The Interpretation Method for Probabilistic Systems

Martin Avanzini[1], Ugo Dal Lago[2], and Akihisa Yamada[3]

1 **INRIA Sophia Antipolis, France**
`martin.avanzini@inria.fr`
2 **INRIA Sophia Antipolis, France & University of Bologna, Italy**
`ugo.dallago@unibo.it`
3 **National Institute of Informatics, Japan**
`akihisa.yamada@uibk.ac.at`

1 Introduction

This abstract extends the abstract *A Simplified Framework for Probabilistic Reductions in the Presence of Nondeterminism* in two directions. First, we follow Bournez and Garnier [4] in defining *probabilistic term rewrite systems* (*PTRSs*) on top of *probabilistic abstract reduction systems* (*PARSs*). Second, we study 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. In correspondence to the non-probabilistic setting (e.g., [3, 9]), suitable restrictions of these methods can be used to assess polynomial bounds on the (expected) runtime of investigated systems. We have also implemented these new techniques into the termination tool `NaTT` [10].

2 Probabilistic Abstract Reduction Systems and Ranking Functions

An *abstract reduction system* (*ARS*) is a binary (transition) relation $\to\ \subseteq A \times A$ on a set of *objects* A, with $a \to b$ indicating that a reduces to b in one step. We follow Bournez and Garnier [4] and define *probabilistic ARSs* (*PARSs*) similar to ARSs, where however reducts b are sampled from a probability distribution d, i.e., mappings from A to the *positive reals* $\mathbb{R}_{\geq 0}$ so that $\sum_{a \in A} d(a) = 1$. For simplicity, we restrict ourselfs to distributions d with finite support $\mathsf{Supp}(d) := \{a \in A \mid d(a) > 0\}$. Such a distribution d is sometimes also denoted by $\{d(a_1) : a_1, \ldots,\ d(a_n) : a_n\}$ where $\mathsf{Supp}(d) = \{a_1, \ldots, a_n\}$.

▶ **Definition 1** (Probabilistic ARS [4])**.** A *probabilistic ARS* (*PARS*) $\mathcal{A}$ over objects A consists of a (usually infinite) set of transitions of the form $a \to d$ for $a \in A$ and d a distribution over A.

The intended meaning of $a \to d \in \mathcal{A}$ is that a reduces $b \in \mathsf{Supp}(d)$ with propability $d(b)$.

▶ **Example 2** (Random walk)**.** A random walk over $\mathbb{N}$ is modeled by the PARS $\mathcal{W}$ consisting of the probabilistic transition

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

In the case $p = \frac{1}{2}$, equally likely, positive natural numbers n are incremented or decremented by one in each transition step, i.e., the walk is unbiased.

8th Workshop on Developments in Implicity Complexity.
Editor: M. Avanzini; Article No. 1; pp. 1:1–1:5
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

In contrast to Bournez and Garnier [4] that describe reductions as stochastic sequences, we give dynamics to PARSs via a reduction relation on *multidistributions*, a generalization of distributions. The two notions are equiexpressive. For motivation and intuitions, we kindly refer the reader to the extended version [1].

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 $\mathsf{FMDist}(A)$. We identify distributions $\{p_1 : a_1, \ldots, p_n : a_n\}$ with multidistribution $\{\!\{p_1 : a_1, \ldots, p_n : a_n\}\!\}$ as no confusion can arise. We define *scalar multiplication* by $p \cdot \{\!\{q_1 : a_1, \ldots, q_n : a_n\}\!\} := \{\!\{p \cdot q_1 : a_1, \ldots, p \cdot q_n : a_n\}\!\}$, which yields a multidistribution whenever $0 \leq p \leq 1$. More generally, multidistributions are closed under *convex multiset unions* $\biguplus_{i=1}^{n} p_i \cdot \mu_i$, where $p_1, \ldots, p_n \geq 0$ and $p_1 + \cdots + p_n \leq 1$.

▶ **Definition 3** (Probabilistic Reduction)**.** Given a PARS $\mathcal{A}$ over objects A, we define the *probabilistic reduction relation* $\rightsquigarrow_{\mathcal{A}} \subseteq \mathsf{FMDist}(A) \times \mathsf{FMDist}(A)$ such that $\mu \rightsquigarrow_{\mathcal{A}} \nu$ holds if $\mu = \{\!\{p_1 : a_1, \ldots, p_n : a_n\}\!\}$, $\nu_i = \biguplus_{i=1}^{n} p_i \cdot \nu_i$ where, for all $1 \leq i \leq n$, either $a_i \to d_i \in \mathcal{A}$ for some distribution d_i and $\nu_i = d_i$, or otherwise, $\nu_i = \varnothing$.

