# Probabilistic Term Rewriting and the Interpretation Method

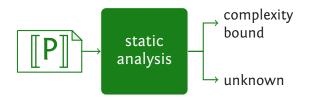
Martin Avanzini and Ugo Dal Lago and Akihisa Yamada





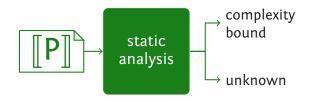


### **Automated Complexity Analysis**





### **Automated Complexity Analysis**



- AProVE (http://aprove.informatik.rwth-aachen.de)
- TcT (http://cl-informatik.uibk.ac.at/software/tct)
- RaML (http://raml.co)
- UIS

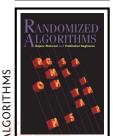
# **Probabilistic Computation is Becoming Pervasive**

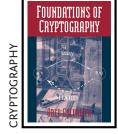
ARTIFICIAL INTELLIGENCE



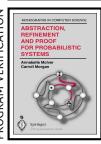








ANGUAGE PROCESSING VATURAL ROGRAM VERIFICATION



#### **Starting Point**

#### Proving Positive Almost-Sure Termination

Olivier Bournez, Florent Garnier



LORIA/INRIA, 615 Rue du Jardin Botanique 54602 Villers lès Nancy Cedex, France

Abstract In order to extend the modeling capabilities of rewriting systems, it is rather natural to consider that the firing of rules can be subject to some probabilistic laws. Considering rewrite rules subject to probabilities leads to numerous questions about the underlying notions and results.

We focus here on the problem of termination of a set of probabilistic rewrite rules. A probabilistic rewrite system is said almost surely terminating if the probability that a derivation leads to a normal form is one. Such a system is said positively almost surely terminating if furthermore the mean length of a derivation is finite. We provide several results and techniques in order to prove positive almost sure termination of a given set of probabilistic rewrite rules. All these techniques subsume classical ones for non-probabilistic systems.

#### 1 Introduction

Since 30 years, term rewriting has shown to be a very powerful tool in several



### **Probabilistic Abstract Reduction Systems**

Definition (PARS – Bournez & Garnier, RTA'05)

Probabilistic abstract reduction system is tuple  $A = (A, \rightarrow)$  s.t.:

- A is countable set of objects
- → ⊆ *A* × Dist(*A*) maps elements from *A* to (discrete) probability distributions over *A*

#### Intuitions:

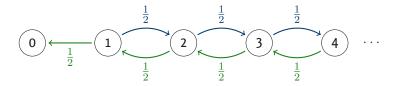
- 1. probabilistic choice:
  - if  $a \to \{p_i : b_i\}_i$  then a reduces to  $b_i$  with probability  $p_i$ ;
- 2. non-deterministic choice:
  - if  $a o d_1$  and  $a o d_2$  ( $d_1
    eq d_2$ ), reduct is chosen from  $d_1$  or  $d_2$ .



PARS  $\mathcal{W} = (\mathbb{N}, \rightarrow)$  where

$$n \to \{\frac{1}{2} \colon n+1; \frac{1}{2} \colon n-1\}$$
 (for all  $n > 0$ ),

defines simple random walk on  $\mathbb{N}$ :

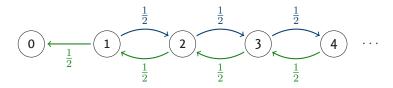




PARS  $\mathcal{W} = (\mathbb{N}, \rightarrow)$  where

$$n \to \{\frac{1}{2} \colon n+1; \frac{1}{2} \colon n-1\}$$
 (for all  $n > 0$ ),

defines simple random walk on  $\mathbb{N}$ :



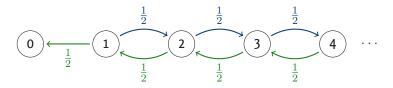
#### Some properties of interest:

· almost-sure termination

PARS  $\mathcal{W} = (\mathbb{N}, \rightarrow)$  where

$$n \to \{\frac{1}{2} \colon n+1; \frac{1}{2} \colon n-1\}$$
 (for all  $n > 0$ ),

defines simple random walk on  $\mathbb{N}$ :



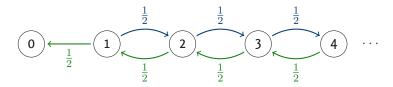
#### Some properties of interest:

