On Probabilistic Term Rewriting

Martin Avanzini[a], Ugo Dal Lago[a,b], Akihisa Yamada[c]

[a]*INRIA Sophia Antipolis, France*
[b]*University of Bologna, Italy*
[c]*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 nontrivial 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 algorithmics [32] or cryptography [19]. All this has stimulated the study of probabilistic computational models and programming languages: probabilistic variations on well-known models like automata [12, 35], Turing machines [37, 17], and the λ-calculus [36, 23] are known from the early days of theoretical computer science.

The simplest way probabilistic choice can be made available in programming languages consists in 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 [11]).

Term rewriting [38] is a well-studied model of computation where no probabilistic behavior is involved. It provides a faithful abstraction of pure functional programming which is, up to a certain extent, also adequate for modeling higher-order parameter passing [27]. 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 Bournez and Garnier [6]. 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 [38] (Section 3). In particular, what plays the role of a reduction sequence, usually a (possibly infinite) sequence $a_1 \rightarrow a_2 \rightarrow \ldots$ of *states*, is a sequence $\mu_1 \rightarrow^{\mathcal{M}} \mu_2 \rightarrow^{\mathcal{M}} \ldots$ 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 behavior of each rule and the nondeterminism in rule selection. Such a 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. Consequently, our framework is not only conceptually simpler, it is closer to ordinary rewriting. Indeed, our reduction relation is simply a reduction system over multidistributions, and can be studied without particular knowledge of probability theory. Nevertheless, the two frameworks turn out to be equiexpressive (Section 3.4).

On top of this framework, we then introduce abstract *embeddings* and more concrete probabilistic ranking functions, sound and complete methods for proving *strong almost sure termination*, a strengthening of *positive almost sure termination* [6]. We moreover show that ranking functions provide bounds on expected runtimes (Section 3.3).

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 (Section 4). 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` [39] (Section 5).

This paper is revised version of the conference paper [3]. Apart from giving all missing proofs and more detailed examples and explanations, we generalized the definition of probabilistic reduction systems so that reducts may be sampled from infinite distributions. Another notable change lies in the reformulation of probabilistic ranking functions in terms of embeddings, leading also to a more

2

elegant formulation of barycentric algebras.

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 [38].

Termination remains an interesting and desirable property in a probabilistic setting, e.g., in probabilistic programming [20] 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, the termination questions for almost-sure and positive almost-sure termination on a *single* input are already Π_2^0 and Σ_2^0 complete, respectively [24].

Many sound verification methodologies for probabilistic termination have recently been introduced (see, e.g., [6, 7, 18, 14, 10]). 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 [26]. More recently, Ngo et al. [33] implemented the `ert`-calculus of Kaminski et al. [25] for reasoning about the expected runtime of imperative integer programs, showing promising results.[1]

Finally, as already mentioned, the current work can be seen as stemming from the work by Bournez et al. [8, 6, 7]. 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. Our completeness result can be seen as a correction to their claim [6, Theorem 3], which was already refuted [14]. In fact, incompleteness claims in [14] also contradict our result, but their counterexample is invalid as part of the reduction steps are not counted.[2] Moreover, an interpretation method inspired by ranking functions is made more general here, this way accommodating not only interpretations over the real numbers,

---

[1]While it is possible to translate such programs into probabilistic term rewrite systems [4], our method will not capture typical termination arguments on probabilistic integer programs, since properties of integers will be unavailable after the encoding.
[2]We thank Luis María Ferrer Fioriti for this analysis.

but also interpretations over vectors, in the spirit of matrix interpretations. Finally, we provide an automation of polynomial and matrix interpretation inference, whereas nothing about implementation was presented in Bournez and Garnier's work.

Some of our results have been independently obtained by Fu and Chatterjee [15], most notably, Theorem 2. Their notion of *bounded termination* is what we call strong almost-sure termination in this work. The main peculiarity of our work is that we focus on probabilistic term rewriting, while their focus is on procedural programs with probabilistic sampling. Besides, the authors also investigate lower bounds on expected runtimes and tail-probabilities, topics not covered in our 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 [6] extended the ARS formalism to probabilistic computations by allowing reducts to be sampled from a *(probability) distribution*. Throughout the paper, we denote by $\mathbb{R}_{\geq 0}$ the set of non-negative reals.

Definition 1 (Distribution). A *distribution* on a (countable) set A is a mapping $d : A \to \mathbb{R}_{\geq 0}$ assigning to each $a \in A$ a probability $d(a)$, such that $\sum_{a \in A} d(a) = 1$. We write $\mathcal{D}(A)$ for the set of distributions over A. The *support* of a distribution d is the set $\mathsf{Supp}(d) := \{a \in A \mid d(a) > 0\}$. We may denote a distribution d by $\{d(a) : a \mid a \in \mathsf{Supp}(d)\}$, or $\{d(a_1) : a_1, \ldots, d(a_n) : a_n\}$ when $\mathsf{Supp}(d)$ is the finite set $\{a_1, \ldots, a_n\}$.

A *probabilistic ARS (PARS)* in terms of Bournez and Garnier [6] is defined as a relation $\to \subseteq A \times \mathcal{D}(A)$, and having $a \to d$ means that d is the distribution of the one-step reducts of a, or a reduces to b with probability $p = d(b)$, in notation $p : a \to b$. We extend this notation to sequences, also called *runs* below. For instance, if further $q : b \to c$, we write $p \cdot q : a \to b \to c$. Notice that since $\to$ is a relation, the PARS may indeed specify more than one distribution of one-step reductions for $a \in A$, i.e., $a \to d_1$ and $a \to d_2$ with $d_1 \neq d_2$. The distribution of one-step reducts of a is nondeterministically chosen from d_1 and d_2 in this case, and $\to$ is called *nondeterministic*.

Example 1 (Random walk)**.** *A random walk over* $\mathbb{N}$ *with* bias *probability p is modeled by the probabilistic ARS* $\xrightarrow[\mathcal{W}_p]{}$ *such that*

$$n + 1 \xrightarrow[\mathcal{W}_p]{} \{p : n, \ 1 - p : n + 2\} \quad \textit{for all } n \in \mathbb{N}.$$

For a PARS $\to$, we aim at defining a reduction relation $\to^{\mathcal{D}}$, as an ARS on distributions. Taking Example 1, we would like to have

$$\{1 : 1\} \xrightarrow[\mathcal{W}_{\frac{1}{2}}]{}^{\mathcal{D}} \{\tfrac{1}{2} : 0, \ \tfrac{1}{2} : 2\} \,,$$

meaning that the distribution of the one-step reducts of 1, or more precisely the singleton distribution assigning probability one to 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): since 0 cannot be reduced, by probability $\frac{1}{2}$ we have no two-step reduct of 1. Hence we consider *subdistributions*, i.e., generalizations of distributions where probabilities may sum up to less than one, allowing

$$\{1:1\} \xrightarrow[\mathcal{W}_{\frac{1}{2}}]{\mathcal{D}} \{\tfrac{1}{2}:0,\ \tfrac{1}{2}:2\} \xrightarrow[\mathcal{W}_{\frac{1}{2}}]{\mathcal{D}} \{\tfrac{1}{4}:1,\ \tfrac{1}{4}:3\}\ .$$

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 due to the run $\frac{1}{8}:1 \to 2 \to 1 \to 2$, and the other half is due to the run $\frac{1}{8}:1 \to 2 \to 3 \to 2$. It turns out that, in the presence of nondeterminism, we should distinguish the two possibilities.

Example 2. *Consider the PARS $\xrightarrow[\mathcal{N}]{}$ such that*

$$\mathsf{a} \xrightarrow[\mathcal{N}]{} \{\tfrac{1}{2}:\mathsf{b}_1,\ \tfrac{1}{2}:\mathsf{b}_2\} \qquad \mathsf{b}_1 \xrightarrow[\mathcal{N}]{} \{1:\mathsf{c}\} \qquad \mathsf{c} \xrightarrow[\mathcal{N}]{} \{1:\mathsf{d}_1\}$$

$$\mathsf{b}_2 \xrightarrow[\mathcal{N}]{} \{1:\mathsf{c}\} \qquad \mathsf{c} \xrightarrow[\mathcal{N}]{} \{1:\mathsf{d}_2\}\ .$$

Reducing a twice always yields c, so the two-step reduct of a seen as a distribution is $\{1:\mathsf{c}\}$. More precisely, there are two runs from a to c, namely, $\frac{1}{2}:\mathsf{a} \to \mathsf{b}_1 \to \mathsf{c}$ and $\frac{1}{2}:\mathsf{a} \to \mathsf{b}_2 \to \mathsf{c}$. Each of them can be nondeterministically continued to d_1 and d_2, so the distribution of three-step reducts of a is the nondeterministic choice among $\{1:\mathsf{d}_1\}$, $\{\frac{1}{2}:\mathsf{d}_1,\frac{1}{2}:\mathsf{d}_2\}$, $\{1:\mathsf{d}_2\}$. On the other hand, whereas it is obvious that $\{1:\mathsf{c}\}$ should reduce to $\{1:\mathsf{d}_1\}$ or $\{1:\mathsf{d}_2\}$, obtaining the third choice $\{\frac{1}{2}:\mathsf{d}_1,\frac{1}{2}:\mathsf{d}_2\}$ would require the reduction relation $\xrightarrow[\mathcal{N}]{\mathcal{D}}$ to be defined in a non-local manner.

To overcome this problem, we refine subdistributions to *multidistributions*, where a single element can be associated with *more than one* probability. With respect to the nondeterministic PARS $\xrightarrow[\mathcal{N}]{}$, this allows us to model that a reduces in two steps to c on two distinct runs, each with probability $\frac{1}{2}$, and continuing this reduction may yield any of the above three-step reducts.

3.1. Probabilistic ARSs and Multidistribution Reductions

We model multidistribution by a special form of (possibly infinite) multisets.

Definition 2 (Multiset). A multiset over a set A is a mapping $M : A \to \mathbb{N}$. The *union* $\biguplus_{i \in I} M_i$ of countably many multisets M_i is defined by

$$\left(\biguplus_{i \in I} M_i\right)(a) := \sum_{i \in I} M_i(a)\ ,$$

which forms a multiset if and only if $\sum_{i \in I} M_i(a)$ is finite for every $a \in A$. The *sum* of a multiset M with respect to $f : A \to \mathbb{R}$ is defined by

$$\sum_{a \in M} f(a) := \sum_{a \in A} M(a) \cdot f(a) \; .$$

The *submultiset* relation is defined by $M \subseteq N :\Longleftrightarrow \forall a \in A. \; M(a) \leq N(a)$.

We use set-like notations for multisets: $\varnothing$ denotes the *empty* multiset $\varnothing(a) := 0$, $\{\!\{a_i \mid i \in I\}\!\}$ is the multiset M with $M(a) = |\{i \in I \mid a_i = a\}|$, and $\{\!\{a_1, \dots, a_n\}\!\}$ is its special case where $I = \{1, \dots, n\}$ is finite. The following lemma is an easy consequence of the definition.

Lemma 1. *For a family* $(\{\!\{a_{i,j} \mid j \in J_i\}\!\})_{i \in I}$ *of multisets we have*

$$\biguplus_{i \in I} \{\!\{a_{i,j} \mid j \in J_i\}\!\} = \{\!\{a_{i,j} \mid i \in I, j \in J_i\}\!\} \; .$$

Definition 3 (Multidistributions). A *(sub)multidistribution* on a set A is a multiset μ of pairs of $a \in A$ and $0 < p \leq 1$, written $p : a$, satisfying

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

We call μ a *proper multidistribution* if $|\mu| = 1$. We denote the set of multidistributions on A by $\mathcal{M}_{\leq 1}(A)$, and proper ones by $\mathcal{M}(A)$.

