Probabilistic Rewriting: Normalization, Termination, and Unique Normal Forms

Claudia Faggian
Université de Paris, IRIF, CNRS, F-75013 Paris, France
faggian@irif.fr

Abstract

While a mature body of work supports the study of rewriting systems, abstract tools for Probabilistic Rewriting are still limited. We study in this setting questions such as uniqueness of the result (unique limit distribution) and normalizing strategies (is there a strategy to find a result with greatest probability?). The goal is to have tools to analyse the operational properties of *probabilistic* calculi (such as probabilistic lambda-calculi) whose evaluation is also non-deterministic, in the sense that different reductions are possible.

2012 ACM Subject Classification Theory of computation → Probabilistic computation; Theory of computation → Rewrite systems; Theory of computation → Logic

Keywords and phrases probabilistic rewriting, PARS, abstract rewriting systems, confluence, probabilistic lambda calculus

Digital Object Identifier 10.4230/LIPIcs.FSCD.2019.19

Related Version `http://arxiv.org/abs/1804.05578`

Acknowledgements This work benefitted of fruitful discussions with U. Dal Lago, B. Valiron, and T. Leventis. I also wish to thank the anonymous referees for valuable comments and suggestions.

1 Introduction

Rewriting Theory [39] is a foundational theory of computing. Its impact extends to both the theoretical side of computer science, and the development of programming languages. A clear example of both aspects is the paradigmatic term rewriting system, λ-calculus, which is also the foundation of functional programming. *Abstract Rewriting Systems (ARS)* are the general theory which captures the common substratum of rewriting theory, independently of the particular structure of the objects. It studies properties of terms transformations, such as normalization, termination, unique normal form, and the relations among them. Such results are a powerful set of tools which can be used when we study the computational and operational properties of any calculus or programming language. Furthermore, the theory provides tools to study and compare strategies, which become extremely important when a system *may* have reductions leading to a normal form, but *not necessarily*. Here we need to know: is there a strategy which is guaranteed to lead to a normal form, if any exists (*normalizing* strategies)? Which strategies diverge if at all possible (*perpetual* strategies)?

Probabilistic Computation models uncertainty. Probabilistic models such as automata [34], Turing machines [37], and the λ-calculus [36] exist since long. The pervasive role it is assuming in areas as diverse as robotics, machine learning, natural language processing, has stimulated the research on probabilistic programming languages, including functional languages [27, 35, 32] whose development is increasingly active. A typical programming language supports at least discrete distributions by providing a probabilistic construct which models sampling from a distribution. This is also the most concrete way to endow the λ-calculus with probabilistic choice [13, 10, 16]. Within the vast research on models of probabilistic systems, we wish to mention that probabilistic rewriting is the explicit base of PMaude [1], a language for specifying probabilistic concurrent systems.

4th International Conference on Formal Structures for Computation and Deduction (FSCD 2019).
Editor: Herman Geuvers; Article No. 19; pp. 19:1–19:25
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Probabilistic Rewriting. Somehow surprisingly, while a large and mature body of work supports the study of rewriting systems – even infinitary ones [12, 24] – work on the abstract theory of *probabilistic* rewriting systems is still sparse. The notion of *Probabilistic* Abstract Reduction Systems (PARS) has been introduced by Bournez and Kirchner in [5], and then extended in [4] to account for non-determinism. Recent work [7, 15, 25, 3] shows an increased research interest. The key element in *probabilistic* rewriting is that even when the probability that a term leads to a normal form is 1 (*almost sure termination*), that degree of certitude is typically not reached in any finite number of steps, but it appears as a limit. Think of a rewrite rule (as in Fig. 1) which rewrites c to either the value $\mathtt{T}$ or c, with equal probability 1/2. We write this $c \to \{c^{1/2}, \mathtt{T}^{1/2}\}$. After n steps, c reduces to $\mathtt{T}$ with probability $\frac{1}{2} + \frac{1}{2^2} + \cdots + \frac{1}{2^n}$. Only at the limit this computation terminates with probability 1 .

The most well-developed literature on PARS is concerned with methods to prove almost sure termination, see e.g. [4, 19, 3] (this interest matches the fact that there is a growing body of methods to establish `AST` [2, 20, 22, 30]). However, considering rewrite rules subject to probabilities opens numerous other questions on PARS, which motivate our investigation.

We study a rewrite relation on distributions, which describes the evolution of a probabilistic system, for example a probabilistic program P. The *result* of the computation is a distribution β over all the possible values of P. The intuition (see [27]) is that the program P is executed, and random choices are made by sampling. This process eventually defines a distribution β over the various outputs that the program can produce. We write this $P \overset{\infty}{\Rrightarrow} \beta$.

What happens if the *evaluation* of a term P is also *non-deterministic*? Remember that non-determinism arises naturally in the λ-calculus, because a term may have several redexes. This aspect has practical relevance to programming. Together with the fact that the result of a terminating computation is unique, it is key to the inherent parallelism of functional programs (see *e.g.* [29]). When assuming non-deterministic evaluation, several questions on PARS arise naturally. For example: (1.) when is the result unique? (naively, if $P \overset{\infty}{\Rrightarrow} \alpha$ and $P \overset{\infty}{\Rrightarrow} \beta$, is $\alpha = \beta$?) (2.) Do all rewrite sequences from the same term have the same probability to reach a result? (3.) If not, does there exist a strategy to find a result with greatest probability?

Such questions are relevant not only to the theory, but also to the practice of computing. We believe that to study them, we can advantageously adapt techniques from Rewrite Theory. However, we *cannot assume that standard properties of ARS hold for PARS.* The game-changer is that termination appears as a *limit.* In Sec. 4.4 we show that a well-known ARS property, Newman's Lemma, does not hold for PARS. This is not surprising; indeed, Newman's Lemma is known not to hold in general for infinitary rewriting [23, 26]. Still, our counter-example points out that moving from ARS to PARS is non-trivial. There are two main issues: we need to find the *right formulation* and the *right proof technique.* It seems especially important to have a collection of proof methods which apply well to PARS.

Content and contributions. Probability is concerned with *asymptotic* behaviour: what happens not after a finite number n of steps, but *when n tends to infinity.* In this paper we focus on the asymptotic behaviour of rewrite sequences *with respect to normal forms.* We study computational properties such as (1.),(2.),(3.) above. We do so with the point of view of ARS, aiming for properties which *hold independently* of the specific nature of the rewritten objects; the purpose is to have tools which apply to any probabilistic rewriting system.

After introducing and motivating our formalism (Sec. 2 and 3), in Sec. 4, we extend to the probabilistic setting the notions of *Normalization (`WN`), Termination (`SN`) and Unique Normal Form* (`UN`). In the rest of the paper, we provide methods and criteria to establish these properties, and we uncover relations between them. In particular, we study *normalizing strategies*. To do so, we extend to the probabilistic setting a proposal by Van Oostrom [40], which is based on Newman's property of Random Descent [31, 40, 41] (see Sec. 1.1). The Random Descent method turns out to provide proof techniques which are well suited to PARS. Specific contributions are the following.

- We propose an analogue of `UN` for PARS. This is not obvious; the question was already studied in [15] for PARS which are `AST`, but their solution does not extend to general PARS.
- We investigate the classical ARS method to prove `UN` via *confluence*. It turns out that the notion of confluence does not need to be as strong as the classical case would suggest, broadening its scope of application. Subtle aspects appear when dealing with limits, and the proof demand specific techniques.
- We develop a probabilistic extension of the ARS notions of Random Descent ($\mathcal{E}$-RD, Sec. 5) and of being *better* ($\mathcal{R}$-better, Sec. 7) as tools to analyze and compare strategies, in analogy to their counterpart in [40]. Both properties are here parametric with respect to a chosen event of interest. $\mathcal{E}$-RD entails that all rewrite sequences from a term lead to the *same result*, in the *same expected number of steps* (the average of number of steps, weighted w.r.t. probability). $\mathcal{R}$-better offers a method to compare strategies ("strategy $\mathcal{S}$ is always better than strategy $\mathcal{T}$") w.r.t. the *probability* of reaching a result and the *expected time* to reach a result. It provides a sufficient criterion to establish that a strategy is *normalizing* (resp. *perpetual*) *i.e.* the strategy is guaranteed to lead to a result with maximal (resp. minimal) probability. A significant technical feature (inherited from [40]) is that both notions of $\mathcal{E}$-RD and $\mathcal{R}$-better come with a characterization via a *local condition* (in ARS, a typical example of a local vs global condition is local confluence vs confluence).

We apply these methods to study a probabilistic λ-calculus, which we discuss below together with the notion of Random Descent. A deeper example of application to probabilistic λ-calculus is in [18]; we discuss it in Sec.8 "Further work and applications".

▶ Remark (On the term *Random* Descent). Please note that in [31], the term *Random* refers to non-determinism (in the choice of the redex), *not to randomized* choice.

Related work. We discuss related work in the context of PARS [4, 5]. We are not aware of any work which investigates *normalizing strategies* (or *normalization* in general, rather than termination). Instead, *confluence* in probabilistic rewriting has already drawn interesting work. A notion of confluence for a probabilistic rewrite system defined over a λ-calculus is studied in [14, 9]; in both case, the probabilistic behavior corresponds to measurement in a quantum system. The work more closely related to our goals is [15]. It studies confluence of non-deterministic PARS in the case of finitary termination (being finitary is the reason why a Newman's Lemma holds), and in the case of `AST`. As we observe in Sec. 4.3, their notion of unique limit distribution (if α, β are limits, then $\alpha = \beta$), while simple, it is not an analogue of `UN` for general PARS; we extend the analysis beyond `AST`, to the general case, which arises naturally when considering probabilistic λ-calculus. On confluence, we also mention [25], whose results however do not cover *non-deterministic PARS*; the probability of the limit distribution is concentrated in a single element, in the spirit of Las Vegas Algorithms. [25] revisits results from [5], while we are in the non-deterministic framework of [4].

The way we define the *evolution of PARS*, via the one-step relation $\rightrightarrows$, follows the approach in [7], which also contains an embryo of the current work (a form of diamond property); the other results and developments are novel. A technical difference with [7] is that for the formalism to be general, a refinement is necessary (see Sec. 2.2); the issue was first pointed out in [15]. Our refinement is a variation of the one introduced (for the same reasons) in [3]; we however do not strictly adopt it, because we prefer to use a standard definition of distribution. [3] demonstrates the equivalence with the approach in [4].

1.1 Key notions

Random Descent. Newman's Random Descent (RD) [31] is an ARS property which guarantees that normalization suffices to establish both termination and uniqueness of normal forms. Precisely, if an ARS has random descent, paths to a normal form do not need to be unique, but they have *unique length.* In its essence: *if a normal form exists, all rewrite sequences lead to it, and all have the* same length[1]. While only few systems directly verify it, RD is a powerful ARS tool; a typical use in the literature is to prove that *a strategy* has RD, to conclude that it is *normalizing.* A well-known property which implies RD is a form of diamond: "$\leftarrow \cdot \rightarrow \subseteq (\rightarrow \cdot \leftarrow) \cup =$".

In [40] Von Oostrom defines a characterization of RD by means of a *local* property and proposes RD as a uniform method to (locally) compare strategies for normalization and minimality (resp. perpetuality and maximality). [41] extends the method and abstracts the notion of length into a notion of measure. In Sec. 5 and 7 we develop similar methods in a *probabilistic* setting. The analogous of *length*, is the *expected number of steps* (Sec. 5.1).

Probabilistic Weak λ-calculus. A notable example of system which satisfies RD is the pure untyped λ-calculus endowed with call-by-value (CbV) weak evaluation. *Weak* [21, 6] means that reduction does not evaluate function bodies (*i.e.* the scope of λ-abstractions). We recall that weak CbV is the basis of the ML/CAML family of functional languages (and of most probabilistic functional languages). Because of RD, weak CbV λ-calculus has *striking properties* (see e.g. [8] for an account). First, if a term M has a normal form N, any rewrite sequence will find it; second, the number n of steps such that $M \rightarrow^n N$ is always the same.

In Sec. 6, we study a probabilistic extension of weak CbV, $\Lambda_\oplus^{\mathtt{weak}}$. We show that it has analogous properties to its classical counterpart: all rewrite sequences converge to the same result, in the same *expected* number of steps.

Local vs global conditions. To work *locally* means to reduce a test problem which is global, *i.e.*, quantified over *all rewrite sequences* from a term, to local properties (quantified only over *one-step reductions* from the term), thus reducing the space of search when testing.

A paradigmatic example of a global property is confluence (CR: $b \ {}^*\!\!\leftarrow a \rightarrow^* c \Rightarrow \exists d$ s.t. $b \rightarrow^* d \ {}^*\!\!\leftarrow c$). Its global nature makes it difficult to establish. A standard way to factorize the problem is: (1.) prove termination and (2.) prove *local* confluence (WCR: $b \leftarrow a \rightarrow c \Rightarrow \exists d$ s.t. $b \rightarrow^* d \ {}^*\!\!\leftarrow c$). This is exactly *Newman's lemma: Termination + WCR ⇒ CR.* The beauty of Newman's lemma is that a global property (CR) is guaranteed by a local property (WCR). Locality is also the strength and beauty of the RD method. While Newman's lemma fails in a probabilistic setting (see Sec. 4.4), RD methods can be adapted (Sec. 5 and 7).

[1] or, in Newman's original terminology: the end-form is reached by *random descent* (whenever $x \rightarrow^k y$ and $x \rightarrow^n u$ with u in normal form, all maximal reductions from y have length $n - k$ and end in u).

1.2 Probabilistic λ-calculus and (Non-)Unique Result

Rewrite theory provides numerous tools to study uniqueness of normal forms, as well as techniques to study and compare strategies. This is not the case in the probabilistic setting. Perhaps a reason is that when extending the λ-calculus with a choice operator, confluence is lost, as was observed early [11]; we illustrate it in Example 1.1 and 1.2, which is adapted from [11, 10]. The way to deal with this issue in probabilistic λ-calculi (e.g. [13, 10, 16]) has been to *fix a deterministic reduction strategy*, typically "leftmost-outermost". To fix a strategy is not satisfactory, neither for the theory nor the practice of computing. To understand why this matters, recall for example that confluence of the λ-calculus is what makes functional programs inherently parallel: every sub-expression can be evaluated in parallel, still, we can reason on a program using a deterministic sequential model, because the result of the computation is independent of the evaluation order (we refer to [29], and to Harper's text "Parallelism is not Concurrency" for discussion on *deterministic* parallelism, and how it differs from concurrency). Let us see what happens in the probabilistic case.