- · almost-sure termination
- positive almost-sure termination

PARS  $\mathcal{W} = (\mathbb{N}, \rightarrow)$  where

$$n \to \{\frac{1}{2} : n+1; \frac{1}{2} : n-1\}$$
 (for all  $n > 0$ ),

defines simple random walk on  $\mathbb{N}$ :



#### Some properties of interest:

- · almost-sure termination
- positive almost-sure termination
- · expected runtime

• Bournez and Garnier model reductions as stochastic processes

$$\mathbf{X} = \mathbf{X}_0, \mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \dots$$

- *n*th random variable  $X_n$  gives the *n*th reduct (or  $\perp$ )



Bournez and Garnier model reductions as stochastic processes

$$\mathbf{X} = \mathbf{X}_0, \mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \dots$$

- *n*th random variable  $X_n$  gives the *n*th reduct (or  $\perp$ )
- random variable  $T_{\mathbf{X}}$  over  $\mathbb{N} \cup \{\infty\}$  measures reduction length
- expected reduction length defined as expectation of  $T_X$ :

$$\mathbb{E}(T_{\mathbf{X}}) \triangleq \sum_{n=1}^{\infty} n \cdot \mathbb{P}(T_{\mathbf{X}} = n) \qquad (= \sum_{n>1} \mathbb{P}(T_{\mathbf{X}} \geq n))$$



Bournez and Garnier model reductions as stochastic processes

$$\mathbf{X} = \mathbf{X}_0, \mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \dots$$

- *n*th random variable  $X_n$  gives the *n*th reduct (or  $\perp$ )
- random variable  $T_{\mathbf{X}}$  over  $\mathbb{N} \cup \{\infty\}$  measures reduction length
- expected reduction length defined as expectation of  $T_X$ :

$$\mathbb{E}(T_{\mathbf{X}}) \triangleq \sum_{n=1}^{\infty} n \cdot \mathbb{P}(T_{\mathbf{X}} = n) \qquad (= \sum_{n \geq 1} \mathbb{P}(T_{\mathbf{X}} \geq n))$$

- +a.s. terminating (or PAST) if  $\mathbb{E}(\mathcal{T}_{\mathbf{X}}) \in \mathbb{N}$  for all  $\mathbf{X}$ 



Bournez and Garnier model reductions as stochastic processes

$$\mathbf{X} = \mathbf{X}_0, \mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \dots$$

- nth random variable  $X_n$  gives the nth reduct (or  $\perp$ )
- random variable  $T_{\mathbf{X}}$  over  $\mathbb{N} \cup \{\infty\}$  measures reduction length
- expected reduction length defined as expectation of  $T_X$ :

$$\mathbb{E}(T_{\mathbf{X}}) \triangleq \sum_{n=1}^{\infty} n \cdot \mathbb{P}(T_{\mathbf{X}} = n) \qquad (= \sum_{n>1} \mathbb{P}(T_{\mathbf{X}} \geq n))$$

- +a.s. terminating (or PAST) if  $\mathbb{E}(T_{\mathbf{X}}) \in \mathbb{N}$  for all  $\mathbf{X}$ 

#### **Pros**

· solid theoretical foundation

#### Cons

· definition of whole machinery involved

$$\mathbf{a} \rightarrow \left\{ \frac{1}{2} : \mathbf{b_1}, \ \frac{1}{2} : \mathbf{b_2} \right\} \qquad \mathbf{b_1} \rightarrow \mathbf{c} \qquad \mathbf{b_2} \rightarrow \mathbf{c} \qquad \mathbf{c} \rightarrow \mathbf{d_1} \qquad \mathbf{c} \rightarrow \mathbf{d_2}$$

