

Strategies for Asymptotic Normalization

Claudia Faggian

IRIF, CNRS, Université de Paris Cité, F-75013 Paris, France

Giulio Guerrieri  

Huawei Research, Edinburgh Research Centre, UK

Abstract

We present an abstract technique to study normalizing strategies when termination is asymptotic, that is, it appears as a limit. Asymptotic termination occurs in several settings, such as effectful, and in particular probabilistic computation – where the limits are distributions over the possible outputs – or infinitary lambda-calculi – where the limits are infinitary terms such as Böhm trees.

As a concrete application, we obtain a result which is of independent interest: a normalization theorem for Call-by-Value (and – in a uniform way – for Call-by-Name) probabilistic lambda-calculus.

2012 ACM Subject Classification Theory of computation → Models of computation; Theory of computation → Equational logic and rewriting; Theory of computation → Lambda calculus

Keywords and phrases rewriting, strategies, normalization, lambda calculus, probabilistic rewriting

Digital Object Identifier 10.4230/LIPIcs.FSCD.2022.17

Related Version *Extended Version*: <http://arxiv.org/abs/2204.08772>

Funding Work supported by the ANR project PPS: ANR-19-CE48-0014.

1 Introduction

Probabilistic computation is an example of computational paradigm where the notion of termination is *asymptotic*, that is, it appears as a *limit*, as opposed to reaching a normal form in a *finite* number of steps. Streams, infinitary λ -calculus, algebraic rewriting systems, effectful computation, are other examples: the notion of asymptotic computation is pervasive. Here, we investigate asymptotic normalization, and propose a technique to prove that a strategy is guaranteed to produce a maximal or – ideally – the best possible result. Our technique is abstract (in the sense of Abstract Rewriting Systems) and so of general application.

Rewriting is a foundation for the operational theory of formal calculi and programming languages – λ -calculus being the paradigmatic example where rewriting is an abstract form of program execution. Even if a programming language is usually defined by a specific *evaluation strategy*, to have a general rewriting theory allows for *program transformations*, *optimizations*, *parallel/distributed implementations*, and provides a base on which to reason about program equivalence. The λ -calculus has a rich theory that studies the properties of reductions. Asymptotic computation is much less understood from a rewriting point of view, with the notable exception of infinitary λ -calculus, whose rewriting theory, pioneered in [9, 22, 23], has been extensively studied.

The process of rewriting describes the *computation of a result*. Normal forms, head normal forms, values, may or must termination, are all possible notions of result. For concreteness, let us focus on normal forms. Operationally, key questions about a system are the existence and uniqueness of normal forms, but also *how* the result is computed. In a *finitary setting* we would ask: may a computation produce a result (*Existence* of normal forms)? If so, is the result unique? Do different computations on the same input lead to the same result (*Uniqueness* of normal forms)? How to compute a result? Is there a reduction strategy that is guaranteed to output a result, if any exists (*Normalizing strategy*)? In the *asymptotic case*, such questions are still relevant, but need to be opportunely formulated. To answer



© Claudia Faggian and Giulio Guerrieri;

licensed under Creative Commons License CC-BY 4.0

7th International Conference on Formal Structures for Computation and Deduction (FSCD 2022).

Editor: Amy P. Felty; Article No. 17; pp. 17:1–17:24

Leibniz International Proceedings in Informatics



LIPICs Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

them, we need suitable tools and techniques, because those for finitary computation do not necessarily transfer (the key game-changer being that asymptotic termination does not provide a well-founded order, see [13] for examples in a probabilistic setting).

Abstract Asymptotic Rewriting. Our approach is to study asymptotic reduction strategies and properties of limits in an *abstract* way (independent of the specific syntax of a calculus) as the theory of Abstract Rewrite Systems (ARS) does for finitary computation, so to isolate proof-techniques which are of general application. For example, in infinitary lambda calculus, the limit is usually a (possibly infinite) limit term, while in probabilistic lambda calculus, the limit is a distribution over (finite) terms. The former is concerned with the depth of the redexes, the latter with the probability of reaching a result. The abstract notions of limit and normalization subsume both, and so abstract results apply to either setting. A further, conceptual advantage of an abstract approach is to display the essence of the arguments, and to neatly discriminate between those properties that rely on specific structure of a concrete setting, and those that belong to any asymptotic notion of computation.

Specifically, we work in the setting of *Quantitative Abstract Rewrite System* (QARS) [14], a framework to study asymptotic rewriting *abstractly* which refines Ariola and Blom’s ARSI [4].

From normal forms to limits. Intuitively, a possibly infinite reduction sequence $\langle t_n \rangle_n$ from $t = t_0$ expresses a computation whose *result* is the maximal amount of information produced by that sequence. This is formalized as a limit. When the reduction is deterministic, it is standard to interpret such a limit as the meaning $\llbracket t \rrbracket$ of t . If however t has *several possible reduction sequences*, each can produce a different outcome (a different limit). It is then natural to define the meaning $\llbracket t \rrbracket$ of a term t as the *greatest* element in the set of limits, if any.¹ Intuitively, this means that the notion of “greatest amount of information produced by any reduction sequence” is well defined. To adopt such a notion demands care – for example, in the case of probabilistic and effectful computation, non-deterministic evaluation brings out issues which do not appear in pure lambda-calculus, not even when infinitary.

Given a term t and a general reduction, the notion of result $\llbracket t \rrbracket$ is not necessarily defined: the set of limits for t may contain different maximal elements, or it may not even have any maximal element (think of \mathbb{N} or $[0, 1)$, which have no maximum). Maximal limits play a role similar to normal forms, and the following questions are then natural.

1. Is there a strategy that produces a maximal amount of information (a maximal limit)?
2. Given a term t , is $\llbracket t \rrbracket$ – the result of computing t – well defined?

In Sect. 3 we provide tools to answer these questions, in this order, as we discuss next.

On the workflow (and the limits of confluence). The λ -calculus has two fundamental syntactical results: *confluence*, which implies **uniqueness of normal forms**, and the *standardization theorem*, which implies **normalization**, namely that a normal form (if any) can be reached by a computable strategy, which is a standard reduction (typically, left-to-right). Uniqueness guarantees that the notion of result is well defined, normalization provides a method to actually compute it.

A common workflow when studying λ -calculi is to first prove uniqueness of normal forms (via confluence), then normalization (via standardization). However, in an asymptotic setting *confluence does not directly imply* that the set of limits has a greatest element, but only that

¹ One could also define $\llbracket t \rrbracket$ as the lub of the set of limits, but this opens the question if there is a strategy that asymptotically computes $\llbracket t \rrbracket$, *internally* to the calculus. Since our focus is developing an *operational* theory, we require that $\llbracket t \rrbracket$ is itself a limit – it is a result that can be (asymptotically) computed.

it has a least upper bound. So, even if confluence is established, one still needs to prove that the lub is itself a limit, which may be a non-trivial task. For example, in the probabilistic λ -calculus [13, 14, 16], such a proof relies on (technical) properties of probability distributions.

In this paper, we *reverse the workflow*, and focus on normalization. In the finitary setting, if a rewriting relation \rightarrow has a strategy $\xrightarrow{e} \subset \rightarrow$ that satisfies a suitable completeness hypothesis and uniqueness of normal forms, so does \rightarrow (see [11]). With opportune definitions, this lifts well to the asymptotic setting. Forgoing confluence and focusing on normalization yields an efficient and uniform method which is easy to apply and which provides simultaneously (1.) existence and uniqueness of maximal limits, and (2.) a strategy to compute it.

Content and contributions. We start by illustrating asymptotic computation with examples (Sect. 1.1). Instances of asymptotic computation are quite diverse, and the syntax of each system may be rather complex. To study rewriting *abstractly*, in the spirit of Abstract Rewriting Systems (ARS), makes possible to analyze asymptotic properties in a way independent of a particular syntax, and to develop *general* proof techniques. In Sect. 2 we present the setting of Quantitative Abstract Rewriting Systems (QARS) [14], which are ARS enriched with a notion of observation. QARS are a natural refinement of ARSI [4].

Our first original contribution, and the heart of this paper, is Sect. 3, which proposes a proof technique to study asymptotic reduction strategies, and properties of the limits. We first introduce *asymptotic normalization*, which gives at the same time a tool to establish the existence of maximal limits – or of a greatest one – and a way to compute it. It formalizes the intuition that a normalizing strategy gradually computes (in a finite or infinite number of steps) the/a maximal amount of information that an element t can produce. We then show (Sect. 3.1) that asymptotic normalization can be established by proving that a strategy is asymptotically complete and has a unique limit. Remarkably, such *infinitary properties* reduce to a finitary one, *factorization* (a simple form of *standardization*) and to some *local, elementary tests*, yielding a practical and versatile proof-technique.

We then apply our method to some representative case studies based on λ -calculus. In order to do so, we first revisit normalization for λ -calculus – uniformly for Call-by-Value and Call-by-Name – so as to have a (*novel*) *normalizing strategy* (Sect. 4.2) which is well-suited to asymptotic normalization, and to deal with (CbV and CbN) probabilistic λ -calculi. The application of our method to *probabilistic λ -calculus* yields a result of independent interest, which was left as open question in [16] (Remark 27 there), namely a theorem of asymptotic normalization for Call-by-Value probabilistic λ -calculus. We develop the CbV case explicitly in Sect. 5.1 – the same results hold *in a uniform way* for Call-by-Name. The same technique applies to other monadic calculi such as calculi with output (as we sketch in Sect. 6), but also to the asymptotic computation of Böhm trees, which can be obtained as the limit of a normalizing strategy (we leave this case to Appendix D.2).

1.1 Three examples of Asymptotic Computation

We illustrate three diverse examples of asymptotic computation, where the result of the computation is the limit of an *infinitary* process. All three examples are built on λ -calculus.

Probabilistic computation. A probabilistic program P is a stochastic model generating a distribution over all possible outputs of P . Even if the termination probability is 1 (*almost sure termination*), that degree of certitude is typically not reached in a finite number of steps, but *as a limit*. A standard example is a term M that reduces to either a normal form or

M itself, with equal probability $1/2$. After n steps, M is in normal form with probability $\frac{1}{2} + \frac{1}{2^2} + \dots + \frac{1}{2^n}$. *Only at the limit* this computation terminates with probability 1. A direct way to model higher-order probabilistic computation is to endow the untyped λ -calculus with a binary operator \oplus which models fair, binary probabilistic choice: $M_1 \oplus M_2$ reduces to either M_1 or M_2 with equal probability $1/2$; we write this as $M_1 \oplus M_2 \rightarrow [\frac{1}{2}M_1, \frac{1}{2}M_2]$. Intuitively, *the result* of evaluating a probabilistic term is a *distribution* on its possible outputs.

► **Example 1.** Let $\Delta_{\oplus} = \lambda x.I \oplus (xx)$, where $I = \lambda x.x$. The term $M := \Delta_{\oplus}\Delta_{\oplus}$ has the behavior we have described above, and evaluates to I with probability 1 only at the limit.

Computations with output. Consider a program that can print an output. Following [17], we can represent this with a pair $\mathbf{s} : M$, where \mathbf{s} is a string over an alphabet \mathbb{A} , and M is a term of the λ -calculus extended with a set of operators $\text{out} = \{\text{out}_c \mid c \in \mathbb{A}\}$. The term $\text{out}_c(P)$ outputs c , adding it to the string, and continues as P . That is, $\langle \mathbf{s} : \text{out}_c(P) \rangle \rightarrow \langle c.\mathbf{s} : P \rangle$.

► **Example 2.** Let $\mathbb{A} = \{0, 1\}$, and $\Delta_0 := \lambda x.\text{out}_0(xx)$. The computation from $\langle \epsilon : \Delta_0\Delta_0 \rangle$ (with ϵ the empty string) produces a stream: a string of 0's whose length tends to infinity.

Infinite Normal Forms. Infinitary λ -calculi [9, 22, 23] model infinite structures in λ -calculi. Terms and reduction sequences need not be finite. An infinite reduction sequence is *strongly convergent* if the depth of the contracted redex tends to infinity. Based on different depth measures, in [23] eight different infinitary λ -calculi are developed. If the calculus is *confluent*, the *infinite normal form of a term N is unique*, and it is the *meaning of N* . Infinite normal forms are well-known in λ -calculus in the form of Böhm trees [7] or Lévy-Longo trees [28].

► **Example 3.** Let $\Delta_z := \lambda x.z(xx)$. In the (infinite) reduction sequence $\Delta_z\Delta_z \rightarrow_{\beta} z(\Delta_z\Delta_z) \rightarrow_{\beta} z(z(\Delta_z\Delta_z)) \rightarrow_{\beta} z(z(z(\Delta_z\Delta_z))) \dots$, the depth of the redex $\Delta_z\Delta_z$ tends to infinity. It is intuitively clear that $\Delta_z\Delta_z$ has an *infinite normal form* $z(z(z\dots))$.

Notation. From now on, we use the following standard notations: $I = \lambda x.x$, $\Delta = \lambda x.xx$, together with: $\Delta_{\oplus} = \lambda x.I \oplus (xx)$, $\Delta_c := \lambda x.\text{out}_c(xx)$, $\Delta_z := \lambda x.z(xx)$.

1.2 Motivations, and necessity, for non-deterministic evaluation

In this paper we are concerned with evaluation towards a limit. We allow the evaluation $\xrightarrow{\epsilon}$ (the normalizing strategy) to be *non-deterministic*. Let us discuss the motivations.

A programming language which is built on a λ -calculus implements a specific evaluation strategy $\xrightarrow{\epsilon}$ of the general reduction \rightarrow . The evaluation strategy $\xrightarrow{\epsilon}$ may or may not be deterministic, as long as *all choices eventually yield the same result*. Non-deterministic evaluation (written NDE) is a useful feature, which for example allows for parallel implementations, but in some cases is also a *necessity* and a *key reasoning tool*, as we discuss.

1. *NDE subsumes different evaluation policies.* A good illustration of this is in Plotkin's Call-by-Value λ -calculus, whose general reduction is \rightarrow_{β_v} . *Weak* evaluation (which does not reduce in the body of a function) evaluates closed terms to values. There are three main weak schemes (see Sect. 4.1): reducing left-to-right, as defined by Plotkin [30], right-to-left, as in Leroy's ZINC abstract machine [26], or in an arbitrary order. While left and right reduction are deterministic, weak reduction in arbitrary order is non-deterministic and *subsumes* both.
2. *NDE supports parallel/distributed implementation.* Non-deterministic evaluation does not define an abstract machine, but it includes *all possible parallel implementations*.