By an abuse of notation, we identify a subdistribution $\{p_i : a_i \mid i \in I\}$ with the multidistribution $\{\!\{p_i : a_i \mid i \in I\}\!\}$. We often lift a function $f : A \to B$ to $f : \mathcal{M}_{\leq 1}(A) \to \mathcal{M}_{\leq 1}(B)$ as follows:

$$f\big(\{\!\{p_i : a_i \mid i \in I\}\!\}\big) := \{\!\{p_i : f(a_i) \mid i \in I\}\!\} \; .$$

Definition 4 (PARS). A *probabilistic ARS (PARS)* over a set A is a (typically infinite) set $\to \; \subseteq A \times \mathcal{M}(A)$. An object $a \in A$ is called *terminal*, or a *normal form* in $\to$, if there is no μ with $a \to \mu$. With $\mathsf{NF}_\to$ we denote the set of normal forms in $\to$.

We remark that in contrast to [3, 6], we allow proper multidistributions rather than distributions as right-hand sides, e.g., $a \to \{\!\{\frac{1}{2} : a, \frac{1}{2} : a\}\!\}$ is permitted in our setting.

Now we lift a PARS $\to \; \subseteq A \times \mathcal{M}(A)$ to the reduction relation $\to^{\mathcal{M}} \; \subseteq \mathcal{M}_{\leq 1}(A) \times \mathcal{M}_{\leq 1}(A)$, an ARS over multidistributions. To this end, we need some basic operations on multidistributions. The *scalar multiplication* of a multidistribution is defined by

$$p \cdot \{\!\{q_i : a_i \mid i \in I\}\!\} := \{\!\{p \cdot q_i : a_i \mid i \in I\}\!\} \; ,$$

which is also a multidistribution if $0 < p \leq 1$. More generally, multidistributions are closed under *subconvex combinations*:

Lemma 2. *For families $(\mu_i)_{i\in I}$ of multidistributions and $(p_i)_{i\in I}$ of positive real numbers, $\biguplus_{i\in I} p_i \cdot \mu_i$ is also a multidistribution if $\sum_{i\in I} p_i \leq 1$. More precisely, $\left|\biguplus_{i\in I} p_i \cdot \mu_i\right| = \sum_{i\in I} p_i \cdot |\mu_i| \leq 1$.*

Proof. Let $\mu_i = \{\!\{q_{i,j} : a_{i,j} \mid j \in J_i\}\!\}$ for each $i \in I$. Using Lemma 1 we have

$$\left|\biguplus_{i\in I} p_i \cdot \mu_i\right| = |\{\!\{p_i \cdot q_{i,j} : a_{i,j} \mid i \in I, j \in J_i\}\!\}|$$

$$= \sum_{i\in I}\sum_{j\in J_i} p_i \cdot q_{i,j}$$

$$= \sum_{i\in I} p_i \cdot \sum_{j\in J_i} q_{i,j} = \sum_{i\in I} p_i \cdot |\mu_i|\,.$$

Since $|\mu_i| \leq 1$, the claim then follows from the assumption $\sum_{i\in I} p_i \leq 1$. $\qquad\square$

Definition 5 (Probabilistic Reduction). Given a PARS $\to\, \subseteq A \times \mathcal{M}(A)$, we define the *probabilistic reduction relation* $\to^{\mathcal{M}}\, \subseteq \mathcal{M}_{\leq 1}(A) \times \mathcal{M}_{\leq 1}(A)$ as follows:

$$\frac{a \in \mathsf{NF}_{\to}}{\{\!\{1 : a\}\!\} \to^{\mathcal{M}} \varnothing} \qquad \frac{a \to \mu}{\{\!\{1 : a\}\!\} \to^{\mathcal{M}} \mu} \qquad \frac{\forall i \in I.\ \mu_i \to^{\mathcal{M}} \nu_i}{\biguplus_{i\in I} p_i \cdot \mu_i \to^{\mathcal{M}} \biguplus_{i\in I} p_i \cdot \nu_i}$$

In the last rule, I is an arbitrary (possibly empty) countable set, $p_i > 0$ for every $i \in I$, and $\sum_{i\in I} p_i \leq 1$. We denote by $\mathsf{red}_{\to}(\mu)$ the set of all possible reduction sequences from μ, i.e., $(\mu_n)_{n\in\mathbb{N}} \in \mathsf{red}_{\to}(\mu)$ iff $\mu_0 = \mu$ and $\mu_n \to^{\mathcal{M}} \mu_{n+1}$ for any $n \in \mathbb{N}$. We overload this notation for $a \in A$ and denote by $\mathsf{red}_{\to}(a)$ the set of all possible reduction sequences from $\{\!\{1 : a\}\!\}$.

In essence, the reduction relation $\to^{\mathcal{M}}$ induced by a PARS $\to$ is given by a pointwise extension of $\to$ to multidistributions, removing normal forms along reductions. We abbreviate $\to^{\mathcal{M}}$ by $\to$ when it is clear from the context.

Let us illustrate the definition on the two examples from above.

Example 3 (Example 1, Revisited)**.** *The informal reduction of the PARS $\xrightarrow{\mathcal{W}_{\frac{1}{2}}}$ on distributions as outlined above is given by the following reduction sequence on multidistributions:*

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

$$\xrightarrow{\mathcal{W}_{\frac{1}{2}}} \{\!\{\tfrac{1}{8}:0,\ \tfrac{1}{8}:2,\ \tfrac{1}{8}:2,\ \tfrac{1}{8}:4\}\!\} \xrightarrow{\mathcal{W}_{\frac{1}{2}}} \cdots$$

Here, the first step follows from the second rule of Definition 5. The second is obtained by a combination of all three rules:

$$\frac{\dfrac{0 \in \mathsf{NF}_{\xrightarrow{\mathcal{W}_{\frac{1}{2}}}}}{\{\!\{1:0\}\!\} \xrightarrow{\mathcal{W}_{\frac{1}{2}}} \varnothing} \qquad \dfrac{2 \xrightarrow{\mathcal{W}_{\frac{1}{2}}} \{\!\{\tfrac{1}{2}:1,\ \tfrac{1}{2}:3\}\!\}}{\{\!\{1:2\}\!\} \xrightarrow{\mathcal{W}_{\frac{1}{2}}} \{\!\{\tfrac{1}{2}:1,\ \tfrac{1}{2}:3\}\!\}}}{\{\!\{\tfrac{1}{2}:0,\ \tfrac{1}{2}:2\}\!\} \xrightarrow{\mathcal{W}_{\frac{1}{2}}} \{\!\{\tfrac{1}{4}:1,\ \tfrac{1}{4}:3\}\!\}}$$

The third step is derived similarly.

In Example 3, the third reduct $\left\{\!\!\left\{\frac{1}{8}:0,\ \frac{1}{8}:2,\ \frac{1}{8}:2,\ \frac{1}{8}:4\right\}\!\!\right\}$ records that 2 derives from 1 via two distinct paths. In consequence, this resolves the issues indicated in Example 2 when dealing with nondeterministic systems.

Example 4 (Example 2, Revisited)**.** *Consider again the nondeterministic PARS $\underset{\mathcal{N}}{\rightarrow}$. Besides the two reductions yielding $\mathsf{d_1}$ and $\mathsf{d_2}$ outlined above, we have the reduction*

$$\left\{\!\!\left\{1:\mathsf{a}\right\}\!\!\right\}\ \underset{\mathcal{N}}{\rightarrow}\ \left\{\!\!\left\{\tfrac{1}{2}:\mathsf{b_1},\tfrac{1}{2}:\mathsf{b_2}\right\}\!\!\right\}\ \underset{\mathcal{N}}{\rightarrow}\ \left\{\!\!\left\{\tfrac{1}{2}:\mathsf{c},\tfrac{1}{2}:\mathsf{c}\right\}\!\!\right\}\ \underset{\mathcal{N}}{\rightarrow}\ \left\{\!\!\left\{\tfrac{1}{2}:\mathsf{d_1},\tfrac{1}{2}:\mathsf{d_2}\right\}\!\!\right\}\ .$$

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

The definition of the reduction relation $\rightarrow^{\mathcal{M}}$ has several consequences. First, any multidistribution μ is "reducible" in $\rightarrow^{\mathcal{M}}$, even if μ consists only of terminal objects. Second, if μ is the subconvex combination of multidistributions ν_i ($i \in I$), then any reduct of μ is given by the subconvex combination of some reducts of ν_i. Further, this observation carries over to reduction sequences. Let us extend subconvex combination pointwise to sequences of multidistributions, i.e., $\vec{\mu} = \biguplus_{i \in I} p_i \cdot (\vec{\nu_i})_{i \in I}$ if $\mu_n = \biguplus_{i \in I} p_i \cdot (\nu_{i,n})_{i \in I}$ for all $n \in \mathbb{N}$, where $\vec{\mu} = (\mu_n)_{n \in \mathbb{N}}$ and $\vec{\nu_i} = (\nu_{i,n})_{n \in \mathbb{N}}$ for all $i \in I$.

Lemma 3. *$\vec{\mu} \in \mathsf{red}_{\rightarrow}\left(\biguplus_{i \in I} p_i \cdot \nu_i\right)$ if and only if there exists $\vec{\rho_i} \in \mathsf{red}_{\rightarrow}(\nu_i)$ for each $i \in I$ such that $\vec{\mu} = \biguplus_{i \in I} p_i \cdot \vec{\rho_i}$.*

Proof. The "if" direction is obvious. For the "only if" direction, consider an arbitrary $\vec{\mu} \in \mathsf{red}_{\rightarrow}\left(\biguplus_{i \in I} p_i \cdot \nu_i\right)$. We inductively construct $\rho_{i,n}$ so that $(\rho_{i,n})_{n \in \mathbb{N}} \in \mathsf{red}_{\rightarrow}(\nu_i)$. For the base case we define $\rho_{i,0} = \nu_i$. For the inductive case, assume by induction hypothesis that $\mu_n = \biguplus_{i \in I} p_i \cdot \rho_{i,n} \rightarrow \mu_{n+1}$. An induction on the derivation of this step yields $\rho_{i,n+1}$ such that $\rho_{i,n} \rightarrow \rho_{i,n+1}$ for every $i \in I$. This concludes the proof. $\qquad\square$

Finally, we remark that PARSs constitute a generalization of ARSs: an ARS can be seen as a PARS whose right-hand sides are *singleton distributions*, i.e., $\left\{\!\!\left\{1:b\right\}\!\!\right\}$ for some b, and the non-probabilistic reduction is simulated via the relation $\left\{\!\!\left\{1:\cdot\right\}\!\!\right\} \rightarrow \left\{\!\!\left\{1:\cdot\right\}\!\!\right\}$. Only a little care is needed as $\left\{\!\!\left\{1:a\right\}\!\!\right\}$ reduces to $\varnothing$ when a is terminal.

Proposition 1. *Let $\rightarrow$ be an ARS and define $\hookrightarrow$ as the smallest PARS such that $a \hookrightarrow \left\{\!\!\left\{1:b\right\}\!\!\right\}$ if $a \rightarrow b$. Then $\left\{\!\!\left\{1:a\right\}\!\!\right\} \hookrightarrow \mu$ iff either $a \rightarrow b$ and $\mu = \left\{\!\!\left\{1:b\right\}\!\!\right\}$ for some b, or a is a normal form in $\rightarrow$ and $\mu = \varnothing$.*

Proof. For $\left\{\!\!\left\{1:a\right\}\!\!\right\} \hookrightarrow \mu$ only the first two rules of Definition 5 are effective. Then the claim directly follows. $\qquad\square$

3.2. Notions of Probabilistic Termination

An ARS $\to$ is called *terminating* if it does not give rise to an infinite sequence $a_1 \to a_2 \to \dots$. In a probabilistic setting, our interest is whether infinite sequences occur with non-zero probability or not. This notion is defined using our notation as follows.