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 $i \in \mathbb{N}$.

Concerning the random walk depicted in Example 2 we have:

$$\{\!\{1 : 1\}\!\} \rightsquigarrow_{\mathcal{W}_{\frac{1}{2}}} \{\!\{\tfrac{1}{2} : 0,\ \tfrac{1}{2} : 2\}\!\} \rightsquigarrow_{\mathcal{W}_{\frac{1}{2}}} \{\!\{\tfrac{1}{4} : 1,\ \tfrac{1}{4} : 3\}\!\} \rightsquigarrow_{\mathcal{W}_{\frac{1}{2}}} \{\!\{\tfrac{1}{8} : 0,\ \tfrac{1}{8} : 2, \tfrac{1}{8} : 2,\ \tfrac{1}{8} : 4\}\!\} \rightsquigarrow_{\mathcal{W}_{\frac{1}{2}}} \cdots$$

To measure the runtime of reductions M, we extend the unitary measure, which attributes a cost of one to a transition $a \to b$, to the probabilistic setting by measuring the mean-length of transition sequences underlying M. The mean-length can actually be expressed as follows:

▶ **Definition 4** (Expected Derivation Length, Height)**.** We define the *expected derivation length* $\mathsf{edl}(M) \in \mathbb{R} \cup \{\infty\}$ of $M = (\mu_0 \rightsquigarrow_{\mathcal{A}} \mu_1 \rightsquigarrow_{\mathcal{A}} \cdots)$ by $\mathsf{edl}(M) := \sum_{i \geq 1} |\mu_i|$ and the *expected derivation height* $\mathsf{edh}_{\mathcal{A}}(a) \in \mathbb{R} \cup \{\infty\}$ of an object $a \in A$ by $\sup_{M \in \mathcal{A}(\{\!\{1:a\}\!\})} \mathsf{edl}(M)$.

PARSs constitute a conservative extention of ARSs, and this observation in particular carries over to the notion of (expected) derivation length and height [1].

Probabilistic Ranking Functions

Bournez and Garnier [4] generalized *ranking functions* to PARS. We give here a simpler but equivalent definition, 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 5.** For a PARS $\mathcal{A}$ on A, we say that a function $f : A \to \mathbb{R}_{\geq 0}$ is a *(probabilistic,* or *Lyapunov) ranking function* if there exists $\epsilon > 0$ such that $a \to d \in \mathcal{A}$ implies $f(a) >_\epsilon \mathbb{E}(f(d))$.

▶ **Theorem 6.** *Let f be a ranking function for a PARS $\mathcal{A}$ over A. Then there is $\epsilon > 0$ such that* $\mathsf{edl}(M) \leq \frac{\mathbb{E}(f(\mu))}{\epsilon}$ *for all $M \in \mathcal{A}(\mu)$. In particular,* $\mathsf{edh}_{\mathcal{A}}(a) \leq \frac{f(a)}{\epsilon}$ *for all $a \in A$.*

3 Probabilistic Term Rewrite Systems and the Interpretation Method

We assume (modest) familiarity with term rewriting [2] and just introduce notations. With $T(F, V)$ we denote the set of terms over a *signature* F and *variables* V. With σ and C

we denote substitutions and contexts, respectively. Substitutions and context applications are extended homomorphically to multidistributions: for $\mu = \{\{p_1 : t_1, \ldots, p_n : t_n\}\}$, $\mu\sigma := \{\{p_1 : t_1\sigma, \ldots, p_n : t_n\sigma\}\}$ and $C[\mu] := \{\{p_1 : C[t_1], \ldots, p_n : C[t_n]\}\}$. Note that, even for a distribution d, $d\sigma$ may fail to be a distribution; e.g., consider $\{\frac{1}{2} : x,\ \frac{1}{2} : y\}\sigma$ with $x\sigma = y\sigma$. Hence, we approximate multidistribution μ by a mapping $\overline{\mu}$ defined by $\overline{\mu}(a) := \sum_{p:a\in\mu} p$, which form a distribution if $|\mu| = 1$.

▶ **Definition 7** (Probabilistic Term Rewriting [4])**.** 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 transition $C[l\sigma] \to C[\,\overline{d\sigma}\,]$ for every probabilistic rewrite rule $l \to d \in \mathcal{R}$, context C, and substitution σ.

▶ **Example 8.** Encoding natural numbers as tally-sequences, the random walk of Example 2 can be modeled by a PTRS consisting of a single rule $\mathtt{s}(x) \to \{p : x,\ 1-p : \mathtt{s}(\mathtt{s}(x))\}$. Notice that for $p \leq \frac{1}{2}$ the expected runtime of this system is unbounded, for $p > \frac{1}{2}$ we derive a bound in Example 13 below.

