is a probabilistic bisimulation. The inclusion $\sim \subseteq (\sim)^*$ is obvious. The other way around, $\sim \supseteq (\sim)^*$, follows by $(\sim)^*$ being a probabilistic bisimulation and so included in the union of them all, that is $\sim$. $\square$

Definition 4.2.7. *We write* $\mathcal{R}(X)$ *for the* $\mathcal{R}$-closure of X, *namely the set*

$$\mathcal{R}(X) = \{y \in \mathcal{S} \mid \exists\, x \in X.\, x \mathrel{\mathcal{R}} y\}.$$

We now define the notion of probabilistic simulation. Here preorders play the role of equivalence relations in Definition 4.2.2 of probabilistic bisimulations.

Definition 4.2.8. *Given a labelled Markov chain* $(\mathcal{S}, \mathcal{L}, \mathcal{P})$, *a* probabilistic simulation *is a preorder relation* $\mathcal{R}$ *on* $\mathcal{S}$ *such that* $(s,t) \in \mathcal{R}$ *implies that for every* $\ell \in \mathcal{L}$ *and for every* $X \subseteq \mathcal{S}$, $\mathcal{P}(s, \ell, X) \leq \mathcal{P}(t, \ell, \mathcal{R}(X))$.

Of course, Lemma 4.2.4 can be adapted to probabilistic simulations.

Lemma 4.2.9. *If* $\{\mathcal{R}_i\}_{i \in I}$, *is a collection of probabilistic simulations, then also their reflexive and transitive closure* $(\bigcup_{i \in I} \mathcal{R}_i)^*$ *is a probabilistic simulation.*

Proof. Relation $\mathcal{T} = (\bigcup_{i \in I} \mathcal{R}_i)^*$ is a preorder by construction. Suppose $(s,t) \in \mathcal{T}$, then there are $n \geq 0$ states $v_0, \ldots, v_n$ such that $v_0 = s$, $v_n = t$ and for every $1 \leq i \leq n$ there is $j_i \in I$ such that $v_{i-1} \mathrel{\mathcal{R}_{j_i}} v_i$. As a consequence, for every $\ell \in \mathcal{L}$ and every $X \subseteq \mathcal{S}$, it holds

$$\begin{aligned}
\mathcal{P}(v_0, \ell, X) &\leq \mathcal{P}(v_1, \ell, \mathcal{R}_{j_1}(X)) \\
&\leq \mathcal{P}(v_2, \ell, \mathcal{R}_{j_2}(\mathcal{R}_{j_1}(X))) \\
&\leq \cdots \leq \mathcal{P}(v_n, \ell, \mathcal{R}_{j_n}(\ldots(\mathcal{R}_{j_2}(\mathcal{R}_{j_1}(X)))))
\end{aligned}$$

Moreover, it is straightforward to verify that

$$\mathcal{R}_{j_n}(\ldots(\mathcal{R}_{j_2}(\mathcal{R}_{j_1}(X)))) \subseteq \mathcal{T}(X),$$

which implies $\mathcal{P}(s, \ell, X) \leq \mathcal{P}(t, \ell, \mathcal{T}(X))$. $\square$

Lemma 4.2.9 allows us to define the largest probabilistic simulation, called *probabilistic similarity*.

Definition 4.2.10. Probabilistic similarity *is the relation* $\lesssim$ *defined as*

$$\lesssim = \bigcup\{\mathcal{R} \mid \mathcal{R} \text{ is a probabilistic simulation}\}.$$

Lemma 4.2.11. *It holds that* $\lesssim = (\lesssim)^*$. *Hence,* $\lesssim$ *is a probabilistic simulation.*

Proof. Similar to the proof of Lemma 4.2.6. □

We now give some standard results concerning probabilistic (bi)simulations. Any symmetric probabilistic simulation is a probabilistic bisimulation.

Proposition 4.2.12. *If* $\mathcal{R}$ *is a symmetric probabilistic simulation, then* $\mathcal{R}$ *is a probabilistic bisimulation.*

Proof. Since $\mathcal{R}$ is a probabilistic simulation, then it is a preorder (*i.e.* it is a reflexive and transitive relation) by definition (Definition 4.2.8). Moreover, $\mathcal{R}$ is symmetric by hypothesis, hence $\mathcal{R}$ is an equivalence relation. We now prove $\mathcal{R}$ is a probabilistic bisimulation, namely that $s \ \mathcal{R} \ t$ implies that for every $\ell \in \mathcal{L}$ and every $\mathtt{E} \in \mathcal{S}/\mathcal{R}$, $\mathcal{P}(s, \ell, \mathtt{E}) = \mathcal{P}(t, \ell, \mathtt{E})$. From the fact that $\mathcal{R}$ is a simulation follows $s \ \mathcal{R} \ t$ implies that for every $\ell \in \mathcal{L}$ and every $\mathtt{E} \in \mathcal{S}/\mathcal{R}$, $\mathcal{P}(s, \ell, \mathtt{E}) \leq \mathcal{P}(t, \ell, \mathcal{R}(\mathtt{E}))$. Since $\mathtt{E} \in \mathcal{S}/\mathcal{R}$ is an $\mathcal{R}$-equivalence class, it holds $\mathcal{R}(\mathtt{E}) = \mathtt{E}$: hence the latter entails $\mathcal{P}(s, \ell, \mathtt{E}) \leq \mathcal{P}(t, \ell, \mathtt{E})$. The other way around follows by the symmetry property of $\mathcal{R}$, which implies that for every $\ell \in \mathcal{L}$ and every $\mathtt{E} \in \mathcal{S}/\mathcal{R}$, $\mathcal{P}(t, \ell, \mathtt{E}) \leq \mathcal{P}(s, \ell, \mathtt{E})$. Hence, $\mathcal{P}(s, \ell, \mathtt{E}) = \mathcal{P}(t, \ell, \mathtt{E})$ which completes the proof. □

Moreover, every probabilistic bisimulation (and its reciprocal) is a probabilistic simulation.

Lemma 4.2.13. *If* $\mathcal{R}$ *is a probabilistic bisimulation, then* $\mathcal{R}$ *and* $\mathcal{R}^{op}$ *are probabilistic simulation.*

Proof. We prove $\mathcal{R}$ is a probabilistic simulation first. Given $X \subseteq \mathcal{S}$, consider the collection $\{X_i\}_{i \in I}$ of equivalence classes of X modulo $\mathcal{R}$. Formally, $X = \bigcup_{i \in I} X_i$ and, for all $i \in I$, $X_i \subseteq \mathtt{E}_i$ with $\mathtt{E}_i$ equivalence class of $\mathcal{S}$ modulo $\mathcal{R}$. As a consequence, $\mathcal{R}(X) = \bigcup_{i \in I} \mathtt{E}_i$. For every $\ell \in \mathcal{L}$ and every $X \subseteq \mathcal{S}$, it follows

$$\begin{aligned}
\mathcal{P}(s, \ell, X) &= \sum_{i \in I} \mathcal{P}(s, \ell, X_i) \\
&\leq \sum_{i \in I} \mathcal{P}(s, \ell, \mathtt{E}_i) \\
&= \sum_{i \in I} \mathcal{P}(t, \ell, \mathtt{E}_i) = \mathcal{P}(t, \ell, \mathcal{R}(X)).
\end{aligned}$$

Finally, $\mathcal{R}^{op}$ is also a probabilistic simulation as a consequence of the symmetry property of $\mathcal{R}$ and the fact, just proved, that $\mathcal{R}$ is a probabilistic simulation. □

Contrary to the non-deterministic case [San11, Pan11], simulation equivalence (*i.e.* the equivalence relation generated by $\lesssim$) coincides with bisimulation.

Theorem 4.2.14. *$\sim$ coincides with $\lesssim \cap \lesssim^{op}$.*

Proof. The fact that $\sim$ is a subset of $\lesssim \cap \lesssim^{op}$ is a straightforward consequence of symmetry property of $\sim$ and the fact that, by Lemma 4.2.13, every probabilistic bisimulation is also a probabilistic simulation. We now prove that $\lesssim \cap \lesssim^{op}$ is a subset of $\sim$, namely the fact that the former is a probabilistic bisimulation. Of course, $\lesssim \cap \lesssim^{op}$ is an equivalence relation because $\lesssim$ is a preorder. Consider any equivalence class $\mathtt{E}$ modulo $\lesssim \cap \lesssim^{op}$, and define the two sets of states $X = \lesssim (\mathtt{E})$ and $Y = X \setminus \mathtt{E}$. Of course, Y and $\mathtt{E}$ are disjoint sets of states whose union is precisely X. Moreover, both X and Y are closed with respect to $\lesssim$:

- On the one hand, if $s \in \lesssim (X)$, then $s \in \lesssim (\lesssim (\mathtt{E})) = \lesssim (\mathtt{E}) = X$;
- On the other hand, if $s \in \lesssim (Y)$, then there is $t \in X$ which is not in $\mathtt{E}$ such that $t \lesssim s$. It follows $s \in \lesssim (X)$, hence $s \in X$ (from the previous point) but $s \notin \mathtt{E}$, that is $s \in Y$. Indeed suppose $s \in \mathtt{E}$, then it would imply s and t members of the same equivalence class modulo $\lesssim \cap \lesssim^{op}$, that is $t \in \mathtt{E}$. Contradiction.

As a consequence, given any $(s,t) \in \lesssim \cap \lesssim^{op}$ and any $\ell \in \mathcal{L}$,

$$\mathcal{P}(s,\ell,X) \leq \mathcal{P}(t,\ell,\lesssim (X)) = \mathcal{P}(t,\ell,X),$$

and

$$\mathcal{P}(t,\ell,X) \leq \mathcal{P}(s,\ell,\lesssim (X)) = \mathcal{P}(s,\ell,X).$$

It follows $\mathcal{P}(s,\ell,X) = \mathcal{P}(t,\ell,X)$ and, by a similar reasoning, $\mathcal{P}(s,\ell,Y) = \mathcal{P}(t,\ell,Y)$. Since $\mathcal{P}(s,\ell,X) \geq \mathcal{P}(s,\ell,Y)$ and $\mathcal{P}(t,\ell,X) \geq \mathcal{P}(t,\ell,Y)$, it follows

$$\begin{aligned}\mathcal{P}(s,\ell,\mathtt{E}) &= \mathcal{P}(s,\ell,X) - \mathcal{P}(s,\ell,Y)\\ &= \mathcal{P}(t,\ell,X) - \mathcal{P}(t,\ell,Y) = \mathcal{P}(t,\ell,\mathtt{E})\end{aligned}$$

which is the thesis. □

For technical reasons that turn apparent in Section 4.4, it is convenient to consider labelled Markov chains in which the state space is partitioned into disjoint sets, in such a way that comparing states coming from different components is not possible.

Definition 4.2.15. *The* disjoint union $\biguplus_{i\in I} X_i$ *of a collection of sets* $\{X_i\}_{i\in I}$ *is defined as* $\biguplus_{i\in I} X_i = \{(a,i) \mid i \in I \wedge a \in X_i\}$. *A labelled Markov chain* $(\mathcal{S},\mathcal{L},\mathcal{P})$ *is said to be* multisorted *whenever* $\mathcal{S} = \biguplus_{i\in I} X_i$ *for some collection of sets* $\{X_i\}_{i\in I}$.

Whenever a labelled Markov chain is multisorted, with the state space $\mathcal{S}$ given by $\biguplus_{i \in I} X_i$, we require that (bi)simulation relations only compare elements coming from the same X_i, namely $(a, i)\ \mathcal{R}\ (b, j)$ implies $i = j$. We details such definitions with respect to *probabilistic applicative (bi)simulations* in Section 4.3, when we set up a multisorted labelled Markov chain on top of $\Lambda_\oplus$.

***Remark* 4.2.16.** The two systems of Example 4.2.3 can be shown to simulate each other, that is $(1, 5) \in \lesssim \cap \lesssim^{op}$. Theorem 4.2.14 confirms that the two are indeed bisimilar. The two are also bisimilar in a non-deterministic setting.

Suppose now that the system on the left is deprived of the transition from state 2 to state 4. The two systems are no longer bisimilar, neither in a probabilistic sense nor in the non-deterministic one. Nonetheless, it is simple to verify that they still simulates each other in a non-deterministic setting.

4.3 Probabilistic applicative bisimulation and Howe's method

In this section we introduce the notions of similarity and bisimilarity for $\Lambda_\oplus$, in the spirit of Abramsky's work on applicative bisimulation [Abr90]. Definitionally, this consists in seeing $\Lambda_\oplus$'s operational semantics (Figure 4.1) as a labelled Markov chain, then giving the Larsen and Skou's notion of (bi)simulation for it. States are terms, while labels are of two kinds: one can either *evaluate* a term, obtaining (a distribution of) values, or *apply* a term to a value.

We show that the resulting bisimulation (probabilistic applicative bisimulation) is a congruence, thus included in probabilistic context equivalence. In order to do so, we provide a non-trivial generalisation of Howe's technique [How96], which is a well-known methodology to get congruence results in presence of higher-order functions, but which has not been applied to probabilistic calculi so far.

4.3.1 Probabilistic applicative bisimulation

Formalising probabilistic applicative bisimulation requires some care. As usual, two values $\lambda x.M$ and $\lambda x.N$ are defined to be bisimilar if for every L, $M\,[L/x]$ and $N\,[L/x]$ are themselves bisimilar. But how if we rather want to compare two arbitrary closed terms M and N? The simplest solution consists in following Larsen and Skou [LS91] and stipulate that every equivalence class of $V\Lambda_\oplus$ modulo bisimulation is attributed the same measure by both $[\![M]\!]$ and $[\![N]\!]$. Values are thus treated in two different ways (they are both terms and values), and this is the reason why each of them corresponds to *two* states in the underlying Markov chain.

Definition 4.3.1. $\Lambda_\oplus$ *can be seen as a multisorted labelled Markov chain* $(\Lambda_\oplus(\emptyset) \uplus V\Lambda_\oplus, \Lambda_\oplus(\emptyset) \uplus \{\tau\}, \mathcal{P}_\oplus)$ *that we denote with* $\Lambda_\oplus$. *Labels are either closed terms, which model parameter passing, or* τ, *that models evaluation. Observe that the states of the labelled Markov chain we have just defined are elements of the disjoint union* $\Lambda_\oplus(\emptyset) \uplus V\Lambda_\oplus$. *Two distinct states correspond to the same value* V, *and to avoid ambiguities, we call the second one (i.e. the one coming from* $V\Lambda_\oplus$*) a* distinguished value. *When we want to insist on the fact that a value* $\lambda x.M$ *is distinguished, we indicate it with* $\nu x.M$. *We define the transition probability matrix* $\mathcal{P}_\oplus$ *as follows:*

- *For every term* M *and for every distinguished value* $\nu x.N$,
$$\mathcal{P}_\oplus(M, \tau, \nu x.N) = [\![M]\!]\,(\nu x.N);$$
- *For every term* M *and for every distinguished value* $\nu x.N$,
$$\mathcal{P}_\oplus(\nu x.N, M, N\,[M/x]) = 1;$$
- *In all other cases,* $\mathcal{P}_\oplus$ *returns* 0.

Terms seen as states only interact with the environment by performing τ, while distinguished values only take other closed terms as parameters.

Simulation and bisimulation relations can be defined for $\Lambda_\oplus$ as for any labelled Markov chain. Even if, strictly speaking, these are binary relations on $\Lambda_\oplus(\emptyset) \uplus V\Lambda_\oplus$, we often see them just as their restrictions to $\Lambda_\oplus(\emptyset)$. Formally,

Definition 4.3.2. *A* probabilistic applicative bisimulation *(a* PAB*) is simply a probabilistic bisimulation on* $\Lambda_\oplus$. *Similarly, a* probabilistic applicative simulation *(a* PAS*) is simply a probabilistic simulation on* $\Lambda_\oplus$.

This way one can define *probabilistic applicative similarity and bisimilarity*, which are denoted $\lesssim$ and $\sim$, respectively.

***Remark* 4.3.3.** Technically, the distinction between terms and values in Definition 4.3.1 means that our bisimulation is in *late* style. In bisimulations for value-passing concurrent languages, *late* indicates the explicit manipulation of functions in the clause for input actions: functions are chosen first, and only later, the input value received is taken into account [SW01]. Late-style is used in contraposition to *early* style, where the order of quantifiers is exchanged, so that the choice of functions may depend on the specific input value received. In our setting, adopting an early style would mean having transitions such as $\lambda x.M \xrightarrow{N} M\,[N/x]$, and then setting up a probabilistic bisimulation on top of the resulting transition system. In this paper, we stick to the late style because easier to deal with, especially under Howe's technique.

Previous works on applicative bisimulation for non-deterministic functions also focus on the late approach [Ong93, Pit11].

Defining applicative bisimulation in terms of multisorted labelled Markov chains turns to be useful when dealing with Howe's method. To spell out the explicit operational details of the definition, a probabilistic applicative bisimulation can be seen as an equivalence relation $\mathcal{R} \subseteq \Lambda_\oplus(\emptyset) \times \Lambda_\oplus(\emptyset)$ such that whenever $M \mathrel{\mathcal{R}} N$, then:

1. $[\![M]\!]\,(\mathtt{E} \cap \mathsf{V}\Lambda_\oplus) = [\![N]\!]\,(\mathtt{E} \cap \mathsf{V}\Lambda_\oplus)$, for any equivalence class $\mathtt{E}$ of $\mathcal{R}$ (*i.e.* the probability of reaching a value in $\mathtt{E}$ is the same for the two terms);
2. if $M, N \in \mathsf{V}\Lambda_\oplus$, say $M = \lambda x.P$ and $N = \lambda x.Q$, then $P\,[L/x] \mathrel{\mathcal{R}} Q\,[L/x]$ for all $L \in \Lambda_\oplus(\emptyset)$.

The special treatment of values in the 2nd clause motivates the use of *multisorted* labelled Markov chains in Definition 4.3.1.

As one can easily guess, terms with the same semantics are indistinguishable:

Lemma 4.3.4. *The following binary relation $\mathcal{R}$ is a* PAB*:*

$$\mathcal{R} = \{(M, N) \in \Lambda_\oplus(\emptyset) \times \Lambda_\oplus(\emptyset) \mid [\![M]\!] = [\![N]\!]\} \uplus \{(V, V) \in \mathsf{V}\Lambda_\oplus \times \mathsf{V}\Lambda_\oplus\}.$$

Proof. The fact $\mathcal{R}$ is an equivalence easily follows from the reflexive, symmetric and transitive properties of set-theoretic equality. Relation $\mathcal{R}$ must enjoy the property that, whenever $M \mathrel{\mathcal{R}} N$, then for every $\mathtt{E} \in \mathsf{V}\Lambda_\oplus/\mathcal{R}$, $\mathcal{P}_\oplus(M, \tau, \mathtt{E}) = \mathcal{P}_\oplus(N, \tau, \mathtt{E})$. Notice that $[\![M]\!] = [\![N]\!]$ clearly entails $\mathcal{P}_\oplus(M, \tau, V) = \mathcal{P}_\oplus(N, \tau, V)$, for every $V \in \mathsf{V}\Lambda_\oplus$. With the same hypothesis,

$$\begin{aligned}\mathcal{P}_\oplus(M, \tau, \mathtt{E}) &= \sum_{V \in \mathtt{E}} \mathcal{P}_\oplus(M, \tau, V) \\ &= \sum_{V \in \mathtt{E}} \mathcal{P}_\oplus(N, \tau, V) = \mathcal{P}_\oplus(N, \tau, \mathtt{E}).\end{aligned}$$

On distinguished values, $\mathcal{R}$ must enjoy the property that, whenever $\nu x.M \mathrel{\mathcal{R}} \nu x.N$, then for every $L \in \Lambda_\oplus(\emptyset)$ and for every $\mathtt{E} \in \Lambda_\oplus(\emptyset)/\mathcal{R}$, $\mathcal{P}_\oplus(\nu x.M, L, \mathtt{E}) = \mathcal{P}_\oplus(\nu x.N, L, \mathtt{E})$. Now, the hypothesis $[\![\nu x.M]\!] = [\![\nu x.N]\!]$ implies $M = N$, hence $\mathcal{P}_\oplus(\nu x.M, L, P) = \mathcal{P}_\oplus(\nu x.N, L, P)$ for every $P \in \Lambda_\oplus(\emptyset)$. With the same hypothesis,

$$\begin{aligned}\mathcal{P}_\oplus(\nu x.M, L, \mathtt{E}) &= \sum_{P \in \mathtt{E}} \mathcal{P}_\oplus(\nu x.M, L, P) \\ &= \sum_{P \in \mathtt{E}} \mathcal{P}_\oplus(\nu x.N, L, P) = \mathcal{P}_\oplus(\nu x.N, L, \mathtt{E}).\end{aligned}$$

This concludes the proof. $\square$

Observe that the previous result yields a nice consequence: for every $M, N \in \Lambda_\oplus(\emptyset)$, $(\lambda x.M)N \sim M[N/x]$. Indeed, Lemma 4.1.6 tells us that the latter terms have the same semantics.

Conversely, knowing that two terms M and N are (bi)similar means knowing quite a lot about their convergence probability.

Lemma 4.3.5. *If $M \sim N$, then $\sum [\![M]\!] = \sum [\![N]\!]$. Moreover, if $M \lesssim N$, then $\sum [\![M]\!] \leq \sum [\![N]\!]$.*

Proof. Straightforward from the definitions of $\sim$ and $\lesssim$. In particular,

$$\begin{aligned}\sum [\![M]\!] &= \sum_{\mathtt{E} \in \mathsf{V}\Lambda_\oplus/\sim} \mathcal{P}_\oplus(M, \tau, \mathtt{E}) \\ &= \sum_{\mathtt{E} \in \mathsf{V}\Lambda_\oplus/\sim} \mathcal{P}_\oplus(N, \tau, \mathtt{E}) = \sum [\![N]\!],\end{aligned}$$

and,

$$\begin{aligned}\sum [\![M]\!] &= \mathcal{P}_\oplus(M, \tau, \mathsf{V}\Lambda_\oplus) \\ &\leq \mathcal{P}_\oplus(N, \tau, \lesssim(\mathsf{V}\Lambda_\oplus)) \\ &= \mathcal{P}_\oplus(N, \tau, \mathsf{V}\Lambda_\oplus) = \sum [\![N]\!].\end{aligned}$$

This concludes the proof. □

In the following example we stress again the crucial point of probabilistic applicative (bi)simulation, namely the fact that we reason modulo equivalence classes of states, rather than on states proper.

***Example* 4.3.6.** Bisimilar terms do not necessarily have the same semantics. After all, this is one reason for using bisimulation, and its proof method, as basis to prove equalities among functions. Let us consider the following terms:

$$\begin{aligned}M &= ((\lambda x.(x \oplus x)) \oplus \lambda x.x) \oplus \mathbf{\Omega}; \\ N &= \mathbf{\Omega} \oplus \lambda x.(\mathbf{I})\, x;\end{aligned}$$

Their semantics differ, as for every value V, we have:

$$\begin{aligned}[\![M]\!](V) &= \begin{cases} \frac{1}{4} & \text{if } V \text{ is } \nu x.(x \oplus x) \text{ or } \nu x.x; \\ 0 & \text{otherwise;} \end{cases} \\ [\![N]\!](V) &= \begin{cases} \frac{1}{2} & \text{if } V \text{ is } \nu x.(\mathbf{I})\, x; \\ 0 & \text{otherwise.} \end{cases}\end{aligned}$$

Nonetheless, we can prove $M \sim N$. Indeed, $\nu x.(x \oplus x) \sim \nu x.x \sim \nu x.(\mathbf{I})\, x$ because, for every $L \in \Lambda_\oplus(\emptyset)$, the three terms L, $L \oplus L$ and $(\mathbf{I})\, L$ all have the same semantics, namely $[\![L]\!]$. Now, consider any equivalence class $\mathtt{E}$ of distinguished values modulo $\sim$. If $\mathtt{E}$ includes the three distinguished values above, then

$$\mathcal{P}_{\oplus}(M,\tau,\mathtt{E}) = \sum_{V\in\mathtt{E}} [\![M]\!]\,(V) = \frac{1}{2} = \sum_{V\in\mathtt{E}} [\![N]\!]\,(V) = \mathcal{P}_{\oplus}(N,\tau,\mathtt{E}).$$

Otherwise, $\mathcal{P}_{\oplus}(M,\tau,\mathtt{E}) = 0 = \mathcal{P}_{\oplus}(N,\tau,\mathtt{E})$.

We now prove the following technical result asserting that bisimilar distinguished values are bisimilar values, and *vice versa*. More importantly, it motivates our limitation in considering (bi)similarities defined on closed terms only.

Lemma 4.3.7. *The following three statements are equivalent:*

1. $\lambda x.M \sim \lambda x.N$;
2. $\nu x.M \sim \nu x.N$;
3. *for all* $L \in \Lambda_{\oplus}(\emptyset)$, $M\,[L/x] \sim N\,[L/x]$.

Proof. The fact that 1 and 2 are equivalent is obvious by the definition of $[\![\cdot]\!]$ and Lemma 4.3.4. For that matter, distinguished values are value terms. Let us now detail that 2 and 3 are equivalent.

($2 \Rightarrow 3$) The fact that $\sim$ is a PAB implies that, for every $L \in \Lambda_{\oplus}(\emptyset)$ and every $\mathtt{E} \in \Lambda_{\oplus}(\emptyset)/\!\sim$, $\mathcal{P}_{\oplus}(\nu x.M, L, \mathtt{E}) = \mathcal{P}_{\oplus}(\nu x.N, L, \mathtt{E})$. Assume, for the sake of contradiction, that $M\,[L/x] \not\sim N\,[L/x]$ for some $L \in \Lambda_{\oplus}(\emptyset)$. In particular, the latter means that there is $\mathtt{F} \in \Lambda_{\oplus}(\emptyset)/\!\sim$ such that $M\,[L/x] \in \mathtt{F}$ and $N\,[L/x] \notin \mathtt{F}$. According to its definition, $\mathcal{P}_{\oplus}(\nu x.M, L, P) = 1$ whenever $P = M\,[L/x]$, and $\mathcal{P}_{\oplus}(\nu x.M, L, P) = 0$ otherwise. Then, since $M\,[L/x] \in \mathtt{F}$, it follows $\mathcal{P}_{\oplus}(\nu x.M, L, \mathtt{F}) = \sum_{P\in\mathtt{F}} \mathcal{P}_{\oplus}(\nu x.M, L, P) \geq \mathcal{P}_{\oplus}(\nu x.M, L, M\,[L/x]) = 1$, which implies $\sum_{P\in\mathtt{F}} \mathcal{P}_{\oplus}(\nu x.M, L, P) = \mathcal{P}_{\oplus}(\nu x.M, L, \mathtt{F}) = 1$. Although $\nu x.N$ is a distinguished value and the starting reasoning we have just made above still holds, $\mathcal{P}_{\oplus}(\nu x.N, L, \mathtt{F}) = \sum_{P\in\mathtt{F}} \mathcal{P}_{\oplus}(\nu x.N, L, P) = 0$, due to the fact that there is no $P \in \mathtt{F}$ of the form $N\,[L/x]$, as $N\,[L/x] \notin \mathtt{F}$ by hypothesis.

From the hypothesis $\nu x.M \sim \nu x.N$ on the equivalence class $\mathtt{F}$, *i.e.* $\mathcal{P}_{\oplus}(\nu x.M, L, \mathtt{F}) = \mathcal{P}_{\oplus}(\nu x.N, L, \mathtt{F})$, we obtain the absurd:

$$1 = \mathcal{P}_{\oplus}(\nu x.M, L, \mathtt{F}) = \mathcal{P}_{\oplus}(\nu x.N, L, \mathtt{F}) = 0.$$

($3 \Rightarrow 2$) We need to prove that, for every $L \in \Lambda_{\oplus}(\emptyset)$ and every $\mathtt{E} \in \Lambda_{\oplus}(\emptyset)/\!\sim$, $\mathcal{P}_{\oplus}(\nu x.M, L, \mathtt{E}) = \mathcal{P}_{\oplus}(\nu x.N, L, \mathtt{E})$ assuming that $M\,[L/x] \sim N\,[L/x]$ holds. First of all, let us rewrite $\mathcal{P}_{\oplus}(\nu x.M, L, \mathtt{E})$ and $\mathcal{P}_{\oplus}(\nu x.N, L, \mathtt{E})$ as $\sum_{P\in\mathtt{E}} \mathcal{P}_{\oplus}(\nu x.M, L, P)$

and $\sum_{P\in\mathtt{E}} \mathcal{P}_{\oplus}(\nu x.N, L, P)$, respectively. Then, from the hypothesis and the same reasoning we have made for (2 ⇒ 3), for every $\mathtt{E} \in \Lambda_{\oplus}(\emptyset)/\sim$:

$$\sum_{P\in\mathtt{E}} \mathcal{P}_{\oplus}(\nu x.M, L, P) = \begin{cases} 1 & \text{if } M\,[L/x] \in \mathtt{E}; \\ 0 & \text{otherwise.} \end{cases}$$
$$= \begin{cases} 1 & \text{if } N\,[L/x] \in \mathtt{E}; \\ 0 & \text{otherwise.} \end{cases} = \sum_{P\in\mathtt{E}} \mathcal{P}_{\oplus}(\nu x.N, L, P)$$

which concludes the case and the proof. □

4.3.2 $\Lambda_{\oplus}$-relations

It is here convenient to work with generalisations of relations called $\Lambda_{\oplus}$-*relations*.

Definition 4.3.8. *A* $\Lambda_{\oplus}$-relation *is a set of triples* $(\overline{x}, M, N)$*, where* $M, N \in \Lambda_{\oplus}(\overline{x})$*.*

Thus if a relation contains the pair (M, N) with $M, N \in \Lambda_{\oplus}(\overline{x})$, then the corresponding $\Lambda_{\oplus}$-relation includes $(\overline{x}, M, N)$.

Notation. Given a $\Lambda_{\oplus}$-relation $\mathcal{R}$, we use mixfix notation and write $\overline{x} \vdash M\ \mathcal{R}\ N$ to indicate that $(\overline{x}, M, N) \in \mathcal{R}$. Moreover, we just write $M\ \mathcal{R}\ N$ whenever $M, N \in \Lambda_{\oplus}(\emptyset)$. When dealing with $\Lambda_{\oplus}$-relations, we let us commit a slight abuse of notation and treat any sequence of terms $\overline{M}$ as a set (*i.e.* we forget the ordering of the terms in $\overline{M}$). Accordingly, we often write $\overline{M} \in \mathcal{P}_{\mathsf{FIN}}(\Lambda_{\oplus}(\overline{x}))$, with $\overline{x} \in \mathcal{P}_{\mathsf{FIN}}(\mathcal{V})$.

As in Pitts [Pit11], we consider the free variables in a term as implicitly λ-bound and use the property of (bi)similarity proved in Lemma 4.3.7. We then call *open extension* of probabilistic applicative (bi)similarity the following $\Lambda_{\oplus}$-relations.

Definition 4.3.9. *Given a finite set of variables* $\overline{x} \in \mathcal{P}_{\mathsf{FIN}}(\mathcal{V})$ *and* $M, N \in \Lambda_{\oplus}(\overline{x})$*, let*

$$\overline{x} \vdash M \ \precsim\ N$$

if $M\left[\overline{L}/\overline{x}\right] \precsim N\left[\overline{L}/\overline{x}\right]$ *holds for all* $\overline{L} \in \mathcal{P}_{\mathsf{FIN}}(\Lambda_{\oplus}(\emptyset))$*. Similarly, let*

$$\overline{x} \vdash M \ \sim\ N$$

if $M\left[\overline{L}/\overline{x}\right] \sim N\left[\overline{L}/\overline{x}\right]$ *holds for all* $\overline{L} \in \mathcal{P}_{\mathsf{FIN}}(\Lambda_{\oplus}(\emptyset))$*.*

We recall the definition of (pre)congruence relations in terms of $\Lambda_{\oplus}$-relations. In general, a precongruence is a compatible preorder relation, and a congruence is a compatible equivalence relation.

Definition 4.3.10. *A* $\Lambda_{\oplus}$*-relation* $\mathcal{R}$ *is*

- Symmetric *if, for every* $\overline{x} \in \mathcal{P}_{\mathsf{FIN}}(\mathcal{V})$ *and every* $M, N \in \Lambda_{\oplus}(\overline{x})$*,*

$$\overline{x} \vdash M \mathrel{\mathcal{R}} N \Rightarrow \overline{x} \vdash N \mathrel{\mathcal{R}} M \tag{Sym}$$

- Transitive *if, for every* $\overline{x} \in \mathcal{P}_{\mathsf{FIN}}(\mathcal{V})$ *and every* $M, N, L \in \Lambda_{\oplus}(\overline{x})$*,*

$$\overline{x} \vdash M \mathrel{\mathcal{R}} N \;\wedge\; \overline{x} \vdash N \mathrel{\mathcal{R}} L \Rightarrow \overline{x} \vdash M \mathrel{\mathcal{R}} L \tag{Tra}$$

- Compatible *if the following four conditions hold:*
  - *for every* $\overline{x} \in \mathcal{P}_{\mathsf{FIN}}(\mathcal{V})$*,*

$$x \in \overline{x} \Rightarrow \overline{x} \vdash x \mathrel{\mathcal{R}} x \tag{Com1}$$

  - *for every* $\overline{x} \in \mathcal{P}_{\mathsf{FIN}}(\mathcal{V})$*,* $x \in \mathcal{V} \setminus \overline{x}$ *and every* $M, N \in \Lambda_{\oplus}(\overline{x} \cup \{x\})$*,*

$$\overline{x} \cup \{x\} \vdash M \mathrel{\mathcal{R}} N \Rightarrow \overline{x} \vdash \lambda x.M \mathrel{\mathcal{R}} \lambda x.N \tag{Com2}$$

  - *for every* $\overline{x} \in \mathcal{P}_{\mathsf{FIN}}(\mathcal{V})$ *and every* $M, N, L, P \in \Lambda_{\oplus}(\overline{x})$*,*

$$\overline{x} \vdash M \mathrel{\mathcal{R}} N \wedge \overline{x} \vdash L \mathrel{\mathcal{R}} P \Rightarrow \overline{x} \vdash (M)\,L \mathrel{\mathcal{R}} \,(N)\,P \tag{Com3}$$

  - *for every* $\overline{x} \in \mathcal{P}_{\mathsf{FIN}}(\mathcal{V})$ *and every* $M, N, L, P \in \Lambda_{\oplus}(\overline{x})$*,*

$$\overline{x} \vdash M \mathrel{\mathcal{R}} N \wedge \overline{x} \vdash L \mathrel{\mathcal{R}} P \Rightarrow \overline{x} \vdash M \oplus L \mathrel{\mathcal{R}} N \oplus P \tag{Com4}$$

A $\Lambda_{\oplus}$*-relation* $\mathcal{R}$ *is a* precongruence *if it exhibits the properties* (Tra), (Com1), (Com2), (Com3) *and* (Com4)*. It is a* congruence *if it also satisfies the property* (Sym).

