



# The annotated dependency pair framework for almost-sure termination of probabilistic term rewriting

Jan-Christoph Kassing \*, Jürgen Giesl 

RWTH Aachen University, Aachen, Germany

## ARTICLE INFO

### Keywords:

Dependency pairs  
Term rewriting  
Probabilistic programming  
Almost-sure termination

## ABSTRACT

Dependency pairs are one of the most powerful techniques to analyze termination of term rewrite systems automatically. We adapt dependency pairs to the probabilistic setting and develop an annotated dependency pair framework for automatically proving almost-sure termination of *probabilistic term rewrite systems*, both for full and innermost rewriting. To evaluate its power, we implemented our framework in the tool AProVE.

## 1. Introduction

Term rewrite systems (TRSs) are used in many areas of computer science, like symbolic computation, automated theorem proving, and automated analysis of programs in different languages. Termination is one of the most crucial properties of programs, and numerous tools have been developed to analyze termination of TRSs, e.g., AProVE [1], Matchbox [2], MultumNonMultum [3], MUTERM [4], NaTT [5], NTI [6], TTT2 [7], etc. One of the most powerful techniques integrated in essentially all current termination tools for TRSs is the *dependency pair* (DP) framework [8–12] which allows modular proofs that apply different techniques in different sub-proofs.

The DP framework also allows for certification of proofs via tools like, e.g., CeTA [13], and DPs have been formalized in several proof assistants (e.g., in Rocq (formerly Coq) [14,15], Isabelle [13], and PVS [16]).

*Probabilistic* programs describe randomized algorithms and probability distributions, with applications in many areas, see, e.g., [17]. For instance, well-known examples are variants of the probabilistic quicksort algorithm, see, e.g., Section 6. If one uses a version where every list element has the same probability of becoming the next pivot element, then the algorithm has an expected runtime of  $\mathcal{O}(n \cdot \log(n))$ , for every possible input list of length  $n$ . In contrast, for a non-probabilistic version of quicksort, there always exists a “worst-case” input of length  $n$  such that the algorithm takes at least  $\mathcal{O}(n^2)$  steps. Several probabilistic programming languages have been developed. An example is the probabilistic guarded command language (pGCL) [18,19], whose basic form however cannot express arbitrary data structures. In contrast, TRSs are especially suitable for modeling and analyzing algorithms like quicksort that operate on (user-defined) data structures like lists, and they were first extended to *probabilistic term rewrite systems* (PTRSs) in [20,21].

In the probabilistic setting, there are several notions of “termination”. One of the most important notions is *almost-sure termination* (AST), which means that the probability of termination is 1. A strictly stronger notion is *positive* AST (PAST), which requires that the expected runtime is finite. Moreover, there are several different rewrite strategies. For example, *innermost* rewriting only allows rewrite steps at innermost positions in a term. This strategy is the easiest to analyze for termination, and it corresponds to a call-by-value strategy as used by many programming languages. Thus, in this paper we will not only investigate AST for PTRSs, but we also consider innermost AST (iAST, allowing only innermost rewrite sequences).

\* Corresponding author.

E-mail addresses: [kassing@cs.rwth-aachen.de](mailto:kassing@cs.rwth-aachen.de) (J.-C. Kassing), [giesl@cs.rwth-aachen.de](mailto:giesl@cs.rwth-aachen.de) (J. Giesl).

**Algorithm 1:** Running Example.

---

```

1  $x \leftarrow 0$ 
2 while  $x = 0$  do
3   {
4      $x \leftarrow 0 \oplus_{1/2} x \leftarrow 1$ ;
5      $y \leftarrow 2 \cdot y$ ;
6   }  $\square$  {
7      $x \leftarrow 0 \oplus_{1/3} x \leftarrow 1$ ;
8      $y \leftarrow 3 \cdot y$ ;
9   }
10 while  $y > 0$  do
11    $y \leftarrow y - 1$ ;

```

---

As a running example in the first part of this paper (Sections 1–4), we transform Algorithm 1 (written in pGCL) into an equivalent PTRS and show how our framework proves AST. Here,  $\oplus_{1/2}$  denotes probabilistic choice, and  $\square$  denotes demonic non-determinism. In every iteration of the first loop, we have a chance of at least  $1/3$  to leave the first loop and enter the second one, which in turn terminates after a finite number of steps. Hence, the termination probability for Algorithm 1 is at least  $1/3 + 2/3 \cdot 1/3 + (2/3)^2 \cdot 1/3 + \dots = 1/3 \cdot \sum_{i=0}^{\infty} (2/3)^i = 1$ . Note that while Algorithm 1 is AST, its expected runtime is infinite, i.e., it is not PAST. This already holds for the program where only the first possibility of the first **while**-loop is considered (i.e., where  $y$  is always doubled in its body). Then for the initial value  $y = 1$ , the expected number of iterations of the second **while**-loop which decrements  $y$  is  $\frac{1}{2} \cdot 2 + \frac{1}{4} \cdot 4 + \frac{1}{8} \cdot 8 + \dots = 1 + 1 + 1 + \dots = \infty$ . In this paper, we use Algorithm 1 as a leading example to illustrate how to handle algorithms with non-determinism. While the non-deterministic choice does not affect the termination behavior of Algorithm 1, in general it can indeed influence termination.

There exist numerous techniques to prove (P)AST of imperative programs on numbers like Algorithm 1 (e.g., [22–33]). In fact, there are proof rules (e.g., [23]) and tools (e.g., [31]) that can prove AST for both loops of Algorithm 1 individually, and hence for the whole algorithm. Moreover, the tool Caesar [34] can prove AST if one provides super-martingales for the two loops. However, there are only few automated methods available to analyze (P)AST for programs with complex non-tail recursive structure [35–37]. The approaches that are suitable for algorithms on recursive data structures [38–40] are mostly tailored to specific structures, making it challenging to adapt them to other (possibly user-defined) data structures, or are not yet fully automated. To the best of our knowledge, there exist no automatic techniques to handle algorithms similar to Algorithm 1 that operate on arbitrary algebraic data structures, i.e., (non-deterministic) algorithms that first create a random data object  $y$  in a first loop and then access or modify it in a second loop using auxiliary functions, whereas this is possible with our *annotated dependency pair* framework.<sup>1</sup> In this paper we develop an approach for fully automatic termination analysis of (arbitrary) probabilistic TRSs that can successfully analyze probabilistic programs like the probabilistic quicksort algorithm mentioned above.

Up to now, there were only two approaches for automatic termination analysis of PTRSs. In [41], orders based on interpretations were adapted to prove PAST, and we presented a related technique to prove AST in [42]. However, already for non-probabilistic TRSs such a direct application of orders is limited in power. To obtain a powerful approach, one should combine such orders in a modular way, as in the DP framework. Thus, we presented a first version of a probabilistic DP framework for innermost rewriting in [42]. However, in contrast to the DP framework for ordinary TRSs, the framework of [42] was *incomplete*, i.e., there are PTRSs which are iAST but where this cannot be proved with DTs.

Therefore, in this paper we introduce the new concept of *annotated dependency pairs* (ADPs) to analyze termination of PTRSs. In this way, we obtain a novel complete criterion for iAST via DPs while maintaining soundness for all processors that were developed in the probabilistic framework of [42]. Moreover, our improvement allows for additional more powerful “transformational” probabilistic DP processors which were not possible in the framework of [42].

In [43] we showed that there exist classes of PTRSs where it suffices to analyze iAST in order to prove AST for full rewriting. However, these classes are quite restrictive (they exclude PTRSs with non-probabilistic non-determinism and impose linearity restrictions on both sides of the rewrite rules), e.g., a PTRS corresponding to Algorithm 1 would not be contained in this class. Thus, up to now there were no powerful techniques to prove full AST for more general forms of PTRSs. Therefore, in the current paper we show that our ADP framework can also be adapted to analyze full (instead of innermost) AST. In particular, our novel ADP framework can also be applied to prove AST for overlapping PTRSs and it weakens the linearity requirements considerably.

We start with preliminaries on (probabilistic) term rewriting and recapitulate the most important notions of the DP framework for non-probabilistic TRSs in Section 2. Then, we present our novel ADPs for probabilistic TRSs in Section 3. In Section 4, we explain the general idea of the probabilistic ADP framework, and present its fundamental processors. In addition, our framework allows for the definition of new processors which *transform* ADPs. To increase its power, we extend the probabilistic ADP framework by *transformational processors* in Section 5. The implementation of our approach in the tool AProVE is evaluated in Section 6. In Appendix A, we illustrate our approach on several examples, including PTRSs with non-numerical data structures like lists or trees. In the main part of the paper, we give proof sketches for all theorems and we refer to Appendix B for the full proofs.

<sup>1</sup> Such examples can be found in our benchmark set, see Section 6 and Appendix A.

The current paper is based on our earlier conference papers [44] and [45], and it extends them by:

- A unified formalization and presentation of the ADP framework with all processors for both innermost and full rewriting, whereas these strategies were previously handled in different frameworks for innermost [44] and full rewriting [45].
- A new more general version of the *reduction pair processor* (Section 4.4) that allows us to use arbitrary algebras where addition and the barycentric operation  $\mathbb{E}$  are defined (whereas this processor was restricted to polynomial interpretations in [44,45]).
- A new processor based on the *subterm criterion* of [12] (Section 4.6) in the ADP framework for innermost AST.
- A new discussion and counterexamples to show why the linearity requirement is needed for the *rewriting processor* in the ADP framework (Section 5.1).
- A new *instantiation processor* (Section 5.2), which is applicable both in the ADP framework for full and innermost AST.
- A new *forward instantiation processor* (Section 5.3), which is also applicable in the ADP framework for full and innermost AST.
- A new *rule overlap instantiation processor* (Section 5.4), which is applicable in the ADP framework for innermost AST.
- Proofs for all our lemmas and theorems.
- Numerous additional explanations, examples, and remarks.
- An improved implementation and evaluation with our tool AProVE.

## 2. Preliminaries

In this section, we recapitulate the basics of term rewriting and the DP framework for ordinary termination in Section 2.1, and probabilistic term rewriting in Section 2.2.

### 2.1. Term rewriting and the DP framework

We assume some familiarity with term rewriting [46], but recapitulate all important notions. We regard finite signatures  $\Sigma = \bigsqcup_{n \in \mathbb{N}} \Sigma_n$ , where  $\Sigma_n$  is the set of  $n$ -ary function symbols and we require  $\Sigma_0 \neq \emptyset$ .<sup>2</sup> Let  $\mathcal{T} = \mathcal{T}(\Sigma, \mathcal{V})$  denote the set of all terms over  $\Sigma$  and a set of variables  $\mathcal{V}$ , which is recursively defined as the smallest set such that  $\mathcal{V} \subseteq \mathcal{T}$ , and if  $f \in \Sigma_n$  and  $t_1, \dots, t_n \in \mathcal{T}$ , then  $f(t_1, \dots, t_n) \in \mathcal{T}$ . A *substitution* is a function  $\sigma : \mathcal{V} \rightarrow \mathcal{T}$  with  $\sigma(x) = x$  for all but finitely many  $x \in \mathcal{V}$ . We lift substitutions to terms by defining  $\sigma(t) = f(\sigma(t_1), \dots, \sigma(t_n))$  for a term  $t = f(t_1, \dots, t_n) \in \mathcal{T}$ , and we often write  $t\sigma$  instead of  $\sigma(t)$ . For a term  $t \in \mathcal{T}$ , the set of *positions*  $\text{Pos}(t)$  is the smallest subset of  $\mathbb{N}^*$  satisfying  $\varepsilon \in \text{Pos}(t)$ , and if  $t = f(t_1, \dots, t_n)$  then for all  $1 \leq j \leq n$  and all  $\pi \in \text{Pos}(t_j)$  we have  $j.\pi \in \text{Pos}(t)$ . Let  $<_{\mathbb{N}^*}$  be the *prefix ordering* on positions and let  $\leq_{\mathbb{N}^*}$  be its reflexive closure. We say that a position  $\pi$  is *above* (strictly above) another position  $\tau$  if  $\pi \leq_{\mathbb{N}^*} \tau$  ( $\pi <_{\mathbb{N}^*} \tau$ ), i.e.,  $\pi$  is a (strict) prefix of  $\tau$ . Two positions  $\pi$  and  $\tau$  are *orthogonal* if both  $\pi \not\leq_{\mathbb{N}^*} \tau$  and  $\tau \not\leq_{\mathbb{N}^*} \pi$  hold.

If  $\pi \in \text{Pos}(t)$  then  $t|_{\pi}$  denotes the *subterm* at position  $\pi$ , i.e.,  $t|_{\varepsilon} = t$  for the *root position*  $\varepsilon$  and  $t|_{j.\pi} = t_j|_{\pi}$  for  $t = f(t_1, \dots, t_n)$ ,  $1 \leq j \leq n$ , and  $\pi \in \text{Pos}(t_j)$ . The *root symbol* (or variable) at position  $\varepsilon$  is also denoted by  $\text{root}(t) \in \Sigma \cup \mathcal{V}$ . We write  $t \triangleright s$  if  $s$  is a subterm of  $t$  and  $t \triangleright s$  if  $s$  is a *proper* subterm of  $t$  (i.e., if  $t \triangleright s$  and  $t \neq s$ ). Let  $t[r]_{\pi}$  denote the term that results from replacing the subterm  $t|_{\pi}$  at position  $\pi$  with the term  $r \in \mathcal{T}$ , i.e.,  $t[r]_{\varepsilon} = r$  and  $t[r]_{j.\pi} = f(t_1, \dots, t_j[r]_{\pi}, \dots, t_n)$  for  $t = f(t_1, \dots, t_n)$ ,  $1 \leq j \leq n$ , and  $\pi \in \text{Pos}(t_j)$ . A *context* is a term  $C \in \mathcal{T}(\Sigma \uplus \{\square\}, \mathcal{V})$  containing the 0-ary function symbol  $\square$  exactly once. For a context  $C$  and a term  $t$ , let  $C[t]$  denote the term resulting from  $C$  by replacing the only occurrence of  $\square$  by  $t$ .

A *rewrite rule*  $\ell \rightarrow r$  is a pair of terms  $\ell, r \in \mathcal{T}$  such that  $\mathcal{V}(r) \subseteq \mathcal{V}(\ell)$  and  $\ell \notin \mathcal{V}$ , where  $\mathcal{V}(t)$  denotes the set of all variables occurring in  $t \in \mathcal{T}$ . The restrictions  $\mathcal{V}(r) \subseteq \mathcal{V}(\ell)$  and  $\ell \notin \mathcal{V}$  are imposed to exclude trivially non-terminating systems. Here,  $\ell$  is called the *left-hand side* and  $r$  the *right-hand side* of the rule. A *term rewrite system* (TRS) is a finite set of rewrite rules. As an example, consider  $\mathcal{R}_{\text{ex}} = \{(1), (2)\}$ .

$$f(s(x)) \rightarrow c(f(g(x))) \quad (1)$$

$$g(x) \rightarrow s(x) \quad (2)$$

A TRS  $\mathcal{R}$  induces a *rewrite relation*  $\rightarrow_{\mathcal{R}} \subseteq \mathcal{T} \times \mathcal{T}$  on terms where  $s \rightarrow_{\mathcal{R}} t$  holds if there is a position  $\pi \in \text{Pos}(s)$ , a rule  $\ell \rightarrow r \in \mathcal{R}$ , and a substitution  $\sigma$  such that  $s|_{\pi} = \ell\sigma$  (i.e.,  $\ell$  matches  $s|_{\pi}$ ) and  $t = s[r\sigma]_{\pi}$ . A (full) rewrite step  $s \rightarrow_{\mathcal{R}} t$  is an *innermost* rewrite step (denoted  $s \xrightarrow{i}_{\mathcal{R}} t$ ) if all proper subterms of the used *redex*  $\ell\sigma$  are in *normal form* w.r.t.  $\mathcal{R}$ , i.e., the proper subterms of  $\ell\sigma$  do not contain redexes themselves and thus, they cannot be reduced with  $\rightarrow_{\mathcal{R}}$ . For example, we have  $f(g(x)) \xrightarrow{i}_{\mathcal{R}_{\text{ex}}} f(s(x))$ . Let  $\text{NF}_{\mathcal{R}}$  denote the set of all terms that are in normal form w.r.t.  $\mathcal{R}$ , and let  $\text{ANF}_{\mathcal{R}}$  denote the set of all terms that are in *argument normal form* w.r.t.  $\mathcal{R}$ , i.e., all terms where every proper subterm is in normal form.

