On Probabilistic Term Rewriting

Martin Avanzini[1] and Ugo Dal Lago[12] and Akihisa Yamada[3]

[1] INRIA Sophia Antipolis, France
[2] Department of Computer Science, University of Bologna, Italy
[3] National Institute of Informatics, Japan

Abstract. We study the termination problem for probabilistic term rewrite systems. We prove that the interpretation method is sound and complete for a strengthening of positive almost sure termination, when abstract reduction systems and term rewrite systems are considered. Two instances of the interpretation method—polynomial and matrix interpretations—are analyzed and shown to capture interesting and non-trivial examples when automated. We capture probabilistic computation in a novel way by means of multidistribution reduction sequences, thus accounting for both the nondeterminism in the choice of the redex and the probabilism intrinsic in firing each rule.

1 Introduction

Interactions between computer science and probability theory are pervasive and extremely useful to the first discipline. Probability theory indeed offers models that enable *abstraction*, but it also suggests a new model *of computation*, like in randomized computation or cryptography [17]. All this has stimulated the study of probabilistic computational models and programming languages: probabilistic variations on well-known models like automata [24], Turing machines [26], and the λ-calculus [25] are known from the early days of theoretical computer science.

The simplest way probabilistic choice can be made available in programming is endowing the language of programs with an operator modeling sampling from (one or many) distributions. Fair, binary, probabilistic choice is for example perfectly sufficient to get universality if the underlying programming language is itself universal (e.g., see [10]).

Term rewriting [27] is a well-studied model of computation when no probabilistic behavior is involved. It provides a faithful model of pure functional programming which is, up to a certain extent, also adequate for modeling higher-order parameter passing [12]. What is peculiar in term rewriting is that, in principle, rule selection turns reduction into a potentially nondeterministic process. The following question is then a natural one: is there a way to generalize term rewriting to a fully-fledged *probabilistic* model of computation? Actually, not much is known about probabilistic term rewriting: we are only aware of the definitions due to Agha et al. [1] and due to Bournez and Garnier [5]. We base our work on the latter, where probabilistic rewriting is captured as a Markov decision process; rule selection remains nondeterministic, but each rule can have one of many possible outcomes, each with its own probability. Rewriting thus becomes a process in which both nondeterministic and probabilistic aspects are present and intermingled. When firing a rule, the reduction process implicitly samples from a distribution, much in the same way as when performing binary probabilistic choice in one of the models mentioned above.

In this paper, we first define a new, simple framework for discrete probabilistic reduction systems, which properly generalizes standard abstract reduction systems [27]. In particular, what plays the role of a reduction sequence, usually a (possibly infinite) sequence $a_1 \to a_2 \to \dots$ of *states*, is a sequence $\mu_1 \rightsquigarrow \mu_2 \rightsquigarrow \dots$ of *(multi)distributions* over the set of states. A multidistribution is not merely a distribution, and this is crucial to appropriately account for both the probabilistic behaviour of each rule and the nondeterminism in rule selection. Such correspondence does not exist in Bournez and Garnier's framework, as nondeterminism has to be resolved by a *strategy*, in order to define reduction sequences. However, the two frameworks turn out to be equiexpressive, at least as far as every rule has finitely many possible outcomes. We then prove that the probabilistic ranking functions [5] are sound and complete for proving *strong almost sure termination*, a strengthening of *positive almost sure termination* [5]. We moreover show that ranking functions provide bounds on expected runtimes.

This paper's main contribution, then, is the definition of a simple framework for *probabilistic term rewrite systems* as an example of this abstract framework. Our main aim is studying whether any of the well-known techniques for termination of term rewrite systems can be generalized to the probabilistic setting, and whether they can be automated. We give positive answers to these two questions, by describing how polynomial and matrix interpretations can indeed be turned into instances of probabilistic ranking functions, thus generalizing them to the more general context of probabilistic term rewriting. We moreover implement these new techniques into the termination tool `NaTT` [28]. The implementation and an extended version of this paper [3] are available at `http://www.trs.cm.is.nagoya-u.ac.jp/NaTT/probabilistic`.

2 Related Work

Termination is a crucial property of programs, and has been widely studied in term rewriting. Tools checking and certifying termination of term rewrite systems are nowadays capable of implementing tens of different techniques, and can prove termination of a wide class of term rewrite systems, although the underlying verification problem is well known to be undecidable [27].

Termination remains an interesting and desirable property in a probabilistic setting, e.g., in probabilistic programming [18] where inference algorithms often rely on the underlying program to terminate. But what does termination *mean* when systems become probabilistic? If one wants to stick to a *qualitative* definition, almost-sure termination is a well-known answer: a probabilistic computation is said to almost surely terminate iff non-termination occurs with null probability. One could even require *positive* almost-sure termination, which asks the expected time to termination to be finite. Recursion-theoretically, checking (positive) almost-sure termination is harder than checking termination of non-probabilistic programs, where termination is at least recursively enumerable, although undecidable: in a universal probabilistic imperative programming language, almost sure termination is Π^0_2 complete, while positive almost-sure termination is Σ^0_2 complete [20].

Many sound verification methodologies for probabilistic termination have recently been introduced (see, e.g., [5,6,16,14,9]). In particular, the use of ranking martingales has turned out to be quite successful when the analyzed program is imperative, and thus does not have an intricate recursive structure. When the latter holds, techniques akin to sized types have been shown to be applicable [11]. Finally, as already mentioned, the current work can be seen as stemming from the work by Bournez et al. [7,5,6]. The added value compared to their work are first of all the notion of multidistribution as a way to give an instantaneous description of the state of the underlying system which exhibits both nondeterministic and probabilistic features. Moreover, an interpretation method inspired by ranking functions is made more general here, this way accommodating not only interpretations over the real numbers, but also interpretations over vectors, in the sense of matrix interpretations. Finally, we provide an automation of polynomial and matrix interpretation inference here, whereas nothing about implementation was presented in Bournez and Garnier's work.

3 Probabilistic Abstract Reduction Systems

An *abstract reduction system (ARS)* on a set A is a binary relation $\to \subseteq A \times A$. Having $a \to b$ means that a reduces to b in one step, or b is a one-step reduct of a. Bournez and Garnier [5] extended the ARS formalism to probabilistic computations, which we will present here using slightly different notations.

We write $\mathbb{R}_{\geq 0}$ for the set of non-negative reals. A *(probability) distribution* on a countable set A is a function $d : A \to \mathbb{R}_{\geq 0}$ such that $\sum_{a \in A} d(a) = 1$. We say a distribution d is *finite* if its *support* $\mathsf{Supp}(d) := \{a \in A \mid d(a) > 0\}$ is finite, and write $\{d(a_1) : a_1, \ldots,\ d(a_n) : a_n\}$ for d if $\mathsf{Supp}(d) = \{a_1, \ldots, a_n\}$ (with pairwise distinct a_is). We write $\mathsf{FDist}(A)$ for the set of finite distributions on A.

Definition 1 (PARS, [5]). *A* probabilistic reduction *over a set A is a pair of $a \in A$ and $d \in \mathsf{FDist}(A)$, written $a \to d$. A* probabilistic ARS (PARS) *$\mathcal{A}$ over A is a (typically infinite) set of probabilistic reductions. An object $a \in A$ is called* terminal *(or a* normal form*) in $\mathcal{A}$, if there is no d with $a \to d \in \mathcal{A}$. With $\mathsf{TRM}(\mathcal{A})$ we denote the set of terminals in $\mathcal{A}$.*

The intended meaning of $a \to d \in \mathcal{A}$ is that "there is a reduction step $a \to_{\mathcal{A}} b$ with probability $d(b)$".