Trajectories

inventeurs du monde numérique

$$\mathtt{a} \to \{\tfrac{1}{2} : \mathtt{b_1}, \ \tfrac{1}{2} : \mathtt{b_2}\} \qquad \mathtt{b_1} \to \mathtt{c} \qquad \mathtt{b_2} \to \mathtt{c} \qquad \mathtt{c} \to \mathtt{d_1} \qquad \mathtt{c} \to \mathtt{d_2}$$

$$b_1 \rightarrow c$$

$$b_2 \rightarrow 0$$

$$\mathtt{c} o \mathtt{d_1}$$

$$\mathsf{c} o \mathsf{d}_2$$

$$\begin{array}{cccc}
1:a \\
\downarrow & \\
\frac{1}{2}:b_1 & \frac{1}{2}:b_2 \\
\downarrow & \downarrow \\
\frac{1}{2}:c & \frac{1}{2}:c \\
\downarrow & \downarrow \\
\frac{1}{2}:d_1 & \frac{1}{2}:d_2
\end{array}$$

$$X_0 = \{1: \mathbf{a}\}$$

$$X_1 = \{\frac{1}{2} : b_1, \frac{1}{2} : b_2\}$$

$$\mathit{X}_2 = \{1:\mathsf{c}\}$$

$$X_3 = \{\frac{1}{2} : d_1, \frac{1}{2} : d_2\}$$

**Trajectories** 

Stochastic Process

parameterised by strategy  $\phi$ :

$$\phi(ab_1c) = (c \rightarrow d_1)$$
  $\phi(ab_2c) = (c \rightarrow d_2)$  ...

Stochastic Process

Multidistribution Reduction

parameterised by strategy  $\phi$ :

**Trajectories** 

... 
$$\phi(ab_1c) = (c \rightarrow d_1)$$
  $\phi(ab_2c) = (c \rightarrow d_2)$  ...

# **Dynamics: As Multidistribution Reductions**

• a (finite) multidistribution over A is a (finite) multiset

$$\mu = \{\!\!\{ p_1: a_1, \ldots, p_n: a_n \}\!\!\}$$
 ,

where  $0 \le p_i \le 1$ ,  $a_i \in A$  and

$$|\mu| \triangleq \sum_{i=1}^n p_i \leq 1.$$



# **Dynamics: As Multidistribution Reductions**

• a (finite) multidistribution over A is a (finite) multiset

$$\mu = \{\!\!\{ p_1 : a_1, \dots, p_n : a_n \}\!\!\}$$
 ,

where  $0 \le p_i \le 1$ ,  $a_i \in A$  and

$$|\mu| \triangleq \sum_{i=1}^n p_i \leq 1.$$

• for PARS  $A = (A, \rightarrow)$ , reduction relation  $\leadsto_A$  is defined s.t.:

$$\{\!\!\{p_1:a_1,\ldots,p_n:a_n\}\!\!\} \leadsto_{\mathcal{A}} p_1\cdot \nu_1 \uplus \cdots \uplus p_n\cdot \nu_n$$
 ,

where either

- $\nu_i = d_i$  for some  $d_i$  with  $a_i \rightarrow d_i$ , or
- $\nu_i = \emptyset$  if  $a_i$  is terminal.



#### **Dynamics: As Multidistribution Reductions**

• a (finite) multidistribution over A is a (finite) multiset

$$\mu = \{ p_1 : a_1, \dots, p_n : a_n \}$$
,

where  $0 \le p_i \le 1$ ,  $a_i \in A$  and

$$|\mu| \triangleq \sum_{i=1}^n p_i \leq 1.$$

• for PARS  $A = (A, \rightarrow)$ , reduction relation  $\leadsto_A$  is defined s.t.:

$$\{\{p_1:a_1,\ldots,p_n:a_n\}\} \leadsto_{\mathcal{A}} p_1\cdot \nu_1 \uplus \cdots \uplus p_n\cdot \nu_n$$
,

where either

- $\nu_i = d_i$  for some  $d_i$  with  $a_i \rightarrow d_i$ , or
- $\nu_i = \emptyset$  if  $a_i$  is terminal.
- expected derivation length of reduction  $M = \mu_0 \leadsto_{\mathcal{A}} \mu_1 \leadsto_{\mathcal{A}} \ldots$ :

$$\operatorname{edl}(M) \triangleq \sum_{n \geq 1} |\mu_n|$$
.

#### **Reductions vs Stochastic Processes**

Theorem (A., Dal Lago & Yamada, FLOPS'18)

Stochastic sequence  $\mathbf{X} = \{X_n\}_{n \in \mathbb{N}}$  and probabilistic reductions  $\mathbf{M} = \mu_0 \leadsto \mu_1 \leadsto \dots$  are in one-to-one correspondence with:

$$\mathbb{P}(X_n=a)=|\mu_n|_a$$
 (for all  $n\in\mathbb{N}$  and  $a\in A$ ),

where  $|\mu|_a \triangleq \sum_{(p:a)\in\mu} p$  denotes the total probability of a in  $\mu$ .