Two rules  $\ell_1 \rightarrow r_1, \ell_2 \rightarrow r_2 \in \mathcal{R}$  with renamed variables such that  $\mathcal{V}(\ell_1) \cap \mathcal{V}(\ell_2) = \emptyset$  are *overlapping* if there exists a non-variable position  $\pi$  of  $\ell_1$  such that  $\ell_1|_{\pi} = \ell_2$  and  $\ell_2$  are *unifiable*, i.e., there exists a substitution  $\sigma$  such that  $\ell_1|_{\pi}\sigma = \ell_2\sigma$ . If  $(\ell_1 \rightarrow r_1) = (\ell_2 \rightarrow r_2)$ , then we require that  $\pi \neq \varepsilon$ .  $\mathcal{R}$  is *non-overlapping* (NO) if it has no overlapping rules. As an example, the TRS  $\mathcal{R}_{\text{ex}}$  is non-overlapping. A TRS is *left linear* (*right linear*) if every variable occurs at most once in the left-hand side (right-hand side) of a rule, and it is *linear* if it is both left and right linear. Finally, a TRS is *non-duplicating* if for every rule, every variable occurs at most as often in the right-hand side as in the left-hand side. As an example,  $\mathcal{R}_{\text{ex}}$  is both left and right linear, and hence non-duplicating.

<sup>2</sup> This requirement ensures that there exist ground terms. This is not a restriction, since we showed in [43] that almost-sure termination of PTRSs is preserved under signature extensions.

We call a TRS  $\mathcal{R}$  *strongly normalizing* (or *terminating*) or SN for short if  $\rightarrow_{\mathcal{R}}$  is well founded.  $\mathcal{R}$  is *strongly innermost normalizing* (or *innermost terminating*) or  $\text{iSN}$  for short if  $\xrightarrow{*}_{\mathcal{R}}$  is well founded.

Next, we recapitulate the DP framework with its core processors to analyze (innermost) termination (see, e.g., [8–12] for further details). We decompose the signature  $\Sigma = D \uplus C$  such that  $f \in D$  if  $f = \text{root}(\ell)$  for some  $\ell \rightarrow r \in \mathcal{R}$ . The symbols in  $D$  are called *defined symbols* (representing executable functions) and the symbols in  $C$  are *constructors* (representing data). For  $\mathcal{R}_{\text{ex}}$  we have  $D = \{f, g\}$  and  $C = \{s, c\}$ . For every  $f \in D$  we introduce a fresh *annotated symbol*  $f^{\sharp}$  of the same arity.<sup>3</sup> Let  $D^{\sharp}$  denote the set of all annotated symbols,  $\Sigma^{\sharp} = D^{\sharp} \uplus \Sigma$ , and  $\mathcal{T}^{\sharp} = \mathcal{T}(\Sigma^{\sharp}, \mathcal{V})$ . For any  $t = f(t_1, \dots, t_n) \in \mathcal{T}$  with  $f \in D$ , let  $t^{\sharp} = f^{\sharp}(t_1, \dots, t_n)$ . The annotated symbols are used to compare the arguments of two subsequent “function calls”. For example, if we want to prove termination of a TRS containing only the rule  $\text{plus}(s(x), y) \rightarrow s(\text{plus}(x, y))$ , then we need to show that the arguments of the defined function symbol  $\text{plus}$  “decrease” w.r.t. some well-founded order in every recursive call. To avoid handling tuples of arguments, we use the annotated symbol  $\text{plus}^{\sharp}$  to combine the arguments of  $\text{plus}$ , i.e., we compare  $\text{plus}^{\sharp}(s(x), y)$  with  $\text{plus}^{\sharp}(x, y)$ . More precisely, for every rule  $\ell \rightarrow r$  and every (not necessarily proper) subterm  $t$  of  $r$  with defined root symbol, one obtains a *dependency pair* (DP)  $\ell^{\sharp} \rightarrow t^{\sharp}$  that represents two possible subsequent “function calls” in a rewrite sequence.  $DP(\mathcal{R})$  denotes the set of all dependency pairs of  $\mathcal{R}$ . As an example, we have  $DP(\mathcal{R}_{\text{ex}}) = \{(3), (4)\}$ . To ease readability, we often use capital letters like  $F$  instead of  $f^{\sharp}$ .

$$F(s(x)) \rightarrow F(g(x)) \tag{3}$$

$$F(s(x)) \rightarrow G(x) \tag{4}$$

The DP framework uses *DP problems*  $(P, \mathcal{R})$  where  $P$  is a (finite) set of DPs and  $\mathcal{R}$  is a TRS. A (possibly infinite) sequence  $t_0^{\sharp}, t_1^{\sharp}, t_2^{\sharp}, \dots$  with  $t_j^{\sharp} \xrightarrow{*}_{\mathcal{R}} t_{j+1}^{\sharp}$  for all  $j$  is a  $(P, \mathcal{R})$ -*chain* which represents a sequence of subsequent “function calls” in evaluations. Here, for any binary relation  $\rightarrow$ , “ $\rightarrow^*$ ” denotes its reflexive and transitive closure. Moreover, “ $\xrightarrow{*}_P$ ” denotes rewrite steps with DPs from  $P$  at the root and “ $\circ$ ” denotes composition of relations. We refer to steps with  $\xrightarrow{*}_P$  as *a-steps* (a for *annotation*, as we rewrite at the position of an annotated symbol, since rewrite steps with  $\xrightarrow{*}_P$  require an annotated symbol at the root), and we call steps with  $\rightarrow_{\mathcal{R}}$  *n-steps* (n for *no-annotation*, as we rewrite at the position of a symbol without annotation). The latter are used to evaluate the arguments of an annotated function symbol. So an infinite chain consists of an infinite number of a-steps with a finite number of n-steps between consecutive a-steps. A (possibly infinite) sequence  $t_0^{\sharp}, t_1^{\sharp}, t_2^{\sharp}, \dots$  with  $t_j^{\sharp} \xrightarrow{*}_{\mathcal{R}} t_{j+1}^{\sharp}$  for all  $j$  is an *innermost*  $(P, \mathcal{R})$ -*chain*. Here, “ $\xrightarrow{*}_{P, \mathcal{R}}$ ” is the restriction of  $\xrightarrow{*}_P$  to rewrite steps where the used redex is in  $\text{NF}_{\mathcal{R}}$ . For example,  $F(s(x)), F(s(x)), \dots$  is an infinite innermost  $(DP(\mathcal{R}_{\text{ex}}), \mathcal{R}_{\text{ex}})$ -chain, as  $F(s(x)) \xrightarrow{*}_{DP(\mathcal{R}_{\text{ex}}), \mathcal{R}_{\text{ex}}} F(g(x)) \xrightarrow{*}_{\mathcal{R}_{\text{ex}}} F(s(x))$ .

A DP problem  $(P, \mathcal{R})$  is called *terminating* (or SN) if there is no infinite  $(P, \mathcal{R})$ -chain. It is *innermost terminating* (or  $\text{iSN}$ ) if there is no infinite innermost  $(P, \mathcal{R})$ -chain. The main result on DPs is the *chain criterion* which states that there is no infinite (innermost) sequence  $t_0 \xrightarrow{\text{i}}_{\mathcal{R}} t_1 \xrightarrow{\text{i}}_{\mathcal{R}} \dots$  (i.e.,  $\mathcal{R}$  is  $\text{iSN}$ ) iff there is no infinite (innermost)  $(DP(\mathcal{R}), \mathcal{R})$ -chain (i.e.,  $(DP(\mathcal{R}), \mathcal{R})$  is  $\text{iSN}$ ). The DP framework is a *divide-and-conquer* approach, which applies *DP processors*  $\text{Proc}$  of the form  $\text{Proc}(P, \mathcal{R}) = \{(P_1, \mathcal{R}_1), \dots, (P_n, \mathcal{R}_n)\}$ , where  $P, P_1, \dots, P_n$  are (finite) sets of DPs and  $\mathcal{R}, \mathcal{R}_1, \dots, \mathcal{R}_n$  are TRSs, to transform DP problems into simpler sub-problems. A processor  $\text{Proc}$  is *sound* for termination (for innermost termination) if  $(P, \mathcal{R})$  is SN ( $\text{iSN}$ ) whenever  $(P_j, \mathcal{R}_j)$  is SN ( $\text{iSN}$ ) for all  $1 \leq j \leq n$ . It is *complete* for termination (for innermost termination) if  $(P_j, \mathcal{R}_j)$  is SN ( $\text{iSN}$ ) for all  $1 \leq j \leq n$  whenever  $(P, \mathcal{R})$  is SN ( $\text{iSN}$ ).

So given a TRS  $\mathcal{R}$ , one starts with the initial DP problem  $(DP(\mathcal{R}), \mathcal{R})$  and applies sound (and preferably complete) DP processors repeatedly until all sub-problems are “solved”, i.e., sound processors transform them to the empty set. This yields a modular framework for termination and innermost termination proofs, as different techniques can be used for different sub-problems  $(P_j, \mathcal{R}_j)$ . To give an intuition for the framework, the following three theorems recapitulate the three most important processors of the DP framework. Later in the paper, we will also introduce additional processors which we adapt to the probabilistic setting.

The (innermost)  $(P, \mathcal{R})$ -*dependency graph* is a control flow graph that indicates which DPs can be used after each other in an (innermost) chain. Its set of nodes is  $P$  and there is an edge from  $\ell_1^{\sharp} \rightarrow t_1^{\sharp}$  to  $\ell_2^{\sharp} \rightarrow t_2^{\sharp}$  if there exist substitutions  $\sigma_1, \sigma_2$  such that  $t_1^{\sharp} \sigma_1 \xrightarrow{*}_{\mathcal{R}} \ell_2^{\sharp} \sigma_2$ ,  $t_1^{\sharp} \sigma_1 \xrightarrow{\text{i}}_{\mathcal{R}} \ell_2^{\sharp} \sigma_2$  and  $\ell_1^{\sharp} \sigma_1, \ell_2^{\sharp} \sigma_2 \in \text{NF}_{\mathcal{R}}$  for the innermost dependency graph). Any infinite (innermost)  $(P, \mathcal{R})$ -chain corresponds to an infinite path in the (innermost) dependency graph, and since the graph is finite, this infinite path must end in some strongly connected component (SCC).<sup>4</sup> Hence, it suffices to consider the SCCs of this graph independently.

**Theorem 1** (Dependency Graph Processor). *Let  $P_1, \dots, P_n$  be the SCCs of the (innermost)  $(P, \mathcal{R})$ -dependency graph. Then  $\text{Proc}_{\text{DG}}(P, \mathcal{R}) = \{(P_1, \mathcal{R}), \dots, (P_n, \mathcal{R})\}$  is sound and complete for (innermost) termination.*

**Example 1** (Dependency Graph). Consider the TRS  $\mathcal{R}_{\text{ffg}} = \{(5)\}$  with  $DP(\mathcal{R}_{\text{ffg}}) = \{(6), (7), (8)\}$  and the (innermost)  $(DP(\mathcal{R}_{\text{ffg}}), \mathcal{R}_{\text{ffg}})$ -dependency graph depicted in Fig. 1. Here,  $\text{Proc}_{\text{DG}}(DP(\mathcal{R}_{\text{ffg}}), \mathcal{R}_{\text{ffg}})$  yields the DP problem  $(\{(8)\}, \mathcal{R}_{\text{ffg}})$ .

$$f(f(g(x))) \rightarrow f(g(f(g(f(x)))))) \tag{5}$$

$$F(f(g(x))) \rightarrow F(g(f(g(f(x)))))) \tag{6}$$

$$F(f(g(x))) \rightarrow F(g(f(x))) \tag{7}$$

$$F(f(g(x))) \rightarrow F(x) \tag{8}$$

<sup>3</sup> The symbols  $f^{\sharp}$  were also called *tuple symbols* (or *marked symbols*) in the literature.

<sup>4</sup> Here, a set  $P'$  of DPs is an SCC if it is a maximal cycle, i.e., it is a maximal set such that for any  $\ell_1^{\sharp} \rightarrow t_1^{\sharp}$  and  $\ell_2^{\sharp} \rightarrow t_2^{\sharp}$  in  $P'$  there is a non-empty path from  $\ell_1^{\sharp} \rightarrow t_1^{\sharp}$  to  $\ell_2^{\sharp} \rightarrow t_2^{\sharp}$  which only traverses nodes from  $P'$ .

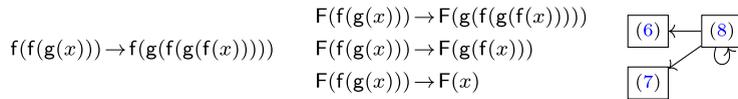


Fig. 1.  $(DP(\mathcal{R}_{ffg}), \mathcal{R}_{ffg})$ -dependency graph.

While the exact (innermost) dependency graph is not computable in general, there exist several techniques to over-approximate it automatically, see, e.g., [8,10,11]. A basic approximation technique that we will also use for the novel transformational processors in Section 5 works as follows. To find out whether there is an edge from  $\ell_1^\# \rightarrow t_1^\#$  to  $\ell_2^\# \rightarrow t_2^\#$  we compute the skeleton  $Cap_{\mathcal{R}}(t_1^\#)$  of  $t_1^\#$  that remains unchanged when reducing  $\ell_1^\# \sigma_1$  to  $\ell_2^\# \sigma_2$  for arbitrary substitutions  $\sigma_1$  and  $\sigma_2$ .  $Cap_{\mathcal{R}}(t_1^\#)$  results from replacing all those subterms of  $t_1^\#$  by different fresh variables whose root is a defined symbol of  $\mathcal{R}$ . Here, multiple occurrences of the same subterm are also replaced by pairwise different variables. So if  $f \in \mathcal{D}$  and  $c \in \mathcal{C}$ , then  $Cap_{\mathcal{R}}(c(f(x), f(x))) = c(x_1, x_2)$ . Afterwards, we check whether the skeleton  $Cap_{\mathcal{R}}(t_1^\#)$  and  $\ell_2^\#$  are unifiable by a substitution  $\sigma$  where  $\ell_1^\# \sigma$  and  $\ell_2^\# \sigma$  are in  $ANF_{\mathcal{R}}$  (and thus, also in  $NF_{\mathcal{R}}$ ). If they are, then we create an edge from  $\ell_1^\# \rightarrow t_1^\#$  to  $\ell_2^\# \rightarrow t_2^\#$  in the estimated innermost dependency graph. If they cannot be unified, then there is no innermost rewrite sequence from an instantiation of  $t_1^\#$  to an instantiation of  $t_2^\#$  such that both instantiated left-hand sides  $\ell_1^\# \sigma$  and  $\ell_2^\# \sigma$  are in argument normal form. Thus, this leads to a sound over-approximation of the innermost dependency graph.

To check whether we can reduce  $t_1^\# \sigma_1$  to  $\ell_2^\# \sigma_2$  in non-innermost chains, we add another transformation  $Ren$  that replaces all variable occurrences (including multiple occurrences of the same variable) by fresh variables. The reason for this additional transformation is that for non-innermost chains, variables in  $t_1^\#$  may be instantiated by terms that are not in normal form. Hence, they can be evaluated to different terms during the reduction of the arguments of  $t_1^\#$ . Thus, we check unifiability of  $Ren(Cap_{\mathcal{R}}(t_1^\#))$  and  $\ell_2^\#$  for the (non-innermost) dependency graph.

The next processor removes rules that cannot be used for right-hand sides of dependency pairs when their variables are instantiated with normal forms.