3. *NDE allows for breadth-first scheduling.* Left-to-right evaluation is inherently depth-first. NDE allows for breadth-first evaluation (favoring redexes at minimal depth), which is a necessity when the reduction graph is infinitary. An example comes from CbN λ -calculus. Thinking of Example 3, the terms $z(\Delta\Delta)(Iz)$ and $z(\Delta\Delta)(\Delta_z\Delta_z)$ do deliver more information than $z(\Delta\Delta)(\Delta\Delta)$. Their respective infinite normal forms are very different ($z\Omega z$ and $z\Omega(zz\dots)$, respectively). Still, for both left evaluation gets stuck at the leftmost redex, $\Delta\Delta$. Similar phenomena appear with effectful computation: in CbV λ -calculus with output of Example 2, the terms $(\Delta\Delta)(\text{out}_0(I))$ and $(\Delta\Delta)(\Delta_0\Delta_0)$ behave similarly to the previous terms. A breadth-first approach allows one to compute the “best” (in some sense) possible result across all settings, uniformly.
4. *NDE facilitates reasoning and proofs.* This point is highly relevant when dealing with complex calculi, such as a probabilistic λ -calculus. Two examples from the literature are [16] and [10] – in both cases moving from the usual deterministic head reduction to its non-deterministic variant (given in Sect. 4.1) is crucial to the results.

2 Quantitative Abstract Rewriting Systems

In this section we present Quantitative Abstract Rewriting Systems (QARS) [4, 14]. QARS are Abstract Rewriting Systems (ARS) enriched with a notion of observation, where we can formalize both finitary and asymptotic rewriting. We first recall some standard notions of rewriting (see [31] or [6]), in particular that of ARS and of *normalizing strategy*.

2.1 Basics in (Finitary) Rewriting

An *abstract rewriting system* (ARS) is a pair $(\mathcal{A}, \rightarrow)$ consisting of a set \mathcal{A} and a binary relation \rightarrow on \mathcal{A} whose pairs are written $t \rightarrow s$ and called *steps*. We denote \rightarrow^* (resp. $\rightarrow^=$, \rightarrow^+) the transitive-reflexive (resp. reflexive, transitive) closure of \rightarrow . We write $t \leftarrow u$ if $u \rightarrow t$. If $\rightarrow_1, \rightarrow_2$ are binary relations on \mathcal{A} then $\rightarrow_1 \cdot \rightarrow_2$ denotes their composition (*i.e.* $t \rightarrow_1 \cdot \rightarrow_2 s$ if there exists $u \in \mathcal{A}$ such that $t \rightarrow_1 u \rightarrow_2 s$). The relation \rightarrow is *confluent* if $\leftarrow^* \cdot \rightarrow^* \subseteq \rightarrow^* \cdot \leftarrow^*$. An element $u \in \mathcal{A}$ is \rightarrow -**normal**, or a \rightarrow -*normal form* (*nf*) if there is no t such that $u \rightarrow t$ (we also write $u \nrightarrow$).

A \rightarrow -**sequence** (or **reduction sequence**) from t is a possibly infinite sequence $t = t_0, t_1, t_2, \dots$ such that $t_i \rightarrow t_{i+1}$. Notice that $t \rightarrow^* s$ holds exactly when there is a *finite* sequence from t to s – we often write $t \rightarrow^* s$ to indicate a finite \rightarrow -sequence. A \rightarrow -sequence from t is *maximal* if it is either infinite or ends in a \rightarrow -*nf*. We write $\langle t_n \rangle_n$ to indicate a maximal \rightarrow -sequence from t_0 ; by convention, if $t_i = u \nrightarrow$ then $t_k = u$ for all $k \geq i$.

Normalization. In general, $t \in \mathcal{A}$ may or may not reduce to a normal form. And if it does, not all reduction sequences necessarily lead to normal form. $(\mathcal{A}, \rightarrow)$ is strongly (weakly, uniformly) normalizing if every $t \in \mathcal{A}$ is, where the normalization notions are as follows.

- t is *strongly* \rightarrow -normalizing: every maximal \rightarrow -sequence from t ends in a normal form;
- t is *weakly* \rightarrow -normalizing: there is a \rightarrow -sequence from t which ends in a normal form;
- t is *uniformly* \rightarrow -normalizing: t weakly \rightarrow -normalizing implies t strongly \rightarrow -normalizing.

Untyped λ -calculus is not strongly normalizing. How do we compute a normal form, or *test* if any exists? This problem is tackled by *normalizing strategies*. By repeatedly performing *only specific steps* \xrightarrow{e} , we are guaranteed that a normal form, if any, will eventually be computed.

A reduction \xrightarrow{e} is a *one-step* (resp. *multi-step*) *strategy* for \rightarrow if $\xrightarrow{e} \subseteq \rightarrow$ (resp. $\xrightarrow{e} \subseteq \rightarrow^+$), and it has the same normal forms as \rightarrow . It is a **normalizing strategy** for \rightarrow if, moreover, whenever t has a \rightarrow -normal form, then *every* maximal \xrightarrow{e} -sequence from t ends in a \rightarrow -normal form. Note that \rightarrow may not have the property of unique normal forms.

► **Remark 4.** A familiar example of calculus where terms may not have a unique normal form is Call-by-Name Weak λ -calculus (weak means no reduction under λ), studied by Abramsky and Ong [1]. The term $M = (\lambda xy.x)(II)$ has two distinct normal forms, $N_1 = \lambda y.II$ and $N_2 = \lambda y.I$. Weak head reduction is a normalizing strategy for it. However the strategy is not complete, in the sense that it produces the normal form N_1 , but it cannot reach N_2 .

A normalizing strategy \xrightarrow{e} need not be deterministic (a reduction \rightarrow is deterministic if for all $t \in \mathcal{A}$ there is at most one $s \in \mathcal{A}$ such that $t \rightarrow s$). However, \xrightarrow{e} is required to be *uniformly normalizing*, i.e., all reduction sequences from the same t have the same behavior.

A property of \xrightarrow{e} which guarantees uniform normalization is Newman's *Random Descent* (RD) [29]: for each $t \in \mathcal{A}$, all maximal sequences from t have the same length and – if it is finite – they all end in the same element. The following property suffices to establish it.

► **Fact 5 (Newman).** *If reduction \xrightarrow{e} is RD-diamond, then it has Random Descent, where RD-diamond: $(t_1 \xleftarrow{e} t \xrightarrow{e} t_2)$ implies $(t_1 = t_2 \text{ or } \exists u. t_1 \xrightarrow{e} u \xleftarrow{e} t_2)$.*

2.2 QARS

Ariola and Blom [4] have introduced the notion of Abstract Rewrite Systems with Information content (ARSI); a rewrite system is associated with a partial order that expresses the “information content” of the elements. ARSI however are tailored to infinite normal forms in the sense of Böhm and Levy-Longo trees: limits are there given by the ideal completion [3, Prop. 1.1.21] of the partial order. QARS [14] move from partial orders to ω -complete partial orders (ω -cpo) – this is enough to capture also effectful computation, such as the probabilistic one. We illustrate the key notions with several examples, including the calculi from Sect. 1.1.

Computation is a process that produces a result by gradually increasing the amount of available information – the standard structure to express a result in terms of partial information is that of an ω -cpo. Recall that a partially ordered set $\mathbb{S} = (\mathbb{S}, \leq)$ is an **ω -complete partial order (ω -cpo)** if every ω -chain $s_0 \leq s_1 \leq \dots$ has a supremum. We assume that \leq has a *least element* \perp . The elements of \mathbb{S} are denoted by bold letters $\mathbf{s}, \mathbf{p}, \mathbf{q}$.

Let $(\mathcal{A}, \rightarrow)$ be an ARS. With each $t \in \mathcal{A}$ is associated a notion of (partial) information, called *observation*, by means of a function from \mathcal{A} to an ω -cpo. Def. 6 formalizes this idea.

► **Definition 6 (QARS).** *A quantitative ARS (QARS) is an ARS $(\mathcal{A}, \rightarrow)$ with a function $\mathbf{obs}: \mathcal{A} \rightarrow \mathbb{S}$ (where \mathbb{S} is an ω -cpo) such that for all $t, s \in \mathcal{A}$, if $t \rightarrow s$ then $\mathbf{obs}(t) \leq \mathbf{obs}(s)$.*

Intuitively, the function \mathbf{obs} observes a specific property of interest about $t \in \mathcal{A}$, and indicates how much stable information t delivers: the information content is monotonically increasing during computation. Notice that \mathbf{obs} may take numerical values, but needs not.

► **Example 7.**

1. λ -calculus: let $\mathbb{S} = \{0 < 1\}$ and $\mathbf{obs}_n(t) = 1$ if t is normal, 0 otherwise.
2. Probabilistic λ -calculus: take $\mathbb{S} = ([0, 1], \leq_{\mathbb{R}})$, and for \mathbf{obs} the probability to be in normal form (we will formalize this in Sect. 5, see $\mathbf{obs}_{pn}(\mathbf{m})$ in Fig. 4.)
3. Infinitary λ -calculus: take $\mathbb{S} = \mathbb{N}^{\infty} = \mathbb{N} \cup \{\infty\}$ with the usual order, and for \mathbf{obs} the function which associates with any term t the minimal depth k of any redex in t .

► **Example 8** (Non-numerical obs).

1. λ -calculus: take for \mathbb{S} the flat order on normal forms, and define $\text{obs}_{\mathcal{N}}(u) = u$ if u is normal, $\text{obs}_{\mathcal{N}}(u) = \perp$ otherwise.
2. Probabilistic λ -calculus: take for \mathbb{S} the ω -cpo of the subdistributions on normal forms $\mathcal{D}(\mathcal{N})$ (we will formalize this in Sect. 5, see Fig. 4).
3. Infinitary λ -calculus: take the ω -cpo of the partial normal forms that are associated with λ -terms (see [3] page 52, and Appendix D.2).

Limits as Results. From now on, let $\mathcal{Q} = ((\mathcal{A}, \rightarrow), \text{obs})$ be an arbitrary but fixed QARS. By definition, given a \rightarrow -sequence $\langle t_n \rangle_n$, its *limit* $\sup_n \{\text{obs}(t_n)\}$ with respect to obs always exists, because \mathbb{S} is an ω -cpo. If \rightarrow is deterministic – hence any t has a unique maximal \rightarrow -sequence – it is standard to interpret the limit as the *meaning* of t . In a QARS, t has *several possible reduction sequences*, and so can produce several outcomes (limits). Following [13]:

► **Definition 9** (obs-limits). *Let $t \in \mathcal{A}$. We write*

- $t \rightarrow_{\text{obs}}^{\infty} \mathbf{p}$, if there exists a \rightarrow -sequence $\langle t_n \rangle_n$ from t whose limit $\sup_n \{\text{obs}(t_n)\} = \mathbf{p}$;
- $\text{Lim}_{\text{obs}}(t, \rightarrow)$ is the set $\{\mathbf{p} \mid t \rightarrow_{\text{obs}}^{\infty} \mathbf{p}\}$ of limits from t ;
- $\llbracket t \rrbracket$ denotes the greatest element of $\text{Lim}_{\text{obs}}(t, \rightarrow)$, if it exists.

The notations omit the subscript obs when the function obs is clear from the context.

Intuitively, $\llbracket t \rrbracket$ is well defined if different reduction sequences from t do not produce *essentially different* results: if $\mathbf{q} \neq \mathbf{p}$ then they both approximate a same result \mathbf{r} (i.e., $\mathbf{q}, \mathbf{p} \leq \mathbf{r}$).

Thinking of usual rewriting, consider $\text{obs}_{\mathcal{N}}$ as in Example 8, point 1: here to have a greatest limit exactly corresponds to uniqueness of normal forms.

► **Example 10.** Let us revisit Example 7 pointwise, using the same notations.

1. λ -calculus: consider $t = (\lambda x.z)(\Delta\Delta)$. This term has infinite possible \rightarrow_{β} -sequences. The set of limits w.r.t. obs_n contains two elements: $\text{Lim}_{\text{obs}_n}(t, \rightarrow_{\beta}) = \{0, 1\}$
2. Probabilistic λ -calculus: consider the term $I \oplus \Delta\Delta$. It has only one reduction sequence $\mathbf{m} = [I \oplus \Delta\Delta] \Rightarrow [\frac{1}{2}I, \frac{1}{2}\Delta\Delta] \Rightarrow [\frac{1}{2}I, \frac{1}{2}\Delta\Delta] \Rightarrow \dots$. Here $\text{Lim}_{\text{obs}_{pn}}(\mathbf{m}, \Rightarrow) = \{\frac{1}{2}\}$.
3. Infinitary λ -calculus: consider the reduction sequence in Example 3. The depth of the redex $(\Delta_z\Delta_z)$ tends to ∞ , which is the limit.

Note that maximal elements of $\text{Lim}_{\text{obs}}(t, \rightarrow)$ need not be maximal elements of \mathbb{S} . For instance, in Example 10.2, the term $I \oplus (\Delta\Delta)$ converges with probability $\frac{1}{2}$ (rather than 1). As a consequence, *the set of limits may or may not have maximal elements*. The fact that $\text{Lim}_{\text{obs}}(t, \rightarrow)$ may have a lub but not a maximum – similarly to \mathbb{N} in \mathbb{N}^{∞} or the real interval $[0, 1)$ – is also easy to realize.

Even if $\text{Lim}_{\text{obs}}(t, \rightarrow)$ has maximal elements, a greatest limit does not necessarily exist: different reduction sequences may lead to different limits. The probabilistic λ -calculus and the λ -calculus with output provide several natural examples. Point 2 in Example 11 below shows moreover that the set of limits is – in general – *uncountable*.

► **Example 11** (Output λ -calculus). Consider the calculus sketched in Example 2. Let $\text{Out}_{\mathbb{A}} = (\mathbb{A}^* \times \Lambda_{\text{out}}, \overrightarrow{\text{w}})$, where reduction is *CbV and weak*, with the obvious definitions. Let \mathbb{S} be the ω -cpo of strings, and let $\text{obs}(\langle \mathbf{s} : M \rangle) = \mathbf{s}$. Clearly, $(\text{Out}_{\mathbb{A}}, \text{obs})$ is a QARS.