Example 2 (Random walk). A random walk over $\mathbb{N}$ with *bias* probability p is modeled by the PARS $\mathcal{W}_p$ consisting of the probabilistic reduction

$$n + 1 \to \{p : n,\ 1 - p : n + 2\} \quad \text{for all } n \in \mathbb{N}.$$

A PARS describes both nondeterministic and probabilistic choice; we say a PARS $\mathcal{A}$ is *nondeterministic* if $a \to d_1, a \to d_2 \in \mathcal{A}$ with $d_1 \neq d_2$. In this case, the distribution of one-step reducts of a is nondeterministically chosen from d_1 and d_2. Bournez and Garnier [5] describe reduction sequences via stochastic sequences, which demand nondeterminism to be resolved by fixing a *strategy* (also called *policies*). In contrast, we capture nondeterminism by defining a reduction relation $\rightsquigarrow_{\mathcal{A}}$ on distributions, and emulate ARSs by $\{1 : a\} \rightsquigarrow_{\mathcal{A}} \{1 : b\}$ when $a \to \{1 : b\} \in \mathcal{A}$. For the probabilistic case, taking Example 2 we would like to have

$$\{1 : 1\} \rightsquigarrow_{\mathcal{W}_{\frac{1}{2}}} \{\tfrac{1}{2} : 0,\ \tfrac{1}{2} : 2\}\ ,$$

meaning that the distribution of one-step reducts of 1 is $\{\frac{1}{2} : 0,\ \frac{1}{2} : 2\}$. Continuing the reduction, what should the distribution of two-step reducts of 1 be? Actually, it cannot be a distribution (on A): by probability $\frac{1}{2}$ we have no two-step reduct of 1. One solution, taken by [5], is to introduce $\bot \notin A$ representing the case where no reduct exists. We take another solution: we consider *subdistributions*, i.e. generalizations of distributions where probabilities may sum up to less than one, allowing

$$\{1 : 1\} \rightsquigarrow_{\mathcal{W}_{\frac{1}{2}}} \{\tfrac{1}{2} : 0,\ \tfrac{1}{2} : 2\} \rightsquigarrow_{\mathcal{W}_{\frac{1}{2}}} \{\tfrac{1}{4} : 1,\ \tfrac{1}{4} : 3\}\ .$$

Further continuing the reduction, one would expect $\{\frac{1}{8} : 0,\ \frac{1}{4} : 2,\ \frac{1}{8} : 4\}$ as the next step, but note that a half of the probability $\frac{1}{4}$ of 2 is the probability of reduction sequence $2 \to_{\mathcal{W}_{\frac{1}{2}}} 1 \to_{\mathcal{W}_{\frac{1}{2}}} 2$, and the other half is of $2 \to_{\mathcal{W}_{\frac{1}{2}}} 3 \to_{\mathcal{W}_{\frac{1}{2}}} 2$.

Example 3. Consider the PARS $\mathcal{N}$ consisting of the following rules:

$$\mathtt{a} \to \{\tfrac{1}{2} : \mathtt{b}_1,\ \tfrac{1}{2} : \mathtt{b}_2\} \qquad \mathtt{b}_1 \to \{1 : \mathtt{c}\} \qquad \mathtt{c} \to \{1 : \mathtt{d}_1\}$$
$$\mathtt{b}_2 \to \{1 : \mathtt{c}\} \qquad \mathtt{c} \to \{1 : \mathtt{d}_2\}\ .$$

Reducing $\mathtt{a}$ twice always yields $\mathtt{c}$, so the distribution of the two-step reduct of $\mathtt{a}$ is $\{1 : \mathtt{c}\}$. More precisely, there are two paths to reach $\mathtt{c}$: $\mathtt{a} \to_{\mathcal{N}} \mathtt{b}_1 \to_{\mathcal{N}} \mathtt{c}$ and $\mathtt{a} \to_{\mathcal{N}} \mathtt{b}_2 \to_{\mathcal{N}} \mathtt{c}$. Each of them can be nondeterministically continued to $\mathtt{d}_1$ and $\mathtt{d}_2$, so the distribution of three-step reducts of $\mathtt{a}$ is the nondeterministic choice among $\{1 : \mathtt{d}_1\}$, $\{\frac{1}{2} : \mathtt{d}_1, \frac{1}{2} : \mathtt{d}_2\}$, $\{1 : \mathtt{d}_2\}$. On the other hand, whereas it is obvious that the two-step reduct $\{1 : \mathtt{c}\}$ of $\mathtt{a}$ should further reduce to $\{1 : \mathtt{d}_1\}$ or $\{1 : \mathtt{d}_2\}$, respectively, obtaining the third choice $\{\frac{1}{2} : \mathtt{d}_1, \frac{1}{2} : \mathtt{d}_2\}$ would require that the reduction relation $\rightsquigarrow_{\mathcal{N}}$ is defined in a non-local manner.

To overcome this problem, we refine distributions to multidistributions.

Definition 4 (Multidistributions). *A* multidistribution *on A is a finite multiset μ of pairs of $a \in A$ and $0 \leq p \leq 1$, written $p : a$, such that*

$$|\mu| := \sum_{p:a \in \mu} p \ \leq\ 1\ .$$

We denote the set of multidistributions on A by $\mathsf{FMDist}(A)$.

Abusing notation, we identify $\{p_1{:}a_1, \ldots,\ p_n{:}a_n\} \in \mathsf{FDist}(A)$ with multidistribution $\{\!\{p_1 : a_1, \ldots,\ p_n : a_n\}\!\}$ as no confusion can arise. For a function $f : A \to B$, we often generalize the domain and range to multidistributions as follows:

$$f\big(\{\!\{p_1 : a_1, \ldots,\ p_n : a_n\}\!\}\big) := \{\!\{p_1 : f(a_1), \ldots,\ p_n : f(a_n)\}\!\} \ .$$

The *scalar multiplication* of a multidistribution is $p \cdot \{\!\{q_1 : a_1, \ldots,\ q_n : a_n\}\!\} := \{\!\{p \cdot q_1 : a_1, \ldots,\ p \cdot q_n : a_n\}\!\}$, which is also a multidistribution if $0 \le p \le 1$. More generally, multidistributions are closed under *convex multiset unions*, defined as $\biguplus_{i=1}^{n} p_i \cdot \mu_i$ with $p_1, \ldots, p_n \ge 0$ and $p_1 + \cdots + p_n \le 1$.

Now we introduce the reduction relation $\rightsquigarrow_{\mathcal{A}}$ over multidistributions.

Definition 5 (Probabilistic Reduction). *Given a PARS $\mathcal{A}$, we define the* probabilistic reduction relation $\rightsquigarrow_{\mathcal{A}} \subseteq \mathsf{FMDist}(A) \times \mathsf{FMDist}(A)$ *as follows:*

$$\frac{a \in \mathsf{TRM}(\mathcal{A})}{\{\!\{1 : a\}\!\} \rightsquigarrow_{\mathcal{A}} \varnothing} \qquad \frac{a \to d \in \mathcal{A}}{\{\!\{1 : a\}\!\} \rightsquigarrow_{\mathcal{A}} d} \qquad \frac{\mu_1 \rightsquigarrow_{\mathcal{A}} \rho_1 \quad \cdots \quad \mu_n \rightsquigarrow_{\mathcal{A}} \rho_n}{\biguplus_{i=1}^{n} p_i \cdot \mu_i \rightsquigarrow_{\mathcal{A}} \biguplus_{i=1}^{n} p_i \cdot \rho_i}$$

In the last rule, we assume $p_1, \ldots, p_n \ge 0$ and $p_1 + \cdots + p_n \le 1$. We denote by $\mathcal{A}(\mu)$ the set of all possible reduction sequences from μ, i.e., $\{\mu_i\}_{i\in\mathbb{N}} \in \mathcal{A}(\mu)$ iff $\mu_0 = \mu$ and $\mu_i \rightsquigarrow_{\mathcal{A}} \mu_{i+1}$ for any $i \in \mathbb{N}$.

Thus $\mu \rightsquigarrow_{\mathcal{A}} \nu$ if ν is obtained from μ by replacing every nonterminal a in μ with all possible reducts with respect to some $a \to d \in \mathcal{A}$, suitably weighted by probabilities, and by removing terminals. The latter implies that $|\mu|$ is not preserved during reduction: it decreases by the probabilities of terminals.

To continue Example 2, we have the following reduction sequence:

$$\begin{aligned}\{\!\{1 : 1\}\!\} &\rightsquigarrow_{\mathcal{W}_{\frac{1}{2}}} \{\!\{\tfrac{1}{2} : 0,\ \tfrac{1}{2} : 2\}\!\} \rightsquigarrow_{\mathcal{W}_{\frac{1}{2}}} \varnothing \uplus \{\!\{\tfrac{1}{4} : 1,\ \tfrac{1}{4} : 3\}\!\} \\ &\rightsquigarrow_{\mathcal{W}_{\frac{1}{2}}} \{\!\{\tfrac{1}{8} : 0,\ \tfrac{1}{8} : 2\}\!\} \uplus \{\!\{\tfrac{1}{8} : 2,\ \tfrac{1}{8} : 4\}\!\} \rightsquigarrow_{\mathcal{W}_{\frac{1}{2}}} \cdots\end{aligned}$$

The use of multidistributions resolves the issues indicated in Example 3 when dealing with nondeterministic systems. We have, besides others, the reduction

$$\{\!\{1 : \mathsf{a}\}\!\} \rightsquigarrow_{\mathcal{N}} \{\!\{\tfrac{1}{2} : \mathsf{b}_1, \tfrac{1}{2} : \mathsf{b}_2\}\!\} \rightsquigarrow_{\mathcal{N}} \{\!\{\tfrac{1}{2} : \mathsf{c}, \tfrac{1}{2} : \mathsf{c}\}\!\} \rightsquigarrow_{\mathcal{N}} \{\!\{\tfrac{1}{2} : \mathsf{d}_1, \tfrac{1}{2} : \mathsf{d}_2\}\!\} \ .$$

The final step is possible because $\{\!\{\frac{1}{2} : \mathsf{c}, \frac{1}{2} : \mathsf{c}\}\!\}$ is not collapsed to $\{\!\{1 : \mathsf{c}\}\!\}$.