**Theorem 2** (Usable Rules Processor). *Let  $\mathcal{R}$  be a TRS. For every  $f \in \Sigma^\#,$  let  $Rules_{\mathcal{R}}(f) = \{\ell \rightarrow r \in \mathcal{R} \mid \text{root}(\ell) = f\}.$  Moreover, for every  $t \in \mathcal{T}(\Sigma^\#, \mathcal{V}),$  the usable rules  $\mathcal{U}_{\mathcal{R}}(t)$  of  $t$  w.r.t.  $\mathcal{R}$  are recursively defined as follows:*

$$\begin{aligned}
 \mathcal{U}_{\mathcal{R}}(t) &= \emptyset, & \text{if } t \in \mathcal{V} \text{ or } \mathcal{R} = \emptyset \\
 \mathcal{U}_{\mathcal{R}}(f(t_1, \dots, t_n)) &= Rules_{\mathcal{R}}(f) \cup \bigcup_{j=1}^n \mathcal{U}_{\mathcal{R}}(t_j) \cup \bigcup_{\ell \rightarrow r \in Rules_{\mathcal{R}}(f)} \mathcal{U}_{\mathcal{R}}(r), & \text{otherwise,}
 \end{aligned}$$

where  $\mathcal{R}' = \mathcal{R} \setminus Rules_{\mathcal{R}}(f).$

The usable rules for the DP problem  $(\mathcal{P}, \mathcal{R})$  are  $\mathcal{U}(\mathcal{P}, \mathcal{R}) = \bigcup_{\ell^\# \rightarrow t^\# \in \mathcal{P}} \mathcal{U}_{\mathcal{R}}(t^\#).$  Then  $Proc_{UR}(\mathcal{P}, \mathcal{R}) = \{(\mathcal{P}, \mathcal{U}(\mathcal{P}, \mathcal{R}))\}$  is sound but not complete for innermost termination.<sup>5</sup>

To continue Example 1 for innermost rewriting,  $Proc_{UR}(\{(8)\}, \mathcal{R}_{ffg})$  yields the problem  $(\{(8)\}, \emptyset),$  i.e., it removes all rules, because the right-hand side of (8) does not contain the defined symbol  $f.$

Finally, the reduction pair processor allows us to use term orders to remove certain DPs. As our probabilistic adaption in Section 4 needs to consider “expected values”, we focus on reduction pairs derived from a  $\Sigma^\#$ -algebra. An (ordered)  $\Sigma^\#$ -algebra is a pair  $(\mathcal{A}, >),$  where  $\mathcal{A}$  maps every function symbol  $f \in \Sigma_n^\#$  of arity  $n$  to a function  $\mathcal{A}(f) = f_{\mathcal{A}} : \mathcal{A}^n \rightarrow \mathcal{A}$  for a non-empty carrier set  $\mathcal{A}.$  The mapping  $\mathcal{A}$  is extended to ground terms, i.e.,  $\mathcal{A}(f(t_1, \dots, t_n)) = f_{\mathcal{A}}(\mathcal{A}(t_1), \dots, \mathcal{A}(t_n))$  for  $f \in \Sigma_n^\#$  and ground terms  $t_1, \dots, t_n \in \mathcal{T}^\#.$  Furthermore,  $>$  is a well-founded order on  $\mathcal{A},$  i.e.,  $>$  is well founded and transitive. Now one can obtain a well-founded order on ground terms by considering  $t_1$  to be greater than  $t_2$  iff  $\mathcal{A}(t_1) > \mathcal{A}(t_2).$  We extend this to a well-founded order on terms with variables by writing  $\mathcal{A}(t_1) > \mathcal{A}(t_2)$  iff  $\mathcal{A}(t_1 \sigma) > \mathcal{A}(t_2 \sigma)$  holds for all substitutions  $\sigma$  that instantiate the variables of  $t_1$  and  $t_2$  by ground terms. Let  $\geq$  be the reflexive closure of  $>.$  We say that  $(\mathcal{A}, >)$  is *monotonic* if  $\mathcal{A}(t_1) > \mathcal{A}(t_2)$  implies  $\mathcal{A}(C[t_1]) > \mathcal{A}(C[t_2]),$  and *weakly monotonic* if  $\mathcal{A}(t_1) \geq \mathcal{A}(t_2)$  implies  $\mathcal{A}(C[t_1]) \geq \mathcal{A}(C[t_2])$  for every context  $C.$  The reduction pair processor searches for a weakly monotonic  $\Sigma^\#$ -algebra  $(\mathcal{A}, >)$  such that all rules and DPs are ordered by  $\geq,$  and it removes those DPs that are ordered by  $>.$

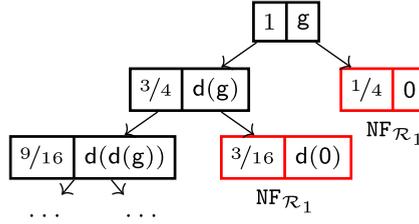
**Theorem 3** (Reduction Pair Processor). *Let  $(\mathcal{A}, >)$  be a weakly monotonic  $\Sigma^\#$ -algebra. Let  $\mathcal{P} = \mathcal{P}_{\geq} \uplus \mathcal{P}_{>} \text{ with } \mathcal{P}_{>} \neq \emptyset \text{ where:}$*

- (1)  $\forall \ell^\# \rightarrow r^\# \in \mathcal{P} : \mathcal{A}(\ell^\#) \geq \mathcal{A}(r^\#)$
- (2)  $\forall \ell^\# \rightarrow r^\# \in \mathcal{P}_{>} : \mathcal{A}(\ell^\#) > \mathcal{A}(r^\#)$
- (3)  $\forall \ell \rightarrow r \in \mathcal{R} : \mathcal{A}(\ell) \geq \mathcal{A}(r)$

Then  $Proc_{RP}(\mathcal{P}, \mathcal{R}) = \{(\mathcal{P}_{\geq}, \mathcal{R})\}$  is sound and complete for both termination and innermost termination.

Examples for  $\Sigma^\#$ -algebras are *polynomial* [47] or *matrix interpretations* [48]. A polynomial interpretation  $(\mathcal{A}, >) = (\text{Pol}, >)$  maps every function symbol  $f \in \Sigma^\#$  to a polynomial  $f_{\text{Pol}} \in \mathbb{N}[\mathcal{V}]$  over the variables  $\mathcal{V}$  with coefficients from  $\mathbb{N},$  see [47]. The order  $>$  is simply the natural order on  $\mathbb{N}.$  The reduction pair processor allows us to use weakly monotonic polynomial interpretations that do not

<sup>5</sup> See [9] for a complete version of this processor. It extends DP problems by an additional set which stores the left-hand sides of all rules (including the non-usable ones) to determine whether a rewrite step is innermost. We omit this refinement here for readability.

Fig. 2. Infinite  $\mathcal{R}_1$ -RST.

have to depend on all of their arguments, and SMT solvers can be used to search for an appropriate interpretation automatically. For  $(\{(8)\}, \emptyset)$ , one can use the reduction pair processor with the polynomial interpretation that maps  $f(x)$  to  $x + 1$  and both  $F(x)$  and  $g(x)$  to  $x$ . Then,  $\text{Proc}_{\text{RP}}(\{(8)\}, \emptyset) = \{(\emptyset, \emptyset)\}$ . As  $\text{Proc}_{\text{DG}}(\emptyset, \emptyset) = \emptyset$  and all processors used are sound, this means that there is no infinite innermost chain for the initial DP problem  $(DP(\mathcal{R}_{\text{ffg}}), \mathcal{R}_{\text{ffg}})$  and thus,  $\mathcal{R}_{\text{ffg}}$  is innermost terminating. This implies that  $\mathcal{R}_{\text{ffg}}$  is terminating, since  $\mathcal{R}_{\text{ffg}}$  is non-overlapping, hence, innermost termination implies termination [49].

## 2.2. Probabilistic rewriting

Next, we recapitulate the basics of probabilistic rewriting as introduced in [20,21,41,42]. A *probabilistic rule* has (finite) multi-distributions on the right-hand sides of its rewrite rules instead of a single term. A finite *multi-distribution*  $\mu$  on a set  $A \neq \emptyset$  is a finite multiset of pairs  $(p : a)$ , where  $0 < p \leq 1$  is a probability and  $a \in A$ , such that  $\sum_{(p:a) \in \mu} p = 1$ .  $\text{FDist}(A)$  is the set of all finite multi-distributions on  $A$ . For  $\mu \in \text{FDist}(A)$ , its *support* is the multiset  $\text{Supp}(\mu) = \{a \mid (p : a) \in \mu \text{ for some } p\}$ . A *probabilistic rewrite rule*  $\ell \rightarrow \mu \in \mathcal{T} \times \text{FDist}(\mathcal{T})$  is a pair such that  $\ell \notin \mathcal{V}$  and  $\mathcal{V}(r) \subseteq \mathcal{V}(\ell)$  for every  $r \in \text{Supp}(\mu)$ . A *probabilistic TRS* (PTRS) is a finite set  $\mathcal{R}$  of probabilistic rewrite rules.

### Example 2.

As running examples, we use the PTRSs  $\mathcal{R}_1 = \{(9)\}$ ,  $\mathcal{R}_2 = \{(9), (11)\}$ , and  $\mathcal{R}_3 = \{(9), (10), (12)\}$  with

$$g \rightarrow \{3/4 : d(g), 1/4 : 0\} \quad (9)$$

$$d(x) \rightarrow \{1 : 0\} \quad (10)$$

$$d(x) \rightarrow \{1 : c(x, x)\} \quad (11)$$

$$d(d(x)) \rightarrow \{1 : c(x, g)\} \quad (12)$$

Similar to TRSs, a PTRS  $\mathcal{R}$  induces a *probabilistic rewrite relation*  $\rightarrow_{\mathcal{R}} \subseteq \mathcal{T} \times \text{FDist}(\mathcal{T})$  where  $s \rightarrow_{\mathcal{R}} \{p_1 : t_1, \dots, p_k : t_k\}$  if there are an  $\ell \rightarrow \{p_1 : r_1, \dots, p_k : r_k\} \in \mathcal{R}$ , a substitution  $\sigma$ , and a  $\pi \in \text{Pos}(s)$  such that  $s|_{\pi} = \ell \sigma$  and  $t_j = s[r_j \sigma]_{\pi}$  for all  $1 \leq j \leq k$ .  $\text{NF}_{\mathcal{R}}$  and  $\text{ANF}_{\mathcal{R}}$  are defined as for TRSs. The rewrite step is *innermost* (denoted  $s \xrightarrow{i}_{\mathcal{R}} \{p_1 : r_1, \dots, p_k : r_k\}$ ) if  $\ell \sigma \in \text{ANF}_{\mathcal{R}}$ . So the PTRS  $\mathcal{R}_1$  can be interpreted as a biased coin flip that terminates in each step with a probability of  $1/4$ .

To track all possible rewrite sequences (up to non-determinism) with their corresponding probabilities, we *lift*  $\rightarrow_{\mathcal{R}}$  to *rewrite sequence trees* (RSTs).

**Definition 1** (Rewrite Sequence Tree, Almost-Sure Termination). Let  $\mathcal{R}$  be a PTRS. We say that  $\mathfrak{T} = (V, E, L)$  is an  $\mathcal{R}$ -*rewrite sequence tree* ( $\mathcal{R}$ -RST) if

- (1)  $(V, E)$  is a (possibly infinite) directed tree with nodes  $V \neq \emptyset$  and directed edges  $E \subseteq V \times V$ , where  $vE = \{w \mid (v, w) \in E\}$  is finite for every  $v \in V$ .
- (2)  $L : V \rightarrow (0, 1] \times \mathcal{T}$  labels every node  $v$  by a probability  $p_v$  and a term  $t_v$ . For the root  $v \in V$  of the tree, we have  $p_v = 1$ .
- (3) For all  $v \in V$ : If  $vE = \{w_1, \dots, w_k\}$ , then  $t_v \rightarrow_{\mathcal{R}} \{\frac{p_{w_1}}{p_v} : t_{w_1}, \dots, \frac{p_{w_k}}{p_v} : t_{w_k}\}$ .

An  $\mathcal{R}$ -RST is an *innermost*  $\mathcal{R}$ -RST if the edge relation represents only innermost steps. We say that an RST  $\mathfrak{T}$  *converges* (or *terminates*) with probability  $|\mathfrak{T}| = \sum_{v \in \text{Leaf}} p_v$ , where  $\text{Leaf}$  is the set of all its leaves. A PTRS  $\mathcal{R}$  is *almost surely (innermost) terminating* (or (i)AST for short) if  $|\mathfrak{T}| = 1$  holds for all (innermost)  $\mathcal{R}$ -RSTs  $\mathfrak{T}$ .

While  $|\mathfrak{T}| = 1$  for every finite RST  $\mathfrak{T}$ , for infinite RSTs  $\mathfrak{T}$  we may have  $|\mathfrak{T}| < 1$  or even  $|\mathfrak{T}| = 0$  if  $\mathfrak{T}$  has no leaf at all. The notion of AST from Definition 1 is equivalent to the ones in [21,41,42], where AST is defined via a lifting of  $\rightarrow_{\mathcal{R}}$  to multisets or via stochastic processes. For example, we have  $|\mathfrak{T}| = 1$  for the infinite  $\mathcal{R}_1$ -RST  $\mathfrak{T}$  in Fig. 2. As this holds for all  $\mathcal{R}_1$ -RSTs,  $\mathcal{R}_1$  is AST. In Fig. 2, we split the nodes  $v$  of the RST into the corresponding probability  $p_v$  and the corresponding term  $t_v$  to ease readability instead of writing  $(p_v : t_v)$  in a single node. Moreover, we indicated nodes  $v$  where the corresponding term  $t_v$  is in normal form.

**Example 3.**  $\mathcal{R}_2$  from Example 2 is not AST. If we always apply (11) directly after (9) (in a non-innermost step), then this corresponds to the rule

$$g \rightarrow \{3/4 : c(g, g), 1/4 : 0\} \quad (13)$$

which represents a random walk on the number of  $g$ 's in a term biased towards non-termination (as  $3/4 > 1/4$ ).  $\mathcal{R}_3$  is not AST either, because if we always apply (12) after two applications of (9), this corresponds to

$$g \rightarrow \{9/16 : c(g, g), 3/16 : 0, 1/4 : 0\}$$

which is a biased random walk towards non-termination (as  $9/16 > 3/16 + 1/4$ ).

However, in innermost evaluations, the  $d$ -rule (11) can only duplicate normal forms, i.e., we cannot apply the rule (11) directly after (9), and hence  $\mathcal{R}_2$  is  $\text{iAST}$ , see [43].  $\mathcal{R}_3$  is  $\text{iAST}$  as well, as (12) is not applicable in innermost evaluations. For both  $\mathcal{R}_2$  and  $\mathcal{R}_3$ , we will see that  $\text{iAST}$  can be proved automatically by our implementation of the ADP framework for  $\text{iAST}$  introduced in Section 4.

In Example 3 we can see the reason for using multi-distributions instead of distributions. Consider  $g \rightarrow \{9/16 : c(g, g), 3/16 : 0, 1/4 : 0\}$  again. If we had two different rules that can rewrite 0, e.g.,  $0 \rightarrow \{1 : a\}$  and  $0 \rightarrow \{1 : b\}$ , then we would be able to rewrite both occurrences of 0 differently. Hence, we do not “merge” those terms in the multi-distribution to  $\{9/16 : c(g, g), 3/16 + 1/4 : 0\}$ .

**Example 4.** The following PTRS  $\mathcal{R}_{\text{alg}}$  corresponds to Algorithm 1. Here, natural numbers are represented via the constructors 0 and  $s$ , and the non-determinism is modeled by the non-deterministic choice between the overlapping rules (14) and (15). In Section 4, we will prove that  $\mathcal{R}_{\text{alg}}$  is AST.

$$\text{loop1}(y) \rightarrow \{1/2 : \text{loop1}(\text{double}(y)), 1/2 : \text{loop2}(\text{double}(y))\} \quad (14)$$

$$\text{loop1}(y) \rightarrow \{1/3 : \text{loop1}(\text{triple}(y)), 2/3 : \text{loop2}(\text{triple}(y))\} \quad (15)$$

$$\text{loop2}(s(y)) \rightarrow \{1 : \text{loop2}(y)\} \quad (16)$$

$$\text{double}(s(y)) \rightarrow \{1 : s(\text{double}(y))\} \quad (17)$$

$$\text{double}(0) \rightarrow \{1 : 0\} \quad (18)$$

$$\text{triple}(s(y)) \rightarrow \{1 : s(s(\text{triple}(y)))\} \quad (19)$$

$$\text{triple}(0) \rightarrow \{1 : 0\} \quad (20)$$

### 3. Probabilistic annotated dependency pairs

In this section, we extend DPs to the probabilistic setting, which results in the notion of *annotated dependency pairs*. Compared to our first adaption of DPs to the probabilistic setting via dependency tuples in [42], ADPs are easier, more elegant, and lead to a sound and *complete* chain criterion. Note that ordinary DPs do not suffice in the probabilistic setting. Similar to the dependency tuples for complexity analysis [50], we cannot consider each defined symbol in a right-hand side individually, but we have to consider all of them simultaneously.