Observe that we do not impose a (pre)congruence $\mathcal{R}$ to be reflexive since a compatible $\Lambda_{\oplus}$-relation already is.

Lemma 4.3.11. *Every compatible* $\Lambda_{\oplus}$*-relation is reflexive.*

The next properties come in handy in the following, where proofs of compatibility results get tricky.

Lemma 4.3.12. *Let* $\mathcal{R}$ *be a* $\Lambda_{\oplus}$*-relation. If* $\mathcal{R}$ *is transitive, then the properties:*

- *for every* $\overline{x} \in \mathcal{P}_{\mathsf{FIN}}(\mathcal{V})$ *and every* $M, N, L \in \Lambda_{\oplus}(\overline{x})$*,*

$$\overline{x} \vdash M \mathrel{\mathcal{R}} N \Rightarrow \overline{x} \vdash (M)\,L \mathrel{\mathcal{R}} \,(N)\,L \tag{Com3L}$$

- *for every* $\overline{x} \in \mathcal{P}_{\mathsf{FIN}}(\mathcal{V})$ *and every* $M, N, L \in \Lambda_{\oplus}(\overline{x})$,

$$\overline{x} \vdash M \mathrel{\mathcal{R}} N \Rightarrow \overline{x} \vdash (L)\, M \mathrel{\mathcal{R}} (L)\, N \tag{Com3R}$$

together imply (Com3).

Proof. We need to show that the hypothesis $\overline{x} \vdash M \mathrel{\mathcal{R}} N$ and $\overline{x} \vdash L \mathrel{\mathcal{R}} P$ imply $\overline{x} \vdash (M)\, L \mathrel{\mathcal{R}} (N)\, P$. (Com3L) on the former, with L as pivotal term, entails $\overline{x} \vdash (M)\, L \mathrel{\mathcal{R}} (N)\, L$. Similarly, (Com3R) on the latter, with N as pivotal term, entails $\overline{x} \vdash (N)\, L \mathrel{\mathcal{R}} (N)\, P$. We conclude by the transitive property of $\mathcal{R}$. □

Lemma 4.3.13. *Let* $\mathcal{R}$ *be a* $\Lambda_{\oplus}$*-relation. If* $\mathcal{R}$ *is transitive, then the properties:*

- *for every* $\overline{x} \in \mathcal{P}_{\mathsf{FIN}}(\mathcal{V})$ *and every* $M, N, L \in \Lambda_{\oplus}(\overline{x})$,

$$\overline{x} \vdash M \mathrel{\mathcal{R}} N \Rightarrow \overline{x} \vdash M \oplus L \mathrel{\mathcal{R}} N \oplus L \tag{Com4L}$$

- *for every* $\overline{x} \in \mathcal{P}_{\mathsf{FIN}}(\mathcal{V})$ *and every* $M, N, L \in \Lambda_{\oplus}(\overline{x})$,

$$\overline{x} \vdash M \mathrel{\mathcal{R}} N \Rightarrow \overline{x} \vdash L \oplus M \mathrel{\mathcal{R}} L \oplus N \tag{Com4R}$$

together imply (Com4).

Proof. All in all similar to the one we provided for Lemma 4.3.12. □

If bisimilarity $\sim$ is a congruence, then $C[M] \sim C[N]$ whenever $M \sim N$ and C is a context. In other words, terms can be replaced by equivalent ones in any context. This is a crucial sanity-check any notion of equivalence is expected to pass.

It is well-known that proving bisimulation to be a congruence may be non-trivial when the underlying language contains higher-order functions. This is also the case here. Proving (Com1), (Com2) and (Com4) just by inspecting the operational semantics of the involved terms is indeed possible, but the method fails for (Com3), when the involved contexts contain applications. This is also related to requiring probabilistic applicative bisimilarity stable with respect to substitution of bisimilar terms, hence not necessarily the same. As a matter of fact, bisimilar terms applied to the *same term* are still bisimilar.

Lemma 4.3.14. *For all* $M, N, L \in \Lambda_{\oplus}(\emptyset)$, $M \sim N$ *implies* $(M)\, L \sim (N)\, L$.

Proof. We need to show that $M \sim N$ implies $\mathcal{P}_{\oplus}((M)\, L, \tau, \mathtt{E}) = \mathcal{P}_{\oplus}((N)\, L, \tau, \mathtt{E})$ for every $\mathtt{E} \in \mathsf{V}\Lambda_{\oplus}/\!\sim$. Definition 4.3.1, of $\Lambda_{\oplus}$ as multisorted labelled Markov chain, justifies the writing

$$\mathcal{P}_{\oplus}((M)\, L, \tau, \mathtt{E}) = \sum_{\nu x.P \in \mathsf{V}\Lambda_{\oplus}} \mathcal{P}_{\oplus}(M, \tau, \nu x.P) \cdot \mathcal{P}_{\oplus}(P\,[L/x]\,, \tau, \mathtt{E}).$$

Since each $\nu x.P$ appears in some $\sim$-equivalence class, it follows

$$
\begin{aligned}
&\sum_{\nu x.P \in \mathsf{V}\Lambda_\oplus} \mathcal{P}_\oplus(M,\tau,\nu x.P)\cdot\mathcal{P}_\oplus(P\,[L/x]\,,\tau,\mathtt{E})\\
&= \sum_{\mathtt{D}\in\mathsf{V}\Lambda_\oplus/\sim}\ \sum_{\nu x.P\in\mathtt{D}} \mathcal{P}_\oplus(M,\tau,\nu x.P)\cdot\mathcal{P}_\oplus(P\,[L/x]\,,\tau,\mathtt{E}).
\end{aligned}
$$

Now observe that in the case of different distinguished values $\nu x.P, \nu x.Q$ of $\mathtt{D}$, with $\nu x.P \sim \nu x.Q$, Lemma 4.3.7 entails $\mathcal{P}_\oplus(P\,[L/x]\,,\tau,\mathtt{E}) = \mathcal{P}_\oplus(Q\,[L/x]\,,\tau,\mathtt{E})$. In particular, this means that the quantity $\mathcal{P}_\oplus(P\,[L/x]\,,\tau,\mathtt{E})$ does not depend on the specific $\nu x.P$ we choose. Let us refer to this quantity as $\mathcal{P}_\oplus^{L,\mathtt{D},\mathtt{E}}$. Then, by using the hypothesis, it follows

$$
\begin{aligned}
\mathcal{P}_\oplus((M)\,L,\tau,\mathtt{E}) &= \sum_{\mathtt{D}\in\mathsf{V}\Lambda_\oplus/\sim}\ \sum_{\nu x.P\in\mathtt{D}} \mathcal{P}_\oplus(M,\tau,\nu x.P)\cdot\mathcal{P}_\oplus(P\,[L/x]\,,\tau,\mathtt{E})\\
&= \sum_{\mathtt{D}\in\mathsf{V}\Lambda_\oplus/\sim}\ \sum_{\nu x.P\in\mathtt{D}} \mathcal{P}_\oplus(M,\tau,\nu x.P)\cdot\mathcal{P}_\oplus^{L,\mathtt{D},\mathtt{E}}\\
&= \sum_{\mathtt{D}\in\mathsf{V}\Lambda_\oplus/\sim} \mathcal{P}_\oplus^{L,\mathtt{D},\mathtt{E}}\cdot\mathcal{P}_\oplus(M,\tau,\mathtt{D})\\
&= \sum_{\mathtt{D}\in\mathsf{V}\Lambda_\oplus/\sim} \mathcal{P}_\oplus^{L,\mathtt{D},\mathtt{E}}\cdot\mathcal{P}_\oplus(N,\tau,\mathtt{D})\\
&= \sum_{\mathtt{D}\in\mathsf{V}\Lambda_\oplus/\sim}\ \sum_{\nu x.P\in\mathtt{D}} \mathcal{P}_\oplus(N,\tau,\nu x.P)\cdot\mathcal{P}_\oplus^{L,\mathtt{D},\mathtt{E}}\\
&= \sum_{\mathtt{D}\in\mathsf{V}\Lambda_\oplus/\sim}\ \sum_{\nu x.P\in\mathtt{D}} \mathcal{P}_\oplus(N,\tau,\nu x.P)\cdot\mathcal{P}_\oplus(P\,[L/x]\,,\tau,\mathtt{E})\\
&= \mathcal{P}_\oplus((N)\,L,\tau,\mathtt{E}),
\end{aligned}
$$

which is the thesis. $\square$

Notice that the above lemma is, in fact, (Com3L). As previously mentioned, the problem turns out to be (Com3R), which is related with the following notion of *term substitutivity*.

Definition 4.3.15. *A $\Lambda_\oplus$-relation $\mathcal{R}$ is called* (term) substitutive *if for all $\overline{x} \in \mathcal{P}_{\mathsf{FIN}}(\mathcal{V})$, $x \in \mathcal{V}\setminus\overline{x}$, $M, N \in \Lambda_\oplus(\overline{x}\cup\{x\})$ and $L, P \in \Lambda_\oplus(\overline{x})$*

$$\overline{x}\cup\{x\} \vdash M\ \mathcal{R}\ N \wedge \overline{x}\vdash L\ \mathcal{R}\ P \Rightarrow \overline{x}\vdash M\,[L/x]\ \ \mathcal{R}\ N\,[P/x]\,. \tag{4.1}$$

Note that if $\mathcal{R}$ is also reflexive, then this implies

$$\overline{x}\cup\{x\} \vdash M\ \mathcal{R}\ N \wedge L\in\Lambda_\oplus(\overline{x}) \Rightarrow \overline{x}\vdash M\,[L/x]\ \ \mathcal{R}\ N\,[L/x]\,. \tag{4.2}$$

A $\Lambda_\oplus$-relations $\mathcal{R}$ is closed under term-substitution *if it satisfies (4.2).*

Because of the way open extensions are defined (Definition 4.3.9), $\sim$ and $\lesssim$ are obviously closed under term-substitution.

Unfortunately, directly prove they also enjoy the *term substitutive* property is not evident. We thus proceed indirectly by defining, starting from $\lesssim$, a new relation $\lesssim^H$, called *Howe's lifting* of $\lesssim$, that has such property by construction and that can be proved equal to $\lesssim$.

4.3.3 Howe's construction

In this section we prove that probabilistic applicative bisimilarity is indeed a congruence relation, and that its non-symmetric sibling is a precongruence. The overall structure of the proof follows Howe [How96], but we mainly refer to Pitts' survey [Pit11] on the subject.

The main idea of Howe "precongruence candidate" construction consists in turning a relation $\mathcal{R}$, on (possibly open) terms, to another relation $\mathcal{R}^H$, in such a way that, if $\mathcal{R}$ satisfies a few simple conditions, then $\mathcal{R}^H$ is a (pre)congruence including $\mathcal{R}$. The key step, then, is to prove that $\mathcal{R}^H$ is indeed a (bi)simulation. In view of Theorem 4.2.14, considering similarity $\lesssim$ suffices.

Definition 4.3.16. Howe's lifting *of any $\Lambda_\oplus$-relation $\mathcal{R}$ is the relation $\mathcal{R}^H$ inductively defined by the rules in Figure 4.4.*

$$\frac{\overline{x} \vdash x \mathrel{\mathcal{R}} M}{\overline{x} \vdash x \mathrel{\mathcal{R}^H} M}\ (\mathsf{How1}) \qquad \frac{\overline{x} \cup \{x\} \vdash M \mathrel{\mathcal{R}^H} L \qquad \overline{x} \vdash \lambda x.L \mathrel{\mathcal{R}} N \qquad x \notin \overline{x}}{\overline{x} \vdash \lambda x.M \mathrel{\mathcal{R}^H} N}\ (\mathsf{How2})$$

$$\frac{\overline{x} \vdash M \mathrel{\mathcal{R}^H} P \qquad \overline{x} \vdash N \mathrel{\mathcal{R}^H} Q \qquad \overline{x} \vdash (P)\,Q \mathrel{\mathcal{R}} L}{\overline{x} \vdash (M)\,N \mathrel{\mathcal{R}^H} L}\ (\mathsf{How3})$$

$$\frac{\overline{x} \vdash M \mathrel{\mathcal{R}^H} P \qquad \overline{x} \vdash N \mathrel{\mathcal{R}^H} Q \qquad \overline{x} \vdash P \oplus Q \mathrel{\mathcal{R}} L}{\overline{x} \vdash M \oplus N \mathrel{\mathcal{R}^H} L}\ (\mathsf{How4})$$

Figure 4.4: Howe's lifting for $\Lambda_\oplus$.

The reader familiar with Howe's method should have a sense of *déjà vu* here: indeed, this is *precisely* the same definition one finds in the realm of *non-deterministic* λ-calculus. After all, the language of terms is the same.

We now prove some properties concerning $(\cdot)^H$, hence non-specific to the current setting. We provide all the details for the sake of completeness.

The following property is the first peculiarity of Howe's method, namely that the Howe lifting $\mathcal{R}^H$ is compatible, whenever $\mathcal{R}$ is a reflexive relation. Since $\lesssim$ is a preorder (hence reflexive), instantiating $\mathcal{R}$ with $\lesssim$ results in $\lesssim^H$ compatible.

Lemma 4.3.17. *If $\mathcal{R}$ is reflexive, then $\mathcal{R}^H$ is compatible.*

Proof. We show that (Com1), (Com2), (Com3) and (Com4) hold for $\mathcal{R}^H$.

- (Com1): since $\mathcal{R}$ is reflexive, for all $x \in \overline{x}$, it holds $\overline{x} \vdash x \; \mathcal{R} \; x$. Hence rule (How1) entails $\overline{x} \vdash x \; \mathcal{R}^H \; x$.

- (Com2): since $\mathcal{R}$ is reflexive, it holds $\overline{x} \vdash \lambda x.N \; \mathcal{R} \; \lambda x.N$ with $N \in \Lambda_{\oplus}(\overline{x})$. Moreover, $\overline{x} \cup \{x\} \vdash M \; \mathcal{R}^H \; N$ as hypothesis of (Com2). Hence rule (How2) entails

$$\frac{\overline{x} \cup \{x\} \vdash M \, \mathcal{R}^H \, N \qquad \overline{x} \vdash \lambda x.N \, \mathcal{R} \, \lambda x.N \qquad x \notin \overline{x}}{\overline{x} \vdash \lambda x.M \, \mathcal{R}^H \, \lambda x.N} \; (\mathsf{How2})$$

- (Com3): since $\mathcal{R}$ is reflexive, it holds $\overline{x} \vdash (N) \, P \; \mathcal{R} \; (N) \, P$ with $N, P \in \Lambda_{\oplus}(\overline{x})$. Moreover, $\overline{x} \vdash M \; \mathcal{R}^H \; N$ and $\overline{x} \vdash L \; \mathcal{R}^H \; P$ as hypothesis of (Com3). Hence rule (How3) entails

$$\frac{\overline{x} \vdash M \, \mathcal{R}^H \, N \qquad \overline{x} \vdash L \, \mathcal{R}^H \, P \qquad \overline{x} \vdash (N) \, P \, \mathcal{R} \; (N) \, P}{\overline{x} \vdash (M) \, L \, \mathcal{R}^H \; (N) \, P} \; (\mathsf{How3})$$

- (Com4): since $\mathcal{R}$ is reflexive, it holds $\overline{x} \vdash N \oplus P \; \mathcal{R} \; N \oplus P$ with $N, P \in \Lambda_{\oplus}(\overline{x})$. Moreover, $\overline{x} \vdash M \; \mathcal{R}^H \; N$ and $\overline{x} \vdash L \; \mathcal{R}^H \; P$ as hypothesis of (Com4). Hence (How4) entails

$$\frac{\overline{x} \vdash M \, \mathcal{R}^H \, N \qquad \overline{x} \vdash L \, \mathcal{R}^H \, P \qquad \overline{x} \vdash N \oplus P \, \mathcal{R} \, N \oplus P}{\overline{x} \vdash M \oplus L \, \mathcal{R}^H \, N \oplus P} \; (\mathsf{How4})$$

This concludes the proof. $\square$

Lemma 4.3.18. *If $\mathcal{R}$ is transitive, then $\overline{x} \vdash M \; \mathcal{R}^H \; N$ and $\overline{x} \vdash N \; \mathcal{R} \; L$ imply $\overline{x} \vdash M \, \mathcal{R}^H \, L$.*

Proof. By case analysis on the last rule used in the derivation of $\overline{x} \vdash M \, \mathcal{R}^H \, N$, thus on the structure of M.

- If $M = x$, with $x \in \overline{x}$, then the hypothesis $\overline{x} \vdash x \; \mathcal{R}^H \; N$ has been derived, by using (How1) as last rule, from the hypothesis $\overline{x} \vdash x \; \mathcal{R} \; N$. The transitive property of $\mathcal{R}$

and rule (How1) together entail

$$\dfrac{\dfrac{\overline{x} \vdash x \mathrel{\mathcal{R}} N \qquad \overline{x} \vdash N \mathrel{\mathcal{R}} L}{\overline{x} \vdash x \mathrel{\mathcal{R}} L}\,(\mathsf{Tra})}{\overline{x} \vdash x \mathrel{\mathcal{R}^H} L}\,(\mathsf{How1})$$

i.e. $\overline{x} \vdash M \mathrel{\mathcal{R}^H} L$.

- If $M = \lambda x.Q$, with $Q \in \Lambda_\oplus(\overline{x} \cup \{x\})$, then the hypothesis $\overline{x} \vdash \lambda x.Q \mathrel{\mathcal{R}^H} N$ has been derived, by using (How2) as last rule, from the hypothesis $\overline{x} \cup \{x\} \vdash Q \mathrel{\mathcal{R}^H} P$ and $\overline{x} \vdash \lambda x.P \mathrel{\mathcal{R}} N$ for some $P \in \Lambda_\oplus(\overline{x} \cup \{x\})$. The transitive property of $\mathcal{R}$ and rule (How2) together entail

$$\dfrac{\overline{x} \cup \{x\} \vdash Q \mathrel{\mathcal{R}^H} P \qquad \dfrac{\overline{x} \vdash \lambda x.P \mathrel{\mathcal{R}} N \qquad \overline{x} \vdash N \mathrel{\mathcal{R}} L}{\overline{x} \vdash \lambda x.P \mathrel{\mathcal{R}} L}\,(\mathsf{Tra})}{\overline{x} \vdash \lambda x.Q \mathrel{\mathcal{R}^H} L}\,(\mathsf{How2})$$

  i.e. $\overline{x} \vdash M \mathrel{\mathcal{R}^H} L$.

- If $M = (R)\,S$, with $R, S \in \Lambda_\oplus(\overline{x})$, then the hypothesis $\overline{x} \vdash (R)\,S \mathrel{\mathcal{R}^H} N$ has been derived, by using (How3) as last rule, from the hypothesis $\overline{x} \vdash R \mathrel{\mathcal{R}^H} P$, $\overline{x} \vdash S \mathrel{\mathcal{R}^H} Q$ and $\overline{x} \vdash (P)\,Q \mathrel{\mathcal{R}} N$ for some $P, Q \in \Lambda_\oplus(\overline{x} \cup \{x\})$. The transitive property of $\mathcal{R}$ and rule (How3) together entail

$$\dfrac{\overline{x} \vdash R \mathrel{\mathcal{R}^H} P \qquad \overline{x} \vdash S \mathrel{\mathcal{R}^H} Q \qquad \dfrac{\overline{x} \vdash (P)\,Q \mathrel{\mathcal{R}} N \qquad \overline{x} \vdash N \mathrel{\mathcal{R}} L}{\overline{x} \vdash (P)\,Q \mathrel{\mathcal{R}} L}\,(\mathsf{Tra})}{\overline{x} \vdash (R)\,S \mathrel{\mathcal{R}^H} L}\,(\mathsf{How3})$$

  i.e. $\overline{x} \vdash M \mathrel{\mathcal{R}^H} L$.

- If $M = R \oplus S$, with $R, S \in \Lambda_\oplus(\overline{x})$, then the hypothesis $\overline{x} \vdash R \oplus S \mathrel{\mathcal{R}^H} N$ has been derived, by using (How4) as last rule, from the hypothesis $\overline{x} \vdash R \mathrel{\mathcal{R}^H} P$, $\overline{x} \vdash S \mathrel{\mathcal{R}^H} Q$ and $\overline{x} \vdash P \oplus Q \mathrel{\mathcal{R}} N$ for some $P, Q \in \Lambda_\oplus(\overline{x} \cup \{x\})$. The transitive property of $\mathcal{R}$ and rule (How4) together entail

$$\dfrac{\overline{x} \vdash R \mathrel{\mathcal{R}^H} P \qquad \overline{x} \vdash S \mathrel{\mathcal{R}^H} Q \qquad \dfrac{\overline{x} \vdash P \oplus Q \mathrel{\mathcal{R}} N \qquad \overline{x} \vdash N \mathrel{\mathcal{R}} L}{\overline{x} \vdash P \oplus Q \mathrel{\mathcal{R}} L}\,(\mathsf{Tra})}{\overline{x} \vdash R \oplus S \mathrel{\mathcal{R}^H} L}\,(\mathsf{How4})$$

  i.e. $\overline{x} \vdash M \mathrel{\mathcal{R}^H} L$.

This concludes the proof. □

The following property asserts that Howe's lifting $\mathcal{R}^H$ encompasses $\mathcal{R}$, whenever this latter is a reflexive relation. Since $\lesssim$ is a preorder (hence reflexive), instantiating $\mathcal{R}$ with $\lesssim$ results in $\lesssim \subseteq \lesssim^H$.

Lemma 4.3.19. *If $\mathcal{R}$ is reflexive, then $\overline{x} \vdash M \mathcal{R} N$ implies $\overline{x} \vdash M \mathcal{R}^H N$.*

Proof. By case analysis on the structure of M. Moreover, since $\mathcal{R}$ reflexive implies $\mathcal{R}^H$ compatible (Lemma 4.3.17), we use the fact that $\mathcal{R}^H$ is reflexive (Lemma 4.3.11).

- If $M = x$, with $x \in \overline{x}$, then $\overline{x} \vdash x \mathcal{R} N$. Hence rule (How1) entails $\overline{x} \vdash x \mathcal{R}^H N$, *i.e.* $\overline{x} \vdash M \mathcal{R}^H N$.

- If $M = \lambda x.L$, with $L \in \Lambda_\oplus(\overline{x} \cup \{x\})$, then $\overline{x} \vdash \lambda x.L \mathcal{R} N$. Hence the reflexive property of $\mathcal{R}^H$ and rule (How2) together entail

$$\frac{\overline{x} \cup \{x\} \vdash L \mathcal{R}^H L \qquad \overline{x} \vdash \lambda x.L \mathcal{R} N \qquad x \notin \overline{x}}{\overline{x} \vdash \lambda x.L \mathcal{R}^H N} \text{(How2)}$$

  i.e. $\overline{x} \vdash M \mathcal{R}^H N$.

- If $M = (L) P$, with $L, P \in \Lambda_\oplus(\overline{x})$, then $\overline{x} \vdash (L) P \mathcal{R} N$. Hence the reflexive property of $\mathcal{R}^H$ and rule (How3) together entail

$$\frac{\overline{x} \vdash L \mathcal{R}^H L \qquad \overline{x} \vdash P \mathcal{R}^H P \qquad \overline{x} \vdash (L) P \mathcal{R} N}{\overline{x} \vdash (L) P \mathcal{R}^H N} \text{(How3)}$$

  i.e. $\overline{x} \vdash M \mathcal{R}^H N$.

- If $M = L \oplus P$, with $L, P \in \Lambda_\oplus(\overline{x})$, then $\overline{x} \vdash L \oplus P \mathcal{R} N$. Hence the reflexive property of $\mathcal{R}^H$ and rule (How4) together entail

$$\frac{\overline{x} \vdash L \mathcal{R}^H L \qquad \overline{x} \vdash P \mathcal{R}^H P \qquad \overline{x} \vdash L \oplus P \mathcal{R} N}{\overline{x} \vdash L \oplus P \mathcal{R}^H N} \text{(How4)}$$

  i.e. $\overline{x} \vdash M \mathcal{R}^H N$.

This concludes the proof. □

The next property is the second peculiarity of Howe's method. It asserts $\mathcal{R}^H$ term substitutive, whenever $\mathcal{R}$ is a preorder and closed under term-substitution.

Lemma 4.3.20. *If $\mathcal{R}$ is reflexive, transitive and closed under term-substitution, then $\mathcal{R}^H$ is (term) substitutive and hence also closed under term-substitution.*

Proof. We show $\mathcal{R}^H$ term substitutive (Definition 4.1) by induction on the derivation of $\overline{x} \cup \{x\} \vdash M \mathcal{R}^H N$, thus on the structure of M.

- If M is a variable, then either $M = x$ or $M \in \overline{x}$. In the latter case, suppose $M = y$. Then $\overline{x} \cup \{x\} \vdash y \ \mathcal{R}^H \ N$, and the only way to deduce it is by rule (How1) from $\overline{x} \cup \{x\} \vdash y \ \mathcal{R} \ N$. Hence, by the fact $\mathcal{R}$ is closed under term-substitution and $P \in \Lambda_\oplus(\overline{x})$, $\overline{x} \vdash y\,[P/x] \ \ \mathcal{R} \ \ N\,[P/x]$ which is equivalent to $\overline{x} \vdash y \ \mathcal{R} \ N\,[P/x]$. Finally, Lemma 4.3.19 entails $\overline{x} \vdash y \ \mathcal{R}^H \ N\,[P/x]$, which is equivalent to $\overline{x} \vdash y\,[L/x] \ \ \mathcal{R}^H \ N\,[P/x]$, *i.e.* $\overline{x} \vdash M\,[L/x] \ \ \mathcal{R}^H \ N\,[P/x]$. Otherwise, $M = x$ and $\overline{x} \cup \{x\} \vdash x \ \mathcal{R}^H \ N$. The only way to deduce the latter is by the rule (How1) from $\overline{x} \cup \{x\} \vdash x \ \mathcal{R} \ N$. Hence, by the fact $\mathcal{R}$ is closed under term-substitution and $P \in \Lambda_\oplus(\overline{x})$, $\overline{x} \vdash x\,[P/x] \ \ \mathcal{R} \ \ N\,[P/x]$ which is equivalent to $\overline{x} \vdash P \ \mathcal{R} \ N\,[P/x]$. Thus Lemma 4.3.18 entails

$$\frac{\overline{x} \vdash L \ \mathcal{R}^H \ P \qquad \overline{x} \vdash P \ \mathcal{R} \ N\,[P/x]}{\overline{x} \vdash L \ \mathcal{R}^H \ N\,[P/x]}$$

which is equivalent to $\overline{x} \vdash x\,[L/x] \ \ \mathcal{R}^H \ N\,[P/x]$, *i.e.* $\overline{x} \vdash M\,[L/x] \ \ \mathcal{R}^H \ N\,[P/x]$.

- If $M = \lambda y.Q$, with $Q \in \Lambda_\oplus(\overline{x} \cup \{x, y\})$, then $\overline{x} \cup \{x\} \vdash \lambda y.Q \ \mathcal{R}^H \ N$. The only way to deduce the latter is by rule (How2) as follows:

$$\frac{\overline{x} \cup \{x,y\} \vdash Q \ \mathcal{R}^H \ R \qquad \overline{x} \cup \{x\} \vdash \lambda y.R \ \mathcal{R} \ N \qquad x, y \notin \overline{x}}{\overline{x} \cup \{x\} \vdash \lambda y.Q \ \mathcal{R}^H \ N} \ (\mathsf{How2})$$

Let us denote $\overline{y} = \overline{x} \cup \{y\}$. By the induction hypothesis on $\overline{y} \cup \{x\} \vdash Q \ \mathcal{R}^H \ R$ follows $\overline{y} \vdash Q\,[L/x] \ \ \mathcal{R}^H \ \ R\,[P/x]$. Moreover, by the fact $\mathcal{R}$ is closed under term-substitution and $P \in \Lambda_\oplus(\overline{x})$, $\overline{x} \vdash (\lambda y.R)\,[P/x] \ \ \mathcal{R} \ \ N\,[P/x]$, *i.e.* $\overline{x} \vdash \lambda y.R\,[P/x] \ \ \mathcal{R} \ N\,[P/x]$. Thus rule (How2) entails

$$\frac{\overline{x} \cup \{y\} \vdash Q\,[L/x] \ \ \mathcal{R}^H \ R\,[P/x] \qquad \overline{x} \vdash \lambda y.R\,[P/x] \ \ \mathcal{R} \ N\,[P/x] \qquad y \notin \overline{x}}{\overline{x} \vdash \lambda y.Q\,[L/x] \ \ \mathcal{R}^H \ N\,[P/x]} \ (\mathsf{How2})$$

which is equivalent to $\overline{x} \vdash (\lambda y.Q)\,[L/x] \ \ \mathcal{R}^H \ N\,[P/x]$, *i.e.* $\overline{x} \vdash M\,[L/x] \ \ \mathcal{R}^H \ N\,[P/x]$.

- If $M = (Q)\,R$, with $Q, R \in \Lambda_\oplus(\overline{x} \cup \{x\})$, then $\overline{x} \cup \{x\} \vdash (Q)\,R \ \mathcal{R}^H \ N$. The only way to deduce the latter is by rule (How3) as follows:

$$\frac{\overline{x} \cup \{x\} \vdash Q \ \mathcal{R}^H \ Q' \qquad \overline{x} \cup \{x\} \vdash R \ \mathcal{R}^H \ R' \qquad \overline{x} \cup \{x\} \vdash (Q')\,R' \ \mathcal{R} \ N}{\overline{x} \cup \{x\} \vdash (Q)\,R \ \mathcal{R}^H \ N} \ (\mathsf{How3})$$

By the induction hypothesis on $\overline{x} \cup \{x\} \vdash Q \ \mathcal{R}^H \ Q'$ and $\overline{x} \cup \{x\} \vdash R \ \mathcal{R}^H \ R'$ follows $\overline{x} \vdash Q\,[L/x] \ \ \mathcal{R}^H \ \ Q'\,[P/x]$ and $\overline{x} \vdash R\,[L/x] \ \ \mathcal{R}^H \ \ R'\,[P/x]$. Moreover, by the fact $\mathcal{R}$ is closed under term-substitution and $P \in \Lambda_\oplus(\overline{x})$, $\overline{x} \vdash ((Q')\,R')\,[P/x] \ \ \mathcal{R} \ N\,[P/x]$, *i.e.* $\overline{x} \vdash (Q'\,[P/x])\,R'\,[P/x] \ \mathcal{R} \ N\,[P/x]$. Rule (How3) on the hypothesis

- $\overline{x} \vdash Q\,[L/x]\ \ \mathcal{R}^H\ Q'\,[P/x]$
- $\overline{x} \vdash R\,[L/x]\ \ \mathcal{R}^H\ R'\,[P/x]$
- $\overline{x} \vdash (Q'\,[P/x])\,R'\,[P/x]\ \mathcal{R}\ N\,[P/x]$

entails $\overline{x} \vdash (Q\,[L/x])\,R\,[L/x]\ \ \mathcal{R}^H\ \ N\,[P/x]$. This latter is equivalent to $\overline{x} \vdash ((Q)\,R)\,[L/x]\ \ \mathcal{R}^H\ N\,[P/x]$, *i.e.* $\overline{x} \vdash M\,[L/x]\ \ \mathcal{R}^H\ N\,[P/x]$.

- If $M = Q \oplus R$, with $Q, R \in \Lambda_{\oplus}(\overline{x} \cup \{x\})$, then $\overline{x} \cup \{x\} \vdash Q \oplus R\ \mathcal{R}^H\ N$. The only way to deduce the latter is by rule (How4) as follows:

$$\frac{\overline{x} \cup \{x\} \vdash Q\ \mathcal{R}^H\ Q' \qquad \overline{x} \cup \{x\} \vdash R\ \mathcal{R}^H\ R' \qquad \overline{x} \cup \{x\} \vdash Q' \oplus R'\ \mathcal{R}\ N}{\overline{x} \cup \{x\} \vdash Q \oplus R\ \mathcal{R}^H\ N}\ (\mathsf{How4})$$

By the induction hypothesis on $\overline{x} \cup \{x\} \vdash Q\ \mathcal{R}^H\ Q'$ and $\overline{x} \cup \{x\} \vdash R\ \mathcal{R}^H\ R'$ follows $\overline{x} \vdash Q\,[L/x]\ \ \mathcal{R}^H\ \ Q'\,[P/x]$ and $\overline{x} \vdash R\,[L/x]\ \ \mathcal{R}^H\ \ R'\,[P/x]$. Moreover, by the fact $\mathcal{R}$ is closed under term-substitution and $P \in \Lambda_{\oplus}(\overline{x})$, $\overline{x} \vdash (Q' \oplus R')\,[P/x]\ \ \mathcal{R}\ N\,[P/x]$, *i.e.* $\overline{x} \vdash Q'\,[P/x] \oplus R'\,[P/x]\ \mathcal{R}\ N\,[P/x]$. Rule (How3) on the hypothesis

- $\overline{x} \vdash Q\,[L/x]\ \ \mathcal{R}^H\ Q'\,[P/x]$
- $\overline{x} \vdash R\,[L/x]\ \ \mathcal{R}^H\ R'\,[P/x]$
- $\overline{x} \vdash Q'\,[P/x] \oplus R'\,[P/x]\ \mathcal{R}\ N\,[P/x]$

entails $\overline{x} \vdash Q\,[L/x] \oplus R\,[L/x]\ \ \mathcal{R}^H\ \ N\,[P/x]$. This latter is equivalent to $\overline{x} \vdash (Q \oplus R)\,[L/x]\ \ \mathcal{R}^H\ N\,[P/x]$, *i.e.* $\overline{x} \vdash M\,[L/x]\ \ \mathcal{R}^H\ N\,[P/x]$.

This concludes the proof. $\square$

Something is missing, however, before we can conclude that $\lesssim^H$ is a precongruence, namely transitivity. We also follow Howe here and consider the transitive closure $(\lesssim^H)^+$, which is a preorder by construction.

Definition 4.3.21. *The* transitive closure *of any $\Lambda_{\oplus}$-relation $\mathcal{R}$ is the relation $\mathcal{R}^+$ inductively defined by the rules in Figure 4.5.*

Obviously, any $\mathcal{R}^+$ is transitive by definition. The crucial point is that it is rather simple to prove $\mathcal{R}^+$ compatible and closed under term-substitution if $\mathcal{R}$ already is. In particular, this means that the good properties we have established by means of Howe's construction remain valid.