#### **Reductions vs Stochastic Processes**

Theorem (A., Dal Lago & Yamada, FLOPS'18)

Stochastic sequence  $\mathbf{X} = \{X_n\}_{n \in \mathbb{N}}$  and probabilistic reductions  $\mathbf{M} = \mu_0 \leadsto \mu_1 \leadsto \dots$  are in one-to-one correspondence with:

$$\mathbb{P}(X_n = a) = |\mu_n|_a$$
 (for all  $n \in \mathbb{N}$  and  $a \in A$ ),

where  $|\mu|_a \triangleq \sum_{(p:a)\in\mu} p$  denotes the total probability of a in  $\mu$ .

Corollary

$$edl(M) = \mathbb{E}(T_{\mathbf{X}})$$
.



### **Probabilistic Ranking Functions**

Definition (Bournez & Garnier, RTA'05)

Function  $[\![\cdot]\!]:A\to\mathbb{R}_{\geq 0}$  is (Lyapunov) ranking function for PARS  $\mathcal{A}=(A,\to)$  if for some  $\epsilon>0$ ,

$$a o d \implies \llbracket a 
rbracket >_{\epsilon} \mathbb{E}(\llbracket d 
rbracket)$$
 ,

where  $\mathbb{E}(\{p_i \cdot x_i\}_i) \triangleq \sum_i p_i \cdot x_i$  and  $x >_{\epsilon} y$  if  $x \geq y + \epsilon$ .



# **Probabilistic Ranking Functions**

Definition (Bournez & Garnier, RTA'05)

Function  $[\![\cdot]\!]:A\to\mathbb{R}_{\geq 0}$  is (Lyapunov) ranking function for PARS  $\mathcal{A}=(A,\to)$  if for some  $\epsilon>0$ ,

$$a o d \implies \llbracket a \rrbracket >_{\epsilon} \mathbb{E}(\llbracket d \rrbracket)$$
 ,

where  $\mathbb{E}(\{p_i \cdot x_i\}_i) \triangleq \sum_i p_i \cdot x_i$  and  $x >_{\epsilon} y$  if  $x \geq y + \epsilon$ .

Example

Consider the *biased* random walk  $\mathcal{W}_{rac{1}{3}}=(\mathbb{N}, o)$  where

$$n \to \{1/3: n+1; 2/3: n-1\}$$
 (for all  $n > 0$ ),

define  $[\![n]\!] \triangleq n$  and take  $\epsilon = \frac{1}{3}$ . Then for all n > 0,

$$[\![n]\!] >_{\epsilon} (1/3) \cdot [\![n+1]\!] + (2/3) \cdot [\![n-1]\!] .$$

#### **Probabilistic Ranking Functions**

Definition (Bournez & Garnier, RTA'05)

Function  $[\![\cdot]\!]:A\to\mathbb{R}_{\geq 0}$  is (Lyapunov) ranking function for PARS  $\mathcal{A}=(A,\to)$  if for some  $\epsilon>0$ ,

$$a o d \implies \llbracket a 
rbracket >_{\epsilon} \mathbb{E}(\llbracket d 
rbracket)$$
 ,

where  $\mathbb{E}(\{p_i \cdot x_i\}_i) \triangleq \sum_i p_i \cdot x_i$  and  $x >_{\epsilon} y$  if  $x \geq y + \epsilon$ .

Theorem (Bournez & Garnier, RTA'05)

- 1. Soundness: If  $\llbracket \cdot \rrbracket$  is a ranking function for A, then A is PAST.
- 2. Completeness: If A is finitely branching and PAST, then there exists a ranking function  $[\cdot]$  for A.



#### Probabilistic Ranking Functions, revisited

Theorem (A., Dal Lago & Yamada, FLOPS'18)

Let  $\llbracket \cdot \rrbracket$  be a ranking function for A.

- 1. Soundness: If  $\llbracket \cdot \rrbracket$  is a ranking function for  $\mathcal{A}$ , then  $\operatorname{edh}_{\mathcal{A}}(a) \leq \llbracket a \rrbracket \cdot \frac{1}{\epsilon}$  for all  $a \in \mathcal{A}$ .
- 2. Completeness: If  $edh_{\mathcal{A}}(a) \in \mathbb{N}$  for all  $a \in A$ , then  $[a] \triangleq edh_{\mathcal{A}}(a)$  is a ranking function for  $\mathcal{A}$ .