**Example 5.** A natural idea to define dependency pairs for a probabilistic rule  $\ell \rightarrow \{p_1 : r_1, \dots, p_k : r_k\} \in \mathcal{R}$  would be (21) or (22):

$$\{\ell^\# \rightarrow \{p_1 : r_1, \dots, p_i : t_j^\#, \dots, p_k : r_k\} \mid \exists j \in \{1, \dots, k\} : t_j \preceq r_j, \text{root}(t_j) \in \mathcal{D}\} \quad (21)$$

$$\{\ell^\# \rightarrow \{p_1 : t_1^\#, \dots, p_k : t_k^\#\} \mid \forall j \in \{1, \dots, k\} : t_j \preceq r_j, \text{root}(t_j) \in \mathcal{D}\} \quad (22)$$

For (22), if some  $r_j$  does not contain any defined symbol, then instead of  $t_j^\#$  we use a fresh constructor  $\text{bot}$  that does not occur in  $\mathcal{R}$  instead. So in both (21) and (22), we replace  $r_j$  by a single term  $t_j^\#$  in the right-hand side, as in the non-probabilistic DP framework. A counterexample to such an approach is the PTRS  $\mathcal{R}_{rw}$  containing only the rule  $g \rightarrow \{3/4 : c(g, g), 1/4 : 0\}$  (13) of Example 3, which is a biased random walk towards non-termination. When using (21) or (22), we would just get the dependency pair  $G \rightarrow \{3/4 : G, 1/4 : t\}$ , where  $t$  is either 0 or  $\text{bot}$ . However, with this dependency pair we lose the information about the second call of the defined symbol  $g$ , which is the reason for non-termination of the random walk. In fact, the PTRS  $\mathcal{R}_1 = \{(9)\} = \{g \rightarrow \{3/4 : d(g), 1/4 : 0\}\}$  from Example 2 would yield the same DP  $G \rightarrow \{3/4 : G, 1/4 : t\}$ , although  $\mathcal{R}_1$  is AST, whereas  $\mathcal{R}_{rw}$  is not. Thus, using such a direct adaption of dependency pairs to the probabilistic setting is unsound.

The counterexample  $\mathcal{R}_{rw}$  shows that having only one annotation in the right-hand side of each dependency pair is not expressive enough, but instead we need a rule where all defined symbols in the right-hand side are annotated, called an *annotated dependency pair*. As we showed in [51], ADPs are not only useful to analyze probabilistic rewriting, but a suitable variant of ADPs can also be used to extend the DP framework to (non-probabilistic) *relative* rewriting. In the following, we first define annotated dependency pairs (Section 3.1) formally. Then we explain how to perform innermost rewriting with ADPs (Section 3.2) and extend this to an arbitrary evaluation strategy (Section 3.3). Finally, we adapt the notion of chains to the probabilistic setting (Section 3.4).

#### 3.1. Annotated dependency pairs

We start with some basic definitions regarding the positions in a term, in order to obtain all positions of defined symbols and to replace certain symbols with their annotated or non-annotated version. For  $t \in \mathcal{T}^\#$  and  $\mathcal{X} \subseteq \Sigma^\# \cup \mathcal{V}$ , let  $\text{Pos}_{\mathcal{X}}(t)$  be all positions of  $t$  with symbols or variables from  $\mathcal{X}$ . For a set of positions  $\Phi \subseteq \text{Pos}_{\mathcal{D} \cup \mathcal{D}^\#}(t)$ , let  $\#_\Phi(t)$  be the variant of  $t$  where the symbols at positions from  $\Phi$  in  $t$  are annotated, and all other annotations are removed. Thus,  $\text{Pos}_{\mathcal{D}^\#}(\#_\Phi(t)) = \Phi$ , and  $\#_\emptyset(t)$  removes all annotations from  $t$ ,

where we often write  $b(\iota)$  instead of  $\sharp_{\theta}(\iota)$ .<sup>6</sup> Moreover, let  $b_{\pi}^{\uparrow}(\iota)$  result from removing all annotations from  $\iota$  that are strictly above the position  $\pi$ . So for  $\mathcal{R}_2$ , we have  $\sharp_{\{1\}}(d(g)) = \sharp_{\{1\}}(D(G)) = d(G)$ ,  $b(D(G)) = d(g)$ , and  $b_1^{\uparrow}(D(G)) = d(G)$ . To transform the rules of a PTRS into ADPs, initially we annotate all defined symbols  $f \in D$  occurring in right-hand sides, as we need to consider all these defined symbols simultaneously. The left-hand side of the resulting ADP is just the left-hand side of the original rule.

The original DP framework for TRSs initially starts with the DP problem  $(\mathcal{P}, \mathcal{R})$ , where  $\mathcal{P}$  is the set of all DPs and  $\mathcal{R}$  contains the original rewrite rules. Instead of additionally considering the original rewrite rules in our ADP framework, every ADP has a flag  $m \in \{\text{true}, \text{false}\}$  to indicate whether this ADP may also be used as an ordinary probabilistic rewrite rule without annotations to rewrite at a position below an annotated symbol. This flag will be modified and used by the processors in Section 4, similar to the ordinary rewrite rules  $\mathcal{R}$  in the DP problem  $(\mathcal{P}, \mathcal{R})$ . So an ADP can be used as a collection of all DPs of a rule, but it can also be used as the original rule itself. We call this the *duality* of ADPs.

**Remark 1.** Instead of the flag  $m$ , one could of course also regard pairs  $(\mathcal{P}, \mathcal{R})$  where  $\mathcal{P}$  is a set of ADPs and  $\mathcal{R}$  contains the same ADPs like  $\mathcal{P}$  but without any annotations. We did not choose this representation since it “duplicates” information (i.e., every non-annotated original rule can be directly obtained from the corresponding ADP in  $\mathcal{P}$  by removing all annotations and thus, it is superfluous to store both the ADP and its corresponding rule). This is in contrast to ordinary DPs where the original rule cannot be reconstructed from the DPs. Moreover, a representation with pairs  $(\mathcal{P}, \mathcal{R})$  has the disadvantage that a processor like the usable rules processor would have to delete rules from  $\mathcal{R}$  which affects the innermost evaluation strategy and would result in a processor that is not complete. Therefore, instead of pairs  $(\mathcal{P}, \mathcal{R})$  we only regard a set of ADPs and use a flag for every ADP to indicate whether this ADP can also be used as an ordinary rule.

**Definition 2** (ADPs, ADP Problem). An *annotated dependency pair* (ADP) has the form  $\ell \rightarrow \{p_1 : r_1, \dots, p_k : r_k\}^m$ , where  $\ell \in \mathcal{T}$  with  $\ell \notin \mathcal{V}$ ,  $m \in \{\text{true}, \text{false}\}$ , and for all  $1 \leq j \leq k$  we have  $r_j \in \mathcal{T}^{\sharp}$  with  $\mathcal{V}(r_j) \subseteq \mathcal{V}(\ell)$ .

For a rule  $\ell \rightarrow \mu = \{p_1 : r_1, \dots, p_k : r_k\}$ , its *canonical annotated dependency pair* is  $DP(\ell \rightarrow \mu) = \ell \rightarrow \{p_1 : \sharp_{\text{Pos}_D(r_1)}(r_1), \dots, p_k : \sharp_{\text{Pos}_D(r_k)}(r_k)\}^{\text{true}}$ . The canonical ADPs of a PTRS  $\mathcal{R}$  are  $DP(\mathcal{R}) = \{DP(\ell \rightarrow \mu) \mid \ell \rightarrow \mu \in \mathcal{R}\}$ . An *ADP problem* is a (finite) set of ADPs  $\mathcal{P}$ .

In the following, we fix an ADP problem  $\mathcal{P}$ . Note that in contrast to ordinary dependency pairs as in Section 2.1, due to the duality of ADPs, the left-hand sides of ADPs are not annotated and the matching used to perform rewriting with ADPs in Definitions 3 and 5 will ignore the annotations as well.

**Example 6.** For  $\mathcal{R}_2 = \{(\#9), (\#11)\}$  and  $\mathcal{R}_3 = \{(\#9), (\#10), (\#12)\}$  from Example 2, we obtain  $DP(\mathcal{R}_2) = \{(\#9), (\#11)\}$  and  $DP(\mathcal{R}_3) = \{(\#9), (\#10), (\#12)\}$  with

$$g \rightarrow \{3/4 : D(G), 1/4 : 0\}^{\text{true}} \quad (\#9)$$

$$d(x) \rightarrow \{1 : 0\}^{\text{true}} \quad (\#10)$$

$$d(x) \rightarrow \{1 : c(x, x)\}^{\text{true}} \quad (\#11)$$

$$d(d(x)) \rightarrow \{1 : c(x, G)\}^{\text{true}} \quad (\#12)$$

Similar to the DPs in the non-probabilistic setting, the annotated subterms of a right-hand side  $r_j$  correspond to those subsequent “function calls” where we need to prove that the arguments are decreasing to conclude AST. For example, for the ADP  $g \rightarrow \{3/4 : D(G), 1/4 : 0\}^{\text{true}}$  (#9), we have to compare the annotated left-hand side  $G$  to both  $D(g)$  and  $G$  from the first term in the right-hand side. When adapting the reduction pair processor to the probabilistic setting in Section 4.4, we will show how to compare a single term with several annotated subterms on the right-hand side.

**Example 7.** For  $\mathcal{R}_{\text{alg}}$  from Example 4, the canonical ADPs are<sup>7</sup>

$$\text{loop1}(y) \rightarrow \{1/2 : L1(D(y)), 1/2 : L2(D(y))\}^{\text{true}} \quad (\#14)$$

$$\text{loop1}(y) \rightarrow \{1/3 : L1(T(y)), 2/3 : L2(T(y))\}^{\text{true}} \quad (\#15)$$

$$\text{loop2}(s(y)) \rightarrow \{1 : L2(y)\}^{\text{true}} \quad (\#16)$$

$$\text{double}(s(y)) \rightarrow \{1 : s(s(D(y)))\}^{\text{true}} \quad (\#17)$$

$$\text{double}(0) \rightarrow \{1 : 0\}^{\text{true}} \quad (\#18)$$

$$\text{triple}(s(y)) \rightarrow \{1 : s(s(s(T(y))))\}^{\text{true}} \quad (\#19)$$

$$\text{triple}(0) \rightarrow \{1 : 0\}^{\text{true}} \quad (\#20)$$

When defining the rewrite relation for ADPs, the essential part is to determine how to handle annotations if we rewrite above or below them. We start with defining innermost rewriting.

<sup>6</sup> Instead of using  $\sharp$  as an inverse operator to  $\sharp$  (like in musical notation), we use  $b$  as in the musical isomorphism from differential geometry.

<sup>7</sup> For reasons of space, we write L1, L2, T, and D, instead of Loop1, Loop2, Triple, and Double, respectively.

### 3.2. Innermost rewriting with ADPs

As in the classical non-probabilistic DP framework, our goal is to track reduction sequences with **a**-steps where the root symbol of the redex is annotated, and between two **a**-steps there can be several **n**-steps. Our novel rewrite relation for ADPs refines **a**-steps into **(at)**- and **(af)**-steps in order to distinguish between rewriting at an **annotation** by an ADP with flag **true** or **false**. Similarly, we also refine **n**-steps into **(nt)**- and **(nf)**-steps depending on the flag of the ADP that is used to rewrite at a non-annotated position. Still, the **(at)**- and **(af)**-steps are the crucial ones, and when extending chains to the probabilistic setting, we prohibit infinite rewrite sequences only consisting of **n**-steps. Similar to the non-probabilistic setting (Section 2.1), this is needed for the reduction pair processor (Section 4.4) to guarantee termination if we orient the **a**-steps strictly with a well-founded order, while only orienting the **n**-steps weakly.

Rewriting with an ADP problem  $\mathcal{P}$  is like ordinary probabilistic term rewriting while considering and modifying annotations that indicate where a non-AST evaluation may arise. If we rewrite a term  $s$  at an innermost position  $\pi$  with an ADP  $\ell \rightarrow \{p_1 : r_1, \dots, p_k : r_k\}^m$ , then the following properties should hold:

- Due to the duality of ADPs, the matching of left-hand sides should ignore the annotations. So for example, we have to be able to rewrite both  $d(d(x))$  and  $D(D(x))$  with an ADP  $d(d(x)) \rightarrow \{1 : c(x, G)\}^{\text{true}}$  at the root.
- Annotations below the position  $\pi$  should be removed, as terms below the redex have to be normal forms due to the innermost evaluation strategy.
- Annotations at positions orthogonal to  $\pi$  should be kept.
- Annotations at positions strictly above  $\pi$  should only be kept if the ADP has the flag  $m = \text{true}$ .
- Annotations in the right-hand sides  $r_1, \dots, r_k$  should only be kept if the redex was annotated at the root, i.e., if  $\pi \in \text{Pos}_{\mathcal{D}^\#}(s)$ .

This leads to the following probabilistic rewrite relation.

**Definition 3** (Innermost Rewriting with ADPs,  $\overset{\downarrow}{\mathcal{C}}_{\mathcal{P}}$ ). A term  $s \in \mathcal{T}^\#$  rewrites innermost with  $\mathcal{P}$  to  $\mu = \{p_1 : t_1, \dots, p_k : t_k\}$  (denoted  $s \overset{\downarrow}{\mathcal{C}}_{\mathcal{P}} \mu$ ) if there are an  $\ell \rightarrow \{p_1 : r_1, \dots, p_k : r_k\}^m \in \mathcal{P}$ , a substitution  $\sigma$ , and a  $\pi \in \text{Pos}_{\mathcal{D}\cup\mathcal{D}^\#}(s)$  such that  $b(s|_\pi) = \ell \sigma \in \text{ANF}_{\mathcal{P}}$ , and for all  $1 \leq j \leq k$  the term  $t_j$  is defined as follows, depending on the flag  $m$  and on whether  $\pi \in \text{Pos}_{\mathcal{D}^\#}(s)$  holds:

$$\begin{array}{l}
 m = \text{true} \\
 m = \text{false}
 \end{array}
 \left| \begin{array}{l}
 \pi \in \text{Pos}_{\mathcal{D}^\#}(s) \\
 t_j = s[r_j \sigma]_\pi \\
 t_j = b_\pi^\uparrow(s[r_j \sigma]_\pi)
 \end{array} \right.
 \begin{array}{l}
 \text{(at)} \\
 \text{(af)}
 \end{array}
 \left| \begin{array}{l}
 \pi \notin \text{Pos}_{\mathcal{D}^\#}(s) \\
 t_j = s[b(r_j) \sigma]_\pi \\
 t_j = b_\pi^\uparrow(s[b(r_j) \sigma]_\pi)
 \end{array} \right.
 \begin{array}{l}
 \text{(nt)} \\
 \text{(nf)}
 \end{array}$$

We sometimes indicate explicitly which case of the above definition is used, so, e.g., we write  $s \overset{\downarrow}{\mathcal{C}}_{\mathcal{P}}^{\text{(at)}} \mu$ .

The above definition indeed corresponds to the properties (a)-(e) mentioned above:

- When computing the matching substitution  $\sigma$ , we only consider the terms  $\ell$  and  $b(s|_\pi)$  without annotations.
- Since all annotations of  $s|_\pi$  below the root correspond to variables of  $\ell$ , the substitution  $\sigma$  removes these annotations.
- All annotations in  $s$  on positions orthogonal to  $\pi$  are not affected when computing  $t_j$ .
- If  $m = \text{false}$ , then  $b_\pi^\uparrow$  removes all annotations at positions strictly above  $\pi$ .
- If  $\pi \notin \text{Pos}_{\mathcal{D}^\#}(s)$ , then we use  $b(r_j)$  to remove all annotations at the right-hand side  $r_j$ .