1. Let $\mathbf{m} = \langle \epsilon : \text{out}_0(I)\text{out}_1(I) \rangle$. $\text{Lim}_{\text{obs}}(\mathbf{m}, \overrightarrow{\text{w}})$ contains two limits, 10 and 01, both maximal, because $\mathbf{m} \overrightarrow{\text{w}} \langle 0 : I\text{out}_1(I) \rangle \overrightarrow{\text{w}} \langle 10 : II \rangle$, but also $\mathbf{m} \overrightarrow{\text{w}} \langle 1 : \text{out}_0(I)I \rangle \overrightarrow{\text{w}} \langle 01 : II \rangle$.
2. Let $\mathbf{m}' = \langle \epsilon : M' \rangle$ for $M' = (\Delta_0\Delta_0)(\Delta_1\Delta_1)$. This produces all possible sequences on the alphabet $\{0, 1\}$. So $\text{Lim}_{\text{obs}}(\mathbf{m}', \overrightarrow{\text{w}})$ has uncountable many elements, all maximal.

We are interested in the case when a greatest limit exists. The reason is that if $\text{Lim}_{\text{obs}}(t, \rightarrow)$ has a sup $\mathbf{s} \in \mathbb{S}$ which does not belong to $\text{Lim}_{\text{obs}}(t, \rightarrow)$, no reduction sequence converges to \mathbf{s} ; that is, we cannot compute \mathbf{s} internally to the calculus.

3 Strategies and Asymptotic Normalization

The question of whether the result $\llbracket t \rrbracket$ of computing an element t is well defined is natural. Equally natural is to wonder if there is a strategy that is guaranteed to compute $\llbracket t \rrbracket$. These two questions are at the core of this section. The existence of unique normal forms is independent of that of a normalizing strategy (see Remark 4). However, the computationally interesting case is (often) when both hold, so we will focus on this case.

We say that a reduction $\rightarrow_{\mathcal{e}} \subseteq \rightarrow$ is (asymptotically) normalizing if *each* $\rightarrow_{\mathcal{e}}$ -sequence from a given t converges *maximally*. We decompose this property in two properties: completeness and uniformity, which we discuss after the formal definition.

► **Definition 12** (Asymptotic properties). *Given a QARS $((\mathcal{A}, \rightarrow), \text{obs})$, a subreduction $\rightarrow_{\mathcal{e}} \subseteq \rightarrow$ is **asymptotically normalizing** for \rightarrow (or **obs-normalizing**) if it is both asymptotically complete and uniform, where*

1. $\rightarrow_{\mathcal{e}}$ is **asymptotically complete** (or **obs-complete**) if

$$(\forall t \in \mathcal{A}) : t \rightarrow_{\text{obs}}^{\infty} \mathbf{q} \text{ implies } t \rightarrow_{\mathcal{e}}^{\infty} \mathbf{p} \text{ for some } \mathbf{p} \text{ such that } \mathbf{q} \leq \mathbf{p};$$

2. $\rightarrow_{\mathcal{e}}$ is **asymptotically uniform** (or **obs-uniform**) if

$$(\forall t \in \mathcal{A}) : \text{all elements in } \text{Lim}_{\text{obs}}(t, \rightarrow_{\mathcal{e}}) \text{ are maximal in } \text{Lim}_{\text{obs}}(t, \rightarrow).$$

All definitions adapt to $\rightarrow_{\mathcal{e}}$ multistep subreduction of \rightarrow .

Let us discuss all components, comparing with their ARS analog.

- *Completeness* guarantees that the strategy $\rightarrow_{\mathcal{e}}$ is as good as \rightarrow in the amount of information it produces.
- *Completeness is not enough*: an asymptotically complete strategy is not guaranteed to find a/the “best” result: in Sect. 5.1 we will study a reduction $\xrightarrow{\mathbb{E}}$ which is complete, but need not converge to the greatest limit (Remark 25). Let us first see a classical example.

► **Example 13.** In the usual λ -calculus (as in Example 10.1), the term $M = (\lambda x.I)(\Delta\Delta)$ has a \rightarrow_{β} -sequence which reaches I , and a diverging one. The leftmost-outermost strategy always produces I (it is complete *and* normalizing). Notice that \rightarrow_{β} is *trivially a complete strategy* for \rightarrow_{β} , but it is not normalizing, because M has a diverging \rightarrow_{β} -sequence. Indeed, \rightarrow_{β} is *complete, but not uniform*.

- *Asymptotic uniformity* expresses that all $\rightarrow_{\mathcal{e}}$ -sequences from a term behave the same way. This corresponds to the ARS notion of *uniform normalization*: the reduction sequences from a term either all diverge, or all terminate (not necessarily in the same normal form).
- *Normalizing strategies*. If we consider usual ARS, and assume **obs** as in Example 7.1, expressing whether t is or is not normal, then a strategy for \rightarrow that is **obs-normalizing** is exactly a normalizing strategy for \rightarrow in the usual sense.

If $\xrightarrow{e} \subseteq \rightarrow$ is *obs-complete*, then $\text{Lim}_{\text{obs}}(t, \rightarrow)$ has maximal elements (resp. a greatest element) if and only if $\text{Lim}_{\text{obs}}(t, \xrightarrow{e})$ does. So we can reduce testing such properties for \rightarrow , to testing the same properties for \xrightarrow{e} , which is often simpler to study. In particular, if we are able to find a reduction $\xrightarrow{e} \subseteq \rightarrow$ which is complete and moreover has a unique limit, then necessarily \rightarrow has a greatest limit. That is, we can simultaneously answer both of our questions: whether $\llbracket t \rrbracket$ is well defined, and if some strategy is guaranteed to compute it.

► **Proposition 14** (Main, abstractly). *If the following hold*

- i. \xrightarrow{e} is asymptotically complete for \rightarrow ;
- ii. $\text{Lim}_{\text{obs}}(t, \xrightarrow{e})$ contains a unique element (i.e. $\text{Lim}_{\text{obs}}(t, \xrightarrow{e}) = \{\mathbf{p}\}$, for some \mathbf{p}).

Then: (1.) $\llbracket t \rrbracket$ is defined, and (2.) $t \xrightarrow{e}_{\text{obs}}^{\infty} \llbracket t \rrbracket$, for each \xrightarrow{e} -sequence.

Notice that condition (ii.) means that *all* \xrightarrow{e} -sequences from the term t have the *same* limit.

► **Remark 15** (Asymptotically normalizing strategies). If a QARS is such that $\llbracket t \rrbracket$ is defined for each t , then the two notions – to be an *obs-normalizing strategy* and to satisfy the conditions in Prop. 14 – coincide. Indeed, any *obs-normalizing strategy* for \rightarrow , if it exists, is forced to have a *unique* limit, that is, $\text{Lim}_{\text{obs}}(t, \xrightarrow{e}) = \{\llbracket t \rrbracket\}$.

3.1 A proof technique for Asymptotic Normalization

The two conditions in Prop. 14 give a method to prove normalization. The crucial step is to prove asymptotic completeness. Remarkably, as we show in this section, this can be reduced to prove a *finitary* property (*factorization*) and an elementary one-step test (*neutrality*).

The other condition in Prop. 14, namely uniqueness of limits, is trivial if the strategy is deterministic. Otherwise, random descent (opportunistically formulated [14]) is a property that guarantees it, and that can also be established via a local test, as we recall below. While it is only a sufficient criterion, it often suffices to deal with non-deterministic evaluation strategies in λ -calculus, and in particular it suffices to deal with strategies in probabilistic λ -calculus.

Asymptotic Completeness via Factorization. The following theorem assumes a partition of the \rightarrow -steps into two classes: essential steps \xrightarrow{e} and internal steps \xrightarrow{e} . Point (i) states that every sequence \rightarrow^* factorizes into a \xrightarrow{e} -sequence followed by a \xrightarrow{e} -sequence. Point (ii) states that the internal steps \xrightarrow{e} do not increase the information content.

► **Theorem 16** (Asymptotic completeness criterion). *Given $((A, \rightarrow), \text{obs})$ a QARS, and a subrelation $\xrightarrow{e} \subseteq \rightarrow$, assume :*

- i. *e-factorization: if $t \rightarrow^* u$ then $t \xrightarrow{e}^* \cdot \xrightarrow{e}^* u$;*
- ii. *\neg e-neutrality: $t \xrightarrow{e} s$ implies $\text{obs}(t) = \text{obs}(s)$.*

Then: $t \xrightarrow{e}_{\text{obs}}^{\infty} \mathbf{p}$ implies $t \xrightarrow{\text{obs}}^{\infty} \mathbf{p}$.

Proof. Let $\langle t_n \rangle_n$ be a \rightarrow -sequence such that $t = t_0$ and $\sup_n \{\text{obs}(t_n)\} = \mathbf{p}$. From t , we inductively build a \xrightarrow{e} -sequence $\langle s_n \rangle_n$ with $s_0 = t$ and such that, for every $k \in \mathbb{N}$, there is an index $j(k)$ such that $t \xrightarrow{e}^* s_{j(k)}$ and $s_{j(k)} \xrightarrow{e}^* t_k$. Case $k = 0$ is trivial (set $s_{j(0)} := t$).

Assume the claim holds for $k \geq 0$, so $t \xrightarrow{e}^* s_{j(k)}$. Observe that we have a sequence $s_{j(k)} \xrightarrow{e}^* t_k \rightarrow t_{k+1}$. By applying assumption (i.) to it, we have $s_{j(k)} \xrightarrow{e}^* u \xrightarrow{e}^* t_{k+1}$. We concatenate $t \xrightarrow{e}^* s_{j(k)}$ and $s_{j(k)} \xrightarrow{e}^* u$ to obtain $t \xrightarrow{e}^* s_{j(k)} \xrightarrow{e}^* s_{j(k+1)} := u$, as desired. By assumption (ii.), $s_{j(k)} \xrightarrow{e}^* t_k$ implies $\text{obs}(t_k) = \text{obs}(s_{j(k)})$. The claim easily follows. ◀

Uniqueness of the limit via Random Descent. To establish that a strategy has a unique limit, Random Descent [29, 32, 33] has already been shown to adapt well and naturally in a probabilistic and asymptotic setting [13, 14].

The property **obs-RD** below states that if t has different reduction sequences, they are all *indistinguishable* if regarded through the lenses of **obs**. Namely, all reduction sequences $\langle t_n \rangle_n$ starting from t induce the same ω -chain $\langle \text{obs}(t_n) \rangle_n$. Thus, they all have the same **obs**-limit.

► **Definition 17** (Weighted Random Descent). *Let $((\mathcal{A}, \rightarrow), \text{obs})$ be a QARS. The relation $\xrightarrow{e} \subseteq \rightarrow$ satisfies the following properties if they hold for each $t \in \mathcal{A}$.*

1. **obs-RD**: for each pair of \xrightarrow{e} -sequences $\langle r_n \rangle_n, \langle s_n \rangle_n$ from t , $\text{obs}(r_n) = \text{obs}(s_n)$ for all n .
2. **obs-diamond**: \xrightarrow{e} satisfies RD-diamond, and if $t \xleftarrow{e} m \xrightarrow{e} s$ then $\text{obs}(s) = \text{obs}(t)$.

► **Proposition 18** ([14]). *With the same notation as in Def. 17: $(\text{obs-diamond}) \Rightarrow (\text{obs-RD}) \Rightarrow \text{Lim}_{\text{obs}}(t, \xrightarrow{e})$ contains a unique element.*

► **Example 19** (CbV Weak reduction). Let us consider Call-by-Value λ -calculus with weak reduction \xrightarrow{w} , where weak means no reduction in the scope of λ -abstractions. The following are two different \xrightarrow{w} -sequences from the term $(II)(Ix)$:

$$(II)(Ix) \xrightarrow{w} I(Ix) \xrightarrow{w} Ix \xrightarrow{w} x \quad \text{and} \quad (II)(Ix) \xrightarrow{w} (II)x \xrightarrow{w} Ix \xrightarrow{w} x.$$

The observations of interest are values. Let $\text{obs}_v : \Lambda \rightarrow \{0, 1\}$ be 1 if the term is a *value* (i.e. a variable or an abstraction), 0 otherwise. Through the lenses of obs_v , both sequences appear as $\langle 0, 0, 0, 1 \rangle$.

4 Normalization in CbV and CbN λ -calculi

In the rest of the paper, we study asymptotic normalization in the setting of λ -calculi – in particular we are interested in probabilistic λ -calculus (Sect. 5).

In this section, after recalling the general syntax of λ -calculus, we define a novel, flexible normalizing strategy, which is uniformly defined for *Call-by-name* (CbN) and *Call-by-Value* (CbV) λ -calculi. Its features – in particular the fact that it support breadth-first reduction – make it suitable to then be extended to asymptotic normalization, in different settings.

4.1 Call-by-Name and Call-by-Value (applied) λ -calculus

We recall the basics of λ -calculus. Our syntax admits *operator symbols* [20, 30], i.e. *constants* with a fixed arity for their arguments. *Terms* and *values* are defined by the grammars below.

$$\begin{aligned} M &::= x \mid \lambda x.M \mid MM \mid \mathbf{o}(M, \dots, M) && (\text{Terms}, \Lambda_{\mathcal{O}}) \\ V &::= x \mid \lambda x.M && (\text{Values}, \mathcal{V}) \end{aligned}$$

where x ranges over a countable set of *variables*, and \mathbf{o} over a disjoint (possibly empty) set \mathcal{O} of operator symbols. If \mathcal{O} is empty, the calculus is *pure* and we set $\Lambda := \Lambda_{\mathcal{O}}$. Terms are identified up to renaming of bound variables, where λx is the only binder constructor. $P\{Q/x\}$ is the capture-avoiding substitution of Q for the free occurrences of x in P .

Contexts (with an hole $(\)$) are defined by the grammar below. $\mathbf{C}(N)$ stands for the term obtained from \mathbf{C} by replacing the hole with N (possibly capturing the free variables of N).

$$\mathbf{C} ::= (\) \mid M\mathbf{C} \mid \mathbf{C}M \mid \lambda x.\mathbf{C} \mid \mathbf{o}(M, \dots, \mathbf{C}, \dots, M) \quad (\text{Contexts})$$

Rules and Reductions. A rule ρ is a binary relation on $\Lambda_{\mathcal{O}}$, which we also denote \mapsto_{ρ} , writing $R \mapsto_{\rho} R'$. R is called a ρ -redex. The best known rule is β : $(\lambda x.M)N \mapsto_{\beta} M\{N/x\}$.

A **reduction step** \rightarrow_{ρ} is the closure under context \mathbf{C} of ρ .

CbN and CbV Calculi. The (pure) **Call-by-Name** calculus $\Lambda^{\text{cbn}} = (\Lambda, \rightarrow_{\beta})$ is the set of terms equipped with the contextual closure of the β -rule, as described *e.g.* in [7]. The (pure) **Call-by-Value** calculus $\Lambda^{\text{cbv}} = (\Lambda, \rightarrow_{\beta_v})$ is the same set equipped with the contextual closure of the β_v -rule: $(\lambda x.M)V \mapsto_{\beta_v} M\{V/x\}$ where $V \in \mathcal{V}$, as introduced by Plotkin [30].