#### Probabilistic Ranking Functions, revisited

Theorem (A., Dal Lago & Yamada, FLOPS'18)

Let  $\llbracket \cdot \rrbracket$  be a ranking function for A.

- 1. Soundness: If  $\llbracket \cdot \rrbracket$  is a ranking function for  $\mathcal{A}$ , then  $\operatorname{edh}_{\mathcal{A}}(a) \leq \llbracket a \rrbracket \cdot \frac{1}{\epsilon}$  for all  $a \in \mathcal{A}$ .
- 2. Completeness: If  $edh_{\mathcal{A}}(a) \in \mathbb{N}$  for all  $a \in A$ , then  $[a] \triangleq edh_{\mathcal{A}}(a)$  is a ranking function for  $\mathcal{A}$ .

#### Example

#### Consider

$$\mathtt{a_n} \to \{\frac{1}{2} : \mathtt{a_{n+1}}; \frac{1}{2} : 0\} \quad \mathtt{a_n} \to 2^n \cdot n \quad n+1 \to n \quad \text{(for all } n \in \mathbb{N} \text{)} \ .$$

- this PARS is PAST, i.e.,  $\operatorname{edl}(M) \in \mathbb{N}$  for every reduction sequence M;
- $\operatorname{edh}(a_0) > \frac{1}{2^n} \cdot (2^n \cdot n) = n$  for all  $n \in \mathbb{N}$ , i.e., is not bounded.

### **Probabilistic Term Rewrite Systems**

• probabilistic TRS  ${\cal R}$  is finite set of probabilistic rewrite rules I o d

$$\begin{split} \operatorname{rand}_{\operatorname{L}}(xs) &\to \{\tfrac{1}{2}: xs, \tfrac{1}{2}: \operatorname{rand}_{\operatorname{N}}(\operatorname{O}) :: \operatorname{rand}_{\operatorname{L}}(xs) \} \\ \operatorname{rand}_{\operatorname{N}}(n) &\to \{\tfrac{1}{2}: n, \tfrac{1}{2}: \operatorname{rand}_{\operatorname{N}}(\operatorname{succ}(n)) \} \end{split}$$

- rewrite relation  $\leadsto_{\mathcal{R}}$  defined in terms of underlying PARS  $\hat{\mathcal{R}}$ 



### **Interpretation Method for Runtime Analysis**

Definition (Hirokawa & Moser, IJCAR'08)

Monotone algebra  $(\llbracket \cdot \rrbracket, \succ)$  on domain X consists of:

- interpretations  $\llbracket \mathtt{f} \rrbracket : \mathsf{X}^k \to \mathsf{X}$  satisfying:
  - monotonicity:  $\mathbf{x} \succ \mathbf{y} \implies \llbracket \mathbf{f} \rrbracket (\dots, \mathbf{x}, \dots) \succ \llbracket \mathbf{f} \rrbracket (\dots, \mathbf{y}, \dots);$
- order  $\succ \subseteq X \times X$  satisfying:
  - collapsibility:  $x \succ y \implies \mathsf{G}(x) >_{\epsilon} \mathsf{G}(y)$  for some  $\mathsf{G}: X \to \mathbb{R}_{\geq 0}$ ;

orients TRS  ${\cal R}$  if

$$l \to r \in \mathcal{R} \implies \llbracket f \rrbracket \alpha \succ \llbracket r \rrbracket \alpha$$
 for all assignments  $\alpha : V \to X$ .



### **Interpretation Method for Runtime Analysis**

Definition (Hirokawa & Moser, IJCAR'08)

Monotone algebra  $(\llbracket \cdot \rrbracket, \succ)$  on domain X consists of:

- interpretations  $\llbracket \mathtt{f} \rrbracket : X^k \to X$  satisfying:
  - monotonicity:  $x \succ y \implies \llbracket \mathtt{f} \rrbracket (\dots, x, \dots) \succ \llbracket \mathtt{f} \rrbracket (\dots, y, \dots);$
- order  $\succ \subseteq X \times X$  satisfying:
  - collapsibility:  $x \succ y \implies \mathsf{G}(x) >_{\epsilon} \mathsf{G}(y)$  for some  $\mathsf{G}: X \to \mathbb{R}_{\geq 0}$ ;

orients TRS  $\mathcal{R}$  if

$$I \to r \in \mathcal{R} \implies \llbracket f \rrbracket \alpha \succ \llbracket r \rrbracket \alpha$$
 for all assignments  $\alpha : V \to X$ .