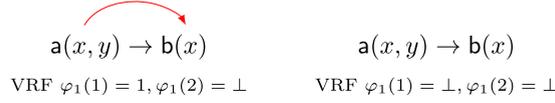
Let us consider an example for each individual case of  $\overset{\downarrow}{\mathcal{C}}_{\mathcal{P}}$ . For this, we fix an ADP problem  $\tilde{\mathcal{P}}$  containing the three rules  $f(x) \rightarrow \{1 : F(x)\}^{\text{true}}$ ,  $g(a) \rightarrow \{1 : G(b)\}^{\text{true}}$ , and  $g(a) \rightarrow \{1 : G(b)\}^{\text{false}}$ .

A step of the form **(at)** (**annotation and true**) is performed at the position of an annotation, i.e., this can potentially lead to a non-AST evaluation. Hence, all annotations from the right-hand side  $r_j$  of the used ADP are kept during the rewrite step. The rewrite step  $F(G(b)) \overset{\downarrow}{\mathcal{C}}_{\tilde{\mathcal{P}}}^{\text{(at)}} \{1 : F(g(b))\}$  is done using the ADP  $f(x) \rightarrow \{1 : F(x)\}^{\text{true}}$  at the root position  $\pi = \varepsilon$  with the substitution  $\sigma(x) = g(b)$ . We get  $b(s|_\pi) = b(F(G(b)))_\varepsilon = f(g(b)) = \ell \sigma$  and  $t_1 = s[r_1]_\pi = F(g(b))$ . Note that annotations below the redex are removed, i.e., we obtain  $g(b)$  instead of  $G(b)$ . This is not a problem because in innermost evaluations, terms below the redex are in normal form anyway.

A step of the form **(af)** (**annotation and false**) is similar but due to the flag  $m = \text{false}$ , this ADP cannot be used below an annotation in a non-AST evaluation. Hence, we remove all annotations above the used redex. Using the ADP  $g(a) \rightarrow \{1 : G(b)\}^{\text{false}}$  on the term  $F(G(a))$  at position 1 yields  $F(G(a)) \overset{\downarrow}{\mathcal{C}}_{\tilde{\mathcal{P}}}^{\text{(af)}} \{1 : f(G(b))\}$ , i.e., we remove the annotation of  $F$  at the root.

A step of the form **(nt)** (**no annotation and true**) is performed at the position of a subterm without annotation. Hence, the subterm cannot lead to a non-AST evaluation, but this rewrite step may be needed for an annotation at a position above. As an example, one could rewrite the non-annotated subterm  $g(a)$  in  $F(g(a)) \overset{\downarrow}{\mathcal{C}}_{\tilde{\mathcal{P}}}^{\text{(nt)}} \{1 : F(g(b))\}$  using the ADP  $g(a) \rightarrow \{1 : G(b)\}^{\text{true}}$ .

Finally, a step of the form **(nf)** (**no annotation and false**) is irrelevant for non-AST evaluations, because the redex is not annotated and due to  $m = \text{false}$ , afterwards one cannot rewrite an annotated term at a position above. For example, using the ADP  $g(a) \rightarrow \{1 : G(b)\}^{\text{false}}$  instead to rewrite the non-annotated subterm  $g(a)$  yields  $F(g(a)) \overset{\downarrow}{\mathcal{C}}_{\tilde{\mathcal{P}}}^{\text{(nf)}} \{1 : f(g(b))\}$ . The case **(nf)** is needed for innermost rewriting to ensure that normal forms always remain the same, even if we remove or add annotations in ADPs or if we modify their flag  $m$ . The flag **false** and steps of the form **(nf)** will allow us to obtain a usable rules processor which is *complete* (see Theorem 11), in contrast to the incomplete usable rules processor from Theorem 2 in the non-probabilistic setting.



**Fig. 3.** The two VRFs for the ADP  $\text{a}(x, y) \rightarrow \{1 : \text{b}(x)\}^{\text{true}}$ .

### 3.3. Full rewriting with ADPs

Next, we consider rewriting with ADPs w.r.t. an arbitrary evaluation strategy. As shown in [Example 3](#), the PTRS  $\mathcal{R}_3$  is  $\text{iAST}$ , but not  $\text{AST}$ . To simulate the corresponding non- $\text{AST}$  reduction from [Example 3](#) with ADPs, it is crucial that if the variables of an ADP are instantiated by annotated terms, then these annotations must be regarded when rewriting with this ADP. For example, if one uses the ADP  $\text{d}(\text{d}(x)) \rightarrow \{1 : \text{c}(x, \text{G})\}^{\text{true}}$  ([#12](#)) to rewrite the redex  $\text{D}(\text{D}(\text{G}))$  (whose subterm  $\text{G}$  starts a non- $\text{AST}$  evaluation), then it should still be possible to continue the evaluation of the subterm  $\text{G}$  afterwards which was “completely inside” the substitution of the applied rewrite step. (We will illustrate this in detail in [Example 9](#).)

So for full rewriting with ADPs, we have to keep certain annotations below the used redex. To this end, we use *variable reposition functions (VRFs)*<sup>8</sup> to relate positions of variables in the left-hand side of an ADP to those positions of the same variables in the right-hand sides where we want to keep the annotations of the instantiated variables. So for an ADP  $\ell \rightarrow \mu^m$  with  $\ell|_\pi = x$ , we indicate which occurrence of  $x$  in  $r \in \text{Supp}(\mu)$  should keep the annotations if one rewrites an instance of  $\ell$  where the subterm at position  $\pi$  contains annotations.

**Definition 4 (Variable Reposition Functions).** Let  $\ell \rightarrow \{p_1 : r_1, \dots, p_k : r_k\}^m$  be an ADP. A family of functions  $(\varphi_j)_{1 \leq j \leq k}$  with  $\varphi_j : \text{Pos}_\mathcal{V}(\ell) \rightarrow \text{Pos}_\mathcal{V}(r_j) \uplus \{\perp\}$  is called a family of *variable reposition functions (VRF)* for the ADP if for all  $1 \leq j \leq k$  we have  $\ell|_\pi = r_j|_{\varphi_j(\pi)}$  whenever  $\varphi_j(\pi) \neq \perp$ .

**Example 8.** Let us consider the ADP  $\text{a}(x, y) \rightarrow \{1 : \text{b}(x)\}^{\text{true}}$ . Here, there are two possible VRFs, viz.,  $\varphi_1(1) = 1$  and  $\varphi_1(2) = \perp$ , or  $\varphi_1(1) = \perp$  and  $\varphi_1(2) = \perp$ . We have to map the position of the variable  $y$  to  $\perp$ , because  $y$  does not occur on the right-hand side. The position of the variable  $x$  can either be mapped to  $\perp$  as well, or to the position 1, which is the position of  $x$  on the right-hand side. This indicates whether we rewrite a term like  $\text{a}(\text{F}, \text{G})$  to  $\text{b}(\text{f})$  or to  $\text{b}(\text{F})$ . See [Fig. 3](#) for a graphical representation of these VRFs.

Note that VRFs do not allow us to duplicate annotations, i.e., we cannot use ADPs to analyze  $\text{AST}$  of duplicating rules like  $\text{d}(x) \rightarrow \{1 : \text{c}(x, x)\}$ . A probabilistic rule or ADP  $\ell \rightarrow \mu$  is *non-duplicating* if all rules in  $\{\ell \rightarrow r \mid r \in \text{Supp}(\mu)\}$  are, and a PTRS or ADP problem is non-duplicating if all of its rules are (disregarding the flag for ADPs).

A VRF indicates at which positions the annotations should be kept. However, since ADPs do not have to be left linear, this movement of annotations from the left- to the right-hand side is a bit more involved. Note that our VRFs do not have to be injective, i.e., for the ADP  $\text{c}(x, x) \rightarrow \{1 : \text{d}(x)\}^{\text{true}}$  a valid VRF would be  $\varphi_1$  with  $\varphi_1(1) = 1$  and  $\varphi_1(2) = 1$ . Now consider the term  $\text{c}(\text{g}(\text{F}), \text{G}(\text{f}))$ . Which annotations should be kept when rewriting at the root position with the ADP  $\text{c}(x, x) \rightarrow \{1 : \text{d}(x)\}^{\text{true}}$ ? Since both the subterm  $\text{G}(\text{f})$  and the subterm  $\text{F}$  may lead to a non- $\text{AST}$  evaluation, this should still be possible after the rewrite step, since both subterms are completely “inside” the substitution used for matching. Hence, our rewrite relation allows for the step  $\text{c}(\text{g}(\text{F}), \text{G}(\text{f})) \hookrightarrow_p \text{d}(\text{G}(\text{F}))$ , i.e., we consider all terms corresponding to the same variable when applying an ADP and *merge* their annotations. In this way, we obtain a term that still contains all annotations that may lead to non- $\text{AST}$  evaluations.

So whenever there is a position  $\rho \in \text{Pos}_\mathcal{V}(\ell)$  of a variable in the left-hand side  $\ell$  of an ADP and  $\rho$  corresponds to a position  $\varphi_j(\rho) \neq \perp$  of the same variable in the right-hand side  $r_j$ , then all annotations below the position  $\rho$  in the redex  $s|_\pi$  must be kept below the position  $\varphi_j(\rho)$  in the resulting term  $r_j\sigma$ . In other words, whenever there is a position  $\tau$  with an annotated symbol at position  $\rho.\tau$  in the redex  $s|_\pi$ , then the symbol at position  $\varphi_j(\rho).\tau$  in the resulting term  $r_j\sigma$  should also be annotated. We collect all these positions in a set  $\Psi_j$ .

In our example with the ADP  $\text{c}(x, x) \rightarrow \{1 : \text{d}(x)\}^{\text{true}}$ , we have  $\ell = \text{c}(x, x)$  and  $r_1 = \text{d}(x)$ . The redex  $s|_\pi$  is  $\text{c}(\text{g}(\text{F}), \text{G}(\text{f}))$  and thus, for the matching substitution  $\sigma$  with  $\ell\sigma = \text{b}(s|_\pi)$ , we obtain  $r_1\sigma = \text{d}(\text{g}(\text{f}))$ . To find out which symbols should be annotated in  $r_1\sigma$ , we consider the VRF  $\varphi_1(1) = 1$  and  $\varphi_1(2) = 1$  which maps variable positions  $\rho$  of  $\ell$  to variable positions of  $r_1$ . For  $\rho = 1$ , the redex  $\text{c}(\text{g}(\text{F}), \text{G}(\text{f}))$  has an annotation at the position  $\rho.\tau$  with  $\tau = 1$ . Hence, the position  $\varphi_1(\rho).\tau = 1.1$  in  $r_1\sigma$  should be annotated. Moreover, for  $\rho' = 2$ , the redex  $\text{c}(\text{g}(\text{F}), \text{G}(\text{f}))$  has an annotation at the position  $\rho'.\tau'$  with  $\tau' = \epsilon$ . Hence, the position  $\varphi_1(\rho').\tau' = 1$  in  $r_1\sigma$  should be annotated as well. Thus, for  $\Psi_1 = \{1.1, 1\}$  we obtain  $\#_{\Psi_1}(r_1\sigma) = \#_{\{1.1, 1\}}(\text{d}(\text{g}(\text{f}))) = \text{d}(\text{G}(\text{F}))$ .

Now we can define (full) rewriting with ADPs. As in the definition of innermost rewriting ([Definition 3](#)), all annotations of  $r_j$  should be deleted whenever  $\pi \notin \text{Pos}_{\mathcal{D}^\#}(s)$ . Thus, we use  $\#_{\Psi_j}(r_j\sigma)$  which removes all annotations from  $r_j$  (and only adds certain annotations on the terms introduced by  $\sigma$ ). In contrast, if  $\pi \in \text{Pos}_{\mathcal{D}^\#}(s)$ , then we use  $\#_{\text{Pos}_{\mathcal{D}^\#}(r_j) \cup \Psi_j}(r_j\sigma)$  in order to keep the annotations of  $r_j$ .

**Definition 5 (Rewriting with ADPs,  $\hookrightarrow_p$ ).** A term  $s \in \mathcal{T}^\#$  rewrites with  $\mathcal{P}$  to  $\mu = \{p_1 : t_1, \dots, p_k : t_k\}$  (denoted  $s \hookrightarrow_p \mu$ ) if there are an  $\ell \rightarrow \{p_1 : r_1, \dots, p_k : r_k\}^m \in \mathcal{P}$ , a VRF  $(\varphi_j)_{1 \leq j \leq k}$  for this ADP, a substitution  $\sigma$ , and a  $\pi \in \text{Pos}_{\mathcal{D} \cup \mathcal{D}^\#}(s)$  such that  $\text{b}(s|_\pi) = \ell\sigma$ , and for all

<sup>8</sup> VRFs were introduced in [\[51\]](#) when adapting ADPs to full *relative* rewriting. However, due to the probabilistic setting, our definition here is slightly different.

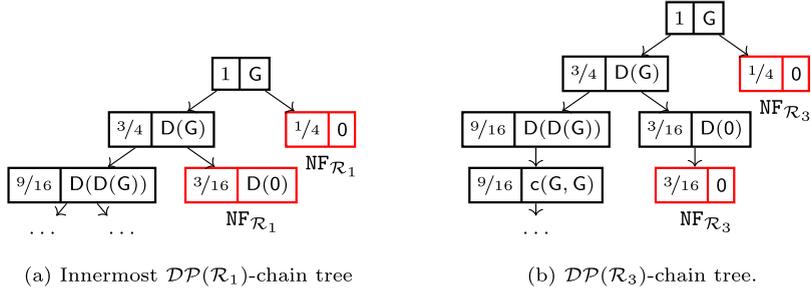


Fig. 4. Two chain trees where we indicated nodes labeled by normal forms.

$1 \leq j \leq k$  the term  $t_j$  is defined as follows, depending on the flag  $m$  and on whether  $\pi \in \text{Pos}_{D^\#}(s)$  holds:

$$\begin{array}{l}
 m = \text{true} \quad \left| \begin{array}{l} \pi \in \text{Pos}_{D^\#}(s) \\ t_j = s[\#_{\text{Pos}_{D^\#}(r_j) \cup \Psi_j}(r_j \sigma)]_\pi \end{array} \right. \quad (\mathbf{at}) \quad \left| \begin{array}{l} \pi \notin \text{Pos}_{D^\#}(s) \\ t_j = s[\#_{\Psi_j}(r_j \sigma)]_\pi \end{array} \right. \quad (\mathbf{nt}) \\
 m = \text{false} \quad \left| \begin{array}{l} t_j = b_\pi^\uparrow(s[\#_{\text{Pos}_{D^\#}(r_j) \cup \Psi_j}(r_j \sigma)]_\pi) \\ t_j = b_\pi^\uparrow(s[\#_{\Psi_j}(r_j \sigma)]_\pi) \end{array} \right. \quad (\mathbf{af}) \quad \left| \begin{array}{l} t_j = b_\pi^\uparrow(s[\#_{\Psi_j}(r_j \sigma)]_\pi) \\ t_j = b_\pi^\uparrow(s[\#_{\Psi_j}(r_j \sigma)]_\pi) \end{array} \right. \quad (\mathbf{nf})
 \end{array}$$

Here,  $\Psi_j = \{\varphi_j(\rho) \cdot \tau \mid \rho \in \text{Pos}_V(\ell), \varphi_j(\rho) \neq \perp, \rho \cdot \tau \in \text{Pos}_{D^\#}(s|_\pi)\}$ .

Thus, innermost rewriting is the special case of Definition 5 where we require that  $b(s|_\pi) = \ell \sigma$  is in argument normal form and where we use the VRF which maps all variable positions to  $\perp$  (i.e., all annotations below the redex are removed).

### 3.4. Chains

For chains in the probabilistic setting we now consider specific RSTs, called chain trees. Chain trees are defined analogously to RSTs, but the crucial requirement is that every infinite path of the tree must contain infinitely many steps of the forms (at) or (af).