▶ **Example 1.1** (Confluence failure)**.** Let us consider the untyped λ-calculus extended with a binary operator $\oplus$ which models probabilistic choice. Here $\oplus$ is just flipping a fair coin: $M \oplus N$ reduces to either M or N with equal probability $1/2$; we write this as $M \oplus N \to \{M^{\frac{1}{2}}, N^{\frac{1}{2}}\}$.

Consider the term PQ, where $P = (\lambda x.x)(\lambda x.x \texttt{ XOR } x)$ and $Q = (\texttt{T} \oplus \texttt{F})$; here XOR is the standard constructs for the exclusive OR, T and F are terms which code the booleans.

- If we evaluate P and Q independently, from P we obtain $\lambda x.(x \texttt{ XOR } x)$, while from Q we have either T or F, with equal probability $1/2$. By composing the partial results, we obtain $\{(\texttt{T XOR T})^{\frac{1}{2}}, (\texttt{F XOR F})^{\frac{1}{2}}\}$, and therefore $\{\texttt{F}^1\}$.
- If we evaluate PQ sequentially, in a standard left-most outer-most fashion, PQ reduces to $(\lambda x.x \texttt{ XOR } x)Q$ which reduces to $(\texttt{T} \oplus \texttt{F}) \texttt{ XOR } (\texttt{T} \oplus \texttt{F})$ and eventually to $\{\texttt{T}^{\frac{1}{2}}, \texttt{F}^{\frac{1}{2}}\}$.

▶ **Example 1.2.** The situation becomes even more complex if we examine also the possibility of diverging; try the same experiment as above on the term PR, with $R = (\texttt{T} \oplus \texttt{F}) \oplus \Delta\Delta$ (where $\Delta = \lambda x.xx$). Proceeding as before, we now obtain either $\{\texttt{F}^{\frac{1}{2}}\}$ or $\{\texttt{T}^{\frac{1}{8}}, \texttt{F}^{\frac{1}{8}}\}$.

We do not need to loose the features of λ-calculus in the *probabilistic* setting. In fact, while some care is needed, determinism of the evaluation *can be relaxed* without giving up uniqueness of the result: the calculus we introduce in Sec. 6 is an example (we relax determinism to RD); we fully develop this direction in further work [18]. To be able to do so, we *need abstract tools and proof techniques* to analyze *probabilistic* rewriting. The same need for theoretical tools holds, more in general, whenever we desire to have a probabilistic language which allows for *deterministic parallel reduction.*

In this paper we focus on *uniqueness of the result*, rather than confluence, which is an important and sufficient, but not necessary property.

2 Probabilistic Abstract Rewriting System

We assume the reader familiar with the basic notions of rewrite theory (such as Ch. 1 of [39]), and of *discrete* probability theory. We review the basic language of both. We then recall the definition of PARS from [5, 4], and explain on examples how a system described by a PARS evolves. This will motivate the formalism which we introduce in Sec. 3.

Basics on ARS. An *abstract rewrite system (ARS)* is a pair $\mathcal{C} = (C, \to)$ consisting of a set C and a binary relation $\to$ on C; $\to^*$ denotes the transitive reflexive closure of $\to$. An element $u \in C$ is in **normal form** if there is no c with $u \to c$; $\mathtt{NF}_{\mathcal{C}}$ denotes the set of the normal forms of $\mathcal{C}$. If $c \to^* u$ and $u \in \mathtt{NF}_{\mathcal{C}}$, we say c has a normal form u. $\mathcal{C}$ has the property of **unique normal form (*with respect to reduction*)**(UN) if $\forall u, v \in \mathtt{NF}_{\mathcal{C}}, \big(c \to^* u \,\&\, c \to^* v \Rightarrow u = v\big)$. $\mathcal{C}$ has the **normal form property** (NFP) if $\forall b, c \in C, \forall u \in \mathtt{NF}_{\mathcal{C}}, \big(b \to^* c \,\&\, b \to^* u \Rightarrow c \to^* u\big)$. NFP implies UN. The fact that an ARS has unique normal forms implies neither that all terms have a normal form, nor that if a term has a normal form, each rewrite sequence converges to it. A term c is **terminating**[2] (aka **strongly normalizing**, SN), if it has no infinite sequence $c \to c_1 \to c_2 ...$; it is **normalizing** (aka **weakly normalizing**, WN), if it has a normal form. These are all important properties to establish about an ARS, as it is important to have a rewrite strategy which finds a normal form, if it exists.

Basics on Probabilities. The intuition is that random phenomena are observed by means of experiments (running a probabilistic program is such an experiment); each experiment results in an outcome. The collection of all possible outcomes is represented by a set, called the **sample space** Ω. When the sample space Ω is *countable*, the theory is simple. A *discrete probability space* is given by a pair (Ω, μ), where Ω is a *countable* set, and μ is a **discrete probability distribution** on Ω, *i.e.* a function $\mu : \Omega \to [0, 1]$ such that $\sum_{\omega \in \Omega} \mu(\omega) = 1$. A probability measure is assigned to any subset $A \subseteq \Omega$ as $\mu(\mathcal{A}) = \sum_{\omega \in A} \mu(\omega)$. In the language of probabilists, a subset of Ω is called an *event*.

▶ **Example 2.1** (Die). Consider tossing a die once. The space of possible outcomes is the set $\Omega = \{1, 2, 3, 4, 5, 6\}$. The probability μ of each outcome is $1/6$. The event *"result is odd"* is the subset $A = \{1, 3, 5\}$, whose probability is $\mu(A) = 1/2$.

Each *function* $F : \Omega \to \Delta$, where Δ is another countable set, **induces a probability distribution** μ^F on Δ by composition: $\mu^F(d) := \mu(F^{-1}(d))$ *i.e.* $\mu\{\omega \in \Omega : F(\omega) = d\}$. Thus (Δ, μ^F) is also a probability space. In the language of probability theory, F is called a *discrete random variable* on (Ω, μ). The **expected value** (also called the expectation or mean) of a random variable F is the weighted (in proportion to probability) average of the possible values of F. Assume $F : \Omega \to \Delta$ discrete and $g : \Delta \to \mathbb{R}$ a non-negative function, then $E(g(F)) = \sum_{d \in \Delta} g(d)\mu_F(d)$.

(Sub)distributions: operations and notation. We need the notion of subdistribution to account for unsuccessful computations and partial results. Given a countable set Ω, a function $\mu : \Omega \to [0, 1]$ is a probability **subdistribution** if $\|\mu\| := \sum_{\omega \in \Omega} \mu(\omega) \le 1$. We write $\mathtt{DST}(\Omega)$ for the set of subdistributions on Ω. With a slight abuse of language, we often use the term distribution also for subdistribution. The *support* of μ is the set $\mathtt{Supp}(\mu) = \{a \in \Omega \mid \mu(a) > 0\}$. $\mathtt{DST}^{\mathtt{F}}(\Omega)$ denotes the set of $\mu \in \mathtt{DST}(\Omega)$ with *finite support*.

$\mathtt{DST}(\Omega)$ is equipped with the **order relation** of functions : $\mu \le \rho$ if $\mu(a) \le \rho(a)$ for each $a \in \Omega$. **Multiplication** for a scalar $(p \cdot \mu)$ and **sum** $(\sigma + \rho)$ are defined as usual, $(p \cdot \mu)(a) = p \cdot \mu(a)$, $(\sigma + \rho)(a) = \sigma(a) + \rho(a)$, provided $p \in [0, 1]$, and $\|\sigma\| + \|\rho\| \le 1$.

We adopt the following **convention**: if $\Omega' \subseteq \Omega$, and $\mu \in \mathtt{DST}(\Omega')$, we also write $\mu \in \mathtt{DST}(\Omega)$, with the implicit assumption that the extension behaves as μ on Ω', and is 0 otherwise. In particular, we identify a subdistribution and its support.

[2] Please observe that the *terminology is community-dependent*. In logic: Strong Normalization, Weak Normalization, Church-Rosser (hence the *standard abbreviations* SN, WN, CR). In computer science: Termination, Normalization, Confluence.

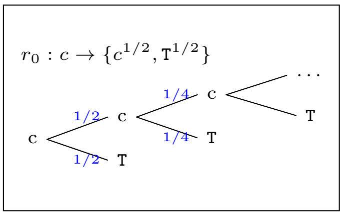

Figure 1 Almost Sure Termination.

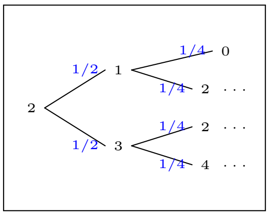

Figure 2 Deterministic PARS.

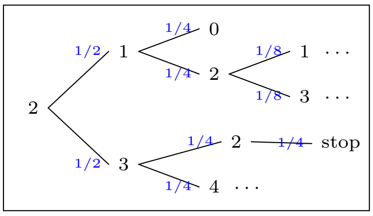

Figure 3 Non-deterministic PARS.

▶ **Notation 2.2** (Representation). *We represent a (sub)distribution by explicitly indicating the support, and (as superscript) the probability assigned to each element by* μ. *We write* $\mu = \{a_0^{p_0}, \ldots, a_n^{p_n}\}$ *if* $\mu(a_0) = p_0, \ldots, \mu(a_n) = p_n$ *and* $\mu(a_j) = 0$ *otherwise.*

2.1 Probabilistic Abstract Rewrite Systems (PARS)

A *probabilistic abstract rewrite system (PARS)* is a pair $\mathcal{A} = (A, \to)$ of a countable set A and a relation $\to \subseteq A \times \mathtt{DST}^{\mathtt{F}}(A)$ such that for each $(a, \beta) \in \to$, $\|\beta\| = 1$. We write $a \to \beta$ for $(a, \beta) \in \to$ and we call it a *rewrite step*, or a *reduction*. An element $a \in A$ is in *normal form* if there is no β with $a \to \beta$. We denote by $\mathtt{NF}_{\mathcal{A}}$ the set of the normal forms of $\mathcal{A}$ (or simply $\mathtt{NF}$ when $\mathcal{A}$ is clear). A PARS is *deterministic* if, for all a, there is at most one β with $a \to \beta$.

▶ Remark. The intuition behind $a \to \beta$ is that the rewrite step $a \to b$ ($b \in A$) has probability $\beta(b)$. The total probability given by the sum of all steps $a \to b$ is 1.

Probabilistic vs Non-deterministic. It is important to have clear the distinction between probabilistic choice (which *globally happens with certitude*) and non-deterministic choice (which leads to different distributions of outcomes.) Let us discuss some examples.

▶ **Example 2.3** (A deterministic PARS). Fig. 2 shows a simple random walk over $\mathbb{N}$, which describes a gambler starting with 2 points and playing a game where every time he either gains 1 point with probablity 1/2 or looses 1 point with probability 1/2. This system is encoded by the following PARS on $\mathbb{N}$: $n + 1 \to \{n^{1/2}, (n+2)^{1/2}\}$. Such a PARS is *deterministic*, because for every element, at most one choice applies. Note that 0 is a normal form.

▶ **Example 2.4** (A non-deterministic PARS). Assume now (Fig. 3) that the gambler of Example 2.3 is also given the possibility to stop at any time. The two choices are here encoded as follows: $n + 1 \to \{n^{1/2}, (n+2)^{1/2}\}$, $\quad n + 1 \to \{\mathtt{stop}^1\}$.

2.2 Evolution of a system described by a PARS

We now need to explain how a system which is described by a PARS evolves. An option is to follow the stochastic evolution of a single run, *a sampling at a time*, as we have done in Fig. 1, 2, and 3. This is the approach in [4], where non-determinism is solved by the use of policies. Here we follow a different (though equivalent) way (see the Related Work Section). We describe the possible states of the system, at a certain time t, *globally*, as a distribution on the space of all terms. The evolution of the system is then a sequence of distributions. Since all the probabilistic choices are taken together, the only source of choice in the evolution is non-determinism. This global approach allows us to deal with non-determinism by using techniques which have been developed in Rewrite Theory. Before introducing the formal definitions, we informally examine some examples, and point out why some care is needed.

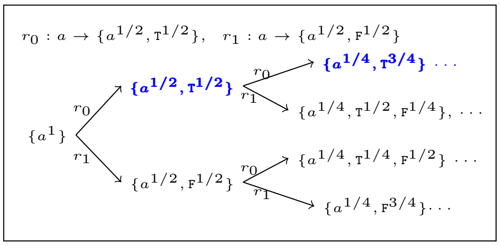

Figure 4 Ex.2.6 (non-deterministic PARS).

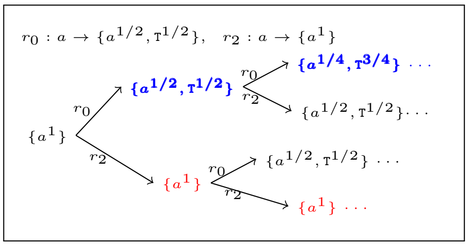

Figure 5 Ex.2.7 (non-deterministic PARS).

▶ **Example 2.5** (Fig.1 continued)**.** The PARS described by the rule $r_0 : c \to \{c^{1/2}, \mathtt{T}^{1/2}\}$ (in Fig. 1) evolves as follows: $\{c\}, \{c^{1/2}, \mathtt{T}^{1/2}\}, \{c^{1/4}, \mathtt{T}^{3/4}\}, \dots$.

▶ **Example 2.6** (Fig.4)**.** Fig. 4 illustrates the possible evolutions of a non-deterministic system which has two rules: $r_0 : a \to \{a^{1/2}, \mathtt{T}^{1/2}\}$ and $r_1 : a \to \{a^{1/2}, \mathtt{F}^{1/2}\}$. The arrows are annotated with the chosen rule.

▶ **Example 2.7** (Fig.5)**.** Fig. 5 illustrates the possible evolutions of a system with rules $r_0 : a \to \{a^{1/2}, \mathtt{T}^{1/2}\}$ and $r_2 : a \to \{a^1\}$.

If we look at Fig. 3, we observe that after two steps, there are *two distinct occurrences* of the element 2, which live in *two different runs* of the program: the run `2.1.2`, and the run `2.3.2`. There are two possible transitions from each 2. The next transition only depends on the fact of having 2, not on the run in which 2 occurs: its history is only a way to distinguish the occurrence. For this reason, given a PARS $(A, \to)$, we keep track of *different occurrences* of an element $a \in A$, but not necessarily of the history. Next section formalizes these ideas.