When every probabilistic reduction in $\mathcal{A}$ is of form $a \to \{1{:}b\}$ for some b, then $\rightsquigarrow_{\mathcal{A}}$ simulates the non-probabilistic ARS via the relation $\{\!\{1 : \cdot\}\!\} \rightsquigarrow_{\mathcal{A}} \{\!\{1 : \cdot\}\!\}$. Only a little care is needed as normal forms are followed by $\varnothing$.

Proposition 6. *Let $\hookrightarrow$ be an ARS and define $\mathcal{A}$ by $a \to \{1 : b\} \in \mathcal{A}$ iff $a \hookrightarrow b$. Then $\{\!\{1 : a\}\!\} \rightsquigarrow_{\mathcal{A}} \mu$ iff either $a \hookrightarrow b$ and $\mu = \{\!\{1 : b\}\!\}$ for some b, or $\mu = \varnothing$ and a is a normal form in $\hookrightarrow$.*

3.1 Notions of Probabilistic Termination

A binary relation $\to$ is called *terminating* if it does not give rise to an infinite sequence $a_1 \to a_2 \to \dots$. In a probabilistic setting, infinite sequences are problematic only if they occur with *non-null* probability.

Definition 7 (AST). *A PARS $\mathcal{A}$ is* almost surely terminating *(*AST*) if for any reduction sequence $\{\mu_i\}_{i\in\mathbb{N}} \in \mathcal{A}(\mu)$, it holds that* $\lim_{n\to\infty} |\mu_n| = 0$.

Intuitively, $|\mu_n|$ is the probability of having n-step reducts, so its tendency towards zero indicates that infinite reductions occur with zero probability.

Example 8 (Example 2 Revisited). The system $\mathcal{W}_p$ is AST for $p \leq \frac{1}{2}$, whereas it is not for $p > \frac{1}{2}$. Note that although $\mathcal{W}_{\frac{1}{2}}$ is AST, the expected number of reductions needed to reach a terminal is infinite.

The notion of *positive almost sure termination (PAST)*, due to Bournez and Garnier [5], constitutes a refinement of AST demanding that in addition, the expected length of reduction is finite for every initial state a, independent of the employed strategy. In particular, $\mathcal{W}_{\frac{1}{2}}$ is not PAST. The expected length of a derivation can be concisely expressed in our setting as follows.

Definition 9 (Expected Derivation Length). *Let $\mathcal{A}$ be a PARS and $\boldsymbol{\mu} = \{\mu_i\}_{i\in\mathbb{N}} \in \mathcal{A}(\mu)$. We define the* expected derivation length $\mathsf{edl}(\boldsymbol{\mu}) \in \mathbb{R} \cup \{\infty\}$ *of $\boldsymbol{\mu}$ by*

$$\mathsf{edl}(\boldsymbol{\mu}) := \sum_{i\geq 1} |\mu_i| \ .$$

A PARS $\mathcal{A}$ is called *PAST* if for every reduction $\boldsymbol{\mu}$ starting from a, $\mathsf{edl}(\boldsymbol{\mu})$ is bounded. Without fixing a strategy, however, this condition does not ensure bounds on the derivation length.

Example 10. Consider the (non-probabilistic) ARS on $\mathbb{N} \cup \{\omega\}$ with reductions $\omega \to n$ and $n+1 \to n$ for every $n \in \mathbb{N}$. It is easy to see that every reduction sequence is of finite length, and thus, this ARS is PAST. There is, however, no global bound on the length of reduction sequences starting from ω.

Hence we introduce a stronger notion, which actually plays a more essential role than PAST. It is based on a natural extension of *derivation height* from complexity analysis of term rewriting.

Definition 11 (Strong AST). *A PARS $\mathcal{A}$ is* strongly almost surely terminating *(*SAST*) if the* expected derivation height $\mathsf{edh}_{\mathcal{A}}(a)$ *of every $a \in A$ is finite, where* $\mathsf{edh}_{\mathcal{A}}(a) \in \mathbb{R} \cup \{\infty\}$ *of a is defined by*

$$\mathsf{edh}_{\mathcal{A}}(a) := \sup_{\boldsymbol{\mu}\in\mathcal{A}(\{\!\{1:a\}\!\})} \mathsf{edl}(\boldsymbol{\mu}) \ .$$

3.2 Probabilistic Ranking Functions

Bournez and Garnier [5] generalized *ranking functions*, a popular and classical method for proving termination of non-probabilistic systems, to PARS. We give here a simpler but equivalent definition of probabilistic ranking function, taking advantage of the notion of multidistribution.

For a (multi)distribution μ over real numbers, the *expected value* of μ is denoted by $\mathbb{E}(\mu) := \sum_{p:x\in\mu} p \cdot x$. A function $f : A \to \mathbb{R}$ is naturally generalized to $f : \mathsf{FMDist}(A) \to \mathsf{FMDist}(\mathbb{R})$, so for $\mu \in \mathsf{FMDist}(A)$, $\mathbb{E}(f(\mu)) = \sum_{p:x\in\mu} p \cdot f(x)$. For $\epsilon > 0$ we define the order $>_\epsilon$ on $\mathbb{R}$ by $x >_\epsilon y$ iff $x \geq \epsilon + y$.

Definition 12. *Given a PARS $\mathcal{A}$ on A, we say that a function $f : A \to \mathbb{R}_{\geq 0}$ is a* (probabilistic) ranking function *(sometimes referred to as* Lyapunov *ranking function), if there exists $\epsilon > 0$ such that $a \to d \in \mathcal{A}$ implies $f(a) >_\epsilon \mathbb{E}(f(d))$.*

The above definition slightly differs from the formulation in [5]: the latter demands the *drift* $\mathbb{E}(f(d)) - f(a)$ to be at least $-\epsilon$, which is equivalent to $f(a) >_\epsilon \mathbb{E}(f(d))$; and allows any lower bound $\inf_{a\in A} f(a) > -\infty$, which can be easily turned into 0 by adding the lower bound to the ranking function.

We prove that a ranking function ensures SAST and gives a bound on expected derivation length. Essentially the same result can be found in [9], but we use only elementary mathematics not requiring notions from probability theory. We moreover show that this method is complete for proving SAST.

Lemma 13. *Let f be a ranking function for a PARS $\mathcal{A}$. Then there exists $\epsilon > 0$ such that $\mathbb{E}(f(\mu)) \geq \mathbb{E}(f(\nu)) + \epsilon \cdot |\nu|$ whenever $\mu \rightsquigarrow_{\mathcal{A}} \nu$.*

Proof. As f is a ranking function for $\mathcal{A}$, we have $\epsilon > 0$ such that $a \to d \in \mathcal{A}$ implies $f(a) >_\epsilon \mathbb{E}(f(d))$. Consider $\mu \rightsquigarrow_{\mathcal{A}} \nu$. We prove the claim by induction on the derivation of $\mu \rightsquigarrow_{\mathcal{A}} \nu$.