Definition 6 (Almost Sure Termination; AST). A PARS $\to$ is said to be *almost surely terminating* (AST) if for any reduction sequence $(\mu_n)_{n \in \mathbb{N}} \in \mathsf{red}_\to(a)$, it holds that $\lim_{n \to \infty} |\mu_n| = 0$.

Intuitively, $|\mu_n|$ is the probability of having a reduct in the nth step, so its tendency towards zero indicates that infinite runs occur with zero probability.

Example 5 (Example 1, Revisited). *The PARS $\xrightarrow[\mathcal{W}_p]{}$ is AST for $p \geq \frac{1}{2}$, whereas it is not for $p < \frac{1}{2}$ (cf. [31]). Note that although $\xrightarrow[\mathcal{W}_{\frac{1}{2}}]{}$ is AST, the expected number of steps needed to reach a terminal is infinite.*

As motivated by Bournez and Garnier [6], it is also interesting and important to further ensure that the *expected length of runs* is finite. We define the notion by means of the following concise definitions. In Section 3.4 we show that our definition is equivalent to the corresponding notion defined via stochastic processes.

Definition 7 (Expected Derivation Length). Let $\to$ be a PARS and $\vec{\mu} = (\mu_n)_{n \in \mathbb{N}} \in \mathsf{red}_\to(\mu)$. We define the *expected derivation length* $\mathsf{edl}(\vec{\mu}) \in \mathbb{R}_{\geq 0} \cup \{\infty\}$ in this multidistribution reduction sequence $\vec{\mu}$ by

$$\mathsf{edl}(\vec{\mu}) := \sum_{n \geq 1} |\mu_n| \ .$$

Bournez and Garnier [6] introduced the notion of *positive almost sure termination (PAST)*, which can be formulated as follows.

Definition 8 (Positive AST; PAST). A PARS $\to$ is said to be *positively almost surely terminating (PAST)* iff for any reduction $\vec{\mu} \in \mathsf{red}_\to(a)$ starting from any $a \in A$, $\mathsf{edl}(\vec{\mu})$ is finite.

One should be careful to notice that the above definition does not ensure a bound on the expected length *of all runs* starting from a given $a \in A$. This phenomenon is already visible in the non-probabilistic setting.

Example 6. *Consider the ARS $\xrightarrow[\omega]{}$ over $\mathbb{N} \cup \{a\}$ given by*

$$a \xrightarrow[\omega]{} n \qquad\qquad n+1 \xrightarrow[\omega]{} n$$

for all $n \in \mathbb{N}$. Every reduction in $\xrightarrow[\omega]{}$ is finite, thus $\xrightarrow[\omega]{}$ seen as a PARS is PAST. However, there is no upper bound on the length of reductions starting from a.

9

In this example, although $\underset{\omega}{\rightarrow}$ is terminating, it is not *finitely branching*: a admits infinitely many one-step reducts. In the non-probabilistic setting, a finitely branching terminating ARS admits a bound on the length of derivation for each starting element. In the probabilistic setting, however, PAST does not ensure such bounds even if the system is finitely branching:[3]

Example 7. *Consider the PARS $\underset{\infty}{\rightarrow}$ over $\mathbb{N} \cup \{a_n \mid n \in \mathbb{N}\}$ defined by*

$$a_n \underset{\infty}{\rightarrow} \{\!\{\tfrac{1}{2} : a_{n+1},\ \tfrac{1}{2} : 0\}\!\} \qquad a_n \underset{\infty}{\rightarrow} \{\!\{1 : 2^n \cdot n\}\!\} \qquad n+1 \underset{\infty}{\rightarrow} \{\!\{1 : n\}\!\}$$

for all $n \in \mathbb{N}$. Obviously $\underset{\infty}{\rightarrow}$ is finitely branching. We show that $\underset{\infty}{\rightarrow}$ is PAST as follows. Observe that every sequence $\vec{\mu} \in \mathsf{red}_{\underset{\infty}{\rightarrow}}(a_n)$ has one of the following two forms:

1. *The first rule is fired infinitely often:*

   $$\vec{\mu} = \{\!\{1 : a_n\}\!\} \underset{\infty}{\rightarrow} \{\!\{\tfrac{1}{2} : a_{n+1},\ \tfrac{1}{2} : 0\}\!\} \underset{\infty}{\rightarrow} \{\!\{\tfrac{1}{4} : a_{n+2},\ \tfrac{1}{4} : 0\}\!\} \underset{\infty}{\rightarrow} \cdots$$

   In this case, $\mathsf{edl}(\vec{\mu}) = 1 + \tfrac{1}{2} + \tfrac{1}{4} + \cdots = \sum_{n \in \mathbb{N}} \tfrac{1}{2^n} = 2$.
2. *The second rule is fired after m applications of the first:*

   $$\vec{\mu} = \{\!\{1 : a_n\}\!\} \underset{\infty}{\rightarrow}^m \{\!\{\tfrac{1}{2^m} : a_{n+m},\ \tfrac{1}{2^m} : 0\}\!\} \underset{\infty}{\rightarrow} \{\!\{\tfrac{1}{2^m} : k\}\!\} \underset{\infty}{\rightarrow}^k \{\!\{\tfrac{1}{2^m} : 0\}\!\}\ ,$$

   where $k = 2^{n+m} \cdot (n+m)$. In this case, $\mathsf{edl}(\vec{\mu}) = 1 + \tfrac{1}{2} + \cdots + \tfrac{1}{2^m} + \tfrac{1}{2^m} \cdot k \geq m$.

The second class of derivations witnesses that, given a_n, there is no bound $b_n < \infty$ such that $\mathsf{edl}(\vec{\mu}) \leq b_n$ for any $\vec{\mu} \in \mathsf{red}_{\rightarrow}(a_n)$.

Therefore we introduce a stronger notion, which ensures a bound on the expected derivation length for each starting element. The definition is based on a natural extension of *derivation height* [22] from complexity analysis of term rewriting.

Definition 9 (Strong AST; SAST). For a PARS $\rightarrow \subseteq A \times \mathcal{M}(A)$, the *expected derivation height* $\mathsf{edh}_{\rightarrow}(\mu) \in \mathbb{R}_{\geq 0} \cup \{\infty\}$ of $\mu \in \mathcal{M}_{\leq 1}(A)$ is defined by

$$\mathsf{edh}_{\rightarrow}(\mu) := \sup \{\mathsf{edl}(\vec{\nu}) \mid \vec{\nu} \in \mathsf{red}_{\rightarrow}(\mu)\}\ .$$

For $a \in A$, we write $\mathsf{edh}_{\rightarrow}(a)$ for $\mathsf{edh}_{\rightarrow}(\{1 : a\})$. We say $\rightarrow$ is *strongly almost surely terminating* (*SAST*) iff $\mathsf{edh}_{\rightarrow}(a)$ of every $a \in A$ is finite.

Notice that on the class of *deterministic PARSs*, i.e., when $a \rightarrow \mu_1$ and $a \rightarrow \mu_2$ implies $\mu_1 = \mu_2$, SAST coincides with PAST. In the nondeterministic case, on the other hand, the notions of PAST and SAST do not coincide. SAST guarantees that the expected length of reductions is bounded, taking a demonic view on nondeterminism. As highlighted in Example 7 such a bound does not necessarily exist for systems that are PAST. This holds even for non-probabilistic ARSs, as Example 6 demonstrates.

---

[3]We are grateful to the anonymous reviewer of an earlier version of this work [3] who pointed us to this example.

Proposition 2. *For every PARS $\rightarrow$, the following implications hold:*

$$\rightarrow \text{ is SAST} \implies \rightarrow \text{ is PAST} \implies \rightarrow \text{ is AST}$$

We conclude the section with the following natural property of expected derivation height. For a (multi)distribution μ over real numbers, the *expected value* of μ is defined by $\mathbb{E}(\mu) := \sum_{p:a\in\mu} p \cdot a$. As a function $f : A \rightarrow \mathbb{R}$ is naturally generalized to $f : \mathcal{M}_{\leq 1}(A) \rightarrow \mathcal{M}_{\leq 1}(\mathbb{R})$, for $\mu \in \mathcal{M}_{\leq 1}(A)$ we have $\mathbb{E}(f(\mu)) = \sum_{p:a\in\mu} p \cdot f(a)$.

Lemma 4. *For every PARS $\rightarrow$, $\mathsf{edh}_\rightarrow(\mu) = \mathbb{E}(\mathsf{edh}_\rightarrow(\mu))$.*

Proof. More generally, we have

$$\mathsf{edh}_\rightarrow\left(\biguplus_{i\in I} p_i \cdot \mu_i\right) = \sup\left\{ \sum_{n\in\mathbb{N}} |\nu_n| \,\Big|\, (\nu_n)_{n\in\mathbb{N}} \in \mathsf{red}_\rightarrow\left(\biguplus_{i\in I} p_i \cdot \mu_i\right)\right\}$$

$$(\text{Lemma 3})$$

$$= \sup\left\{ \sum_{n\in\mathbb{N}} \left|\biguplus_{i\in I} p_i \cdot \nu_{n,i}\right| \,\Big|\, \forall i \in I.\ (\nu_{n,i})_{n\in\mathbb{N}} \in \mathsf{red}_\rightarrow(\mu_i)\right\}$$

$$(\text{Lemma 2})$$

$$= \sup\left\{ \sum_{i\in I} p_i \cdot \sum_{n\in\mathbb{N}} |\nu_{n,i}| \,\Big|\, \forall i \in I.\ (\nu_{n,i})_{n\in\mathbb{N}} \in \mathsf{red}_\rightarrow(\mu_i)\right\}$$

$$= \sum_{i\in I} p_i \cdot \sup\left\{ \sum_{n\in\mathbb{N}} |\nu_n| \,\Big|\, (\nu_n)_{n\in\mathbb{N}} \in \mathsf{red}_\rightarrow(\mu_i)\right\}$$

$$= \sum_{i\in I} p_i \cdot \mathsf{edh}_\rightarrow(\mu_i)\,.$$

The main claim then follows, since

$$\mu = \biguplus_{p:a\in\mu} p \cdot \{\!\{1 : a\}\!\} \quad \text{and} \quad \sum_{p:a\in\mu} p \cdot \mathsf{edh}_\rightarrow(\{\!\{1 : a\}\!\}) = \mathbb{E}(\mathsf{edh}_\rightarrow(\mu))\,. \qquad \square$$

3.3. Proving Probabilistic Termination

A popular way of proving termination of non-probabilistic systems is via *embedding* reductions into a well-founded set $(B, >)$ [29]. The latter can be seen also as a terminating ARS, giving rise to the following general statement.

Proposition 3. *An ARS $\rightarrow\ \subseteq A \times A$ is terminating if and only if there exists a mapping $f : A \rightarrow B$ and a terminating ARS $\succ\ \subseteq B \times B$ such that $a \rightarrow b$ implies $f(a) \succ f(b)$.*

The function f is often called a *ranking function*, typically B is $\mathbb{N}$, and $\succ$ is the standard order on $\mathbb{N}$. Notice that the "only if" direction is trivial by taking the identity as f and $\rightarrow$ as $\succ$. We give an analog of Proposition 3 for probabilistic systems.

11

Definition 10 (Embedding). Let $\to\; \subseteq A \times \mathcal{M}(A)$ and $\sqsupset\; \subseteq B \times \mathcal{M}(B)$ be two PARSs. We say that a mapping $f : A \to B$ is an *embedding of $\to$ into $\sqsupset$*, if $a \to \mu$ implies $f(a) \sqsupset f(\mu)$.