Lemma 4.3.22. *If $\mathcal{R}$ is compatible, then so is $\mathcal{R}^+$.*

Proof. We show that (Com1), (Com2), (Com3) and (Com4) hold for $\mathcal{R}^+$.

$$\frac{\overline{x} \vdash M \mathrel{\mathcal{R}} N}{\overline{x} \vdash M \mathrel{\mathcal{R}^+} N} \;(\mathsf{TC1})$$

$$\frac{\overline{x} \vdash M \mathrel{\mathcal{R}^+} N \qquad \overline{x} \vdash N \mathrel{\mathcal{R}^+} L}{\overline{x} \vdash M \mathrel{\mathcal{R}^+} L} \;(\mathsf{TC2})$$

Figure 4.5: Transitive closure for $\Lambda_\oplus$.

- ($\mathsf{Com1}$): since $\mathcal{R}$ is compatible (hence reflexive), it holds $\overline{x} \vdash x \mathrel{\mathcal{R}} x$. Hence rule $(\mathsf{TC1})$ entails $\overline{x} \vdash x \mathrel{\mathcal{R}^+} x$.
- ($\mathsf{Com2}$): by induction on the derivation of $\overline{x} \cup \{x\} \vdash M \mathrel{\mathcal{R}^+} N$ (Definition 4.3.10 *w.r.t.* $\mathcal{R}^+$), looking at the last rule used. The base case has $(\mathsf{TC1})$ as last rule, hence $\overline{x} \cup \{x\} \vdash M \mathrel{\mathcal{R}} N$. Since $\mathcal{R}$ is compatible, it follows $\overline{x} \vdash \lambda x.M \mathrel{\mathcal{R}} \lambda x.N$. Hence rule $(\mathsf{TC1})$ entails $\overline{x} \vdash \lambda x.M \mathrel{\mathcal{R}^+} \lambda x.N$. Otherwise, $(\mathsf{TC2})$ is the last rule used and, for some $L \in \Lambda_\oplus(\overline{x} \cup \{x\})$, $\overline{x} \cup \{x\} \vdash M \mathrel{\mathcal{R}^+} L$ and $\overline{x} \cup \{x\} \vdash L \mathrel{\mathcal{R}^+} N$. By the induction hypothesis on both of these latter, it follows $\overline{x} \vdash \lambda x.M \mathrel{\mathcal{R}^+} \lambda x.L$ and $\overline{x} \vdash \lambda x.L \mathrel{\mathcal{R}^+} \lambda x.N$. Hence rule $(\mathsf{TC2})$ entails $\overline{x} \vdash \lambda x.M \mathrel{\mathcal{R}^+} \lambda x.N$.
- ($\mathsf{Com3}$): we first prove the following two properties: for every $\overline{x} \in \mathcal{P}_{\mathsf{FIN}}(\mathcal{V})$ and every $M, N, L, P \in \Lambda_\oplus(\overline{x})$,

$$\overline{x} \vdash M \mathrel{\mathcal{R}^+} N \;\wedge\; \overline{x} \vdash L \mathrel{\mathcal{R}} P \Rightarrow \overline{x} \vdash (M)\,L \mathrel{\mathcal{R}^+} (N)\,P, \tag{4.3}$$

$$\overline{x} \vdash M \mathrel{\mathcal{R}} N \;\wedge\; \overline{x} \vdash L \mathrel{\mathcal{R}^+} P \Rightarrow \overline{x} \vdash (M)\,L \mathrel{\mathcal{R}^+} (N)\,P. \tag{4.4}$$

  In particular, we only detail (4.3), as (4.4) is similarly provable. We show (4.3) by induction on the derivation $\overline{x} \vdash M \mathrel{\mathcal{R}^+} N$, looking at the last rule used. The base case has $(\mathsf{TC1})$ as last rule, hence $\overline{x} \vdash M \mathrel{\mathcal{R}} N$. Since $\mathcal{R}$ is compatible and $\overline{x} \vdash L \mathrel{\mathcal{R}} P$, it follows $\overline{x} \vdash (M)\,L \mathrel{\mathcal{R}} (N)\,P$. Hence rule $(\mathsf{TC1})$ entails $\overline{x} \vdash (M)\,L \mathrel{\mathcal{R}^+} (N)\,P$. Otherwise, if $(\mathsf{TC2})$ is the last rule used and, for some $Q \in \Lambda_\oplus$, $\overline{x} \vdash M \mathrel{\mathcal{R}^+} Q$ and $\overline{x} \vdash Q \mathrel{\mathcal{R}^+} N$. By the induction hypothesis on $\overline{x} \vdash M \mathrel{\mathcal{R}^+} Q$, along with $\overline{x} \vdash L \mathrel{\mathcal{R}} P$, it follows $\overline{x} \vdash (M)\,L \mathrel{\mathcal{R}^+} (Q)\,P$. Since $\mathcal{R}$ is compatible (hence reflexive), $\overline{x} \vdash P \mathrel{\mathcal{R}} P$. By the induction hypothesis on $\overline{x} \vdash Q \mathrel{\mathcal{R}^+} N$, along with this latter, it follows $\overline{x} \vdash (Q)\,P \mathrel{\mathcal{R}^+} (N)\,P$. Hence rule $(\mathsf{TC2})$ on $\overline{x} \vdash (M)\,L \mathrel{\mathcal{R}^+} (Q)\,P$ and $\overline{x} \vdash (Q)\,P \mathrel{\mathcal{R}^+} (N)\,P$ entails $\overline{x} \vdash (M)\,L \mathrel{\mathcal{R}^+} (N)\,P$.

  Let us prove ($\mathsf{Com3}$) by induction on the two derivations $\overline{x} \vdash M \mathrel{\mathcal{R}^+} N$ and $\overline{x} \vdash L \mathrel{\mathcal{R}^+} P$ (Definition 4.3.10 *w.r.t.* $\mathcal{R}^+$), which we name here as π and ρ respectively. Looking at the last rules used, there are four possible cases as four are the combinations that allow to conclude with π and ρ:

1. $(\mathsf{TC1})$ for both π and ρ;
2. $(\mathsf{TC1})$ for π and $(\mathsf{TC2})$ for ρ;
3. $(\mathsf{TC2})$ for π and $(\mathsf{TC1})$ for ρ;
4. $(\mathsf{TC2})$ for both π and ρ.

Observe now that the first three cases are addressed by (4.3) and (4.4). Hence, it remains to prove the last case, where both derivations are concluded by applying rule $(\mathsf{TC2})$. According to rule $(\mathsf{TC2})$ definition, two additional hypothesis follow from each derivation: in the case of π, it follows that, for some $Q \in \Lambda_\oplus(\overline{x})$, $\overline{x} \vdash M\ \mathcal{R}^+\ Q$ and $\overline{x} \vdash Q\ \mathcal{R}^+\ N$; in the case of ρ, it follows, for some $R \in \Lambda_\oplus(\overline{x})$, $\overline{x} \vdash L\ \mathcal{R}^+\ R$ and $\overline{x} \vdash R\ \mathcal{R}^+\ P$. By a double induction hypothesis, first on $\overline{x} \vdash M\ \mathcal{R}^+\ Q$, $\overline{x} \vdash L\ \mathcal{R}^+\ R$ and then on $\overline{x} \vdash Q\ \mathcal{R}^+\ N$, $\overline{x} \vdash R\ \mathcal{R}^+\ P$, it follows $\overline{x} \vdash (M)\,L\ \mathcal{R}^+\ (Q)\,R$ and $\overline{x} \vdash (Q)\,R\ \mathcal{R}^+\ (N)\,P$ respectively. Hence rule $(\mathsf{TC2})$ entails $\overline{x} \vdash (M)\,L\ \mathcal{R}^+\ (N)\,P$.

- We do not detail the proof of $(\mathsf{Com4})$ since it is similar to that of $(\mathsf{Com3})$, where probabilistic sum operator plays the role of application constructor.

This concludes the proof. $\square$

Lemma 4.3.23. *If $\mathcal{R}$ is closed under term-substitution, then so is $\mathcal{R}^+$.*

Proof. We prove $\mathcal{R}^+$ closed under term-substitution (Definition 4.1) by induction on the derivation of $\overline{x} \cup \{x\} \vdash M\ \mathcal{R}^+\ N$, looking at the last rule used. The base case has $(\mathsf{TC1})$ as last rule, hence $\overline{x} \cup \{x\} \vdash M\ \mathcal{R}\ N$. Since $\mathcal{R}$ is closed under term-substitution, it follows $\overline{x} \vdash M\,[L/x]\ \mathcal{R}\ N\,[L/x]$. Hence rule $(\mathsf{TC1})$ entails $\overline{x} \vdash M\,[L/x]\ \mathcal{R}^+\ N\,[L/x]$. Otherwise, if $(\mathsf{TC2})$ is the last rule used and, for some $P \in \Lambda_\oplus(\overline{x} \cup \{x\})$, $\overline{x} \cup \{x\} \vdash M\ \mathcal{R}^+\ P$ and $\overline{x} \cup \{x\} \vdash P\ \mathcal{R}^+\ N$. By the induction hypothesis on both of these latter, it follows $\overline{x} \vdash M\,[L/x]\ \mathcal{R}^+\ P\,[L/x]$ and $\overline{x} \vdash P\,[L/x]\ \mathcal{R}^+\ N\,[L/x]$. Hence rule $(\mathsf{TC2})$ entails $\overline{x} \vdash M\,[L/x]\ \mathcal{R}^+\ N\,[L/x]$. $\square$

Lemma 4.3.24. *If a $\Lambda_\oplus$-relation $\mathcal{R}$ is a preorder relation, then so is $(\mathcal{R}^H)^+$.*

Proof. We show $(\mathcal{R}^H)^+$ reflexive and transitive. Of course, being a transitive closure, $(\mathcal{R}^H)^+$ is a a transitive relation. Since $\mathcal{R}$ is reflexive, Lemma 4.3.17 implies $\mathcal{R}^H$ compatible, hence reflexive. Lemma 4.3.22 entails the same for $(\mathcal{R}^H)^+$. $\square$

Probabilistic applicative bisimilarity is a congruence

As previously mentioned, we ultimately need to prove that $(\lesssim^H)^+$ is a simulation. Since we already know the latter is a preorder (Lemma 4.3.24), the following *Key Lemma* gives us the missing bit.

Notation. Given a $\Lambda_{\oplus}$-relation $\mathcal{R}$, by a slight abuse of notation, we write $\lambda x.\mathcal{R}(X)$ for the set of values $\lambda x.\mathcal{R}(X) = \{\lambda x.M \mid \exists N \in X.\, N \;\mathcal{R}\; M\}$. Similarly, in the case of distinguished values, $\nu x.\mathcal{R}(X) = \{\nu x.M \mid \exists N \in X.\, N \;\mathcal{R}\; M\}$

Lemma 4.3.25. *If $M \precsim^{H} N$, then for every $X \subseteq \Lambda_{\oplus}(x)$ it holds:*

$$[\![M]\!]\,(\lambda x.X) \leq [\![N]\!]\,(\lambda x.(\precsim^{H}(X))).$$

The proof of this lemma is delicate and is discussed in the next section. From the lemma, using a standard argument we derive the needed substitutivity results, and ultimately the most important result of this section.

Theorem 4.3.26. *On $\Lambda_{\oplus}$-terms, $\precsim$ is a precongruence.*

Proof. We prove $\precsim$ is a precongruence by observing that $(\precsim^{H})^{+}$ is a precongruence and by showing that $\precsim \;=\; (\precsim^{H})^{+}$. Lemma 4.3.17 and Lemma 4.3.22 imply that $(\precsim^{H})^{+}$ is compatible and Lemma 4.3.24 tells us that $(\precsim^{H})^{+}$ is a preorder. As a consequence, $(\precsim^{H})^{+}$ is a precongruence. Consider now the inclusion $\precsim \,\subseteq (\precsim^{H})^{+}$. By Lemma 4.3.19 and definition of transitive closure operator $(\cdot)^{+}$, it follows that $\precsim \,\subseteq\, (\precsim^{H}) \,\subseteq\, (\precsim^{H})^{+}$. We show the converse by proving that $(\precsim^{H})^{+}$ is included in a relation $\mathcal{R}$ that is a probabilistic applicative simulation, therefore contained in the largest one $\precsim$. In particular, since $(\precsim^{H})^{+}$ is closed under term-substitution (Lemma 4.3.20 and Lemma 4.3.23), it suffices to show the latter only on the closed version of terms and cloned values. $\mathcal{R}$ acts like $(\precsim^{H})^{+}$ on terms, while given two cloned values $\nu x.M$ and $\nu x.N$, $\nu x.M \;\mathcal{R}\; \nu x.N$ if and only if $M \;(\precsim^{H})^{+}\; N$. Since we already know that $(\precsim^{H})^{+}$ is a preorder, which implies the same for $\mathcal{R}$, it remains to prove the clauses for establishing $\mathcal{R}$ simulation:

- For every $M, N \in \Lambda_{\oplus}(\emptyset)$ with $M \;\mathcal{R}\; N$, and every $X \subseteq \Lambda_{\oplus}(x)$, we show

$$\mathcal{P}_{\oplus}(M, \tau, \nu x.X) \leq \mathcal{P}_{\oplus}(N, \tau, \mathcal{R}(\nu x.X)). \tag{4.5}$$

  Observe that $M \;\mathcal{R}\; N$ implies $M \;(\precsim^{H})^{+}\; N$ by definition, hence we proceed by induction on the structure of the derivation of $M \;(\precsim^{H})^{+}\; N$:

  – If $(\mathsf{TC1})$ is the last rule, then $\emptyset \vdash M \precsim^{H} N$ holds. It follows

$$\begin{aligned}
\mathcal{P}_{\oplus}(M, \tau, \nu x.X) &= [\![M]\!]\,(\nu x.X) \\
&\leq [\![N]\!]\,(\nu x.\precsim^{H}(X)) && [\text{Lemma 4.3.25}] \\
&\leq [\![N]\!]\,(\nu x.(\precsim^{H})^{+}(X)) && [(\mathsf{TC1})] \\
&\leq [\![N]\!]\,(\mathcal{R}(\nu x.X)) && [\text{Definition of } \mathcal{R}] \\
&= \mathcal{P}_{\oplus}(N, \tau, \mathcal{R}(\nu x.X)).
\end{aligned}$$

– If (TC2) is the last rule used, then there is $P \in \Lambda_{\oplus}(\emptyset)$ with $\emptyset \vdash M\ (\lesssim^H)^+\ P$ and $\emptyset \vdash P\ (\lesssim^H)^+\ N$. By the induction hypothesis, we get

$$\mathcal{P}_{\oplus}(M,\tau,X) \leq \mathcal{P}_{\oplus}(P,\tau,\mathcal{R}(X)),$$
$$\mathcal{P}_{\oplus}(P,\tau,\mathcal{R}(X)) \leq \mathcal{P}_{\oplus}(N,\tau,\mathcal{R}(\mathcal{R}(X))).$$

Since $\mathcal{R}(\mathcal{R}(X)) \subseteq \mathcal{R}(X)$, the above two inequality together imply (4.5), *i.e.* $\mathcal{P}_{\oplus}(M,\tau,X) \leq \mathcal{P}_{\oplus}(N,\tau,\mathcal{R}(X))$.

- For every $\nu x.M, \nu x.N \in \mathsf{V}\Lambda_{\oplus}(\emptyset)$ with $\nu x.M\ \mathcal{R}\ \nu x.N$, for every $L \in \Lambda_{\oplus}(\emptyset)$ and every $X \subseteq \Lambda_{\oplus}(\emptyset)$, we show

$$\mathcal{P}_{\oplus}(\nu x.M, L, X) \leq \mathcal{P}_{\oplus}(\nu x.N, L, \mathcal{R}(X)). \tag{4.6}$$

Observe that $\nu x.M\ \mathcal{R}\ \nu x.N$ implies $M\ (\lesssim^H)^+\ N$ by definition, hence $M[L/x]\ (\lesssim^H)^+\ N[L/x]$ by the fact that $(\lesssim^H)^+$ is closed under term-substitution. This means that whenever $M[L/x] \in X$, then $N[L/x] \in \lesssim^H(X)$ and so $N[L/x] \in (\lesssim^H)^+(X)$, ultimately entailing that the inequality (4.6) is satisfied as

$$\begin{aligned}\mathcal{P}_{\oplus}(\nu x.M, L, X) &= 1 \\ &= \mathcal{P}_{\oplus}(\nu x.N, L, (\lesssim^H)^+(X)) \\ &= \mathcal{P}_{\oplus}(\nu x.N, L, \mathcal{R}(X)),\end{aligned}$$

On the other hand, if $M[L/x] \notin X$, then the inequality (4.6) is satisfied as $\mathcal{P}_{\oplus}(\nu x.M, L, X) = 0 \leq \mathcal{P}_{\oplus}(\nu x.N, L, \mathcal{R}(X))$. □

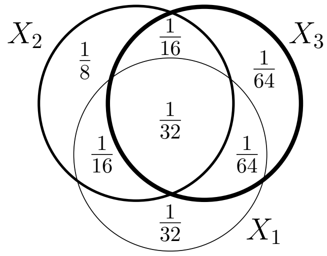
Corollary 4.3.27. *On $\Lambda_{\oplus}$-terms, $\sim$ is a congruence.*

Proof. Relation $\sim$ is an equivalence by definition, in particular a symmetric relation. Since $\sim = \lesssim \cap \lesssim^{op}$ (Theorem 4.2.14), $\sim$ is also compatible as a consequence of Theorem 4.3.26. □

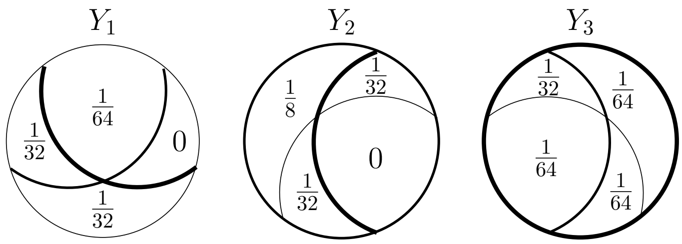
4.4 Proof of the Key Lemma

We devote this section only to proving the Key Lemma 4.3.25. The proof turns out to be much more difficult than the corresponding ones for deterministic or non-deterministic cases [Pit11]. In particular, the case when M is an application relies on another technical lemma we give in the following, which itself can be proved by tools from linear programming.

The combinatorial problem we face while proving the Key Lemma can actually be decontextualized and understood independently. Suppose we have $n = 3$



(a)



(b)

Figure 4.6: Disentangling sets

non-disjoint sets X_1, X_2, X_3 whose elements are labelled with real numbers. As an example, we could be in a situation like the one in Figure 4.6a (where for the sake of simplicity only the labels are indicated). We fix three real numbers $p_1 = \frac{5}{64}$, $p_2 = \frac{3}{16}$, $p_3 = \frac{5}{64}$. It is routine to check that for every $I \subseteq \{1, 2, 3\}$ it holds that

$$\sum_{i \in I} p_i \leq \left\| \bigcup_{i \in I} X_i \right\|,$$

where $\|X\|$ is the sum of the labels of the elements of X. Observe that it is of course possible to turn the three sets X_1, X_2, X_3 into three disjoint sets Y_1, Y_2 and Y_3 where each Y_i contains (copies of) the elements of X_i whose labels, however, are obtained by splitting the ones of the original elements. Examples of those sets are in Figure 4.6b: if we superpose the three sets, we obtain the Venn diagram we started from. Quite remarkably, however, the examples from Figure 4.6b have an additional property, namely that for every $i \in \{1, 2, 3\}$ it holds that $p_i \leq \|Y_i\|$. We now show that finding sets satisfying the properties above is always possible, even when n is arbitrary.

4.4.1 Disentangling probability assignments

The scenario just described can be formally defined as follows:

Definition 4.4.1. *Let $p_1, \ldots, p_n \in \mathbb{R}_{[0,1]}$, and for each $I \subseteq \{1, \ldots, n\}$ let $r_I \in \mathbb{R}_{[0,1]}$ be defined such that for every such I it holds that $\sum_{i \in I} p_i \leq \sum_{J \cap I \neq \emptyset} r_J \leq 1$. Then $(\{p_i\}_{1 \leq i \leq n}, \{r_I\}_{I \subseteq \{1,\ldots,n\}})$ is said to be a* probability assignment *for $\{1, \ldots, n\}$.*

Notation. Probability assignments are written as $\mathbf{P}, \mathbf{Q}$.

We now focus on showing that it is always possible to "disentangle" probability assignments. The idea is to model any probabilitity assignment $\mathbf{P}$ as a flow network $\mathcal{N}_{\mathbf{P}}$, on which we apply the well-known *Max-flow Min-cut theorem*.

In our case a flow network is a special class of $\mathbb{R}_{[0,1]}$-weighted digraph $\mathcal{N} = (\mathcal{V}, \mathcal{E})$ with $\mathcal{V}$ vertices and edges $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$. The graph exhibits two special vertices s and t: the former, called *source*, does not have ingoing edges whereas the latter, called *target*, does not have outgoing edges. A *flow* from s to t is any function $f : \mathcal{E} \to \mathbb{R}_{[0,1]}$ such that for all vertices v, different from s and t, the following conservation principle holds:

$$\sum_{w \in \mathcal{E}^+(v)} f(v, w) = \sum_{z \in \mathcal{E}^-(v)} f(z, v)$$

where $\mathcal{E}^+(v) = \{w \in \mathcal{V} \mid (v, w) \in \mathcal{E}\}$ and $\mathcal{E}^-(v) = \{z \in \mathcal{V} \mid (z, v) \in \mathcal{E}\}$. The flow passing through an edge is bounded by the *capacity* of the edge, which is specified as a function $c : \mathcal{E} \to \mathbb{R}_{[0,1]}$. The *value* of a flow is defined as $\sum_{v \in \mathcal{V}} f(s, v)$. A *cut* $C = (S, T)$ is a partition of $\mathcal{V}$ such that $s \in S$ and $t \in T$, with a *cut capacity* defined as $\sum\{c(v, w) \mid v \in S \text{ and } w \in T\}$.

It is rather simple to observe that there exists a cut whose cut capacity is minimal, and that there exists a flow with maximal value. Moreover, the value of this maximal flow is bounded by the value of the minimal cut. The Max-flow Min-cut theorem asserts that this inequality is, in fact, an equality.

Theorem 4.4.2. *For any flow network, the value of the maximal flow is equal to the capacity of the minimal cut.*

The following is the crucial technical lemma of this section, which allows us to conclude the case of application terms in the proof of Key Lemma (Lemma 4.3.25).

Lemma 4.4.3. *Let $\mathbf{P} = (\{p_i\}_{1 \leq i \leq n}, \{r_I\}_{I \subseteq \{1,\ldots,n\}})$ be a probability assignment. Then for every non-empty $I \subseteq \{1, \ldots, n\}$ and for every $k \in I$ there is $s_{k,I} \in \mathbb{R}_{[0,1]}$ such that the following conditions all hold:*

1. *for every I, it holds that $\sum_{k \in I} s_{k,I} \leq 1$;*
2. *for every $k \in \{1, \ldots, n\}$, it holds that $p_k \leq \sum_{k \in I} s_{k,I} \cdot r_I$.*

Proof. We first detail how a flow network is defined on the basis of the probability assignment $\mathbf{P}$.

Definition. *The* flow network of $\mathbf{P}$ *is the digraph* $\mathcal{N}_{\mathbf{P}} = (\mathcal{V}_{\mathbf{P}}, \mathcal{E}_{\mathbf{P}})$ *defined as follows:*

- $\mathcal{V}_{\mathbf{P}} = (\mathcal{P}(\{1, \ldots, n\}) \setminus \varnothing) \cup \{s, t\}$*;*
- $\mathcal{E}_{\mathbf{P}}$ *is composed by three kinds of edges:*
  - $(s, \{i\})$ *for every* $i \in \{1, \ldots, n\}$*, with an assigned capacity of* p_i*;*
  - $(I, I \cup \{i\})$*, for every non-empty* $I \subseteq \{1, \ldots, n\}$ *and* $i \notin I$*, with an assigned capacity of* 1*;*
  - (I, t)*, for every non-empty* $I \subseteq \{1, \ldots, n\}$*, with an assigned capacity of* r_I*.*

We now prove the following two lemmas on $\mathcal{N}_{\mathbf{P}}$ which together directly entail the result. In the following, we write p for $\sum_{i=1}^{n} p_i$.

- The first lemma proves the result of Lemma 4.4.3 under the hypothesis that the flow network $\mathcal{N}_{\mathbf{P}}$ can bear a flow of value p.

  Lemma. *If* $\mathcal{N}_{\mathbf{P}}$ *admits a flow of value* p*, then the* $s_{k,I}$*'s exist for which conditions 1 and 2 hold.*

  Proof. We start by splitting the flow of value p, which by hypothesis is admitted by $\mathcal{N}_{\mathbf{P}}$, into n flows of value p_i going from the source vertex s to singleton vertices $\{i\}$, for every $i \in I$. Afterwards, for every other vertex $I \subseteq \{1, \ldots, n\}$, the values of flows on the incoming edges are added up and then distributed to the outgoing adges as one wishes, thanks to conservation principle. In general, this can be formalised by turning a flow $f : \mathcal{E}_{\mathbf{P}} \to \mathbb{R}_{[0,1]}$ into a function $\overline{f} : \mathcal{E}_{\mathbf{P}} \to (\mathbb{R}_{[0,1]})^n$ defined as follows:

  - For every $i \in \{1, \ldots, n\}$, $\overline{f}(s, \{i\}) = (0, \ldots, f(s, \{i\}), \ldots, 0)$, where the only possibly nonnull component is exactly the i-th;
  - For every non-empty $I \subseteq \{1, \ldots, n\}$, as soon as $\overline{f}$ has been defined on all the ingoing edges of I, we can define it on all its outgoing ones, by just splitting each component as we want. Of course, this is possible because f is a flow and, as such, ingoing and outgoing values are the same. More formally, let us fix $\overline{f}(*, I) = \sum_{K \in \mathcal{E}^{-}_{\mathbf{P}}(I)} \overline{f}(K, I)$ and indicate with $\overline{f}(*, I)_k$ its k-th component. Then, for every $i \notin I$, we set

    $$\overline{f}(I, I \cup \{i\}) = (q_{1,i} \cdot \overline{f}(*, I)_1, \ldots, q_{n,i} \cdot \overline{f}(*, I)_n),$$

$$\overline{f}(I,t) = (q_{1,t} \cdot \overline{f}(*,I)_1, \ldots, q_{n,t} \cdot \overline{f}(*,I)_n)$$

where, for every $i,j \in \{1,\ldots,n\}$, $q_{j,i}, q_{j,t} \in \mathbb{R}_{[0,1]}$ and such that the conservation principle with respect to components holds

$$\sum_{i \notin I} q_{j,i} \cdot \overline{f}(*,I)_j + q_{j,t} \cdot \overline{f}(*,I)_j = \overline{f}(*,I)_j,$$

along with the soundness of construction

$$\sum_{j=1}^{n} q_{j,i} \cdot \overline{f}(*,I)_j = f(I, I \cup \{i\}),$$
$$\sum_{j=1}^{n} q_{j,t} \cdot \overline{f}(*,I)_j = f(I,t).$$

Notice that, the way we have just defined $\overline{f}$ guarantees that the sum of all components of $\overline{f}(v,w)$ is always equal to $f(v,w)$, for every $v,w \in \mathcal{V}_{\mathbf{P}}$. Now, for every non-empty $I \subseteq \{1,\ldots,n\}$, let $s_{k,I}$ be the k-th component of $\overline{f}(I,t)$ (or 0, if the first is itself 0). On the one hand, $\sum_{k \in I} s_{k,I}$ is obviously less or equal to 1, hence condition 1 holds. On the other hand, each component of $\overline{f}$ is itself a flow as it satisfies the capacity and conservation constraints. Moreover, $\mathcal{N}_{\mathbf{P}}$ is structured in such a way that the k-th component of $\overline{f}(I,t)$ is 0 whenever $k \notin I$. As a consequence, since $\overline{f}$ is sound with respect to f and this latter satisfies the capacity constraint, for every $I \subseteq \{1,\ldots,n\}$ and every $k \in \{1,\ldots,n\}$,

$$p_k \leq \sum_{k \in I} s_{k,I} \cdot \overline{f}(I,t) \leq \sum_{k \in I} s_{k,I} \cdot r_I$$

and so condition 2 holds as well. ■

- The second lemma proves that $\mathcal{N}_{\mathbf{P}}$ admits, indeed, a flow of value p.

Lemma. *$\mathcal{N}_{\mathbf{P}}$ admits a flow of value p.*

Proof. We prove the result by means of Theorem 4.4.2. In particular, we just prove that the capacity of any cut must be at least p.

Definition. *A cut (S,A) is said to be* degenerate *if there are $I \subseteq \{1,\ldots,n\}$ and $i \in \{1,\ldots,n\}$ such that $I \in S$ and $I \cup \{i\} \in A$.*

It is easy to verify that every degenerate cut has capacity at least 1, thus greater or equal to p. As a consequence, we can just concentrate on non-degenerate cuts and prove that all of them have capacity at least p.

Given two cuts $C = (S,A)$ and $D = (T,B)$, we write $C \leq D$ iff $S \subseteq T$.

Definition. *Given $I \subseteq \{1, \ldots, n\}$, an I-*cut *is any cut (S, A) such that $\bigcup_{\{i\} \in S} \{i\} = I$. The* canonical *$I$-cut is the unique I-cut $C_I = (S, A)$ such that $S = \{s\} \cup \{J \subseteq \{1, \ldots, n\} \mid J \cap I \neq \emptyset\}$.*

Observe that, by definition, C_I is non-degenerate and that the capacity $c(C_I)$ of C_I is at least p, because the forward edges in C_I (those connecting elements of S to those of A) are those going from s to the singletons not in S, plus the edges going from any $J \in S$ to t. The sum of the capacities of such edges are greater or equal to p by construction.

We now show the following two lemmas to complete this proof.

Lemma. *For every non-degenerate I-cuts C, D such that $C > D$, there is a non-degenerate I-cut E such that $C \geq E > D$ and $c(E) \geq c(D)$.*

Proof. Let $C = (S, A)$, $D = (T, B)$ and let J be any element of S not in T. Consider the cut $E = (T \cup \{K \subseteq \{1, \ldots, n\} \mid J \subseteq K\}, B \backslash \{K \subseteq \{1, \ldots, n\} \mid J \subseteq K\})$ and verify that E is the cut we are looking for. Indeed, E is non-degenerate because it is obtained from D, which is non-degenerate by hypothesis, by adding to it J and all its supersets. Of course, $E > D$. Moreover, $C \geq E$ holds since $J \in S$ and C is non-degenerate, which implies C contains all supersets of J as well. It is also easy to check that $c(E) \geq c(D)$. In fact, in the process of constructing E from D we do not lose any forward edges coming from s, since J cannot be a singleton with C and D both I-cuts, or any other edge coming from some element of T, since D is non-degenerate. ■

Lemma. *For every non-degenerate I-cuts C, D such that $C \geq D$, $c(C) \geq c(D)$.*

Proof. Let $C = (S, A)$ and $D = (T, B)$. We prove the result by induction on the $n = |S| - |T|$. If $n = 0$, then $C = D$ and the thesis follows. If $n > 0$, then $C > D$ and, by the above lemma, there is a non-degenerate I-cut E such that $C \geq E > D$ and $c(E) \geq c(D)$. By induction hypothesis on $C \geq E$, it follows that $c(C) \geq c(E)$. Thus, $c(C) \geq c(D)$. ■

The two lemmas above allow to conclude. Indeed, for every non-degenerate cut D, there is of course a I such that D is a I-cut (possibly with I as the empty set). Now consider the canonical C_I: on the one hand $c(C_I) \geq p$; on the other hand, since C_I is non-degenerate, $c(D) \geq c(C_I)$. Hence, $c(D) \geq p$. ■

This concludes the main proof. □

In the coming proof of Key Lemma 4.3.25 we implicitly appeal to the following technical lemmas.

Lemma 4.4.4. *For every* $X \subseteq \Lambda_{\oplus}(x)$, $\precsim (\nu x.X) = \nu x.{\precsim}(X)$.

Proof. Refer to (4.7). The first and last double implications are given by definitions of $\precsim (\nu x.X)$ and $\nu x.\precsim (X)$. The second double implication is given by Lemma 4.3.7 for $\precsim$, using the fact that $\precsim$ is closed under term-substitution (Definition 4.3.15) and the definition of its open extension (Definition 4.3.9).

$$\begin{aligned} \nu x.M \in \precsim (\nu x.X) &\Leftrightarrow \exists N \in X.\, \nu x.N \precsim \nu x.M \\ &\Leftrightarrow \exists N \in X.\, N \precsim M \\ &\Leftrightarrow \nu x.M \in \nu x.\precsim (X). \end{aligned} \tag{4.7}$$

This concludes the proof. □

Remark **4.4.5.** The property established by Lemma 4.4.4 is precisely the reason why we have formulated $\Lambda_{\oplus}$ as a multisorted labelled Markov chain: ${\precsim}(\nu x.X)$ consists of distinguished values only, and is nothing but $\nu x.{\precsim}(X)$.

Lemma 4.4.6. *If* $M \precsim N$*, then* $[\![M]\!]\,(\nu x.X) \leq [\![N]\!]\,(\nu x.\precsim (X))$ *for every* $X \in \Lambda_{\oplus}(x)$.

Proof. If $M \precsim N$, then by definition $[\![M]\!]\,(\nu x.X) \leq [\![N]\!]\,(\precsim (\nu x.X))$. Lemma 4.4.4 entails $[\![N]\!]\,(\precsim (\nu x.X)) = [\![N]\!]\,(\nu x.{\precsim}(X))$, hence $[\![M]\!]\,(\nu x.X) \leq [\![N]\!]\,(\nu x.\precsim (X))$. □

We now proceed to prove the Key Lemma, after having recalled its statement.

Lemma. *If* $M \precsim^H N$*, then* $[\![M]\!]\,(\lambda x.X) \leq [\![N]\!]\,(\lambda x.(\precsim^H(X)))$ *for every* $X \subseteq \Lambda_{\oplus}(x)$.

Proof. This is equivalent to proving that if $M \precsim^H N$, then for every $X \subseteq \Lambda_{\oplus}(x)$ the following implication holds: if $M \Downarrow \mathscr{D}$, then $\mathscr{D}(\lambda x.X) \leq [\![N]\!]\,(\lambda x.(\precsim^H(X)))$. This is an induction on the structure of the derivation of $M \Downarrow \mathscr{D}$.