**Definition 6** (Chain Tree,  $|\mathfrak{T}|$ ). We say that  $\mathfrak{T} = (V, E, L)$  is a  $\mathcal{P}$ -chain tree (CT) if

- (1)  $(V, E)$  is a (possibly infinite) directed tree with nodes  $V \neq \emptyset$  and directed edges  $E \subseteq V \times V$  where  $vE = \{w \mid (v, w) \in E\}$  is finite for every  $v \in V$ .
- (2)  $L : V \rightarrow (0, 1] \times \mathcal{T}^\#$  labels every node  $v$  by a probability  $p_v$  and a term  $t_v$ . For the root  $v \in V$  of the tree, we have  $p_v = 1$ .
- (3) If  $vE = \{w_1, \dots, w_k\}$ , then  $t_v \hookrightarrow_{\mathcal{P}} \{\frac{p_{w_1}}{p_v} : t_{w_1}, \dots, \frac{p_{w_k}}{p_v} : t_{w_k}\}$ .
- (4) Every infinite path in  $\mathfrak{T}$  contains infinitely many rewrite steps with Case (at) or (af).

$\mathfrak{T}$  is an innermost chain tree if the edges represent innermost rewrite steps. As for RSTs, we say that a CT converges (or terminates) with probability  $|\mathfrak{T}| = \sum_{v \in \text{Leaf}} p_v$ , where Leaf is the set of its leaves.

With this notion of chain trees, we can now define almost-sure termination for ADP problems.

**Definition 7** (AST and iAST for ADP Problems). An ADP problem  $\mathcal{P}$  is called almost surely terminating (AST) if we have  $|\mathfrak{T}| = 1$  for every  $\mathcal{P}$ -CT  $\mathfrak{T}$ . It is called almost surely innermost terminating (iAST) if we have  $|\mathfrak{T}| = 1$  for every innermost  $\mathcal{P}$ -CT  $\mathfrak{T}$ .

The corresponding  $DP(\mathcal{R}_1)$ -chain tree  $\mathfrak{T}$  for the  $\mathcal{R}_1$ -RST from Fig. 2 is shown in Fig. 4(a). Here, we again have  $|\mathfrak{T}| = 1$ .

**Example 9.** To see why VRFs are needed to obtain a sound chain criterion for full rewriting, consider  $\mathcal{R}_3$  with  $DP(\mathcal{R}_3) = \{(\#9), (\#10), (\#12)\}$  from Example 6:

$$\begin{array}{ll}
 g \rightarrow \{3/4 : D(G), 1/4 : 0\}^{\text{true}} & (\#9) \\
 d(x) \rightarrow \{1 : 0\}^{\text{true}} & (\#10) \\
 d(d(x)) \rightarrow \{1 : c(x, G)\}^{\text{true}} & (\#12)
 \end{array}$$

As noted in Example 3,  $\mathcal{R}_3$  is iAST, but not AST. For  $DP(\mathcal{R}_3)$ , applying two rewrite steps with the ADP (#9) to G would result in a chain tree with the leaves  $9/16 : D(D(G))$ ,  $3/16 : D(0)$  (which can be extended by the child  $3/16 : 0$ ), and  $1/4 : 0$ . Now it is important that the next application of the ADP (#12) to  $9/16 : D(D(G))$  allows us to keep the annotation of its argument G, i.e., it can yield  $9/16 : c(G, G)$ . The corresponding chain tree is depicted in Fig. 4(b). Otherwise (i.e., if it could only yield  $9/16 : c(g, G)$ ), the number of G-symbols would never be increased. Since non-annotated symbols like g do not result in any (at)- or (af)-steps, for all such chain trees  $\mathfrak{T}$  we would have  $|\mathfrak{T}| = 1$ , i.e., then the chain criterion would falsely conclude that  $\mathcal{R}_3$  is AST.

For innermost rewriting, we now indeed obtain a sound and complete chain criterion as in the non-probabilistic setting, i.e., to analyze iAST for a PTRS  $\mathcal{R}$  one can analyze iAST of its canonical ADP problem  $DP(\mathcal{R})$  instead.

**Theorem 4** (Chain Criterion for iAST). A PTRS  $\mathcal{R}$  is iAST iff  $DP(\mathcal{R})$  is iAST.

*Proof Sketch. Soundness:* Each innermost  $\mathcal{R}$ -RST  $\mathfrak{T}$  can be transformed into an innermost  $DP(\mathcal{R})$ -CT  $\mathfrak{T}'$  with  $|\mathfrak{T}| = |\mathfrak{T}'|$  by annotating at least all defined symbols at the root of subterms that are not in normal form. This is possible since the relation  $\hookrightarrow_{DP(\mathcal{R})}$  only removes annotations of normal forms if all ADPs have the flag true.

*Completeness:* Each innermost  $DP(\mathcal{R})$ -CT  $\mathfrak{T}$  can be transformed into an innermost  $\mathcal{R}$ -RST  $\mathfrak{T}'$  with  $|\mathfrak{T}| = |\mathfrak{T}'|$  by removing all annotations.  $\square$

For the chain criterion for full AST, we have to impose a restriction on the initial PTRS, namely the PTRS must be non-duplicating. As mentioned before, our VRFs (and the corresponding full rewrite relation for ADPs) cannot handle duplicating rules correctly. For example, for the duplicating ADP  $d(x) \rightarrow \{1 : c(x, x)\}^{\text{true}}$  (#11), we have three different VRFs which map position 1 to either  $\perp$ , 1, or 2, but we cannot map it to both positions 1 and 2. Hence,  $DP(\mathcal{R}_2)$  would be considered to be AST, as  $D(G)$  only rewrites to  $\{1 : c(G, g)\}$  or  $\{1 : c(g, G)\}$ , but the annotation cannot be duplicated. Thus, the chain criterion would be unsound for duplicating PTRSs like  $\mathcal{R}_2$ . To handle duplicating rules, one can adapt the direct application of orders to prove AST from [42] and try to remove the duplicating rules of the PTRS before constructing the canonical ADPs.

Alternatively, one could modify the definition of the rewrite relation  $\hookrightarrow_p$  and use *generalized* VRFs (GVRFs) instead of VRFs which can duplicate annotations. This would yield a sound and complete chain criterion for AST of possibly duplicating PTRSs, but then one would also have to consider this modified definition of  $\hookrightarrow_p$  for the processors of the ADP framework in Section 4. Unfortunately, the most important processors would become unsound when defining the rewrite relation  $\hookrightarrow_p$  via GVRFs (see Examples 12 and 19). Therefore, we use VRFs instead and restrict ourselves to non-duplicating PTRSs for the soundness of the chain criterion for full probabilistic rewriting.<sup>9</sup>

**Theorem 5** (Chain Criterion for AST). *A non-duplicating PTRS  $\mathcal{R}$  is AST iff  $DP(\mathcal{R})$  is AST.*

*Proof Sketch. Soundness:* Each  $\mathcal{R}$ -RST  $\mathfrak{T}$  can be transformed into a  $DP(\mathcal{R})$ -CT  $\mathfrak{T}'$  with  $|\mathfrak{T}| = |\mathfrak{T}'|$  by annotating all defined symbols. Since  $\mathcal{R}$  is non-duplicating and all ADPs in  $DP(\mathcal{R})$  have the flag true, we do not remove any annotations when rewriting with  $\hookrightarrow_{DP(\mathcal{R})}$ .

*Completeness:* Each  $DP(\mathcal{R})$ -CT  $\mathfrak{T}$  can be transformed into an  $\mathcal{R}$ -RST  $\mathfrak{T}'$  with  $|\mathfrak{T}| = |\mathfrak{T}'|$  by removing all annotations.  $\square$

**Example 10.** As an example, the PTRS  $\mathcal{R}_{\text{alg}}$  from Example 4 is non-duplicating. Hence, we can not only analyze  $\text{iAST}$  but also AST via the canonical ADP problem  $DP(\mathcal{R}_{\text{alg}})$  from Example 7.

**Remark 2.** In the chain criterion for non-probabilistic DPs, it suffices to regard only instantiations where all terms below an annotated symbol are terminating. The reason is the *minimality property* of non-probabilistic term rewriting, i.e., whenever a term starts an infinite rewrite sequence, then it also starts an infinite sequence, where all proper subterms of every used redex are terminating. However, in the probabilistic setting such a minimality property does not hold [43]. For  $\mathcal{R}_3$  from Example 2,  $g$  starts a non-AST RST, but in this RST, one has to apply Rule (12) to the redex  $d(d(g))$ , although it contains the proper subterm  $g$  that starts a non-AST RST.

#### 4. The probabilistic ADP framework

The idea of the DP framework for non-probabilistic TRSs is to apply *DP processors* repeatedly which transform a DP problem into simpler sub-problems as we have seen in Section 2.1. The same idea is used in the ADP framework, which yields the first *modular* approach for proving AST of probabilistic TRSs automatically. An *ADP processor* Proc has the form  $\text{Proc}(\mathcal{P}) = \{\mathcal{P}_1, \dots, \mathcal{P}_n\}$  for ADP problems  $\mathcal{P}, \mathcal{P}_1, \dots, \mathcal{P}_n$ . Let  $\mathcal{Z} \in \{\text{AST}, \text{iAST}\}$ . Proc is *sound* for  $\mathcal{Z}$  if  $\mathcal{P}$  is  $\mathcal{Z}$  whenever  $\mathcal{P}_j$  is  $\mathcal{Z}$  for all  $1 \leq j \leq n$ . It is *complete* for  $\mathcal{Z}$  if  $\mathcal{P}_j$  is  $\mathcal{Z}$  for all  $1 \leq j \leq n$  whenever  $\mathcal{P}$  is  $\mathcal{Z}$ . Thus, one starts with the canonical ADP problem and applies sound (and preferably complete) ADP processors repeatedly until there are no more remaining ADP problems. This implies that the canonical ADP problem is  $\mathcal{Z}$  and by the chain criterion, the original PTRS is  $\mathcal{Z}$  as well.

An ADP problem  $\mathcal{P}$  without annotations is always AST and  $\text{iAST}$ , because then no rewrite step with  $\hookrightarrow_p$  increases the number of annotations (recall that VRFs cannot duplicate annotations). Hence, then any term with  $n$  annotations only starts rewrite sequences with at most  $n$  steps of the form (at) or (af), i.e., all  $\mathcal{P}$ -CTs are finite.

In Sections 4.1–4.4, we adapt the main processors from Section 2.1 to our new ADP framework for AST and  $\text{iAST}$ . Whenever the respective processor can be used for proving AST, we explain the variant of the processor for full rewriting first and afterwards, we show how one can improve the processor for  $\text{iAST}$ . Moreover, we introduce an additional processor to remove all occurring probabilities (Section 4.5), a novel processor based on the *subterm criterion* of [12] (Section 4.6), and we discuss the possibility of introducing a processor to switch from full AST to innermost AST (Section 4.7).

All processors in this section are complete, because they only remove annotations and only change flags of ADPs from true to false.

**Theorem 6** (Completeness of Processors). *Let Proc be an ADP processor which satisfies the following for all ADP problems  $\mathcal{P}$ : Whenever there is an ADP  $\ell \rightarrow \{p_1 : r'_1, \dots, p_k : r'_k\}^{m'} \in \mathcal{P}'$  for some  $\mathcal{P}' \in \text{Proc}(\mathcal{P})$ , then there is an ADP  $\ell \rightarrow \{p_1 : r_1, \dots, p_k : r_k\}^m \in \mathcal{P}$  such that  $b(r'_j) = b(r_j)$  and  $\text{Pos}_{D\#}(r'_j) \subseteq \text{Pos}_{D\#}(r_j)$  for all  $1 \leq j \leq k$ , and such that  $m' = \text{true}$  implies  $m = \text{true}$ . Then Proc is complete for both AST and  $\text{iAST}$ .*

**Proof.** Each (innermost)  $\mathcal{P}'$ -CT  $\mathfrak{T}'$  can be transformed into an (innermost)  $\mathcal{P}$ -CT  $\mathfrak{T}$  with  $|\mathfrak{T}'| = |\mathfrak{T}|$  by performing the respective rewrite steps with  $\ell \rightarrow \{p_1 : r_1, \dots, p_k : r_k\}^m \in \mathcal{P}$  instead of  $\ell \rightarrow \{p_1 : r'_1, \dots, p_k : r'_k\}^{m'} \in \mathcal{P}'$ . Since we have  $b(r'_j) = b(r_j)$  and  $\text{Pos}_{D\#}(r'_j) \subseteq \text{Pos}_{D\#}(r_j)$  for all  $1 \leq j \leq k$ , and  $m = \text{true}$  if  $m' = \text{true}$ ,  $\mathfrak{T}$  has the same terms as  $\mathfrak{T}'$  but possibly with more annotations.

<sup>9</sup> A related restriction is needed when analyzing (non-probabilistic) relative termination due to the VRFs [51].

Since every infinite path in  $\mathfrak{T}'$  contains infinitely many rewrite steps with Case (at) or (af), this ensures that  $\mathfrak{T}$  satisfies this property as well.  $\square$

The above proof structure will often be used for the following soundness proofs as well. Given some  $\mathcal{P}$ -CT  $\mathfrak{T}$ , we use a  $\mathcal{P}' \in \text{Proc}(\mathcal{P})$  and construct a  $\mathcal{P}'$ -CT  $\mathfrak{T}'$  such that  $|\mathfrak{T}'| \leq |\mathfrak{T}|$ . If the processor Proc removes annotations, then one has to ensure that every infinite path in  $\mathfrak{T}'$  still contains infinitely many rewrite steps with Case (at) or (af).

Moreover, the following two lemmas are important for the soundness proofs. The first lemma states that if we can partition the set of ADPs  $\mathcal{P} = \mathcal{P}_1 \uplus \mathcal{P}_2$  such that  $\mathcal{P}_1 \cup \mathfrak{b}(\mathcal{P}_2)$  is AST, then it suffices to consider  $\mathcal{P}$ -CTs where every infinite path has an infinite number of annotated rewrite steps with nodes from  $\mathcal{P}_2$  (i.e., infinitely many rewrite steps where ADPs from  $\mathcal{P}_2$  are used with Case (at) or (af)).

**Lemma 1** ( *$\mathcal{P}$ -Partition Lemma*). *Let  $\mathcal{P}$  be an ADP problem and let  $\mathfrak{T}$  be a  $\mathcal{P}$ -CT that converges with probability  $< 1$ . Assume that  $\mathcal{P}_1 \cup \mathfrak{b}(\mathcal{P}_2)$  is AST. Then there exists a  $\mathcal{P}$ -CT  $\mathfrak{T}'$  that converges with probability  $< 1$  such that every infinite path uses an infinite number of annotated rewrite steps with  $\mathcal{P}_2$ .*

The second crucial lemma is the *Starting Lemma*. It shows that w.l.o.g., we can assume that the root of a CT is labeled with  $(1 : t)$  for a term  $t \in \mathcal{T}^\#$  with  $\text{Pos}_{\mathcal{D}^\#}(t) = \{\varepsilon\}$  and  $\mathfrak{b}(t) = s\theta$  for a substitution  $\theta$  and an ADP  $s \rightarrow \mu^m \in \mathcal{P}$ . So  $t$  is an instantiated left-hand side where only the root is annotated. The proof for this lemma requires the above  $\mathcal{P}$ -Partition Lemma.

**Lemma 2** (*Starting Lemma*). *If an ADP problem  $\mathcal{P}$  is not AST, then there exists a  $\mathcal{P}$ -CT  $\mathfrak{T}$  with  $|\mathfrak{T}| < 1$  that starts with  $(1 : t)$  for a term  $t \in \mathcal{T}^\#$  where  $\text{Pos}_{\mathcal{D}^\#}(t) = \{\varepsilon\}$  and  $\mathfrak{b}(t) = s\theta$  for a substitution  $\theta$  and an ADP  $s \rightarrow \dots \in \mathcal{P}$ . If  $\mathcal{P}$  is not iAST, then we can further assume that  $t \in \text{ANF}_{\mathcal{P}}$ .*

#### 4.1. Dependency graph processor

As in the non-probabilistic setting, the  $\mathcal{P}$ -dependency graph is a control flow graph whose nodes are now the ADPs from  $\mathcal{P}$ . Since a term can now contain multiple annotations, the definition is a bit more involved than in the non-probabilistic setting. An edge of the dependency graph now indicates whether an ADP  $\alpha$  may lead to an application of another ADP  $\alpha'$  on an annotated subterm whose annotation was introduced by  $\alpha$ . This possibility is not related to the probabilities. Hence, here we use the *non-probabilistic variant*  $\text{np}(\mathcal{P}) = \{\ell \rightarrow \mathfrak{b}(r_j) \mid \ell \rightarrow \{p_1 : r_1, \dots, p_k : r_k\}^{\text{true}} \in \mathcal{P}, 1 \leq j \leq k\}$ , which is an ordinary TRS over the original signature  $\Sigma$ . For  $\text{np}(\mathcal{P})$  we only consider rules with the flag true, as only they can be used for rewriting below annotations without removing them. We define  $t \trianglelefteq_{\#} s$  if there is a  $\pi \in \text{Pos}_{\mathcal{D}^\#}(s)$  and  $t = \mathfrak{b}(s|_{\pi})$ , i.e.,  $t$  results from a subterm of  $s$  with annotated root symbol by removing its annotations.