- Suppose $\mu = \{\!\{1 : a\}\!\}$ and $a \in \mathsf{TRM}(\mathcal{A})$. Then $\nu = \varnothing$ and $\mathbb{E}(f(\mu)) \geq 0 = \mathbb{E}(f(\nu)) + \epsilon \cdot |\nu|$ since $\mathbb{E}(f(\varnothing)) = |\varnothing| = 0$.
- Suppose $\mu = \{\!\{1 : a\}\!\}$ and $a \to \nu \in \mathcal{A}$. From the assumption $\mathbb{E}(f(\mu)) = f(a) >_\epsilon \mathbb{E}(f(\nu))$, and as $|\nu| = 1$ we conclude $\mathbb{E}(f(\mu)) \geq \mathbb{E}(f(\nu)) + \epsilon \cdot |\nu|$.
- Suppose $\mu = \biguplus_{i=1}^n p_i \cdot \mu_i$, $\nu = \biguplus_{i=1}^n p_i \cdot \nu_i$, and $\mu_i \rightsquigarrow_{\mathcal{A}} \nu_i$ for all $1 \leq i \leq n$. Induction hypothesis gives $\mathbb{E}(f(\mu_i)) \geq \mathbb{E}(f(\nu_i)) + \epsilon \cdot |\nu_i|$. Thus,

$$\begin{aligned}\mathbb{E}(f(\mu)) = \sum_{i=1}^n p_i \cdot \mathbb{E}(f(\mu_i)) &\geq \sum_{i=1}^n p_i \cdot (\mathbb{E}(f(\nu_i)) + \epsilon \cdot |\nu_i|) \\ &= \sum_{i=1}^n p_i \cdot \mathbb{E}(f(\nu_i)) + \epsilon \cdot \sum_{i=1}^n p_i \cdot |\nu_i| = \mathbb{E}(f(\nu)) + \epsilon \cdot |\nu| \,. \qquad \square\end{aligned}$$

Lemma 14. *Let f be a ranking function for PARS $\mathcal{A}$. Then there is $\epsilon > 0$ such that $\mathbb{E}(f(\mu_0)) \geq \epsilon \cdot \mathsf{edl}(\boldsymbol{\mu})$ for every $\boldsymbol{\mu} = \{\mu_i\}_{i\in\mathbb{N}} \in \mathcal{A}(\mu_0)$.*

Proof. We first show $\mathbb{E}(f(\mu_m)) \geq \sum_{i=m+1}^{n} |\mu_i|$ for every $n \geq m$, by induction on $m - n$. Let ϵ be given by Lemma 13. The base case is trivial, so let us consider the inductive step. By Lemma 13 and induction hypothesis we get

$$\begin{aligned}\mathbb{E}(f(\mu_m)) &\geq \mathbb{E}(f(\mu_{m+1})) + \epsilon \cdot |\mu_{m+1}| \\ &\geq \epsilon \cdot \sum_{i=m+2}^{n} |\mu_i| + \epsilon \cdot |\mu_{m+1}| = \epsilon \cdot \sum_{i=m+1}^{n} |\mu_i| \;.\end{aligned}$$

By fixing $m = 0$, we conclude that the sequence $\left\{\epsilon \cdot \sum_{i=1}^{n} |\mu_i|\right\}_{n\geq 1}$ is bounded by $\mathbb{E}(f(\mu_0))$, and so is its limit $\epsilon \cdot \sum_{i\geq 1} |\mu_i| = \epsilon \cdot \mathsf{edl}(\boldsymbol{\mu})$. ☐

Theorem 15. *Ranking functions are sound and complete for proving SAST.*

Proof. For soundness, let f be a ranking function for a PARS $\mathcal{A}$. For every derivation $\boldsymbol{\mu}$ starting from $\{\!\{1 : a\}\!\}$, we have $\mathsf{edl}(\boldsymbol{\mu}) \leq \frac{f(a)}{\epsilon}$ by Lemma 14. Hence, $\mathsf{edh}_{\mathcal{A}}(a) \leq \frac{f(a)}{\epsilon}$, concluding that $\mathcal{A}$ is SAST.

For completeness, suppose that $\mathcal{A}$ is SAST, and let $a \to d \in \mathcal{A}$. Then we have $\mathsf{edh}_{\mathcal{A}}(a) \in \mathbb{R}$, and

$$\begin{aligned}\mathsf{edh}_{\mathcal{A}}(a) \;=\; \sup_{\boldsymbol{\mu}\in\mathcal{A}(\{\!\{1:a\}\!\})} \mathsf{edl}(\boldsymbol{\mu}) \;&\geq\; \sup_{\boldsymbol{\mu}\in\mathcal{A}(d)} (1 + \mathsf{edl}(\boldsymbol{\mu})) \\ &=\; 1 + \sup_{\boldsymbol{\mu}\in\mathcal{A}(d)} \mathsf{edl}(\boldsymbol{\mu}) \;=\; 1 + \mathbb{E}(\mathsf{edh}_{\mathcal{A}}(d)) \;,\end{aligned}$$

concluding $\mathsf{edh}_{\mathcal{A}}(a) >_1 \mathbb{E}(\mathsf{edh}_{\mathcal{A}}(d))$. Thus, taking $\epsilon = 1$, $\mathsf{edh}_{\mathcal{A}}$ is a ranking function according to Definition 12. ☐

Bournez and Garnier claimed that ranking functions are complete for proving PAST, if the system is finitely branching [5, Theorem 3]. The claim does not hold,[4] as the following example illustrates that PAST and SAST do not coincide even for finitely branching systems.[5]

Example 16. Consider PARS $\mathcal{A}$ over $\mathbb{N} \cup \{a_n \mid n \in \mathbb{N}\}$, consisting of

$$a_n \to \{\tfrac{1}{2} : a_{n+1},\ \tfrac{1}{2} : 0\} \qquad a_n \to \{1 : 2^n \cdot n\} \qquad n + 1 \to \{1 : n\} \;.$$

Then P is finitely branching and PAST, because every reduction sequence from $\{\!\{1 : a_n\}\!\}$ with $n \in \mathbb{N}$ is one of the following forms:

- $\boldsymbol{\mu}_{n,0} = \{\!\{1 : a_n\}\!\} \rightsquigarrow \{\!\{1 : 2^n \cdot n\}\!\} \rightsquigarrow^{2^n\cdot n} \{\!\{1 : 0\}\!\}$;
- $\boldsymbol{\mu}_{n,m} = \{\!\{1 : a_n\}\!\} \rightsquigarrow^m \{\!\{\frac{1}{2^m} : a_{n+m},\ \frac{1}{2^m} : 0\}\!\} \rightsquigarrow \{\!\{\frac{1}{2^m} : 2^{n+m} \cdot (n + m)\}\!\}$ $\rightsquigarrow^{2^{n+m}\cdot(n+m)} \{\!\{\frac{1}{2^m} : 0\}\!\}$ with $m = 1, 2, \ldots$;
- $\boldsymbol{\mu}_{n,\infty} = \{\!\{1 : a_n\}\!\} \rightsquigarrow \{\!\{\frac{1}{2} : a_{n+1},\ \frac{1}{2} : 0\}\!\} \rightsquigarrow \{\!\{\frac{1}{4} : a_{n+2},\ \frac{1}{4} : 0\}\!\} \rightsquigarrow \cdots$,

and $\mathsf{edl}(\boldsymbol{\mu}_{n,\alpha})$ is finite for each $n \in \mathbb{N}$ and $\alpha \in \mathbb{N} \cup \{\infty\}$. However, e.g., $\mathsf{edh}_{\mathcal{A}}(a_0)$ is not bounded, since $\mathsf{edl}(\boldsymbol{\mu}_{0,m}) = \frac{1}{2^0} + \cdots + \frac{1}{2^{m-1}} + \frac{1}{2^m} + \frac{1}{2^m} \cdot (2^m \cdot m) \geq m$ for every $m \in \mathbb{N}$.

---

[4] The completeness claim of [5] has already been refuted in [14], but [14] also contradicts our completeness result. The counterexample there is invalid since a part of reduction steps are not counted. We thank Luis María Ferrer Fioriti for this analysis.

[5] We are grateful to the anonymous reviewer who pointed us to this example.

3.3 Relation to Formulation by Bournez and Garnier

As done by Bournez and Garnier [5], the dynamics of probabilistic systems are commonly defined as stochastic sequences, i.e., infinite sequences of random variables whose n-th variable represents the n-th reduct. A disadvantage of this approach is that nondeterministic choices have to be *a priori* resolved by means of strategies. In this section, we establish a precise correspondence between our formulation and the one of Bournez and Garnier. In particular, we show that the corresponding notions of AST and PAST coincide.

We shortly recap central definitions from [5]. We assume basic familiarity with stochastic processes, see e.g. [23]. Here we fix a PARS $\mathcal{A}$ on A. A *history* (of length $n+1$) is a finite sequence $\boldsymbol{a} = a_0, a_1, \dots, a_n$ of objects from A, and such a sequence is called *terminal* if a_n is. A *strategy* ϕ is a function from nonterminal histories to distributions such that $a_n \to \phi(a_0, a_1, \dots, a_n) \in \mathcal{A}$. A history $a_0, a_1, \dots, a_n$ is called *realizable under* ϕ iff for every $0 \le i < n$, it holds that $\phi(a_0, a_1, \dots, a_i)(a_{i+1}) > 0$.