- If $\mathscr{D} = \emptyset$, then of course $\mathscr{D}(\lambda x.X) = 0 \leq [\![N]\!]\,(\lambda x.Y)$ for every $X, Y \subseteq \Lambda_{\oplus}(x)$.
- If M is a value $\lambda x.L$ and $\mathscr{D}(\lambda x.L) = 1$, then the proof of $M \precsim^H N$ necessarily ends as follows:

$$\frac{\{x\} \vdash L \precsim^H P \qquad \emptyset \vdash \lambda x.P \precsim N}{\emptyset \vdash \lambda x.L \precsim^H N}$$

  Let X be any subset of $\Lambda_{\oplus}(x)$. Now, if $L \notin X$, then $\mathscr{D}(\lambda x.X) = 0$ and the inequality trivially holds. If, on the contrary, $L \in X$, then $P \in \precsim^H(X)$. Consider $\precsim (P)$, the set of terms that are in relation with P via $\precsim$. We have that for every $Q \in \precsim (P)$, both $\{x\} \vdash L \precsim^H P$ and $\{x\} \vdash P \precsim Q$ hold, hence

$\{x\} \vdash P \lesssim^H Q$, and as a consequence $\{x\} \vdash L \lesssim^H Q$ does. In other words, $\lesssim (P) \subseteq \lesssim^H(X)$. Therefore, by Lemma 4.4.6 and Lemma 4.3.7 for $\lesssim$,

$$[\![N]\!]\,(\lambda x.\lesssim^H(X)) \geq [\![N]\!]\,(\lambda x.\lesssim (P)) \geq [\![\lambda x.P]\!]\,(\lambda x.P) = 1.$$

- If M is an application LP, then $M \Downarrow \mathscr{D}$ is obtained as follows:

$$\frac{L \Downarrow \mathscr{F} \qquad \{Q\,[P/x] \Downarrow \mathscr{H}_{Q,P}\}_{Q,P}}{LP \Downarrow \sum_Q \mathscr{F}(\lambda x.Q) \cdot \mathscr{H}_{Q,P}}$$

  Moreover, the proof of $\emptyset \vdash M \lesssim^H N$ must end as follows:

$$\frac{\emptyset \vdash L \lesssim^H R \qquad \emptyset \vdash P \lesssim^H S \qquad \emptyset \vdash (R)\,S \;\lesssim\; N}{\emptyset \vdash (L)\,P \lesssim^H N}$$

  Now, since $L \Downarrow \mathscr{F}$ and $\emptyset \vdash L \lesssim^H R$, by the induction hypothesis follows that, for every $Y \subseteq \Lambda_\oplus(x)$, $\mathscr{F}(\lambda x.Y) \leq [\![R]\!]\,(\lambda x.\lesssim^H(Y))$. Let us now take a look at the distribution

$$\mathscr{D} = \sum_Q \mathscr{F}(\lambda x.Q) \cdot \mathscr{H}_{Q,P}.$$

  Since $\mathscr{F}$ is a *finite* distribution, the sum above is actually the sum of finitely many summands. Let the support $\mathsf{Supp}(\mathscr{F})$ of $\mathscr{F}$ be $\{\lambda x.Q_1, \ldots, \lambda x.Q_n\}$. It is now time to put the above into a form that is amenable to treatment by Lemma 4.4.3. Let us consider the n sets $\lesssim^H(Q_1), \ldots, \lesssim^H(Q_n)$; to each term U in them we can associate the probability $[\![R]\!]\,(\lambda x.U)$. We are then in the scope of Lemma 4.4.3, since by the induction hypothesis, we know that for every $Y \subseteq \Lambda_\oplus(x)$, $\mathscr{F}(\lambda x.Y) \leq [\![R]\!]\,(\lambda x.\lesssim^H(Y))$. We can then conclude that for every

$$U \in \lesssim^H(\{Q_1, \ldots, Q_n\}) = \bigcup_{1 \leq i \leq n} \lesssim^H(Q_i)$$

  there are n real numbers $r_1^{U,R}, \ldots, r_n^{U,R}$ such that

$$\begin{aligned} [\![R]\!]\,(\lambda x.U) &\geq \sum_{1 \leq i \leq n} r_i^{U,R} \qquad \forall\, U \in \bigcup_{1 \leq i \leq n} \lesssim^H(Q_i); \\ \mathscr{F}(\lambda x.Q_i) &\leq \sum_{U \in \lesssim^H(Q_i)} r_i^{U,R} \qquad \forall\, 1 \leq i \leq n. \end{aligned}$$

  So, we can conclude that

$$\mathscr{D} \leq \sum_{1 \leq i \leq n} \left(\sum_{U \in \lesssim^H(Q_i)} r_i^{U,R} \right) \cdot \mathscr{H}_{Q_i,P}$$

$$= \sum_{1\leq i\leq n} \sum_{U\in\lesssim^H(Q_i)} r_i^{U,R} \cdot \mathscr{H}_{Q_i,P}.$$

Now, whenever $Q_i \lesssim^H U$ and $P \lesssim^H S$, by the substitutive property of $\lesssim^H$ (Lemma 4.3.20) follows $Q_i\,[P/x] \lesssim^H U\,[S/x]$. We can then apply the induction hypothesis to the n derivations of $Q_i\,[P/x] \Downarrow \mathscr{H}_{Q_i,P}$, obtaining that, for every $X \subseteq \Lambda_\oplus(x)$,

$$\begin{aligned}
\mathscr{D}(\lambda x.X) &\leq \sum_{1\leq i\leq n} \sum_{U\in\lesssim^H(Q_i)} r_i^{U,R} \cdot [\![U\,[S/x]]\!]\,(\lambda x.\lesssim^H(X)) \\
&\leq \sum_{1\leq i\leq n} \sum_{U\in\lesssim^H(\{Q_1,\ldots,Q_n\})} r_i^{U,R} \cdot [\![U\,[S/x]]\!]\,(\lambda x.\lesssim^H(X)) \\
&= \sum_{U\in\lesssim^H(\{Q_1,\ldots,Q_n\})} \sum_{1\leq i\leq n} r_i^{U,R} \cdot [\![U\,[S/x]]\!]\,(\lambda x.\lesssim^H(X)) \\
&= \sum_{U\in\lesssim^H(\{Q_1,\ldots,Q_n\})} \left(\sum_{1\leq i\leq n} r_i^{U,R}\right) \cdot [\![U\,[S/x]]\!]\,(\lambda x.\lesssim^H(X)) \\
&\leq \sum_{U\in\lesssim^H(\{Q_1,\ldots,Q_n\})} [\![R]\!]\,(\lambda x.U) \cdot [\![U\,[S/x]]\!]\,(\lambda x.\lesssim^H(X)) \\
&\leq \sum_{U\in\Lambda_\oplus(x)} [\![R]\!]\,(\lambda x.U) \cdot [\![U\,[S/x]]\!]\,(\lambda x.\lesssim^H(X)) \\
&= [\![RS]\!]\,(\lambda x.\lesssim^H(X)) \leq [\![N]\!]\,(\lambda x.\lesssim((\lesssim^H)(X))) \\
&\leq [\![N]\!]\,(\lambda x.\lesssim^H(X)).
\end{aligned}$$

- If M is a probabilistic sum $L \oplus P$, then $M \Downarrow \mathscr{D}$ is obtained as follows:

$$\frac{L \Downarrow \mathscr{F} \qquad P \Downarrow \mathscr{G}}{L \oplus P \Downarrow \frac{1}{2}\mathscr{F} + \frac{1}{2}\mathscr{G}}$$

  Moreover, the proof of $\emptyset \vdash M \lesssim^H N$ must end as follows:

$$\frac{\emptyset \vdash L \lesssim^H R \qquad \emptyset \vdash P \lesssim^H S \qquad \emptyset \vdash R \oplus S \lesssim N}{\emptyset \vdash L \oplus P \lesssim^H N}$$

  Now:

  - Since $L \Downarrow \mathscr{F}$ and $\emptyset \vdash L \lesssim^H R$, by the induction hypothesis follows that, for every $Y \subseteq \Lambda_\oplus(x)$, $\mathscr{F}(\lambda x.Y) \leq [\![R]\!]\,(\lambda x.\lesssim^H(Y))$;
  - Similarly, since $P \Downarrow \mathscr{G}$ and $\emptyset \vdash P \lesssim^H S$, by the induction hypothesis follows that, for every $Y \subseteq \Lambda_\oplus(x)$, $\mathscr{G}(\lambda x.Y) \leq [\![S]\!]\,(\lambda x.\lesssim^H(Y))$.

Let us now take a look at the distribution

$$\mathscr{D} = \frac{1}{2}\mathscr{F} + \frac{1}{2}\mathscr{G}.$$

It suffices to prove that, for every $X \subseteq \Lambda_{\oplus}(x)$, $\mathscr{D}(\lambda x.X) \leq [\![R \oplus S]\!]\,(\lambda x.{\lesssim^H}(X))$, since $[\![R \oplus S]\!]\,(\lambda x.{\lesssim^H}(X)) \leq [\![N]\!]\,(\lambda x.{\lesssim^H}(X))$, thus implying the thesis. The induction hypothesis, along with Lemma 4.1.6, entails

$$\begin{aligned}
\mathscr{D}(\lambda x.X) &= \frac{1}{2}\mathscr{F}(\lambda x.X) + \frac{1}{2}\mathscr{G}(\lambda x.X) \\
&\leq \frac{1}{2}\,[\![R]\!]\,(\lambda x.{\lesssim^H}(X)) + \frac{1}{2}\,[\![S]\!]\,(\lambda x.{\lesssim^H}(X)) \\
&= [\![R \oplus S]\!]\,(\lambda x.{\lesssim^H}(X)).
\end{aligned}$$

This concludes the proof. □

4.5 Relating Applicative bisimilarity and Context equivalence

In this section we follow Pitts [Pit11] in the process of proving that probabilistic context equivalence is itself a congruence. Hence, the congruence of applicative bisimilarity yields the inclusion in context equivalence.

The converse inclusion fails, as the well-known non-deterministic counterexample works here as well. In order to do so, we appeal to *CIU-equivalence*, a relation that can be shown to coincide with context equivalence by a *context lemma*, itself proved by Howe's technique.

4.5.1 Probabilistic context equivalence

We now formally introduce probabilistic context equivalence and prove it to be coarser than probabilistic applicative bisimilarity.

Definition 4.5.1. $\Lambda_{\oplus}$-term contexts $C\Lambda_{\oplus}$ *are syntax trees with a unique "hole"* $\langle\cdot\rangle$ *given by the following grammar (where* $M \in \Lambda_{\oplus}$*):*

$$C, D ::= \langle\cdot\rangle \quad | \quad \lambda x.C \quad | \quad (C)\,M \quad | \quad (M)\,C \quad | \quad C \oplus M \quad | \quad M \oplus C.$$

$C\,\langle N\rangle$ *denotes the* $\Lambda_{\oplus}$*-term that results from filling the hole with a* $\Lambda_{\oplus}$*-term* N*:*

$$\begin{aligned}
\langle\cdot\rangle\,\langle N\rangle &= N; \\
(\lambda x.C)\,\langle N\rangle &= \lambda x.C\,\langle N\rangle;
\end{aligned}$$

$$\frac{}{\langle\cdot\rangle \in \mathsf{C}\Lambda_{\oplus}(\overline{x}\,;\,\overline{x})}\ (\mathsf{Ctx1})$$

$$\frac{C \in \mathsf{C}\Lambda_{\oplus}(\overline{x}\,;\,\overline{y} \cup \{x\}) \qquad x \notin \overline{y}}{\lambda x.C \in \mathsf{C}\Lambda_{\oplus}(\overline{x}\,;\,\overline{y})}\ (\mathsf{Ctx2})$$

$$\frac{C \in \mathsf{C}\Lambda_{\oplus}(\overline{x}\,;\,\overline{y}) \qquad M \in \Lambda_{\oplus}(\overline{y})}{(C)\, M \in \mathsf{C}\Lambda_{\oplus}(\overline{x}\,;\,\overline{y})}\ (\mathsf{Ctx3})$$

$$\frac{M \in \Lambda_{\oplus}(\overline{y}) \qquad C \in \mathsf{C}\Lambda_{\oplus}(\overline{x}\,;\,\overline{y})}{(M)\, C \in \mathsf{C}\Lambda_{\oplus}(\overline{x}\,;\,\overline{y})}\ (\mathsf{Ctx4})$$

$$\frac{C \in \mathsf{C}\Lambda_{\oplus}(\overline{x}\,;\,\overline{y}) \qquad M \in \Lambda_{\oplus}(\overline{y})}{C \oplus M \in \mathsf{C}\Lambda_{\oplus}(\overline{x}\,;\,\overline{y})}\ (\mathsf{Ctx5})$$

$$\frac{M \in \Lambda_{\oplus}(\overline{y}) \qquad C \in \mathsf{C}\Lambda_{\oplus}(\overline{x}\,;\,\overline{y})}{M \oplus C \in \mathsf{C}\Lambda_{\oplus}(\overline{x}\,;\,\overline{y})}\ (\mathsf{Ctx6})$$

Figure 4.7: Rules for $\mathsf{C}\Lambda_{\oplus}(\overline{x}\,;\,\overline{y})$.

$$\begin{aligned}((C)\, M)\, \langle N\rangle &= (C\, \langle N\rangle)\, M;\\ ((M)\, C)\, \langle N\rangle &= (M)\, C\, \langle N\rangle;\\ (C \oplus M)\, \langle N\rangle &= C\, \langle N\rangle \oplus M;\\ (M \oplus C)\, \langle N\rangle &= M \oplus C\, \langle N\rangle.\end{aligned}$$

Notation. We write $C\,\langle D\rangle$ for the context resulting from replacing the occurrence of $\langle\cdot\rangle$ in the syntax tree C by the tree D.

We continue to keep track of free variables by sets $\overline{x}$ of variables and we inductively define sets $\mathsf{C}\Lambda_{\oplus}(\overline{x}\,;\,\overline{y})$ of contexts by the rules in Figure 4.7. We use double indexing over $\overline{x}$ and $\overline{y}$ to indicate the sets of free variables before and after the filling of the hole by a term. The following two properties explain this idea.

Lemma 4.5.2. *If $M \in \Lambda_{\oplus}(\overline{x})$ and $C \in \mathsf{C}\Lambda_{\oplus}(\overline{x}\,;\,\overline{y})$, then $C\,\langle M\rangle \in \Lambda_{\oplus}(\overline{y})$.*

Proof. Simple induction on the derivation of $C \in \mathsf{C}\Lambda_{\oplus}(\overline{x}\,;\,\overline{y})$. $\square$

Lemma 4.5.3. *If $C \in \mathsf{C}\Lambda_{\oplus}(\overline{x}\,;\,\overline{y})$ and $D \in \mathsf{C}\Lambda_{\oplus}(\overline{y}\,;\,\overline{z})$, then $D\,\langle C\rangle \in \mathsf{C}\Lambda_{\oplus}(\overline{x}\,;\,\overline{z})$.*

Proof. Simple induction on the derivation of $D \in \mathsf{C}\Lambda_\oplus(\overline{y}\,;\overline{z})$. $\square$

We now define the notion of context equivalence for this probabilistic setting. Differently from a qualitative scenario, where terms are considered context equivalent if they both converge or diverge, here terms with different convergence probabilities are considered different in an essential way. This agrees with the intuition that taking into account quantitative information allows to be more precise.

Definition 4.5.4. *A* closing context *is a* $\Lambda_\oplus$*-term context* C *such that* $C \in \mathsf{C}\Lambda_\oplus(\overline{x}\,;\emptyset)$.

Definition 4.5.5. *The expression* $M{\Downarrow_p}$ *stands for* $\sum[\![M]\!] = p$, i.e. *the term* M *converges with probability* p. *The* probabilistic context preorder $\leq_\oplus$ *stipulates* $\overline{x} \vdash M \leq_\oplus N$ *if* $C\langle M\rangle \Downarrow_p$ *implies* $C\langle N\rangle \Downarrow_q$ *with* $p \leq q$, *for every closing context* $C \in \mathsf{C}\Lambda_\oplus(\overline{x}\,;\emptyset)$. Probabilistic context equivalence, *denoted as* $\simeq_\oplus$, *is the equivalence induced by* $\leq_\oplus$.

***Remark* 4.5.6.** Observe that context preorders and equivalences are defined on open terms, whereas (bi)similarities are defined on closed terms. This is not a problem as long as we consider their extensions defined on open terms by requiring the usual closure under term-substitutions.

Lemma 4.5.7. *Probabilistic context preorder* $\leq_\oplus$ *is a precongruence.*

Proof. We prove $\leq_\oplus$ to be transitive and compatible. We start with the former, namely that for every $\overline{x} \in \mathcal{P}_{\mathsf{FIN}}(\mathcal{V})$ and for every $M, N, L \in \Lambda_\oplus(\overline{x})$, $\overline{x} \vdash M \leq_\oplus N$ and $\overline{x} \vdash N \leq_\oplus L$ imply $\overline{x} \vdash M \leq_\oplus L$. This directly follows by Definition 4.5.5 and the transitive property of the usual ordering of $\mathbb{R}_{[0,1]}$.

We prove $\leq_\oplus$ to be a compatible relation starting from the $(\mathsf{Com2})$ property because $(\mathsf{Com1})$ is trivially valid. In particular, we must show that, for every $\overline{x} \in \mathcal{P}_{\mathsf{FIN}}(\mathcal{V})$, $x \in \mathcal{V} \setminus \overline{x}$, and $M, N \in \Lambda_\oplus(\overline{x} \cup \{x\})$, if $\overline{x} \cup \{x\} \vdash M \leq_\oplus N$ then $\overline{x} \vdash \lambda x.M \leq_\oplus \lambda x.N$. By Definition 4.5.5, the latter boils down to prove that the hypothesis

- for every closing context C, $C\langle M\rangle \Downarrow_p$ implies $C\langle N\rangle \Downarrow_q$, with $p \leq q$;
- $D\langle \lambda x.M\rangle \Downarrow_r$

imply $D\langle \lambda x.N\rangle \Downarrow_s$, with $r \leq s$. Since $D \in \mathsf{C}\Lambda_\oplus(\overline{x}\,;\emptyset)$, consider the context $\lambda x.\langle\cdot\rangle \in \mathsf{C}\Lambda_\oplus(\overline{x} \cup \{x\}\,;\overline{x})$. Lemma 4.5.3 implies that the context $E = D\langle \lambda x.\langle\cdot\rangle\rangle$ is in $\mathsf{C}\Lambda_\oplus(\overline{x} \cup \{x\}\,;\emptyset)$. Moreover, observe that $D\langle \lambda x.M\rangle = E\langle M\rangle$ and, therefore, the second hypothesis can be rewritten as $E\langle M\rangle \Downarrow_r$. The first hypothesis on this latter entails $E\langle N\rangle \Downarrow_s$, namely $D\langle \lambda x.N\rangle \Downarrow_s$, with $r \leq s$. Since $\leq_\oplus$ is transitive, we prove the $(\mathsf{Com3})$ property by showing that $(\mathsf{Com3L})$ and $(\mathsf{Com3R})$ hold (recall that, by Lemma 4.3.12, the latter two imply the former). In the case of $(\mathsf{Com3L})$ we must show that, for every $\overline{x} \in \mathcal{P}_{\mathsf{FIN}}(\mathcal{V})$ and $M, N, L \in \Lambda_{\oplus}(\overline{x})$, if $\overline{x} \vdash M \leq_{\oplus} N$ then $\overline{x} \vdash (M)\, L \leq_{\oplus} (N)\, L$. By Definition 4.5.5, the latter boils down to prove that the hypothesis

- for every closing context C, $C\langle M \rangle \Downarrow_p$ implies $C\langle N \rangle \Downarrow_q$, with $p \leq q$;
- $D\langle (M)\, L \rangle \Downarrow_r$

imply $D\langle (N)\, L \rangle \Downarrow_s$, with $r \leq s$. Since $D \in \mathsf{C}\Lambda_{\oplus}(\overline{x}\,;\emptyset)$, consider the context $(\langle\cdot\rangle)\, L \in \mathsf{C}\Lambda_{\oplus}(\overline{x}\,;\overline{x})$. Lemma 4.5.3 implies that the context $E = D\langle (\langle\cdot\rangle)\, L \rangle$ is in $\mathsf{C}\Lambda_{\oplus}(\overline{x}\,;\emptyset)$. Moreover, observe that $D\langle (M)\, L \rangle = E\langle M \rangle$ and, therefore, the second hypothesis can be rewritten as $E\langle M \rangle \Downarrow_r$. The first hypothesis on this latter entails $E\langle N \rangle \Downarrow_s$, namely $D\langle (N)\, L \rangle \Downarrow_s$, with $r \leq s$. The proof of $(\mathsf{Com3R})$ follows the same reasoning we have just detailed for $(\mathsf{Com3L})$, considering E as the context $D\langle (L)\, \langle\cdot\rangle \rangle$. We do not detail the proof of $(\mathsf{Com4})$ either, as it follows the guidelines of that for $(\mathsf{Com3})$. ☐

Corollary 4.5.8. *Probabilistic context equivalence $\simeq_{\oplus}$ is a congruence.*

Proof. Straightforward consequence of Lemma 4.5.7 as $\simeq_{\oplus} = \leq_{\oplus} \cap \leq_{\oplus}{}^{op}$. ☐

Lemma 4.5.9. *Let $\mathcal{R}$ be a compatible $\Lambda_{\oplus}$-relation. If $\overline{x} \vdash M \mathrel{\mathcal{R}} N$ and $C \in \mathsf{C}\Lambda_{\oplus}(\overline{x}\,;\overline{y})$, then $\overline{y} \vdash C\langle M \rangle \mathrel{\mathcal{R}} C\langle N \rangle$.*

Proof. By induction on the derivation of $C \in \mathsf{C}\Lambda_{\oplus}(\overline{x}\,;\overline{y})$:

- If C is due to $(\mathsf{Ctx1})$, then $C = \langle\cdot\rangle$. Thus, $C\langle M \rangle = M$ and $C\langle N \rangle = N$. The thesis follows from the hypothesis.
- If $(\mathsf{Ctx2})$ is the last rule used, then $C = \lambda x.D$, with $D \in \mathsf{C}\Lambda_{\oplus}(\overline{x}\,;\overline{y} \cup \{x\})$. By the induction hypothesis follows $\overline{y} \cup \{x\} \vdash D\langle M \rangle \mathrel{\mathcal{R}} D\langle N \rangle$. Since $\mathcal{R}$ is a compatible relation, it follows $\overline{y} \vdash \lambda x.D\langle M \rangle \mathrel{\mathcal{R}} \lambda x.D\langle N \rangle$, hence the result $\overline{y} \vdash C\langle M \rangle \mathrel{\mathcal{R}} C\langle N \rangle$ holds.
- If $(\mathsf{Ctx3})$ is the last rule used, then $C = (D)\, L$, with $D \in \mathsf{C}\Lambda_{\oplus}(\overline{x}\,;\overline{y})$ and $L \in \Lambda_{\oplus}(\overline{y})$. By the induction hypothesis, it holds that $\overline{y} \vdash D\langle M \rangle \mathrel{\mathcal{R}} D\langle N \rangle$. Since $\mathcal{R}$ is a compatible relation, it follows $\overline{y} \vdash (D\langle M \rangle)\, L \mathrel{\mathcal{R}} (D\langle N \rangle)\, L$, which by Definition 4.5.1 means $\overline{y} \vdash ((D)\, L)\langle M \rangle \mathrel{\mathcal{R}} ((D)\, L)\langle N \rangle$. Hence, the result $\overline{y} \vdash C\langle M \rangle \mathrel{\mathcal{R}} C\langle N \rangle$ holds. The case of rule $(\mathsf{Ctx4})$ holds by a similar reasoning.
- If $(\mathsf{Ctx5})$ is the last rule used, then $C = D \oplus L$, with $D \in \mathsf{C}\Lambda_{\oplus}(\overline{x}\,;\overline{y})$ and $L \in \Lambda_{\oplus}(\overline{y})$. By the induction hypothesis, it holds that $\overline{y} \vdash D\langle M \rangle \mathrel{\mathcal{R}} D\langle N \rangle$. Since $\mathcal{R}$ is a compatible relation, it follows $\overline{y} \vdash D\langle M \rangle \oplus L \mathrel{\mathcal{R}} D\langle N \rangle \oplus L$,

which by Definition 4.5.1 means $\overline{y} \vdash (D \oplus L)\langle M\rangle \ \mathcal{R}\ (D \oplus L)\langle N\rangle$. Hence, the result $\overline{y} \vdash C\langle M\rangle \ \mathcal{R}\ C\langle N\rangle$ holds. The case of rule (Ctx6) holds by a similar reasoning.

This concludes the proof. $\square$

Lemma 4.5.10. *If $\overline{x} \vdash M \ \sim\ N$ and $C \in \mathsf{C}\Lambda_{\oplus}(\overline{x}\,;\overline{y})$, then $\overline{y} \vdash C\langle M\rangle \ \sim\ C\langle N\rangle$.*

Proof. Since $\sim = \lesssim \cap \lesssim^{op}$ (Theorem 4.2.14), $\overline{x} \vdash M \ \sim\ N$ implies $\overline{x} \vdash M \ \lesssim\ N$ and $\overline{x} \vdash N \ \lesssim\ M$ (recall Definition 4.3.9 of open extension for $\lesssim$ and $\sim$). From the fact that $\lesssim$ is a precongruence (Theorem 4.3.26) follows that $\lesssim$ is a compatible relation, hence both $\overline{y} \vdash C\langle M\rangle \ \lesssim\ C\langle N\rangle$ and $\overline{y} \vdash C\langle N\rangle \ \lesssim\ C\langle M\rangle$ follow by Lemma 4.5.9. The latter two imply $\overline{y} \vdash C\langle M\rangle \ \sim\ C\langle N\rangle$. $\square$

Theorem 4.5.11. *For all $M, N \in \Lambda_{\oplus}$, $M \sim N$ implies $M \simeq_{\oplus} N$.*

Proof. If $M \ \sim\ N$, then Lemma 4.5.10 entails $C\langle M\rangle \ \sim\ C\langle N\rangle$, for every closing context C. Hence, Lemma 4.3.5 implies $\sum [\![C\langle M\rangle]\!] = p = \sum [\![C\langle N\rangle]\!]$. This means, in particular, that $C\langle M\rangle \Downarrow_p$ if and only if $C\langle N\rangle \Downarrow_p$, which is equivalent to $M \simeq_{\oplus} N$ (Definition 4.5.5). $\square$

The converse inclusion fails.

***Counterexample* 4.5.12.** For $M = \lambda x.\lambda y.(\mathbf{\Omega} \oplus \mathbf{I})$ and $N = \lambda x.((\lambda y.\mathbf{\Omega}) \oplus (\lambda y.\mathbf{I}))$, we have $N \not\lesssim M$, hence $M \not\sim N$, but $M \simeq_{\oplus} N$.

Proof. We only prove that $N \not\lesssim M$, whereas $M \simeq_{\oplus} N$ is shown by means of CIU-equivalence at the end of Section 4.5.3.

For the sake of contradiction, suppose $N \lesssim M$. Due to the definition of the underlying labelled Markov chain (Definition 4.3.1), this is equivalent to assume

$$(\lambda y.\mathbf{\Omega}) \oplus (\lambda y.\mathbf{I}) \lesssim \lambda y.(\mathbf{\Omega} \oplus \mathbf{I}).$$

If the latter holds, then it follows (by taking $X = \{\nu y.\mathbf{I}\}$ in Definition 4.2.8)

$$\frac{1}{2} = \mathcal{P}_{\oplus}((\lambda y.\mathbf{\Omega}) \oplus (\lambda y.\mathbf{I}), \tau, \nu y.\mathbf{I}) \leq \mathcal{P}_{\oplus}(\lambda y.(\mathbf{\Omega} \oplus \mathbf{I}), \tau, \lesssim (\nu y.\mathbf{I})). \tag{4.8}$$

Definition 4.3.1 of $\mathcal{P}_{\oplus}$ on the term $\lambda y.(\mathbf{\Omega} \oplus \mathbf{I})$ entails

$$\mathcal{P}_{\oplus}(\lambda y.(\mathbf{\Omega} \oplus \mathbf{I}), \tau, V) = [\![\lambda y.(\mathbf{\Omega} \oplus \mathbf{I})]\!]\,(V) = \begin{cases} 1 & \text{if } V \text{ is } \nu y.(\mathbf{\Omega} \oplus \mathbf{I}); \\ 0 & \text{otherwise.} \end{cases}$$

Since (4.8) implies $\mathcal{P}_{\oplus}(\lambda y.(\mathbf{\Omega} \oplus \mathbf{I}), \tau, \lesssim (\nu y.\mathbf{I})) \geq 0$, it follows that

$$\nu y.(\mathbf{\Omega} \oplus \mathbf{I}) \in \lesssim (\nu y.\mathbf{I}),$$

hence $\nu y.\mathbf{I} \lesssim \nu y.(\mathbf{\Omega} \oplus \mathbf{I})$, that is $\mathbf{I} \lesssim \mathbf{\Omega} \oplus \mathbf{I}$. From this latter we deduce the absurd as

$$1 = \mathcal{P}_{\oplus}(\mathbf{I}, \tau, \mathbf{I}) \leq \mathcal{P}_{\oplus}(\mathbf{\Omega} \oplus \mathbf{I}, \tau, \lesssim (\mathbf{I})) \leq \mathcal{P}_{\oplus}(\mathbf{\Omega} \oplus \mathbf{I}, \tau, \mathsf{V}\Lambda_{\oplus}) = \frac{1}{2}.$$

This concludes the proof. $\square$

4.5.2 "Context-free" context equivalence

We continue following Pitts [Pit11] and present a coinductive characterisation of probabilistic context preorder phrased in terms of $\Lambda_{\oplus}$-relations. Indeed, the notion of context we defined is extremely concrete, in that it prevents from working up-to α-equivalence classes of syntax trees. In Section 4.5.3, such characterisation comes in handy when dealing with probabilistic CIU-equivalence.

Definition 4.5.13. *A $\Lambda_{\oplus}$-relation $\mathcal{R}$ is said to be adequate if, for every $M, N \in \Lambda_{\oplus}(\emptyset)$, $\emptyset \vdash M \mathrel{\mathcal{R}} N$ implies $M \Downarrow_p$ and $N \Downarrow_q$, with $p \leq q$.*

Definition 4.5.14. *Let $\mathbb{CA}$ be the collection of all compatible and adequate $\Lambda_{\oplus}$-relations. Let $\leq_{\oplus}^{\mathsf{ca}}$ be defined as $\bigcup \mathbb{CA}$, that is $\leq_{\oplus}^{\mathsf{ca}} = \bigcup \mathbb{CA}$.*

We first prove that $\leq_{\oplus}^{\mathsf{ca}} \in \mathbb{CA}$, hence that $\leq_{\oplus}^{\mathsf{ca}}$ is a compatible and adequate $\Lambda_{\oplus}$-relation. Later we show that $\leq_{\oplus} = \leq_{\oplus}^{\mathsf{ca}}$.

Lemma 4.5.15. *For every $\mathcal{R}, \mathcal{T} \in \mathbb{CA}$, $\mathcal{R} \circ \mathcal{T} \in \mathbb{CA}$.*

Proof. We need to show that $\mathcal{R} \circ \mathcal{T} = \{(M, N) \mid \exists L \in \Lambda_{\oplus}(\overline{x}).\, \overline{x} \vdash M \mathrel{\mathcal{R}} L \wedge \overline{x} \vdash L \mathrel{\mathcal{T}} N\}$ is a compatible and adequate $\Lambda_{\oplus}$-relation. Obviously, $\mathcal{R} \circ \mathcal{T}$ is adequate: for every $(M, N) \in \mathcal{R} \circ \mathcal{T}$, there exists a term L such that $M \Downarrow_p \Rightarrow L \Downarrow_q \Rightarrow N \Downarrow_r$, with $p \leq q \leq r$. Then, $M \Downarrow_p \Rightarrow N \Downarrow_r$, with $p \leq r$.

Note that the identity relation $\mathsf{ID} = \{(M, M) \mid M \in \Lambda_{\oplus}(\overline{x})\}$ is in $\mathcal{R} \circ \mathcal{T}$. Therefore $\mathcal{R} \circ \mathcal{T}$ is reflexive and, in particular, it satisfies compatibility property ($\mathsf{Com1}$). Proving ($\mathsf{Com2}$) means to show that, if $\overline{x} \cup \{x\} \vdash M\ (\mathcal{R} \circ \mathcal{T})\ N$, then $\overline{x} \vdash \lambda x.M\ (\mathcal{R} \circ \mathcal{T})\ \lambda x.N$. From the hypothesis, it follows that there exists a term L such that $\overline{x} \cup \{x\} \vdash M \mathrel{\mathcal{R}} L$ and $\overline{x} \cup \{x\} \vdash L \mathrel{\mathcal{T}} N$. Since both $\mathcal{R}$ and $\mathcal{T}$ are in $\mathbb{CA}$, hence compatible, it holds $\overline{x} \vdash \lambda x.M \mathrel{\mathcal{R}} \lambda x.L$ and $\overline{x} \vdash \lambda x.L \mathrel{\mathcal{T}} \lambda x.N$. The latter together imply $\overline{x} \vdash \lambda x.M\ (\mathcal{R} \circ \mathcal{T})\ \lambda x.N$. Proving ($\mathsf{Com3}$) means to show that, if $\overline{x} \vdash M\ (\mathcal{R} \circ \mathcal{T})\ N$ and $\overline{x} \vdash P\ (\mathcal{R} \circ \mathcal{T})\ R$, then $\overline{x} \vdash (M)\, P\ (\mathcal{R} \circ \mathcal{T})\ (N)\, R$. From the hypothesis, it follows that there exist two terms L, O such that, on the one hand, $\overline{x} \vdash M \mathrel{\mathcal{R}} L$ and $\overline{x} \vdash L \mathrel{\mathcal{T}} N$, and on the other hand, $\overline{x} \vdash P \mathrel{\mathcal{R}} O$ and $\overline{x} \vdash O \mathrel{\mathcal{T}} R$. Since both $\mathcal{R}$ and $\mathcal{T}$ are in $\mathbb{CA}$, hence compatible, it holds

$$\overline{x} \vdash M \mathrel{\mathcal{R}} L \wedge \overline{x} \vdash P \mathrel{\mathcal{R}} O \Rightarrow \overline{x} \vdash (M)\, P \mathrel{\mathcal{R}}\ (L)\, O,$$

and

$$\overline{x} \vdash L\ \mathcal{T}\ N \wedge \overline{x} \vdash O\ \mathcal{T}\ R \Rightarrow \overline{x} \vdash (L)\, O\ \mathcal{T}\ \ (N)\, R.$$

The two imply $\overline{x} \vdash\ (M)\, P\ \ (\mathcal{R} \circ \mathcal{T})\ \ (N)\, R$. One can prove property (Com4) by a similar reasoning. $\square$

Lemma 4.5.16. *$\Lambda_\oplus$-relation $\leq_\oplus^{\mathsf{ca}}$ is adequate.*

Proof. It suffices to notice that the property of being adequate is closed under taking unions of relations. Indeed, if $\mathcal{R}, \mathcal{T}$ are adequate relations, then it is simple to see that the union $\mathcal{R} \cup \mathcal{T}$ also is: for every $(M, N) \in \mathcal{R} \cup \mathcal{T}$, either $\overline{x} \vdash\ M\ \mathcal{R}\ N$ or $\overline{x} \vdash M\ \mathcal{T}\ N$. Either way, $M{\Downarrow_p} \Rightarrow N{\Downarrow_q}$ with $p \leq q$, implying $\mathcal{R} \cup \mathcal{T}$ adequate. $\square$

Lemma 4.5.17. *$\Lambda_\oplus$-relation $\leq_\oplus^{\mathsf{ca}}$ is a precongruence.*

Proof. We show that $\leq_\oplus^{\mathsf{ca}}$ is a transitive and compatible relation. Lemma 4.5.15 entails $\leq_\oplus^{\mathsf{ca}} \circ \leq_\oplus^{\mathsf{ca}} \subseteq \leq_\oplus^{\mathsf{ca}}$, implying $\leq_\oplus^{\mathsf{ca}}$ transitive.