CbN and CbV applied calculi are obtained by associating to operators (the contextual closure of) a family of rules of the form $\mathbf{o}(M_1, \dots, M_k) \mapsto_{\mathbf{o}} N$. This is a standard way to enrich λ -calculus with new computational features, such as probabilistic choice or output.

Weak reductions in CbV. In CbV λ -calculus, various restrictions of \rightarrow_{β_v} are studied. If the result of interest are values, the reduction is *weak*, that is, it does not reduce in the body of a function. There are three main weak schemes: left, right and in arbitrary order. *Left* contexts \mathbf{L} , *right* contexts \mathbf{R} , and (arbitrary order) *weak* contexts \mathbf{W} are defined by

$$\mathbf{L} ::= () \mid \mathbf{L}M \mid \mathbf{V}\mathbf{L} \quad \mathbf{R} ::= () \mid \mathbf{M}\mathbf{R} \mid \mathbf{R}\mathbf{V} \quad \mathbf{W} ::= () \mid \mathbf{W}M \mid \mathbf{M}\mathbf{W}$$

Given a rule \mapsto on Λ , *weak reduction* $\xrightarrow{\mathbf{W}}$ is the closure of \mapsto under context \mathbf{W} . A step $T \rightarrow S$ is *non-weak*, noted $T \xrightarrow{\text{nw}} S$ if it is not weak. Similarly for left ($\xrightarrow{\mathbf{L}}$ and $\xrightarrow{\mathbf{R}}$), and right ($\xrightarrow{\mathbf{R}}$ and $\xrightarrow{\mathbf{L}}$). Left and right reduction are *deterministic*. Reduction $\xrightarrow{\mathbf{W}\beta_v}$ *subsumes* both. The choice of a redex is non-deterministic, but irrelevant w.r.t. reaching a value and the number of steps to do so, because $\xrightarrow{\mathbf{W}\beta_v}$ is RD-diamond (Fact 5). We can fire any arbitrary redex in weak position – or *all of them* in parallel. A *parallel variant* can easily be defined.

Weak factorization holds for the three reductions: $\rightarrow_{\beta_v}^* \subseteq \xrightarrow{\mathbf{L}\beta_v}^* \cdot \xrightarrow{\mathbf{R}\beta_v}^*$, for $s \in \{\mathbf{w}, \mathbf{l}, \mathbf{r}\}$.

Head reduction in CbN. Head reduction [7] is the closure of β under head context $\lambda x_1 \dots x_n. () \mid M_1 \dots M_k$. *Head normal forms (hnf)*, whose set is denoted by \mathcal{H} , are its normal forms. The literature of linear logic often uses a variant of head context which includes the standard one, and induces exactly the same set \mathcal{H} of normal forms. Given a rule ρ , we write $\xrightarrow{\mathbf{H}\rho}$ for its closure under context \mathbf{H} .

$$\mathbf{H} ::= () \mid \lambda x.\mathbf{H} \mid \mathbf{H}M \quad (\text{Head contexts})$$

Head factorization (see [7, Lemma 11.4.6]) and head normalization (see [7, Thm. 8.3.11]) are classical results, which hold also when the calculus includes constants, *i.e.* for $(\Lambda_{\mathcal{O}}, \rightarrow_{\beta})$.

4.2 A strategy for finitary normalization in CbV and CbN λ -calculus

We revisit normalization for λ -calculus – uniformly for CbV and CbN – and define a strategy which is well-suited to be extended to probabilistic λ -calculi, and to asymptotic normalization. It supports non-deterministic head and weak reduction (as needed in the probabilistic case) and breadth-first evaluation of redexes (as needed to deal with infinitary reduction graphs).

We call *surface reduction* weak reduction in CbV and head reduction in CbN, because they only fire redexes at *depth 0*, where in CbV the depth of a redex R is the number of *abstractions* in which R is nested, and in CbN is the number of arguments. Normal forms for β and β_v can be computed by iterating surface reduction in a suitable way, as we show below.

17:12 Strategies for Asymptotic Normalization

Normalizing strategies. In Λ^{cbn} , a paradigmatic normalizing strategy is leftmost-outermost reduction. It can be described as: first apply head reduction $\xrightarrow{\mathfrak{h}}_{\beta}$ until *hnf*, and then iterate the process, in left-to-right order. Normalization in Λ^{cbv} is less established: one can iterate $\xrightarrow{\mathfrak{r}}_{\beta_v}$ left to right (as in Plotkin's standard reduction [30]), but also iterate $\xrightarrow{\mathfrak{r}}_{\beta_v}$ right to left, as in Grégoire and Leroy's implementation [19]. In all cases, once a head or weak normal form is reached (think of $xM_1 \dots M_k$ in CbN) no interaction is possible among the subterms M_i, \dots, M_k , so in fact the process can be iterated in any *arbitrary order*.

We define a rather liberal normalizing strategy, *uniformly* for CbN and CbV, and *parametrically* in the choice of surface reduction. Unlike leftmost-outermost reduction, which is sequential and *inherently depth-first*, the *unbiased* reduction $\xrightarrow{\mathfrak{v}}$ is non-deterministic in the choice of the outermost redex, and can support a breadth-first reduction policy. It persistently performs surface steps, as long as it is possible, and then iterates the process in the subterms, in arbitrary order.

► **Definition 20** (Unbiased iteration of surface reduction). *Given $(\Lambda_{\mathcal{O}}, \rightarrow)$, where \rightarrow is the contextual closure of a rule $\mathfrak{b} \in \{\beta, \beta_v\}$, let $\xrightarrow{\mathfrak{s}} \subseteq \rightarrow$ be as follows:*

$$\xrightarrow{\mathfrak{s}} = \xrightarrow{\mathfrak{h}} \text{ if } \mathfrak{b} = \beta \text{ (CbN)} \quad \xrightarrow{\mathfrak{s}} \in \{\xrightarrow{\mathfrak{w}}, \xrightarrow{\mathfrak{r}}, \xrightarrow{\mathfrak{r}'}\} \text{ if } \mathfrak{b} = \beta_v \text{ (CbV)} .$$

The relation $\xrightarrow{\mathfrak{v}} \subseteq \rightarrow$ is inductively defined as follows:

- if $M \xrightarrow{\mathfrak{s}} M'$ then $M \xrightarrow{\mathfrak{v}} M'$;
- if $M \not\xrightarrow{\mathfrak{s}}$ then $M \xrightarrow{\mathfrak{v}} M'$ is defined according the rules below.

$$\frac{P \xrightarrow{\mathfrak{v}} P'}{(\lambda x.P) \xrightarrow{\mathfrak{v}} (\lambda x.P')} \quad \frac{P \xrightarrow{\mathfrak{v}} P' \quad Q \xrightarrow{\mathfrak{v}} Q'}{PQ \xrightarrow{\mathfrak{v}} P'Q'} \quad \frac{P_i \xrightarrow{\mathfrak{v}} P'_i}{\mathfrak{o}(P_1, \dots, P_i, \dots, P_k) \xrightarrow{\mathfrak{v}} \mathfrak{o}(P_1, \dots, P'_i, \dots, P_k)}$$

The same definition of $\xrightarrow{\mathfrak{v}} \subseteq \rightarrow$ still applies if \rightarrow is the contextual closure of $\mapsto_{\mathfrak{b}} \cup \mapsto_{\rho}$, i.e. of the rule $\mapsto_{\mathfrak{b}}$ extended with some other rule \mapsto_{ρ} on $\Lambda_{\mathcal{O}}$.

We study $\xrightarrow{\mathfrak{v}}_{\mathfrak{b}}$. It is RD-diamond (see Fact 5) and is a normalizing strategy for both CbN and CbV λ -calculi. Note that in CbN, $\xrightarrow{\mathfrak{v}}_{\beta}$ *subsumes* usual *leftmost-outermost reduction*.

► **Proposition 21** (U-Factorization). *Let $\mathfrak{b} \in \{\beta, \beta_v\}$.*

$$M \rightarrow_{\mathfrak{b}}^* N \text{ implies } M \xrightarrow{\mathfrak{v}}_{\mathfrak{b}}^* \cdot \xrightarrow{\mathfrak{v}}_{\mathfrak{b}}^* N \quad (\text{U-Factorization})$$

► **Proposition 22.** *With the same assumptions as in Def. 20, let $\mathfrak{b} \in \{\beta, \beta_v\}$. Then:*

1. $\xrightarrow{\mathfrak{v}}_{\mathfrak{b}}$ is RD-diamond.
2. $\xrightarrow{\mathfrak{v}}_{\mathfrak{b}}$ has the same normal forms as $\rightarrow_{\mathfrak{b}}$.
3. Let N be \mathfrak{b} -normal. $M \rightarrow_{\mathfrak{b}}^* N$ implies $M \xrightarrow{\mathfrak{v}}_{\mathfrak{b}}^* N$.

Normalization for both CbN and CbV follows from the points above.

► **Theorem 23** (Normalization). *For $\mathfrak{b} \in \{\beta, \beta_v\}$, $\xrightarrow{\mathfrak{v}}_{\mathfrak{b}}$ is a normalizing strategy for $\rightarrow_{\mathfrak{b}}$.*

Depth-first vs Breadth-first. Leftmost-outermost reduction fires redexes in a depth-first way. Instead, $\xrightarrow{\mathfrak{v}}$ evaluates in a breadth-first style, which is more suitable to deal with possibly infinitary reductions. For example, in CbN think of $z(\Delta\Delta)(\Delta_z\Delta_z)$. Leftmost-outermost reduction never leaves the redex $\Delta\Delta$, while $\xrightarrow{\mathfrak{v}}$ can also fire $(\Delta_z\Delta_z)$ yielding $z(z(\dots))$.

A parallel variant. Once a term is $\xrightarrow{\mathfrak{s}}$ -normal, the process can be iterated in any arbitrary order, or in *parallel*. Parallel (multi-step) reduction $\xrightarrow{\mathfrak{U}}$ is easily defined (Appendix B.1).

5 Probabilistic λ -calculi and Asymptotic Normalization

A standard way to model *probabilistic choice* (a fair coin) is by means of a binary operator \oplus . We write $M \oplus N$ for $\oplus(M, N)$. Intuitively, $M \oplus N$ reduces to either M or N , *with equal probability* $\frac{1}{2}$. Reduction is then defined not simply on terms but on (monadic) structures representing probability distributions over terms. Here we follow [16], which defines both a CbV and a CbN calculus $\Lambda_{\oplus}^{\text{cbv}}$ and $\Lambda_{\oplus}^{\text{cbn}}$, where β or β_v reduction are “as usual”, so if a term contains no probabilistic operator, it behaves the same as in the usual λ -calculus (*i.e.* the extension is *conservative*). Probabilistic reduction instead needs to be constrained in order to have good properties such as confluence (see [16], and [12, 25] for a discussion of the issues).

Discrete Probability Distributions. Given a countable set Ω , a function $\mu: \Omega \rightarrow [0, 1]$ is a probability *subdistribution* if $\|\mu\| := \sum_{\omega \in \Omega} \mu(\omega) \leq 1$ (a *distribution* if $\|\mu\| = 1$). Subdistributions allow us to deal with partial results. We write $\mathcal{D}(\Omega)$ for the set of subdistributions on Ω , equipped with the pointwise order on functions: $\mu \leq \rho$ if $\mu(\omega) \leq \rho(\omega)$ for all $\omega \in \Omega$. $\mathcal{D}(\Omega)$ has a bottom element (the subdistribution $\mathbf{0}$) and maximal elements (all distributions).

Multi-distributions. We use multi-distributions [5] to syntactically represent distributions, A *multi-distribution* $\mathfrak{m} = [p_i M_i]_{i \in I}$ on the set of terms $\Lambda_{\mathcal{O}}$ is a finite multiset of pairs of the form pM , with $p \in]0, 1]$, $M \in \Lambda_{\mathcal{O}}$, and $\sum_i p_i \leq 1$. The set of all multi-distributions on $\Lambda_{\mathcal{O}}$ is $\mathcal{M}(\Lambda_{\mathcal{O}})$. The sum of multi-distributions is noted $+$. The product $q \cdot \mathfrak{m}$ of a scalar q and a multi-distribution \mathfrak{m} is defined pointwise $q[p_i M_i]_{i \in I} := [(qp_i)M_i]_{i \in I}$. We write $[M]$ for $[1M]$.

Syntax. *Terms* (Λ_{\oplus}) and *values* are as in Sect. 4.1, with the operator \mathbf{o} being here \oplus .

Call-by-Value. The calculus $\Lambda_{\oplus}^{\text{cbv}}$ is the rewrite system $(\mathcal{M}(\Lambda_{\oplus}), \Rightarrow)$ where $\mathcal{M}(\Lambda_{\oplus})$ is the set of *multi-distributions* on Λ_{\oplus} and the relation $\Rightarrow \subseteq \mathcal{M}(\Lambda_{\oplus}) \times \mathcal{M}(\Lambda_{\oplus})$ is defined in Fig. 1 and Fig. 2. First, define one-step reductions from terms to multi-distributions – so for example, $M \oplus N \rightarrow [\frac{1}{2}M, \frac{1}{2}N]$. Then, lift the definition of reduction to a binary relation on $\mathcal{M}(\Lambda_{\oplus})$, in the natural way – for instance $[\frac{1}{2}(\lambda x.x)z, \frac{1}{2}(M \oplus N)] \Rightarrow [\frac{1}{2}z, \frac{1}{4}M, \frac{1}{4}N]$. Precisely:

1. The reductions $\rightarrow_{\beta_v}, \rightarrow_{\oplus} \subseteq \Lambda_{\oplus} \times \mathcal{M}(\Lambda_{\oplus})$ are defined in Fig. 1. *Contexts \mathbf{C} and \mathbf{W}* are as in Sect. 4.1. Note that β_v is closed under arbitrary context, while the \oplus rule – probabilistic choice – is closed under weak contexts \mathbf{W} (no reduction in the scope of λ or \oplus). We write $\xrightarrow{\mathfrak{s}}_{\beta_v}$ for the closure of β_v under context \mathbf{W} . The relation \rightarrow is $\rightarrow_{\beta_v} \cup \rightarrow_{\oplus}$. *Surface reduction* is $\xrightarrow{\mathfrak{s}} = \xrightarrow{\mathfrak{s}}_{\beta_v} \cup \rightarrow_{\oplus}$. A \rightarrow -step which is not surface is noted $\xrightarrow{\mathfrak{s}}$.
2. The lifting of a relation $\rightarrow_r \subseteq \Lambda_{\oplus} \times \mathcal{M}(\Lambda_{\oplus})$ to a reduction on multi-distributions is defined in Fig. 2. In particular, $\rightarrow, \rightarrow_{\beta_v}, \rightarrow_{\oplus}, \xrightarrow{\mathfrak{s}}, \xrightarrow{\mathfrak{s}}$ lift to $\Rightarrow, \Rightarrow_{\beta_v}, \Rightarrow_{\oplus}, \xRightarrow{\mathfrak{s}}, \xRightarrow{\mathfrak{s}}$.