With Theorem 1 we will show that such embeddings give a sound and complete method for proving SAST. The proof of this theorem is based on the following simulation result.

Lemma 5. *Let $f : A \to B$ be an embedding of a PARS $\to$ into a PARS $\sqsupset$. If $\mu \to \nu$ then $f(\mu) \sqsupset f(\nu) \uplus \xi$ for some multidistribution ξ.*

Proof. We proceed by induction on the derivation of $\mu \to \nu$.

- If $\mu = \{\!\{1 : a\}\!\} \to \varnothing = \nu$ with $a \in \mathsf{NF}_\to$, then the claim is satisfied by any ξ with $f(\mu) \sqsupset \xi$. Note that by the definition of $\to^{\mathcal{M}}$, such a ξ always exists.

- If $\mu = \{\!\{1 : a\}\!\} \to \nu$ with $a \to \nu$, then the assumption gives $f(a) \sqsupset f(\nu)$ and thus $f(\mu) = \{\!\{1 : f(a)\}\!\} \sqsupset f(\nu)$.

- Finally, if $\mu = \biguplus_{i \in I} p_i \cdot \mu_i \to \biguplus_{i \in I} p_i \cdot \nu_i = \nu$ with $\mu_i \to \nu_i$, then the claim is an easy consequence of the induction hypothesis. $\square$

Lemma 6. *Let $f : A \to B$ be an embedding of a PARS $\to$ into a PARS $\sqsupset$. For every reduction sequence $(\mu_n)_{n \in \mathbb{N}} \in \mathsf{red}_\to(\mu)$ there exists a reduction sequence $(\nu_n)_{n \in \mathbb{N}} \in \mathsf{red}_\sqsupset(f(\mu))$ such that $f(\mu_n) \subseteq \nu_n$ for all $n \in \mathbb{N}$. In particular, $\mathsf{edl}\big((\mu_n)_{n \in \mathbb{N}}\big) \leq \mathsf{edl}\big((\nu_n)_{n \in \mathbb{N}}\big)$.*

Proof. Given $(\mu_n)_{n \in \mathbb{N}} \in \mathsf{red}_\to(\mu)$, we inductively construct the desired $\nu_0, \nu_1, \nu_2, \dots$ as follows. We set $\nu_0 = f(\mu_0)$. Suppose $f(\mu_n) \subseteq \nu_n$, and thus $\nu_n = f(\mu_n) \uplus \xi$ for some multidistribution ξ. Lemma 5 yields $f(\mu_n) \sqsupset f(\mu_{n+1}) \uplus \rho$ for some ρ. Take an arbitrary ξ' with $\xi \sqsupset \xi'$ and define $\nu_{n+1} = f(\mu_{n+1}) \uplus \rho \uplus \xi'$. We have

$$\nu_n \;=\; f(\mu_n) \uplus \xi \;\sqsupset\; f(\mu_{n+1}) \uplus \rho \uplus \xi' \;=\; \nu_{n+1}$$

as desired. From this, $\mathsf{edl}\big((\mu_n)_{n \in \mathbb{N}}\big) \leq \mathsf{edl}\big((\nu_n)_{n \in \mathbb{N}}\big)$ follows as we have $|\mu_n| \leq |\nu_n|$ for every $n \in \mathbb{N}$. $\square$

Theorem 1. *A PARS $\to\; \subseteq A \times \mathcal{M}(A)$ is SAST if and only if there exists an embedding $f : A \to B$ into a PARS $\sqsupset\; \subseteq B \times \mathcal{M}(B)$ which is SAST. Moreover, in this case $\mathsf{edh}_\to(a) \leq \mathsf{edh}_\sqsupset(f(a))$ for any $a \in A$.*

Proof. The "only if" direction is trivial by taking $\to$ as $\sqsupset$ and the identity as f. Concerning the "if" direction, suppose that f is an embedding of $\to$ into $\sqsupset$ and $\sqsupset$ is SAST. Consider an arbitrary $a \in A$. By Lemma 6, for every $\vec{\mu} \in \mathsf{red}_\to(a)$ there is a "longer" reduction $\vec{\nu} \in \mathsf{red}_\sqsupset(f(a))$ in the sense $\mathsf{edl}(\vec{\mu}) \leq \mathsf{edl}(\vec{\nu})$. Consequently,

$$\begin{aligned}
\mathsf{edh}_\to(a) &= \sup\{\mathsf{edl}(\vec{\mu}) \mid \vec{\mu} \in \mathsf{red}_\to(a)\} \\
&\leq \sup\{\mathsf{edl}(\vec{\nu}) \mid \vec{\nu} \in \mathsf{red}_\sqsupset(f(a))\} \\
&= \mathsf{edh}_\sqsupset(f(a)) < \infty \;.
\end{aligned}$$
$\square$

Probabilistic ranking functions [6] or *ranking supermartingales* [9, 14] are instances of Theorem 1. To demonstrate this, we introduce the following PARS over $\mathbb{R}_{\geq 0}$, which is *canonically* SAST.

Definition 11 (Probabilistic Ranking Function). Let $\epsilon > 0$. We define the ARS $[\geq \epsilon +] \subseteq \mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0}$ by

$$a \; [\geq \epsilon +] \; b :\Longleftrightarrow \; a \geq \epsilon + b \,,$$

and the PARS $[\geq \epsilon + \mathbb{E}] \subseteq \mathbb{R}_{\geq 0} \times \mathcal{M}(\mathbb{R}_{\geq 0})$ by

$$a \; [\geq \epsilon + \mathbb{E}] \; \mu :\Longleftrightarrow \; a \geq \epsilon + \mathbb{E}(\mu) \,.$$

We call an embedding $f : A \to \mathbb{R}_{\geq 0}$ of a PARS $\to \; \subseteq A \times \mathcal{M}(A)$ into $[\geq \epsilon + \mathbb{E}]$ a *probabilistic ranking function* for $\to$.

Intuitively, a mapping f is a probabilistic ranking function for $\to$ if the value of f decreases by ϵ in expectation whenever there is a reduction in $\to$. Lemma 9 will confirm that $[\geq \epsilon + \mathbb{E}]$ is SAST, and hence, probabilistic ranking functions are sound for proving SAST, by Theorem 1. Moreover, we will also see that they are complete. Towards the soundness result, we analyze reductions of $[\geq \epsilon + \mathbb{E}]$, in particular, how they evolve in expectation. We start with the analysis of a single reduction.

Lemma 7. *If $\mu \; [\geq \epsilon + \mathbb{E}]^{\mathcal{M}} \; \nu$, then $\mathbb{E}(\mu) \geq \epsilon \cdot |\nu| + \mathbb{E}(\nu)$.*

Proof. We prove the claim by induction on the derivation of $\mu \; [\geq \epsilon + \mathbb{E}]^{\mathcal{M}} \; \nu$.

- Suppose $\mu = \{\!\{1 : a\}\!\}$ and $a \in \mathsf{NF}_{[\geq \epsilon + \mathbb{E}]}$, that is, $a < \epsilon$. Then $\nu = \varnothing$ and $\mathbb{E}(\mu) \geq 0 = \epsilon \cdot |\nu| + \mathbb{E}(\nu)$ since $\mathbb{E}(\varnothing) = |\varnothing| = 0$.

- Suppose $\mu = \{\!\{1 : a\}\!\}$ and $a \; [\geq \epsilon + \mathbb{E}] \; \nu$, that is, $a \geq \epsilon + \mathbb{E}(\nu)$. As $|\nu| = 1$ we have $\mathbb{E}(\mu) = a \geq \epsilon \cdot |\nu| + \mathbb{E}(\nu)$.

- Suppose $\mu = \biguplus_{i \in I} p_i \cdot \mu_i$, $\nu = \biguplus_{i \in I} p_i \cdot \nu_i$, and $\mu_i \; [\geq \epsilon + \mathbb{E}]^{\mathcal{M}} \; \nu_i$ for every $i \in I$. The induction hypothesis gives $\mathbb{E}(\mu_i) \geq \epsilon \cdot |\nu_i| + \mathbb{E}(\nu_i)$. Thus,

$$\mathbb{E}(\mu) = \sum_{i \in I} p_i \cdot \mathbb{E}(\mu_i) \geq \sum_{i \in I} p_i \cdot (\epsilon \cdot |\nu_i| + \mathbb{E}(\nu_i))$$
$$= \epsilon \cdot \sum_{i \in I} p_i \cdot |\nu_i| + \sum_{i \in I} p_i \cdot \mathbb{E}(\nu_i) = \epsilon \cdot |\nu| + \mathbb{E}(\nu) \,. \qquad \square$$

Lemma 8. *For every $\vec{\mu} = (\mu_n)_{n \in \mathbb{N}} \in \mathsf{red}_{[\geq \epsilon + \mathbb{E}]}(\mu_0)$, $\mathbb{E}(\mu_0) \geq \epsilon \cdot \mathsf{edl}(\vec{\mu})$.*

Proof. We first show $\mathbb{E}(\mu_m) \geq \epsilon \cdot \sum_{i=m+1}^{n} |\mu_i|$ for every $n \geq m$, by induction on $n - m$. The base case is trivial, so let us consider the inductive step. By Lemma 7 and the induction hypothesis we get

$$\mathbb{E}(\mu_m) \geq \epsilon \cdot |\mu_{m+1}| + \mathbb{E}(\mu_{m+1})$$
$$\geq \epsilon \cdot |\mu_{m+1}| + \epsilon \cdot \sum_{i=m+2}^{n} |\mu_i| = \epsilon \cdot \sum_{i=m+1}^{n} |\mu_i| \,.$$

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 $\mathbb{E}(\mu_0)$, and so is its limit $\epsilon \cdot \sum_{i \geq 1} |\mu_i| = \epsilon \cdot \mathsf{edl}(\vec{\mu})$. $\qquad\square$

Lemma 9. *If $\epsilon > 0$ then $\mathsf{edh}_{[\geq \epsilon + \mathbb{E}]}(a) \leq \frac{a}{\epsilon}$ for every $a \in \mathbb{R}_{\geq 0}$, and thus the PARS $[\geq \epsilon + \mathbb{E}]$ is SAST.*

Proof. By Lemma 8, we have $\mathsf{edl}(\vec{\mu}) \leq \frac{\mathbb{E}(\{1:a\})}{\epsilon} = \frac{a}{\epsilon}$ for every $\vec{\mu} \in \mathsf{red}_{[\geq \epsilon + \mathbb{E}]}(a)$. Hence, $\mathsf{edh}_{[\geq \epsilon + \mathbb{E}]}(a) \leq \frac{a}{\epsilon}$, concluding that $[\geq \epsilon + \mathbb{E}]$ is SAST. $\qquad\square$

Theorem 2. *A PARS $\to$ is SAST if and only if there is a probabilistic ranking function f for $\to$, i.e., f embeds $\to$ into $[\geq \epsilon + \mathbb{E}]$ for some $\epsilon > 0$. Moreover, in this case $\mathsf{edh}_{\to}(a) \leq \frac{f(a)}{\epsilon}$ for any $a \in A$.*

Proof. The "if" direction follows from Theorem 1 and Lemma 9. Concerning the "only if" direction, suppose that $\to$ is SAST. Thus, we have $\mathsf{edh}_{\to} : A \to \mathbb{R}_{\geq 0}$. Consider $a \to \mu$. Then we have $|\mu| = 1$ and

$$
\begin{aligned}
\mathsf{edh}_{\to}(a) \ &= \sup\left\{\mathsf{edl}(\vec{\mu}) \mid \vec{\mu} \in \mathsf{red}_{\to}(a)\right\} \\
&\geq \sup\left\{|\mu| + \mathsf{edl}(\vec{\mu}) \mid \vec{\mu} \in \mathsf{red}_{\to}(\mu)\right\} \\
&= |\mu| + \sup\left\{\mathsf{edl}(\vec{\mu}) \mid \vec{\mu} \in \mathsf{red}_{\to}(\mu)\right\} \\
&= |\mu| + \mathsf{edh}_{\to}(\mu) & \text{(Lemma 4)} \\
&= |\mu| + \mathbb{E}(\mathsf{edh}_{\to}(\mu)) \,,
\end{aligned}
$$

concluding $\mathsf{edh}_{\to}(a) \, [\geq 1 + \mathbb{E}] \, \mathsf{edh}_{\to}(\mu)$. Thus $\mathsf{edh}_{\to}$ is an embedding of $\to$ into $[\geq 1 + \mathbb{E}]$, i.e., a ranking function. $\qquad\square$

3.4. Relation to Formulation by Bournez and Garnier

The dynamics of probabilistic systems are commonly defined as stochastic processes, so that the nth random variable represents the nth reduct. Bournez and Garnier [6] follow this approach. In this section, we establish a precise correspondence between their formulation and ours. In particular, we show that the corresponding notions of AST and PAST coincide.