Definition 17 (Stochastic Reduction, [5]). *Let $\mathcal{A}$ be a PARS on A and $\bot \notin A$ a special symbol. A sequence of random variables $\boldsymbol{X} = \{X_n\}_{n\in\mathbb{N}}$ over $A \cup \{\bot\}$ is a* (stochastic) reduction *in $\mathcal{A}$ (under strategy ϕ) if*

$$\begin{aligned} \mathbb{P}(X_{n+1} = \bot \mid X_n = \bot) &= 1; \\ \mathbb{P}(X_{n+1} = \bot \mid X_n = a) &= 1 && \text{if } a \text{ is terminal;} \\ \mathbb{P}(X_{n+1} = \bot \mid X_n = a) &= 0 && \text{if } a \text{ is nonterminal;} \\ \mathbb{P}(X_{n+1} = a \mid X_n = a_n, \dots, X_0 = a_0) &= d(a) && \text{if } \phi(a_0, \dots, a_n) = d, \end{aligned}$$

where $a_0, \dots, a_n$ is a realizable nonterminal history under ϕ.

Thus, $\boldsymbol{X}$ is set up so that trajectories correspond to reductions $a_0 \to_{\mathcal{A}} a_1 \to_{\mathcal{A}} \cdots$, and $\bot$ signals termination. In correspondence, the derivation length is given by the *first hitting time* to $\bot$:

Definition 18 ((P)AST of [5]). *For $\boldsymbol{X} = \{X_n\}_{n\in\mathbb{N}}$ define the random variable $T_{\boldsymbol{X}} := \min\{n \in \mathbb{N} \mid X_n = \bot\}$, where $\min \varnothing = \infty$ by convention. A PARS $\mathcal{A}$ is* stochastically AST *(resp. PAST) if for every stochastic reduction $\boldsymbol{X}$ in $\mathcal{A}$, $\mathbb{P}(T_{\boldsymbol{X}} = \infty) = 0$ (resp. $\mathbb{E}(T_{\boldsymbol{X}}) < \infty$).*

A proof of the following correspondence is available in the extened version [3].

Lemma 19. *For each stochastic reduction $\{X_n\}_{n\in\mathbb{N}}$ in a PARS $\mathcal{A}$ there exists a corresponding reduction sequence $\mu_0 \rightsquigarrow_{\mathcal{A}} \mu_1 \rightsquigarrow_{\mathcal{A}} \cdots$ where μ_0 is a distribution and $\mathbb{P}(X_n = a) = \sum_{p:a\in\mu_n} p$ for all $n \in \mathbb{N}$ and $a \in A$, and vice versa.*

As the above lemma relates $T_{\boldsymbol{X}}$ with the n-th reduction μ_n of the corresponding reduction so that $\mathbb{P}(T_{\boldsymbol{X}} \geq n) = \mathbb{P}(X_n \neq \bot) = |\mu_n|$, using that $\mathbb{E}(T_{\boldsymbol{X}}) = \sum_{n\in\mathbb{N}\cup\{\infty\}} \mathbb{P}(T_{\boldsymbol{X}} \geq n)$ [8], it is not difficult to derive the central result of this section:

Theorem 20. *A PARS $\mathcal{A}$ is (P)AST if and only if it is stochastically (P)AST.*

4 Probabilistic Term Rewrite Systems

Now we formulate probabilistic term rewriting following [5], and then lift the interpretation method for term rewriting to the probabilistic case.

We briefly recap notions from rewriting; see [4] for an introduction to rewriting. A *signature* F is a set of *function symbols* $\mathtt{f}$ associated with their *arity* $\mathsf{ar}(\mathtt{f}) \in \mathbb{N}$. The set $T(F,V)$ of *terms* over a signature F and a set V of variables (disjoint with F) is the least set such that $x \in T(F,V)$ if $x \in V$ and $\mathtt{f}(t_1,\dots,t_{\mathsf{ar}(\mathtt{f})}) \in T(F,V)$ whenever $\mathtt{f} \in F$ and $t_i \in T(F,V)$ for all $1 \le i \le \mathsf{ar}(\mathtt{f})$. A *substitution* is a mapping $\sigma : V \to T(F,V)$, which is extended homomorphically to terms. We write $t\sigma$ instead of $\sigma(t)$. A *context* is a term $C \in T(F, V \cup \{\Box\})$ containing exactly one occurrence of a special variable $\Box$. With $C[t]$ we denote the term obtained by replacing $\Box$ in C with t. We extend substitutions and contexts to multidistributions: $\mu\sigma := \{\{p_1 : t_1\sigma, \dots,\ p_n : t_n\sigma\}\}$ and $C[\mu] := \{\{p_1 : C[t_1], \dots,\ p_n : C[t_n]\}\}$ for $\mu = \{\{p_1 : t_1, \dots,\ p_n : t_n\}\}$. Given a multidistribution μ over A, we define a mapping $\overline{\mu} : A \to \mathbb{R}_{\ge 0}$ by $\overline{\mu}(a) := \sum_{p:a\in\mu} p$, which forms a distribution if $|\mu| = 1$.

Definition 21 (Probabilistic Term Rewriting). *A* probabilistic rewrite rule *is a pair of $l \in T(F,V)$ and $d \in \mathsf{FDist}(T(F,V))$, written $l \to d$. A* probabilistic term rewrite system (PTRS) *is a (typically finite) set of probabilistic rewrite rules. We write $\widehat{\mathcal{R}}$ for the PARS consisting of a probabilistic reduction $C[l\sigma] \to C[\overline{d\sigma}]$ for every probabilistic rewrite rule $l \to d \in \mathcal{R}$, context C, and substitution σ. We say a PTRS $\mathcal{R}$ is AST/SAST if $\widehat{\mathcal{R}}$ is.*

Note that, for a distribution d over terms, $d\sigma$ is in general a multidistribution; e.g., consider $\{\frac{1}{2}{:}x,\ \frac{1}{2}{:}y\}\sigma$ with $x\sigma = y\sigma$. This explains why we use $C[\overline{d\sigma}]$, which is a distribution, to obtain a probabilistic reduction above.

Example 22. The random walk of Example 2 can be modeled by a PTRS consisting of a single rule $\mathtt{s}(x) \to \{p : x,\ 1-p : \mathtt{s}(\mathtt{s}(x))\}$. To rewrite a term, there are typically multiple choices of a subterm to reduce (i.e., *redexes*). For instance, $\mathtt{s}(\mathtt{f}(\mathtt{s}(\mathtt{0})))$ has two redexes and consequently two possible reducts:

$$\{\{p : \mathtt{f}(\mathtt{s}(\mathtt{0})),\ 1-p : \mathtt{s}(\mathtt{s}(\mathtt{f}(\mathtt{s}(\mathtt{0}))))\}\} \quad \text{and} \quad \{\{p : \mathtt{s}(\mathtt{f}(\mathtt{0})),\ 1-p : \mathtt{s}(\mathtt{f}(\mathtt{s}(\mathtt{s}(\mathtt{0}))))\}\}\ .$$

4.1 Interpretation Methods for Proving SAST

We now generalise the *interpretation method* for term rewrite systems to the probabilistic setting. The following notion is standard.

Definition 23 (F-Algebra). *An F-*algebra *$\mathcal{X}$ on a non-empty carrier set X specifies the* interpretation *$\mathtt{f}_{\mathcal{X}} : X^{\mathsf{ar}(\mathtt{f})} \to X$ of each function symbol $\mathtt{f} \in F$. We say $\mathcal{X}$ is* monotone *with respect to a binary relation $\succ \subseteq X \times X$ if $x \succ y$ implies $\mathtt{f}_{\mathcal{X}}(\dots,x,\dots) \succ \mathtt{f}_{\mathcal{X}}(\dots,y,\dots)$ for every $\mathtt{f} \in F$. Given an* assignment *$\alpha : V \to X$, the interpretation of a term is defined as follows:*

$$[\![t]\!]^{\alpha}_{\mathcal{X}} := \begin{cases} \alpha(t) & \text{if } t \in V, \\ \mathtt{f}_{\mathcal{X}}([\![t_1]\!]^{\alpha}_{\mathcal{X}}, \dots, [\![t_n]\!]^{\alpha}_{\mathcal{X}}) & \text{if } t = \mathtt{f}(t_1,\dots,t_n). \end{cases}$$

We write $s \succ_{\mathcal{X}} t$ iff $[\![s]\!]^{\alpha}_{\mathcal{X}} \succ [\![t]\!]^{\alpha}_{\mathcal{X}}$ for every assignment α.

Theorem 24 (cf. [27]). *A TRS $\mathcal{R}$ is terminating iff there exists an F-algebra $\mathcal{X}$ which is monotone with respect to a well-founded order $\succ$ and satisfies $\mathcal{R} \subseteq \succ_{\mathcal{X}}$.*

In a proof of the completeness of the above theorem, the *term algebra $\mathcal{T}$*, an F-algebra on $T(F,V)$ such that $\mathtt{f}_{\mathcal{T}}(t_1,\dots,t_n) := \mathtt{f}(t_1,\dots,t_n)$, plays a crucial role. In this term algebra, assignments are substitutions, and $[\![t]\!]^{\sigma}_{\mathcal{T}} = t\sigma$. We will also use the term algebra when proving the completeness of the probabilistic version of interpertation method for proving SAST.