A term M is \rightarrow -**normal** if there is no \mathfrak{m} such that $M \rightarrow \mathfrak{m}$. We also write $M \not\rightarrow$. We denote by \mathcal{N}_v the set of the **normal forms** of $\rightarrow = (\rightarrow_{\beta_v} \cup \rightarrow_{\oplus})$.

Call-by-Name. The calculus $\Lambda_{\oplus}^{\text{cbn}}$ is defined in a similar way, by replacing β_v with β and weak contexts with head contexts \mathbf{H} (as defined in Sect. 4.1).

$$\begin{array}{l}
 \mathbf{C}((\lambda x.M)V) \rightarrow_{\beta_v} [\mathbf{C}(M\{V/x\})] \\
 \mathbf{W}(M \oplus N) \rightarrow_{\oplus} [\frac{1}{2}\mathbf{W}(M), \frac{1}{2}\mathbf{W}(N)] \\
 \rightarrow := \rightarrow_{\beta_v} \cup \rightarrow_{\oplus} \quad \xrightarrow{s} := \xrightarrow{s}_{\beta_v} \cup \xrightarrow{s}_{\oplus}
 \end{array}$$

■ **Figure 1** \rightarrow -steps for the calculus $\Lambda_{\oplus}^{\text{cbv}}$.

$$\frac{M \rightarrow \mathfrak{m}}{[M] \Rightarrow [M]} \quad \frac{M \rightarrow \mathfrak{m}}{[M] \Rightarrow \mathfrak{m}} \quad \frac{([M_i] \Rightarrow \mathfrak{m}_i)_{i \in I}}{[p_i M_i \mid i \in I] \Rightarrow \sum_{i \in I} p_i \cdot \mathfrak{m}_i}$$

■ **Figure 2** Lifting of \rightarrow .

$$\frac{M \not\rightarrow}{[M] \not\Rightarrow [M]} \quad \frac{M \rightarrow \mathfrak{m}}{[M] \not\Rightarrow \mathfrak{m}} \quad \frac{([M_i] \not\Rightarrow \mathfrak{m}_i)_{i \in I}}{[p_i M_i \mid i \in I] \not\Rightarrow \sum_{i \in I} p_i \cdot \mathfrak{m}_i}$$

■ **Figure 3** Full lifting of \rightarrow .

$$\begin{array}{l}
 \text{obs}_{\mathcal{N}}: \mathcal{M}(\Lambda_{\oplus}) \rightarrow \mathcal{D}(\mathcal{N}_v) \quad [p_i M_i]_{i \in I} \mapsto \mu \quad \text{where } \forall N \in \mathcal{N}_v, \mu(N) = \sum_{i \in I} p_i \text{ s.t. } M_i = N \\
 \text{obs}_{pn}: \mathcal{M}(\Lambda_{\oplus}) \rightarrow [0, 1] \quad [p_i M_i]_{i \in I} \mapsto \|\mu\|
 \end{array}$$

■ **Figure 4** CbV observations on multi-distributions.

Observations on multi-distributions. In **CbV**, events of interest are the set \mathcal{V} of values and the set \mathcal{N}_v of \rightarrow -normal forms (for $\rightarrow = \rightarrow_{\beta_v} \cup \rightarrow_{\oplus}$). Focusing on \mathcal{N}_v , we can define:

- $\text{obs}_{\mathcal{N}}$ extracts from $\mathfrak{m} = [p_i M_i]_{i \in I}$ a *subdistribution* μ over normal forms. For example, if $\mathfrak{m} = [\frac{1}{4}\mathbf{T}, \frac{1}{8}\mathbf{T}, \frac{1}{4}\mathbf{F}, \frac{1}{4}\mathbf{II}]$, $\text{obs}_{\mathcal{N}}(\mathfrak{m})$ is the subdistribution $\{\mathbf{T}^{\frac{3}{8}}, \mathbf{F}^{\frac{1}{4}}\}$, i.e. $\mu(\mathbf{T}) = \frac{3}{8}$, $\mu(\mathbf{F}) = \frac{1}{4}$.
- obs_{pn} observes the probability that \mathfrak{m} has reached a normal form. For example, with \mathfrak{m} as above, $\text{obs}_{pn}(\mathfrak{m}) = \frac{5}{8}$.

In **CbN**, events of interest are the set of normal forms (w.r.t. $\rightarrow_{\beta} \cup \rightarrow_{\oplus}$), and the set \mathcal{H} of head normal forms. The corresponding observations are defined in the obvious way.

5.1 Asymptotic Normalization for Probabilistic λ -Calculi

We can now revisit the probabilistic calculi $\Lambda_{\oplus}^{\text{cbv}}$ and $\Lambda_{\oplus}^{\text{cbn}}$ as QARS, and define for them an asymptotically normalizing strategy. We develop explicitly only the CbV case, but similar definitions and results hold for CbN, taking into account that \rightarrow_{β_v} is replaced by \rightarrow_{β} and surface reduction is $\xrightarrow{\text{h}}$. Method and proofs are exactly the same.

The QARS framework allows us to express and analyze the asymptotic behaviour of the calculus $\Lambda_{\oplus}^{\text{cbv}} = (\mathcal{M}(\Lambda_{\oplus}), \Rightarrow)$. Here we are interested in $\text{obs}_{\mathcal{N}}: \mathcal{M}(\Lambda_{\oplus}) \rightarrow \mathcal{D}(\mathcal{N})$ as defined in Fig. 4. It is immediate that $\mathfrak{m} \rightarrow \mathfrak{m}'$ implies $\text{obs}_{\mathcal{N}}(\mathfrak{m}) \leq \text{obs}_{\mathcal{N}}(\mathfrak{m}')$. So, $(\Lambda_{\oplus}^{\text{cbv}}, \text{obs}_{\mathcal{N}})$ is a QARS. We prove (Thm. 31) that $\Lambda_{\oplus}^{\text{cbv}}$ satisfies the following properties: (1) the result $\llbracket \mathfrak{m} \rrbracket$ of computing \mathfrak{m} is well defined; (2) there exists a strategy that is guaranteed to produce $\llbracket \mathfrak{m} \rrbracket$.

Beyond the surface. We define a reduction $\xrightarrow{\text{E}} \subseteq \Lambda_{\oplus} \times \mathcal{M}(\Lambda_{\oplus})$ which performs surface steps ($\xrightarrow{s} = \xrightarrow{s}_{\beta_v} \cup \rightarrow_{\oplus}$, see Sect. 5) as much as possible, and then iterates the process on the subterms. There are two subtleties here. First: $M \not\xrightarrow{s}$ if and only if $(M \not\xrightarrow{s}_{\beta_v} \text{ and } M \not\xrightarrow{s}_{\oplus})$. Second: an occurrence of \oplus -redex can only be fired when it is a surface redex. By keeping this into account, Def. 20 updates as follows. We denote by \mathcal{S} the set of \xrightarrow{s} -normal forms.

► **Definition 24** (Unbiased evaluation $\xrightarrow{\text{E}}, \xrightarrow{\text{E}})$.

- The relation $\xrightarrow{\text{E}} \subseteq \Lambda_{\oplus} \times \mathcal{M}(\Lambda_{\oplus})$ is defined by the following rules, depending if $M \notin \mathcal{S}$ or $M \in \mathcal{S}$. The relation $\xrightarrow{\text{V}}_{\beta_v}$ is as in Def. 20.

$$\frac{M \xrightarrow{s} \mathfrak{m}}{M \xrightarrow{\text{E}} \mathfrak{m}} \quad (M \notin \mathcal{S}) \quad \frac{M \not\xrightarrow{s} \quad M \xrightarrow{\text{V}}_{\beta_v} M'}{M \xrightarrow{\text{E}} [M']} \quad (M \in \mathcal{S})$$

- $\xrightarrow{\text{E}}, \xrightarrow{\text{E}} \subseteq \mathcal{M}(\Lambda_{\oplus}) \times \mathcal{M}(\Lambda_{\oplus})$ are respectively the lifting and full lifting of $\xrightarrow{\text{E}}$ (Figs. 2 and 3).

Clearly, $\xrightarrow[E]{\subseteq} \rightarrow$, and moreover \rightarrow and $\xrightarrow[E]{\rightarrow}$ have the *same normal forms*.

► **Remark 25.** $\xrightarrow[E]{\Rightarrow}$ is $\text{obs}_{\mathcal{N}}$ -complete, but not $\text{obs}_{\mathcal{N}}$ -normalizing for \Rightarrow . Indeed, the sequence $\mathfrak{m} = [II \oplus \Delta\Delta] \xrightarrow[E]{\Rightarrow} [\frac{1}{2}II, \frac{1}{2}\Delta\Delta] \xrightarrow[E]{\Rightarrow} [\frac{1}{2}II, \frac{1}{2}\Delta\Delta] \xrightarrow[E]{\Rightarrow} \dots$ never fires II . The solution is to move to $\xrightarrow[E]{\Rightarrow}$, which forces all non-normal terms to reduce. Note that $\xrightarrow[E]{\Rightarrow}$ does not factorize \Rightarrow .

We show that $\xrightarrow[E]{\Rightarrow}$ is an $\text{obs}_{\mathcal{N}}$ -normalizing strategy for \Rightarrow . The pillars of our construction are E-factorization and weighted Random Descent. The former holds for $\xrightarrow[E]{\Rightarrow}$, the latter for $\xrightarrow[E]{\Rightarrow}$.

► **Proposition 26** (Factorization and $\text{obs}_{\mathcal{N}}$ -neutrality).

1. E-factorization: $\mathfrak{m} \Rightarrow^* \mathfrak{n}$ implies $\mathfrak{m} \xrightarrow[E]{\Rightarrow}^* \cdot \xrightarrow[E]{\Rightarrow}^* \mathfrak{n}$.
2. $\text{obs}_{\mathcal{N}}$ -neutrality: $\mathfrak{m} \xrightarrow[E]{\Rightarrow} \mathfrak{n}$ implies $\text{obs}_{\mathcal{N}}(\mathfrak{m}) = \text{obs}_{\mathcal{N}}(\mathfrak{n})$.

► **Proposition 27** (Diamond). $\xrightarrow[E]{\Rightarrow}$ is $\text{obs}_{\mathcal{N}}$ -diamond.

We are now ready to prove that, in $(\Lambda_{\oplus}^{\text{cbv}}, \text{obs}_{\mathcal{N}})$, the reduction $\xrightarrow[E]{\subseteq} \Rightarrow$ (i.e., the full lifting of $\xrightarrow[E]{\rightarrow}$) is guaranteed to compute the best possible result from each $\mathfrak{m} \in \mathcal{M}(\Lambda_{\oplus})$.

Asymptotic Completeness. We have that $\xrightarrow[E]{\Rightarrow}$ is asymptotically complete for \Rightarrow , because it satisfies the conditions of Thm. 16 (by Prop. 26).

► **Lemma 28.** If $\mathfrak{m} \Rightarrow^{\infty} \mathfrak{r}$ then $\mathfrak{m} \xrightarrow[E]{\Rightarrow}^{\infty} \mathfrak{r}$.

In turn, $\xrightarrow[E]{\Rightarrow}$ is asymptotically complete for $\xrightarrow[E]{\Rightarrow}$ (immediate). So via Lemma 28 we have:

► **Theorem 29.** $\xrightarrow[E]{\Rightarrow}$ is asymptotically complete for \Rightarrow : if $\mathfrak{m} \Rightarrow^{\infty} \mathfrak{r}$ then $\mathfrak{m} \xrightarrow[E]{\Rightarrow}^{\infty} \mathfrak{s}$ and $\mathfrak{r} \leq \mathfrak{s}$.

Unique Result. All $\xrightarrow[E]{\Rightarrow}$ -sequences from \mathfrak{m} converge to the same limit, by Prop. 18 and 27.

► **Theorem 30.** $\text{Lim}_{\text{obs}_{\mathcal{N}}}(\mathfrak{m}, \xrightarrow[E]{\Rightarrow})$ contains a unique element.

Asymptotic Normalization. By Prop. 14, the main result follows from Thm. 29 and 30.

► **Theorem 31** (Main, probabilistic CbV). For each $\mathfrak{m} \in \mathcal{M}(\Lambda_{\oplus})$:

1. $\llbracket \mathfrak{m} \rrbracket$ is defined;
2. $\mathfrak{m} \xrightarrow[E]{\Rightarrow}^{\infty} \mathfrak{r}$ if and only if $\mathfrak{r} = \llbracket \mathfrak{m} \rrbracket$.

Hence $\xrightarrow[E]{\Rightarrow}$ is an $\text{obs}_{\mathcal{N}}$ -normalizing strategy for \Rightarrow (see Remark 15).

Some simple examples will help to see how the normalizing strategy works, and how it differs from surface reduction.

► **Example 32.** Recall that β_v -reduction is unrestricted, so for example $M = \lambda z.(Iz) \rightarrow_{\beta_v} \lambda z.z$. Instead, $\lambda z.(Iz) \not\rightarrow_{\mathfrak{S}}$, because surface reduction cannot fire under abstraction. So *surface reduction is not a complete strategy* w.r.t. β_v -normal forms.

A direct consequence is that surface reduction is not informative about normalization, as it produces “false positive”. For example, $N = \lambda z.\Delta\Delta$ is diverging w.r.t. β_v -reduction, but it is a surface normal form. Let us now incept probability (with the terms M and N as above).

1. Let $R = (\lambda x.M \oplus xx)(\lambda x.M \oplus xx)$. Then $[R] \xrightarrow[E]{\Rightarrow} [M \oplus R] \xrightarrow[E]{\Rightarrow} [\frac{1}{2}M, \frac{1}{2}R] \xrightarrow[E]{\Rightarrow} [\frac{1}{2}I, \frac{1}{2}M \oplus R] \xrightarrow[E]{\Rightarrow} [\frac{1}{2}I, \frac{1}{4}M, \frac{1}{4}R] \xrightarrow[E]{\Rightarrow} \dots$. At the limit, R converges with probability 1 to I , as wanted.