We assume familiarity with *stochastic processes*, see e.g. [34]. We briefly fix central notions and notations. A *measurable space* is a tuple (Ω, Σ) consisting of a set Ω and a *sigma-algebra* Σ on it, i.e., Σ is a collection of subsets of Ω that contains the empty set and is closed under complement and countable unions. A *probability space* is a triple $(\Omega, \Sigma, \mathbb{P})$ with (Ω, Σ) a measurable space and $\mathbb{P} : \Sigma \to \mathbb{R}_{\geq 0}$ a countable additive function such that $\mathbb{P}(\Omega) = 1$. Given $A, B \in \Sigma$ with $\mathbb{P}(B) > 0$, the *conditional probability* $\mathbb{P}(A|B)$ is defined as $\mathbb{P}(A \cap B)/\mathbb{P}(B)$. The law of total probability states that $\mathbb{P}(A) = \sum_i \mathbb{P}(A \cap B_i)$ for any $A \in \Sigma$ and finite or countable partition $(B_i)_{i \in \mathbb{N}} \in \Sigma$ of Ω.

A *random variable* X over a countable set A in a probability space $(\Omega, \Sigma, \mathbb{P})$ is a *measurable function* $X : \Omega \to A$, that is, $X^{-1}(a) := \{\omega \mid X(\omega) = a\} \in \Sigma$ for all $a \in A$. We follow the usual conventions concerning random variables, in particular, with $\mathbb{P}(X = a)$ we denote the probability $\mathbb{P}(X^{-1}(a))$. The *probability distribution* of X is the probability distribution over A that assigns to every

$a \in A$ the probability $\mathbb{P}(X = a)$. The *expected value* of a random variable X over $\mathbb{N} \cup \{\infty\}$ is defined by $\mathbb{E}(X) := \sum_{n \in \mathbb{N} \cup \{\infty\}} \mathbb{P}(X = n) \cdot n$, where $0 \cdot \infty = 0$ and $p \cdot \infty = \infty$ for $p > 0$.

A *stochastic process* on A is an infinite sequence $\vec{X} = (X_n)_{n \in \mathbb{N}}$ of random variables over A all defined on some probability space $(\Omega, \Sigma, \mathbb{P})$. A *stopping time* with respect to $\vec{X}$ is a random variable S, taking values in $\mathbb{N} \cup \{\infty\}$, with the property that for each $n \in \mathbb{N} \cup \{\infty\}$, the occurrence or non-occurrence of the event $S = n$ depends only on the values of $X_0, \dots, X_n$. Every stopping time S satisfies

$$\mathbb{E}(S) := \sum_{n=1}^{\infty} n \cdot \mathbb{P}(S = n) = \sum_{n=1}^{\infty} \mathbb{P}(S \geq n) . \tag{1}$$

An instance of a stopping time is the *first hitting time* with respect to a set $H \subseteq A$, which is defined as $\tau_H(\omega) := \min\{n \mid X_n(\omega) \in H\}$ for all $\omega \in \Omega$, where $\min \varnothing = \infty$.

For the remaining of the section, we fix a PARS $\rightarrow$ on A such that all right-hand sides in $\rightarrow$ are distributions. In order to define reduction sequences of $\rightarrow$ as stochastic processes, first, all nondeterministic choices are resolved by fixing a *strategy*, also called *policy*.

Definition 12 ([6]). An *n-step history* is a nonempty finite sequence $a_{0..n} = (a_0, \dots, a_n) \in A^+$. We say $a_{0..n}$ is *(non)terminal* iff a_n is. A *strategy* is a mapping $\phi : A^+ \rightarrow \mathcal{D}(A)$ that satisfies $a_n \rightarrow \phi(a_{0..n})$ whenever a_n is nonterminal. We say $a_{0..n}$ is *realizable under ϕ* iff $\phi(a_{0..i})(a_{i+1}) > 0$ for every $0 \leq i < n$.

Given an initial element or distribution, a strategy completely defines a stochastic process that corresponds to a particular reduction in the PARS. As [6] does not allow subdistributions, in the following we fix a special symbol $\perp \notin A$ to denote termination.

Definition 13 (Stochastic Reduction, [6]). A *stochastic reduction* under a strategy ϕ is a sequence $\vec{X} = (X_n)_{n \in \mathbb{N}}$ of random variables over $A \uplus \{\perp\}$ such that

$$\mathbb{P}(X_{n+1} = \perp \mid X_n = \perp) = 1 ;$$
$$\mathbb{P}(X_{n+1} = \perp \mid X_n = a) = 1 \qquad \text{if } a \in \mathsf{NF}_{\rightarrow};$$
$$\mathbb{P}(X_{n+1} = \perp \mid X_n = a) = 0 \qquad \text{if } a \notin \mathsf{NF}_{\rightarrow};$$
$$\mathbb{P}(X_{n+1} = a \mid X_n = a_n, \dots, X_0 = a_0) = \phi(a_{0..n})(a) ,$$

whenever $a_{0..n}$ is a nonterminal realizable history under ϕ. The *initial distribution* of $\vec{X}$ is given by the probability distribution of X_0.

As an immediate consequence of the law of total probability, by Definition 13 any stochastic reduction $\vec{X} = (X_n)_{n \in \mathbb{N}}$ satisfies

$$\mathbb{P}(X_n = a_n) = \sum_{a_{0..n} \in A^{n+1}} \mathbb{P}(X_0 = a_0, \dots, X_n = a_n) . \tag{2}$$

Notice that a_i ranges over A in this equality, i.e., $a_i \neq \bot$. As $\bot$ signals termination, the derivation length of a stochastic reduction is given by the *first hitting time* to $\bot$.

Definition 14 (AST and PAST of [6]). For $\vec{X} = (X_n)_{n \in \mathbb{N}}$ define the random variable $T := \min\{n \in \mathbb{N} \mid X_n = \bot\}$ where $\min \varnothing = \infty$ by convention. The random variable T is called the *stopping time* of $\vec{X}$. A PARS $\rightarrow$ is *stochastically AST* if for every stochastic reduction $\vec{X}$ in $\rightarrow$, $\mathbb{P}(T = \infty) = 0$. We say $\rightarrow$ is *stocastically PAST* if for every stochastic reduction $\vec{X}$ in $\rightarrow$ starting from $a \in A$, $\mathbb{E}(T) < \infty$.

We will now see that stochastic (P)AST coincides with (P)AST. To this end, we first clarify the correspondence of stochastic reductions and multidistribution reduction sequences. Given a stochastic reduction $\vec{X} = (X_n)_{n \in \mathbb{N}}$, for each $n \in \mathbb{N}$ we define the random variable $X_{0..n}$ ranging over $(A \cup \{\bot\})^{n+1}$ by

$$\mathbb{P}(X_{0..n} = a_{0..n}) = \mathbb{P}(X_0 = a_0, \dots, X_n = a_n) \, ,$$

where $a_0, \dots, a_n$ range over $A \cup \{\bot\}$. Now consider projecting to the last element in $X_{0..n}$. We define the following multidistribution over A:

$$X_n^{\mathcal{M}} := \{\!\!\{ p : a_n \mid p = \mathbb{P}(X_{0..n} = a_{0..n}) > 0 \}\!\!\} \, .$$

The following two lemmas state the one-to-one correspondence between multidistribution reduction sequences and stochastic reductions in each direction.

Lemma 10 (Stochastic Reductions to Reductions). *Let $\vec{X} = (X_n)_{n \in \mathbb{N}}$ be the stochastic reduction in $\rightarrow$ under strategy ϕ. Then $X_n^{\mathcal{M}} \rightarrow^{\mathcal{M}} X_{n+1}^{\mathcal{M}}$ for every $n \in \mathbb{N}$, i.e., $\left(X_n^{\mathcal{M}} \right)_{n \in \mathbb{N}} \in \mathsf{red}_{\rightarrow}(X_0)$.*

Proof.

$$
\begin{aligned}
X_n^{\mathcal{M}} &= \{\!\!\{ p : a_n \mid p = \mathbb{P}(X_{0..n} = a_{0..n}) > 0 \}\!\!\} \\
&\rightarrow^{\mathcal{M}} \biguplus \{ p \cdot \phi(a_{0..n}) \mid p = \mathbb{P}(X_{0..n} = a_{0..n}) > 0, \ a_n \notin \mathsf{NF}_{\rightarrow} \} \\
&= \{\!\!\{ q : a_{n+1} \mid q = \mathbb{P}(X_{0..n} = a_{0..n}) \cdot \phi(a_{0..n})(a_{n+1}) > 0, \ a_n \notin \mathsf{NF}_{\rightarrow} \}\!\!\} \\
&= \{\!\!\{ q : a_{n+1} \mid q = \mathbb{P}(X_{0..n} = a_{0..n}, X_{n+1} = a_{n+1}) > 0 \}\!\!\} \\
&= \{\!\!\{ q : a_{n+1} \mid q = \mathbb{P}(X_{0..n+1} = a_{0..n+1}) > 0 \}\!\!\} \\
&= X_{n+1}^{\mathcal{M}} \, . \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \square
\end{aligned}
$$

Lemma 11 (Reduction to Stochastic Reductions). *Let $\vec{\mu} = (\mu_n)_{n \in \mathbb{N}} \in \mathsf{red}_{\rightarrow}(d)$ for some distribution d. There exists a strategy ϕ which induces a stochastic reduction $(X_n)_{n \in \mathbb{N}}$ such that $\mu_n = X_n^{\mathcal{M}}$ for all $n \in \mathbb{N}$.*

Proof. First, we inductively define $\phi_n : A^n \rightarrow A \rightarrow \mathbb{R}_{\geq 0}$ such that

$$\mu_n = \{\!\!\{ \phi_n(a_{0..n-1})(a_n) : a_n \mid \phi_n(a_{0..n-1})(a_n) > 0 \}\!\!\} \, . \tag{3}$$

16

Then by defining $\phi(a_{0..n}) = \phi_{n+1}(a_{0..n})$ we obtain the desired strategy. Since $d = \mu_0$ is a distribution, we define $\phi_0()(a) := d(a)$. To obtain ϕ_{n+1} from ϕ_n, suppose by induction hypothesis that (3) holds. Let

$$I := \{a_{0..n} \mid \phi_n(a_{0..n-1})(a_n) > 0, a_n \notin \mathsf{NF}_\rightarrow\} \ .$$

As $\mu_n \rightarrow^{\mathcal{M}} \mu_{n+1}$, for each $a_{0..n} \in I$ there is $a_n \rightarrow d_{a_{0..n}}$ such that

$$\mu_{n+1} = \biguplus_{a_{0..n} \in I} \phi_n(a_{0..n-1})(a_n) \cdot d_{a_{0..n}}$$
$$= \{\!\{\phi_n(a_{0..n-1})(a_n) \cdot d_{a_{0..n}}(a) : a \mid \phi_{n+1}(a_{0..n})(a) > 0\}\!\} \ .$$

Hence, defining

$$\phi_{n+1}(a_{0..n}) := \begin{cases} \phi_n(a_{0..n-1})(a_n) \cdot d_{a_{0..n}} & \text{if } (a_{0..n}) \in I \ , \\ 0 & \text{otherwise,} \end{cases}$$

yields the desired ϕ_{n+1}. $\qquad\square$

Now we relate the expected derivation length of multidistribution reduction sequences and the expected stopping time of stochastic reductions. For a multidistribution $\mu \in \mathcal{M}_{\leq 1}(A)$, we define the distribution $\overline{\mu} \in \mathcal{D}(A \cup \{\bot\})$ by $\overline{\mu}(a) := \sum_{p:a \in \mu} p$ for $a \in A$ and $\overline{\mu}(\bot) := 1 - |\mu|$. We have the following natural correspondence between $X_n^{\mathcal{M}}$ and X_n.