The identity relation $\mathsf{ID} = \{(M, M) \mid M \in \Lambda_\oplus(\overline{x})\}$ is in $\mathbb{CA}$, which implies $\leq_\oplus^{\mathsf{ca}}$ reflexive, hence $\leq_\oplus^{\mathsf{ca}}$ enjoys the property (Com1). It is clear that property (Com2) is closed under taking unions of relations, so that $\leq_\oplus^{\mathsf{ca}}$ satisfies (Com2) as well. The same is not true for properties (Com3) and (Com4). By Lemma 4.3.12, for (Com3) it suffices to show that $\leq_\oplus^{\mathsf{ca}}$ satisfies (Com3L) and (Com3R), and these latter clearly are closed under taking unions of relations. Using Lemma 4.3.13, the same holds for (Com4). $\square$

Corollary 4.5.18. *$\leq_\oplus^{\mathsf{ca}}$ is the largest compatible and adequate $\Lambda_\oplus$-relation.*

Proof. Straightforward consequence of Lemma 4.5.16 and Lemma 4.5.17. $\square$

Lemma 4.5.19. *$\Lambda_\oplus$-relations $\leq_\oplus$ and $\leq_\oplus^{\mathsf{ca}}$ coincide.*

Proof. Relation $\leq_\oplus$ is adequate (Definition 4.5.5) and a precongruence (Lemma 4.5.7). Therefore $\leq_\oplus \in \mathbb{CA}$, implying $\leq_\oplus \subseteq \leq_\oplus^{\mathsf{ca}}$.

We now prove the converse. Since $\leq_\oplus^{\mathsf{ca}}$ is a precongruence (Lemma 4.5.17), hence a compatible relation, it holds that, for every $M, N \in \Lambda_\oplus(\overline{x})$ and every $C \in \mathsf{C}\Lambda_\oplus(\overline{x}\,;\,\overline{y})$, $\overline{x} \vdash M \leq_\oplus^{\mathsf{ca}} N$ implies $\overline{y} \vdash C\,\langle M\rangle\ \leq_\oplus^{\mathsf{ca}} C\,\langle N\rangle$. Therefore, for every $M, N \in \Lambda_\oplus(\overline{x})$ and every $C \in \mathsf{C}\Lambda_\oplus(\overline{x}\,;\,\emptyset)$, it follows

$$\overline{x} \vdash M \leq_\oplus^{\mathsf{ca}} N \Rightarrow \emptyset \vdash C\,\langle M\rangle\ \leq_\oplus^{\mathsf{ca}} C\,\langle N\rangle$$

which implies, by the fact that $\leq_\oplus^{\mathsf{ca}}$ is adequate,

$$C\,\langle M\rangle \Downarrow_p \Rightarrow C\,\langle N\rangle \Downarrow_q \text{ with } p \leq q,$$

namely, $\overline{x} \vdash M \leq_\oplus N$ by Definition 4.5.5. $\square$

4.5.3 Probabilistic CIU-equivalence

Probabilistic CIU-equivalence (acronym for "Uses of Closed Instantiations") is a simpler characterisation of that kind of program equivalence we are interested in, namely probabilistic context equivalence. Here we prove the two notions to coincide.

In general, while context equivalences envisage a quantification over all contexts, CIU-equivalences are limited to a restricted class of contexts without affecting the associated notion of program equivalence. In CbN, such class of contexts is that of *applicative contexts*, namely contexts of the form $(\ldots(((\langle\cdot\rangle)\, M_1)\, M_2 \ldots)\, M_n$. In particular, we follow Pitts [Pit11] in representing applicative contexts as a stack of (applicative) frames.

Definition 4.5.20. *The set $\mathcal{FS}$ of* frame stacks *is given by the following grammar:*

$$\mathbf{S}, \mathbf{T} ::= \mathtt{nil} \;\;|\;\; ((\langle\cdot\rangle)\, M :: \mathbf{S}.$$

Notation. We denote the set of free variables of a frame stack $\mathbf{S}$ as $\mathsf{FV}(\mathbf{S})$, which can be simply defined as the union of the variables occurring free in the terms appearing into $\mathbf{S}$. Then $\mathcal{FS}(\overline{x})$ is the set of frame stacks whose free variables are all from $\overline{x}$.

Definition 4.5.21. *Let $\mathbf{S} \in \mathcal{FS}(\overline{x})$ and $M \in \Lambda_{\oplus}(\overline{x})$. The $\Lambda_{\oplus}$-term $E_{\mathbf{S}}(M) \in \Lambda_{\oplus}(\overline{x})$ is defined by induction on $\mathbf{S}$ as follows:*

$$\begin{aligned} E_{\mathtt{nil}}(M) &= M; \\ E_{((\langle\cdot\rangle)N::\mathbf{S}}(M) &= E_{\mathbf{S}}((M)\, N). \end{aligned}$$

We now lift CbN reduction to a relation $\rightsquigarrow$ between pairs of frame stacks and close $\Lambda_{\oplus}$-terms $(\mathbf{S}, M)$, and *sequences* of pairs of the same kind.

Definition 4.5.22. *Reduction $\rightsquigarrow$ is the least binary relation on $(\mathcal{FS} \times \Lambda_{\oplus}) \times (\mathcal{FS} \times \Lambda_{\oplus})^*$ such that:*

- $(\mathbf{S}, (M)\, N) \rightsquigarrow (((\langle\cdot\rangle)\, N :: \mathbf{S}, M)$;
- $(\mathbf{S}, M \oplus N) \rightsquigarrow (\mathbf{S}, M), (\mathbf{S}, N)$;
- $(((\langle\cdot\rangle)\, N :: \mathbf{S}, \lambda x.M) \rightsquigarrow (\mathbf{S}, M\,[N/x])$.

The idea of small-step CbN approximation semantics, introduced in Section 4.1.2, is adapted to this case, resulting in a formal system whose judgements are in the form $(\mathbf{S}, M)\Downarrow^p$ (Figure 4.8). This time, however, the semantics approximates a real value in $\mathbb{R}_{[0,1]}$.

$$\frac{}{(\mathbf{S}, M)\downarrow^0}\ (\mathsf{empty})$$

$$\frac{}{(\mathtt{nil}, V)\downarrow^1}\ (\mathsf{value})$$

$$\frac{(\mathbf{S}, M) \rightsquigarrow (\mathbf{T}_1, N_1), \ldots, (\mathbf{T}_n, N_n) \quad (\mathbf{T}_i, N_i)\downarrow^{p_i}}{(\mathbf{S}, M)\downarrow^{\frac{1}{n}\sum_{i=1}^n p_i}}\ (\mathsf{term})$$

Figure 4.8: Small-step CbN approximation semantics for $(\mathcal{FS} \times \Lambda_\oplus(\emptyset))$.

Definition 4.5.23. *The* CbN probability of CIU-convergence *of* $(\mathbf{S}, M)$ *is the real number* $\mathbb{C}(\mathbf{S}, M)$ *defined as:*

$$\mathbb{C}(\mathbf{S}, M) = \sup_{p \in \mathbb{R}_{[0,1]}} (\mathbf{S}, M)\downarrow^p.$$

This notion of convergence is related to the one expressed in terms of value distributions (Definition 4.5.5) as follows. We first show two technical lemmas, which put to use the small-step CbN approximation semantics for value distributions.

Lemma 4.5.24. *Let* $\mathbf{S} \in \mathcal{FS}(\emptyset)$ *and* $M \in \Lambda_\oplus(\emptyset)$. *If* $(\mathbf{S}, M)\downarrow^p$ *then there is* $\mathscr{D} \in \mathcal{P}_{\mathsf{V}\Lambda_\oplus}$ *such that* $E_{\mathbf{S}}(M) \Rightarrow \mathscr{D}$ *with* $\sum \mathscr{D} = p$.

Proof. By induction on the derivation of $(\mathbf{S}, M)\downarrow^p$, looking at the last rule used.

- (empty) rule entails $(\mathbf{S}, M)\downarrow^0$. Consider the empty distribution $\mathscr{D} = \emptyset$ and observe that $E_{\mathbf{S}}(M) \Rightarrow \mathscr{D}$ by the rule (se). Of course, $\sum \mathscr{D} = 0 = p$.

- (value) rule entails $(\mathbf{S}, M)\downarrow^1$, implying $\mathbf{S} = \mathtt{nil}$ and M a value V. Consider the distribution $\mathscr{D} = \{V^1\}$ and observe that $E_{\mathtt{nil}}(V) = V \Rightarrow \mathscr{D}$ by the rule (sv). Of course, $\sum \mathscr{D} = 1 = p$.

- (term) rule entails $(\mathbf{S}, M)\downarrow^{\frac{1}{n}\sum_{i=1}^n p_i}$, implying that $(\mathbf{S}, M) \rightsquigarrow (\mathbf{T}_1, N_1), \ldots, (\mathbf{T}_n, N_n)$ and $(\mathbf{T}_i, N_i)\downarrow^{p_i}$, for every $i \in \{1, \ldots, n\}$. The induction hypothesis entails $\mathscr{E}_1, \ldots, \mathscr{E}_n$ such that $E_{\mathbf{T}_i}(N_i) \Rightarrow \mathscr{E}_i$ with $\sum \mathscr{E}_i = p_i$.

  We now proceed by cases according to the structure of M.

  – If $M = \lambda x.L$, then $\mathbf{S} = (\langle\cdot\rangle)\, P :: \mathbf{T}$ implying $n = 1$, $\mathbf{T}_1 = \mathbf{T}$ and $N_1 = L\,[P/x]$. Consider the distribution $\mathscr{D} = \mathscr{E}_1$ and observe that $E_{\mathbf{S}}(M) = E_{(\langle\cdot\rangle)P::\mathbf{T}}(\lambda x.L) = E_{\mathbf{T}}((\lambda x.L)\, P) \mapsto E_{\mathbf{T}}(L\,[P/x]) = E_{\mathbf{T}_1}(N_1)$. Hence, $E_{\mathbf{S}}(M) \Rightarrow \mathscr{D}$ by the rule (st). Moreover, $\sum \mathscr{D} = \sum \mathscr{E}_1 = p_1 = \frac{1}{n}\sum_{i=1}^n p_i = p$.

- If $M = L \oplus P$, then $n = 2$, $\mathbf{T}_1 = \mathbf{T}_2 = \mathbf{S}$, $N_1 = L$ and $N_2 = P$. Consider the distribution $\mathscr{D} = \sum_{i=1}^{2} \frac{1}{2}\mathscr{E}_i$ and observe that $E_{\mathbf{S}}(M) = E_{\mathbf{S}}(L \oplus P) \mapsto E_{\mathbf{S}}(L), E_{\mathbf{S}}(P) = E_{\mathbf{T}_1}(N_1), E_{\mathbf{T}_2}(N_2)$. Hence, $E_{\mathbf{S}}(M) \Rightarrow \mathscr{D}$ by the rule (st). Moreover, $\sum \mathscr{D} = \sum \sum_{i=1}^{2} \frac{1}{2}\mathscr{E}_i = \frac{1}{2}\sum_{i=1}^{2} \sum \mathscr{E}_i = \frac{1}{2}\sum_{i=1}^{2} p_i = p$.
- If $M = (L)\,P$, then $n = 1$, $\mathbf{T}_1 = (\langle\cdot\rangle)\,P :: \mathbf{S}$ and $N_1 = L$. Consider the distribution $\mathscr{D} = \mathscr{E}_1$ and observe that $E_{(\langle\cdot\rangle)P::\mathbf{S}}(L) \Rightarrow \mathscr{E}_1$ implies $E_{\mathbf{S}}(M) \Rightarrow \mathscr{D}$. Moreover, $\sum \mathscr{D} = \sum \mathscr{E}_1 = p_1 = \frac{1}{n}\sum_{i=1}^{n} p_i = p$.

This concludes the proof. $\square$

Lemma 4.5.25. *For every $\mathscr{D} \in \mathcal{P}_{\mathsf{V}\Lambda_\oplus}$ and $M \in \Lambda_\oplus(\emptyset)$, if $M \Rightarrow \mathscr{D}$ then there is $\mathbf{S} \in \mathcal{FS}(\emptyset)$ and $N \in \Lambda_\oplus(\emptyset)$ such that $E_{\mathbf{S}}(N) = M$ and $(\mathbf{S}, N)\downarrow^p$ with $\sum \mathscr{D} = p$.*

Proof. By induction on the derivation of $M \Rightarrow \mathscr{D}$, looking at the last rule used.

- (se) rule entails $M \Rightarrow \emptyset$. Then, for every $\mathbf{S}$ and every N such that $E_{\mathbf{S}}(N) = M$, $(\mathbf{S}, N)\downarrow^0$ by the rule (empty). Of course, $\sum \mathscr{D} = 0 = p$.

- (sv) rule intails $M = V$, and $\mathscr{D} = \{V^1\}$ with $V \Rightarrow \{V^1\}$. Consider $\mathbf{S} = \mathtt{nil}$ and $N = V$ and verify that $E_{\mathbf{S}}(N) = E_{\mathtt{nil}}(V) = V = M$. The rule (value) implies $(\mathtt{nil}, V)\downarrow^1$, hence $\sum \mathscr{D} = 1 = p$.

- (st) rule entails $M \Rightarrow \sum_{i=1}^{n} \frac{1}{n}\mathscr{E}_i$, implying $M \mapsto Q_1, \ldots, Q_n$ with $Q_i \Rightarrow \mathscr{E}_i$, for every $i \in \{1, \ldots, n\}$. The induction hypothesis entails $\mathbf{T}_i$ and N_i such that $E_{\mathbf{T}_i}(N_i) = Q_i$ and $(\mathbf{T}_i, N_i)\downarrow^{p_i}$ with $\sum \mathscr{E}_i = p_i$.

  We now proceed by cases according to the structure of M.

  - If $M = (\lambda x.L)\,P$, then $n = 1$ and $Q_1 = L\,[P/x]$. Consider $\mathbf{S} = (\langle\cdot\rangle)\,P :: \mathtt{nil}$, $N = \lambda x.L$ and verify that $E_{\mathbf{S}}(N) = E_{(\langle\cdot\rangle)P::\mathtt{nil}}(\lambda x.L) = E_{\mathtt{nil}}(((\lambda x.L))\,P) = ((\lambda x.L))\,P = M$. The rule (term) on the hypothesis $(\mathbf{S}, N) = ((\langle\cdot\rangle)\,P :: \mathtt{nil}, \lambda x.L) \rightsquigarrow (\mathtt{nil}, L\,[P/x]) = E_{\mathtt{nil}}(L\,[P/x])$ and $(\mathtt{nil}, L\,[P/x])\downarrow^{p_1}$, implies $(\mathbf{S}, N)\downarrow^{p_1}$. Moreover, $\sum \mathscr{D} = \sum \sum_{i=1}^{n} \frac{1}{n}\mathscr{E}_i = \sum \mathscr{E}_1 = p_1 = p$.
  - If $M = L \oplus P$, then $n = 2$, $Q_1 = L$ and $Q_2 = P$. Consider $\mathbf{S} = \mathtt{nil}$, $N = L \oplus P$ and verify that $E_{\mathbf{S}}(N) = E_{\mathtt{nil}}(L \oplus P) = L \oplus P = M$. The rule (term) on the hypothesis $(\mathbf{S}, N) = (\mathtt{nil}, L \oplus P) \rightsquigarrow (\mathtt{nil}, L), (\mathtt{nil}, P) = E_{\mathtt{nil}}(L), E_{\mathtt{nil}}(P)$ and $(\mathtt{nil}, L)\downarrow^{p_1}$ and $(\mathtt{nil}, P)\downarrow^{p_2}$, implies $(\mathbf{S}, N)\downarrow^{\frac{1}{2}\sum_{i=1}^{2} p_i}$. Moreover, $\sum \mathscr{D} = \sum \sum_{i=1}^{n} \frac{1}{n}\mathscr{E}_i = \sum \sum_{i=1}^{n} \frac{1}{2}\mathscr{E}_i = \frac{1}{2}\sum_{i=1}^{2} \sum \mathscr{E}_i = \frac{1}{2}\sum_{i=1}^{2} p_i = p$.
  - If $M = (L)\,P$ and $L \mapsto R_1, \ldots, R_n$, then $Q_i = (R_i)\,P$ for every $i \in \{1, \ldots, n\}$. Consider $\mathbf{S} = (\langle\cdot\rangle)\,P :: \mathtt{nil}$, $N = L$ and verify that $E_{\mathbf{S}}(N) = E_{(\langle\cdot\rangle)P::\mathtt{nil}}(L) = E_{\mathtt{nil}}((L)\,P) = (L)\,P = M$. The rule (term) on the hypothesis $(\mathbf{S}, N) = ((\langle\cdot\rangle)\,P :: \mathtt{nil}, L) \rightsquigarrow ((\langle\cdot\rangle)\,P :: \mathtt{nil}, R_1), \ldots, ((\langle\cdot\rangle)\,P ::

$\mathtt{nil}, R_n) = (E_{(\langle\cdot\rangle) P::\mathtt{nil}}(R_1), \ldots, E_{(\langle\cdot\rangle) P::\mathtt{nil}}(R_n))$ and, for every $i \in \{1, \ldots, n\}$, $((\langle\cdot\rangle)\, P :: \mathtt{nil}, R_i){\downarrow}^{p_i}$, implies $(\mathbf{S}, N){\downarrow}^{\frac{1}{n}\sum_{i=1}^{n} p_i}$. Moreover, $\sum \mathscr{D} = \sum \sum_{i=1}^{n} \frac{1}{n} \mathscr{E}_i = \frac{1}{n} \sum_{i=1}^{n} \sum \mathscr{E}_i = \frac{1}{n} \sum_{i=1}^{n} p_i = p$.

This concludes the proof. □

Recall that big-step and small-step CbN semantics for value distributions can simulate each other [DLZ12]. Therefore, the previous two results directly entail the following:

Corollary 4.5.26. *$(\mathbf{S}, M){\downarrow}^p$ if and only if $E_{\mathbf{S}}(M) \Downarrow \mathscr{D}$ with $\sum \mathscr{D} = p$.*

Lemma 4.5.27. *For every $\mathbf{S} \in \mathcal{FS}(\emptyset)$ and $M \in \Lambda_{\oplus}(\emptyset)$, $\mathbb{C}(\mathbf{S}, M) = p$ if and only if $E_{\mathbf{S}}(M){\Downarrow}_p$. In particular, $M{\Downarrow}_p$ holds if and only if $\mathbb{C}(\mathtt{nil}, M) = p$.*

Proof. Straightforward by Corollary 4.5.26 and the property of ω-completeness concerning value distributions:

$$
\begin{aligned}
p = \mathbb{C}(\mathbf{S}, M) &= \sup_{q \in \mathbb{R}_{[0,1]}} (\mathbf{S}, M){\downarrow}^q \\
&= \sup_{E_{\mathbf{S}}(M) \Downarrow \mathscr{D}} \sum \mathscr{D} \\
&= \sum \sup_{E_{\mathbf{S}}(M) \Downarrow \mathscr{D}} \mathscr{D} = \sum [\![E_{\mathbf{S}}(M)]\!] = E_{\mathbf{S}}(M){\Downarrow}_p.
\end{aligned}
$$
□

Probabilistic CIU-preorder and equivalence are defined on closed terms. As in the case of (bi)similarities, it is intended that they are extended to open terms via closing term-substitutions.

Definition 4.5.28. *The* probabilistic CIU-preorder $\preceq^{\mathsf{CIU}}$ *stipulates $M \preceq^{\mathsf{CIU}} N$ if $\mathbb{C}(\mathbf{S}, M) \leq \mathbb{C}(\mathbf{S}, N)$ for every frame stack $\mathbf{S} \in \mathcal{FS}(\emptyset)$. The equivalence induced by $\preceq^{\mathsf{CIU}}$ is* probabilistic CIU-equivalence, *denoted $\cong^{\mathsf{CIU}}$.*

The following lemma asserts that β-reduction is validated by $\cong^{\mathsf{CIU}}$.

Lemma 4.5.29. *For every $\overline{x} \in \mathcal{P}_{\mathsf{FIN}}(\mathcal{V})$, $\overline{x} \vdash (\lambda x.M)\, N \;\cong^{\mathsf{CIU}}\; M\,[N/x]$.*

Proof. We show that both $\overline{x} \vdash (\lambda x.M)\, N \;\preceq^{\mathsf{CIU}}\; M\,[N/y]$ and $\overline{x} \vdash M\,[N/x] \;\preceq^{\mathsf{CIU}}$ $(\lambda x.M)\, N$ hold. Since $\preceq^{\mathsf{CIU}}$ is defined on open terms by requiring the closure under term-substitutions, it suffices to show the result for close $\Lambda_{\oplus}$-terms only: we therefore show $(\lambda x.M)\, N \preceq^{\mathsf{CIU}} M\,[N/x]$ and $M\,[N/x] \preceq^{\mathsf{CIU}} (\lambda x.M)\, N$.

We start with $(\lambda x.M)\, N \preceq^{\mathsf{CIU}} M\,[N/x]$ and prove that, for every close frame stack $\mathbf{S}$, $\mathbb{C}(\mathbf{S}, (\lambda x.M)\, N) \leq \mathbb{C}(\mathbf{S}, M\,[N/x])$. The latter is an obvious consequence of the fact that $(\mathbf{S}, (\lambda x.M)\, N)$ reduces to $(\mathbf{S}, M\,[N/x])$. We give the details by distinguishing two cases:

- If $\mathbf{S} = \mathtt{nil}$, then $(\mathbf{S}, (\lambda x.M)\, N) \rightsquigarrow ((\langle\cdot\rangle)\, N :: \mathbf{S}, \lambda x.M) \rightsquigarrow (\mathbf{S}, M\,[N/x])$ which implies

$$\begin{aligned}\mathbb{C}(\mathbf{S}, (\lambda x.M)\, N) &= \sup_{p \in \mathbb{R}_{[0,1]}} (\mathbf{S}, (\lambda x.M)\, N)\downarrow^{p} \\ &= \sup_{p \in \mathbb{R}_{[0,1]}} (\mathbf{S}, M\,[N/x])\downarrow^{p} = \mathbb{C}(\mathbf{S}, M\,[N/x]).\end{aligned}$$

- The case $\mathbf{S} = (\langle\cdot\rangle)\, L :: \mathbf{T}$ follows by a similar reasoning.

We now prove $M\,[N/x] \preceq^{\mathsf{CIU}} (\lambda x.M)\, N$. Consider $p \in \mathbb{R}_{[0,1]}$ as $(\mathbf{S}, M\,[N/x])\downarrow^{p}$ and distinguish two cases:

- If $\mathbf{S} = \mathtt{nil}$ and $p = 0$, then $(\mathbf{S}, (\lambda x.M)\, N)\downarrow^{0}$ holds by the rule (empty). Hence $\mathbb{C}(\mathbf{S}, M\,[N/x]) = 0 \leq \mathbb{C}(\mathbf{S}, (\lambda x.M)\, N)$. Otherwise, from

$$\frac{((\langle\cdot\rangle)\, N :: \mathbf{S}, \lambda x.M) \rightsquigarrow (\mathbf{S}, M\,[N/x]) \quad (\mathbf{S}, M\,[N/x])\downarrow^{p}}{((\langle\cdot\rangle)\, N :: \mathbf{S}, \lambda x.M)\downarrow^{p}}\ (\mathsf{term})$$

  follows

$$\frac{(\mathbf{S}, M) \rightsquigarrow ((\langle\cdot\rangle)\, N :: \mathbf{S}, \lambda x.M) \quad ((\langle\cdot\rangle)\, N :: \mathbf{S}, \lambda x.M)\downarrow^{p}}{(\mathbf{S}, (\lambda x.M)\, N)\downarrow^{p}}\ (\mathsf{term})$$

  which implies

$$\begin{aligned}\mathbb{C}(\mathbf{S}, M\,[N/x]) &= \sup_{p \in \mathbb{R}_{[0,1]}} (\mathbf{S}, M\,[N/x])\downarrow^{p} \\ &= \sup_{p \in \mathbb{R}_{[0,1]}} (\mathbf{S}, (\lambda x.M)\, N)\downarrow^{p} = \mathbb{C}(\mathbf{S}, (\lambda x.M)\, N).\end{aligned}$$

- The case $\mathbf{S} = (\langle\cdot\rangle)\, L :: \mathbf{T}$ follows by a similar reasoning.

This concludes the proof. $\square$

Definition 4.5.28 implies $\preceq^{\mathsf{CIU}}$ preorder, that is reflexive and transitive. We show that $\preceq^{\mathsf{CIU}}$ is a compatible $\Lambda_{\oplus}$-relation by showing that $\preceq^{\mathsf{CIU}} = (\preceq^{\mathsf{CIU}})^{H}$ (as we already know that $(\preceq^{\mathsf{CIU}})^{H}$ is compatible by Lemma 4.3.17), which ultimately needs Howe's construction to be extented to (closed) frame stacks. This implies that $\preceq^{\mathsf{CIU}}$ coincides with $\leq_{\oplus}$, and hence that $\cong^{\mathsf{CIU}}$ coincides with $\simeq_{\oplus}$.

$$\frac{}{\texttt{nil}\ \mathcal{R}^H\ \texttt{nil}}\ (\mathsf{Howstk1})$$

$$\frac{\emptyset \vdash M\,\mathcal{R}^H\,N \qquad \mathbf{S}\,\mathcal{R}^H\,\mathbf{T}}{((\langle\cdot\rangle)\,M :: \mathbf{S})\ \mathcal{R}^H\ ((\langle\cdot\rangle)\,N :: \mathbf{T})}\ (\mathsf{Howstk2})$$

Figure 4.9: Howe's rules on frame stacks.

In fact, since $\preceq^{\mathsf{CIU}}$ is reflexive, we only have to show $(\preceq^{\mathsf{CIU}})^H \subseteq \preceq^{\mathsf{CIU}}$, as the converse inclusion is a consequence of Lemma 4.3.19. Moreover, due to the fact that $\preceq^{\mathsf{CIU}}$ is defined on open terms by taking closing term-substitutions, $\preceq^{\mathsf{CIU}}$ and its lifting $(\preceq^{\mathsf{CIU}})^H$ are closed under term-substitution (Lemma 4.3.20), so that the above inclusion need to be proved on closed terms only.

Howe's construction is extended to (closed) frame stacks by the rules in Figure 4.9. We then provide the following context lemma, asserting that $M\ (\preceq^{\mathsf{CIU}})^H\ N$ implies $M \preceq^{\mathsf{CIU}} N$ (this latter under the form $(\mathbf{S}, M)\downarrow^p$ and $p \leq \mathbb{C}(\mathbf{S}, N)$).

Lemma 4.5.30. *For every $\mathbf{S}, \mathbf{T} \in \mathcal{FS}(\emptyset)$ and $M, N \in \Lambda_\oplus(\emptyset)$, if $\mathbf{S}\ (\preceq^{\mathsf{CIU}})^H\ \mathbf{T}$ and $M\ (\preceq^{\mathsf{CIU}})^H\ N$ and $(\mathbf{S}, M)\downarrow^p$, then $p \leq \mathbb{C}(\mathbf{T}, N)$.*

Proof. By induction on the structure of the derivation of $(\mathbf{S}, M)\downarrow^p$.

- (empty) entails $(\mathbf{S}, M)\downarrow^0$, hence $0 \leq \mathbb{C}(\mathbf{T}, N)$.
- (value) entails $\mathbf{S} = \texttt{nil}$, $M = \lambda x.L$ and $p = 1$, then $\mathbf{S}\ (\preceq^{\mathsf{CIU}})^H\ \mathbf{T}$ entails $\mathbf{T} = \texttt{nil}$. From $M\ (\preceq^{\mathsf{CIU}})^H\ N$ follows P such that $x \vdash L\ (\preceq^{\mathsf{CIU}})^H\ P$ and $\emptyset \vdash \lambda x.P\ \preceq^{\mathsf{CIU}}\ N$. The latter implies that $p = 1 \leq \mathbb{C}(\texttt{nil}, N) = \mathbb{C}(\mathbf{T}, N)$.
- (term) entails the last step of the derivation of $(\mathbf{S}, M)\downarrow^p$ to be as follows:

  $$\frac{(\mathbf{S}, M) \rightsquigarrow (\mathbf{U}_1, L_1), \ldots, (\mathbf{U}_n, L_n) \quad (\mathbf{U}_i, L_i)\downarrow^{p_i}}{(\mathbf{S}, M)\downarrow^{\frac{1}{n}\sum_{i=1}^n p_i}}\ (\mathsf{term})$$

  We distinguish three cases as three are the clauses of Definition 4.5.22 for $\rightsquigarrow$:

  - If $M = PQ$, then $n = 1$, $\mathbf{U}_1 = (\langle\cdot\rangle)\,Q :: \mathbf{S}$ and $L_1 = P$. From $M\ (\preceq^{\mathsf{CIU}})^H\ N$ follows R, S such that $P\ (\preceq^{\mathsf{CIU}})^H\ R$, $Q\ (\preceq^{\mathsf{CIU}})^H\ S$ and $(R)\,S \preceq^{\mathsf{CIU}} N$. The hypothesis $\mathbf{S}\ (\preceq^{\mathsf{CIU}})^H\ \mathbf{T}$ and $Q\ (\preceq^{\mathsf{CIU}})^H\ S$ entail $\mathbf{U}_1 = (\langle\cdot\rangle)\,Q :: \mathbf{S}\ (\preceq^{\mathsf{CIU}})^H\ (\langle\cdot\rangle)\,S :: \mathbf{T}$. By the induction hypothesis, $p \leq \mathbb{C}((\langle\cdot\rangle)\,S :: \mathbf{T}, R)$. Observe that $(\mathbf{T}, (R)\,S) \rightsquigarrow ((\langle\cdot\rangle)\,S :: \mathbf{T}, R)$, and as a consequence $p \leq \mathbb{C}(\mathbf{T}, (R)\,S)$. Since $(R)\,S \preceq^{\mathsf{CIU}} N$, it follows $\mathbb{C}(\mathbf{T}, (R)\,S) \leq \mathbb{C}(\mathbf{T}, N)$ and $p \leq \mathbb{C}(\mathbf{T}, N)$.

– If $M = P \oplus Q$, then $n = 2$, $\mathbf{U}_1 = \mathbf{U}_2 = \mathbf{S}$ and $L_1 = P$, $L_2 = Q$. The hypothesis $\mathbf{S}\ (\preceq^{\mathsf{CIU}})^H\ \mathbf{T}$ entails $\mathbf{U}_1(\preceq^{\mathsf{CIU}})^H\mathbf{T}$ and $\mathbf{U}_2(\preceq^{\mathsf{CIU}})^H\mathbf{T}$. From $M\ (\preceq^{\mathsf{CIU}})^H\ N$ follows R, S such that $P\ (\preceq^{\mathsf{CIU}})^H\ R$, $Q\ (\preceq^{\mathsf{CIU}})^H\ S$ and $R \oplus S\ \preceq^{\mathsf{CIU}}\ N$. By a double induction hypothesis, $p \leq \mathbb{C}(\mathbf{T}, R)$ and $p \leq \mathbb{C}(\mathbf{T}, S)$. Observe that $(\mathbf{T}, (R)\, S) \rightsquigarrow (\mathbf{T}, R), (\mathbf{T}, S)$, and as a consequence $p \leq \mathbb{C}(\mathbf{T}, R \oplus S)$. Since $R \oplus S \preceq^{\mathsf{CIU}} N$, it follows $\mathbb{C}(\mathbf{T}, R \oplus S) \leq \mathbb{C}(\mathbf{T}, N)$ and $p \leq \mathbb{C}(\mathbf{T}, N)$.

– If $M = \lambda x.P$, then $\mathbf{S} = (\langle \cdot \rangle)\, Q :: \mathbf{U}$ as it is the only case remaining. Hence $n = 1$, $\mathbf{U}_1 = \mathbf{U}$ and $L_1 = P\,[Q/x]$. The hypothesis $\mathbf{S}\ (\preceq^{\mathsf{CIU}})^H\ \mathbf{T}$ entails $\mathbf{T} = (\langle \cdot \rangle)\, R :: \mathbf{V}$, with $Q\ (\preceq^{\mathsf{CIU}})^H\ R$ and $\mathbf{U}\ (\preceq^{\mathsf{CIU}})^H\ \mathbf{V}$. From $M(\preceq^{\mathsf{CIU}})^H N$ follows S such that $x \vdash P\ (\preceq^{\mathsf{CIU}})^H\ S$ and $\emptyset \vdash \lambda x.S\ \ \preceq^{\mathsf{CIU}}\ \ N$. From $x \vdash P\ (\preceq^{\mathsf{CIU}})^H\ S$ and $\emptyset \vdash Q\ (\preceq^{\mathsf{CIU}})^H\ R$, by substitutivity of $(\preceq^{\mathsf{CIU}})^H$, follow $\emptyset \vdash P\,[Q/x]\ \ (\preceq^{\mathsf{CIU}})^H\ \ S\,[R/x]$. By the induction hypothesis, $p \leq \mathbb{C}(\mathbf{V}, S\,[R/x])$. Observe that $(\mathbf{T}, \lambda x.S) = ((\langle \cdot \rangle)\, R :: \mathbf{V}, \lambda x.S) \rightsquigarrow (\mathbf{V}, S\,[R/x])$, and as a consequence $p \leq \mathbb{C}(\mathbf{V}, S\,[R/x]) \leq \mathbb{C}(\mathbf{T}, \lambda S.x)$. Since $\lambda x.S \preceq^{\mathsf{CIU}} N$, it follows $\mathbb{C}(\mathbf{T}, \lambda x.S) \leq \mathbb{C}(\mathbf{T}, N)$ and $p \leq \mathbb{C}(\mathbf{T}, N)$.