The Interpretation Method

In the non-probabilistic case, the interpretation method refers to a class of termination methods that use monotone F-algebras to embed reduction sequences into a well-founded order $\succ$. This method is not only sound and complete for proving strong normalisation, but, using suitable F-algebras they can also be used to assess polynomial bounds on runtimes (e.g., [3, 9]). The following notion is standard.

▶ **Definition 9** (monotone, F-Algebra)**.** A *monotone F-algebra* $(\mathcal{X}, \succ)$ on a non-empty carrier set X consists of an F-algebra $\mathcal{X}$, i.e., *interpretation* $\mathtt{f}_{\mathcal{X}} : X^{\mathsf{ar}(\mathtt{f})} \to X$ of each function symbol $\mathtt{f} \in F$, and an order $\succ$ on X so that all interpretations are monotone wrt. $\succ$: $x \succ y$ 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 by $[\![x]\!]^{\alpha}_{\mathcal{X}} := \alpha(x)$ if x is a variable, and $[\![\mathtt{f}(t_1, \ldots, t_n)]\!]^{\alpha}_{\mathcal{X}} := \mathtt{f}_{\mathcal{X}}([\![t_1]\!]^{\alpha}_{\mathcal{X}}, \ldots, [\![t_n]\!]^{\alpha}_{\mathcal{X}})$. If t is ground, i.e., contains no variables, we may drop the assignment α. We write $s \succ_{\mathcal{X}} t$ iff $[\![s]\!]^{\alpha}_{\mathcal{X}} \succ [\![t]\!]^{\alpha}_{\mathcal{X}}$ for every α.

Suppose a TRS $\mathcal{R}$ satisfies the *orientation condition* $l \succ_{\mathcal{X}} r$ for all $l \to r \in \mathcal{R}$. Then, for every (ground) term t, the interpretation $[\![\cdot]\!]_{\mathcal{X}}$ embeds $\mathcal{R}$-reduction steps into $\succ$, and consequently the length of reduction sequences starting from t can be estimated via an analysis of $\succ$-descending sequences starting from $[\![t]\!]_{\mathcal{X}}$. To assess the latter, a standard way is to require an embedding of $\succ$ into the natural order, i.e., $x \succ y$ implies $\mathsf{G}(x) >_{\mathbb{N}} \mathsf{G}(y)$ for some $\mathsf{G} : \mathcal{X} \to \mathbb{N}$ (cf. [6]). Consequently, $\mathsf{G}([\![t]\!]^{\alpha}_{\mathcal{X}})$ binds the runtime of $\mathcal{R}$ on t. Slightly generalising this idea, we call the order $\succ$ *ϵ-collapsible* if there is a function $\mathsf{G} : \mathcal{X} \to \mathbb{R}_{\geq 0}$ embedding $\succ$ into $>_{\epsilon}$. To lift the interpretation method to the probabilistic setting, we equip the carrier set $\mathcal{X}$ with a *barycentric operation* $\mathbb{E}_X : \mathsf{FDist}(\mathcal{X}) \to \mathcal{X}$. We call $\mathcal{X}$ a *barycentric domain.* Of particular interest in this work will be the barycentric domains $\mathbb{R}$ and $\mathbb{R}^m$ with barycentric operations $\mathbb{E}(\{p_1 : a_1, \ldots, p_n : a_n\}) = \sum_{i=1}^{n} p_i \cdot a_i$.

Recall that a function $f : \mathbb{R} \to \mathbb{R}$ is *concave* if for all $0 \leq p \leq 1$, $f(p \cdot x + (1-p) \cdot y) \geq p \cdot f(x) + (1-p) \cdot f(y)$ for all x, y. For instance, affine functions are (trivially) concave. This notion can be generalized to functions $f : X \to Y$ with barycentric domains X and Y: f is *concave* wrt. 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 $\succ$.

▶ **Definition 10** (Barycentric F-Algebra)**.** A *barycentric F-algebra* $(\mathcal{X}, \succ)$ consists 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 define the relation $\succ^{\mathbb{E}} \subseteq X \times \mathsf{FDist}(X)$ by $x \succ^{\mathbb{E}} d$ iff $x \succ \mathbb{E}_X(d)$.

▶ **Theorem 11.** *Suppose $\mathcal{R}$ is oriented by a barycentric F-algebra $(\mathcal{X}, \succ)$, i.e., $l \succ^{\mathbb{E}} d$ for all rules $l \to d \in \mathcal{R}$, and that $\succ$ is ϵ-collapsible by a concave function G. Then $\mathsf{edh}_{\widehat{\mathcal{R}}}(t) \leq \frac{\mathsf{G}([\![t]\!]_{\mathcal{X}})}{\epsilon}$ for every ground term t.*

This result can be proven with Theorem 6. We remark that its conditions are all necessary.