Lemma 12. $\overline{X_n^{\mathcal{M}}}(a) = \mathbb{P}(X_n = a)$.

Proof. The lemma follows from the definition of $\overline{(\cdot)}$ and Equation (2). $\qquad\square$

Lemma 13. Let $\vec{X} = (X_n)_{n \in \mathbb{N}}$ be the stochastic derivation in $\rightarrow$ under strategy ϕ from the initial distribution d and let T denote its stopping time. The following two properties hold:

1. $\mathbb{P}(T \geq n) = \left|X_n^{\mathcal{M}}\right|$ for every $n \in \mathbb{N}$.
2. $\mathbb{P}(T = \infty) = \lim_{n \to \infty} \left|X_n^{\mathcal{M}}\right|$.

Proof. Concerning the first property, we have

$$\mathbb{P}(T \geq n) = \mathbb{P}(X_n \in A) = \sum_{a \in A} \mathbb{P}(X_n = a) = \sum_{a \in A} \overline{X_n^{\mathcal{M}}}(a) = \left|X_n^{\mathcal{M}}\right| \ ,$$

for all $n \in N$, where the penultimate equation follows from Lemma 12. As we have

$$\mathbb{P}(T = \infty) = \lim_{n \to \infty} \mathbb{P}(T \geq n) \ ,$$

the second property follows from the first. $\qquad\square$

Now we arrive at the main theorem of this section.

Theorem 3. *A PARS $\rightarrow$ is (P)AST if and only if it is stochastically (P)AST.*

Proof. We consider the "if" direction first. Suppose $\to$ is AST. Lemma 10 translates an arbitrary stochastic derivation $\vec{X} = (X_n)_{n \in \mathbb{N}}$ in $\to$ to a reduction $\left(X_n^{\mathcal{M}}\right)_{n \in \mathbb{N}} \in \mathsf{red}_\to(a)$, for which we have $\mathbb{P}(T = \infty) = \lim_{n \to \infty} |\mu_n| = 0$ by Lemma 13(2). Hence, $\to$ is stochastically AST. If $\to$ is moreover PAST, using Lemma 13(1) and (1) we get

$$\infty > \mathsf{edl}\left(\left(X_n^{\mathcal{M}}\right)_{n \in \mathbb{N}}\right) = \sum_{n \geq 1} \left|X_n^{\mathcal{M}}\right| = \sum_{n \geq 1} \mathbb{P}(T \geq n) = \mathbb{E}(T) \ ,$$

where we tacitly employ $\mathbb{P}(T = \infty) = 0$. The "only if" direction is proven dually, using Lemma 11. $\qquad\square$

Finally, we relate our probabilistic ranking functions (Theorem 2) to the following formulation by Bournez and Garnier [6].

Proposition 4 ([6]). *A PARS $\to$ over A is PAST if there exists a function $V : A \to \mathbb{R}$, with $\inf_{a \in A} V(a) > -\infty$, and $\epsilon > 0$ such that for all $a \in A$, for all μ with $a \to \mu$, the drift in a according to μ defined by*

$$\Delta_\mu V(a) := \sum_{p:b \in \mu} p \cdot V(b) - V(a) \ ,$$

satisfies

$$\Delta_\mu V(a) \leq -\epsilon \ . \tag{4}$$

Formulated differently, condition (4) states that $V(a) \geq \epsilon + \mathbb{E}(V(\mu))$ holds whenever $a \to \mu$. Further, $\inf_{a \in A} V(a) > -\infty$ means that $f(a) := V(a) - \inf_{a \in A} V(a)$ gives a mapping $f : A \to \mathbb{R}_{\geq 0}$. Hence, f is a probabilistic ranking function in the sense of Theorem 2.

Proposition 5. *Let $\to$ be a PARS. There exists $V : A \to \mathbb{R}$ satisfying Proposition 4 if and only if there is a probabilistic ranking function for $\to$.*

As a consequence, Proposition 4 can be strengthened from PAST to SAST. Moreover, the method is not only sound but also *complete* for SAST. Bournez and Garnier claimed with [6, Theorem 3] that Proposition 4 is complete for proving finitely branching PARSs to be PAST. Example 7 is a counterexample to this claim, depicting a finitely branching system that is PAST but not SAST.

4. Probabilistic Term Rewriting

We now formulate probabilistic term rewriting, and then lift the interpretation method for term rewriting to the probabilistic case.

We briefly recap notions from term rewriting; see [5] for more details. A *signature F* is a set of *function symbols*, each associated with a natural number called *arity*. 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 $\mathtt{f}(t_1, \ldots, t_n) \in T(F, V)$ whenever $\mathtt{f} \in F$ is of arity n and $t_1, \ldots, t_n \in T(F, V)$. 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 over terms as follows: $\mu\sigma := \{\!\{p_i : t_i\sigma \mid i \in I\}\!\}$ and $C[\mu] := \{\!\{p_i : C[t_i] \mid i \in I\}\!\}$ for $\mu = \{\!\{p_i : t_i \mid i \in I\}\!\}$.

Definition 15 (Probabilistic Term Rewriting). A *probabilistic rewrite rule* is a pair of $l \in T(F, V)$ and $\mu \in \mathcal{M}(T(F, V))$, written $l \to \mu$. A *probabilistic term rewrite system (PTRS)* $\mathcal{R}$ is a (typically finite) set of probabilistic rewrite rules. We write $\underset{\mathcal{R}}{\to}$ for the least PARS such that $C[l\sigma] \underset{\mathcal{R}}{\to} C[\mu\sigma]$ for every probabilistic rewrite rule $l \to \mu \in \mathcal{R}$, context C, and substitution σ. We say a PTRS $\mathcal{R}$ is AST/PAST/SAST if $\underset{\mathcal{R}}{\to}$ is.

Note here that we deviate from [6]: there $\underset{\mathcal{R}}{\to}$ is defined by $C[l\sigma] \underset{\mathcal{R}}{\to} C[\overline{\mu\sigma}]$ for each $l \to \mu \in \mathcal{R}$ in our notation, where $\overline{\mu\sigma}$ is the distribution corresponding to the multidistribution $\mu\sigma$. Notice that even if μ is a distribution over terms, $\mu\sigma$ is in general a (proper) multidistribution; e.g., consider $\{\frac{1}{2} : x, \; \frac{1}{2} : y\}\sigma$ with $x\sigma = y\sigma$. On the other hand, Bournez and Garnier [6] demand that right-hand sides of PARSs to be distributions. Consequently, the extra operation $\overline{\cdot}$ is necessary, in order for $\underset{\mathcal{R}}{\to}$ to be a PARS of their definition. One consequence of this difference is that we allow strategies to know which rule was applied even if the resulting term was identical, while in the setting of Bournez and Garnier [6] strategies can be defined only in terms of the resulting term.

Example 8. *The random walk of Example 1 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, $\mathtt{s}(\mathtt{f}(\mathtt{s}(0)))$ has two redexes in the above PTRS, and consequently two possible reducts:

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

4.1. Interpretation Methods for Proving SAST

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

Definition 16 (F-Algebra, cf. [38]). An F-*algebra* $\mathcal{X}$ on a non-empty carrier set X specifies the *interpretation* $\mathtt{f}_{\mathcal{X}} : X^n \to X$ of each function symbol $\mathtt{f} \in F$ of arity n. We say $\mathcal{X}$ is *monotone* with respect to a binary relation (an ARS) $\succ \; \subseteq X \times X$ if $x \succ y$ implies $\mathtt{f}_{\mathcal{X}}(\ldots, x, \ldots) \succ \mathtt{f}_{\mathcal{X}}(\ldots, y, \ldots)$ for every $\mathtt{f} \in F$. Given an *assignment* $\alpha : V \to X$, the interpretation of a term is defined as follows:

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

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

19

Proposition 6 (cf. [38]). *A TRS $\mathcal{R}$ is terminating if and only if there exists an F-algebra $\mathcal{X}$ which is monotone with respect to a terminating relation $\succ$ and satisfies $\mathcal{R} \subseteq \succ_\mathcal{X}$.*

In essence, this proposition is a consequence of Proposition 3: the imposed conditions witness that the interpretation embeds the ARS $\xrightarrow{\mathcal{R}}$ underlying $\mathcal{R}$ into the well-founded order $\succ$. Conversely, completeness can be proven by taking $\xrightarrow{\mathcal{R}}$ as the well-founded order and for $\mathcal{X}$ the *term algebra* $\mathcal{T}$, an F-algebra on terms such that $\mathtt{f}_\mathcal{T}(t_1, \ldots, t_n) := \mathtt{f}(t_1, \ldots, t_n)$. Note that in the term algebra, assignments are substitutions, and $[\![t]\!]^\sigma_\mathcal{T} = t\sigma$.

We now generalize Proposition 6 for probabilistic TRSs via Theorem 1. For an F-algebra $\mathcal{X}$, we lift the interpretation of terms to multidistributions as follows:

$$[\![\{\!\{p_i : t_i \mid i \in I\}\!\}]\!]^\alpha_\mathcal{X} := \{\!\{p_i : [\![t_i]\!]^\alpha_\mathcal{X} \mid i \in I\}\!\} .$$

Definition 17 (Probabilistic Monotone F-Algebra). A *probabilistic monotone F-algebra* $(\mathcal{X}, \sqsupset)$ is an F-algebra $\mathcal{X}$ equipped with a relation $\sqsupset \subseteq X \times \mathcal{M}(X)$, such that $x \sqsupset \mu$ implies $\mathtt{f}_\mathcal{X}(\ldots, x, \ldots) \sqsupset \mathtt{f}_\mathcal{X}(\ldots, \mu, \ldots)$ for every $\mathtt{f} \in F$, where

$$\mathtt{f}_\mathcal{X}(\ldots, \mu, \ldots) := \{\!\{p : \mathtt{f}_\mathcal{X}(\ldots, t, \ldots) \mid p : t \in \mu\}\!\} .$$

We write $t \sqsupset_\mathcal{X} \mu$ iff $[\![t]\!]^\alpha_\mathcal{X} \sqsupset [\![\mu]\!]^\alpha_\mathcal{X}$ for every assignment $\alpha : V \to X$.