The following definition gives our probabilistic version of the interpretation method. It is sound and complete for proving SAST. To achieve completeness, we first keep the technique as general as possible. For an F-algebra $\mathcal{X}$, we lift the interpretation of terms to multidistributions as before, i.e.,

$$[\![\{\!\{p_1 : t_1, \dots,\ p_n : t_n\}\!\}]\!]^{\alpha}_{\mathcal{X}} := \{\!\{p_1 : [\![t_1]\!]^{\alpha}_{\mathcal{X}}, \dots,\ p : [\![t_n]\!]^{\alpha}_{\mathcal{X}}\}\!\} \ .$$

Definition 25 (Probabilistic F-Algebra). *A* probabilistic monotone *F-algebra $(\mathcal{X}, \sqsupset)$ is an F-algebra $\mathcal{X}$ equipped with a relation $\sqsupset \subseteq X \times \mathsf{FDist}(X)$, such that for every $\mathtt{f} \in F$, $\mathtt{f}_{\mathcal{X}}$ is monotone with respect to $\sqsupset$, i.e., $x \sqsupset d$ implies $\mathtt{f}_{\mathcal{X}}(\dots,x,\dots) \sqsupset \overline{\mathtt{f}_{\mathcal{X}}(\dots,d,\dots)}$ where $\mathtt{f}_{\mathcal{X}}(\dots,\cdot,\dots)$ is extended to (multi-) distributions. We say it is* collapsible *(cf. [19]) if there exist a function $\mathsf{G} : X \to \mathbb{R}_{\geq 0}$ and $\epsilon > 0$ such that $x \sqsupset d$ implies $\mathsf{G}(x) >_{\epsilon} \mathbb{E}(\mathsf{G}(d))$.*

For a relation $\sqsupset \subseteq X \times \mathsf{FDist}(X)$, we define the relation $\sqsupset_{\mathcal{X}} \subseteq T(F,V) \times \mathsf{FDist}(T(F,V))$ by $t \sqsupset_{\mathcal{X}} d$ iff $[\![t]\!]^{\alpha}_{\mathcal{X}} \sqsupset \overline{[\![d]\!]^{\alpha}_{\mathcal{X}}}$ for every assignment $\alpha : V \to X$. The following property is easily proven by induction.

Lemma 26. *Let $(\mathcal{X}, \sqsupset)$ be a probabilistic monotone F-algebra. If $s \sqsupset_{\mathcal{X}} d$ then $[\![s\sigma]\!]^{\alpha}_{\mathcal{X}} \sqsupset \overline{[\![d\sigma]\!]^{\alpha}_{\mathcal{X}}}$ and $[\![C[s]]\!]^{\alpha}_{\mathcal{X}} \sqsupset \overline{[\![C[d]]\!]^{\alpha}_{\mathcal{X}}}$ for arbitrary α, σ, and C.*

Theorem 27 (Soundness and Completeness). *A PTRS $\mathcal{R}$ is SAST iff there exists a collapsible monotone F-algebra $(\mathcal{X}, \sqsupset)$ such that $\mathcal{R} \subseteq \sqsupset_{\mathcal{X}}$.*

Proof. For the "if" direction, we show that the PARS $\widehat{\mathcal{R}}$ is SAST using Theorem 15. Let $\alpha : V \to X$ be an arbitrary assignment, which exists as X is nonempty. Consider $s \to d \in \widehat{\mathcal{R}}$. Then we have $s = C[l\sigma]$ and $d = C[\overline{d'\sigma}]$ for some σ, C, and $l \to d' \in \mathcal{R}$. By assumption we have $l \sqsupset_{\mathcal{X}} d'$, and thus $[\![s]\!]^{\alpha}_{\mathcal{X}} \sqsupset \overline{[\![d]\!]^{\alpha}_{\mathcal{X}}}$ by Lemma 26. The collapsibility of $\sqsupset$ gives a function $\mathsf{G} : X \to \mathbb{R}_{\geq 0}$ and $\epsilon > 0$ such that $\mathsf{G}([\![s]\!]^{\alpha}_{\mathcal{X}}) >_{\epsilon} \mathbb{E}(\mathsf{G}(\overline{[\![d]\!]^{\alpha}_{\mathcal{X}}}))$, and by extending definitions we easily see $\mathbb{E}(\mathsf{G}(\overline{[\![d]\!]^{\alpha}_{\mathcal{X}}})) = \mathbb{E}(\mathsf{G}([\![d]\!]^{\alpha}_{\mathcal{X}}))$. Thus $\mathsf{G}([\![\cdot]\!]^{\alpha}_{\mathcal{X}})$ is a ranking function.

For the "only if" direction, suppose that $\mathcal{R}$ is SAST. We show $(\mathcal{T}, \widehat{\mathcal{R}})$ forms a collapsible probabilistic monotone F-algebra orienting $\mathcal{R}$.

- Since $\mathcal{R}$ is SAST, Theorem 15 gives a ranking function $f : T(F,V) \to \mathbb{R}_{\geq 0}$ and $\epsilon > 0$ for the underlying PARS $\widehat{\mathcal{R}}$. Taking $\mathsf{G} = f$, $\widehat{\mathcal{R}}$ is collapsible.

– Suppose $s \mathrel{\widehat{\mathcal{R}}} d$. Then we have $s = C[l\sigma]$ and $d = C[\,\overline{d'\sigma}\,]$ for some C, σ, and $l \to d' \in \mathcal{R}$. As $f(\dots, C, \dots)$ is also a context, $f(\dots, s, \dots) \mathrel{\widehat{\mathcal{R}}} f(\dots, d, \dots)$, concluding monotonicity.
– For every probabilistic rewrite rule $l \to d \in \mathcal{R}$ and every assignment (i.e., substitution) $\sigma : V \to T(F, V)$, we have $[\![l]\!]^{\sigma}_{\mathcal{T}} = l\sigma \mathrel{\widehat{\mathcal{R}}} \overline{d\sigma} = \overline{[\![d]\!]^{\sigma}_{\mathcal{T}}}$, and hence $l \mathrel{\widehat{\mathcal{R}}_{\mathcal{T}}} d$. This concludes $\mathcal{R} \subseteq \widehat{\mathcal{R}}_{\mathcal{T}}$. □

4.2 Barycentric Algebras

As probabilistic F-algebras are defined so generally, it is not yet clear how to search them for ones that prove the termination of a given PTRS. Now we make one step towards finding probabilistic algebras, by imposing some conditions to (non-probabilistic) F-algebras, so that the relation $\sqsupset$ can be defined from orderings which we are more familiar with.

Definition 28 (Barycentric Domain). *A* barycentric domain *is a set X equipped with the* barycentric operation $\mathbb{E}_X : \mathsf{FDist}(X) \to X$.

Of particular interest in this work will be the barycentric domains $\mathbb{R}_{\geq 0}$ and $\mathbb{R}^m_{\geq 0}$ with barycentric operations $\mathbb{E}(\{p_1 : a_1, \dots,\ p_n : a_n\}) = \sum_{i=1}^{n} p_i \cdot a_i$.

We naturally generalize the following notions from standard mathematics.

Definition 29 (Concavity, Affinity). *Let $f : X \to Y$ be a function from and to barycentric domains. We say f is* concave *with respect to an order $\succ$ on Y if $f(\mathbb{E}_X(d)) \succcurlyeq \mathbb{E}_Y(\overline{f(d)})$ where $\succcurlyeq$ is the reflexive closure of $\succ$. We say f is* affine *if it satisfies $f(\mathbb{E}_X(d)) = \mathbb{E}_Y(\overline{f(d)})$.*

Clearly, every affine function is concave.

Now we arrive at the main definition and theorem of this section.

Definition 30 (Barycentric F-Algebra). *A* barycentric F-algebra *is a pair $(\mathcal{X}, \succ)$ of an F-algebra $\mathcal{X}$ on a barycentric domain X and an order $\succ$ on X, such that for every $\mathtt{f} \in F$, $\mathtt{f}_{\mathcal{X}}$ is monotone and concave with respect to $\succ$. We say it is* collapsible *if there exist a concave function $\mathsf{G}\colon X \to \mathbb{R}_{\geq 0}$ (with respect to $>$) and $\epsilon > 0$ such that $\mathsf{G}(x) >_{\epsilon} \mathsf{G}(y)$ whenever $x \succ y$.*

We define the relation $\succ^{\mathbb{E}} \subseteq X \times \mathsf{FDist}(X)$ by $x \succ^{\mathbb{E}} d$ iff $x \succ \mathbb{E}_X(d)$.