Markov Decision Processes. To understand our distinction between occurrences of $a \in A$ in different paths, it is helpful to think how a system is described in the framework of Markov Decision Processes (MDP) [33]. Indeed, in the same way as ARS correspond to transition systems, PARS correspond to probabilistic transitions. Let us regard a PARS step $r : a \to \beta$ as a probabilistic transition (r is here a name for the rule). Let assume $a_0 \in \mathcal{A}$ is an initial state. In the setting of MDP, a typical element (called *sample path*) of the sample space Ω is a sequence $\omega = (a_0, r_0, a_1, r_1 \dots)$ where $r_0 : a_0 \to \beta_1$ is a rule, $a_1 \in \mathtt{Supp}(\beta_1)$ an element, $r_1 : a_1 \to \beta_1$, and so on. The index $t = 0, 1, 2, \dots, n, \dots$ is interpreted as *time*. On Ω various random variables are defined; for example, $X_t = a_t$, which represents the state at time t. The sequence $\langle X_t \rangle$ is called a stochastic process.

3 A Formalism for Probabilistic Rewriting

We introduce a formalism to describe the evolution of a system described by a PARS. From now on, we assume A to be a countable set on which a PARS $(A, \to)$ is defined.

The sample space. Let $\mathtt{m}$ be a list over A, and mA the collection of all such lists. More formally, we fix a countable *index set* $\mathbb{S}$, and let $\mathtt{m} = \{(j, \mathtt{m}_j) \mid j \in \mathbb{S}, \mathtt{m}_j \in A\}$ be the graph of a function from $J \subseteq \mathbb{S}$ to A ($j \mapsto \mathtt{m}_j$). We denote by mA the collection of all such $\mathtt{m}$. $\mathtt{DST}^{\mathtt{F}}(mA) := \bigcup_{\mathtt{m} \in mA} \mathtt{DST}^{\mathtt{F}}(\mathtt{m})$ is the collection of finitely supported distributions μ on $\mathtt{m} \in mA$ (*i.e.* $\mu : \mathtt{m} \to [0, 1]$, with $(j, a) \mapsto p$). For concreteness, here we assume $\mathbb{S} = \mathbb{N}$. Hence, if J is finite, $\mathtt{m}$ is simply a *list* over A.

$$\texttt{flat} : (j, a) \mapsto a$$

$$(-)^{\texttt{flat}} : \begin{array}{ccc} \texttt{DST}^{\texttt{F}}(mA) & \to & \texttt{DST}^{\texttt{F}}(A) \\ \mu & \mapsto & \mu^{\texttt{flat}} \end{array}$$

where $\mu^{\texttt{flat}}(a) = \mu\{\texttt{flat}^{-1}(a)\} = \sum_{(j,a)\in \texttt{m}} \mu(j, a)$

Figure 6 Flattening.

$$\mathcal{I} : a \mapsto (j, a)$$

$$(-)^{\mathcal{I}} : \begin{array}{ccc} \texttt{DST}^{\texttt{F}}(A) & \to & \texttt{DST}^{\texttt{F}}(mA) \\ \beta & \mapsto & \beta^{\mathcal{I}} \end{array}$$

where $\beta^{\mathcal{I}}(j, a) = \beta\{\mathcal{I}^{-1}(j, a)\} = \beta(a)$

Figure 7 Embedding.

▶ **Notation 3.1.** *If $\mu \in \texttt{DST}^{\texttt{F}}(mA)$, we write its support as a list. We write $[a, a, b, b]$ for $\{(1, a), (2, a), (3, b), (4, b)\}$ and $[a^{1/4}, a^{1/4}, b^{1/6}, b^{1/3}]$ for $\{(1, a)^{1/4}, (2, a)^{1/4}, (3, b)^{1/6}, (4, b)^{1/3}\}$*

▶ Remark 3.2 (Index Set). The role of indexing is only to distinguish different occurrences; the specific order is irrelevant. We use $\mathbb{N}$ as index set for simplicity. Another natural instance of $\mathbb{S}$ is A^* *i.e.* the set of finite sequences on A. This way, occurrences are labelled by their path, which allows a direct connection with the sample space of Markov Decision Processes [33] we mention in 2.2 (see Appendix).

Given the PARS $\mathcal{A} = (A, \to)$, we work with two families of probability spaces: (A, β), where $\beta \in \texttt{DST}(A)$ (used *e.g.* to describe a rewrite step) and $(\texttt{m}, \mu)$, where $\texttt{m} \in mA$ and $\mu \in \texttt{DST}^{\texttt{F}}(\texttt{m})$.

Letters Convention. we reserve the letters α, β, γ for distributions in $\texttt{DST}(A)$, and the letters $\mu, \nu, \sigma, \tau, \rho, \xi$ for distributions in $\texttt{DST}^{\texttt{F}}(mA)$.

Embedding and Flattening. we move between A and subsets of $\mathbb{N} \times A$ via the maps $\texttt{flat}(-) : \texttt{m} \to A$ and $\mathcal{I} : A \to \texttt{n}$ (Fig. 6 and 7), where to define an injection $\mathcal{I}$, we fix an enumeration $\texttt{n} : \mathbb{N} \to A$, and identify $\texttt{n}$ with its graph. Given a distribution $\mu \in \texttt{DST}^{\texttt{F}}(\texttt{m})$, the function $\texttt{flat}$ induces the distribution $\mu^{\texttt{flat}} \in \texttt{DST}^{\texttt{F}}(A)$ (Fig. 6); conversely, given $\beta \in \texttt{DST}^{\texttt{F}}(A)$, the function $\mathcal{I} : A \to \texttt{n} \in mA$ induces the distributions $\beta^{\mathcal{I}} \in \texttt{DST}^{\texttt{F}}(mA)$ (Fig. 7). Recall that in Sec. 2 we already reviewed how functions induce distributions; indeed, with that language, $\texttt{flat}(-) : \texttt{m} \to A$ and $\mathcal{I} : A \to \texttt{n}$ are random variables.

▶ **Example 3.3.** Assume $\beta = \{a^{0.3}, b^{0.2}, c^{0.5}\}$, and an enumeration of $\{a, b, c\}$. Then $\beta^{\mathcal{I}} = \{(a, 1)^{0.3}, (b, 2)^{0.2}, (c, 3)^{0.5}\}$ which we also write $[a^{0.3}, b^{0.2}, c^{0.5}]$.

Disjoint sum $\biguplus$. The disjoint sum of lists is simply their concatenation. The disjoint sum of sets in mA and of the corresponding distributions is easily defined.

The rewriting relation $\rightrightarrows$. Let $\mathcal{A} = (A, \to)$ be a PARS. We now define a binary relation $\rightrightarrows$ on $\texttt{DST}^{\texttt{F}}(mA)$, which is obtained by lifting the relation $\to$. Several natural choices are possible. Here, we choose a lifting which forces *all* non-terminal elements to be reduced. This plays an important role for the development of the paper, as it corresponds to the the key notion of *one step* reduction in classical ARS (see discussion in Sec. 8).

▶ **Definition 3.4** (Lifting). *Given a relation $\to \subseteq A \times \texttt{DST}(A)$, its lifting to a relation $\rightrightarrows \subseteq \texttt{DST}^{\texttt{F}}(mA) \times \texttt{DST}^{\texttt{F}}(mA)$ is defined by the following rules, where for readibility we use Notation 3.1.*

$$\frac{a \in \mathit{NF}_{\mathcal{A}}}{[a^1] \rightrightarrows [a^1]}\ L1 \qquad \frac{a \to \beta \in \mathcal{A}}{[a^1] \rightrightarrows \beta^{\mathcal{I}}}\ L2 \qquad \frac{\left([\texttt{m}_j^1] \rightrightarrows \mu_j\right)_{j\in J}}{[\texttt{m}_j^{p_j} \mid j \in J] \rightrightarrows \biguplus_{j\in J} p_j \cdot \mu_j}\ L3$$

In rule $(L2)$, $\beta^{\mathcal{I}}$ is the result of embedding $\beta \in \mathtt{DST}^{\mathtt{F}}(A)$ in $\mathtt{DST}^{\mathtt{F}}(mA)$ (see Fig. 7 and Example 3.3). To apply rule $(L3)$, we choose a reduction step from $\mathtt{m}_j$ for *each* $j \in J$. The disjoint sum of all μ_j $(j \in J)$ is weighted with the probability of each $\mathtt{m}_j$.

▶ **Example 3.5.** Let us derive the reduction in Fig. 3.

$$\frac{2 \to \{1^{1/2}, 3^{1/2}\}}{[2^1] \rightrightarrows [1^{1/2}, 3^{1/2}]} \quad \frac{1 \to \{0^{1/2}, 2^{1/2}\} \quad 3 \to \{2^{1/2}, 4^{1/2}\}}{[1^{1/2}, 3^{1/2}] \rightrightarrows [0^{1/4}, 2^{1/4}, 2^{1/4}, 4^{1/4}]} \quad \frac{\dots \quad 2 \to \{stop^1\} \quad 2 \to \{1^{1/2}, 3^{1/2}\} \quad \dots}{[0^{1/4}, 2^{1/4}, 2^{1/4}, 4^{1/4}] \rightrightarrows [..., stop^{1/4}, 1^{1/8}, 3^{1/8}, ...]}$$

Rewrite sequences. We write $\mu_0 \rightrightarrows^* \mu_n$ to indicate that there is a *finite sequence* $\mu_0, \dots, \mu_n$ such that $\mu_i \rightrightarrows \mu_{i+1}$ for all $0 \le i < n$ (and $\mu_0 \rightrightarrows^k \mu_k$ to specify its length k). We write $\langle \mu_n \rangle_{n \in \mathbb{N}}$ to indicate an *infinite rewrite sequence.*

Figures conventions. We depict *any* rewrite relation simply as $\to$; as it is standard, we use $\twoheadrightarrow$ for $\to^*$; solid arrows are universally quantified, dashed arrows are existentially quantified.

Normal Forms. The intuition is that a rewrite sequence describes a computation; a distribution μ_i such that $\mu \rightrightarrows^i \mu_i$ represents a state (precisely, the state at time i) in the evolution of the system with initial state μ. Let $\mu \in \mathtt{DST}^{\mathtt{F}}(mA)$ represents a state of the system. The **probability that the system is in normal form** is described by $\mu^{\mathtt{flat}}(\mathtt{NF}_{\mathcal{A}})$ (recall Example 2.1); the probability that the system is in a specific normal form t is described by $\mu^{\mathtt{flat}}(t)$. It is convenient to denote by μ^{NF} the restriction of $\mu^{\mathtt{flat}}$ to $\mathtt{NF}_{\mathcal{A}}$. Observe that $\|\mu^{\mathsf{NF}}\| = \mu^{\mathtt{flat}}(\mathtt{NF}_{\mathcal{A}}) = \mu^{\mathsf{NF}}(\mathtt{NF}_{\mathcal{A}})$. The probability of reaching a normal form t can only increase in a rewrite sequence (becaluse of (L1) in Def. 3.4). Therefore the following key lemma holds.

▶ **Lemma 3.6.** *If* $\sigma \rightrightarrows \tau$ *then* $\sigma^{\mathsf{NF}} \le \tau^{\mathsf{NF}}$.

Equivalences and Order. In this paper we do not need, and do not define, any equality on lists. If we wanted, the natural one would be equality up to reordering, making lists into multisets; however, here we are rather interested in observing specific events. Given $\mu, \rho \in \mathtt{DST}^{\mathtt{F}}(mA)$, we only conside equivalence and order relations w.r.t. the associated (flat) distribution in $\mathtt{DST}(A)$ and in $\mathtt{DST}(\mathtt{NF}_{\mathcal{A}})$. The order on $\mathtt{DST}(A)$ is the pointwise order (Sec. 2).

▶ **Definition 3.7** (Equivalence and Order). *Let* $\mu, \rho \in \mathtt{DST}^{\mathtt{F}}(mA)$.

1. Flat Equivalence*:* $\mathcal{E}_{flat}(\mu, \rho)$*, if* $\mu^{flat} = \rho^{flat}$*. Similarly,* $\le_{flat} (\mu, \rho)$ *if* $\mu^{flat} \le \rho^{flat}$*.*
2. Equivalence in Normal Form*:* $\mathcal{E}_{\mathsf{NF}}(\mu, \rho)$*, if* $\mu^{\mathsf{NF}} = \rho^{\mathsf{NF}}$*. Similarly,* $\le_{\mathsf{NF}} (\mu, \rho)$*, if* $\mu^{\mathsf{NF}} \le \rho^{\mathsf{NF}}$
3. Equivalence in the $\mathtt{NF}$-norm*:* $\mathcal{E}_{\|\|_N}(\mu, \rho)$*, if* $\|\mu^{\mathsf{NF}}\| = \|\rho^{\mathsf{NF}}\|$*, and* $\le_{\|\|_N} (\mu, \rho)$*, if* $\|\mu^{\mathsf{NF}}\| \le \|\rho^{\mathsf{NF}}\|$

Observe that (2.) and (3.) compare μ and ρ abstracting from any term which is not in normal form; these two will be the relations which matter to us.

▶ **Example 3.8.** Assume $\mathtt{T}$ is a normal form and $a \neq c$ are not. **(1.)** Let $\mu = [\mathtt{T}^{1/2}, \mathtt{T}^{1/2}]$, $\rho = [\mathtt{T}^1]$. $\mathcal{E}(\mu, \rho)$ holds for $\mathcal{E} \in \{\mathcal{E}_{\mathtt{flat}}, \mathcal{E}_{\mathsf{NF}}, \mathcal{E}_{\|\|_{\mathsf{N}}}\}$ because $\mu^{\mathtt{flat}} = \rho^{\mathtt{flat}} = \{\mathtt{T}^1\}$. **(2.)** Let $\mu = [a^{1/2}, \mathtt{T}^{1/2}]$, $\rho = [c^{1/2}, \mathtt{T}^{1/6}, \mathtt{T}^{2/6}]$. $\mathcal{E}_{\mathsf{NF}}(\mu, \rho), \mathcal{E}_{\|\|_{\mathsf{N}}}(\mu, \rho)$ both *hold*, $\mathcal{E}_{\mathtt{flat}}(\mu, \rho)$ does *not.*

The above example illustrates also the following.

▶ **Fact 3.9.** $\mathcal{E}_{flat}(\mu, \rho) \;\Rightarrow\; \mathcal{E}_{\mathsf{NF}}(\mu, \rho) \;\Rightarrow\; \mathcal{E}_{\|\|_N}(\mu, \rho)$*. Similarly for the order relations.*

4 Asymptotic Behaviour and Normal Forms

We examine the asymptotic behaviour of rewrite sequences *with respect to normal forms.* If a rewrite sequence describes a computation, the *result* of the computation is a distribution on the possible outputs of the probabilistic program. We are interested in the result *at the limit*, which is formalized by the (standard) notion of *limit distribution* (Def. 4.2). What is less standard here, and demands care, is that each termhas a set of limits. In the section we investigate the notions of normalization, termination and unique normal form for PARS.