Notice that $\sqsupset$ is a PARS over $\mathcal{X}$. As in the non-probabilistic case, it is closed under substitutions and contexts:

Lemma 14. *Let $(\mathcal{X}, \sqsupset)$ be a probabilistic monotone F-algebra. If $s \sqsupset_\mathcal{X} \mu$ then $[\![s\sigma]\!]^\alpha_\mathcal{X} \sqsupset [\![\mu\sigma]\!]^\alpha_\mathcal{X}$ and $[\![C[s]]\!]^\alpha_\mathcal{X} \sqsupset [\![C[\mu]]\!]^\alpha_\mathcal{X}$ for arbitrary α, σ, and C.*

Proof. Let $\mu = \{\!\{p_i : t_i \mid i \in I\}\!\}$. Concerning the first property, define the assignment β by $\beta(x) = [\![x\sigma]\!]^\alpha_\mathcal{X}$ for every $x \in V$. By structural induction on t, one can verify $[\![t]\!]^\beta_\mathcal{X} = [\![t\sigma]\!]^\alpha_\mathcal{X}$ for any term t. Thus, from the assumption we get

$$[\![s\sigma]\!]^\alpha_\mathcal{X} = [\![s]\!]^\beta_\mathcal{X} \sqsupset [\![\mu]\!]^\beta_\mathcal{X} = \{\!\{p_i : [\![t_i]\!]^\beta_\mathcal{X} \mid i \in I\}\!\}$$
$$= \{\!\{p_i : [\![t_i\sigma]\!]^\alpha_\mathcal{X} \mid i \in I\}\!\} = [\![\mu\sigma]\!]^\alpha_\mathcal{X} .$$

The second property is proven by induction on C, where the base case follows directly from the assumption, and the inductive step from monotonicity. $\square$

Theorem 4. *A PTRS $\mathcal{R}$ is SAST if and only if there exists a probabilistic monotone F-algebra $(\mathcal{X}, \sqsupset)$ such that $\sqsupset$ is SAST and $\mathcal{R} \subseteq \sqsupset_\mathcal{X}$.*

Proof. For the "if" direction, let $\alpha : V \to X$ be an arbitrary assignment, which exists as X is non-empty. We show that $[\![\cdot]\!]^\alpha_\mathcal{X}$ is an embedding of $\xrightarrow{\mathcal{R}}$ into $\sqsupset$, so that the claim follows from Theorem 1. Consider $s \xrightarrow{\mathcal{R}} \mu$. Then we have $s = C[l\sigma]$ and $\mu = C[\nu\sigma]$ for some σ, C, and $l \to \nu \in \mathcal{R}$. By assumption we have $l \sqsupset_\mathcal{X} \nu$, and thus $[\![s]\!]^\alpha_\mathcal{X} \sqsupset [\![\mu]\!]^\alpha_\mathcal{X}$ by Lemma 14.

For the "only if" direction, it suffices to show that $(\mathcal{T}, \xrightarrow{\mathcal{R}})$ forms a probabilistic monotone F-algebra, which easily follows from Lemma 14. $\square$

4.2. Barycentric Algebras

As probabilistic F-algebras are defined so generally, it is not yet clear how to search for ones for proving that a given PARS is SAST. Now we make one step towards finding probabilistic algebras, by imposing some conditions on (non-probabilistic) F-algebras, so that the relation $\sqsupset$ can be defined from orderings which we are more familiar with. The following generalizes the PARS $[\geq \epsilon + \mathbb{E}]$ from Definition 11.

Definition 18 (Barycentric Domain). A *barycentric domain* is a set X where the *barycentric operation* $\mathbb{E} : \mathcal{M}(X) \to X$ is defined. From a binary relation $\succ$ on X we define the PARS $[\succ \mathbb{E}] \subseteq X \times \mathcal{M}(X)$ by

$$x \ [\succ \mathbb{E}] \ \mu :\Longleftrightarrow \ x \succ \mathbb{E}(\mu) \ .$$

Of particular interest in this work will be the barycentric domains $\mathbb{R}_{\geq 0}$ and $\mathbb{R}_{\geq 0}^m$ with barycentric operations $\mathbb{E}(\{\!\{p_i : a_i \mid i \in I\}\!\}) = \sum_{i \in I} p_i \cdot a_i$. We may write $\mathbb{E}_X$ when we would like to clarify the domain X. The following generalizes standard notions from mathematics.

Definition 19 (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 $f(\mathbb{E}_X(\mu)) \succcurlyeq \mathbb{E}_Y(f(\mu))$ where $\succcurlyeq$ is the reflexive closure of $\succ$. We say f is *affine* if it satisfies $f(\mathbb{E}_X(\mu)) = \mathbb{E}_Y(f(\mu))$.

Clearly, every affine function is concave.

Definition 20 (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 $\succ$ in every argument.

Note that the following theorem claims soundness but not completeness, in contrast to Theorem 4.

Theorem 5. *A PTRS $\mathcal{R}$ is SAST if there exists a barycentric F-algebra $(\mathcal{X}, \succ)$ such that $\mathcal{R} \subseteq [\succ \mathbb{E}]_{\mathcal{X}}$ and $[\succ \mathbb{E}]$ is SAST.*

Proof. Due to Theorem 4, it suffices to show that $(\mathcal{X}, [\succ \mathbb{E}])$ is a probabilistic monotone F-algebra. To this end, suppose that $x \succ \mathbb{E}_X(\mu)$, and let $\mathtt{f} \in F$. Since $\mathtt{f}_{\mathcal{X}}$ is monotone and concave with respect to $\succ$, we have

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

Extending a notion from Hirokawa and Moser [21], we say a relation $\succ \subseteq X \times X$ on a barycentric domain is *collapsible* if there is a concave function $f : X \to \mathbb{R}_{\geq 0}$ that embeds $\succ$ into $[\geq \epsilon +]$.

Lemma 15. *If $\succ \subseteq X \times X$ is collapsible then $[\succ \mathbb{E}]$ is SAST.*

Proof. Suppose $x \, [\succ \mathbb{E}] \, \mu$, i.e., $x \succ \mathbb{E}_X(\mu)$. This implies

$$f(x) \geq \epsilon + f(\mathbb{E}_X(\mu)) \geq \epsilon + \mathbb{E}(f(\mu)) \, ,$$

for some $f : X \to \mathbb{R}_{\geq 0}$. Thus $f(x) \, [\geq \epsilon + \mathbb{E}] \, f(\mu)$, i.e., f is a probabilistic ranking function, and hence the lemma follows from Theorem 2. $\qquad\square$

The following is thus an immediate consequence of Theorem 1.

Corollary 1. *A PTRS $\mathcal{R}$ is SAST if there exists a barycentric F-algebra $(\mathcal{X}, \succ)$ such that $\mathcal{R} \subseteq [\succ \mathbb{E}]_{\mathcal{X}}$ for a collapsible order $\succ$.*

In the rest of this section we recast two popular interpretation methods, polynomial and matrix interpretations (over the reals), as barycentric F-algebras.

4.3. Polynomial Interpretations

Polynomial interpretations were introduced (on natural numbers [28] and real numbers [30]) for the termination analysis of non-probabilistic rewrite systems. Various techniques for synthesizing polynomial interpretations (e.g., [16]) exist, and these techniques are easily applicable in our setting.

Definition 21 (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 $\mathtt{c}_V \in \mathbb{R}_{\geq 0}$:

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

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

Proposition 7. *A PTRS $\mathcal{R}$ is SAST if there exist $\epsilon > 0$ and a monotone multilinear polynomial interpretation $\mathcal{X}$ such that $\mathcal{R} \subseteq [\geq \epsilon + \mathbb{E}]_{\mathcal{X}}$.*

Proof. We show that $(\mathcal{X}, [\geq \epsilon +])$ forms a collapsible barycentric F-algebra, and thus Theorem 5 shows that $\mathcal{R}$ is SAST. Collapsibility is trivial with $\mathsf{G}(x) = x$ and monotonicity is by assumption. Further, every multilinear polynomial is affine and thus concave in all variables. $\qquad\square$

An observation by Lucas [30] 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 $l \, [> \mathbb{E}]_{\mathcal{X}} \, \mu$ for all rules $l \to \mu \in \mathcal{R}$. Define $\epsilon_{l \to \mu} := [\![l]\!]_{\mathcal{X}}^{\alpha} - \mathbb{E}([\![\mu]\!]_{\mathcal{X}}^{\alpha})$ for such α that $\alpha(x) = 0$. Then for any other α, we can show $[\![l]\!]_{\mathcal{X}}^{\alpha} - \mathbb{E}([\![\mu]\!]_{\mathcal{X}}^{\alpha}) \geq \epsilon_{l \to \mu} > 0$. As $\mathcal{R}$ is finite, we take $\epsilon := \min\{\epsilon_{l \to \mu} \mid l \to \mu \in \mathcal{R}\} > 0$.

Corollary 2. *A finite PTRS $\mathcal{R}$ is SAST if there exists a monotone multilinear polynomial interpretation $\mathcal{X}$ such that $\mathcal{R} \subseteq [> \mathbb{E}]_{\mathcal{X}}$.*

Example 9 (Example 8 Revisited). *Consider again the PTRS consisting of the single rule $\mathtt{s}(x) \to \{p\mathord{:}x, \; 1-p\mathord{:}\mathtt{s}(\mathtt{s}(x))\}$. Define the polynomial interpretation $\mathcal{X}$ by $0_{\mathcal{X}} := 0$ 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 Corollary 2.

Corollary 2 constitutes a generalization of [6, Theorem 5]. The latter does not cover linear interpretations, since *context decrease* [6, Definition 8] demands e.g. $[\![\mathtt{f}(t)]\!]^{\alpha}_{\mathcal{X}} - [\![\mathtt{f}(t')]\!]^{\alpha}_{\mathcal{X}} \leq [\![t]\!]^{\alpha}_{\mathcal{X}} - [\![t']\!]^{\alpha}_{\mathcal{X}}$ and thus excludes interpretations such as $\mathtt{f}_{\mathcal{X}}(x) = 2x$.

4.4. Matrix Interpretations

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

Definition 22 (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

$$\mathtt{f}_{\mathcal{X}}(\vec{x}_1, \ldots, \vec{x}_n) = \sum_{i=1}^{n} C_i \cdot \vec{x}_i + \vec{c} \,, \tag{5}$$

where $\vec{c} \in \mathbb{R}^{m}_{\geq 0}$, 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

$$\begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix} \gg_{\epsilon} \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} :\Longleftrightarrow x_1 \geq \epsilon + y_1 \text{ and } x_i \geq y_i \text{ for all } i = 2, \ldots, m.$$

Monotonicity of matrix interpretations can be ensured by requiring (5) to satisfy $(C_i)_{1,1} \geq 1$ for all i, cf. [13]. It is easy to derive the following from Theorem 5:

Proposition 8. *A PTRS $\mathcal{R}$ is SAST if there exist $\epsilon > 0$ and a monotone matrix interpretation $\mathcal{X}$ such that $\mathcal{R} \subseteq [\gg_{\epsilon} \mathbb{E}]_{\mathcal{X}}$.*

Proof. The order $\gg_{\epsilon}$ is collapsible with $f((x_1, \ldots, x_m)^T) = x_1$. It is well known that (5) is affine and thus concave. $\square$

As for polynomial interpretations, for finite systems we do not have to find ϵ. Below $\gg$ is defined in the same manner as $\gg_{\epsilon}$ but replacing "$\geq \epsilon +$" by $>$.