17:16 Strategies for Asymptotic Normalization

$$\frac{}{\langle n : \mathbf{W}(\mathbf{tick}.P) \rangle \xrightarrow{\mathbf{W}\mathbf{tick}} \langle n+1 : \mathbf{W}(P) \rangle} \quad \frac{M \xrightarrow{\mathbf{W}\beta_v} M'}{\langle n : M \rangle \xrightarrow{\mathbf{W}\beta_v} \langle n : M' \rangle} \quad \frac{M \rightarrow_{\beta_v} M'}{\langle n : M \rangle \rightarrow_{\beta_v} \langle n : M' \rangle}$$

■ **Figure 5** Payoff reductions $\xrightarrow{\mathbf{W}\mathbf{tick}}, \rightarrow_{\beta_v}, \xrightarrow{\mathbf{W}\beta_v} \subseteq (\mathbb{N} \times \Lambda_{\mathbf{tick}})$.

$$\frac{\frac{M \mapsto_{\beta_v} M'}{\langle 0 : M \rangle \xrightarrow{\mathbf{W}} \langle 0 : M' \rangle} \quad \frac{M \mapsto_{\mathbf{tick}} M'}{\langle 0 : M \rangle \xrightarrow{\mathbf{W}} \langle 1 : M' \rangle} \quad \overline{\langle 0 : V \rangle \xrightarrow{\mathbf{W}} \langle 0 : V \rangle}}{\langle 0 : P_1 \rangle \xrightarrow{\mathbf{W}} \langle k_1 : P_1 \rangle \quad \langle 0 : P_2 \rangle \xrightarrow{\mathbf{W}} \langle k_2 : P_2 \rangle \quad n > 0 \quad \langle 0 : M \rangle \xrightarrow{\mathbf{W}} \langle k : M' \rangle} \quad \frac{}{\langle 0 : P_1 P_2 \rangle \xrightarrow{\mathbf{W}} \langle k_1 + k_2 : P_1 P_2 \rangle} \quad \frac{}{\langle n : M \rangle \xrightarrow{\mathbf{W}} \langle n + k : M' \rangle}$$

■ **Figure 6** Parallel weak reduction in the payoff calculus.

2. The term $S = (\lambda x.N \oplus xx)(\lambda x.N \oplus xx)$ converges to normal form with probability 0. One can easily check that $[S] \xrightarrow{\mathbb{E}} \infty \mathbf{0}$.
3. The term $S' = (\lambda x.(N \oplus I) \oplus xx)(\lambda x.(N \oplus I) \oplus xx)$ converges with probability $\frac{1}{2}$ to the normal form I . One can easily check that $[S'] \xrightarrow{\mathbb{E}} \infty \{I^{\frac{1}{2}}\}$.

► **Example 33.** One can easily build probabilistic terms with a more interesting behaviour than those in Example 32. First, observe that for $F = \lambda x.I$ (encoding the boolean *false*), we have that $(\lambda z.FF) \rightarrow_{\beta_v} F$. Now let $U = \lambda xy.(y \oplus xx(\lambda z.yy))$ and consider the term UUF , which converges with probability 1 to F . Indeed $[UUF] \xrightarrow{\mathbb{E}} \infty \{F^1\}$. In contrast, surface reduction converges to a distribution over *countably many different surface normal forms*, since each iteration produces a new *snf*: $\frac{1}{2}F, \frac{1}{4}\lambda z.FF, \frac{1}{8}\lambda z.(\lambda z.FF)(\lambda z.FF), \dots$

6 Asymptotic Normalization: More Case Studies

Our method applies – uniformly – to the other examples in Sect. 1.1. In this section we consider a CbV λ -calculus extended with an output operator. For the sake of a compact presentation, we take as output not a string, but simply an integer (think of it as a string on a single character). Albeit simple, this case study allows us to illustrate the subtleties related to limits with output calculi, and the use of our method. In a similar way, one can revisit Böhm Trees as the limit of a specific asymptotic strategy – we leave this to Appendix D.2.

λ -calculus with output: the payoff calculus. The payoff λ -calculus (called cost λ -calculus in [18, 24]) extends the λ -calculus with a ticking operation. Its intrinsic purpose is to facilitate an intensional analysis of programs, endowing terms with constructs to perform cost analysis.

Let $\Lambda_{\mathbf{tick}}$ denote the set of λ -terms extended with a unary operator \mathbf{tick} . The elements of the payoff calculus are pairs $\mathfrak{m} = \langle n : M \rangle$ of a counter $n \in \mathbb{N}$ and a closed term $M \in \Lambda_{\mathbf{tick}}$. Intuitively, the term $\mathbf{tick}(P)$ increments the counter by 1, and continues as P . Following [18], in $(\mathbb{N} \times \Lambda_{\mathbf{tick}})$ we define the full reduction \rightarrow and the weak reduction $\xrightarrow{\mathbf{W}}$ as follows:

$$\rightarrow := \rightarrow_{\beta_v} \cup \xrightarrow{\mathbf{W}\mathbf{tick}} \quad \xrightarrow{\mathbf{W}} := \xrightarrow{\mathbf{W}\beta_v} \cup \xrightarrow{\mathbf{W}\mathbf{tick}}$$

where $\xrightarrow{\mathbf{W}\mathbf{tick}}, \rightarrow_{\beta_v}, \xrightarrow{\mathbf{W}\beta_v} \subseteq (\mathbb{N} \times \Lambda_{\mathbf{tick}})$ are given in Fig. 5. Note that weak effectful reduction $\xrightarrow{\mathbf{W}\mathbf{tick}}$ is the closure under weak context \mathbf{W} of the rule $(\mathbf{tick}.P) \mapsto_{\mathbf{tick}} P$ (effects are only allowed under weak context). Left and right reductions $\xrightarrow{\mathbf{L}}$ and $\xrightarrow{\mathbf{R}}$ can be defined similarly.

The pair $((\Lambda_{\text{tick}}, \rightarrow), \text{obs})$ is a QARS where we observe the payoff, *i.e.* $\text{obs}\langle n : M \rangle = n$. We now prove (using Thm. 16) that $\xrightarrow{w} = \xrightarrow{w}_{\beta_v} \cup \xrightarrow{w}_{\text{tick}}$ is asymptotically complete for \rightarrow .

► **Lemma 34.** *For every pair $m = \langle n : M \rangle$, $m \rightarrow^\infty n$ implies $m \xrightarrow{w}^\infty n$, because*

- *w-factorization of \rightarrow : if $m \rightarrow^* n$ then $m \xrightarrow{w}^* \cdot \xrightarrow{w}^* n$;*
- *obs-neutrality: if $m \xrightarrow{w} m'$ then $\text{obs}(m) = \text{obs}(m')$.*

Weak reduction \xrightarrow{w} however does *not* have a unique limit, as Example 35 below illustrates. An unsatisfactory solution would be to fix a deterministic evaluation order (left or right, as in point 1. below), making the limit easy to predict but also rather *arbitrary*.

► **Example 35.** Consider $M = (\Delta\Delta)(\Delta_\vee\Delta_\vee)$, where $\Delta_\vee = \lambda x.\text{tick}(xx)$, and let $m = \langle 0 : M \rangle$.

1. By fixing left (resp. right) evaluation, $\text{Lim}_{\text{obs}}(m, \xrightarrow{\uparrow}) = \{0\}$ (resp. $\text{Lim}_{\text{obs}}(m, \xrightarrow{\downarrow}) = \{\infty\}$).
2. By choosing a redex in unspecified order, we have an uncountable number of \xrightarrow{w} -sequences, leading to $\text{Lim}_{\text{obs}}(m, \xrightarrow{w}) = \{0, 1, \dots, \infty\} = \mathbb{N}^\infty$.

A way out is to proceed somehow similarly to Sect. 5.1. If we examine more closely the set of limits associated with \xrightarrow{w} , we realize that $\text{Lim}_{\text{obs}}(m, \xrightarrow{w})$ does have a greatest element. Thus $\llbracket m \rrbracket$ can naturally be defined as the best possible payoff from m . We prove that parallel reduction $\xrightarrow{\parallel/w}$ (given in Fig. 6) is a (*multistep*) strategy which is guaranteed to compute m . Indeed, it is easy to verify that $\xrightarrow{\parallel/w}$ is asymptotically complete for \xrightarrow{w} . By composing with Lemma 34 we have that $\xrightarrow{\parallel/w}$ is asymptotically complete for \rightarrow (point 1. below).

► **Lemma 36.**

1. Asymptotic Completeness. *If $\langle k : M \rangle \xrightarrow{w}^\infty n$ then $\langle k : M \rangle \xrightarrow{\parallel/w}^\infty n'$ and $n \leq n'$. That is, $\xrightarrow{\parallel/w}$ is asymptotically complete for \xrightarrow{w} and (by Lemma 34) for \rightarrow .*
2. Unique Limit. *The reduction $\xrightarrow{\parallel/w}$ is deterministic.*

Since (by points 1. and 2. in Lemma 36) both conditions of Prop. 14 are verified, we have:

► **Theorem 37 (Main, payoff).** *Given the QARS $((\mathbb{N} \times \Lambda_{\text{tick}}, \rightarrow), \text{obs})$, for each pair $m = \langle k : M \rangle$, $\llbracket m \rrbracket$ is defined, and $m \xrightarrow{\parallel/w}^\infty \llbracket m \rrbracket$. Hence, multistep reduction $\xrightarrow{\parallel/w}$ is asymptotically normalizing for \xrightarrow{w} and for \rightarrow (Remark 15).*

λ -calculus with outputs. The calculus in Example 2 can be formalized in a similar way to the payoff calculus. We can define $\text{obs}\langle s : M \rangle = s$. As already noted, \xrightarrow{w} is not confluent, and given a pair m , the set of limits may contain uncountably many different elements. Still, the reduction has interesting properties, which appear when looking not directly at the string s itself, but at its length $|s|$. This way, one can transfer the results from the payoff calculus.

7 Conclusions

We propose a method to study completeness and normalization when the result of computation is *asymptotic*. Our techniques abstract from details specific to the calculus under study – they are therefore of *general* application. The robustness of the method is witnessed by its ability to deal with different settings and *different notions of asymptotic computation*.

The application to probabilistic λ -calculus yields a result of independent interest: a theorem of *asymptotic normalization*, both for CbV and CbN probabilistic λ -calculi. Remarkably, the same definitions and proof techniques apply *uniformly* to both. In the paper we prefer to give the details for the CbV calculus, which is arguably a more natural one in presence of effects.

Related work. QARS, proposed in [14] in the setting of probabilistic rewriting, refine Ariola and Blom’s ARSI [4]. The techniques in Sect. 3 are an original contribution of this paper. Our Thm. 16 generalizes an ARS technique for finitary normalization (studied in [2, 21, 33]) to asymptotic computation, refining it for arbitrary observations.

The study of reduction strategies in a probabilistic λ -calculus where the notion of reduction is general – rather than simply fixing a deterministic reduction – started in [16] (CbV and CbN) and [27] (CbN). Asymptotic completeness is there established only for *surface* normal forms (values in closed CbV, hnf’s in CbN). Strategies that are complete for *full normal forms* (which we treat and solve here) are more difficult to study than head or weak reduction, especially in the CbV setting. The question of defining such a strategy was left open in [16, Remark 27]. We stress that our technique would also yield a simpler proof of the results in [16, 27], where confluence is used to establish that a greatest limit exists. The (non-trivial) proofs there use properties that are *specific* to probability distributions. The method we propose here avoids technical issues, it is much simpler, and it is general, in that it can be applied to other settings.

Finally, we mention that forgoing confluence and studying uniqueness of normal forms via a complete subreduction is a route already employed in the context of infinitary λ -calculi [8, 9].

References

- 1 Samson Abramsky and C.-H. Luke Ong. Full Abstraction in the Lazy Lambda Calculus. *Inf. Comput.*, 105(2):159–267, 1993. doi:10.1006/inco.1993.1044.
- 2 Beniamino Accattoli, Claudia Faggian, and Giulio Guerrieri. Factorization and Normalization, Essentially. In *Programming Languages and Systems - 17th Asian Symposium, APLAS 2019*, volume 11893 of *Lecture Notes in Computer Science*, pages 159–180. Springer, 2019. doi:10.1007/978-3-030-34175-6_9.
- 3 Roberto M. Amadio and Pierre-Louis Curien. *Domains and Lambda-Calculi*, volume 46 of *Cambridge Tracts in Theoretical Computer Science*. Cambridge University Press, 1998. doi:10.1017/CB09780511983504.
- 4 Zena M. Ariola and Stefan Blom. Skew confluence and the lambda calculus with letrec. *Annals of Pure and Applied Logic*, 117(1):95–168, 2002. doi:10.1016/S0168-0072(01)00104-X.
- 5 Martin Avanzini, Ugo Dal Lago, and Akihisa Yamada. On probabilistic term rewriting. *Sci. Comput. Program.*, 185, 2020. doi:10.1016/j.scico.2019.102338.
- 6 Franz Baader and Tobias Nipkow. *Term Rewriting and All That*. Cambridge University Press, 1998. doi:10.1017/CB09781139172752.
- 7 Hendrik Pieter Barendregt. *The Lambda Calculus – Its Syntax and Semantics*, volume 103 of *Studies in logic and the foundations of mathematics*. North-Holland, 1984.
- 8 Henk Barendregt and Jan Willem Klop. Applications of infinitary lambda calculus. *Inf. Comput.*, 207(5):559–582, 2009. doi:10.1016/j.ic.2008.09.003.
- 9 Alessandro Berarducci and Benedetto Intrigila. Church-Rosser λ -theories, infinite λ -calculus and consistency problems. In W. Hodges, M. Hyland, and et. al., editors, *Logic: From Foundations to Applications (European Logic Colloquium)*, pages 33–58. Oxford Sci. Publ., 1996.
- 10 Gianluca Curzi and Michele Pagani. The Benefit of Being Non-Lazy in Probabilistic λ -calculus: Applicative Bisimulation is Fully Abstract for Non-Lazy Probabilistic Call-by-Name. In *LICS ’20: 35th Annual ACM/IEEE Symposium on Logic in Computer Science*, pages 327–340. ACM, 2020. doi:10.1145/3373718.3394806.
- 11 Roel C. de Vrijer. Conditional linearization. *Indagationes Mathematicae*, 10(1):145–159, 1999. doi:10.1016/S0019-3577(99)80012-3.
- 12 Ugo de’Liguoro and Adolfo Piperno. Non Deterministic Extensions of Untyped Lambda-Calculus. *Inf. Comput.*, 122(2):149–177, 1995. doi:10.1006/inco.1995.1145.