4.1 Limit Distributions

Before introducing limit distributions, we revisit some facts on sequences of bounded functions.

Monotone Convergence. Let $\langle \alpha_n \rangle_{n\in\mathbb{N}}$ be a *non-decreasing sequence* of (sub)distributions over a countable set X (the order on subdistributions is defined pointwise, Sec. 2). For each $t \in X$, the sequence $\langle \alpha_n(t) \rangle_{n\in\mathbb{N}}$ of real numbers is *nondecreasing and bounded*, therefore the sequence has a limit, which is the supremum: $\lim_{n\to\infty} \alpha_n(t) = \sup_n\{\alpha_n(t)\}$. Observe that if $\alpha < \alpha'$ then $\|\alpha\| < \|\alpha'\|$, where we recall that $\|\alpha\| := \sum_{x\in X} \alpha(x)$.

▶ **Fact 4.1.** *Given* $\langle \alpha_n \rangle_{n\in\mathbb{N}}$ *as above, the following properties hold. Define*

$$\beta(t) = \lim_{n\to\infty} \alpha_n(t), \ \forall t \in X$$

1. $\lim_{n\to\infty} \|\alpha_n\| \ = \ \|\beta\|$
2. $\lim_{n\to\infty} \|\alpha_n\| = \sup_n\{\|\alpha_n\|\} \leq 1$
3. β *is a* subdistribution *over* X*.*

Proof.
1. follows from the fact that $\langle \alpha_n \rangle_{n\in\mathbb{N}}$ is a nondecreasing sequence of functions, hence (by Monotone Convergence, see Thm. A.1 in Appendix) we have :

$$\lim_{n\to\infty} \sum_{t\in X} \alpha_n(t) \ = \ \sum_{t\in X} \lim_{n\to\infty} \alpha_n(t)$$

2. is immediate, because the sequence $\langle \|\alpha_n\| \rangle_{n\in\mathbb{N}}$ is nondecreasing and bounded.
3. follows from (1.) and (2.). Since $\|\beta\| = \sup_n \|\alpha_n\| \leq 1$, then β is a subdistribution. ◀

Limit distributions. Let $\langle \mu_n \rangle_{n\in\mathbb{N}}$ be a rewrite sequence. If $t \in \mathtt{NF}_{\mathcal{A}}$, then $\langle \mu_n^{\mathsf{NF}}(t) \rangle_{n\in\mathbb{N}}$ is nondecreasing (by Lemma 3.6); so we can apply Fact 4.1, with $\langle \alpha_n \rangle_{n\in\mathbb{N}}$ now being $\langle \mu_n^{\mathsf{NF}} \rangle_{n\in\mathbb{N}}$.

▶ **Definition 4.2** (Limits). *Let* $\langle \mu_n \rangle_{n\in\mathbb{N}}$ *be a rewrite sequence from* $\mu \in \mathtt{DST}^{\mathsf{F}}(mA)$*. We say*
1. $\langle \mu_n \rangle_{n\in\mathbb{N}}$ ***converges with probability*** $p = \sup_n \|\mu_n^{\mathsf{NF}}\|$.
2. $\langle \mu_n \rangle_{n\in\mathbb{N}}$ ***converges to*** $\beta \in \mathtt{DST}^{\mathsf{F}}(\mathtt{NF}_{\mathcal{A}})$ $\big($*written* $\langle \mu_n \rangle_{n\in\mathbb{N}} \stackrel{\infty}{\Rightarrow} \beta$ $\big)$*, where for* $t \in \mathtt{NF}_{\mathcal{A}}$

$$\beta(t) = sup_n\{\mu_n^{\mathsf{NF}}(t)\}$$

We call β *a* ***limit distribution*** *of* μ*. We write* $\mu \stackrel{\infty}{\Rightarrow} \beta$ *if* μ *has a sequence converging to* β*, and define* $\mathtt{Lim}(\mu) := \{\beta \mid \mu \stackrel{\infty}{\Rightarrow} \beta\}$.

4.2 Normalization and Termination

Non-determinism implies that several rewrite sequences are possible from the same $\mu \in \mathtt{DST}^{\mathtt{F}}(mA)$. In the setting of ARS, the notion of reaching a result from a term c comes in two flavours (see Sec. 2): (1.) *there exists* a rewrite sequence from c which leads to a normal form (*normalization*, WN); (2.) *each* rewrite sequence from c leads to a normal form (*termination*, SN). Below, we do a similar $\exists/\forall$ distinction . Instead of reaching a normal form or not, a sequence does so with a probability q.

▶ **Definition 4.3** (Normalization and Termination)**.** *Let $\mu \in \mathtt{DST}^{\mathtt{F}}(mA)$, $q \in [0,1]$. We write $\mu \overset{\infty}{\Rrightarrow}_p$ if* there exists a sequence *from μ which converges with probability p.*

- *μ is p-$\mathtt{WN}^\infty$ (μ **normalizes** with probability p) if p is the greatest probability to which a sequence from μ can converge.*
- *μ is p-$\mathtt{SN}^\infty$ (μ **terminates** with probability p) if* each *sequence from μ converges with probability p. μ is **Almost Sure Terminating (AST)** if it terminates with probability* 1*.*

A PARS is p-$\mathtt{WN}^\infty$, p-$\mathtt{SN}^\infty$, AST, if each μ satisfies that property.

▶ **Example 4.4.** The system in Fig. 5 is 1-$\mathtt{WN}^\infty$, but not 1-$\mathtt{SN}^\infty$. The top rewrite sequence (in blue) converges to $1 = \lim_{n\to\infty} \sum_1^n \frac{1}{2^n}$. The bottom rewrite sequence (in red) converges to 0. In between, we have all dyadic possibilities. In contrast, the system in Fig. 4 is AST.

▶ Remark (Not only AST). Many natural examples are not limited to termination and AST, such as those in Fig. 5, in Example 1.2 and 6.3. For this reason, we go beyond AST, and moreover make a distinction between weak and strong normalization.

4.3 On Unique Normal Forms

How do different rewrite sequences from the same initial μ compare w.r.t. the result they compute? Assume $[M^1] \overset{\infty}{\Rrightarrow} \alpha$ and $[M^1] \overset{\infty}{\Rrightarrow} \beta$, it is natural to wonder how β and α relate. Normalization and termination are *quantitative yes/no* properties - we are only interested in the measure $\|\beta\|$, for β limit distribution; for example, if $\mu \overset{\infty}{\Rrightarrow} \{\mathtt{F}^1\}$ and $\mu \overset{\infty}{\Rrightarrow} \{\mathtt{T}^{1/2}, \mathtt{F}^{1/2}\}$, then μ converges with probability 1, but we make no distinction between the two -very different- results. Similarly, consider again Fig. 4. The system is AST, however the limit distributions are *not unique*: they span the continuum $\{\mathtt{T}^p, \mathtt{F}^{1-p}\}$, for $p \in [0,1]$. These observations motivate attention to finer-grained properties.

In Sec. 2 we reviewed the ARS notion of *unique normal form* (UN). Let us now examine an analogue of UN in a probabilistic setting. An intuitive candidate is the following :

$$\mathtt{ULD} : \text{if } \alpha, \beta \in \mathtt{Lim}(\mu), \text{then } \alpha = \beta$$

which was first proposed in [15], where is shown that, in the case of AST, confluence implies ULD. However, ULD is not a good analogue in general, because a PARS does not need to be AST (or $\mathtt{SN}^\infty$); it may well be that $\mu \overset{\infty}{\Rrightarrow} \alpha$ and $\mu \overset{\infty}{\Rrightarrow} \beta$, with $\|\alpha\| \neq \|\beta\|$, as in Ex. 1.2 and in Fig. 5; similar examples are natural in an untyped probabilistic λ-calculus (recall that the λ-calculus is not SN!). In the general case, ULD is not implied by confluence: the system in Fig. 5 is indeed confluent. We then would like to say that it satisfies UN.

We propose as probabilistic analogue of UN the following property

$\mathtt{UN}^\infty$: $\mathtt{Lim}(\mu)$ has a *unique maximal* element.

▶ Remark. In the case of $\mathtt{SN}^\infty$ (and AST), all limits are maximal, hence $\mathtt{UN}^\infty$ becomes ULD.

4.3.1 Confluence and $\mathtt{UN}^\infty$

We justify that $\mathtt{UN}^\infty$ is an appropriate generalization of the $\mathtt{UN}$ property, by showing that it satisfies an analogue of standard ARS results: "Confluence implies $\mathtt{UN}$" (see Thm. 4.7) and "the Normal Form Property implies $\mathtt{UN}$" (Lemma 4.6). While the statements are similar to the classical ones, the content is not. To understand why is different, and non-trivial, observe that $\mathtt{Lim}(\mu)$ is in general uncountable, hence there is not even reason to believe that $\mathtt{Lim}(\mu)$ has maximal elements, for the same reason as $[0,1)$ has no max, even if it has a sup.

▶ Remark 4.5 (Which notion of Confluence?). To guarantee $\mathtt{UN}^\infty$, it suffices a weaker form of confluence than one would expect. Assume $\sigma\ {}^*\!\!\leftleftarrows \mu \rightrightarrows^* \rho$; with the standard notion of confluence in mind, we may require that $\exists\xi$ such that $\sigma \rightrightarrows^* \xi$, $\rho \rightrightarrows^* \xi$ or that $\exists\xi,\xi'$ such that $\sigma \rightrightarrows^* \xi$, $\rho \rightrightarrows^* \xi'$ and $\mathcal{E}_{\mathtt{flat}}(\xi,\xi')$. Both are fine, but a weaker notion of equivalence suffices: $\mathbb{NF}$-Confluence (defined below), which only regards normal forms. Obviously, the two stronger notions of confluence which we just discussed, imply it.

A PARS satisfies the following properties if they hold for each $\mu \in \mathtt{DST}^{\mathtt{F}}(mA)$:

- **$\mathbb{NF}$-Confluence(Confluence in Normal Form):** $\forall\sigma,\rho$ with $\sigma\ {}^*\!\!\leftleftarrows \mu \rightrightarrows^* \rho$, $\exists\xi,\tau$ such that $\sigma \rightrightarrows^* \xi$, $\rho \rightrightarrows^* \tau$, and $\xi^{\mathsf{NF}} = \tau^{\mathsf{NF}}$.
- $\mathtt{NFP}^\infty$ **(Normal Form Property):** if α is maximal in $\mathtt{Lim}(\mu)$, and $\mu \rightrightarrows^* \sigma$ then $\sigma \overset{\infty}{\Rrightarrow} \alpha$.
- $\mathtt{LimP}$ **(Limit Distributions Property):** if $\alpha \in \mathtt{Lim}(\mu)$ and $\mu \rightrightarrows^* \sigma$, there exists $\beta \in \mathtt{Lim}(\mu)$ such that $\sigma \overset{\infty}{\Rrightarrow} \beta$ and $\alpha \le \beta$.

The following result (which is standard for ARS) is easy, and independent from confluence.

▶ **Lemma 4.6.** *For each PARS such that* $\mathtt{Lim}(\mu)$ *has maximal elements,* $\mathtt{NFP}^\infty \Rightarrow \mathtt{UN}^\infty$*.*

Proof. Let $\alpha \in \mathtt{Lim}(\mu)$ be maximal. If $\beta \in \mathtt{Lim}(\mu)$, there is a sequence $\langle\tau_n\rangle_{n\in\mathbb{N}}$ from μ such that $\beta = \sup_n\{\tau_n^{\mathsf{NF}}\}$. $\mathtt{NFP}^\infty$ implies that $\forall n$, $\tau_n \overset{\infty}{\Rrightarrow} \alpha$, and therefore $\tau_n^{\mathsf{NF}} \le \alpha$. We conclude that $\beta \le \alpha$; hence if β is maximal, $\beta = \alpha$. ◀

To prove that $\mathbb{NF}$-Confluence implies $\mathtt{UN}^\infty$ is more delicate; the proof is in Appendix. We need to prove that confluence implies existence and uniqueness of maximal elements of $\mathtt{Lim}(\mu)$.

▶ **Theorem 4.7.** *For each PARS, $\mathbb{NF}$-Confluence implies* $\mathtt{UN}^\infty$*.*

Note that the proofs refines those for the analogous ARS properties in a way similar to the generalization to infinitary rewriting, by approximation; the quantitative character of probability add specific elements which are reminiscent of calculus.

4.4 Newman's Lemma Failure, and Proof Technique for PARS

In Prop. 4.6 and 4.7, the statement has the *same flavour* as similar ones for ARS, but the *notions are not the same.* The notion of limit (and therefore that of $\mathtt{UN}^\infty$, $\mathtt{SN}^\infty$, and $\mathtt{WN}^\infty$) does not belong to ARS. For this reason, the rewrite system $(mA, \rightrightarrows)$ which we are studying is not simply an ARS, and one should not assume that standard ARS properties hold. An illustration of this is **Newman's Lemma**. Given a PARS, let us assume $\mathtt{AST}$ and observe that in this case, confluence *at the limit* can be identified with $\mathtt{UN}^\infty$. *A wrong attempt:* $\mathtt{AST}$ + $\mathtt{WCR}^\infty \Rightarrow \mathtt{UN}^\infty$, where $\mathtt{WCR}^\infty$: if $\mu \rightrightarrows \sigma_1$ and $\mu \rightrightarrows \sigma_2$, then $\exists\rho$, with $\sigma_1 \overset{\infty}{\Rrightarrow} \rho$, $\sigma_2 \overset{\infty}{\Rrightarrow} \rho$. This does not hold. A counterexample is the PARS in Fig. 4, which does satisfy $\mathtt{WCR}^\infty$. (More in the Appendix.)

What is at play here is that the notion of *termination* is not the same in ARS and in PARS. A fundamental fact of ARS (on which all proofs of Newman's Lemma rely) is: termination implies that the rewriting relation is well-founded. All terminating ARS allow

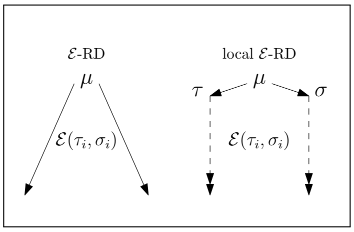

Figure 8 Random Descent.

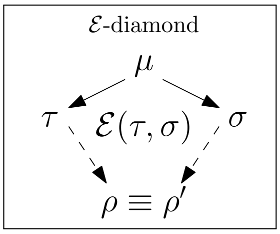

Figure 9 Diamond.

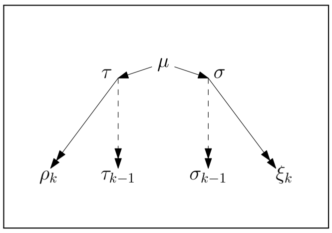

Figure 10 Proof of 5.4.

well-founded induction as proof technique; this is *not the case* for probabilistic termination. To transfer properties from ARS to PARS there are two issues: we need to find the *right formulation* and the *right proof technique*.