Note that the following theorem claims soundness but not completeness, in contrast to Theorem 27.

Theorem 31. *A PTRS $\mathcal{R}$ is SAST if $\mathcal{R} \subseteq \succ^{\mathbb{E}}_{\mathcal{X}}$ for a collapsible barycentric F-algebra $(\mathcal{X}, \succ)$.*

Proof. Due to Theorem 27, it suffices to show that $(\mathcal{X}, \succ^{\mathbb{E}})$ is a collapsible probabilistic monotone F-algebra. Concerning monotonicity, suppose $x \succ^{\mathbb{E}} d$, i.e., $x \succ \mathbb{E}_X(d)$, and let $\mathtt{f} \in F$. Since $\mathtt{f}_{\mathcal{X}}$ is monotone and concave with respect to $\succ$ in every argument, we have

$$\mathtt{f}_{\mathcal{X}}(\dots, x, \dots) \succ \mathtt{f}_{\mathcal{X}}(\dots, \mathbb{E}_X(d), \dots) \succcurlyeq \mathbb{E}_X(\overline{\mathtt{f}_{\mathcal{X}}(\dots, d, \dots)})\ .$$

Concerning collapsibility, whenever $x \succ \mathbb{E}_X(d)$ we have

$$\begin{aligned}
\mathsf{G}(x) &>_\epsilon \mathsf{G}(\mathbb{E}_X(d)) && \text{by assumption on } \mathsf{G}, \\
&\geq \mathbb{E}(\overline{\mathsf{G}(d)}) && \text{as } \mathsf{G} : X \to \mathbb{R} \text{ is concave with respect to } >, \\
&= \mathbb{E}(\mathsf{G}(d)) && \text{by the definition of } \mathbb{E} \text{ on multidistributions.}
\end{aligned}$$

□

The rest of the section recasts two popular interpretation methods, polynomial and matrix interpretations (over the reals), as barycentric F-algebras.

Polynomial interpretations were introduced (on natural numbers [21] and real numbers [22]) for the termination analysis of non-probabilistic rewrite systems. Various techniques for synthesizing polynomial interpretations (e.g., [15]) exist, and these techniques are easily applicable in our setting.

Definition 32 (Polynomial Interpretation). *A* polynomial interpretation *is an F-algebra $\mathcal{X}$ on $\mathbb{R}_{\geq 0}$ such that $\mathtt{f}_\mathcal{X}$ is a polynomial for every $\mathtt{f} \in F$. We say $\mathcal{X}$ is* multilinear *if every $\mathtt{f}_\mathcal{X}$ is of the following form with $\mathsf{c}_V \in \mathbb{R}_{\geq 0}$:*

$$\mathtt{f}_\mathcal{X}(x_1, \ldots, x_n) = \sum_{V \subseteq \{x_1,\ldots,x_n\}} \mathsf{c}_V \cdot \prod_{x_i \in V} x_i \ .$$

In order to use polynomial interpretations for probabilistic termination, multilinearity is necessary for satisfying the concavity condition.

Proposition 33. *Let $\mathcal{X}$ be a monotone multilinear polynomial interpretation and $\epsilon > 0$. If $[\![l]\!]^\alpha_\mathcal{X} >_\epsilon \mathbb{E}([\![d]\!]^\alpha_\mathcal{X})$ for every $l \to d \in \mathcal{R}$ and α, then the PTRS $\mathcal{R}$ is SAST.*

Proof. The order $>_\epsilon$ is trivially collapsible with $\mathsf{G}(x) = x$. Further, every multilinear polynomial is affine and thus concave in all variables. Hence $(\mathcal{X}, >_\epsilon)$ forms a barycentric F-algebra, and thus Theorem 31 shows that $\mathcal{R}$ is SAST. □

An observation by Lucas [22] also holds in probabilistic case: To prove a finite PTRS $\mathcal{R}$ SAST with polynomial interpretations, we do not have to find ϵ, but it is sufficient to check $l >^\mathbb{E}_\mathcal{X} d$ for all rules $l \to d \in \mathcal{R}$. Define $\epsilon_{l \to d} := \mathbb{E}([\![d]\!]^\alpha_\mathcal{X}) - [\![l]\!]^\alpha_\mathcal{X}$ for such α that $\alpha(x) = 0$. Then for any other α, we can show $\mathbb{E}([\![d]\!]^\alpha_\mathcal{X}) - [\![l]\!]^\alpha_\mathcal{X} \geq \epsilon_{l \to d} > 0$. As $\mathcal{R}$ is finite, we can take $\epsilon := \min\{\epsilon_{l \to d} \mid l \to d \in \mathcal{R}\} > 0$.

Example 34 (Example 22 Continued). Consider again the PTRS consisting of the single rule $\mathtt{s}(x) \to \{p : x,\ 1-p : \mathtt{s}(\mathtt{s}(x))\}$. Define the polynomial interpretation $\mathcal{X}$ by $\mathtt{0}_\mathcal{X} := 0$ and $\mathtt{s}_\mathcal{X}(x) := x + 1$. Then whenever $p > \frac{1}{2}$ we have

$$[\![\mathtt{s}(x)]\!]^\alpha_\mathcal{X} = x + 1 > p \cdot x + (1-p) \cdot (x+2) = \mathbb{E}([\![\{p : x,\ 1-p : \mathtt{s}(\mathtt{s}(x))\}]\!]^\alpha_\mathcal{X}) \ .$$

Thus, when $p > \frac{1}{2}$ the PTRS is SAST by Proposition 33.

We remark that polynomial interpretations are not covered by [5, Theorem 5], since *context decrease* [5, Definition 8] demands $[\![\mathtt{f}(t)]\!]^\alpha_\mathcal{X} - [\![\mathtt{f}(t')]\!]^\alpha_\mathcal{X} \leq [\![t]\!]^\alpha_\mathcal{X} - [\![t']\!]^\alpha_\mathcal{X}$, which excludes interpretations such as $\mathtt{f}_\mathcal{X}(x) = 2x$.

Matrix interpretations are introduced for the termination analysis of term rewriting [13]. Now we extend them for probabilistic term rewriting.

Definition 35 (Matrix Interpretation). *A* (real) matrix interpretation *is an F-algebra $\mathcal{X}$ on $\mathbb{R}^m_{\geq 0}$ such that for every $f \in F$, $\mathtt{f}_\mathcal{X}$ is of the form*

$$\mathtt{f}_\mathcal{X}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) = \sum_{i=1}^{n} C_i \cdot \boldsymbol{x}_i + \boldsymbol{c} \ , \tag{1}$$

where $\boldsymbol{c} \in \mathbb{R}^m_{\geq 0}$, and $C_i \in \mathbb{R}^{m \times m}_{\geq 0}$. The order $\gg_\epsilon \subseteq \mathbb{R}^m_{\geq 0} \times \mathbb{R}^m_{\geq 0}$ is defined by

$$(x_1, \ldots, x_m)^T \gg_\epsilon (y_1, \ldots, y_m)^T :\Longleftrightarrow x_1 >_\epsilon y_1 \text{ and } x_i \geq y_i \text{ for all } i = 2, \ldots, m.$$

It is easy to derive the following from Theorem 31:

Proposition 36. *Let $\mathcal{X}$ be a monotone matrix interpretation and $\epsilon > 0$. If $[\![l]\!]^\alpha_\mathcal{X} \gg_\epsilon \mathbb{E}([\![d]\!]^\alpha_\mathcal{X})$ for every $l \to d \in \mathcal{R}$ and α, then the PTRS $\mathcal{R}$ is SAST.*

As in polynomial interpretations, for finite systems we do not have to find ϵ. Monotonicity can be ensured if (1) satisfies $(C_i)_{1,1} \geq 1$ for all i, cf. [13].

Example 37. Consider the PTRS consisting of the single probabilistic rule

$$\mathtt{a}(\mathtt{a}(x)) \to \{p : \mathtt{a}(\mathtt{a}(\mathtt{a}(x))),\ 1-p : \mathtt{a}(\mathtt{b}(\mathtt{a}(x)))\} \ .$$

Consider the two-dimensional matrix interpretation

$$[\![\mathtt{a}]\!](\boldsymbol{x}) = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} \cdot \boldsymbol{x} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \ , \qquad [\![\mathtt{b}]\!](\boldsymbol{x}) = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \cdot \boldsymbol{x} \ .$$

Then we have

$$\begin{aligned} [\![\mathtt{a}(\mathtt{a}(x))]\!]^\alpha = \begin{bmatrix} x_1 + x_2 + 1 \\ 1 \end{bmatrix} &\gg_{1-2p} \begin{bmatrix} x_1 + x_2 + 2p \\ 1 \end{bmatrix} \\ &= p \cdot [\![\mathtt{a}(\mathtt{a}(\mathtt{a}(x)))]\!]^\alpha + (1-p) \cdot [\![\mathtt{a}(\mathtt{b}(\mathtt{a}(x)))]\!]^\alpha \end{aligned}$$

where $\alpha(x) = (x_1, x_2)^T$. Hence this PARS is SAST if $p < \frac{1}{2}$, by Proposition 36.