- 13 Claudia Faggian. Probabilistic Rewriting: Normalization, Termination, and Unique Normal Forms. In *4th International Conference on Formal Structures for Computation and Deduction, FSCD 2019*, volume 131 of *LIPICs*, pages 19:1–19:25. Schloss Dagstuhl, 2019. doi:10.4230/LIPICs.FSCD.2019.19.
- 14 Claudia Faggian. Probabilistic Rewriting and Asymptotic Behaviour: on Termination and Unique Normal Forms. *Log. Methods Comput. Sci.*, vol. 18, issue 2, 2022.
- 15 Claudia Faggian and Giulio Guerrieri. Strategies for Asymptotic Normalization (long version). *CoRR*, 2022. arXiv:2204.08772.
- 16 Claudia Faggian and Simona Ronchi Della Rocca. Lambda Calculus and Probabilistic Computation. In *34th Annual ACM/IEEE Symposium on Logic in Computer Science, LICS 2019*, pages 1–13. IEEE, 2019. doi:10.1109/LICS.2019.8785699.
- 17 Francesco Gavazzo. *Coinductive Equivalences and Metrics for Higher-order Languages with Algebraic Effects*. PhD thesis, Università di Bologna, Italy, 2019. URL: <https://tel.archives-ouvertes.fr/tel-02386201>.
- 18 Francesco Gavazzo and Claudia Faggian. A Relational Theory of Monadic Rewriting Systems, Part I. In *36th Annual ACM/IEEE Symposium on Logic in Computer Science, LICS 2021*, pages 1–14. IEEE, 2021. doi:10.1109/LICS52264.2021.9470633.
- 19 Benjamin Grégoire and Xavier Leroy. A compiled implementation of strong reduction. In Mitchell Wand and Simon L. Peyton Jones, editors, *Proceedings of the Seventh ACM SIGPLAN International Conference on Functional Programming (ICFP '02)*, pages 235–246. ACM, 2002. doi:10.1145/581478.581501.
- 20 J. Roger Hindley and Jonathan P. Seldin. *Introduction to Combinators and Lambda-Calculus*. Cambridge University Press, 1986.
- 21 Nao Hirokawa, Aart Middeldorp, and Georg Moser. Leftmost Outermost Revisited. In *26th International Conference on Rewriting Techniques and Applications, RTA 2015*, volume 36 of *LIPICs*, pages 209–222. Schloss Dagstuhl, 2015. doi:10.4230/LIPICs.RTA.2015.209.
- 22 J. Richard Kennaway, Jan Willem Klop, M. Ronan Sleep, and Fer-Jan de Vries. Transfinite Reductions in Orthogonal Term Rewriting Systems. *Inf. Comput.*, 119(1):18–38, 1995. doi:10.1006/inco.1995.1075.
- 23 J. Richard Kennaway, Jan Willem Klop, M. Ronan Sleep, and Fer-Jan de Vries. Infinite lambda calculus. *Theoretical Computer Science*, 175(1):93–125, 1997. doi:10.1016/S0304-3975(96)00171-5.
- 24 Ugo Dal Lago and Francesco Gavazzo. Effectful Normal Form Bisimulation. In Luís Caires, editor, *Programming Languages and Systems - 28th European Symposium on Programming, ESOP 2019, Proceedings*, volume 11423 of *Lecture Notes in Computer Science*, pages 263–292. Springer, 2019. doi:10.1007/978-3-030-17184-1_10.
- 25 Ugo Dal Lago and Margherita Zorzi. Probabilistic operational semantics for the lambda calculus. *RAIRO Theor. Informatics Appl.*, 46(3):413–450, 2012. doi:10.1051/ita/2012012.
- 26 Xavier Leroy. The ZINC experiment: an economical implementation of the ML language. Technical report 117, INRIA, 1990. URL: <http://gallium.inria.fr/~xleroy/publi/ZINC.pdf>.
- 27 Thomas Leventis. A deterministic rewrite system for the probabilistic λ -calculus. *Mathematical Structures in Computer Science*, 29(10):1479–1512, 2019. doi:10.1017/S0960129519000045.
- 28 Jean-Jacques Lévy. *Réductions correctes et optimales dans le lambda calcul*. PhD thesis, Université Paris 7, 1978. URL: <http://pauillac.inria.fr/~levy/pubs/74phd-cycle3.pdf>.
- 29 Maxwell H. A. Newman. On Theories with a Combinatorial Definition of Equivalence. *Annals of Mathematics*, 43(2), 1942.
- 30 Gordon D. Plotkin. Call-by-Name, Call-by-Value and the lambda-Calculus. *Theor. Comput. Sci.*, 1(2):125–159, 1975. doi:10.1016/0304-3975(75)90017-1.
- 31 Terese. *Term Rewriting Systems*, volume 55 of *Cambridge Tracts in Theoretical Computer Science*. Cambridge University Press, 2003.

- 32 Vincent van Oostrom. Random Descent. In *Term Rewriting and Applications, 18th International Conference, RTA 2007*, volume 4533 of *Lecture Notes in Computer Science*, pages 314–328. Springer, 2007. doi:10.1007/978-3-540-73449-9_24.
- 33 Vincent van Oostrom and Yoshihito Toyama. Normalisation by Random Descent. In *1st International Conference on Formal Structures for Computation and Deduction, FSCD 2016*, volume 52 of *LIPICs*, pages 32:1–32:18. Schloss Dagstuhl, 2016. doi:10.4230/LIPICs.FSCD.2016.32.

APPENDIX

We include some proofs and details that have been omitted in the article.

One more example. $\text{Lim}_{\text{obs}}(t, \rightarrow)$ may have a lub but not a maximum – similarly to \mathbb{N} .

► **Example 38** (Sect. 2, QARS). We revisit Example 11, allowing full reduction \rightarrow_{β_v} . Let $\text{obs}_p(\langle s : M \rangle) = s$ if $M \in \mathcal{V}$, \perp otherwise. The pair $\mathbf{m} = \langle \epsilon : (\lambda z.I)(\lambda z.\Delta_0\Delta_0) \rangle$ has countably many limits, but not a greatest one, because all strings in $\text{Lim}_{\text{obs}_p}(\mathbf{m}, \rightarrow_{\beta_v})$ are finite.

Surface reduction. Everywhere in the appendix, we fix surface reduction to be as follows.

- CbN ($\mathbf{b} = \beta$): $\mathbf{s} = \mathbf{h}$ (the contextual closure of \mathbf{H}).
- CbV ($\mathbf{b} = \beta_v$): $\mathbf{s} = \mathbf{w}$ (the contextual closure of \mathbf{W}).

A Properties of surface normal forms

We will use extensively the following easy fact.

► **Lemma 39** (Surface normal forms). *M is w-normal (resp. h-normal) if there is no redex R such that $M = \mathbf{W}(R)$ (resp. $M = \mathbf{H}(R)$).*

1. CbV. Assume $M \xrightarrow{\mathbf{w}}_{\beta_v} M'$. *M is w-normal $\Leftrightarrow M'$ is w-normal.*
2. CbN. Assume $M \xrightarrow{\mathbf{h}}_{\beta} M'$. *M is h-normal $\Leftrightarrow M'$ is h-normal.*

B Sect. 4.2: properties of unbiased reduction

The properties of $\xrightarrow{\mathbf{v}}$ (Prop. 21, Prop. 22, Thm. 23, and those stted here) are proved in [15].

► **Lemma 40** (Normal forms). *If $U \xrightarrow{\mathbf{v}}_{\mathbf{b}} N$, then N is not b-normal.*

B.1 A parallel variant of unbiased reduction

Given $(\Lambda, \rightarrow_{\mathbf{b}})$ and $\xrightarrow{\mathbf{s}}$ as in Def. 20, a parallel version $\xrightarrow{\mathbf{v}}_{\mathbf{b}}$ is easily defined. The idea here is that once a term is $\xrightarrow{\mathbf{s}}$ -normal, iteration of the reduction process can be performed in any arbitrary order, or in *parallel*. Recall that here $(\mathbf{b}, \mathbf{s}) \in \{(\beta, \mathbf{h}), (\beta_v, \mathbf{w})\}$.

1. If $M \xrightarrow{\mathbf{s}} M'$ then $M \xrightarrow{\mathbf{v}}_{\mathbf{b}} M'$ (*M is not s-normal*);
2. If $M \not\xrightarrow{\mathbf{s}}$ then $M \xrightarrow{\mathbf{v}}_{\mathbf{b}} M$ (*M is \rightarrow -normal*);
3. Otherwise:

$$\frac{P \xrightarrow{\mathbf{v}}_{\mathbf{b}} P'}{M := \lambda x.P \xrightarrow{\mathbf{v}}_{\mathbf{b}} \lambda x.P'} \quad \frac{P_1 \xrightarrow{\mathbf{v}}_{\mathbf{b}} P'_1 \quad P_2 \xrightarrow{\mathbf{v}}_{\mathbf{b}} P'_2}{M := P_1 P_2 \xrightarrow{\mathbf{v}}_{\mathbf{b}} P'_1 P'_2} \quad \frac{(P_i \xrightarrow{\mathbf{v}}_{\mathbf{b}} P'_i)_{1 \leq i \leq k}}{M := \mathbf{o}(P_1, \dots, P_k) \xrightarrow{\mathbf{v}}_{\mathbf{b}} \mathbf{o}(P'_1, \dots, P'_k)}$$

Rule 2. makes the relation reflexive on normal forms and *only* on normal forms – this is a harmless shortcut in order to give a compact and neat formulation.

The (multistep) reduction $\xrightarrow{\beta} \mathfrak{b}$ is guaranteed to reach the $\rightarrow_{\mathfrak{b}} \mathit{nf}$, if any exists.

► **Lemma 41.** *Let $\mathfrak{b} \in \{\beta, \beta_v\}$*

1. *If $M \xrightarrow{\beta} \mathfrak{b} N$ then $M \xrightarrow{\beta} \mathfrak{b}^* N$. Therefore, $M \xrightarrow{\beta} \mathfrak{b}^* N$ implies $M \xrightarrow{\beta} \mathfrak{b}^* N$.*
2. *If $M \xrightarrow{\beta} \mathfrak{b}^* N$ then there exists N' such that $M \xrightarrow{\beta} \mathfrak{b}^* N'$ and $N \xrightarrow{\beta} \mathfrak{b}^* N'$.*

► **Corollary 42** (\mathcal{N} -completeness). *Let $\mathfrak{b} \in \{\beta, \beta_v\}$ and N be $\rightarrow_{\mathfrak{b}}$ -normal. $M \rightarrow_{\mathfrak{b}}^* N$ if and only if $M \xrightarrow{\beta} \mathfrak{b}^* N$.*

C

 Proofs of Sect. 5.1: Asymptotic Normalization for $\Lambda_{\oplus}^{\text{cbv}}$

Notice also that the definition of the reductions $\xrightarrow{\beta}, \xRightarrow{\beta}$ can be given in the same way also in CbN, by replacing β_v with β (surface steps are here head steps).

Properties. We freely use the following fact.

► **Fact 43.** $[M] \Rightarrow_{\beta_v} [M']$ iff $M \rightarrow_{\beta_v} M'$, where

- on the l.h.s. we have $(\mathcal{M}(\Lambda_{\oplus}), \Rightarrow)$, and
- on the r.h.s. the CbV λ -calculus $(\Lambda_{\oplus}, \rightarrow_{\beta_v})$, as defined in Sect. 4.1.

Factorization and Neutrality. Recall that \mathcal{S} denotes the set of the **surface normal forms** of $\rightarrow = (\rightarrow_{\beta_v} \cup \rightarrow_{\oplus})$. $M \in \Lambda_{\oplus}$ is **s-normal** if $M \not\rightarrow_{\mathcal{S}}$. That is $M \not\rightarrow_{\oplus}$ and $M \not\rightarrow_{\mathcal{S}\beta_v}$.

► **Lemma 44** (*snf propagation*). *If M is s-normal and $M \rightarrow M'$, then M' is s-normal.*

► **Proposition 45** (\mathbb{E} -Factorization of \Rightarrow). *In $\Lambda_{\oplus}^{\text{cbv}}$: $\mathfrak{m} \Rightarrow^* \mathfrak{n}$ implies $\mathfrak{m} \xRightarrow{\mathbb{E}}^* \cdot \xRightarrow{\mathbb{E}}^* \mathfrak{n}$*

Proof. In the proof, we use freely Fact 43. By surface factorization of \Rightarrow (proved in [16]), $\mathfrak{m} \Rightarrow^* \mathfrak{n}$ implies $\mathfrak{m} \xRightarrow{\mathcal{S}}^* \mathfrak{t} \xRightarrow{\mathcal{S}}^* \mathfrak{n}$ for some \mathfrak{t} . From this we have:

- $\mathfrak{m} \xRightarrow{\mathbb{E}}^* \mathfrak{t}$. Because if $M \xrightarrow{\mathcal{S}} \mathfrak{r}$ then also $M \xrightarrow{\mathbb{E}} \mathfrak{r}$.
- $\mathfrak{t} \xRightarrow{\mathcal{S}\beta_v}^* \mathfrak{n}$. Because if $M \xrightarrow{\mathcal{S}} \mathfrak{r}$ then necessarily $M \xrightarrow{\mathcal{S}\beta_v} \mathfrak{r}$. Moreover, $\mathfrak{r} = [M']$.

Let $\mathfrak{t} = [\dots p_i T_i \dots]_{i \in I}$. Then necessarily, $\mathfrak{n} = [\dots p_i N_i \dots]_{i \in I}$ and $[T_i] \xRightarrow{\mathcal{S}\beta_v}^* [N_i]$ and so also $T_i \xrightarrow{\mathcal{S}\beta_v}^* N_i$. For each T_i , we examine if T_i is s-normal or not (\mathcal{S} being the set of *snf*'s).

1. $T_i \in \mathcal{S}$. By $T_i \xrightarrow{\mathcal{S}\beta_v}^* N_i$ and \mathbb{U} -factorization of \rightarrow_{β_v} (Prop. 21), $T_i \xrightarrow{\mathbb{U}} \beta_v^* U_i \xrightarrow{\rightarrow} \beta_v^* N_i$. By Lemma 44, each term in the sequence $T_i \xrightarrow{\mathbb{U}} \beta_v^* U_i \xrightarrow{\rightarrow} \beta_v^* N_i$ is s-normal. Hence, by Def. 24 (since only the second rule can apply), we conclude that $[T_i] \xRightarrow{\mathbb{E}\beta_v}^* [U_i] \xRightarrow{\mathbb{E}\beta_v}^* [N_i]$.
2. $T_i \notin \mathcal{S}$. By Lemma 39, each term in the sequence $T_i \xrightarrow{\mathcal{S}\beta_v}^* N_i$ is not s-normal. By definition of $\xRightarrow{\mathbb{E}}$ (since only the first rule can apply) we conclude that $[T_i] \xRightarrow{\mathbb{E}\beta_v}^* [N_i]$.