Our counter-example still leaves open the question "Are there *local properties* which guarantee $\mathtt{UN}^\infty$?" In the rest of the paper, we develop proof techniques to study $\mathtt{UN}^\infty$, $\mathtt{WN}^\infty$, $\mathtt{SN}^\infty$ and their relations. We will always aim at *local conditions*.

5 Random Descent (RD)

In this section we introduce $\mathcal{E}$ Random Descent ($\mathcal{E}$-RD), a tool which is able to guarantee some remarkable properties : $\mathtt{UN}^\infty$, *p-termination* as soon as *there exists a sequence* which converges to p, and also the fact that all rewrite sequences from a term have the same *expected* number of steps. $\mathcal{E}$-RD generalizes to PARS the notion of Random Descent: after any k steps, non-determinism is *irrelevant up to a chosen equivalence* $\mathcal{E}$. Indeed $\mathcal{E}$-RD is defined parametrically over an equivalence relation $\mathcal{E}$ on $\mathtt{DST}^\mathtt{F}(mA)$. For concreteness, assume $\mathcal{E}$ to be either $\mathcal{E}_{\mathsf{NF}}$ or $\mathcal{E}_{\|\|_\mathsf{N}}$ (see Def. 3.7). Then $\mathcal{E}$-RD implies that all rewrite sequences from μ:

- have the same probability of reaching a normal form after k steps (for each $k \in \mathbb{N}$);
- converge to the same limit;
- have the same expected number of steps.

Main technical result is a *local characterization* of the property (Thm 5.4), similarly to [40].

▶ **Definition 5.1** ($\mathcal{E}$ Random Descent)**.** *Let $\mathcal{E}$ be an equivalence relation on $\mathtt{DST}^\mathtt{F}(mA)$. The PARS $\mathcal{A}$ satisfies the following properties (in Fig. 8) if they hold for each $\mu \in \mathtt{DST}^\mathtt{F}(mA)$.*

- *$\mathcal{E}$-RD: for each pair of sequences $\langle\sigma_n\rangle_{n\in\mathbb{N}}$, $\langle\tau_n\rangle_{n\in\mathbb{N}}$ from μ, $\mathcal{E}(\tau_k, \sigma_k)$ holds, $\forall k$.*
- ***local** $\mathcal{E}$-RD ($\mathcal{E}$-LRD): if $\tau \leftleftarrows \mu \rightrightarrows \sigma$, then for each k there exist σ_k, τ_k with $\sigma \rightrightarrows^k \sigma_k$, $\tau \rightrightarrows^k \tau_k$, and $\mathcal{E}(\sigma_k, \tau_k)$.*

▶ **Example 5.2.** In Fig. 4 $\mathcal{E}$-RD holds for $\mathcal{E} = \mathcal{E}_{\|\|_\mathsf{N}}$, but not for $\mathcal{E} = \mathcal{E}_{\mathsf{NF}}$.

When $\mathcal{E} \in \{\mathcal{E}_{\|\|_\mathsf{N}}, \mathcal{E}_{\mathsf{NF}}\}$, it is easy to check that $\mathcal{E}$-RD guarantees the following.

▶ **Proposition 5.3.**

1. *$\mathcal{E}_{\|\|_N}$-RD implies **Uniformity**: p-$\mathtt{WN}^\infty \Rightarrow p$-$\mathtt{SN}^\infty$.*
2. *$\mathcal{E}_{NF}$-RD implies Uniformity and $\mathtt{UN}^\infty$.*

Proof. Uniformity is immediate; $\mathtt{UN}^\infty$ follows from Prop. 4.7. ◀

While expressive, $\mathcal{E}$-RD is of little practical use, as it is a property which is *universally quantified* on the sequences from μ. The property $\mathcal{E}$-LRD is instead *local*. Somehow surprisingly, *the local property characterizes* $\mathcal{E}$-RD.

▶ **Theorem 5.4** (Characterization). *The following properties are equivalent:*

1. *$\mathcal{E}$-LRD;*
2. *$\forall k, \mu, \xi, \rho$ if $\mu \rightrightarrows^k \xi$ and $\mu \rightrightarrows^k \rho$, then $\mathcal{E}(\xi, \rho)$;*
3. *$\mathcal{E}$-RD.*

Proof. ($1 \Rightarrow 2$). See Fig. 10. We prove that (2) holds by induction on k. If $k = 0$, the claim is trivial. If $k > 0$, let σ be the first step from μ to ξ and τ the first step from μ to ρ. By $\mathcal{E}$-LRD, there exists σ_k such that $\sigma \rightrightarrows^{k-1} \sigma_k$ and τ_k such that $\tau \rightrightarrows^{k-1} \tau_k$, with $\mathcal{E}(\sigma_k, \tau_k)$. Since $\sigma \rightrightarrows^{k-1} \xi$, we can apply the inductive hypothesis, and conclude that $\mathcal{E}(\sigma_k, \xi)$. By using the induction hypothesis on τ, we have that $\mathcal{E}(\tau_k, \rho)$ and conclude that $\mathcal{E}(\rho, \xi)$. ($2 \Rightarrow 3$). Immediate. ($3 \Rightarrow 1$). Assume $\tau \leftleftarrows \mu \rightrightarrows \sigma$. Take a sequence $\langle \tau_n \rangle_{n \in \mathbb{N}}$ from τ and a sequence $\langle \sigma_n \rangle_{n \in \mathbb{N}}$ from σ. By (3), $\mathcal{E}(\tau_k, \sigma_k)$ $\forall k$. ◀

A diamond. Let $\mathcal{E} \in \{\mathcal{E}_{\mathsf{NF}}, \mathcal{E}_{\|\|_{\mathsf{N}}}\}$. A useful case of $\mathcal{E}$-LRD is the ***$\mathcal{E}$*-diamond** property (Fig. 9): $\forall \mu, \sigma, \tau$, if $\tau \leftleftarrows \mu \rightrightarrows \sigma$, then $\mathcal{E}(\sigma, \tau)$, and $\exists \rho, \rho'$ s.t. $\big(\tau \rightrightarrows \rho, \sigma \rightrightarrows \rho'$ and $\mathcal{E}_{\mathtt{flat}}(\rho, \rho')\big)$.

▶ **Proposition 5.5.** *$\mathcal{E}$-diamond $\Rightarrow$ $\mathcal{E}$-LRD.*

Observe that while $\mathcal{E}$-LRD *characterizes* $\mathcal{E}$-RD, $\mathcal{E}$-diamond is only a *sufficient condition*.

5.1 Expected Termination Time

Random Descent captures the property (`Length`) "*all maximal rewrite sequences from a term have the same length.*" By looking at ARS as a special case of PARS (with $a \to [b^1]$ for $a \to b$), $\mathcal{E}_{\|\|_{\mathsf{N}}}$-RD does trivialize to RD. More interesting is that $\mathcal{E}_{\|\|_{\mathsf{N}}}$-RD also implies a property similar to (`Length`) for PARS, where we consider not the number of steps of the rewrite sequences, but its probabilistic analogue, the *expected number of steps*.

In an ARS, if a maximal rewrite sequence terminates, the number of steps is finite; we interpret this number as *time to termination*. In the case of PARS, a system may have infinite runs even if it is `AST`; the number of rewrite steps $\to$ from an initial state is (in general) infinite. However, what interests us is its *expected value*, *i.e.* the weighted average w.r.t. probability (see Sec. 2) which we write $\mathtt{MeanTime}(\langle \mu_n \rangle_{n \in \mathbb{N}})$. This expected value can be finite; in this case, not only the PARS is `AST`, but is said `PAST` *(Positively `AST`)* (see [4]).

▶ **Example 5.6.** In Example 2.5, the sequence from $[a^1]$ has `MeanTime` 2 (see Appendix).

[3] makes a nice observation: the mean number of steps of a rewrite sequence $\langle \mu_n \rangle_{n \in \mathbb{N}}$ admits a very simple formulation, as follows: $\mathtt{MeanTime}(\langle \mu_n \rangle_{n \in \mathbb{N}}) = 1 + \sum_{i \geq 1} (1 - \|\mu_i^{\mathsf{NF}}\|)$. Intuitively, each tick in time (*i.e.* each $\rightrightarrows$ step) is weighted with its probability to take place, which is $\mu_i^{\mathtt{flat}}\{c \mid c \notin \mathsf{NF}_{\mathcal{A}}\} = 1 - \|\mu_i^{\mathsf{NF}}\|$. Using this formulation, the following result is immediate.

▶ **Corollary 5.7.** *Let $\mu \in \mathtt{DST}^{\mathsf{F}}(mA)$. $\mathcal{E}_{\|\|_N}$-RD implies that all maximal rewrite sequences from μ have the* same `MeanTime`*.*

Observe that $\sum_{i \geq 1} (1 - \|\mu_i^{\mathsf{NF}}\|) < \infty$ implies $\lim_{n \to \infty} (1 - \|\mu_n^{\mathsf{NF}}\|) = 0$, hence $\lim_{n \to \infty} \|\mu_n^{\mathsf{NF}}\| = 1$. Therefore, Cor. 5.7 means that if a sequence from μ with *finite* `MeanTime` exists, μ is `PAST`.

6 Analysis of a probabilistic calculus: *weak* CbV λ-calculus

We introduce $\Lambda_\oplus^{\mathtt{weak}}$, a probabilistic analogue of call-by-value λ-calculus (see Sec. 1.1). Evaluation is non-deterministic, because in the case of an application there is no fixed order in the evaluation of the left and right subterms (see Example 6.1). We show that $\Lambda_\oplus^{\mathtt{weak}}$ satisfies

$\mathcal{E}_{NF}$-RD. Therefore it has remarkable properties (Cor. 6.5), analogous to those of its classical counter-part: the choice of the redex is irrelevant with respect to the *final result*, to its *approximants*, and to the *expected number of steps*.

Syntax. Terms (M, N, P, Q) and values (V, W) are defined as follows:

$$M \quad ::= \quad x \mid \lambda x.M \mid MM \mid M \oplus M \qquad\qquad V \quad ::= \quad x \mid \lambda x.M$$

Free variables are defined as usual. A term M is closed if it has no free variable. The substitution of N for the free occurrences of x in M is denoted $M[x := N]$.

Reductions. Weak call-by-value reduction $\to$ is given as a PARS, and inductively defined by the rules below; its lifting $\rightrightarrows$ is as in Def. 3.4.

$$\begin{array}{l|cc} (\lambda x.M)V \to \{M[x := V]^1\} & \dfrac{N \to \{N_i^{p_i} \mid i \in I\}}{MN \to \{MN_i^{p_i} \mid i \in I\}} & \dfrac{M \to \{M_i^{p_i} \mid i \in I\}}{MN \to \{M_iN^{p_i} \mid i \in I\}} \\ P \oplus Q \to \{P^{1/2}, Q^{1/2}\} & & \end{array}$$

▶ **Example 6.1** (Non-deterministic evaluation)**.** A term may have several redexes. The two reductions here join in one step: $[P[x := Q](A \oplus B)^1] \leftleftarrows [((\lambda x.P)Q)(A \oplus B)^1] \rightrightarrows [(\lambda x.P)QA^{1/2}, (\lambda x.P)QB^{1/2}]$.

▶ **Example 6.2** (Infinitary reduction)**.** Let $R = (\lambda x.xx \oplus \mathtt{T})(\lambda x.xx \oplus \mathtt{T})$. We have $[R^1] \overset{\infty}{\Rrightarrow} \{\mathtt{T}^1\}$. This term models the behaviour we discussed in Fig.1.

▶ **Example 6.3.** The term PR in Example 1.2 has the following reduction. $[PR^1] \rightrightarrows [P(\mathtt{T} \oplus \mathtt{F})^{1/2}, P(\Delta\Delta)^{1/2}] \rightrightarrows [P(\mathtt{T})^{1/4}, P(\mathtt{F})^{1/4}, P(\Delta\Delta)^{1/2}] \rightrightarrows^* [(\mathtt{T\ XOR\ T})^{1/4}, (\mathtt{F\ XOR\ F})^{1/4}, \Delta\Delta^{1/2}] \rightrightarrows [\mathtt{F}^{1/4}, \mathtt{F}^{1/4}, \Delta\Delta^{1/2}] \dots$ We conclude that $PR \overset{\infty}{\Rrightarrow} \{\mathtt{F}^{1/2}\}$.

▶ **Theorem 6.4.** $\Lambda_{\oplus}^{weak}$ *satisfies* $\mathcal{E}$*-RD, with* $\mathcal{E} = \mathcal{E}_{NF}$

Proof. We prove the $\mathcal{E}_{NF}$-Diamond property, using the definition of lifting and induction on the structure of the terms (see Appendix). ◀

Therefore, by Sec. 5 (and the fact that $\mathcal{E}_{NF} \Rightarrow \mathcal{E}_{\|\|_N}$), each μ satisfies the following properties:

▶ **Corollary 6.5.**

- All *rewrite sequences from* μ *converge to the same limit distribution.*
- All *rewrite sequences from* μ *have the same* expected *termination time* `MeanTime`*.*
- *If* $\mu \rightrightarrows^k \sigma$ *and* $\mu \rightrightarrows^k \tau$*, then* $\sigma^{NF} = \tau^{NF}$*,* $\forall \sigma, \tau, k$*.*

More diamonds. Other instances of ARS which satisfy Random Descent are surface reduction in Simpson's linear λ-calculus [38], and Lafont's interaction nets [28]. We do expect that their extension with a probabilistic choice satisfy the same properties as $\Lambda_{\oplus}^{\mathtt{weak}}$.

7 Comparing Strategies

In this section we provide a method to compare strategies, and a criterion to establish that a strategy is normalizing or perpetual (Cor. 7.4). *When strategy* $\mathcal{S}$ *is* better *than strategy* $\mathcal{T}$*?* To study this question we introduce a notion (parametric w.r.t. a relation $\mathcal{R}$) which generalizes the ARS notion of "*better*" introduced in [40]. Like in the case of $\mathcal{E}$-RD, we

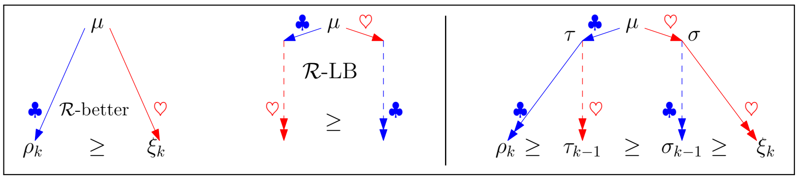

Figure 11 $\mathcal{R}$-better.

will provide a local characterization (Th. 7.3). We obtain criteria which concern both normalization/perpetuity of strategies, and the *expected number of steps* of rewrite sequences.