Corollary 3. *A finite PTRS $\mathcal{R}$ is SAST if there exists a monotone matrix interpretation $\mathcal{X}$ such that $\mathcal{R} \subseteq [\gg \mathbb{E}]_{\mathcal{X}}$.*

Example 10. *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

$$\mathtt{a}_{\mathcal{X}}(\vec{x}) = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} \cdot \vec{x} + \begin{bmatrix} 0 \\ 1 \end{bmatrix}\ , \qquad \mathtt{b}_{\mathcal{X}}(\vec{x}) = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \cdot \vec{x}\ .$$

Then we have

$$[\![\mathtt{a}(\mathtt{a}(x))]\!]_{\mathcal{X}}^{\alpha} = \begin{bmatrix} x_1 + x_2 + 1 \\ 1 \end{bmatrix} \gg_{1-2p} \begin{bmatrix} x_1 + x_2 + 2p \\ 1 \end{bmatrix}$$

$$= p \cdot [\![\mathtt{a}(\mathtt{a}(\mathtt{a}(x)))]\!]_{\mathcal{X}}^{\alpha} + (1 - p) \cdot [\![\mathtt{a}(\mathtt{b}(\mathtt{a}(x)))]\!]_{\mathcal{X}}^{\alpha}$$

where $\alpha(x) = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$. *Hence this PARS is SAST if* $p < \frac{1}{2}$, *by Corollary 3.*

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, since multilinear polynomials are not expressible in the form of (5), although linear ones are.

5. Implementation

We extended the termination prover `NaTT` [39] with a syntax for probabilistic rules, and implemented the probabilistic versions of polynomial and matrix interpretations as `NaTT` version 1.9.

The input format extends the WST format.[4] A probabilistic rewrite rule is specified by

$$l \ \text{->}\ w_1\ :\ r_1\ ||\ \ldots\ ||\ w_n\ :\ r_n$$

indicating the probabilistic rewrite rule

$$l \to \{\!\{ \tfrac{w_1}{w} : r_1, \ldots, \tfrac{w_n}{w} : r_n \}\!\} \quad \text{with } w = \textstyle\sum_{j=1}^{n} w_j. \tag{6}$$

The problem of finding interpretations is encoded as a *satisfiability modulo theory* (SMT) problem and solved by an SMT solver. We already have an implementation to encode that $[\![l]\!]_{\mathcal{X}}^{\alpha} > [\![r]\!]_{\mathcal{X}}^{\alpha}$ holds for arbitrary α, so we only need a little extension to encode

$$w \cdot [\![l]\!]_{\mathcal{X}}^{\alpha} > w_1 \cdot [\![r_1]\!]_{\mathcal{X}}^{\alpha} + \cdots + w_n \cdot [\![r_n]\!]_{\mathcal{X}}^{\alpha}$$

which expresses the desired orientation condition of a probabilistic rule (6).

---

[4]`https://www.lri.fr/~marche/tpdb/format.html`, accessed November 14, 2017.

Example 11. *The bound depicted in Example 10 is found by our implementation in* NaTT.

The following example deserves some attention.

Example 12. *Consider the following encoding of [14, Figure 1]:*

$$?(x) \to \{\!\{\tfrac{1}{2} : ?(\mathtt{s}(x)), \; \tfrac{1}{2} : \$(\mathtt{g}(x))\}\!\} \qquad\qquad \$(\mathtt{0}) \to \{\!\{1 : \mathtt{0}\}\!\}$$
$$?(x) \to \{\!\{1 : \$(\mathtt{f}(x))\}\!\} \qquad\qquad\qquad\quad \$(\mathtt{s}(x)) \to \{\!\{1 : \$(x)\}\!\}$$

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

When f *and* 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* $\mathtt{g}(x) = \lfloor\tfrac{x}{2}\rfloor$*,* NaTT *(combined with the SMT solver* z3 *version 4.4.1) found the following polynomial interpretation proving SAST:*

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

While the above example is merely an encoding of a procedural probabilistic program, PTRSs are well-suited to model probabilistic *functional* programs featuring datatypes such as list, trees, etc.

Example 13. *The following PTRS encodes a probabilistic function* rlist *that samples a list over naturals, where elements and the length of the list follow a geometric distribution.*

$$\mathtt{rlist}(xs) \to \{\!\{\tfrac{1}{2} : xs, \; \tfrac{1}{2} : \mathtt{cons}(\mathtt{rnat}(\mathtt{0}), \mathtt{rlist}(xs))\}\!\}$$
$$\mathtt{rnat}(x) \to \{\!\{\tfrac{1}{2} : x, \; \tfrac{1}{2} : \mathtt{rnat}(\mathtt{s}(x))\}\!\}$$

The system is SAST, as NaTT *finds the following polynomial interpretation:*

$$\mathtt{0}_\mathcal{X}() = 1 \qquad\qquad \mathtt{s}_\mathcal{X}(x) = x \qquad \mathtt{cons}_\mathcal{X}(x, xs) = x + xs$$
$$\mathtt{rnat}_\mathcal{X}(x) = x + 1 \qquad \mathtt{rlist}_\mathcal{X}(xs) = xs + 3 \;.$$

Finally, we also remark that our methods cannot handle the case where the expected derivation length is not bounded by multilinear polynomials.

Example 14. *Consider Example 12 with rules* $\mathtt{f}(x) \to \mathtt{mul}(x,x)$ *and other rules that demand* $\mathtt{mul}(x,x)$ *to take reduction length* x^2*. For any (multi)linear polynomial* $\mathtt{f}_\mathcal{X}$*, it is impossible to have* $\mathtt{f}_\mathcal{X}(x) > x^2$*. Hence we cannot prove the PTRS to be SAST by the multilinear polynomial interpretation method. The same holds for the matrix interpretation.*

6. 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 future work includes clarifying the place of SAST in the arithmetical hierarchy, and extending our result for proving AST, not only SAST.

Acknowledgments

References

[1] G. Agha, J. Meseguer, and K. Sen. PMaude: Rewrite-based specification language for probabilistic object systems. *Electr. Notes in Theor. Comput. Sci.*, 153(2):213–239, 2006.

[2] M. Avanzini. *Verifying Polytime Computability Automatically*. PhD thesis, University of Innsbruck, 2013.

[3] M. Avanzini, U. Dal Lago, and A. Yamada. On Probabilistic Term Rewriting. In *Proc. of 14th FLOPS*, volume 10818 of *LNCS*, pages 132–148. Springer, 2018.

[4] M. Avanzini, M. Schaper, and G. Moser. Modular Runtime Complexity Analysis of Probabilistic While Programs. In *Proc. of 10th DICE and 6th FOPARA*, 2019.

[5] F. Baader and T. Nipkow. *Term Rewriting and All That*. Cambridge University Press, 1998.

[6] O. Bournez and F. Garnier. Proving positive almost-sure termination. In *Proc. of 16th RTA*, volume 3467 of *LNCS*, pages 323–337. Springer, 2005.

[7] O. Bournez and F. Garnier. Proving positive almost sure termination under strategies. In *Proc. of 17th RTA*, volume 4098 of *LNCS*, pages 357–371. Springer, 2006.

[8] O. Bournez and C. Kirchner. Probabilistic rewrite strategies. Applications to ELAN. In *Proc. of 13th RTA*, pages 252–266, 2002.

[9] A. Chakarov and S. Sankaranarayanan. Probabilistic Program Analysis with Martingales. In *Proc. of 25th CAV*, volume 8044 of *LNCS*, pages 511–526. Springer, 2013.

[10] K. Chatterjee, H. Fu, and A. K. Goharshady. Termination Analysis of Probabilistic Programs Through Positivstellensatz's. In *Proc. of 28th CAV*, volume 9779 of *LNCS*, pages 3–22. Springer, 2016.

[11] U. Dal Lago and M. Zorzi. Probabilistic operational semantics for the lambda calculus. *RAIRO - TIA*, 46(3):413–450, 2012.

[12] K. De Leeuw, E. F. Moore, C. E. Shannon, and N. Shapiro. Computability by probabilistic machines. *Automata Studies*, 34:183–198, 1956.

[13] J. Endrullis, J. Waldmann, and H. Zantema. Matrix interpretations for proving termination of term rewriting. *JAR*, 40(3):195–220, 2008.

[14] L. M. Ferrer Fioriti and H. Hermanns. Probabilistic termination: Soundness, completeness, and compositionality. In *Proc. of 42nd POPL*, pages 489–501. ACM, 2015.

[15] H. Fu and K. Chatterjee. Termination of Nondeterministic Probabilistic Programs. In *VMCAI*, volume 11388 of *LNCS*, pages 468–490. Springer, 2019.

[16] C. Fuhs, J. Giesl, A. Middeldorp, P. Schneider-Kamp, R. Thiemann, and H. Zankl. SAT solving for termination analysis with polynomial interpretations. In *Proc. of 10th SAT*, volume 4501 of *LNCS*, pages 340–354. Springer, 2007.

[17] J. Gill. Computational complexity of probabilistic Turing machines. *SIAM JC*, 6(4):675–695, 1977.

[18] I. Gnaedig. Induction for positive almost sure termination. In *PPDP 2007*, pages 167–178. ACM, 2007.

[19] S. Goldwasser and S. Micali. Probabilistic encryption. *JCSS*, 28(2):270–299, 1984.

[20] N. D. Goodman, V. K. Mansinghka, D. M. Roy, K. Bonawitz, and J. B. Tenenbaum. Church: a language for generative models. In *Proc. of 24th UAI*, pages 220–229. AUAI Press, 2008.

[21] N. Hirokawa and G. Moser. Automated complexity analysis based on context-sensitive rewriting. In *RTA-TLCA 2014*, volume 8560 of *LNCS*, pages 257–271, 2014.

[22] D. Hofbauer and C. Lautemann. Termination Proofs and the Length of Derivations. In *Proc. of 3rd RTA*, volume 355 of *LNCS*, pages 167–177. Springer, 1989.

[23] C. Jones and G. D. Plotkin. A probabilistic powerdomain of evaluations. In *Proc. of 4th LICS*, pages 186–195, 1989.

[24] B. L. Kaminski and J.-P. Katoen. On the hardness of almost-sure termination. In *MFCS 2015, Part I*, volume 9234 of *LNCS*, pages 307–318. Springer, 2015.

[25] B. L. Kaminski, J.-P. Katoen, C. Matheja, and F. Olmedo. Weakest Precondition Reasoning for Expected Runtimes of Randomized Algorithms. *J. ACM*, 65(5):30:1–30:68, 2018.

[26] U. Dal Lago and C. Grellois. Probabilistic termination by monadic affine sized typing. In *Proc. of 26th ESOP*, pages 393–419, 2017.

[27] U. Dal Lago and S. Martini. On constructor rewrite systems and the lambda calculus. *LMCS*, 8(3), 2012.

[28] D. Lankford. Canonical algebraic simplification in computational logic. Technical Report ATP-25, University of Texas, 1975.

[29] D.S. Lankford. On Proving Term Rewriting Systems are Noetherian. Technical Report MTP-3, Louisiana Technical University, 1979.

[30] S. Lucas. Polynomials over the reals in proofs of termination: From theory to practice. *ITA*, 39(3):547–586, 2005.

[31] A. McIver, C. Morgan, B. L. Kaminski, and J.-P. Katoen. A New Proof Rule for Almost-sure Termination. *PACMPL*, 2(POPL):33:1–33:28, 2018.

[32] R. Motwani and P. Raghavan. *Randomized algorithms*. Cambridge University Press, 1995.

[33] N. C. Ngo, Q. Carbonneaux, and J. Hoffmann. Bounded Expectations: Resource Analysis for Probabilistic Programs. In *Proc. of 39th PLDI*, pages 496–512, 2018.

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

[35] M. O. Rabin. Probabilistic automata. *Information and Control*, 6(3):230–245, 1963.

[36] N. Saheb-Djahromi. Probabilistic LCF. In *MFCS*, volume 64 of *LNCS*, pages 442–451. Springer, 1978.

[37] E. S. Santos. Probabilistic Turing machines and computability. *Proc. of the American Mathematical Society*, 22(3):704–710, 1969.

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

[39] A. Yamada, K. Kusakari, and T. Sakabe. Nagoya Termination Tool. In *RTA-TLCA 2014*, volume 8560 of *LNCS*, pages 466–475, 2014.