Theorem (Hirokawa & Moser, IJCAR'08)

Suppose  $(\llbracket \cdot \rrbracket, \succ)$  orients the TRS  $\mathcal{R}.$  Then

$$\mathsf{dh}_{\mathcal{R}}(t) \leq \mathsf{G}(\llbracket t \rrbracket) \cdot \frac{1}{\epsilon} \ .$$

# **Interpretation Method for Runtime Analysis**

Definition (A., Dal Lago & Yamada, FLOPS'18)

Barycentric, monotone algebra  $(\llbracket \cdot \rrbracket, \mathbb{E}, \succ)$  on domain *X* consists of:

- barycentric operation  $\mathbb{E}: \operatorname{Dist}(X) \to X$
- interpretations  $[\![\mathtt{f}]\!]: X^k \to X$  satisfying:
  - monotonicity:  $\mathbf{x} \succ \mathbf{y} \implies \llbracket \mathbf{f} \rrbracket (\dots, \mathbf{x}, \dots) \succ \llbracket \mathbf{f} \rrbracket (\dots, \mathbf{y}, \dots);$
  - concavity:  $\llbracket \mathbf{f} \rrbracket (\dots, \mathbb{E}(\{p_i : x_i\}_i), \dots) \succeq \mathbb{E}(\{p_i : \llbracket \mathbf{f} \rrbracket (\dots, x_i, \dots)\}_i)$ .
- order  $\succ \subseteq X \times X$  satisfying:
  - collapsibility:  $x \succ y \implies \mathsf{G}(x) >_{\epsilon} \mathsf{G}(y)$  for some  $\mathsf{G}: X \to \mathbb{R}_{\geq 0}$ ;

orients TRS  ${\cal R}$  if

$$I \to d \in \mathcal{R} \implies \llbracket f \rVert \alpha \succ \mathbb{E}(\llbracket d \rrbracket \alpha) \text{ for all assignments } \alpha : V \to X.$$

Theorem (A., Dal Lago & Yamada, FLOPS'18)

Suppose  $(\llbracket \cdot \rrbracket, \mathbb{E}, \succ)$  orients the PTRS  $\mathcal{R}.$  Then

$$\mathsf{edh}_{\mathcal{R}}(t) \leq \mathsf{G}(\llbracket t \rrbracket) \cdot \frac{1}{\epsilon} \ .$$

#### **Instances of Barycentric Algebras**

1. multi-linear polynomial interpretations  $(\llbracket \cdot \rrbracket, \mathbb{E}, >_{\epsilon})$  where

$$\llbracket \mathbf{f} \rrbracket (x_1, \dots, x_n) = \sum_{V \subseteq \{x_1, \dots, x_n\}} c_V \cdot \prod_{x_i \in V} x_i \qquad (c_V \in \mathbb{N}/\mathbb{Q}/\mathbb{R})$$

2. matrix interpretations  $(\llbracket \cdot \rrbracket, \mathbb{E}, \gg_{\epsilon})$  where

$$[\![\mathbf{f}]\!](x_1,\ldots,x_n) = \sum_{i=1}^n C_i \cdot \vec{x}_i + \vec{c} \qquad (C_i \in \mathbb{N}^{m \times m}/\mathbb{Q}^{m \times m}/\mathbb{R}^{m \times m})$$

and

$$(\vec{x})^T \gg_{\epsilon} (\vec{y})^T :\iff \vec{x}_1 >_{\epsilon} \vec{y}_1 \text{ and } (\vec{x})^T \geq (\vec{y})^T.$$



#### Conclusion

- simple notion of reduction for probabilistic ARSs / TRSs based on multidistributions
- recovered the completeness proof of Lyapunov ranking functions
- barycentric algebras for reasoning about expected runtimes of probabilistic TRSs
- implementation of polynomial & matrix interpretations (over  $\mathbb{R}^n_{>0}$ ) in NaTT