Given $\mathcal{A} = (A, \to)$, a **rewrite strategy** for $\to$ is a relation $\to_{\mathcal{S}} \subseteq \to$ such that $\mathtt{NF}_{(A,\to_{\mathcal{S}})} = \mathtt{NF}_{\mathcal{A}}$. Let $\rightrightarrows$ (resp. $\rightrightarrows_{\mathcal{S}}$) be the lifting of $\to$ (resp. $\to_{\mathcal{S}}$); we call $\rightrightarrows_{\mathcal{S}}$ a rewrite strategy for $\rightrightarrows$. We indicate by colored arrows $\rightrightarrows_{\clubsuit}$ and $\rightrightarrows_{\heartsuit}$ strategies for $\rightrightarrows$.

▶ **Definition 7.1.** *Given μ, let $p_{\max}(\mu)$ and $p_{\min}(\mu)$ be respectively the* greatest *and* least *value in $\{p \mid \mu \overset{\infty}{\rightrightarrows}_p\}$. A strategy $\rightrightarrows_{\clubsuit}$ is* ***normalizing*** *if for each μ,* each *$\rightrightarrows_{\clubsuit}$-sequence starting from μ converges with probability $p_{\max}(\mu)$. A strategy $\rightrightarrows_{\heartsuit}$ is* ***perpetual*** *if for each μ, each $\rightrightarrows_{\heartsuit}$ sequence from μ converges with probability $p_{\min}(\mu)$.*

Let $\mathcal{R}$ be a *reflexive and transitive relation* on $\mathtt{DST}^{\mathtt{F}}(mA)$. For concreteness, assume $\mathcal{R}$ to be either $\geq_{\|\|_{\mathsf{N}}}$ or $\geq_{\mathsf{NF}}$ (see Def. 3.7).

▶ **Definition 7.2** ($\mathcal{R}$-better)**.** *Let $\mathcal{R}$ be a relation as stipulated above. We define the following properties, which are illustrated in Fig. 11.*

- *$\rightrightarrows_{\clubsuit}$* ***is $\mathcal{R}$-better*** *than $\rightrightarrows_{\heartsuit}$ ($\mathcal{R}$-better($\rightrightarrows_{\clubsuit}, \rightrightarrows_{\heartsuit}$)): for each μ and for each pair of a $\rightrightarrows_{\clubsuit}$-sequence $\langle\rho_n\rangle_{n\in\mathbb{N}}$ and a $\rightrightarrows_{\heartsuit}$-sequence $\langle\xi_n\rangle_{n\in\mathbb{N}}$ from μ, $\mathcal{R}(\rho_k, \xi_k)$ holds ($\forall k$).*
- *$\rightrightarrows_{\clubsuit}$* ***is locally $\mathcal{R}$-better*** *than $\rightrightarrows_{\heartsuit}$ (written $\mathcal{R}$-LB($\rightrightarrows_{\clubsuit}, \rightrightarrows_{\heartsuit}$)): if $\tau\ {}_{\clubsuit}\!\leftleftarrows \mu \rightrightarrows_{\heartsuit} \sigma$, then for each $k \geq 0$, $\exists \sigma_k, \tau_k$, such that $\sigma \rightrightarrows_{\clubsuit}^{k} \sigma_k$, $\tau \rightrightarrows_{\heartsuit}{}^{k}\ \tau_k$, and $\mathcal{R}(\tau_k, \sigma_k)$*

By taking $\mathcal{R}$ to be $\geq_{\|\|_{\mathsf{N}}}$, it is immediate that $\mathcal{R}$-better($\rightrightarrows_{\clubsuit}, \rightrightarrows$) implies that $\rightrightarrows_{\clubsuit}$ is normalizing. We prove that $\mathcal{R}$-LB is sufficient (and under conditions even necessary) to establish $\mathcal{R}$-better.

▶ **Theorem 7.3.** *Let $\mathcal{R}$ be transitive and reflexive. $\mathcal{R}$-LB($\rightrightarrows_{\clubsuit}, \rightrightarrows_{\heartsuit}$) implies that $\mathcal{R}$-better($\rightrightarrows_{\clubsuit}, \rightrightarrows_{\heartsuit}$). The reverse holds if either $\rightrightarrows_{\clubsuit}$ or $\rightrightarrows_{\heartsuit}$ is $\rightrightarrows$.*

Proof. The proof is illustrated in Fig. 11. The details are in Appendix. ◀

As a consequence, we obtain a method to prove that a strategy is normalizing or perpetual by means of a local condition.

▶ **Corollary 7.4** (Normalizing criterion)**.** *Let $\mathcal{R}(\tau, \sigma)$ be $\geq_{\|\|_N}(\tau, \sigma)$ it holds that:*

1. *$\mathcal{R}$-LB($\rightrightarrows_{\clubsuit}, \rightrightarrows$) implies that $\rightrightarrows_{\clubsuit}$ is* normalizing.
2. *$\mathcal{R}$-LB($\rightrightarrows, \rightrightarrows_{\heartsuit}$) implies that $\rightrightarrows_{\heartsuit}$ is* perpetual.

It is easy to check that if $\mathcal{R}$-better($\rightrightarrows_{\clubsuit}, \rightrightarrows$), with $\mathcal{R}$ as above, and $\mathfrak{s}$ is a $\rightrightarrows_{\clubsuit}$-sequence, then $\mathtt{MeanTime}(\mathfrak{s}) \leq \mathtt{MeanTime}(\mathfrak{t})$, for each $\mathfrak{t}$ $\rightrightarrows$-sequence. Therefore, with a similar argument as in Sec.5.1, $\mathcal{R}$-better provides a criterion to establish not only that a strategy is normalizing (resp. perpetual), but also minimality (resp. maximality) of the *expected termination time*.

8 Conclusion and Further Work

We have investigated two properties which are computationally important when studying a calculus whose evaluation is both probabilistic and non-deterministic: *uniqueness of the result* and existence of a *normalizing strategy.* We have defined a probabilistic analogue $\mathtt{UN}^{\infty}$ of the notion of unique normal form, we have studied conditions which guarantee $\mathtt{UN}^{\infty}$, and relations with normalization ($\mathtt{WN}^{\infty}$) and termination ($\mathtt{SN}^{\infty}$), and between these notions. We have introduced $\mathcal{E}$-RD and $\mathcal{R}$-better as tools to analyze and compare PARS strategies. $\mathcal{E}$-RD is an alternative to strict determinism, analogous to Random Descent for ARS (non-determinism is irrelevant w.r.t. a chosen event of interest). The notion of $\mathcal{R}$-better provides a sufficient criterion to establish that a strategy is *normalizing* (resp. *perpetual*) *i.e.* the strategy is guaranteed to lead to a result with maximal (resp. minimal) probability. We have illustrated the method by studying a probabilistic extension of weak call-by-value λ-calculus; it has analogous properties to its classical counterpart: all rewrite sequences converge to the *same result*, in the same *expected number of steps.*

One-Step Reduction and Expectations. In this paper, we focus on *normal forms* and properties related to the event $\mathbb{NF}_{\mathcal{A}}$. However, we believe that the methods would allow us to compare strategies w.r.t. other properties and random variables of the system. The formalism seems especially well-suited to express the *expected value* of random variables. A key feature of the binary relation $\rightrightarrows$ is to exactly capture the ARS notion of *one-step reduction* (in contrast to *one or no step*), with a gain which is two-folded.

1. *Probability Theory.* Because all terms in the distribution are forced to reduce at the same pace, a rewrite sequence faithfully represents the evolution in time of the system (*i.e.* if $\mu \rightrightarrows^i \mu_i$, then μ_i captures the state at time i of all possible paths $a_0 \to \ldots \to a_i$). This makes the formalism well suited to express the expected value of stochastic processes.
2. *Rewrite Theory.* The results in Sections 5,6,7, crucially rely on *exactly* one-step reduction.

Further work and applications. The motivation behind this work is the need for theoretical tools to support the study of operational properties in probabilistic computation. As an example of application, we mention further work [18] where for each, the Call-by-Value, Call-by-Name, and a linear λ-calculus, a fully fledged probabilistic extension is developed. In each calculus, once establish that given a term, there exists a unique maximal result (the greatest limit distribution), [18] studies the question *"is there a strategy which is guaranteed to reach the unique result (*asymptotic standardization*)?".* Key elements in [18] rely on the abstract tools developed here; in particular, Sec. 5 and 6 allow us to demonstrate, for both the CbV and CbN *probabilistic* λ-calculi, that the leftmost-outermost strategy reaches the best possible limit distribution. This is remarkable for two reasons. First -as we already observed- the leftmost strategy is the deterministic strategy which is typically adopted in the literature of probabilistic λ-calculus, in either its CbV ([27, 9]) or its CbN version ([13, 16]), but without any completeness result with respect to *probabilistic* computation. [18] offers an “a posteriori” justification for its use. Second, the result is non-trivial, because in the probabilistic case, a standardization result for finite sequences using the leftmost strategy fails for both CbV and CbN. The tools in Sec. 5 allow for an elegant solution.

[40] makes a convincing case of the power of the RD methods for ARS, by using a large range of examples from the literature, to elegantly and uniformly revisit normalization results of various λ-calculi. We cannot here, because the rich development of strategies for λ-calculus has not yet an analogue in the probabilistic case. Nevertheless, we hope that the availability of tools to analyze PARS strategies will contribute to their development.

References

1 Gul A. Agha, José Meseguer, and Koushik Sen. PMaude: Rewrite-based specification language for probabilistic object systems. *Electr. Notes Theor. Comput. Sci.*, 153(2):213–239, 2006.

2 Sheshansh Agrawal, Krishnendu Chatterjee, and Petr Novotný. Lexicographic ranking supermartingales: an efficient approach to termination of probabilistic programs. *PACMPL*, 2(POPL):34:1–34:32, 2018.

3 Martin Avanzini, U. Dal Lago, and Akihisa Yamada. On Probabilistic Term Rewriting. In *Symposium on Functional and Logic Programming, FLOP*, pages 132–148, 2018.

4 Olivier Bournez and Florent Garnier. Proving Positive Almost Sure Termination Under Strategies. In *Rewriting Techniques and Applications, RTA*, pages 357–371, 2006.

5 Olivier Bournez and Claude Kirchner. Probabilistic Rewrite Strategies. Applications to ELAN. In *Rewriting Techniques and Applications, RTA*, pages 252–266, 2002.

6 N. Cagman and J.R. Hindley. Combinatory weak reduction in lambda calculus. *Theor. Comput. Sci.*, 1998.

7 Ugo Dal Lago, Claudia Faggian, Benoît Valiron, and Akira Yoshimizu. The geometry of parallelism: classical, probabilistic, and quantum effects. In *POPL*, pages 833–845, 2017.

8 Ugo Dal Lago and Simone Martini. The weak lambda calculus as a reasonable machine. *Theor. Comput. Sci.*, 398(1-3):32–50, 2008.

9 Ugo Dal Lago, Andrea Masini, and Margherita Zorzi. Confluence Results for a Quantum Lambda Calculus with Measurements. *Electr. Notes Theor. Comput. Sci.*, 270(2):251–261, 2011.

10 Ugo Dal Lago and Margherita Zorzi. Probabilistic operational semantics for the lambda calculus. *RAIRO - Theor. Inf. and Applic.*, 46(3):413–450, 2012.

11 Ugo de'Liguoro and Adolfo Piperno. Non Deterministic Extensions of Untyped Lambda-Calculus. *Inf. Comput.*, 122(2):149–177, 1995.

12 Nachum Dershowitz, Stéphane Kaplan, and David A. Plaisted. Rewrite, Rewrite, Rewrite, Rewrite, Rewrite, . . *Theor. Comput. Sci.*, 83(1):71–96, 1991.

13 Alessandra Di Pierro, Chris Hankin, and Herbert Wiklicky. Probabilistic lambda-calculus and Quantitative Program Analysis. *J. Log. Comput.*, 15(2):159–179, 2005.

14 Alejandro Díaz-Caro, Pablo Arrighi, Manuel Gadella, and Jonathan Grattage. Measurements and Confluence in Quantum Lambda Calculi With Explicit Qubits. *Electr. Notes Theor. Comput. Sci.*, 270(1):59–74, 2011.

15 Alejandro Díaz-Caro and Guido Martinez. Confluence in Probabilistic Rewriting. *Electr. Notes Theor. Comput. Sci.*, 338:115–131, 2018.

16 Thomas Ehrhard, Michele Pagani, and Christine Tasson. The Computational Meaning of Probabilistic Coherence Spaces. In *LICS*, pages 87–96, 2011.

17 Claudia Faggian. Probabilistic Rewriting: On Normalization, Termination, and Unique Normal Forms (Extended Version). Available at `http://arxiv.org/abs/1804.05578`.

18 Claudia Faggian and Simona Ronchi Della Rocca. Lambda Calculus and Probabilistic Computation. In *LICS*, 2019.

19 Luis María Ferrer Fioriti and Holger Hermanns. Probabilistic Termination: Soundness, Completeness, and Compositionality. In *POPL*, pages 489–501, 2015.

20 Hongfei Fu and Krishnendu Chatterjee. Termination of Nondeterministic Probabilistic Programs. In *Verification, Model Checking, and Abstract Interpretation VMCAI*, pages 468–490, 2019.

21 W.A. Howard. Assignment of ordinals to terms for primitive recursive functionals of finite type. In *Intuitionism and Proof Theory*, 1970.

22 Benjamin Lucien Kaminski, Joost-Pieter Katoen, Christoph Matheja, and Federico Olmedo. Weakest Precondition Reasoning for Expected Runtimes of Randomized Algorithms. *J. ACM*, 65(5):30:1–30:68, 2018.

23 Richard Kennaway. On transfinite abstract reduction systems. Tech. rep., CWI, Amsterdam, 1992.

24 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.

25 Maja H. Kirkeby and Henning Christiansen. Confluence and Convergence in Probabilistically Terminating Reduction Systems. In *Logic-Based Program Synthesis and Transformation - 27th International Symposium, LOPSTR 2017*, pages 164–179, 2017.

26 Jan Willem Klop and Roel C. de Vrijer. Infinitary Normalization. In *We Will Show Them! Essays in Honour of Dov Gabbay, Volume Two*, pages 169–192, 2005.

27 Daphne Koller, David A. McAllester, and Avi Pfeffer. Effective Bayesian Inference for Stochastic Programs. In *National Conference on Artificial Intelligence and Innovative Applications of Artificial Intelligence Conference, AAAI 97, IAAI 97*, pages 740–747, 1997.