Instances of Prebarycentric Algebras

▶ **Definition 12** (Polynomial Interpretation [7, 8])**.** 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 *restricted* if every $\mathtt{f}_{\mathcal{X}}$ is of the following form with $\mathtt{c}_V \in \mathbb{R}_{\geq 0}$:

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

Restricted polynomial interpretations $\mathcal{X}$ are monotone and concave by definition, hence $(\mathcal{X}, >_\epsilon)$ forms a barycentric F-Algebra. These subsume a similar notion introduced in [4]. Noteworthy, the context decrease condition in [4] rules out interpretations such as $\mathtt{f}_{\mathcal{X}}(x) = 2x$. Note that $>_\epsilon$ is ϵ-collapsible using as G the identity.

▶ **Example 13** (Example 8 Continued)**.** Define $\mathcal{X}$ such that $\mathtt{0}_{\mathcal{X}} := 0$ and $\mathtt{s}_{\mathcal{X}}(x) := x + 1$. Fix $p > \frac{1}{2}$ and pick $\epsilon \leq 1 - 2p$. Then

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

Hence the tally representation $\underline{n}$ of $n \in \mathbb{N}$ reduces to normal form in at most $\frac{[\![\underline{n}]\!]_{\mathcal{X}}}{\epsilon} = \frac{n}{\epsilon}$ steps on average, by Theorem 11.

▶ **Definition 14** (Matrix Interpretation [5])**.** 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

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

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

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

Since matrix interpretations are affine, they are always concave. Furthermore, monotonicity can be ensured if (1) satisfies $(C_i)_{1,1} \geq 1$ for all i, cf. [5]. Thus, if $\mathcal{Y}$ is such a matrix interpretation, $(\mathcal{X}, \gg_\epsilon)$ forms a barycentric F-algebra, with $\gg_\epsilon$ ϵ-collapsible using as mapping $\mathsf{G} : \mathbb{R}^m_{\geq 0} \to \mathbb{R}_{\geq 0}$ the projection to the first argument.

▶ **Example 15.** Consider the PTRS consisting of the single probabilistic rule

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

It can be shown that the two-dimensional matrix interpretation $\mathcal{Y}$ with

$$[\![\mathsf{a}]\!]^{\alpha}_{\mathcal{Y}}(\vec{x}) = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} \cdot \vec{x} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \ , \qquad [\![\mathsf{b}]\!]^{\alpha}_{\mathcal{Y}}(\vec{x}) = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \cdot \vec{x} \ ,$$

orients the PTRS for $0 < \epsilon \leq 1 - 2p$. The interpretation falls into a class defined in [9, Theorem 16] with $([\![t]\!]_{\mathcal{Y}})_1$ linearly related to the size of t. Consequently, the expected runtime of the PTRS is bounded by a linear function, by Theorem 11.

4 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.

References

1 M. Avanzini, U. Dal Lago, and A. Yamada. On probabilistic term rewriting. Technical report, INRIA, 2017. Available at `http://cl-informatik.uibk.ac.at/users/zini/PTRS.pdf`.
2 F. Baader and T. Nipkow. *Term Rewriting and All That*. Cambridge University Press, 1998.
3 G. Bonfante, A. Cichon, J.-Y. Marion, and H. Touzet. Algorithms with Polynomial Interpretation Termination Proof. *JFP*, 11(1):33–53, 2001.
4 O. Bournez and F. Garnier. Proving positive almost-sure termination. In *Proc. of 16th RTA*, volume 3467 of *LNCS*, pages 323–337. Springer, 2005.
5 J. Endrullis, J. Waldmann, and H. Zantema. Matrix Interpretations for Proving Termination of Term Rewriting. *JAR*, 40(3):195–220, 2008.
6 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.
7 D. Lankford. Canonical algebraic simplification in computational logic. Technical Report ATP-25, University of Texas, 1975.
8 Salvador Lucas. Polynomials Over the Reals in Proofs of Termination: From Theory to Practice. *ITA*, 39(3):547–586, 2005.
9 A. Middeldorp, G. Moser, F. Neurauter, J. Waldmann, and H. Zankl. Joint Spectral Radius Theory for Automated Complexity Analysis of Rewrite Systems. In *Proc. of 4th CAI*, volume 6742 of *LNCS*, pages 1–20. Springer, 2011.
10 Akihisa Yamada, Keiichirou Kusakari, and Toshiki Sakabe. Nagoya Termination Tool. In *RTA-TLCA 2014*, volume 8560 of *LNCS*, pages 466–475, 2014.