##### 4.1.1. Dependency graph processor for AST

We start with the dependency graph processor for full AST.

**Definition 8** (*Dependency Graph*). The  $\mathcal{P}$ -dependency graph has the nodes  $\mathcal{P}$  and there is an edge from  $\ell_1 \rightarrow \{p_1 : r_1, \dots, p_k : r_k\}^m$  to  $\ell_2 \rightarrow \dots$  if there are substitutions  $\sigma_1, \sigma_2$  and a  $t \trianglelefteq_{\#} r_j$  for some  $1 \leq j \leq k$  with  $t^{\#}\sigma_1 \rightarrow_{\text{np}(\mathcal{P})}^* \ell_2^{\#}\sigma_2$ .

So there is an edge from an ADP  $\alpha$  to an ADP  $\alpha'$  if after a  $\hookrightarrow_{\mathcal{P}}$ -step of the form (at) or (af) with  $\alpha$  at position  $\pi$  there may eventually come another  $\hookrightarrow_{\mathcal{P}}$ -step of the form (at) or (af) with  $\alpha'$  at a position  $\tau$  on or below  $\pi$ , where the root symbol of the subterm at position  $\tau$  was introduced by  $\alpha$ . Since every infinite path in a  $\mathcal{P}$ -CT contains infinitely many rewrite steps with Case (at) or (af), every such path traverses a cycle of the dependency graph infinitely often. Thus, as in the non-probabilistic setting, it suffices to consider its strongly connected components separately. In the ADP framework, this means that we remove the annotations from all ADPs except the ones in the SCC that we want to analyze.

To automate the following processor, the same over-approximation techniques as for the non-probabilistic dependency graph can be used, e.g., to estimate whether  $t^{\#}\sigma_1 \rightarrow_{\text{np}(\mathcal{P})}^* \ell_2^{\#}\sigma_2$  holds for some  $\sigma_1, \sigma_2$ , one checks whether  $\text{Ren}(\text{Cap}_{\mathcal{P}}(t^{\#}))$  and  $\ell_2^{\#}$  are unifiable. As in Section 2.1,  $\text{Ren}$  replaces all variable occurrences by fresh variables and  $\text{Cap}_{\mathcal{P}}(t^{\#})$  replaces all those subterms  $f(\dots)$  of  $t^{\#}$  by different fresh variables where  $f$  occurs on the root position of some left-hand side of  $\mathcal{P}$ . One could also improve  $\text{Cap}_{\mathcal{P}}$  by only regarding the root symbols  $f$  of left-hand sides of ADPs with the flag  $m = \text{true}$  as “defined”. In the following, we extend  $\mathfrak{b}$  to multi-distributions, ADPs, and ADP problems by removing the annotations of all occurring terms. So  $\mathfrak{b}(\{p_1 : r_1, \dots, p_k : r_k\}) = \{p_1 : \mathfrak{b}(r_1), \dots, p_k : \mathfrak{b}(r_k)\}$ ,  $\mathfrak{b}(\ell \rightarrow \mu) = \ell \rightarrow \mathfrak{b}(\mu)$ , and  $\mathfrak{b}(\mathcal{P}) = \{\mathfrak{b}(\ell \rightarrow \mu) \mid \ell \rightarrow \mu \in \mathcal{P}\}$ .

**Theorem 7** (*Dependency Graph Processor for AST*). *Let  $\mathcal{P}_1, \dots, \mathcal{P}_n$  be the SCCs of the  $\mathcal{P}$ -dependency graph. Then  $\text{Proc}_{\text{DG}}(\mathcal{P}) = \{\mathcal{P}_1 \cup \mathfrak{b}(\mathcal{P} \setminus \mathcal{P}_1), \dots, \mathcal{P}_n \cup \mathfrak{b}(\mathcal{P} \setminus \mathcal{P}_n)\}$  is sound and complete for AST.*

*Proof Sketch. Soundness:* For any subset  $X \subseteq \mathcal{P}$ , let  $\overline{X} = X \cup \mathfrak{b}(\mathcal{P} \setminus X)$ . Suppose that every  $\overline{\mathcal{P}_i}$ -CT converges with probability 1 for all  $1 \leq i \leq n$ . We prove that then also every  $\mathcal{P}$ -CT converges with probability 1. Let  $\mathfrak{G}$  be the  $\mathcal{P}$ -dependency graph and let  $\mathfrak{B} = \{\mathcal{P}_1, \dots, \mathcal{P}_n\} \cup \{\{v\} \subseteq \mathcal{P} \mid v \text{ is not in an SCC of } \mathfrak{G}\}$  be the set of all SCCs and all singleton sets of nodes that do not belong to any SCC. For two  $X_1, X_2 \in \mathfrak{B}$  we say that  $X_2$  is a *direct successor* of  $X_1$  (denoted  $X_1 >_{\mathfrak{G}} X_2$ ) if there exist nodes  $v \in X_1$  and  $w \in X_2$  such that there is an edge from  $v$  to  $w$  in  $\mathfrak{G}$ . The core steps of the proof are the following (where the Starting Lemma (Lemma 2) is used in Step 1 and 2, and the  $\mathcal{P}$ -Partition Lemma (Lemma 1) is used in Step 2):

1. We first show that every ADP problem  $\overline{X}$  with  $X \in \mathfrak{B}$  is AST. This also holds for the singleton sets. The reason is that if there is only a single ADP with annotations and this ADP is not contained in any cycle of the dependency graph, then there cannot be infinitely many steps at annotated subterms, i.e., all paths are finite.

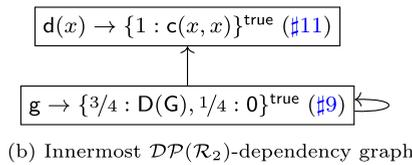
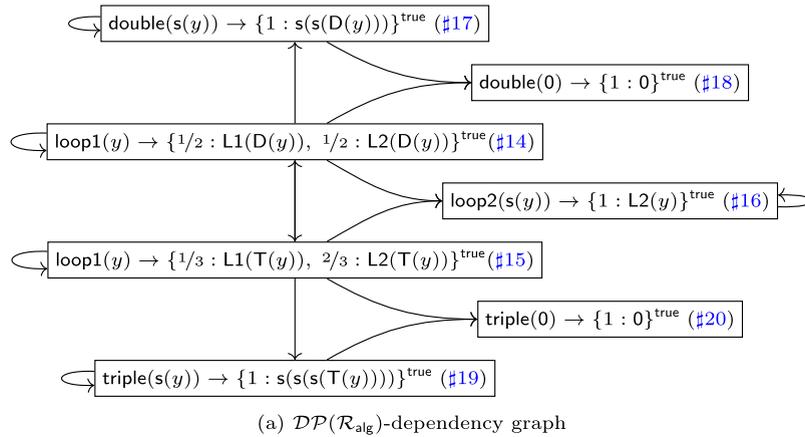


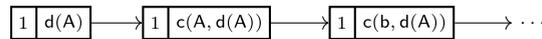
Fig. 5. Two examples of dependency graphs.

2. Then we show that composing SCCs maintains the AST property.
3. Finally, by induction on the direct successor relation we show that for every  $X \in \mathfrak{B}$ , the ADP problem  $\overline{\bigcup_{X >_{\mathfrak{B}}^* Y} Y}$  is AST (where Step 1 is needed for the induction base and Step 2 is needed for the induction step). This implies that  $\mathcal{P}$  must be AST as well.

Completeness: By Theorem 6.  $\square$

**Example 11.** Consider  $\mathcal{R}_{\text{alg}}$  and its canonical ADPs from Example 7. The  $DP(\mathcal{R}_{\text{alg}})$ -dependency graph is shown in Fig. 5(a). Its SCCs are  $\{(\#14), (\#15)\}$ ,  $\{(\#16)\}$ ,  $\{(\#17)\}$ , and  $\{(\#19)\}$ . For each SCC we create a separate ADP problem, where all annotations outside the SCC are removed. This leads to the ADP problems  $\{(\#14), (\#15), b(\#16) - b(\#20)\}$ ,  $\{(\#16), b(\#14), b(\#15), b(\#17) - b(\#20)\}$ ,  $\{(\#17), b(\#14) - b(\#16), b(\#18) - b(\#20)\}$ , and  $\{(\#19), b(\#14) - b(\#18), b(\#20)\}$ .

**Example 12.** If we used GVRFs that can duplicate annotations, then the dependency graph processor would not be sound. The reason is that  $\text{Proc}_{\text{DG}}$  maps ADP problems without annotations to the empty set. However, this would be unsound if we had GVRFs, because then the ADP problem consisting of  $a \rightarrow \{1 : b\}^{\text{true}}$  and  $d(x) \rightarrow \{1 : c(x, d(x))\}^{\text{true}}$  would not be AST. Here, the use of GVRFs would lead to the following CT with an infinite number of (at) steps that rewrite A to b.



#### 4.1.2. Dependency graph processor for iAST

As in Section 2.1, we can integrate the restriction to innermost rewriting into the dependency graph. So instead of arbitrary rewrite steps, we now only allow innermost steps. Moreover, the substitutions  $\sigma_1, \sigma_2$  need to be chosen in such a way that both ADPs can be applied in innermost reductions, i.e., we must have  $\ell_1 \sigma_1, \ell_2 \sigma_2 \in \text{ANF}_{\mathcal{P}}$ .

**Definition 9** (Innermost Dep. Graph). The innermost  $\mathcal{P}$ -dependency graph has the nodes  $\mathcal{P}$ , and there is an edge from  $\ell_1 \rightarrow \{p_1 : r_1, \dots, p_k : r_k\}^m$  to  $\ell_2 \rightarrow \dots$  if there are substitutions  $\sigma_1, \sigma_2$  and a  $t \trianglelefteq_{\#} r_j$  for some  $1 \leq j \leq k$  such that  $t \# \sigma_1 \xrightarrow{i}_{\text{np}(\mathcal{P})} \ell_2 \# \sigma_2$  and both  $\ell_1 \sigma_1$  and  $\ell_2 \sigma_2$  are in  $\text{ANF}_{\mathcal{P}}$ .

Again, the same over-approximation techniques as in Section 2.1 can also be used to estimate the innermost dependency graph in the probabilistic setting.

**Theorem 8** (Dependency Graph Processor for iAST). Let  $\mathcal{P}_1, \dots, \mathcal{P}_n$  be the SCCs of the innermost  $\mathcal{P}$ -dependency graph. Then  $\text{Proc}_{\text{DG}}^{\text{i}}(\mathcal{P}) = \{\mathcal{P}_1 \cup b(\mathcal{P} \setminus \mathcal{P}_1), \dots, \mathcal{P}_n \cup b(\mathcal{P} \setminus \mathcal{P}_n)\}$  is sound and complete for iAST.

*Proof Sketch. Soundness:* Similar to the proof of Theorem 7, just considering innermost rewrite steps.

*Completeness:* By Theorem 6.  $\square$

**Example 13.** For  $\mathcal{R}_{\text{alg}}$ , the innermost and the full  $DP(\mathcal{R}_{\text{alg}})$ -dependency graph coincide. Thus, for  $\text{iAST}$  we result in the same ADP problems as in [Example 11](#).

For the duplicating PTRS  $\mathcal{R}_2$  and its canonical ADPs from [Example 6](#) we can only analyze  $\text{iAST}$ . The innermost  $DP(\mathcal{R}_2)$ -dependency graph can be seen in [Fig. 5\(b\)](#). As the only SCC  $\{\#9\}$  does not contain  $\{\#11\}$ , we can remove all annotations from  $\{\#11\}$ . However, since  $\{\#11\}$  already has no annotations,  $\text{Proc}_{\text{DG}}^{\text{i}}$  does not change  $DP(\mathcal{R}_2)$ .

#### 4.2. Usable terms processor

The dependency graph processor either removes all annotations from an ADP or none. But an ADP can still contain terms  $t^\#$  where no instance  $t^\# \sigma_1$  rewrites to an instance  $\ell^\# \sigma_2$  of a left-hand side  $\ell$  of an ADP with annotations, i.e., it can only lead to ADPs outside the SCC we consider. The *usable terms processor* removes the annotation from the root of such *non-usable* terms like  $D(\dots)$  in  $DP(\mathcal{R}_2) = \{g \rightarrow \{3/4 : D(G), 1/4 : 0\}^{\text{true}} \#9\}, d(x) \rightarrow \{1 : c(x, x)\}^{\text{true}} \#11\}$ . So instead of whole ADPs, here we consider the subterms in the right-hand sides of an ADP individually. Such a processor was not needed in the non-probabilistic setting, as ordinary dependency pairs only contain a single annotation.

##### 4.2.1. Usable terms processor for AST

**Theorem 9** (Usable Terms Processor for AST). *We call  $t \in \mathcal{T}^\#$  with  $\text{root}(t) \in D^\#$  usable w.r.t. an ADP problem  $\mathcal{P}$  if there are substitutions  $\sigma_1, \sigma_2$  and an  $\ell_2 \rightarrow \mu_2 \in \mathcal{P}$  where  $\mu_2$  contains an annotated symbol, such that  $\#_{\{\varepsilon\}}(t)\sigma_1 \xrightarrow[\text{np}(\mathcal{P})]{*} \ell_2^\# \sigma_2$ . Let  $\Delta_{\mathcal{P}}(s) = \{\pi \in \text{Pos}_{D^\#}(s) \mid s|_\pi \text{ is usable w.r.t. } \mathcal{P}\}$  and let  $\mathcal{T}_{\text{UT}}(\mathcal{P}) = \{\ell \rightarrow \{p_1 : \#_{\Delta_{\mathcal{P}}(r_1)}(r_1), \dots, p_k : \#_{\Delta_{\mathcal{P}}(r_k)}(r_k)\}^m \mid \ell \rightarrow \{p_1 : r_1, \dots, p_k : r_k\}^m \in \mathcal{P}\}$  be the transformation that removes the annotations from the roots of all non-usable terms in the right-hand sides. Then  $\text{Proc}_{\text{UT}}(\mathcal{P}) = \{\mathcal{T}_{\text{UT}}(\mathcal{P})\}$  is sound and complete for AST.*

*Proof Sketch. Soundness:* For every  $\mathcal{P}$ -CT  $\mathfrak{A}$  (satisfying the restrictions of the Starting Lemma ([Lemma 2](#))) we generate a  $\mathcal{T}_{\text{UT}}(\mathcal{P})$ -CT  $\mathfrak{A}'$  with  $|\mathfrak{A}'| = |\mathfrak{A}|$ . The CT  $\mathfrak{A}'$  has the same nodes and edges as  $\mathfrak{A}$ . If a node  $v$  is labeled by  $(p_v : t_v)$  in  $\mathfrak{A}$ , then it is labeled by  $(p_v : t'_v)$  in  $\mathfrak{A}'$  where we construct the labeling  $t'_v$  recursively and ensure that for all nodes  $v$  we have

$$b(t_v) = b(t'_v) \text{ and } \text{Pos}_{D^\#}(t_v) \setminus \text{Junk}(t_v) \subseteq \text{Pos}_{D^\#}(t'_v). \quad (23)$$

Here, for any term  $t_v$ , let  $\text{Junk}(t_v)$  be the set of positions that can never be used for a rewrite step with an ADP that contains annotations. Due to (23), every infinite path of  $\mathfrak{A}'$  still contains infinitely many rewrite steps at annotations.