28 Yves Lafont. Interaction Nets. In *POPL*, pages 95–108, 1990.

29 Simon Marlow. *Parallel and Concurrent Programming in Haskell.* O'Reilly Media, 2013.

30 Annabelle McIver, Carroll Morgan, Benjamin Lucien Kaminski, and Joost-Pieter Katoen. A new proof rule for almost-sure termination. *PACMPL*, 2(POPL):33:1–33:28, 2018.

31 Mark Newman. On Theories with a Combinatorial Definition of "Equivalence". *Annals of Mathematics*, 43(2):223–243, 1942.

32 Sungwoo Park, Frank Pfenning, and Sebastian Thrun. A probabilistic language based upon sampling functions. In *POPL*, pages 171–182, 2005.

33 Martin L. Puterman. *Markov Decision Processes: Discrete Stochastic Dynamic Programming.* John Wiley & Sons, Inc., New York, NY, USA, 1st edition, 1994.

34 Michael O. Rabin. Probabilistic Automata. *Information and Control*, 6(3):230–245, 1963.

35 Norman Ramsey and Avi Pfeffer. Stochastic lambda calculus and monads of probability distributions. In *POPL*, pages 154–165, 2002.

36 N. Saheb-Djahromi. Probabilistic LCF. In *Mathematical Foundations of Computer Science*, pages 442–451, 1978.

37 Eugene S. Santos. Computability by Probabilistic turing Machines. In *Transactions of the American Mathematical Society*, pages 159:165–184, 1971.

38 Alex K. Simpson. Reduction in a Linear Lambda-Calculus with Applications to Operational Semantics. In *Rewriting Techniques and Applications, RTA*, pages 219–234, 2005.

39 Terese. *Term Rewriting Systems*, volume 55 of *Cambridge Tracts in Theoretical Computer Science*. Cambridge University Press, 2003.

40 Vincent van Oostrom. Random Descent. In *Term Rewriting and Applications, RTA*, page 314–328, 2007.

41 Vincent van Oostrom and Yoshihito Toyama. Normalisation by Random Descent. In *Formal Structures for Computation and Deduction, FSCD*, pages 32:1–32:18, 2016.

A Omitted Proofs and Technical Details

A.1 Appendix to Section 3. A Formalism for Probabilistic Rewriting

The Index Set. A natural choice for the index set $\mathbb{S}$ is $\mathbb{N}$. Another natural choice for the index set $\mathbb{S}$ is A^* *i.e.* the set of finite sequences on A. This way, occurrences of $a \in A$ are labelled by their derivation path. This establishes a direct connection with the sample space of Markov Decision Processes we did mention in Sec. 2.2.

This choice implies that the **embedding** of A in $A^* \times A$ is naturally built-in in the definition of $\rightrightarrows$. Let us rewrite explicitly the definition of lifting. The key point is that the index j in (j, a) records the rewriting path of that occurrence of a. In rule (L2), since we use the rule $a \to \beta$, each $b \in \mathtt{Supp}(\beta)$ is given as index the path $j.a$; $\beta(b)$ is the probability assigned to b by β. Observe also that all occurrences are automatically distinct.

$$\frac{a \in \mathtt{NF}_{\mathcal{A}}}{\{(j,a)^1\} \rightrightarrows \{(j,a)^1\}} L1 \qquad \frac{a \to \beta}{\{(j,a)^1\} \rightrightarrows \{(j.a,b)^{\beta(b)} \mid b \in Supp(\beta)\}} L2 \qquad \frac{\left(\{(j,a)^1\} \rightrightarrows \alpha_j\right)_{j \in J}}{\{(j,a)^{p_j} \mid j \in J\} \rightrightarrows \biguplus_{j \in J} p_j \cdot \alpha_j} L3$$

A.2 Appendix to Section 4. Asymptotic Behaviour and Normal Forms

Monotone Convergence. We recall the following standard result.

▶ **Theorem A.1** (Monotone Convergence for Sums). *Let X be a countable set, $f_n : X \to [0,\infty]$ a non-decreasing sequence of functions, such that $f(x) := \lim_{n\to\infty} f_n(x) = \sup_n f_n(x)$ exists for each $x \in X$. Then*

$$\lim_{n\to\infty} \sum_{x\in X} f_n(x) \;=\; \sum_{x\in\mathcal{X}} f(x)$$

A.2.1 Confluence and $\mathtt{UN}^\infty$: Greatest Limit Distribution

We prove that $\mathtt{NF}$-Confluence implies $\mathtt{UN}^\infty$(Thm. 4.7) *i.e.* confluence implies both *existence and uniqueness of maximal elements* of $\mathtt{Lim}(\mu)$. Both are consequences of $\mathtt{LimP}$ and of the main lemma, Lemma A.2. We recall the definitions.

- $\mathtt{UN}^\infty$: $\mathtt{Lim}(\mu)$ has a *unique maximal* element.
- $\mathtt{NF}$-Confluence: $\forall \sigma, \rho$ with $\sigma \;{}^*\!\leftleftarrows \mu \rightrightarrows^* \rho$, $\exists \xi, \tau$ such that $\sigma \rightrightarrows^* \xi$, $\rho \rightrightarrows^* \tau$, and $\xi^{\mathsf{NF}} = \tau^{\mathsf{NF}}$.
- $\mathtt{LimP}$: if $\alpha \in \mathtt{Lim}(\mu)$ and $\mu \rightrightarrows^* \sigma$, there exists $\beta \in \mathtt{Lim}(\mu)$ such that $\sigma \overset{\infty}{\Rrightarrow} \beta$ and $\alpha \le \beta$.
- $\mathtt{NFP}^\infty$: if α is maximal in $\mathtt{Lim}(\mu)$, and $\mu \rightrightarrows^* \sigma$ then $\sigma \overset{\infty}{\Rrightarrow} \alpha$.

▶ **Lemma A.2** (Main Lemma). *$\mathtt{NF}$-Confluence implies property* $\mathtt{LimP}$.

Proof. Fig. 12 illustrates the proof. Let $\mu = \rho_0 \in \mathtt{DST}^{\mathsf{F}}(mA)$, and $\langle \rho_n \rangle_{n\in\mathbb{N}}$ be a sequence which converges to α. Assume $\rho_0 \rightrightarrows^* \sigma$. As illustrated in Fig. 12, starting from σ, we build a sequence $\sigma = \sigma_{\rho_0} \rightrightarrows^* \sigma_{\rho_1} \rightrightarrows^* \sigma_{\rho_2} \dots$, where σ_{ρ_i}, $i \ge 1$ is given by $\mathtt{NF}$-confluence : from $\rho_0 \rightrightarrows^* \sigma_{\rho_{i-1}}$ and $\rho_0 \rightrightarrows^* \rho_i$ we obtain $\sigma_{\rho_{i-1}} \rightrightarrows^* \sigma_{\rho_i}$ and $\rho_i \rightrightarrows^* \tau_i$ with $(\sigma_{\rho_i})^{\mathsf{NF}} = (\tau_i)^{\mathsf{NF}}$. Let β be the limit of the sequence so obtained; observe that $\beta \in \mathtt{Lim}(\rho_0)$. By construction, $\rho_i^{\mathsf{NF}} \le \tau_i^{\mathsf{NF}} = \sigma_{\rho_i}^{\mathsf{NF}}$; hence $\forall i$, it holds $\rho_i^{\mathsf{NF}} \le \sigma_{\rho_i}^{\mathsf{NF}} \le \beta$. From $\alpha = \sup \langle \rho_n^{\mathsf{NF}} \rangle_{n\in\mathbb{N}}$ it follows $\alpha \le \beta$. ◀

We already established (Lemma 4.6) that $\mathtt{NFP}^\infty$ implies uniqueness of maximal elements (if they exist).

▶ **Corollary A.3** (Uniqueness). $\mathtt{LimP}$ *implies* $\mathtt{NFP}^\infty$.

Proof. Immediate. If α is maximal, then $\beta = \alpha$. ◀

▶ **Lemma A.4** (Existence). $\mathtt{LimP}$ *implies that:*
1. *$\mathtt{Norms}(\mu) = \{\|\beta\| \mid \beta \in \mathtt{Lim}(\mu)\}$ has a greatest element;*
2. *$\mathtt{Lim}(\mu)$ has maximal elements.*

Proof. (1.) Let $p = \sup \mathtt{Norms}(\mu)$. We show that $p \in \mathtt{Norms}(\mu)$, by building a rewrite sequence $\langle \mu_n \rangle_{n\in\mathbb{N}}$ from μ such that $\langle \mu_n \rangle_{n\in\mathbb{N}} \overset{\infty}{\Rrightarrow} \beta$ and $\|\beta\| = p$.

The following facts are all easy to check:

a. If $\alpha < \beta$ then $\|\alpha\| < \|\beta\|$.
b. If $p \notin \mathtt{Norms}(\mu)$, then for each ϵ, there exists $\alpha \in \mathtt{Lim}(\mu)$ such that $\|\alpha\| \ge p - \epsilon$.

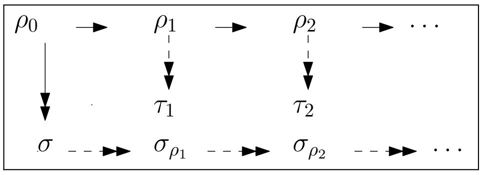

Figure 12 NF-Confluence implies LimP.

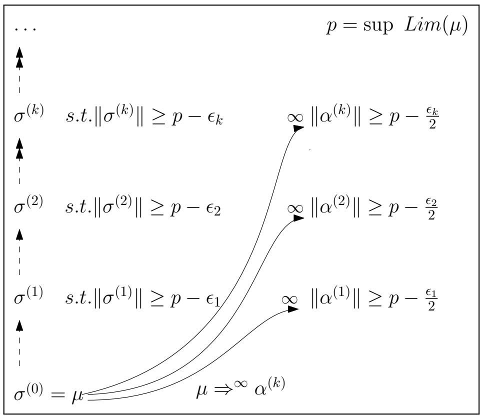

Figure 13 A sequence whose limit distribution is a maximal element of $\texttt{Lim}(\mu)$.

c. $\texttt{LimP}$ implies that, fixed ϵ, if $\mu \overset{\infty}{\Rrightarrow} \alpha$ with $\|\alpha\| \geq (p-\epsilon)$, and $\mu \rightrightarrows^* \sigma$, then there exists σ_{m_ϵ}, such that $\sigma \rightrightarrows^* \sigma_{m_\epsilon}$ and $\|\sigma_{m_\epsilon}\| \geq (p - 2\epsilon)$.
(**Proof**: $\texttt{LimP}$ implies that there is a rewrite sequence $\langle\sigma_n\rangle_{n\in\mathbb{N}}$ from σ which converges to $\gamma \geq \alpha$. Therefore $\langle\sigma_n\rangle_{n\in\mathbb{N}} \overset{\infty}{\Rrightarrow} \gamma$ where $\|\gamma\| \geq (p-\epsilon)$ and $\lim_{n\to\infty}\langle\|\sigma_n\|\rangle = \|\gamma\|$ (by Fact 4.1, point 1.). By definition of limit of a sequence, fixed ϵ, there is an index m_ϵ such that if $m \geq m_\epsilon$ then $\|\sigma_m\| \geq (\|\gamma\| - \epsilon)$, hence $\|\sigma_{m_\epsilon}\| \geq p - 2\epsilon$. Observe that $\sigma \rightrightarrows^* \sigma_{m_\epsilon}$, as a finite rewrite sequence is given by the the first m_ϵ elements of $\langle\sigma_n\rangle_{n\in\mathbb{N}}$.)

d. $\forall\delta \in \mathbb{R}^+$ there exists k such that $\frac{p}{2^k} \leq \delta$.

For each $k \in \mathbb{N}$, let $\epsilon_k = \frac{p}{2^k}$. Let $\sigma^{(0)} = \mu$. From here, we build a sequence of reductions $\mu \rightrightarrows^* \sigma^{(1)} \rightrightarrows^* \sigma^{(2)} \rightrightarrows^* \ldots$ whose limit has norm p, as illustrated in Fig. 13.

For each $k > 0$, we observe that:

- By (b.) there exists $\alpha^{(k)} \in \texttt{Lim}(\mu)$ such that $\|\alpha^{(k)}\| \geq (p - \frac{1}{2}\frac{p}{2^k})$.
- From $\mu \rightrightarrows^* \sigma^{(k-1)}$, we use (c.) to establish that there exists $\sigma^{(k)}$ such that $\sigma^{(k-1)} \rightrightarrows^* \sigma^{(k)}$ and $\|\sigma^{(k)}\| \geq (p - \frac{p}{2^k})$. Observe that $\alpha^{(k)}$, $\sigma^{(k-1)}$, $\sigma^{(k)}$ resp. instantiate α, σ, σ_{m_ϵ} of (c.).

Let $\langle\mu_n\rangle_{n\in\mathbb{N}}$ be the concatenation of all the finite sequences $\sigma^{(k-1)} \rightrightarrows^* \sigma^{(k)}$, and let β be its limit distribution. By construction, $\lim_{n\to\infty}\langle\|\mu_n\|\rangle = \|\beta\| = p$, hence $p \in \texttt{Norms}(\mu)$.

(1. ⇒ 2.) We observe that if $\langle\mu_n\rangle_{n\in\mathbb{N}} \overset{\infty}{\Rrightarrow} \alpha$ and $\|\alpha\|$ is maximal in $\texttt{Norms}(\mu)$, then α is maximal in $\texttt{Lim}(\mu)$ (because if $\gamma \in \texttt{Lim}(\mu)$ and $\gamma > \alpha$, then $\|\gamma\| > \|\alpha\|$). ◀

▶ **Theorem 4.7** (restated). *For each PARS, NF-Confluence implies* $\texttt{UN}^\infty$.

Proof. The claim follows from the Main Lemma, A.3 and A.4 (point 2.), using Lemma 4.6. ◀

Semi-Confluence. The proof of Thm. 4.7 shows we can weaken confluence even further.