This concludes the proof. □

Corollary 4.5.31. *It holds that* $\preceq^{\mathsf{CIU}} = (\preceq^{\mathsf{CIU}})^H$*. Hence,* $\preceq^{\mathsf{CIU}}$ *is a compatible* $\Lambda_\oplus$*-relation.*

Proof. Since $\preceq^{\mathsf{CIU}}$ is reflexive, $\preceq^{\mathsf{CIU}} \subseteq (\preceq^{\mathsf{CIU}})^H$ by Lemma 4.3.19. Since $(\preceq^{\mathsf{CIU}})^H$ is reflexive, the converse inclusion $(\preceq^{\mathsf{CIU}})^H \subseteq \preceq^{\mathsf{CIU}}$ is a straightforward consequence of Lemma 4.5.30, considering $\mathbf{T} = \mathbf{S}$. It follows $\preceq^{\mathsf{CIU}}$ compatible due to the fact that $(\preceq^{\mathsf{CIU}})^H$ already is (Lemma 4.3.17). □

We now prove that $\preceq^{\mathsf{CIU}}$ coincides with $\leq_\oplus$, hence that $\cong^{\mathsf{CIU}}$ coincides with $\simeq_\oplus$.

Theorem 4.5.32. *For all* $\overline{x} \in \mathcal{P}_{\mathsf{FIN}}(\mathcal{V})$ *and* $M, N \in \Lambda_\oplus(\overline{x})$*,* $\overline{x} \vdash M\ \ \preceq^{\mathsf{CIU}}\ \ N$ *if and only if* $\overline{x} \vdash M \leq_\oplus N$*.*

Proof. ($\preceq^{\mathsf{CIU}} \subseteq \leq_\oplus$) Since $\preceq^{\mathsf{CIU}}$ is closed under term-substitution, it suffices to show the result for closed $\Lambda_\oplus$-terms only: *i.e.* for all M, $N \in \Lambda_\oplus(\emptyset)$, $M \preceq^{\mathsf{CIU}} N$ implies $M \leq_\oplus N$. Corollary 4.5.31 establishes $\preceq^{\mathsf{CIU}}$ compatible $\Lambda_\oplus$-relation. Moreover, from Lemma 4.5.27 immediately follows that $\preceq^{\mathsf{CIU}}$ is also adequate. Thus, $\preceq^{\mathsf{CIU}}$ is contained in the largest compatible adequate $\Lambda_\oplus$-relation, that is $\leq_\oplus^{\mathsf{ca}}$. From $\leq_\oplus^{\mathsf{ca}} = \simeq_\oplus$ (Lemma 4.5.19) follows that $\preceq^{\mathsf{CIU}}$ is actually contained in $\leq_\oplus$. In particular, the latter means that $M \preceq^{\mathsf{CIU}} N$ implies $M \leq_\oplus N$.

($\leq_\oplus \subseteq \preceq^{\mathsf{CIU}}$) Observe that, since context preorder is compatible, $M \leq_\oplus N$ entails $E_{\mathbf{S}}(M) \leq_\oplus E_{\mathbf{S}}(N)$, for all $\mathbf{S} \in \mathcal{FS}(\emptyset)$ (using Lemma 4.5.9 along with an induction on the length $\mathbf{S}$). By the adequacy property of $\leq_\oplus$ and Lemma 4.5.27, the latter implies $M \preceq^{\mathsf{CIU}} N$. All together, $M \leq_\oplus N$ implies $M \preceq^{\mathsf{CIU}} N$.

We now address the case of open terms. If $\overline{x} \vdash M \leq_\oplus N$, then by compatibility property of $\leq_\oplus$, it follows $\emptyset \vdash \lambda\overline{x}.M \leq_\oplus \lambda\overline{x}.N$ and hence $\emptyset \vdash \lambda\overline{x}.M \preceq^{\mathsf{CIU}} \lambda\overline{x}.N$. Then, from the fact that $\preceq^{\mathsf{CIU}}$ is compatible (Corollary 4.5.31) and Lemma 4.5.29, for every suitable $\overline{L} \subseteq \Lambda_\oplus(\emptyset)$, $(\lambda\overline{x}.M)\,\overline{L} \preceq^{\mathsf{CIU}} (\lambda\overline{x}.N)\,\overline{L}$ implies $M\left[\overline{L}/\overline{x}\right] \preceq^{\mathsf{CIU}} N\left[\overline{L}/\overline{x}\right]$, i.e. $\overline{x} \vdash M \preceq^{\mathsf{CIU}} N$. □

Corollary 4.5.33. $\cong^{\mathsf{CIU}}$ *coincides with* $\simeq_\oplus$.

Proof. Straightforward consequence of Theorem 4.5.32. □

The following is the major result of this section, which highlights the reason why we developed probabilistic CIU-equivalence.

Proposition 4.5.34. $\leq_\oplus$ *and* $\lesssim$ *do not coincide.*

Proof. We prove that $N \preceq^{\mathsf{CIU}} M$ but $N \not\lesssim M$, where

$$M = \lambda x.\lambda y.(\mathbf{\Omega} \oplus \mathbf{I});$$
$$N = \lambda x.((\lambda y.\mathbf{\Omega}) \oplus (\lambda y.\mathbf{I})).$$

(in fact, the two terms are CIU-equivalent.) We have shown in Counterexample 4.5.12 that $N \not\lesssim M$, so we concentrate on $N \preceq^{\mathsf{CIU}} M$, and prove that for every $\mathbf{S} \in \mathcal{FS}(\emptyset)$, $\mathbb{C}(\mathbf{S}, N) \leq \mathbb{C}(\mathbf{S}, M)$. We distinguish three cases:

- If $\mathbf{S} = \texttt{nil}$, then $(\texttt{nil}, N)$ cannot be further reduced, and $(\texttt{nil}, N)\downarrow^1$. The same holds for $(\texttt{nil}, M)$, hence $\mathbb{C}(\mathbf{S}, N) = 1 = \mathbb{C}(\mathbf{S}, M)$.

- If $\mathbf{S} = (\langle\cdot\rangle)\, L :: \texttt{nil}$, then observe that

  $$\begin{aligned}(\mathbf{S}, N) = ((\langle\cdot\rangle)\, L :: \texttt{nil}, N) &\rightsquigarrow (\texttt{nil}, (\lambda y.\mathbf{\Omega}) \oplus (\lambda y.\mathbf{I}))\\ &\rightsquigarrow (\texttt{nil}, \lambda y.\mathbf{\Omega}), (\texttt{nil}, \lambda y.\mathbf{I}),\end{aligned}$$

  and these last two cannot be further reducted, obtaining $\mathbb{C}(\texttt{nil}, \lambda y.\mathbf{\Omega}) = 1$ and $\mathbb{C}(\texttt{nil}, \lambda y.\mathbf{I}) = 1$. Moreover, $(\mathbf{S}, M) \rightsquigarrow (\texttt{nil}, \lambda y.\mathbf{\Omega} \oplus \mathbf{I})$ which cannot be further reduced, so that $\mathbb{C}(\texttt{nil}, \lambda y.\mathbf{\Omega} \oplus \mathbf{I}) = 1$. It follows,

  $$\mathbb{C}(\mathbf{S}, N) = \frac{1}{2}\mathbb{C}(\texttt{nil}, \lambda y.\mathbf{\Omega}) + \frac{1}{2}\mathbb{C}(\texttt{nil}, \lambda y.\mathbf{I}) = \mathbb{C}(\mathbf{S}, M).$$

- If $\mathbf{S} = (\langle\cdot\rangle)\, L :: (\langle\cdot\rangle)\, P :: \mathbf{T}$, then observe that

  $$(\mathbf{S}, N) = ((\langle\cdot\rangle)\, L :: (\langle\cdot\rangle)\, P :: \mathbf{T}, N) \rightsquigarrow ((\langle\cdot\rangle)\, P :: \mathbf{T}, (\lambda y.\mathbf{\Omega}) \oplus (\lambda y.\mathbf{I}))$$

$$\rightsquigarrow (((\langle\cdot\rangle)\, P :: \mathbf{T}, \lambda y.\mathbf{\Omega}), (((\langle\cdot\rangle)\, P :: \mathbf{T}, \lambda y.\mathbf{I}),$$

and these latter reduce as

$$\begin{aligned} (((\langle\cdot\rangle)\, P :: \mathbf{T}, \lambda y.\mathbf{\Omega}) &\rightsquigarrow (\mathbf{T}, \mathbf{\Omega}); \\ (((\langle\cdot\rangle)\, P :: \mathbf{T}, \lambda y.\mathbf{I}) &\rightsquigarrow (\mathbf{T}, \mathbf{I}). \end{aligned}$$

Moreover,

$$\begin{aligned} (\mathbf{S}, M) = (((\langle\cdot\rangle)\, L :: (\langle\cdot\rangle)\, P :: \mathbf{T}, N) &\rightsquigarrow (((\langle\cdot\rangle)\, P :: \mathbf{T}, \lambda y.\mathbf{\Omega} \oplus \mathbf{I}) \\ &\rightsquigarrow (\mathbf{T}, \mathbf{\Omega} \oplus \mathbf{I}) \\ &\rightsquigarrow (\mathbf{T}, \mathbf{\Omega}), (\mathbf{T}, \mathbf{I}). \end{aligned}$$

As a consequence,

$$\mathbb{C}(\mathbf{S}, N) = \frac{1}{2}\mathbb{C}(\mathbf{T}, \mathbf{\Omega}) + \frac{1}{2}\mathbb{C}(\mathbf{T}, \mathbf{I}) = \mathbb{C}(\mathbf{S}, M).$$

This concludes the proof. □

***Remark* 4.5.35.** Notice that we always obtain equalities in the above proof, which actually implies $\lambda x.\lambda y.(\mathbf{\Omega} \oplus \mathbf{I}) \cong^{\mathsf{CIU}} \lambda x.((\lambda y.\mathbf{\Omega}) \oplus (\lambda y.\mathbf{I}))$. This latter latter entails, by Corollary 4.5.33, $\lambda x.\lambda y.(\mathbf{\Omega} \oplus \mathbf{I}) \simeq_{\oplus} \lambda x.((\lambda y.\mathbf{\Omega}) \oplus (\lambda y.\mathbf{I}))$.

Chapter 5

The discriminating power of probabilistic contexts

This chapter presents the last contribution of this thesis.

We show here that probabilistic applicative bisimilarity and probabilistic context equivalence collapse if the tested terms are pure, *deterministic*, λ-terms. In other words, if the probabilistic choices are brought into the terms only through the inputs supplied to the tested functions, applicative bisimilarity and context equivalence yield exactly the same discriminating power. To show this, we prove that, on pure λ-terms, both relations coincide with the *Lévy-Longo tree equality*, which equates terms with the same Lévy-Longo tree [Lév75, Lon83, DCG01].

Contents

5.1 Lévy-Longo trees **161**

5.2 Pure λ-terms in probabilistic contexts **164**

5.2.1 Context equivalence implies Lévy-Longo equality 164

5.2.2 Lévy-Longo equality implies Applicative bisimilarity . . 169

The first section recalls the definition of the *Lévy-Longo tree* of a λ-term. These latter play the role of Böhm trees in the *lazy* regime, in that they capture the computational content of λ-terms when [Abstraction] terms are treated as values. First examples are provided to show that probabilities can be put to use to attain *Lévy-Longo tree equality*, generally accepted as the finest extensional equivalence on pure λ-terms under a lazy regime.

The second section develops the technique needed to show that bisimilarity and context equivalence coincide with the Lévy-Longo tree equality. In particular, a result akin to the original Böhm-out is given, showing that tailored probabilistic contexts can be used to separate λ-terms with different Lévy-Longo tree.

5.1 Lévy-Longo trees

The standard theory of pure λ-calculus, essence of Barendregt's monograph [Bar84], has its roots in the notion of *head normal form* as the meaning of λ-terms, which corresponds to the notion of *Böhm tree* (briefly, BT) via the principle of unsolvability.

Recall that a pure λ-term in head normal form is of the form

$$\lambda x_1 \ldots \lambda x_n.(\ldots((y)\, M_1)\, M_2 \ldots)\, M_m,$$

with $n, m \geq 0$, and the *head variable* y is either free or $y = x_i$ for some $i \in \{1, \ldots, n\}$. The notion of β-normal form can be inductively formulated on the basis of that of head normal form.

Definition 5.1.1. β-normal forms *are inductively defined as follows:*

- *$\lambda x_1 \ldots \lambda x_n.y$ is a β-normal form (and also a head normal form);*
- *a head normal form $\lambda x_1 \ldots \lambda x_n.(\ldots((y)\, M_1)\, M_2 \ldots)\, M_m$, with $n, m \geq 0$, is a β-normal form if $M_1, \ldots, M_m$ also are.*

The above definition of nested head normal forms directly provides a procedure to define the Böhm tree of a β-normal form. The extension to a notion of generalised Böhm tree for all λ-terms, hence of a possibly infinite "β-normal form", is obtained by first reading the same definition coinductively and then appealing to the principle of unsolvability [Bar84]. Indeed, unsolvable terms (*i.e.* terms with no head normal form) are identified in the standard theory, which is reflected at the level of Böhm trees by the symbol $\perp$. Definition 5.1.2 is the classical one of Barendregt.

Notation. Throughout the chapter we need to manipulate nested [Application] terms, and so we urge a lighter notation. We continue following Krivine's [Kri93], so that nested [Application] like $(\ldots((M)\, N_1)\, N_2 \ldots)\, N_n$ are often written as $(M)\, N_1 \ldots N_n$ whenever there is no ambiguity. As a convention, a sequence of terms $N_1 \ldots N_n$ refers to the last application $(M)\, \langle\cdot\rangle$ when reading the whole term from left to right.

Definition 5.1.2. *The* Böhm tree of M*, denoted $BT(M)$, is the potentially infinite tree structure coinductively constructed as follows:*

- *if M is an* unsolvable*, then $BT(M) = \perp$;*
- *if M has* (principal) head normal form *$\lambda x_1 \ldots \lambda x_n.(y)\, M_1 \ldots M_m$ then $BT(M)$ is a tree with root $\lambda x_1 \ldots \lambda x_n.y$ and with $BT(M_1), \ldots, BT(M_m)$ as subtrees.*

BT's are the most popular tree structure in the λ-calculus, but they only correctly express the computational content of λ-terms in a *strong* regime, where, in particular, reduction can take place inside [Abstraction] terms. For instance, the terms $\lambda x.\mathbf{\Omega}$ and $\mathbf{\Omega}$, as both unsolvable, have identical BT's.

However, as we have seen in Chapter 4, in the *lazy* regime we would always distinguish between them. The former is a weak head normal form, hence a value, whereas the latter is not and never will. The tree structure capturing this difference is that of *Lévy-Longo trees* (briefly, LLT), which are the lazy variant of Böhm trees. LLT's were introduced by Longo [Lon83], developing an original idea by Lévy [Lév75]. The starting point is an inductive characterisation of β-normal forms by means of weak head reduction:

Definition 5.1.3. β-normal forms *are inductively defined as follows:*

- *$x \in \mathcal{V}$ is a β-normal form (and also a weak head normal form);*
- *a weak head normal form $\lambda x.M$ is a β-normal form if M also is;*
- *a weak head normal form $(x)\, M_1 \ldots M_m$, with $m \geq 1$, is a β-normal form if $M_1, \ldots, M_m$ also are.*

Again, the above definition directly provides a procedure to define the Lévy-Longo tree of a β-normal form. If read coinductively, Definition 5.1.3 entails a generalised, possibly infinite, "β-normal form" existing for every λ-term. However, LLT's are finer than BT's as they introduce a *decomposition* of unsolvability.

Definition 5.1.4. *A pure λ-term M is an* unsolvable of order $n \in \mathbb{N}$, *whenever n is the greatest $i \in \mathbb{N}$ such that $M \to_\beta^* \lambda x_1. \ldots \lambda x_i.N$, for some unsolvable N. If there is no such n, M is an* unsolvable of order ∞.

A term has an order of unsolvability n whether, after exhibiting n λ-abstractions, it behaves as $\mathbf{\Omega}$. Therefore, the LLT structure reflects this decomposition as follows.

Definition 5.1.5. *The* Lévy-Longo tree of *M, denoted $LT(M)$, is the potentially infinite tree structure coinductively constructed as follows:*

- *if M is an* unsolvable of order n, *then $LT(M) = \lambda x_1.\ldots \lambda x_n.\bot$;*
- *if M is an* unsolvable of order ∞, *then $LT(M) = \top$;*
- *if M has* (principal) head normal form $\lambda x_1.\ldots \lambda x_n.(y)\, N_1 \ldots N_m$ *then $LT(M)$ is a tree with root $\lambda x_1.\ldots \lambda x_n.y$ and with $LT(N_1), \ldots, LT(N_m)$ as subtrees.*

$$
LT(\mathbf{Y}) = \begin{array}{c} \lambda f.f \\ | \\ f \\ | \\ f \\ | \\ \vdots \end{array}
$$

Figure 5.1: Böhm tree of $\mathbf{Y}$.

On recursively solvable terms, BT's matches LLT's, and *vice versa*. A typical example is the infinite tree structure related to the fixpoint combinator $\mathbf{Y} = \lambda f.(\omega_f)\,\omega_f$, where $\omega_f = \lambda x.(f)\,(x)\,x$. Since $(\omega_f)\,\omega_f$ produces an infinite amount of f, the $BT(\mathbf{Y})$ is the infinite one in Figure 5.1. The same tree is also the LLT of $\mathbf{Y}$.

This is no longer valid when dealing with unsolvable terms, that is a $\perp$ node of Böhm trees may correspond to a non-$\perp$ node of Lévy-Longo trees. Consider, for instance, the *ogre* term Ξ defined as $(\lambda xy.(x)\,x)\,\lambda xy.(x)\,x$. This latter does not converge, but exhibits the unbound number of nested abstractions $\lambda y.\lambda y\ldots$, that is Ξ is an unsolvable of order ∞. It follows that $BT(\Xi) = \perp$, whereas $LT(\Xi) = \top$.

Definition 5.1.6. Lévy-Longo equality *is equality of Lévy-Longo trees:* i.e.

$$
M =_{\mathsf{LL}} N \textit{ if and only if } LT(M) = LT(N).
$$

In view of the previous discussion, if two pure λ-terms M, N have the same BT but $M \neq_{\mathsf{LL}} N$, then they may only differ in an unsolvable node.

***Example* 5.1.7.** Consider the terms

$$
\begin{aligned}
M &= \lambda x.((x)\,\lambda y.(x)\,\Xi\boldsymbol{\Omega} y)\,\Xi; \\
N &= \lambda x.((x)\,(x)\,\Xi\boldsymbol{\Omega})\,\Xi.
\end{aligned}
$$

These terms have been used to prove non-full-abstraction results in a canonical model for the lazy λ-calculus by Abramsky and Ong [AO93]. For this, they show that in the model the convergence test is definable (this operator, when it receives an argument, would return the identity function if the supplied argument is convergent, and would diverge otherwise). The convergence test, denoted ∇, can distinguish between the two terms, as $(M)\,\nabla$ reduces to an abstraction, whereas $(N)\,\nabla$ diverges. However, no pure λ-term can make the same distinction. The two terms also have different LLT's:

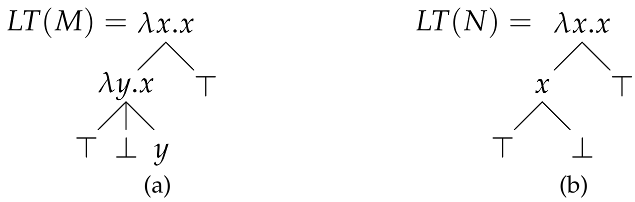


Figure 5.2: Lévy-Longo trees of M and N.

Although in $\Lambda_\oplus$, as in pure λ-calculus, the convergence test operator is not definable, M and N can be separated using probabilities by running them in a context C that would feed $\mathbf{\Omega} \oplus \lambda z.\lambda w.z$ as argument; then $C\langle M\rangle \Downarrow_{\frac{1}{2}}$ whereas $C\langle N\rangle \Downarrow_{\frac{1}{4}}$.

***Example* 5.1.8.** Abramsky's canonical model is itself coarser than LLT equality. Indeed, under the condition of (*may*-)convergence, the former validates η-conversion. For instance, the two η-distant terms $M = \lambda x.(x)\, x$ and $N = \lambda x.(x)\, \lambda y.(x)\, y$, have different LLT's but are equal in Abramsky's model (and hence equivalent for context equivalence in pure λ-calculus). They are separated by context equivalence in $\Lambda_\oplus$, for instance using the context $C = (\langle\cdot\rangle)\,(\mathbf{I} \oplus \mathbf{\Omega})$, since $C\langle M\rangle \Downarrow_{\frac{1}{4}}$ whereas $C\langle N\rangle \Downarrow_{\frac{1}{2}}$.

5.2 Pure λ-terms in probabilistic contexts

In this section we show that applicative bisimilarity and context equivalence coincides with Lévy-Longo equality *on pure λ-terms*. For this, as we already know that on full $\Lambda_\oplus$ applicative bisimilarity ($\sim$) implies context equivalence ($\simeq_\oplus$), it suffices to prove that, on pure terms, $\simeq_\oplus$ implies $=_{\mathsf{LL}}$, and that $=_{\mathsf{LL}}$ implies $\sim$.

5.2.1 Context equivalence implies Lévy-Longo equality

The first implication is obtained by a variation on the Böhm-out technique [Bar84], a powerful methodology for separation results in the pure λ-calculus, often employed in proofs about local structure characterisation theorems of λ-models. From the technical point of view, we revisit the method originally developed by Sangiorgi [San94] to characterise, using non-determinism, the discriminating power of Milner's [Mil92] π-calculus encoding of λ-calculus.

In particular, we first exploit an inductive characterisation of LLT equality via stratification approximants (Definition 5.2.7), and later we show in Lemma 5.2.9 that any difference on the trees of two λ-terms within level n can be observed by a suitable context of the probabilistic λ-calculus.

For this purpose, we do not need all the expressiveness of $\Lambda_{\oplus}$ here (Section 4.1), but only an approximation. Technically, we need to extend the set of pure λ-terms with the particular instance of probabilistic choice $\mathbf{\Omega} \oplus M$.

Notation. We write the specialised form of probabilistic choice $\mathbf{\Omega} \oplus M$ as $\uplus M$.

We briefly provide the set of terms $\Lambda^{\uplus}$ and its operational semantics. The related notions of set of free variables and term substitution can be defined as usual.

Definition 5.2.1. $\Lambda^{\uplus}$ *is the set of pure λ-terms extended with the $\uplus$ operator:*

- *if M is a pure λ-term, then $M \in \Lambda^{\uplus}$;*
- *if $M \in \Lambda^{\uplus}$, then $\uplus M \in \Lambda^{\uplus}$.*

On the operational side, $\Lambda^{\uplus}$ is endowed with the following CbN reduction relation $\longrightarrow \subseteq \Lambda^{\uplus} \times \Lambda^{\uplus}$.

Definition 5.2.2. Leftmost (weak) CbN reduction $\longrightarrow_p$, *with $p \in \left\{\frac{1}{2}, 1\right\}$, is the least binary relation on $\Lambda^{\uplus} \times \Lambda^{\uplus}$ such that (p is omitted whenever $p = 1$):*

- $(\lambda x.M)\, N \longrightarrow M\,[N/x]$;
- ($\uplus L$ rule) $\uplus M \longrightarrow_{\frac{1}{2}} \mathbf{\Omega}$;
- ($\uplus R$ rule) $\uplus M \longrightarrow_{\frac{1}{2}} M$;
- *if $M \longrightarrow_p L$, then $(M)\, N \longrightarrow_p (L)\, N$.*

Then $\longrightarrow^*$ is defined as usual, moreover, multiplying the corresponding probabilities.

Notation. We denote $\Rrightarrow_p$ a sequence $\longrightarrow_p^*$ in which only $\uplus R$, but not $\uplus L$, is applied.

We now recall the definition of *Böhm permutators*, which are the key technical ingredients of the Böhm-out technique.

Definition 5.2.3. *The* Böhm permutator of degree n, *with $n > 0$, is the pure λ-term*

$$\mathbf{P}_n = \lambda x_1. \ldots \lambda x_n.(x_n)\, x_1 x_2 \ldots x_{n-1}.$$

Here we consider a variant of Böhm permutators, the $\uplus$-permutators, which play a pivotal role in Lemma 5.2.9 below.

Definition 5.2.4. *A term $M \in \Lambda^{\uplus}$ is a* $\uplus$-permutator of degree n *if either $M = \mathbf{P}_n$, or there exists $0 \leq r < n$ such that*

$$M = \lambda x_1. \ldots \lambda x_r. \uplus \lambda x_{r+1}. \ldots \lambda x_n.(x_n)\, x_1 x_2 \ldots x_{n-1}.$$

*A function f from the positive integers to $\Lambda^{\uplus}$-terms is a $\uplus$-*permutator function *if, for all n, $f(n)$ is a $\uplus$-permutator of degree n.*

Recall that $\Lambda^{\uplus} \subset \Lambda_{\oplus}$. The following result establishes that context equivalence (Definition 4.5.5) is stable with respect to reduction $\Rrightarrow_p$. Notice that the proof put to use *applicative contexts*, which are contexts C either of the form $\langle\cdot\rangle$ or $(C)\,M$, with $M \in \Lambda^{\uplus}$. A formal definition presents no difficulties.

Lemma 5.2.5. *Let M, N, L, P be closed $\Lambda^{\uplus}$-terms, with $M \simeq_{\oplus} N$. If $M \Rrightarrow_p L$ and $N \Rrightarrow_p P$, then $L \simeq_{\oplus} P$.*

Proof. We prove the contrapositive, namely that $L \not\simeq_{\oplus} P$ and reductions $M \Rrightarrow_p L$, $N \Rrightarrow_p P$ imply $M \not\simeq_{\oplus} N$. Definition 4.5.5 entails C applicative context such that $C\langle L\rangle \Downarrow_r$ and $C\langle P\rangle \Downarrow_s$ imply $r \neq s$. Notice that we can limit ourselves to show the result only in the case applicative contexts because $\simeq_{\oplus} = \cong^{\mathsf{CIU}}$ (Corollary 4.5.33), and this latter is defined on applicative contexts only. Consider, then, $C\langle M\rangle$ and $C\langle N\rangle$. Since C is an applicative context and $\Rrightarrow_p$ is a left-to-right reduction sequence, both $C\langle M\rangle \Rrightarrow_p C\langle L\rangle$ and $C\langle N\rangle \Rrightarrow_p C\langle P\rangle$ hold. It follows $M \not\simeq_{\oplus} N$. $\square$

The proof of Lemma 5.2.9 below makes essential use of a characterisation of $=_{\mathsf{LL}}$ by a bisimulation-like form of relation:

Definition 5.2.6. *A relation $\mathcal{R}$ on pure λ-terms is an* open bisimulation *if $M \;\mathcal{R}\; N$ implies:*

1. *if $M \longrightarrow^* \lambda x.L$, then $N \longrightarrow^* \lambda x.P$ and $L \;\mathcal{R}\; P$;*
2. *if $M \longrightarrow^* (x)\,L_1 \ldots L_m$, then $P_1, \ldots, P_m$ exist such that $N \longrightarrow^* (x)\,P_1 \ldots P_m$ and, for all $i \in \{1, \ldots, m\}$, $L_i \;\mathcal{R}\; P_i$;*

and conversely on reductions from N. Open bisimilarity, *written $\sim^{O}$, is the union of all open bisimulations.*

Open bisimulation has the advantage of very easily providing a notion of approximation:

Definition 5.2.7. *The approximants of $\sim^{O}$ are defined as follows:*

- $\sim^{O}_{0} = \Lambda \times \Lambda$;
- $M \sim^{O}_{n+1} N$ *when*
    1. *if $M \longrightarrow^* \lambda x.L$, then P exists such that $N \longrightarrow^* \lambda x.P$ and $L \sim^{O}_{n} P$;*

2. *if* $M \longrightarrow^* (x)\, L_1 \dots L_m$*, then* $P_1, \dots, P_m$ *exist such that* $N \longrightarrow^* (x)\, P_1 \dots P_m$ *and, for all* $i \in \{1, \dots, m\}$, $L_i \sim_n^O P_i$*;*

and conversely on the reductions from N*.*

The following is a well-known fact [San94]:

Proposition 5.2.8. *On pure* λ*-terms, the relations* $=_{\mathsf{LL}}$, $\sim^O$ *and* $(\bigcap_{n\in\mathbb{N}} \sim_n^O)$ *all coincide.*

We are now ready to state and prove the Böhm-out lemma.

Lemma 5.2.9. *Suppose* $M \not\sim_n^O N$ *for some* n*, and let* $\{x_1, \dots, x_r\} \subseteq \mathsf{FV}(M) \cup \mathsf{FV}(N)$*. Then there are* $m_{x_1}, \dots, m_{x_r}, k \in \mathbb{N}$*, and permutator functions* $f_{x_1}, \dots, f_{x_r}$ *such that, for all* $m > k$*,*

$$\begin{array}{c} M\,[f_{x_1}(m+m_{x_1})/x_1] \dots [f_{x_r}(m+m_{x_r})/x_r] \\ \not\simeq_{\oplus} \\ N\,[f_{x_1}(m+m_{x_1})/x_1] \dots [f_{x_r}(m+m_{x_r})/x_r]. \end{array}$$

Proof. We prove the result by exhibiting an applicative context C such that,

- $C\,\langle M\,[f_{x_1}(m+m_{x_1})/x_1] \dots [f_{x_r}(m+m_{x_r})/x_r]\rangle \Downarrow_p$ and
- $C\,\langle N\,[f_{x_1}(m+m_{x_1})/x_1] \dots [f_{x_r}(m+m_{x_r})/x_r]\rangle \Downarrow_q$,

imply $p \neq q$. Let us first set some notations for the sake of readability of this proof. Given M term, $M^{\overline{f}}$ stands for $M\,[f_{x_1}(m+m_{x_1})/x_1] \dots [f_{x_r}(m+m_{x_r})/x_r]$ where $x_1, \dots, x_r \in \mathsf{FV}(M)$. Moreover, $\mathbf{\Omega}^n$ stands for the sequence of n arguments $\mathbf{\Omega}$: for instance, $(M)\,\mathbf{\Omega}^3$ is $(M)\,\mathbf{\Omega\Omega\Omega}$. Finally, we write $M\Uparrow$ whenever M diverges.

The proof proceeds by induction on the least n such that $M \not\sim_n^O N$. Being symmetric, we need to consider only one direction of each clause of $\sim_n^O$.

- The base case $M \not\sim_1^O N$ needs us to address few cases:
  - The case where only one of the two terms diverges is simple.
  - $M \longrightarrow^* (x)\, L_1 \dots L_t$ and $N \longrightarrow^* (x)\, P_1 \dots P_s$ with $t < s$. Consider $m_x = s$ and $f_x(n) = \mathbf{P}_n$. The values of the other integers (k, m_y for every $y \neq x$) and of the other permutation functions are irrelevant. Set $C = (\langle\cdot\rangle)\,\mathbf{\Omega}^m$. It follows
  $$C\left\langle M^{\overline{f}}\right\rangle = \left(M^{\overline{f}}\right)\mathbf{\Omega}^m \longrightarrow^* (\mathbf{P}_{m+s})\, L_1^{\overline{f}} \dots L_t^{\overline{f}}\mathbf{\Omega}^m \Downarrow_1$$
  since $t + m < s + m$. Moreover,
  $$C\left\langle N^{\overline{f}}\right\rangle = \left(N^{\overline{f}}\right)\mathbf{\Omega}^m \longrightarrow^* (\mathbf{P}_{m+s})\, P_1^{\overline{f}} \dots P_s^{\overline{f}}\mathbf{\Omega}^m \Uparrow$$
  since $m > 0$, and so an $\mathbf{\Omega}$ ends up at the head of the term.

– $M \longrightarrow^* (x)\, L_1 \ldots L_t$ and $M \longrightarrow^* (y)\, P_1 \ldots P_s$ with $x \neq y$. Assume $t \leq s$ without loss of generality. Consider $m_x = s+1$, $m_y = s$, and $f_x(n) = f_y(n) = \mathbf{P}_n$. The values of the other integers and permutation functions are irrelevant. Set $C = (\langle\cdot\rangle)\, \mathbf{\Omega}^m$. It follows

$$C \left\langle M^{\overline{f}} \right\rangle = \left(M^{\overline{f}} \right) \mathbf{\Omega}^m \longrightarrow^* (\mathbf{P}_{m+s+1})\, L_1^{\overline{f}} \ldots L_t^{\overline{f}} \mathbf{\Omega}^m \Downarrow_1$$

since $m+s+1 > t+m$. Moreover,

$$C \left\langle N^{\overline{f}} \right\rangle = \left(N^{\overline{f}} \right) \mathbf{\Omega}^m \longrightarrow^* (\mathbf{P}_{m+s})\, P_1^{\overline{f}} \ldots P_s^{\overline{f}} \mathbf{\Omega}^m \Uparrow$$

since $m > 0$, and so an $\mathbf{\Omega}$ ends up at the head of the term.

– $M \longrightarrow^* \lambda x.L$ and $N \longrightarrow^* (y)\, P_1 \ldots P_s$, for some y and $s \geq 0$. The values of the integers and permutator functions are irrelevant. Set $C = \langle\cdot\rangle$, and $f_y(n) = \uplus\mathbf{P}_n$. It follows $C \left\langle M^{\overline{f}} \right\rangle = M^{\overline{f}} \longrightarrow^* \lambda x.L^{\overline{f}} \Downarrow_1$, whereas

$$C \left\langle N^{\overline{f}} \right\rangle = N^{\overline{f}} \longrightarrow^* \left(\uplus\mathbf{P}_{m+m_y} \right) P_1^{\overline{f}} \ldots P_s^{\overline{f}} \Downarrow_{<1}.$$

The above cases entails $M^{\overline{f}} \not\simeq_{\oplus} N^{\overline{f}}$ by the contrapositive of Lemma 5.2.5.