Let us partition \mathfrak{t} into two multi-distributions, collecting in \mathfrak{t}_1 the terms of case 1. and in \mathfrak{t}_2 the terms of case 2. We partition \mathfrak{n} so that $\mathfrak{t}_1 \xRightarrow{\mathcal{S}}^* \mathfrak{n}_1$ and $\mathfrak{t}_2 \xRightarrow{\mathcal{S}}^* \mathfrak{n}_2$. We have $\mathfrak{t}_1 \xRightarrow{\mathbb{E}\beta_v}^* \mathfrak{u} \xRightarrow{\mathbb{E}\beta_v}^* \mathfrak{n}_1$ and $\mathfrak{t}_2 \xRightarrow{\mathbb{E}\beta_v}^* \mathfrak{n}_2$. Therefore $\mathfrak{m} \xRightarrow{\mathbb{E}}^* (\mathfrak{t}_1 + \mathfrak{t}_2) \xRightarrow{\mathbb{E}}^* (\mathfrak{u} + \mathfrak{t}_2) \xRightarrow{\mathbb{E}}^* (\mathfrak{n}_1 + \mathfrak{n}_2) = \mathfrak{n}$. which proves the claim. ◀

► **Proposition 46** (*neutrality*). *If $\mathfrak{m} \xRightarrow{\mathbb{E}} \mathfrak{n}$ then $\text{obs}_{\mathcal{N}}(\mathfrak{m}) = \text{obs}_{\mathcal{N}}(\mathfrak{n})$.*

Proof. Consequence of the fact that if $U \xrightarrow{\rightarrow} \beta_v N$, then N is not β_v -normal (Lemma 40). Indeed $\xrightarrow{\rightarrow} \subseteq \xrightarrow{\mathcal{S}}$ and so $M \xrightarrow{\rightarrow} \mathfrak{r}$ iff $(M \xrightarrow{\rightarrow} \beta_v M')$ with $\mathfrak{r} = [M']$. ◀

17:22 Strategies for Asymptotic Normalization

Diamonds. Prop. 27 (the relation $\xrightarrow[E]{\Rightarrow}$ is $\text{obs}_{\mathcal{N}}$ -diamond) follows from the following key lemma. Notice that Point (2.) implies that $\text{obs}_{\mathcal{N}}(\mathbf{m}_1) = \text{obs}_{\mathcal{N}}(\mathbf{m}_2)$.

► **Lemma 47 (Pointed Diamond).** *Let $\alpha, \gamma \in \{\beta_v, \oplus\}$. Assume M has two distinct redexes, such that $M \xrightarrow[E]{\rightarrow} \alpha \mathbf{m}_1$ and $M \xrightarrow[E]{\rightarrow} \gamma \mathbf{m}_2$. Then*

1. *exists \mathbf{t} such that $\mathbf{m}_1 \xrightarrow[E]{\Rightarrow} \gamma \mathbf{t}$ and $\mathbf{m}_2 \xrightarrow[E]{\Rightarrow} \alpha \mathbf{t}$.*
2. *Moreover, no M_i in $\mathbf{m}_1 = [p_i M_i]_i$ and no M_j in $\mathbf{m}_2 = [q_j M_j]_j$ is \rightarrow normal.*

Proof.

- If M is \mathbf{s} -normal, then by definition of $\xrightarrow[E]{\rightarrow}$, $M \xrightarrow[\mathbf{v}]{\rightarrow} \beta_v \mathbf{m}_1$ and $M \xrightarrow[\mathbf{v}]{\rightarrow} \beta_v \mathbf{m}_2$, and we conclude by using Fact 43 and Prop. 22, point 1.
- If M is \mathbf{s} -reducible, then by definition of $\xrightarrow[E]{\rightarrow}$, $M \xrightarrow[\mathbf{s}]{\rightarrow} \beta_v \mathbf{m}_1$ and $M \xrightarrow[\mathbf{s}]{\rightarrow} \beta_v \mathbf{m}_2$. We easily conclude by case analysis. ◀

► **Remark 48** ($\xrightarrow[E]{\rightarrow} \neq \xrightarrow[\mathbf{v}]{\rightarrow} \beta_v \cup \rightarrow_{\oplus}$). It is useful to notice that $\xrightarrow[E]{\rightarrow} \neq \xrightarrow[\mathbf{v}]{\rightarrow} \beta_v \cup \rightarrow_{\oplus}$. Such a relation is neither diamond nor confluent.

- The lifting of $\xrightarrow[\mathbf{v}]{\rightarrow} \beta_v \cup \rightarrow_{\oplus}$ is neither diamond nor confluent. Consider $(\Delta \oplus \Delta \Delta)(\lambda z. Iz)$. Then $\mathbf{m}_1 = [\frac{1}{2} \Delta(\lambda z. Iz), \frac{1}{2} (\Delta \Delta)(\lambda z. Iz)] \leftarrow_{\oplus} (\Delta \oplus \Delta \Delta)(\lambda z. Iz) \xrightarrow[\mathbf{v}]{\rightarrow} \beta_v [(\Delta \oplus \Delta \Delta)(\lambda z. z)] = \mathbf{m}_2$. The elements \mathbf{m}_1 and \mathbf{m}_2 cannot join, because no $\xrightarrow[\mathbf{v}]{\rightarrow} \beta_v$ -step can fire the underlined (Iz) .
- Similarly in CbN, for the lifting of $\xrightarrow[\mathbf{v}]{\rightarrow} \beta \cup \rightarrow_{\oplus}$. Consider $(\Delta \oplus \Delta \Delta)(x(Iz))$.

D Details for Sect. 6: more case studies

D.1 Asymptotic Normalization for a calculus with outputs

Proof of Lemma 34. The w -factorization of \rightarrow is proved in [18], where it is called surface factorization, and proved in general for all CbV monadic calculi, including the payoff calculus which we discuss here. obs -neutrality is straightforward to verify, by case analysis.

D.2 Asymptotic Normalization and Böhm Trees

We show that the Böhm Tree of a term M is the (unique) limit of an asymptotically normalizing strategy, *i.e.* the limit of a *single* reduction sequence.

Böhm Trees and Partial Normal Forms. Following [3], the Böhm Tree of a term M is (the downward closure of) the set of the partial normal forms of all reducts of M .

► **Definition 49 (Partial Normal Forms and Böhm Trees).** *The set \mathcal{N}_{ω} of **partial normal forms** is defined as follows;*

$$\frac{}{\Omega \in \mathcal{N}_{\omega}} \quad \frac{A_1 \in \mathcal{N}_{\omega} \dots A_n \in \mathcal{N}_{\omega}}{\lambda x_1 \dots x_n. x A_1 \dots A_n \in \mathcal{N}_{\omega}}$$

\mathcal{N}_{ω} is a subset of the set of partial λ -terms, defined by $P := \Omega \mid x \mid PP \mid \lambda x. P$, and inherits its **order** \leq , which is generated by the following rules:

$$\frac{}{\Omega \leq P} \quad \frac{P_1 \leq P'_1 \quad P_2 \leq P'_2}{P_1 P_2 \leq P'_1 P'_2} \quad \frac{P \leq P'}{\lambda x. P \leq \lambda x. P'}$$

The elements of the ideal completion $\mathcal{N}_{\omega}^{\infty}$ of \mathcal{N}_{ω} are called **Böhm Trees**. Precisely:

1. The function $\omega : \Lambda \rightarrow \mathcal{N}_\omega$ associates to each term $M \in \Lambda$ its partial normal form $\omega(M)$:

$$\omega(M) = \begin{cases} \Omega & \text{if } M \notin \mathcal{H} \\ \lambda \vec{x}.x\omega(M_1) \dots \omega(M_p) & \text{if } M = \lambda \vec{x}.xM_1 \dots M_p \end{cases}$$

2. The **Böhm Tree of M** is defined as below For a set \mathcal{S} , $\downarrow \mathcal{S} = \{Q \in \mathcal{N}_\omega \mid Q \leq S \in \mathcal{S}\}$.

$$\text{BT}(M) := \bigcup_{M \rightarrow^* N} \downarrow \{\omega(N)\} = \downarrow \{\omega(N) \mid M \rightarrow^* N\}$$

The following property is standard and easy-to-check (see [3, Lemma 2.3.2].)

► **Lemma 50.** *Let $M, M' \in \Lambda$. If $M \rightarrow_\beta M'$ then $\omega(M) \leq \omega(M')$.*

Lemma 50 guarantees that $((\Lambda, \rightarrow_\beta), \text{obs})$ is a QARS where $\text{obs} : \Lambda \rightarrow \mathcal{N}_\omega^\infty$ is defined as $\text{obs}(M) = \downarrow \{\omega(M)\}$.

Asymptotic Normalization. Let us define $\text{obs} : \Lambda \rightarrow \mathcal{N}_\omega^\infty$ as $\text{obs}(M) = \downarrow \{\omega(M)\}$. It is easily checked that $((\Lambda, \rightarrow_\beta), \text{obs})$ is a QARS. We show that the Böhm Tree of a term M can be obtained by asymptotic normalization, as the limit a $\xrightarrow[\text{U}]{\beta}$ reduction sequence, which is an **obs-normalizing** strategy for \rightarrow_β (Thm. 54).

The **obs-limit** of a reduction sequence $\langle M_n \rangle_n$ is then $\sup_i \{\text{obs}(M_i)\} = \bigcup_i \downarrow \{\omega(M_i)\}$. $\text{BT}(M)$ is clearly the sup of the set $\text{Lim}_{\text{obs}}(M, \rightarrow_\beta)$. We show that $\text{BT}(M)$ belongs to that set, by proving that $\text{Lim}_{\text{obs}}(M, \rightarrow_\beta)$ has a greatest element $\llbracket M \rrbracket$; this necessarily is $\text{BT}(M)$.

We first show that $\xrightarrow[\text{U}]{\beta}$ is **obs-complete** for \rightarrow_β (Point 1 in Prop. 52 below). Reduction $\xrightarrow[\text{U}]{\beta}$ is not **obs-normalizing** for \rightarrow_β (for example, it admits the sequence $x(\Delta\Delta)(Iz) \xrightarrow[\text{U}]{\beta} x(\Delta\Delta)(Iz) \xrightarrow[\text{U}]{\beta} \dots$) but its parallel version $\xrightarrow[\text{U}]{\beta}$ (Appendix B.1) is.

We proceed similarly to Sect. 5.1 (think $\xrightarrow[\text{E}]{\beta}$ vs $\xrightarrow[\text{E}]{\beta}$). We consider the reduction $\xrightarrow[\text{U}]{\beta}$ (the explicit definition is in Appendix B.1) which has **obs-Random Descent** (trivially) and is asymptotically complete for $\xrightarrow[\text{U}]{\beta}$ (Point 2 in Prop. 52 below), and so for \rightarrow_β .

► **Lemma 51.** *If $M \xrightarrow[\text{U}]{\beta} M'$ then $\omega(M) = \omega(M')$.*

Proof. First, observe that $M \xrightarrow[\text{h}]{\beta} M'$, because $\xrightarrow[\text{U}]{\beta} \subseteq \xrightarrow[\text{h}]{\beta}$.

- If M is not **h-normal** ($M \notin \mathcal{H}$), neither is M' , by (Lemma 39). Therefore, $\omega(M) = \Omega = \omega(M')$.
- Otherwise, M is **h-normal**, that is, $M = \lambda \vec{x}.xM_1 \dots M_p$. As $M \xrightarrow[\text{h}]{\beta} M'$, necessarily $M' = \lambda \vec{x}.xM_1 \dots M'_i \dots M_p$ (which is head normal) and $M_i \rightarrow_\beta M'_i$ for some $1 \leq i \leq p$. It is impossible that $M_i \xrightarrow[\text{U}]{\beta} M'_i$, otherwise $M \xrightarrow[\text{U}]{\beta} M'$ according to the definition of $\xrightarrow[\text{U}]{\beta}$ (Def. 20). Therefore, $M_i \xrightarrow[\text{U}]{\beta} M'_i$ and so, by *i.h.*, $\omega(M_i) = \omega(M'_i)$. Thus, $\omega(M) = \lambda \vec{x}.x\omega(M_1) \dots \omega(M_i) \dots \omega(M_p) = \lambda \vec{x}.x\omega(M_1) \dots \omega(M'_i) \dots \omega(M_p) = \omega(M')$.

► **Proposition 52 (Asymptotic Completeness).**

1. $M \rightarrow_\beta^\infty \mathbf{r}$ implies $M \xrightarrow[\text{U}]{\beta}^\infty \mathbf{r}$, because
 - *U-Factorization of \rightarrow_β* : $M \rightarrow_\beta^* N$ implies $M \xrightarrow[\text{U}]{\beta}^* \cdot \xrightarrow[\text{U}]{\beta}^* N$
 - *obs-neutrality*: $M \xrightarrow[\text{U}]{\beta} M'$ then $\text{obs}M = \text{obs}M'$.
2. $M \xrightarrow[\text{U}]{\beta}^\infty \mathbf{r}$ implies $(M \xrightarrow[\text{U}]{\beta}^\infty \mathbf{s}$ and $\mathbf{r} \leq \mathbf{s})$ ◀

17:24 Strategies for Asymptotic Normalization

Proof.

1. Factorization is that for $\xrightarrow{v}\beta$ (details of the proof are in [15]). **obs**-neutrality is immediate consequence of Lemma 51.
2. It follows by Lemma 41, and the fact that **obs** is monotonic. ◀

► **Remark 53 (Unique Limit).** If we take for head reduction the standard one (as in [7]), then \xrightarrow{h} is deterministic. Otherwise, if we take $\xrightarrow{h}\beta$ as defined in Sect. 4.1, it is easily verified that \xrightarrow{h} has the **obs**-diamond property.

\xrightarrow{h} is **obs**-complete for \rightarrow_β (by Points 1. and 2.). Hence we conclude by Thm. 16:

► **Theorem 54 (Main, Böhm Trees).** \xrightarrow{h} is a (multi-step) **obs**-normalizing strategy for \rightarrow_β , and $M \xrightarrow{h}^\infty \text{BT}(M)$.