It is worthy of note that the above example cannot be handled with polynomial interpretations, intuitively because monotonicity enforces the interpretation of the probable reducts $\mathtt{a}(\mathtt{a}(\mathtt{a}(x)))$ and $\mathtt{a}(\mathtt{b}(\mathtt{a}(x)))$ to be greater than that of the left-hand side $\mathtt{a}(\mathtt{a}(x))$. Generally, polynomial and matrix interpretations are incomparable in strength.

5 Conclusion

This is a study on how much of the classic interpretation-based techniques well known in term rewriting can be extended to *probabilistic* term rewriting, and to what extent they remain automatable. The obtained results are quite encouraging, although finding ways to combine techniques is crucial if one wants to capture a reasonably large class of systems, similarly to what happens in ordinary term rewriting [2]. Another hopeful future work includes extending our result for proving AST, not only SAST.

We extended the termination prover `NaTT` [28] with a syntax for probabilistic rules, and implemented the probabilistic versions of polynomial and matrix interpretations. For usage and implementation details, we refer to the extended version of this paper. Here we only report that we tested the implementation on the examples presented in the paper and successfully found termination proofs.

The following example would deserve some attention.

Example 38. Consider the following encoding of [14, Figure 1]:

$$\begin{aligned} \mathtt{?}(x) &\to \{\tfrac{1}{2} : \mathtt{?}(\mathtt{s}(x)),\ \tfrac{1}{2} : \mathtt{\$}(\mathtt{g}(x))\} & \qquad \mathtt{\$}(\mathtt{0}) &\to \{1 : \mathtt{0}\} \\ \mathtt{?}(x) &\to \{1 : \mathtt{\$}(\mathtt{f}(x))\} & \mathtt{\$}(\mathtt{s}(x)) &\to \{1 : \mathtt{\$}(x)\} \end{aligned}$$

describing a game where the player (strategy) can choose either to quit the game and ensure prize $\mathtt{\$}(\mathtt{f}(x))$, or to try a coin-toss which on success increments the score and on failure ends the game with consolation prize $\mathtt{\$}(\mathtt{g}(x))$.

When $\mathtt{f}$ and $\mathtt{g}$ can be bounded by linear polynomials, it is possible to automatically prove that the system is SAST. For instance, with rules for $\mathtt{f}(x) = 2x$ and $\mathtt{g}(x) = \lfloor \frac{x}{2} \rfloor$, `NaTT` (combined with the SMT solver `z3` version 4.4.1) found the following polynomial interpretation proving SAST:

$$\begin{aligned} \mathtt{?}_{\mathcal{X}}(x) &= 7x + 11 & \qquad \mathtt{s}_{\mathcal{X}}(x) &= x + 1 & \qquad \mathtt{0}_{\mathcal{X}} &= 1 \\ \mathtt{f}_{\mathcal{X}}(x) &= 3x + 1 & \mathtt{g}_{\mathcal{X}}(x) &= 2x + 1 & \mathtt{\$}_{\mathcal{X}}(x) &= 2x + 1\ . \end{aligned}$$

Acknowledgments

We thank the anonymous reviewers for their constructive remarks that improved the paper. Example 12 is due to one of them. We thank Luis María Ferrer Fioriti for the analysis of a counterexample in [14]. This work is partially supported by the ANR projects 14CE250005 ELICA and 16CE250011 REPAS, the FWF project Y757, and JST ERATO HASUO Metamathematics for Systems Design Project (No. JPMJER1603).

References

1. Agha, G., Meseguer, J., Sen, K.: Pmaude: Rewrite-based specification language for probabilistic object systems. Electr. Notes in Theor. Comput. Sci. 153(2), 213–239 (2006)
2. Avanzini, M.: Verifying Polytime Computability Automatically. Ph.D. thesis, University of Innsbruck (2013)
3. Avanzini, M., Dal Lago, U., Yamada, A.: On Probabilistic Term Rewriting (Technical Report). CoRR cs/CC/1802.09774 (2018), available at `http://www.arxiv.org/abs/1802.09774`

4. Baader, F., Nipkow, T.: Term Rewriting and All That. Cambridge University Press (1998)
5. Bournez, O., Garnier, F.: Proving positive almost-sure termination. In: Proc. of 16^{th} RTA. LNCS, vol. 3467, pp. 323–337. Springer (2005)
6. Bournez, O., Garnier, F.: Proving positive almost sure termination under strategies. In: Proc. of 17^{th} RTA. LNCS, vol. 4098, pp. 357–371. Springer (2006)
7. Bournez, O., Kirchner, C.: Probabilistic Rewrite Strategies. Applications to ELAN. In: Proc. of 13^{th} RTA. pp. 252–266 (2002)
8. Brmaud, P.: Marcov Chains. Springer (1999)
9. Chatterjee, K., Fu, H., Goharshady, A.K.: Termination Analysis of Probabilistic Programs Through Positivstellensatz's. In: Proc. of 28^{th} CAV. LNCS, vol. 9779, pp. 3–22. Springer (2016)
10. Dal Lago, U., Zorzi, M.: Probabilistic Operational Semantics for the Lambda Calculus. RAIRO - TIA 46(3), 413–450 (2012)
11. Dal Lago, U., Grellois, C.: Probabilistic Termination by Monadic Affine Sized Typing. In: Proc. of 26^{th} ESOP. pp. 393–419 (2017)
12. Dal Lago, U., Martini, S.: On Constructor Rewrite Systems and the Lambda Calculus. LMCS 8(3) (2012)
13. Endrullis, J., Waldmann, J., Zantema, H.: Matrix Interpretations for Proving Termination of Term Rewriting. JAR 40(3), 195–220 (2008)
14. Ferrer Fioriti, L.M., Hermanns, H.: Probabilistic Termination: Soundness, Completeness, and Compositionality. In: Proc. of 42^{nd} POPL. pp. 489–501. ACM (2015)
15. Fuhs, C., Giesl, J., Middeldorp, A., Schneider-Kamp, P., Thiemann, R., Zankl, H.: SAT Solving for Termination Analysis with Polynomial Interpretations. In: Proc. of 10^{th} SAT. LNCS, vol. 4501, pp. 340–354. Springer (2007)
16. Gnaedig, I.: Induction for positive almost sure termination. In: PPDP 2017. pp. 167–178. ACM (2007)
17. Goldwasser, S., Micali, S.: Probabilistic Encryption. JCSS 28(2), 270–299 (1984)
18. Goodman, N.D., Mansinghka, V.K., Roy, D.M., Bonawitz, K., Tenenbaum, J.B.: Church: a Language for Generative Models. In: Proc. of 24^{th} UAI. pp. 220–229. AUAI Press (2008)
19. Hirokawa, N., Moser, G.: Automated complexity analysis based on context-sensitive rewriting. In: RTA-TLCA 2014. LNCS, vol. 8560, pp. 257–271 (2014)
20. Kaminski, B.L., Katoen, J.: On the hardness of almost-sure termination. In: MFCS 2015, Milan, Italy, August 24-28, 2015, Proceedings, Part I. pp. 307–318 (2015)
21. Lankford, D.: Canonical algebraic simplification in computational logic. Tech. Rep. ATP-25, University of Texas (1975)
22. Lucas, S.: Polynomials Over the Reals in Proofs of Termination: From Theory to Practice. ITA 39(3), 547–586 (2005)
23. Puterman, M.L.: Markov Decision Processes: Discrete Stochastic Dynamic Programming. John Wiley & Sons, Inc., New York, NY, USA, 1st edn. (1994)
24. Rabin, M.O.: Probabilistic Automata. Information and Control 6(3), 230–245 (1963)
25. Saheb-Djahromi, N.: Probabilistic LCF. In: MFCS. pp. 442–451 (1978)
26. Santos, E.S.: Probabilistic Turing Machines and Computability. Proc. of the American Mathematical Society 22(3), 704–710 (1969)
27. Terese: Term Rewriting Systems, Cambridge Tracts in Theoretical Computer Science, vol. 55. Cambridge University Press (2003)
28. Yamada, A., Kusakari, K., Sakabe, T.: Nagoya Termination Tool. In: RTA-TLCA 2014. LNCS, vol. 8560, pp. 466–475 (2014)