- The inductive case $M \not\sim^O_{n+1} N$ needs us to address the following two cases:

– $M \longrightarrow^* (x)\, L_1 \ldots L_s$, $N \longrightarrow^* (x)\, P_1 \ldots P_s$ and $L_i \not\sim^O_n P_i$ for some $i \in \{1, \ldots, s\}$. By the induction hypothesis, there are integers m_z, k and permutator functions f_z such that, for all $m > k$, there is an applicative context D such that $D \left\langle L_i^{\overline{f}} \right\rangle \Downarrow_v$ and $D \left\langle P_i^{\overline{f}} \right\rangle \Downarrow_w$ imply $v \neq w$. Redefine k, if necessary, in order to obtain $k > s$. Set $C = D \left\langle (\langle\cdot\rangle)\, \mathbf{\Omega}^{m+m_x-s-1}(\lambda x_1 \ldots x_{m+m_x}.x_i) \right\rangle$. It follows

$$\begin{aligned} C \left\langle M^{\overline{f}} \right\rangle &= D \left\langle \left(M^{\overline{f}} \right) \mathbf{\Omega}^{m+m_x-s-1}(\lambda x_1 \ldots x_{m+m_x}.x_i) \right\rangle \\ &\longrightarrow^* D \left\langle (f_x(m+m_x))\, L_1^{\overline{f}} \ldots L_s^{\overline{f}} \mathbf{\Omega}^{m+m_x-s-1}(\lambda x_1 \ldots . x_{m+m_x}.x_i) \right\rangle \\ &\longrightarrow_p^* D \left\langle L_i^{\overline{f}} \right\rangle, \end{aligned}$$

whereas

$$\begin{aligned} C \left\langle N^{\overline{f}} \right\rangle &= D \left\langle \left(N^{\overline{f}} \right) \mathbf{\Omega}^{m+m_x-s-1}(\lambda x_1 \ldots x_{m+m_x}.x_i) \right\rangle \\ &\longrightarrow^* D \left\langle (f_x(m+m_x))\, P_1^{\overline{f}} \ldots P_s^{\overline{f}} \mathbf{\Omega}^{m+m_x-s-1}(\lambda x_1 \ldots . x_{m+m_x}.x_i) \right\rangle \\ &\longrightarrow_p^* D \left\langle P_i^{\overline{f}} \right\rangle, \end{aligned}$$

where p is $\frac{1}{2}$ or 1 depending on whether f_x contains $\uplus$ or not. In any case, in both derivations, rule $\uplus L$ has not been used. The contrapositive of Lemma 5.2.5, on the inductive result, entails $M^{\overline{f}} \not\precsim_{\oplus} N^{\overline{f}}$.

– $M \longrightarrow^* \lambda x.L$, $N \longrightarrow^* \lambda x.P$ and $L \not\precsim_n^O P$. By the induction hypothesis, there are integers m_z, k and permutator functions f_z such that, for all $m > k$, there is an applicative context D such that $D\left\langle L^{\overline{f}}\right\rangle \Downarrow_v$ and $D\left\langle P^{\overline{f}}\right\rangle \Downarrow_w$ imply $v \neq w$. Set $C = D\left\langle (\langle\cdot\rangle)\, f_x(m+m_x)\right\rangle$. Below, given Q term, $Q^{\overline{f\backslash x}}$ is defined as $Q^{\overline{f}}$ except that variable x is left uninstantiated. It follows

$$
\begin{aligned}
C\left\langle M^{\overline{f}}\right\rangle &= D\left\langle \left(M^{\overline{f}}\right) f_x(m+m_x)\right\rangle \\
&\longrightarrow^* D\left\langle \left(\lambda x.L^{\overline{f\backslash x}}\right) f_x(m+m_x)\right\rangle \\
&\longrightarrow D\left\langle L^{\overline{f\backslash x}}\left[f_x(m+m_x)/x\right]\right\rangle = D\left\langle L^{\overline{f}}\right\rangle,
\end{aligned}
$$

whereas

$$
\begin{aligned}
C\left\langle N^{\overline{f}}\right\rangle &= D\left\langle \left(N^{\overline{f}}\right) f_x(m+m_x)\right\rangle \\
&\longrightarrow^* D\left\langle \left(\lambda x.P^{\overline{f\backslash x}}\right) f_x(m+m_x)\right\rangle \\
&\longrightarrow D\left\langle P^{\overline{f\backslash x}}\left[f_x(m+m_x)/x\right]\right\rangle = D\left\langle P^{\overline{f}}\right\rangle.
\end{aligned}
$$

Once again, the contrapositive of Lemma 5.2.5, on the inductive result, entails $M^{\overline{f}} \not\precsim_{\oplus} N^{\overline{f}}$.

This concludes the proof. □

The fact the Böhm-out technique actually works implies that the discriminating power of probabilistic contexts is at least as strong as the one of LLT's.

Corollary 5.2.10. *For all M, N pure λ-terms, $M \simeq_{\oplus} N$ implies $M =_{\mathsf{LL}} N$.*

5.2.2 Lévy-Longo equality implies Applicative bisimilarity

To show that LLT equality is included in probabilistic applicative bisimilarity, we proceed as follows. First we define a refinement of the latter, essentially one in which we *observe* all probabilistic choices. As a consequence, the underlying bisimulation game may ignore probabilities. Then we show that the obtained notion of equivalence is strictly finer than probabilistic applicative bisimilarity.

The advantage of the refinement is that *both* the inclusion of LLT equality in the refinement, and the inclusion of the latter in probabilistic applicative bisimilarity turn out to be relatively easy to prove. A *direct* proof of the inclusion of LLT equality in probabilistic applicative bisimilarity would have been harder, as it would have required extending the notion of a Lévy-Longo tree to $\Lambda_\oplus$, then reasoning on substitution closures of such trees.

Lévy-Longo equality implies Strict applicative bisimilarity

We start by redefining reduction relation $\longrightarrow$ on *all* $\Lambda_\oplus$-terms.

Definition 5.2.11. Leftmost (weak) CbN reduction $\longrightarrow_p$*, with* $p \in \left\{\frac{1}{2}, 1\right\}$*, is the least binary relation on* $\Lambda_\oplus \times \Lambda_\oplus$ *such that (*p *is omitted whenever* $p = 1$*):*

- $(\lambda x.M)\, N \longrightarrow M\, [N/x]$;
- $M \oplus N \longrightarrow_{\frac{1}{2}} M$;
- $M \oplus N \longrightarrow_{\frac{1}{2}} N$;
- *if* $M \longrightarrow_p L$*, then* $(M)\, N \longrightarrow_p (L)\, N$.

Then $\longrightarrow^*$ *is defined as usual, moreover, multiplying the corresponding probabilities.*

We now prove that LLT equality is encompassed by a notion of *strict* applicative bisimilarity. Observe the more usual nature of this bisimulation definition.

Definition 5.2.12. *A relation* $\mathcal{R} \subseteq \Lambda_\oplus(\emptyset) \times \Lambda_\oplus(\emptyset)$ *is a* strict applicative bisimulation *whenever* $M\ \mathcal{R}\ N$ *implies*

1. *if* $M \longrightarrow P$*, then* $N \longrightarrow^* Q$ *and* $P\ \mathcal{R}\ Q$*;*
2. *if* $M \longrightarrow_{\frac{1}{2}} P$*, then* $N \longrightarrow^*_{\frac{1}{2}} Q$ *and* $P\ \mathcal{R}\ Q$*;*
3. *if* $M = \lambda x.P$*, then* $N \longrightarrow^* \lambda x.Q$ *and* $P\, [L/x]\ \mathcal{R}\ Q\, [L/x]$ *for all* $L \in \Lambda_\oplus(\emptyset)$*;*
4. *the converse of 1, 2 and 3.*

Strict applicative bisimilarity $\sim$ *is the union of all strict applicative bisimulations.*

***Remark* 5.2.13.** This notion of strict applicative bisimulation is extended to *open terms* as usual, that is by considering the closure under term-substitutions. Formally, given $\mathcal{R}$ strict applicative bisimulation, we write $\overline{x} \vdash M\ \mathcal{R}\ N$ if $\emptyset \vdash M\left[\overline{L}/\overline{x}\right]\ \mathcal{R}\ N\left[\overline{L}/\overline{x}\right]$ holds for all $\overline{L} \in \mathcal{P}_{\mathsf{FIN}}(\Lambda_\oplus(\emptyset))$.

$$\frac{\emptyset \vdash \overline{A} \mathrel{\Diamond} \overline{B} \qquad M \in \Lambda_{\oplus}(\overline{x})}{\emptyset \vdash M\left[\overline{A}/\overline{x}\right] \mathrel{\Diamond} M\left[\overline{B}/\overline{x}\right]}\ (\Diamond\mathsf{var})$$

$$\frac{\emptyset \vdash \overline{A} \mathrel{\Diamond} \overline{B} \qquad \overline{x} \vdash M =_{\mathsf{LL}} N \qquad M \notin \mathcal{V} \text{ or } N \notin \mathcal{V}}{\emptyset \vdash M\left[\overline{A}/\overline{x}\right] \mathrel{\Diamond} N\left[\overline{B}/\overline{x}\right]}\ (\Diamond\mathsf{term})$$

Figure 5.3: $\Lambda_{\oplus}$-relation $\Diamond$ rules.

The crucial point is that, if two pure λ-terms have the same LLT, then passing them the same argument $M \in \Lambda_{\oplus}$ produces *exactly* the same [Choice] structure: intuitively, whenever the first term finds (a copy of) M in head position, also the second will find M.

Below we show that the closure of terms whose LLT is equal is included in a $\Lambda_{\oplus}$-relation $\Diamond$, which we later prove to be a strict applicative bisimulation.

Definition 5.2.14. *$\Diamond$ is the $\Lambda_{\oplus}$-relation inductively defined by the rules in Figure 5.3.*

Notice that $\Diamond$ is defined on closed $\Lambda_{\oplus}$-terms only. However, it is immediate to see that the closure under term-substitution (Remark 5.2.13) of terms whose LLT is equal is encompassed by the $\Lambda_{\oplus}$-relation $\Diamond$.

We only need to show that $\Diamond$ is a strict applicative bisimulation, namely that it validates the clauses of Definition 5.2.12. As customary in bisimulation proofs, we check every clause only once (*i.e.* one direction, as the other is analogous). Moreover, since $\Diamond$ is defined on close terms only, we often omit the heavy notation for $\Lambda_{\oplus}$-relations.

This first lemma is obvious by the definition of $\Diamond$:

Lemma 5.2.15. *$\Lambda_{\oplus}$-relation $\Diamond$ is reflexive and symmetric.*

The following result shows that $\Diamond$ enjoys the property expressed by the 3rd clause of Definition 5.2.12.

Lemma 5.2.16. *If $\lambda x.M \mathrel{\Diamond} N$ and $A \mathrel{\Diamond} B$, then there is $L \in \Lambda_{\oplus}(x)$ such that $N \longrightarrow^{*} \lambda x.L$ and $M[A/x] \mathrel{\Diamond} L[B/x]$.*

Proof. By induction on the proof of $\lambda x.M \mathrel{\Diamond} N$, and case analysis on last rule used.

- Rule $(\Diamond\mathsf{var})$ entails $\lambda x.M = P\left[\overline{C}/\overline{y}\right]$, $N = P\left[\overline{D}/\overline{y}\right]$ with the additional hypothesis $\overline{C} \mathrel{\Diamond} \overline{D}$ and $P \in \Lambda_{\oplus}(\overline{y})$. Then:

– If $P \in \mathcal{V}$, then $P = y_i$ with $y_i \in \overline{y}$. Hence $\lambda x.M = P\left[\overline{C}/\overline{y}\right] = C_i$, $P\left[\overline{D}/\overline{y}\right] = D_i$ and $C_i \diamondsuit D_i$. It follows $\lambda x.M \diamondsuit D_i$. By the induction hypothesis, $D_i \longrightarrow^* \lambda x.R$ and $M[A/x] \diamondsuit R[B/x]$. Consider $L = R$ and verify that $N \longrightarrow^* \lambda x.L$ and $M[A/x] \diamondsuit L[B/x]$.

– If $P = \lambda x.R$, then $\lambda x.M = P\left[\overline{C}/\overline{y}\right] = \lambda x.R\left[\overline{C}/\overline{y}\right]$ and $N = P\left[\overline{D}/\overline{y}\right] = \lambda x.R\left[\overline{D}/\overline{y}\right]$. It follows $R\left[\overline{C}[A/x], A/\overline{y}, x\right] \diamondsuit R\left[\overline{D}[B/x], B/\overline{y}, x\right]$ by rule ($\diamondsuit$var), that is $R\left[\overline{C}/\overline{y}\right][A/x] \diamondsuit R\left[\overline{D}/\overline{y}\right][B/x]$. Consider $L = R\left[\overline{D}/\overline{y}\right]$ and verify that $M[A/x] \diamondsuit L[B/x]$.

- Rule ($\diamondsuit$term) entails $\lambda x.M = P\left[\overline{C}/\overline{y}\right]$, $N = Q\left[\overline{D}/\overline{y}\right]$ with the additional hypothesis $\overline{y} \vdash P =_{\mathsf{LL}} Q$ and $\overline{C} \diamondsuit \overline{D}$. Then:

  – If $P = y_i$, with $y_i \in \overline{y}$, then $Q \longrightarrow^* y_i$ due to the fact that $P =_{\mathsf{LL}} Q$. Hence $P\left[\overline{C}/\overline{y}\right] = C_i$ and $Q\left[\overline{D}/\overline{y}\right] = D_i$. Since $\lambda x.M = P\left[\overline{C}/\overline{y}\right] = C_i$, $N = Q\left[\overline{D}/\overline{y}\right] = D_i$ and $C_i \diamondsuit D_i$, it follows $\lambda x.M \diamondsuit D_i$. By the induction hypothesis, $D_i \longrightarrow^* \lambda x.R$ and $M[A/x] \diamondsuit R[B/x]$. Consider $L = R$ and verify that $N \longrightarrow^* \lambda x.L$ and $M[A/x] \diamondsuit L[B/x]$.

  – If $P = \lambda x.R$, then $Q \longrightarrow^* \lambda x.S$ due to the fact that $P =_{\mathsf{LL}} Q$. Hence $\lambda x.M = \lambda x.R\left[\overline{C}/\overline{y}\right]$, $N \longrightarrow^* \lambda x.S\left[\overline{D}/\overline{y}\right]$ and $\lambda x.R =_{\mathsf{LL}} \lambda x.S$. Since $\lambda x.R =_{\mathsf{LL}} \lambda x.S$ entails $R =_{\mathsf{LL}} S$, it follows $R\left[\overline{C}[A/x], A/\overline{y}, x\right] \diamondsuit S\left[\overline{D}[A/x], B/\overline{y}, x\right]$ by rule ($\diamondsuit$term), that is $R\left[\overline{C}/\overline{y}\right][A/x] \diamondsuit S\left[\overline{D}/\overline{y}\right][B/x]$. Consider $L = S\left[\overline{D}/\overline{y}\right]$ and verify that $M[A/x] \diamondsuit L[B/x]$.

This concludes the proof. □

Lemma 5.2.17. *Suppose $M_i \diamondsuit N_i$, for every $i \in \{1, \ldots, n\}$ with $n \geq 1$. Then $(M_1)\, M_2 \ldots M_n \diamondsuit (N_1)\, N_2 \ldots N_n$.*

Proof. The proof simply inspects the last rule used in $M_i \diamondsuit N_i$ derivations. The result directly follows when $n = 1$. Hence, we detail the case of $n = 2$.

There are three main cases:

- Rule ($\diamondsuit$var) used for both $(M_i \diamondsuit N_i)_{i\{1,2\}}$. It follows $(M_i = L_i\left[\overline{A_i}/\overline{x_i}\right])$ and $(N_i = L_i\left[\overline{B_i}/\overline{x_i}\right])$, with $L_i \in \Lambda_{\oplus}(\overline{x_i})$ and $\overline{x}_1 \cap \overline{x}_2 = \emptyset$. Consider $(L_1)\, L_2 \in \Lambda_{\oplus}(\overline{x}_1 \cup \overline{x}_2)$ and conclude $((L_1)\, L_2)\left[\overline{A}_1, \overline{A}_2/\overline{x}_1, \overline{x}_2\right] \diamondsuit ((L_1)\, L_2)\left[\overline{B}_1, \overline{B}_2/\overline{x}_1, \overline{x}_2\right]$ by rule ($\diamondsuit$var). The latter is exactly $(M_1)\, M_2 \diamondsuit (N_1)\, N_2$.

- Rule ($\diamondsuit$term) used for both $(M_i \diamondsuit N_i)_{i\{1,2\}}$. It follows $(M_i = L_i\left[\overline{A_i}/\overline{x_i}\right])$ and $(N_i = P_i\left[\overline{B_i}/\overline{x_i}\right])$, with $L_i =_{\mathsf{LL}} P_i$ and $L_i, P_i \in \Lambda_{\oplus}(\overline{x_i})$, $\overline{x}_1 \cap \overline{x}_2 = \emptyset$. Consider $(L_1)\, L_2, (P_1)\, P_2 \in \Lambda_{\oplus}(\overline{x} \cup \overline{y})$ and verify that $(L_1)\, L_2 =_{\mathsf{LL}} (P_1)\, P_2$ ($=_{\mathsf{LL}}$ is an equality, hence a congruence). It follows $(L_1)\, L_2\left[\overline{A}_1, \overline{A}_2/\overline{x}_1, \overline{x}_2\right] \diamondsuit (P_1)\, P_2\left[\overline{B}_1, \overline{B}_2/\overline{x}_1, \overline{x}_2\right]$ by rule ($\diamondsuit$term), which is exactly $(M_1)\, M_2 \diamondsuit (N_1)\, N_2$.

- the two cases in which rules $(\diamondsuit\mathsf{var})$ and $(\diamondsuit\mathsf{term})$ are interleaved follow by exploiting the fact that $=_{\mathsf{LL}}$ is a congruence relation.

The generalisation of this proof to any $n > 2$ is obvious. $\square$

The following lemma shows that $\diamondsuit$ enjoys the properties expressed by the 1st and the 2nd clauses of Definition 5.2.12.

Lemma 5.2.18. *Suppose $M \diamondsuit N$. If $M \longrightarrow_p P$, then $N \longrightarrow_p^* Q$ and $P \diamondsuit Q$.*

Proof. By induction on the proof of $M \diamondsuit N$, and case analysis on last rule used.

- Rule $(\diamondsuit\mathsf{var})$ entails $M = L\left[\overline{A}/\overline{y}\right]$, $N = L\left[\overline{B}/\overline{y}\right]$ with the additional hypothesis $\overline{A} \diamondsuit \overline{B}$ and $L \in \Lambda_{\oplus}(\overline{y})$. Then:
  - If $L \in \mathcal{V}$, then $L = y_i$ with $y_i \in \overline{y}$. Hence $M = L\left[\overline{A}/\overline{y}\right] = A_i$ and $A_i \longrightarrow_p P$. Moreover, $N = L\left[\overline{B}/\overline{y}\right] = B_i$ and $A_i \diamondsuit B_i$. By the induction hypothesis, $B_i \longrightarrow_p^* Q$ and $P \diamondsuit Q$.
  - If $L = (R)\,S$ and $R \in \mathcal{V}$. Hence $R = y_i$ with $y_i \in \overline{y}$, and $M = L\left[\overline{A}/\overline{y}\right] = (A_i)\,S\left[\overline{A}/\overline{y}\right]$. The interesting case is $A_i = \lambda x.T$ and $M \longrightarrow T\left[S\left[\overline{A}/\overline{y}\right]/x\right] = P$ (the other is subsumed by the next case). Notice that $N = L\left[\overline{B}/\overline{y}\right] = (B_i)\,S\left[\overline{B}/\overline{y}\right]$. Since $A_i \diamondsuit B_i$ and $S\left[\overline{A}/\overline{y}\right] \diamondsuit S\left[\overline{B}/\overline{y}\right]$ (rule $(\diamondsuit\mathsf{var})$), Lemma 5.2.16 entails $U \in \Lambda_{\oplus}(x)$ such that $B_i \longrightarrow^* \lambda x.U$ and $T\left[S\left[\overline{A}/\overline{y}\right]/x\right] \diamondsuit U\left[S\left[\overline{B}/\overline{y}\right]/x\right]$. Consider $Q = U\left[S\left[\overline{B}/\overline{y}\right]/x\right]$ and verify that $N \longrightarrow^* Q$ and $P \diamondsuit Q$.
  - If $L = (R)\,S$ and $R\left[\overline{A}/\overline{y}\right] \longrightarrow_p T$. Hence $M = L\left[\overline{A}/\overline{y}\right] \longrightarrow_p (T)\,S\left[\overline{A}/\overline{y}\right] = P$. Since $R\left[\overline{A}/\overline{y}\right] \diamondsuit R\left[\overline{B}/\overline{y}\right]$ (rule $(\diamondsuit\mathsf{var})$), by the induction hypothesis follows $R\left[\overline{B}/\overline{y}\right] \longrightarrow_p^* U$ and $T \diamondsuit U$. Notice that $N = \left(R\left[\overline{B}/\overline{y}\right]\right)S\left[\overline{B}/\overline{y}\right] \longrightarrow_p^* (U)\,S\left[\overline{B}/\overline{y}\right] = Q$ and $S\left[\overline{A}/\overline{y}\right] \diamondsuit S\left[\overline{B}/\overline{y}\right]$ (rule $(\diamondsuit\mathsf{var})$), so Lemma 5.2.17 entails $(T)\,S\left[\overline{A}/\overline{y}\right] \diamondsuit (U)\,S\left[\overline{B}/\overline{y}\right]$, that is $P \diamondsuit Q$.
  - If $L = (\lambda x.R)\,S$, then $M = L\left[\overline{A}/\overline{y}\right] \longrightarrow \left(R\left[\overline{A}/\overline{y}\right]\right)\left[S\left[\overline{A}/\overline{y}\right]/x\right] = (R\,[S/x])\left[\overline{A}/\overline{y}\right] = P$. Similarly, $N = L\left[\overline{B}/\overline{y}\right] \longrightarrow (R\,[S/x])\left[\overline{B}/\overline{y}\right] = Q$. Rule $(\diamondsuit\mathsf{var})$ entails $P \diamondsuit Q$.
  - If $L = R \oplus S$ and $R\left[\overline{A}/\overline{y}\right] \longrightarrow_p T$, then $M = L\left[\overline{A}/\overline{y}\right] \longrightarrow_{\frac{p}{2}} T \oplus S\left[\overline{A}/\overline{y}\right] = P$. Notice that $N = L\left[\overline{B}/\overline{y}\right] = R\left[\overline{B}/\overline{y}\right] \oplus S\left[\overline{B}/\overline{y}\right]$, with $R\left[\overline{A}/\overline{y}\right] \diamondsuit R\left[\overline{B}/\overline{y}\right]$ and $S\left[\overline{A}/\overline{y}\right] \diamondsuit S\left[\overline{B}/\overline{y}\right]$ (rule $(\diamondsuit\mathsf{var})$). By the induction hypothesis, $R\left[\overline{B}/\overline{y}\right] \longrightarrow_p^* U$ and $T \diamondsuit U$. Then, $N = R\left[\overline{B}/\overline{y}\right] \oplus S\left[\overline{B}/\overline{y}\right] \longrightarrow_{\frac{p}{2}}^* U \oplus S\left[\overline{B}/\overline{y}\right] = Q$ and $P \diamondsuit Q$. The case in which $S\left[\overline{A}/\overline{y}\right]$ reduces follows analogously.

- Rule ($\diamondsuit$term) entails $M = L\left[\overline{A}/\overline{y}\right]$, $N = R\left[\overline{B}/\overline{y}\right]$ with the additional hypothesis $\overline{y} \vdash L =_{\mathsf{LL}} R$ and $\overline{A} \diamondsuit \overline{B}$. Then:
  - If $L = (\lambda x.S)\, T\overline{L}$, then $M = L\left[\overline{A}/\overline{y}\right] \longrightarrow \left(\left(S\left[\overline{A}/\overline{y}\right]\right)\left[T\left[\overline{A}/\overline{y}\right]/x\right]\right)\overline{L}\left[\overline{A}/\overline{y}\right] = \left(S\left[T/x\right]\right)\overline{L}\left[\overline{A}/\overline{y}\right] = P$. Since $\left(S\left[T/x\right]\right)\overline{L} =_{\mathsf{LL}} L =_{\mathsf{LL}} R$, rule ($\diamondsuit$term) entails $\left(S\left[T/x\right]\right)\overline{L}\left[\overline{A}/\overline{y}\right] \diamondsuit R\left[\overline{B}/\overline{y}\right]$. Consider $Q = R\left[\overline{B}/\overline{y}\right]$ and verify that $P \diamondsuit Q$. Notice that this case encompasses the more general case of $L = (S)\, T$ with S an [Application] term.
  - If $L = (y_i)\,\overline{L}$, with $y_i \in \overline{y}$ (otherwise, we would have no reduction), then $L\left[\overline{A}/\overline{y}\right] = (A_i)\,\overline{L}\left[\overline{A}/\overline{y}\right]$. Since $L =_{\mathsf{LL}} R$, it follows $R \longrightarrow^* (y_i)\,\overline{P}$ and $\overline{L} =_{\mathsf{LL}} \overline{P}$, so that $R\left[\overline{B}/\overline{y}\right] \longrightarrow^* (B_i)\,\overline{P}\left[\overline{B}/\overline{y}\right]$. There are two cases to distinguish:
    * If $A_i \longrightarrow_p C$, then $P = (C)\,\overline{L}\left[\overline{A}/\overline{y}\right]$. By the induction hypothesis, $B_i \longrightarrow_p^* D$ and $C \diamondsuit D$. Consider $Q = (D)\,\overline{P}\left[\overline{B}/\overline{y}\right]$ and, since $\overline{L}\left[\overline{A}/\overline{y}\right] \diamondsuit \overline{P}\left[\overline{B}/\overline{y}\right]$, Lemma 5.2.17 entails $P \diamondsuit Q$.
    * If $A_i = \lambda x.C$, then $\overline{L} = T\overline{L}'$ (otherwise, we would have no reduction) and $\overline{P} = U\overline{P}'$ with $T =_{\mathsf{LL}} U$ and $\overline{L}' =_{\mathsf{LL}} \overline{P}'$. It follows $(A_i)\,\overline{L}\left[\overline{A}/\overline{y}\right] = \left((\lambda x.C)\, T\left[\overline{A}/\overline{y}\right]\right)\overline{L}'\left[\overline{A}/\overline{y}\right] \longrightarrow \left(C\left[T\left[\overline{A}/\overline{y}\right]/x\right]\right)\overline{L}'\left[\overline{A}/\overline{y}\right] = P$. Since $T\left[\overline{A}/\overline{y}\right] \diamondsuit U\left[\overline{B}/\overline{y}\right]$ (rule ($\diamondsuit$term)), Lemma 5.2.16 entails $D \in \Lambda_{\oplus}(x)$ such that $B_i \longrightarrow^* \lambda x.D$ and $C\left[T\left[\overline{A}/\overline{y}\right]/x\right] \diamondsuit D\left[U\left[\overline{B}/\overline{y}\right]/x\right]$. Consider $Q = \left(D\left[U\left[\overline{B}/\overline{y}\right]/x\right]\right)\overline{P}'\left[\overline{B}/\overline{y}\right]$ and, since $\overline{L}'\left[\overline{A}/\overline{y}\right] \diamondsuit \overline{P}'\left[\overline{B}/\overline{y}\right]$, Lemma 5.2.17 entails $P \diamondsuit Q$.

This concludes the proof. $\square$

Lemma 5.2.18 and Lemma 5.2.16 together imply the intended result.

Corollary 5.2.19. *The set of pairs of closed $\Lambda_{\oplus}$-terms (M, N) with $M \diamondsuit N$ is a strict applicative bisimulation.*

The above corollary justifies the following result:

Lemma 5.2.20. *For all M, N pure λ-terms, if $\overline{x} \vdash M =_{\mathsf{LL}} N$ then $\overline{x} \vdash M \diamondsuit N$.*

Proof. Immediate from Corollary 5.2.19 and the way strict applicative bisimulations $\mathcal{R}$ are extended to open terms. In particular, the hypothesis $\overline{x} \vdash M =_{\mathsf{LL}} N$ directly entails $\emptyset \vdash \lambda\overline{x}.M =_{\mathsf{LL}} \lambda\overline{x}.N$ (indeed, LLT equality is a congruence on pure λ-terms). Then, Corollary 5.2.19 implies $\emptyset \vdash \lambda\overline{x}.M \diamondsuit \lambda\overline{x}.N$, which is equivalent to $\emptyset \vdash M\left[\overline{L}/\overline{x}\right] \diamondsuit N\left[\overline{L}/\overline{x}\right]$ for all $\overline{L} \in \mathcal{P}_{\mathsf{FIN}}(\Lambda_{\oplus}\emptyset)$ (3rd clause of Definition 5.2.12). This latter is, by definition of closure under term-substitution, $\overline{x} \vdash M \diamondsuit N$. $\square$

Terms which are strict applicative bisimilar cannot be distinguished by applicative bisimilarity proper, since the requirements induced by the latter are less strict than the ones the former imposes. To show this formally, we first need to show that $\dot{\sim}$ is an equivalence relation.

Lemma 5.2.21. *On $\Lambda_\oplus$-terms, strict applicative bisimilarity is an equivalence relation.*

Proof. Reflexivity and symmetry of $\dot{\sim}$ are obvious. We now detail the transitive property of $\dot{\sim}$. In particular, given $\mathcal{R}, \mathcal{T}$ strict applicative bisimulation, we show that $\mathcal{R} \circ \mathcal{T} = \{(M, N) \mid \exists L.\, M \;\mathcal{R}\; L \;\wedge\; L \;\mathcal{T}\; N\}$ is a strict applicative bisimulation. We check that the clauses of Definition 5.2.12 hold only in the case M reduces. The other way around, when N reduces, is analogous.

Suppose $M \longrightarrow_p P$. Since $\mathcal{R}$ is a strict applicative bisimulation, $L \longrightarrow_p^* Q$ and $P \;\mathcal{R}\; Q$. In details, there are $L_1, \ldots, L_n \in \Lambda_\oplus$ and $q_1, \ldots, q_n \in \mathbb{R}_{[0,1]}$ such that

- $L \longrightarrow_{q_1} L_1$,
- $L_i \longrightarrow_{q_{i+1}} L_{i+1}$ with $i \in \{1, \ldots, n-1\}$,
- $L_n \longrightarrow_{q_n} Q$,

with $p = q_1 \cdot q_2 \cdots q_n$. Since $\mathcal{T}$ is a strict applicative bisimulation, there are $N_1, \ldots, N_n \in \Lambda_\oplus$ such that

- $N \longrightarrow_{q_1} N_1$,
- $N_i \longrightarrow_{q_{i+1}} N_{i+1}$ with $i \in \{1, \ldots, n-1\}$,
- $N_n \longrightarrow_{q_n} R$,

with, for all $j \in \{1, \ldots, n\}$, $L_j \;\mathcal{T}\; N_j$ and $Q \;\mathcal{T}\; R$. It follows $N \longrightarrow_p^* R$ with $P \;(\mathcal{R} \circ \mathcal{T})\; R$, which is the thesis. □

Moreover, we prove that $\dot{\sim}$ validates β-reduction. This result is crucial in the proof of Proposition 5.2.24.

Lemma 5.2.22. *For all $M, N \in \Lambda_\oplus(\emptyset)$, $(\lambda x.M)\, N \;\dot{\sim}\; M\,[N/x]$.*

Proof. On the one hand, Definition 5.2.12 entails $(\lambda x.M)\, N \longrightarrow M\,[N/x]$, and so the result follows by the reflexive property of $\dot{\sim}$ (Lemma 5.2.21). On the other hand, if $M\,[N/x] \longrightarrow_p L$, then $(\lambda x.M)\, N \longrightarrow M\,[N/x] \longrightarrow_p L$, *i.e.* $(\lambda x.M)\, N \longrightarrow_p^* L$, and the result follows by the reflexive property of $\dot{\sim}$ (Lemma 5.2.21). □

Lemma 5.2.23. *If $M \longrightarrow^* N$, then $M \;\dot{\sim}\; N$.*

Proof. Direct induction on the length of $M \longrightarrow^* N$, using Lemma 5.2.22. □

As previously mentioned, the crucial point in testing deterministic, pure λ-terms in a probabilistic context is that they exhibit the same [Choice] structure (modulo β-reductions). This is captured by the definition of strict applicative bisimulation (Definition 5.2.12), which intuitively can be turned into the classical notion of bisimulation for non-deterministic transition systems, *i.e.* into a one-to-one transitions matching relation. This is exploited in the proof of the following result.

Proposition 5.2.24. *On $\Lambda_\oplus$-terms, $\dot{\sim} \subseteq \sim$.*

Proof. We show that $\dot{\sim}$ is a probabilistic applicative bisimulation, hence included in the greatest one $\sim$. Lemma 5.2.21 entails $\dot{\sim}$ is an equivalence relation. Hence, we only need to verify that the following two hold:

1. for every $M \dot{\sim} N$ and every $\mathtt{E} \in \mathsf{V}\Lambda_\oplus / \dot{\sim}$, $\mathcal{P}_\oplus(M, \tau, \mathtt{E}) = \mathcal{P}_\oplus(N, \tau, \mathtt{E})$;

2. for every $\nu x.M \dot{\sim} \nu x.N$ and every $L \in \Lambda_\oplus(\emptyset)$, $M[L/x] \dot{\sim} N[L/x]$.

The 2nd clause directly holds by the 3rd condition of Definition 5.2.12. The 1st one is obviously more delicate. We show that M and N induce isomorphic [Choice] tree structure $\mathsf{CT}_\oplus(M)$ and $\mathsf{CT}_\oplus(N)$ respectively, where the [Abstraction] leaves are in a one-to-one relation due to the bisimulation game provided by the strict applicative bisimulation (Definition 5.2.12). A [Choice] tree structure is constructed as follows.