▶ **Proposition A.5.** *For each PARS, the property below implies* $\texttt{LimP}$.
Semi-Confluence*: $\forall\mu, \sigma, \rho$ with $\sigma\ {}^*\!\leftleftarrows \mu \rightrightarrows^* \rho$, $\exists\sigma'$ such that $\sigma \rightrightarrows^* \sigma'$ and $\rho^{\mathsf{NF}} \leq \sigma'^{\mathsf{NF}}$.*

Proof. Proof A.2 really only uses Semi-Confluence. Let us revisit the proof, which is now illustrated in Fig. 14. Let $\rho_0 \in \texttt{DST}^{\texttt{F}}(mA)$, and α maximal in $\texttt{Lim}(\rho_0)$. Assume $\rho_0 \rightrightarrows^* \sigma$. Let $\langle\rho_n\rangle_{n\in\mathbb{N}}$ be a sequence which converges to α. As illustrated in Fig. 14, starting from σ, we build a sequence $\sigma = \sigma_{\rho_0} \rightrightarrows^* \sigma_{\rho_1} \rightrightarrows^* \sigma_{\rho_2} \ldots$, where σ_{ρ_i}, $i \geq 1$ is given by Semi-Confluence:

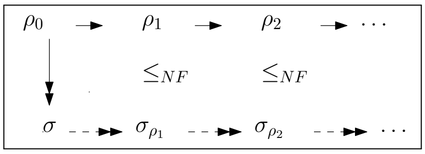

Figure 14 Semi-Confluence implies $\mathtt{LimP}$.

from $\rho_0 \rightrightarrows^* \rho_i$ and $\rho_0 \rightrightarrows^* \sigma_{\rho_{i-1}}$ we obtain $\sigma_{\rho_{i-1}} \rightrightarrows^* \sigma_{\rho_i}$ with $\rho_i^{\mathsf{NF}} \leq \sigma_{\rho_i}^{\mathsf{NF}}$. Let β be the limit of the sequence so obtained; observe that $\beta \in \mathtt{Lim}(\rho_0)$. By construction, $\forall i$, it holds $\rho_i^{\mathsf{NF}} \leq (\sigma_{\rho_i})^{\mathsf{NF}} \leq \beta$. From $\alpha = \sup \langle \rho_n^{\mathsf{NF}} \rangle_{n \in \mathbb{N}}$ it follows that $\alpha \leq \beta$. ◀

A.2.2 More on Newman's Lemma and Proof Techniques

We pointed out that to transfer properties from ARS to PARS there are two issues: find the *right formulation* and the *right proof technique*. Newman's Lemma illustrates both. Can a different formulation uncover properties similar to Newman Lamma? Another "candidate" statement we can attempt is : $\mathtt{AST}$ + WCR $\Rightarrow \mathtt{UN}^\infty$. I have no answer here. This property is indeed an interesting case study. It is not hard to show that this property holds when $\mathtt{Lim}(\mu)$ is finite, or uniformly discrete, which also means that a counterexample (if any) cannot be trivial. On the other side, if the property holds, the difficulty is which proof technique to use, since well-founded induction is not available to us.

A.3 Appendix to Section 5. Random Descent

$\mathcal{E}$-diamond implies $\mathcal{E}$-LRD. In this section, we assume $\mathcal{E} \in \{\mathcal{E}_{\mathsf{NF}}, \mathcal{E}_{\|\,\|_{\mathsf{N}}}\}$.

▶ **Lemma A.6.** *If $\mathcal{E}_{flat}(\mu, \rho,)$, there exists a rewrite sequence $\langle \mu_n \rangle_{n \in \mathbb{N}}$ and a rewrite sequence $\langle \rho_n \rangle_{n \in \mathbb{N}}$, with $\mathcal{E}_{flat}(\mu_i, \rho_i,)$*

Proof. By easy induction on n. It is enough, at each $\rightrightarrows$ step as defined in Def. 3.4, to choose the same reduction $c \to \beta$ for all $(j, \mathtt{m}_j)$ such that $\mathtt{m}_j = c$. ◀

▶ **Proposition A.7.** *$\mathcal{E}$-diamond $\Rightarrow$ $\mathcal{E}$-LRD.*

Proof. By using Lemma A.6. ◀

Point-wise formulation. In Section 6, we exploit the fact that not only $\mathcal{E}$-RD admits a local characterization, but also that the properties $\mathcal{E}$-LRD and $\mathcal{E}$-diamond can be expressed point-wise, making the condition easier to verify.

1. pointed $\mathcal{E}$-LRD: $\forall a \in A$, if $\tau \leftleftarrows [a^1] \rightrightarrows \sigma$, then $\forall k, \exists \sigma_k, \tau_k$ with $\sigma \rightrightarrows^k \sigma_k$, $\tau \rightrightarrows^k \tau_k$, and $\mathcal{E}(\sigma_k, \tau_k)$.
2. pointed $\mathcal{E}$-diamond: $\forall a \in A$, if $\tau \leftleftarrows [a^1] \rightrightarrows \sigma$, then it holds that $\mathcal{E}(\sigma, \tau)$, and $\exists \rho, \rho'$ such that $\tau \rightrightarrows \rho, \sigma \rightrightarrows \rho'$ and $\mathcal{E}_{\mathtt{flat}}(\rho, \rho',)$.

▶ **Proposition A.8** (point-wise $\mathcal{E}$-LRD)**.** *The following hold*
- *$\mathcal{E}$-LRD $\iff$ pointed $\mathcal{E}$-LRD;*
- *$\mathcal{E}$-diamond $\iff$ pointed $\mathcal{E}$-diamond.*

Proof. Immediate, by the definition of $\rightrightarrows$. Given $\mu = [a_i^{p_i} \mid i \in I]$, we establish the result for each a_i, and put all the resulting distributions together. ◀

A.3.1 Section 5.1. Finite expected time to termination

An example of PARS with finite expected time to termination is the one in Fig. 1. We can see this informally, recalling Sec. 2. Let the sample space Ω be the set of paths ending in a normal form, and let μ be the probability distribution on Ω. What is the expected value of the random variable $\texttt{length} : \Omega \to \mathbb{N}$? We have $E(\texttt{length}) = \sum_\omega \texttt{length}(\omega) \cdot \mu(\omega) = \sum_{n\in\mathbb{N}} n \cdot \mu\{\omega \mid \texttt{length}(\omega) = n\} = \sum n \cdot \frac{1}{2^n} = 2$.

It is immediate to check that in Example 2.5, the (only) rewrite sequence from $[a^1]$ has $\texttt{MeanTime}$ 2 by using the definition of mean number of steps of a rewrite sequence $\langle\mu_n\rangle_{n\in\mathbb{N}}$ as

$$\texttt{MeanTime}(\langle\mu_n\rangle_{n\in\mathbb{N}}) = 1 + \sum_{i\geq 1}(1 - \|\mu_i^{\mathsf{NF}}\|)$$

as formulated in [3] (to which we refer for the details).

A.4 Appendix to Section 6. Weak CbV λ-calculus

We prove Thm. 6.4.

▶ **Theorem 6.4** (restated). *$\Lambda_\oplus^{weak}$ satisfies the $\mathcal{E}$-diamond property, with $\mathcal{E} = \mathcal{E}_{\mathsf{NF}}$.*

Proof. We show by induction on the structure of the term M that for all pairs of one-step reductions $\tau \leftleftarrows [M^1] \rightrightarrows \sigma$, the following hold: (1.) $\sigma^{\mathsf{NF}} = \tau^{\mathsf{NF}}$ is 0 everywhere (2.) exists ρ, ρ' such that $\tau \rightrightarrows \rho$, $\sigma \rightrightarrows \rho'$ and $\mathcal{E}_{\texttt{flat}}(\rho, \rho',)$.

It is convenient to introduce the following notation. If $\mu = [M_i^{m_i} \mid i \in I]$ we define

$$\mu@Q := [(M_iQ)^{m_i} \mid i \in I] \quad Q@\mu := [(QM_i)^{m_i} \mid i \in I]$$

We write the Dirac distribution $[\mathtt{m}_i^1]$ simply as $[\mathtt{m}_i]$. We write $\rho \equiv \rho'$ for $\mathcal{E}_{\texttt{flat}}(\rho, \rho',)$.

- If $M = x$ or $M = \lambda x.P$, no reduction is possible.
- If $M = P \oplus_p Q$, only one reduction is possible.
- If $M = PQ$ is a redex, then $P = (\lambda x.N)$, Q is a value, and no other reduction is possible inside either P or Q.
- If $M = PQ$ has two different reductions, two cases are possible.
  - Assume that both P and Q reduce; PQ has the following reductions.

    $$\frac{P \to \{P_i^{p_i} \mid i \in I\}}{PQ \to \sigma = \{P_iQ^{p_i} \mid i \in I\}} \quad \text{and} \quad \frac{Q \to \{Q_j^{q_j} \mid j \in J\}}{PQ \to \tau = \{PQ_j^{q_j} \mid j \in J\}}$$

    Observe that none of the P_iQ or PQ_j is a normal form, hence (1.) holds. By the definition of reduction, the following holds

    $$\frac{Q \to \{Q_j^{q_j} \mid j \in J\}}{P_iQ \to \{P_iQ_j^{q_j} \mid j \in J\}}$$

    and therefore by Lifting we have $\biguplus_i p_i \cdot [P_iQ] \rightrightarrows \biguplus_i p_i \cdot (\biguplus_j q_j \cdot [P_iQ_j]) = \biguplus_{i,j} p_iq_j \cdot [P_iQ_j]$. Similarly we obtain $\biguplus_j q_j \cdot [PQ_j] \rightrightarrows \biguplus_{i,j} p_iq_j \cdot [P_iQ_j]$.
  - Assume that one subterm has two different redexes; let assume it is the subterm P (the case of Q is similar):

    $$P \to \sigma = \{S_i^{s_i} \mid i \in I\} \text{and} P \to \tau = \{T_j^{t_j} \mid j \in J\}$$

By inductive hypothesis, two facts hold: (1.) $\sigma^{\mathsf{NF}} = \tau^{\mathsf{NF}}$ is 0 everywhere, therefore no S_i and no T_j in the support is a normal form; (2.) exists ρ, ρ' with $\rho \equiv \rho'$ and such that $[S_i] \rightrightarrows \rho_i$ with $\sum_i s_i \cdot \rho_i = \rho$, and $[T_j] \rightrightarrows \rho_j$ with $\sum_j t_j \cdot \rho_j = \rho'$. For PQ we have

$$\frac{P \to \{S_i^{s_i} \mid i \in I\}}{PQ \to \{(S_iQ)^{s_i} \mid i \in I\}} \quad \text{and} \quad \frac{P \to \{T_j^{t_j} \mid j \in J\}}{PQ \to \{(T_jQ)^{t_j} \mid j \in J\}}$$

First of all, we observe that no S_iQ and no T_jQ is a normal form, hence property (1.) is verified. Moreover, it holds that $S_iQ \to \rho_i@Q$ and $T_jQ \to \rho_j@Q$. We conclude by Lifting that $[(S_iQ)^{s_i} \mid i \in I] \rightrightarrows \biguplus_i s_i \cdot \rho_i@Q$. It is easy to check that, $\biguplus_i s_i \cdot \rho_i@Q = \rho@Q$, and $[(T_jQ)^{t_j} \mid j \in J] \rightrightarrows \biguplus t_j \cdot \rho_j@Q = \rho@Q$. It is immediate also that $\rho@Q \equiv \rho'@Q$; hence property (2.) is also verified. ◀

A.5 Appendix to Section 7. Comparing Strategies

▶ **Theorem 7.3** (restated). *Let $\mathcal{R}$ be transitive and reflexive. $\mathcal{R}$-LB($\rightrightarrows_\clubsuit, \rightrightarrows_\heartsuit$) implies that $\mathcal{R}$-better($\rightrightarrows_\clubsuit, \rightrightarrows_\heartsuit$). The reverse holds if either $\rightrightarrows_\clubsuit$ or $\rightrightarrows_\heartsuit$ is $\rightrightarrows$.*

Proof. $\Rightarrow$. See Fig. 11. We prove by induction on k the following: "*$\mathcal{R}$-LB($\rightrightarrows_\clubsuit, \rightrightarrows_\heartsuit$) implies ($\forall \mu, \rho, \xi$, if $\mu \rightrightarrows_\clubsuit{}^k \rho$ and $\mu \rightrightarrows_\heartsuit{}^k \xi$, then $\mathcal{R}(\rho, \xi)$)*". If $k = 0$, the claim is trivial. If $k \geq 1$, let σ be the first step from μ to ξ, and τ the first step from μ to ρ, as in Fig. 11. $\mathcal{R}$-LB implies that exist σ_{k-1} and τ_{k-1} such that $\sigma \rightrightarrows_\clubsuit{}^{k-1} \sigma_{k-1}$, $\tau \rightrightarrows_\heartsuit{}^{k-1} \tau_{k-1}$, with $\mathcal{R}(\tau_{k-1}, \sigma_{k-1})$. Since $\sigma \rightrightarrows_\heartsuit{}^{k-1} \xi$ we can apply the inductive hypothesis, and obtain that $\mathcal{R}(\sigma_{k-1}, \xi)$. Again by inductive hypothesis, from $\tau \rightrightarrows_\clubsuit{}^{k-1} \rho$ we obtain $\mathcal{R}(\rho, \tau_{k-1})$. By transitivity, it holds that $\mathcal{R}(\rho, \xi)$. $\Leftarrow$. Assume $\rightrightarrows_\heartsuit = \rightrightarrows$, and $\tau \; {}_\clubsuit\!\leftleftarrows \mu \rightrightarrows \sigma$. Let $\langle \tau_n \rangle_{n \in \mathbb{N}}$ and $\langle \sigma_n \rangle_{n \in \mathbb{N}}$ be obtained by extending τ and σ with a maximal $\rightrightarrows_\clubsuit$ sequence. The claim follows from the hypothesis that $\rightrightarrows_\clubsuit$ dominates $\rightrightarrows$, by viewing the $\rightrightarrows_\clubsuit$ steps in $\langle \sigma_n \rangle_{n \in \mathbb{N}}$ as $\rightrightarrows$ steps. ◀

▶ Remark. Observe that $\mathcal{E}$-RD (resp. $\mathcal{E}$-LRD) is a special case of $\mathcal{R}$-better (resp. $\mathcal{R}$-LB). We have preferred to treat it independently, for the sake of presentation.

A.6 Comments

Finite Approximants. $\mathcal{E}$-RD characterizes the case when (not only at the limit, but also at the level of the approximants) the non-deterministic choices are irrelevant. The notion of approximant which we have studied here is "stop after a number k of steps" ($k \in \mathbb{N}$). We can consider different notion of approximants. For example, we could also wish to stop the evolution of the system when it reaches a normal form with probability p. Our method can easily be adapted to analyze this case (see [17]). We believe it is also possible to extend to the probabilistic setting the results in [41], which would allow to further push this direction.

The beauty of local. In Sec. 5 and 7, by local conditions we mean the following: to show that a property P holds globally (*i.e. for each* two rewrite sequences, P holds), we can show that P holds locally (*i.e.* for each pair of one-step reductions, *there exist* two rewrite sequences such that P holds). This reduces the space of search for testing the property, a fact that we exploit in the proofs of Sec. 6.