Consider the two labelled transition systems that describe the evaluation of M and N *w.r.t.* reduction $\longrightarrow$: rooted in M and N, these are transition systems where the set of states ranges over $\Lambda_\oplus(\emptyset) \uplus \mathsf{V}\Lambda_\oplus$, labels range over $\left\{\frac{1}{2}, 1\right\}$ and transitions are those given by definition of $\longrightarrow$ (Definition 5.2.11), in that a transition corresponding to a β-reduction is labelled with 1, whereas the two transitions caused by a probabilistic choices are both labelled with $\frac{1}{2}$. Observe that these labelled transition systems are no more than trees (recall $\oplus$ is a binary operator) such that nodes are [Application] or [Choice] terms, whereas leaves are [Abstraction] terms. Nonetheless, we can simplify these constructions and consider [Choice] trees only, *i.e.* where the nodes are just [Choice] terms. Indeed, Lemma 5.2.23 tells us that β-convertible terms are $\dot{\sim}$-equivalent, hence a sequence of transitions corresponding to a sequence of β-reduction can be squeezed altoghether, resulting in a [Choice] node or an [Abstraction] leaf. Observe that $\dot{\sim}$-equivalent nodes remain equivalent by the transitive property of $\dot{\sim}$ (Lemma 5.2.21). Given a term $L \in \Lambda_\oplus(\emptyset)$, we denote as $\mathsf{CT}_\oplus(L)$ the function giving such [Choice] tree structure of L. Then, the following lemma follows straightforwardly.

Lemma. *If $M \dot{\sim} N$, then $\mathsf{CT}_\oplus(M)$ is isomorphic to $\mathsf{CT}_\oplus(N)$.*

Proof. Suppose $\mathsf{CT}_\oplus(M)$ and $\mathsf{CT}_\oplus(N)$ are not isomorphic. Without loss of generality, at some level of the trees, there is a [Choice] node L in $\mathsf{CT}_\oplus(M)$ which is an [Abstraction] leaf P in $\mathsf{CT}_\oplus(N)$. Then, there exists a same path accross [Choice] $\dot{\sim}$-equivalent terms that leads to L in $\mathsf{CT}_\oplus(M)$ and to P in $\mathsf{CT}_\oplus(N)$. Of course $L \not\dot{\sim} P$, as $L \longrightarrow_{\frac{1}{2}} Q$ while the [Abstraction] leaf P cannot perform any transition. It follows $M \not\dot{\sim} N$, against the hypothesis $M \dot{\sim} N$. Hence $\mathsf{CT}_\oplus(M)$ and $\mathsf{CT}_\oplus(N)$ are isomorphic trees. ■

The tree isomorphism between $\mathsf{CT}_\oplus(M)$ and $\mathsf{CT}_\oplus(N)$ entails that their [Abstraction] leaves are in a bijection and strict bisimilar (recall that the $\mathsf{CT}_\oplus$'s are built following strict bisimulation game). In particular, this means that, for all $\mathtt{E} \in \mathsf{V}\Lambda_\oplus/\dot{\sim}$, if $V \in [\![M]\!] \cap \mathtt{E}$, then there is a unique $W \in [\![N]\!] \cap \mathtt{E}$ such that $M \longrightarrow_p V$, $N \longrightarrow_p W$ and $V \dot{\sim} W$. It follows

$$
\begin{aligned}
\mathcal{P}_\oplus(M,\tau,\mathtt{E}) &= \sum_{X\in\mathtt{E}} \mathcal{P}_\oplus(M,\tau,X) \\
&= \sum_{X\in\mathtt{E}} [\![M]\!]\,(X) \\
&= \sum_{X\in\mathtt{E}} \sum_{V\in[\![M]\!]\cap\mathtt{E}, V=X} p \\
&= \sum_{X\in\mathtt{E}} \sum_{W\in[\![N]\!]\cap\mathtt{E}, W=X} p \\
&= \sum_{X\in\mathtt{E}} [\![N]\!]\,(X) \\
&= \sum_{X\in\mathtt{E}} \mathcal{P}_\oplus(N,\tau,X) = \mathcal{P}_\oplus(N,\tau,\mathtt{E}).
\end{aligned}
$$

This concludes the proof. □

Since we now know that for *pure, deterministic* λ-terms, $=_{\mathsf{LL}}$ is included in $\sim$ (Lemma 5.2.20 and Proposition 5.2.24), that $\sim$ is included in $\simeq_\oplus$ (Theorem 4.5.11) and that the latter is included in $=_{\mathsf{LL}}$ (Corollary 5.2.10), we can conclude:

Corollary 5.2.25. *The relations $=_{\mathsf{LL}}$, $\sim$, and $\simeq_\oplus$ coincide in pure λ-calculus.*

Conclusions and Perspectives

This thesis presents two main contributions towards the understanding of the operational aspects of quantitative λ-calculi:

1. **Normal forms for the algebraic λ-calculus.** We have investigated the reduction theory of the algebraic λ-calculus [Vau09] with the intent of understanding how β-reduction, and computation in general, fits into the particular setting of a module of terms. In particular, we have focused on the open problem of normal forms for the algebraic λ-calculus.

   We have proposed a very first notion of unique normal form for this setting, even in the case where β-reduction collapses. From the technical point of view, this has required a full development of a *weak normalisation scheme* [ER03, Vau09], for the first time, without appealing to any type system. For this, we have put in relation $[\![\cdot]\!]$ any module of terms $\mathsf{R}\langle\Delta_{\mathsf{R}}\rangle$ with the module of terms $\mathbb{P}\langle\Delta_{\mathbb{P}}\rangle$ and, by non-trivial term rewriting techniques only, we have proved such relation stable with respect to normalisability. This permits to conceive a partial (since valid on normalisable terms only), but consistent, term equivalence validating β-reduction.

   Afterwards, we have studied the problem of attaining such normal forms directly, without appealing to the module of terms $\mathbb{P}\langle\Delta_{\mathbb{P}}\rangle$. To the end of conceiving consistent reduction notions, we have been forced to consider canonical terms only. We have proved that a notion of parallel β-reduction on these latter is sufficient to characterise the previously established notion of normal form. In order to do this, we have provided a method to translate reductions taking place in $\mathbb{P}\langle\Delta_{\mathbb{P}}\rangle$ into this latter notion of parallel β-reduction on canonical terms. However, we do not know whether it is also necessary. In fact, as we have failed to achieve the same characterisation with a notion of (one-step) β-reduction on canonical terms, we have not been able to show a counterexample proving the contrary. Nonetheless, we have discovered some fundamental properties, and deficiencies, of reduction relations on canonical terms only: canonical β-reduction does not validate β-rule in general, and no notion of reduction can be contextual, hence none enjoys Church-Rosser.

Finally, the quest for characterising normal forms by means of canonical β-reduction has guided us to prove a *factorisation theorem* for the algebraic λ-calculus. In the process, the classical stronger *standardisation theorem* is found unfeasible. Even though our result is similar to the one established for the resource λ-calculus [PT09] (in particular, when dealing with perpetual resources only), the technicalities required to handle the algebraic setting are crucially different. We believe our result sheds new light on the distinction between reduction related redex duplications, and those due to the algebraic component of the calculus. Indeed, the resource λ-calculus is bilinear, in that sums of terms arise only at top-level, and so the second kind of duplications cannot happen. Moreover, our result is not limited to terms with coefficients in $\mathbb{N}$.

We have confidence that our work confirms the complexity of studying β-reduction theory in pure λ-calculi endowed with a structure of module, moreover providing some answers on the subject. In particular, we have shown that, on the one hand, β-reduction needs to be contextual in order to exhibit good rewriting properties (*e.g.* Church-Rosser); on the other hand, contextuality with respect to sums leads to inconsistency.

2. **Coinductive equivalences in a probabilistic scenario.** We have studied *applicative bisimulation* and *context equivalence* in a probabilistic λ-calculus. In particular, we have endowed non-deterministic λ-calculus with a probabilistic call-by-name operational semantics [DLZ12]. The one presented here is the first investigation in which bisimulation techniques for program equivalence are shown to be applicable to probabilistic computation.

   In the spirit of applicative bisimulation for pure λ-calculus, we have shown a technique for proving congruence of a probabilistic notion of applicative bisimilarity. Technically, probabilistic applicative bisimulation is obtained by setting up a labelled Markov chain on top of terms, then adapting to it the coinductive scheme already well-established in a first order setting [LS91]. While the technique of congruence here proposed follows Howe's method, the proof has turned to be more complicated than in the cases of deterministic and non-deterministic setting. In particular, we have proved some non-trivial "disentangling" properties for sets of real numbers, themselves proved by modeling the problem as a flow network and then apply the Max-flow Min-cut theorem. We have shown that the congruence of applicative bisimilarity yields soundness with respect to context equivalence. Completeness, however, fails: applicative bisimilarity is proved to be finer.

Finally, we have proved that applicative bisimilarity is fully abstract with respect to context equivalence when pure terms are tested in contexts providing higher-order functions and probabilistic choice. For this, we have proved that, on pure terms, both applicative bisimilarity and context equivalence collapses to the *Lévy-Longo tree equality* [DCG01], which equates terms with the same Lévy-Longo tree. Along the way, we have shown a Böhm-out result where the presence of probabilistic information (*i.e.* probabilities) is exploited in an essential manner.

We believe that a nice contribution has been provided towards the understanding of observational equivalences in presence of quantitative information, such as probabilities of convergence. In particular, when we compare non-deterministic context equivalence and probabilistic context equivalence, we see that the two are incomparable. As an example of terms that are context equivalent in the *must* sense but not probabilistically, we can take $\mathbf{I} \oplus (\mathbf{I} \oplus \mathbf{\Omega})$ and $\mathbf{I} \oplus \mathbf{\Omega}$. Conversely, $\mathbf{I}$ is probabilistically equivalent to any term M that reduces to $\mathbf{I} \oplus M$ (which can be defined using fixpoint operator), while $\mathbf{I}$ and M are not equivalent in the must sense, since the latter can diverge (the divergence is irrelevant probabilistically because it has probability zero). *May* context equivalence, in contrast, is coarser than probabilistic context equivalence. Despite the differences, the two semantics have similarities, in that applicative bisimulation and context equivalence do not coincide either in the non-determistic setting or in the probabilistic one, at least if call-by-name is considered.

Topics for future work abound. We report on some of the most interesting ones:

- Technically speaking, our work on the algebraic λ-calculus goes on the original idea of studying β-reduction in a *module of λ-terms*, where the syntax of the calculus was conceived just as a notation for objects of certain denotational models of linear logic [Ehr02, Ehr05]. However, in presence of non-positive semiring, β-reduction collapses, resulting in a trivial term equivalence (even if we restrict reduction on canonical terms only).

  This seems to suggest a rethinking of the algebraic equality (Definition 1.3.5) in order to avoid the decomposition of $\mathbf{0}$ into $\infty_S - \infty_S$. But how should we modify algebraic equality without giving up on the notion of module of terms completely?

  An answer might be Arrighi and Dowek's way of developing *linear algebraic λ-calculus* [AD08], namely by orienting *all* the identities of the algebraic equality.

This turns the calculus into a proper term rewriting system where rewriting rules are admitted under certain conditions. Except that this latter cannot be properly considered an extension of pure λ-calculus with a structure of R-module, it is technically cumbersome and difficult to deal with.

This opens interesting questions. Is there a way of orienting only *few* identities of algebraic equality so as to avoid the aforementioned collapse? What are the reduction properties of the related calculus? Is the notion of normal form the same we gave in Section 2.3?

To the end of preventing the collapse of β-reduction, typing seems a plausible idea to prevent arbitrary fixpoints. However, Vaux has shown that this is a non-trivial option [Vau09], as typability is not preserved under our notion of reduction, in general. The literature presents already some work on similar issues [ADC12, ADCV13].

- In Chapter 3, we have established some properties of head reduction in the algebraic λ-calculus. Despite the fact that the classical formulation of standardisation theorem is out of reach, it is not clear whether there is a strategy, different from the leftmost one, which can be considered *standard* in our setting. Indeed, the justifications reported by Pagani and Tranquilli against the possibility of a standard strategy for the resource λ-calculus [PT09] do not apply here. In that case, they rely on the non-deterministic nature of some aspects of the calculus, something missing in the algebraic λ-calculus.

  We have found that head reduction does not commute with internal ones, in general. Therefore, we have decomposed β-reduction in function and argument reductions, and we have shown that the former characterises head normalisability. Is it possible to translate a function reduction leading to an head normal form into an head reduction sequence?

  Related to this, there is the open question on the kind of normalisability enjoyed by function reduction. On normalisable terms, is function reduction strongly normalising?

  We believe so with respect to both questions, but we lack a formal, technical argument. For this, we most likely need a further decomposition of function reduction (which is non-deterministic), something that recall us of the work recently done on *linear head reduction*.

- In Chapter 4, we have proved that probabilistic applicative bisimilarity is strictly finer than probabilistic context equivalence. Surprisingly, Crubillé and Dal Lago [CL14] have proved that this is a peculiarity of call-by-name evaluation strategy. By considering the call-by-value variant of our setting, they have shown that applicative bisimilarity coincides with context equivalence, and that, even more surprisingly, full-abstraction holds only in the symmetric setting (*i.e.* not in the case of applicative similarity). This is a phenomenon in sharp contrast with what happen in both deterministic [Abr90] and non-deterministic cases [Las98], where the choice of evaluation does not affect the relation between applicative bisimilarity and context equivalence.

To prove full-abstraction, Crubillé and Dal Lago exploit a well-known [DEP02b] generalisation of applicative (bi)similarity to labelled Markov processes (*i.e.* labelled probabilistic transition systems with continuous state space), where a relatively simple characterisation in terms of *testing* is available [vBMOW05]. Finally, they prove that context equivalence coincides with testing equivalence, hence with applicative bisimilarity, by showing that every test can be turned into an appropriate context.

A similar characterisation of bisimulation via testing is available in call-by-name setting as well. However, this latter is not expressive enough to implement every test as a context. As the authors have highlighted, what call-by-name evaluation misses is the capability to feed in arguments to functions after having evaluated them, something peculiar of call-by-value strategy instead.

The gap between call-by-name and call-by-value evaluations opens interesting horizons. In particular, Crubillé and Dal Lago's questions on the subject are intriguing. Is an operator of *sequencing* adequate to recover full-abstraction in a call-by-name setting? What is the subclass of tests implementable by a call-by-name setting? What about call-by-need evaluation?

- Our work is based on the notion of value distribution, which is an essentially infinitary operational semantics where the meaning of a term is obtained as the least upper bound of all its finite approximations. Would it be possible to define an *effective* notion of bisimulation in terms of approximations, without getting too fine grained?

  Related to this question is the quest to understand how to relate the algebraic λ-calculus with the probabilistic one. Obviously the two are close, at least at the level of terms, as the latter can be considered the instance of the former when considering the particular module of terms $\mathbb{R}_{[0,1]}\langle\Delta_{\mathbb{R}_{[0,1]}}\rangle$. Is it possible that $\mathbb{R}_{[0,1]}\langle\Delta_{\mathbb{R}_{[0,1]}}\rangle$ is the object language for studying such notions of effective equivalences?

  In particular, the idea of directly capturing the notion of distributions by possibly infinite formal sums of terms (hence, a subset of $\mathbb{R}_{[0,1]}\langle\Delta_{\mathbb{R}_{[0,1]}}\rangle$)

has been conceived for studying the notion of probabilistic *coupled logical bisimulation* [LSA14].

More in general, how does the works on probabilistic denotational models of linear logic [EPT11, EPT14] relate to ours? Does (algebraic) linearity in $\Lambda_\oplus$ affect our results?

The last question is intriguing and deserves further study. Observe, indeed, that endowing $\Lambda_\oplus$ with linearity has the immediate consequence that Counterexample 4.5.12 is no longer valid as M and N turn to the same term. Is it only a coincidence or linearity is a real issue?

- Applicative bisimulation and Howe's method have been successfully adapted in several settings [Pit97, Gor98, Gor99]. However, some problems arise when dealing, for instance, with state [JR99], concurrency [JR00, GH05] and infinitary syntax [Lev06]. For these reasons, new theories of bisimulation for higher-order languages have been proposed under the names of *logical bisimulations* [SKS07] and *environmental bisimulation* [SP07b, SP07a, SKS11]. These latter introduce a more complex machinery, especially on the environment in which functions are tested, which is nonetheless essential [KLS11].

  The further analysis of such techniques in our setting is of great interest. Related to this is the development of *up-to techniques*, which are a well-established method permitting to mitigate the technical burden of the proof method associated to bisimulation.

Bibliography

[ABM14] Beniamino Accattoli, Pablo Barenbaum, and Damiano Mazza. Distilling abstract machines. In *Proceedings of the 19th ACM SIGPLAN International Conference on Functional Programming*, ICFP '14, pages 363–376, New York, NY, USA, 2014. ACM.

[Abr90] Samson Abramsky. The lazy λ-calculus. In D. Turner, editor, *Research Topics in Functional Programming*, pages 65–117. Addison Wesley, 1990.

[AC84] Egidio Astesiano and Gerardo Costa. Distributive semantics for nondeterministic typed lambda-calculi. *Theor. Comput. Sci.*, 32:121–156, 1984.

[AD08] Pablo Arrighi and Gilles Dowek. Linear-algebraic lambda-calculus: higher-order, encodings, and confluence. In *Rewriting Techniques and Applications, 19th International Conference, RTA 2008, Hagenberg, Austria, July 15-17, 2008, Proceedings*, pages 17–31, 2008.

[ADC12] Pablo Arrighi and Alejandro Díaz-Caro. A System F accounting for scalars. *Logical Methods in Computer Science*, 8(1:11), 2012.

[ADCV13] Pablo Arrighi, Alejandro Díaz-Caro, and Benoît Valiron. The vectorial lambda-calculus. *CoRR*, abs/1308.1138, 2013.

[ADL14] Beniamino Accattoli and Ugo Dal Lago. Beta reduction is invariant, indeed. In *Proceedings of the Joint Meeting of the Twenty-Third EACSL Annual Conference on Computer Science Logic (CSL) and the Twenty-Ninth Annual ACM/IEEE Symposium on Logic in Computer Science (LICS)*, CSL-LICS '14, pages 8:1–8:10, New York, NY, USA, 2014. ACM.

[Alb13] Michele Alberti. Normal forms for the algebraic lambda-calculus. In Damien Pous and Christine Tasson, editors, *JFLA - Journées francophones des langages applicatifs*, Aussois, France, 2013.

[AO93] Samson Abramsky and C.-H. Luke Ong. Full abstraction in the lazy lambda calculus. *Inf. Comput.*, 105(2):159–267, 1993.

[Bar84] Hendrik Pieter Barendregt. *The Lambda Calculus – Its Syntax and Semantics*, volume 103 of *Studies in Logic and the Foundations of Mathematics*. North-Holland, 1984.

[BDCJ12] Pablo Buiras, Alejandro Díaz-Caro, and Mauro Jaskelioff. Confluence via strong normalisation in an algebraic λ-calculus with rewriting. In Simona Ronchi della Rocca and Elaine Pimentel, editors, *Proceedings of the 6th Workshop on Logical and Semantic Frameworks with Applications (LSFA 2011)*, volume 81 of *EPTCS*, pages 16–29, 2012.

[BDL13] Marco Bernardo, Rocco De Nicola, and Michele Loreti. A uniform framework for modeling nondeterministic, probabilistic, stochastic, or mixed processes and their behavioral equivalences. *Inf. Comput.*, 225:29–82, 2013.

[BL96] G. Boudol and C. Laneve. The discriminating power of the λ-calculus with multiplicities. *Inf. Comput.*, 126(1):83–102, 1996.

[Bou93] Gérard Boudol. The lambda-calculus with multiplicities (abstract). In *Proceedings of the 4th International Conference on Concurrency Theory*, CONCUR '93, pages 1–6, 1993.

[Bou94] Gérard Boudol. Lambda-calculi for (strict) parallel functions. *Inf. Comput.*, 108(1):51–127, 1994.

[CL14] Raphaëlle Crubillé and Ugo Dal Lago. On probabilistic applicative bisimulation and call-by-value λ-calculi. In *Programming Languages and Systems - 23rd European Symposium on Programming, ESOP 2014, Held as Part of the European Joint Conferences on Theory and Practice of Software, ETAPS 2014, Grenoble, France, April 5-13, 2014, Proceedings*, pages 209–228, 2014.

[DCG01] M. Dezani-Ciancaglini and E. Giovannetti. From Bohm's theorem to observational equivalences: an informal account. *Electr. Notes Theor. Comput. Sci.*, 50(2):83–116, 2001.

[DCPTV10] Alejandro Díaz-Caro, Simon Perdrix, Christine Tasson, and Benoît Valiron. Equivalence of algebraic lambda-calculi - work in progress -. *CoRR*, abs/1005.2897, 2010.

[DCTU99] M. Dezani-Ciancaglini, J. Tiuryn, and P. Urzyczyn. Discrimination by parallel observers: The algorithm. *Inf. Comput.*, 150(2):153–186, 1999.

[DEP02a] Josee Desharnais, Abbas Edalat, and Prakash Panangaden. Bisimulation for labelled markov processes. *Inf. Comput.*, 179(2):163–193, 2002.

[DEP02b] Josee Desharnais, Abbas Edalat, and Prakash Panangaden. Bisimulation for labelled markov processes. *Inf. Comput.*, 179(2):163–193, 2002.

[DH84] Rocco De Nicola and Matthew Hennessy. Testing equivalences for processes. *Theor. Comput. Sci.*, 34:83–133, 1984.

[DH02] Vincent Danos and Russell Harmer. Probabilistic game semantics. *ACM Trans. Comput. Log.*, 3(3):359–382, 2002.

[DHR96] Vincent Danos, Hugo Herbelin, and Laurent Regnier. Game semantics & abstract machines. In *Proceedings, 11th Annual IEEE Symposium on Logic in Computer Science, New Brunswick, New Jersey, USA, July 27-30, 1996*, pages 394–405, 1996.

[DLZ12] Ugo Dal Lago and Margherita Zorzi. Probabilistic operational semantics for the lambda calculus. *RAIRO - Theor. Inf. and Applic.*, 46(3):413–450, 2012.

[dP95] Ugo de'Liguoro and Adolfo Piperno. Nondeterministic extensions of untyped lambda-calculus. *Inf. Comput.*, 122(2):149–177, 1995.

[DR04] Vincent Danos and Laurent Regnier. Head linear reduction. Technical report, 2004.

[Ehr02] Thomas Ehrhard. On Köthe sequence spaces and linear logic. *Mathematical Structures in Computer Science*, 12(5), 2002.

[Ehr05] Thomas Ehrhard. Finiteness spaces. *Mathematical Structures in Computer Science*, 15(4):615–646, 2005.

[EL10a] Thomas Ehrhard and Olivier Laurent. Acyclic solos and differential interaction nets. *Logical Methods in Computer Science*, 6(3), 2010.

[EL10b] Thomas Ehrhard and Olivier Laurent. Interpreting a finitary pi-calculus in differential interaction nets. *Inf. Comput.*, 208(6):606–633, 2010.

[EPT11] Thomas Ehrhard, Michele Pagani, and Christine Tasson. The Computational Meaning of Probabilistic Coherence Spaces. In Martin Grohe, editor, *Proceedings of the 26th Annual IEEE Symposium on Logic in Computer Science (LICS 2011)*, IEEE Computer Society Press, pages 87–96, 2011.

[EPT14] Thomas Ehrhard, Michele Pagani, and Christine Tasson. Probabilistic Coherence Spaces are Fully Abstract for Probabilistic PCF. In P. Sewell, editor, *The 41th Annual ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages, POPL14, San Diego, USA*. ACM, 2014.

[ER03] Thomas Ehrhard and Laurent Regnier. The differential lambda-calculus. *Theor. Comput. Sci.*, 309(1), 2003.

[ER06a] Thomas Ehrhard and Laurent Regnier. *Böhm trees, Krivine's machine and the Taylor expansion of lambda-terms*, volume 3988 of *Lecture Notes in Computer Science*, pages 186–197. Springer, 2006.

[ER06b] Thomas Ehrhard and Laurent Regnier. Differential interaction nets. *Theor. Comput. Sci.*, 364(2):166–195, 2006.

[GAB$^+$13] Andrew D. Gordon, Mihhail Aizatulin, Johannes Borgström, Guillaume Claret, Thore Graepel, Aditya V. Nori, Sriram K. Rajamani, and Claudio V. Russo. A model-learner pattern for bayesian reasoning. In *The 40th Annual ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages, POPL '13, Rome, Italy - January 23 - 25, 2013*, pages 403–416, 2013.

[GH05] Jens Chr. Godskesen and Thomas T. Hildebrandt. Extending Howe's method to early bisimulations for typed mobile embedded resources with local names. In *FSTTCS 2005: Foundations of Software Technology and Theoretical Computer Science, 25th International Conference, Hyderabad, India, December 15-18, 2005, Proceedings*, pages 140–151, 2005.

[Gir87] Jean-Yves Girard. Linear logic. *Theoretical Computer Science*, (50):1–102, 1987.

[Gir88] Jean-Yves Girard. Normal functors, power series and lambda-calculus. *Ann. Pure Appl. Logic*, 37(2):129–177, 1988.

[GM84] Shafi Goldwasser and Silvio Micali. Probabilistic encryption. *J. Comput. Syst. Sci.*, 28(2):270–299, 1984.

[Goo13] Noah D. Goodman. The principles and practice of probabilistic programming. In *The 40th Annual ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages, POPL '13, Rome, Italy - January 23 - 25, 2013*, pages 399–402, 2013.

[Gor98] A. D. Gordon. Operational equivalences for untyped and polymorphic object calculi. In Andrew D. Gordon and Andrew M. Pitts, editors, *Higher Order Operational Techniques in Semantics*, pages 9–54. 1998.

[Gor99] Andrew D. Gordon. Bisimilarity as a theory of functional programming. *Theor. Comput. Sci.*, 228(1-2):5–47, 1999.

[Hen12] Matthew Hennessy. Exploring probabilistic bisimulations, part I. *Formal Asp. Comput.*, 24(4-6):749–768, 2012.

[How96] Douglas J. Howe. Proving congruence of bisimulation in functional programming languages. *Inf. Comput.*, 124(2):103–112, 1996.

[JP89] C. Jones and Gordon D. Plotkin. A probabilistic powerdomain of evaluations. In *Proceedings, Fourth Annual Symposium on Logic in Computer Science, 5-8 June, 1989, Asilomar Conference Center, Pacific Grove, California, USA*, pages 186–195, 1989.

[JP90] Radha Jagadeesan and Prakash Panangaden. A domain-theoretic model for a higher-order process calculus. In *Automata, Languages and Programming, 17th International Colloquium, ICALP90, Warwick University, England, July 16-20, 1990, Proceedings*, pages 181–194, 1990.

[JR99] Alan Jeffrey and Julian Rathke. Towards a theory of bisimulation for local names. In *14th Annual IEEE Symposium on Logic in Computer Science, Trento, Italy, July 2-5, 1999*, pages 56–66, 1999.

[JR00] Alan Jeffrey and Julian Rathke. A theory of bisimulation for a fragment of concurrent ML with local names. In *15th Annual IEEE Symposium on Logic in Computer Science, Santa Barbara, California, USA, June 26-29, 2000*, pages 311–321, 2000.

[KLS11] Vasileios Koutavas, Paul Blain Levy, and Eijiro Sumii. From applicative to environmental bisimulation. *Electr. Notes Theor. Comput. Sci.*, 276:215–235, 2011.

[Kri93] Jean-Louis Krivine. *Lambda-calculus, types and models*. Ellis Horwood series in computers and their applications. Masson, 1993.

[Las98] S. B. Lassen. *Relational Reasoning about Functions and Nondeterminism*. PhD thesis, University of Aarhus, 1998.

[Lév75] Jean-Jacques Lévy. An algebraic interpretation of equality in some models of the lambda calculus. In C. Böhm, editor, *Lambda Calculus and Computer Science Theory*, volume 37 of *LNCS*, pages 147–165. Springer-Verlag, 1975.

[Lev06] Paul Blain Levy. Infinitary howe's method. *Electr. Notes Theor. Comput. Sci.*, 164(1):85–104, 2006.

[Lon83] Giuseppe Longo. Set-theoretical models of lambda calculus: Theories, expansions and isomorphisms. *Ann. Pure Appl. Logic*, 24:153–188, 1983.

[LS91] Kim Guldstrand Larsen and Arne Skou. Bisimulation through probabilistic testing. *Inf. Comput.*, 94(1):1–28, 1991.

[LSA14] Ugo Dal Lago, Davide Sangiorgi, and Michele Alberti. On coinductive equivalences for higher-order probabilistic functional programs. In *The 41st Annual ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages, POPL '14, San Diego, CA, USA, January 20-21, 2014*, pages 297–308, 2014.

[Mel97] Paul-André Melliès. A factorisation theorem in rewriting theory. In *Category Theory and Computer Science*, pages 49–68, 1997.

[Mil92] Robin Milner. Functions as processes. *Mathematical Structures in Computer Science*, 2(2):119–141, 1992.

[MP94] Gianfranco Mascari and Marco Pedicini. Head linear reduction and pure proof net extraction. *Theoretical Computer Science*, 135(1):111–137, 1994.

[Ong93] C.-H. Luke Ong. Non-determinism in a functional setting. In *Proceedings, Eighth Annual IEEE Symposium on Logic in Computer Science, 19-23 June 1993, Montreal, Canada*, pages 275–286, 1993.

[Pan09] Prakash Panangaden. *Labelled Markov Processes*. Imperial College Press, 2009.

[Pan11] Prakash Panangaden. Probabilistic bisimulation. In D. Sangiorgi and J. Rutten, editors, *Advanced Topics in Bisimulation and Coinduction*, pages 290–326. Cambridge University Press, 2011.

[Par03] Sungwoo Park. A calculus for probabilistic languages. In *Proceedings of the 2003 ACM SIGPLAN International Workshop on Types in Languages Design and Implementation*, TLDI '03, pages 38–49, 2003.

[Pfe01] Avi Pfeffer. IBAL: A probabilistic rational programming language. In *Proceedings of the Seventeenth International Joint Conference on Artificial Intelligence, IJCAI 2001, Seattle, Washington, USA, August 4-10, 2001*, pages 733–740. Morgan Kaufmann, 2001.

[Pit97] A. M. Pitts. Operationally-based theories of program equivalence. In P. Dybjer and A. M. Pitts, editors, *Semantics and Logics of Computation*, Publications of the Newton Institute, pages 241–298. Cambridge University Press, 1997.

[Pit11] Andrew M. Pitts. Howe's method for higher-order languages. In D. Sangiorgi and J. Rutten, editors, *Advanced Topics in Bisimulation and Coinduction*, pages 197–232. Cambridge University Press, 2011.

[Plo75] Gordon D. Plotkin. Call-by-name, call-by-value and the lambda-calculus. *Theor. Comput. Sci.*, 1(2):125–159, 1975.

[PPT08] Sungwoo Park, Frank Pfenning, and Sebastian Thrun. A probabilistic language based on sampling functions. *ACM Trans. Program. Lang. Syst.*, 31(1), 2008.

[PR10] Michele Pagani and Simona Ronchi Della Rocca. Linearity, non-determinism and solvability. *Fundamenta Informaticae*, 103(1-4):173–202, 2010.

[PT09] Michele Pagani and Paolo Tranquilli. Parallel reduction in resource lambda-calculus. In *Programming Languages and Systems, 7th Asian Symposium, APLAS 2009, Seoul, Korea, December 14-16, 2009. Proceedings*, pages 226–242, 2009.

[Reg94] Laurent Regnier. Une équivalence sur les lambda-termes. *Theor. Comput. Sci.*, 126(2):281–292, 1994.

[RP02] Norman Ramsey and Avi Pfeffer. Stochastic lambda calculus and monads of probability distributions. In *Conference Record of POPL 2002: The 29th SIGPLAN-SIGACT Symposium on Principles of Programming Languages, Portland, OR, USA, January 16-18, 2002*, pages 154–165, 2002.

[San94] Davide Sangiorgi. The lazy lambda calculus in a concurrency scenario. *Inf. Comput.*, 111(1):120–153, 1994.

[San11] Davide Sangiorgi. *Introduction to Bisimulation and Coinduction*. Cambridge University Press, New York, NY, USA, 2011.

[SD78] N. Saheb-Djahromi. Probabilistic LCF. In *Mathematical Foundations of Computer Science 1978, Proceedings, 7th Symposium, Zakopane, Poland, September 4-8, 1978*, volume 64 of *LNCS*, pages 442–451, 1978.

[Sel04] Peter Selinger. Towards a quantum programming language. *Mathematical Structures in Computer Science*, 14(4):527–586, 2004.

[Sie93] Kurt Sieber. Call-by-value and nondeterminism. In *Typed Lambda Calculi and Applications, International Conference on Typed Lambda Calculi and Applications, TLCA '93, Utrecht, The Netherlands, March 16-18, 1993, Proceedings*, volume 664 of *LNCS*, pages 376–390, 1993.

[SKS07] Davide Sangiorgi, Naoki Kobayashi, and Eijiro Sumii. Logical bisimulations and functional languages. In *International Symposium on Fundamentals of Software Engineering, International Symposium, FSEN 2007, Tehran, Iran, April 17-19, 2007, Proceedings*, pages 364–379, 2007.

[SKS11] Davide Sangiorgi, Naoki Kobayashi, and Eijiro Sumii. Environmental bisimulations for higher-order languages. *ACM Trans. Program. Lang. Syst.*, 33(1):5, 2011.

[SP07a] Eijiro Sumii and Benjamin C. Pierce. A bisimulation for dynamic sealing. *Theor. Comput. Sci.*, 375(1-3):169–192, 2007.

[SP07b] Eijiro Sumii and Benjamin C. Pierce. A bisimulation for type abstraction and recursion. *J. ACM*, 54(5), 2007.

[SW01] Davide Sangiorgi and David Walker. *The pi-calculus – a theory of mobile processes*. Cambridge University Press, 2001.

[Tak95] Masako Takahashi. Parallel reductions in λ-calculus. *Inf. Comput.*, 118(1), 1995.

[Vau07] Lionel Vaux. On linear combinations of λ-terms. In *Term Rewriting and Applications, 18th International Conference, RTA 2007, Paris, France, June 26-28, 2007, Proceedings*, pages 374–388, 2007.

[Vau09] Lionel Vaux. The algebraic lambda calculus. *Mathematical Structures in Computer Science*, 19(5):1029–1059, 2009.

[vBMOW05] Franck van Breugel, Michael W. Mislove, Joël Ouaknine, and James Worrell. Domain theory, testing and simulation for labelled markov processes. *Theor. Comput. Sci.*, 333(1-2):171–197, 2005.

[vT04] André van Tonder. A lambda calculus for quantum computation. *SIAM J. Comput.*, 33(5):1109–1135, 2004.