*Completeness:* By [Theorem 6](#).  $\square$

**Example 14.** For AST,  $\text{Proc}_{\text{UT}}$  transforms  $\{\#14, \#15, b\#16 - b\#20\}$  from [Example 11](#) into  $\{\#14', \#15', b\#16 - b\#20\}$  with

$$\begin{aligned} \text{loop1}(y) &\rightarrow \{1/2 : L1(\text{double}(y)), 1/2 : \text{loop2}(\text{double}(y))\}^{\text{true}} \#14' \\ \text{loop1}(y) &\rightarrow \{1/3 : L1(\text{triple}(y)), 2/3 : \text{loop2}(\text{triple}(y))\}^{\text{true}} \#15' \end{aligned}$$

The reason is that the left-hand sides of the only ADPs with annotations in the ADP problem have the root  $\text{loop1}$ . Thus,  $L2$ -,  $D$ -, or  $T$ -terms are not usable.

##### 4.2.2. Usable terms processor for $\text{iAST}$

As for the dependency graph processor, for  $\text{iAST}$  we only consider innermost rewriting in the definitions.

**Theorem 10** (Usable Terms Processor for  $\text{iAST}$ ). *Let  $\ell_1 \in \mathcal{T}$  and  $\mathcal{P}$  be an ADP problem. We call  $t \in \mathcal{T}^\#$  with  $\text{root}(t) \in D^\#$  innermost usable w.r.t.  $\ell_1$  and  $\mathcal{P}$  if there are substitutions  $\sigma_1, \sigma_2$  and an  $\ell_2 \rightarrow \mu_2 \in \mathcal{P}$  where  $\mu_2$  contains an annotated symbol, such that  $\#_{\{\varepsilon\}}(t)\sigma_1 \xrightarrow[\text{np}(\mathcal{P})]{*} \ell_2^\# \sigma_2$  and both  $\ell_1 \sigma_1$  and  $\ell_2 \sigma_2$  are in  $\text{ANF}_{\mathcal{P}}$ . Let  $\Delta_{\ell, \mathcal{P}}^{\text{i}}(s) = \{\pi \in \text{Pos}_{D^\#}(s) \mid s|_\pi \text{ is innermost usable w.r.t. } \ell \text{ and } \mathcal{P}\}$  and  $\mathcal{T}_{\text{UT}}^{\text{i}}(\mathcal{P}) = \{\ell \rightarrow \{p_1 : \#_{\Delta_{\ell, \mathcal{P}}^{\text{i}}(r_1)}(r_1), \dots, p_k : \#_{\Delta_{\ell, \mathcal{P}}^{\text{i}}(r_k)}(r_k)\}^m \mid \ell \rightarrow \{p_1 : r_1, \dots, p_k : r_k\}^m \in \mathcal{P}\}$ . Then  $\text{Proc}_{\text{UT}}^{\text{i}}(\mathcal{P}) = \{\mathcal{T}_{\text{UT}}^{\text{i}}(\mathcal{P})\}$  is sound and complete for  $\text{iAST}$ .*

*Proof Sketch. Soundness:* Similar to the proof of [Theorem 9](#), just considering innermost rewrite steps.

*Completeness:* By [Theorem 6](#).  $\square$

**Example 15.** For  $\mathcal{R}_{\text{alg}}$ , up to now the proof of  $\text{iAST}$  is identical to the proof of AST, i.e., here the results of  $\text{Proc}_{\text{UT}}^{\text{i}}$  on the ADP problems resulting from the (innermost) dependency graph processor are the same as the results of  $\text{Proc}_{\text{UT}}$ .

When considering  $\mathcal{R}_2$  and its ADPs  $DP(\mathcal{R}_2)$  from [Example 6](#) again,  $\text{Proc}_{\text{UT}}^{\text{i}}$  replaces  $\#9$  by  $g \rightarrow \{3/4 : d(G), 1/4 : 0\}^{\text{true}} \#9'$ .

#### 4.3. Usable rules processor

Next, we consider the *usable rules processor*. Here, we will see a major difference between the ADP frameworks for AST and  $\text{iAST}$ , since it will turn out that this processor is only sound for  $\text{iAST}$ , but not for AST. Therefore, this time we consider innermost rewriting first.

### 4.3.1. Usable rules processor for iAST

In an innermost rewrite step, all variables of the used rule are instantiated with normal forms. The *usable rules processor* detects rules that cannot be used below annotations in right-hand sides of ADPs when their variables are instantiated with normal forms. For these rules we can set their flag to false, indicating that the annotated subterms on their right-hand sides may still lead to a non-iAST sequence, but the context of these annotations is irrelevant.

**Theorem 11** (Usable Rules Processor for iAST). *For every  $f \in \Sigma^\sharp$  and ADP problem  $\mathcal{P}$ , let  $\text{Rules}_{\mathcal{P}}(f) = \{\ell \rightarrow \mu^{\text{true}} \in \mathcal{P} \mid \text{root}(\ell) = f\}$ . Moreover, for every  $t \in \mathcal{T}^\sharp$ , the usable rules  $\mathcal{U}_{\mathcal{P}}(t)$  of  $t$  w.r.t.  $\mathcal{P}$  are recursively defined as follows:*

$$\begin{aligned} \mathcal{U}_{\mathcal{P}}(t) &= \emptyset, & \text{if } t \in \mathcal{V} \text{ or } \mathcal{P} = \emptyset \\ \mathcal{U}_{\mathcal{P}}(f(t_1, \dots, t_n)) &= \text{Rules}_{\mathcal{P}}(f) \cup \bigcup_{j=1}^n \mathcal{U}_{\mathcal{P}'}(t_j) \cup \bigcup_{\substack{\ell \rightarrow \mu^{\text{true}} \in \text{Rules}_{\mathcal{P}}(f), \\ r \in \text{Supp}(\mu)}} \mathcal{U}_{\mathcal{P}'}(b(r)), & \text{otherwise,} \end{aligned}$$

where  $\mathcal{P}' = \mathcal{P} \setminus \text{Rules}_{\mathcal{P}}(f)$ .

The usable rules of  $\mathcal{P}$  are  $\mathcal{U}(\mathcal{P}) = \bigcup_{\ell \rightarrow \mu^m \in \mathcal{P}, r \in \text{Supp}(\mu), t \in \mathcal{T}^\sharp} \mathcal{U}_{\mathcal{P}}(t^\sharp)$ . Then, the processor  $\text{Proc}_{\text{UR}}^1(\mathcal{P}) = \{\mathcal{U}(\mathcal{P}) \cup \{\ell \rightarrow \mu^{\text{false}} \mid \ell \rightarrow \mu^m \in \mathcal{P} \setminus \mathcal{U}(\mathcal{P})\}\}$  is sound and complete for iAST, i.e., we turn the flag  $m$  of all non-usable rules to false.

*Proof Sketch. Soundness:* Due to the Starting Lemma (Lemma 2) we can restrict ourselves to innermost  $\mathcal{P}$ -CTs that start with a term  $t \in \mathcal{T}^\sharp$  such that  $b(t) = s\theta \in \text{ANF}_{\mathcal{P}}$  for some  $s \rightarrow \dots \in \mathcal{P}$ . By the definition of usable rules, rules  $\ell \rightarrow \mu^m \in \mathcal{P}$  that are not usable are never applied below an annotated symbol in such  $\mathcal{P}$ -CTs. Hence, we can also regard  $\mathfrak{A}$  as a  $\overline{\mathcal{P}}$ -CT for  $\overline{\mathcal{P}} = \mathcal{U}(\mathcal{P}) \cup \{\ell \rightarrow \mu^{\text{false}} \mid \ell \rightarrow \mu^m \in \mathcal{P} \setminus \mathcal{U}(\mathcal{P})\}$ .

*Completeness:* By Theorem 6.  $\square$

Note that the innermost evaluation strategy is not affected by our usable rules processor, since it does not remove rules but only changes their flag to false. This is different from the non-probabilistic DP framework, where the usable rules processor from Theorem 2 reduces the number of rules. This may result in new redexes that are allowed for innermost rewriting. So in [9] one has to extend DP problems by an additional component to achieve completeness of this processor (see Footnote 5), whereas the usable rules processor in our new ADP framework is complete.

**Example 16.** The ADP problem  $\{(\#9)', (\#11)\}$  which we obtained for the iAST proof of  $\mathcal{R}_2$  in Example 15 has no subterms below annotations. So both rules are not usable and we set their flags to false which leads to

$$\begin{aligned} G &\rightarrow \{3/4 : d(G), 1/4 : 0\}^{\text{false}} & (\#9'') \\ D(x) &\rightarrow \{1 : c(x, x)\}^{\text{false}} & (\#11') \end{aligned}$$

**Example 17.** Consider the ADP problem  $\{(\#14'), (\#15'), b(\#16) - b(\#20)\}$  from Example 14 that resulted from applying the innermost dependency graph processor and the innermost usable terms processor to  $DP(\mathcal{R}_{\text{alg}})$ . When proving iAST of  $\mathcal{R}_{\text{alg}}$ , only the double- and triple-ADPs  $b(\#17) - b(\#20)$  are usable, since the defined symbols double and triple occur below an annotated symbol L1. So we can set the flag of all other ADPs in this problem to false. All other ADP problems resulting from the innermost dependency graph and the innermost usable terms processor in this example have no usable rules, hence we can turn the flag of all ADPs in these ADP problems to false.

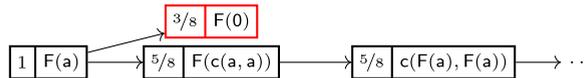
### 4.3.2. Usable rules processor for AST

For full rewriting, the usable rules processor is unsound. This is already the case for non-probabilistic rewriting, but in the classical DP framework there nevertheless exist processors for full rewriting based on usable rules which rely on taking the  $C_{\mathcal{E}}$ -rules  $h(x, y) \rightarrow x$  and  $h(x, y) \rightarrow y$  for a fresh function symbol  $h$  into account, see, e.g., [9, 10, 12, 52]. However, the following example shows that this is not possible for AST.

**Example 18.** Consider the following ADP problem.

$$\begin{aligned} a &\rightarrow \{5/8 : c(a, a), 3/8 : 0\}^{\text{true}} \\ f(c(x_1, x_2)) &\rightarrow \{1 : c(F(x_1), F(x_2))\}^{\text{true}} \end{aligned}$$

Although the first ADP has no annotations, the ADP problem is not AST:



This represents a random walk biased towards non-termination, where the number of  $F(a)$  subterms increases by one with probability  $5/8$  or decreases by one with probability  $3/8$ .

The ADP problem has no usable rules and thus, the usable rule processor would turn the flag of all ADPs to false. However, if we now start with  $F(a)$  (or  $F(A)$ ), rewriting  $a$  (or  $A$ ) would remove the annotation of  $F$  above, since the flag indicates whether the ADP can be used to rewrite arguments of an annotated symbol, i.e.,  $F(a) \hookrightarrow \{5/8 : f(c(a, a)), 3/8 : f(0)\}$ . Hence, then all CTs are finite. This also holds when adding the  $C_{\mathcal{E}}$ -ADPs  $h(x, y) \rightarrow \{1 : x\}^{\text{true}}$  and  $h(x, y) \rightarrow \{1 : y\}^{\text{true}}$ .

Thus, even integrating the  $C_{\mathcal{E}}$ -rules to represent non-determinism would not allow a usable rule processor for AST. Moreover, the corresponding proofs in the non-probabilistic setting rely on the minimality property, which does not hold in the probabilistic setting, see Remark 2.

#### 4.4. Reduction pair processor

Next we consider the reduction pair processor, whose adaption is the same for AST and  $\mathfrak{i}$ AST. In the probabilistic setting, we cannot use arbitrary term orders, but as in [41], we have to consider algebras that allow the definition of an *expected value*, called *barycentric*  $\Sigma^\sharp$ -algebras. Compared to [41], we have to require even more structure, i.e., we have to restrict ourselves to algebras that form an  $\mathbb{R}_{\geq 0}$ -semimodule. We will discuss the reasons for requiring  $\mathbb{R}_{\geq 0}$ -semimodules below.

**Definition 10** ( $\mathbb{R}_{\geq 0}$ -Semimodule). An  $\mathbb{R}_{\geq 0}$ -semimodule  $(A, +, \cdot, 0_A)$  (or simply  $A$ ) is a set  $A$  (called the universe) together with two binary functions  $+$  :  $(A \times A) \rightarrow A$  and  $\cdot$  :  $(\mathbb{R}_{\geq 0} \times A) \rightarrow A$  such that  $(A, +, 0_A)$  is a commutative monoid (i.e.,  $+$  is commutative and associative with identity element  $0_A$ ), and such that for all  $a, b \in A, r, s \in \mathbb{R}_{\geq 0}$  we have:  $r \cdot (a + b) = r \cdot a + r \cdot b$ ,  $(r + s) \cdot a = r \cdot a + s \cdot a$ ,  $r \cdot (s \cdot a) = (r \cdot s) \cdot a$ ,  $1 \cdot a = a$ , and  $0 \cdot a = r \cdot 0_A = 0_A$ .

**Definition 11** (Barycentric  $\Sigma^\sharp$ -Algebra,  $\mathbb{N}$ -collapsible). Let  $A$  be an  $\mathbb{R}_{\geq 0}$ -semimodule. We define the barycentric operation  $\mathbb{E}_A : \text{FDist}(A) \rightarrow A$  as  $\mathbb{E}_A(\{p_1 : a_1, \dots, p_k : a_k\}) = p_1 \cdot a_1 + \dots + p_k \cdot a_k$ . We write  $\mathbb{E}$  instead of  $\mathbb{E}_A$  if  $A$  is clear from the context. A function  $f : A^n \rightarrow B$  on two  $\mathbb{R}_{\geq 0}$ -semimodules  $A$  and  $B$  is called *concave* w.r.t. a relation  $\succcurlyeq$  on  $B$  if we have  $f(\dots, \mathbb{E}_A(\mu), \dots) \succcurlyeq \mathbb{E}_B(f(\dots, \mu, \dots))$ , where we define  $f(\dots, \{p_1 : a_1, \dots, p_k : a_k\}, \dots) = \{p_1 : f(\dots, a_1, \dots), \dots, p_k : f(\dots, a_k, \dots)\}$  for every  $\mu \in \text{FDist}(A)$ .

A *barycentric*  $\Sigma^\sharp$ -algebra is a  $\Sigma^\sharp$ -algebra  $(A, >)$  on the universe of a  $\mathbb{R}_{\geq 0}$ -semimodule  $A$  such that  $>$  is compatible with the semimodule structure and the addition is monotonic w.r.t.  $>$ .<sup>10</sup> Moreover, for every  $f \in \Sigma^\sharp$ ,  $f_A$  must be concave with respect to the reflexive closure  $\succcurlyeq$  of  $>$ . We say that  $(A, >)$  is  $\mathbb{N}$ -collapsible if there exists a concave function  $g : A \rightarrow \mathbb{N}$  that embeds  $>$  into  $>$  on the natural numbers, i.e., for all ground terms  $t_1, t_2 \in \mathcal{T}^\sharp$ ,  $\mathcal{A}(t_1) > \mathcal{A}(t_2)$  implies  $g(\mathcal{A}(t_1)) > g(\mathcal{A}(t_2))$ .

For example, both polynomial interpretations and matrix interpretations [48] are  $\mathbb{N}$ -collapsible barycentric  $\Sigma^\sharp$ -algebras, provided that  $\text{Pol}$  maps each function symbol to a *multilinear* polynomial, i.e., a polynomial whose monomials have the form  $c \cdot x_1^{e_1} \cdot \dots \cdot x_n^{e_n}$  with  $c \in \mathbb{N}$  and  $e_1, \dots, e_n \in \{0, 1\}$ , to ensure concavity. (Again, we will discuss the reasons for requiring concavity of  $f_A$  and for  $\mathbb{N}$ -collapsibility below.)

In the following, let  $(\mathcal{A}, >)$  be an  $\mathbb{N}$ -collapsible, barycentric  $\Sigma^\sharp$ -algebra. The constraints of the reduction pair processor of [Theorem]12 are based on the conditions of a ranking function for AST as in [23]. If one proves AST by considering the rules  $\ell \rightarrow \mu^m$  of a PTRS directly, then one needs a *monotonic* algebra and requires a weak decrease when comparing  $\mathcal{A}(\ell)$  to the expected value  $\mathbb{E}(\mathcal{A}(\mu))$  of the right-hand side, where we extend  $\mathcal{A}$  to multi-distributions by defining  $\mathcal{A}(\{p_1 : t_1, \dots, p_k : t_k\}) = \{p_1 : \mathcal{A}(t_1), \dots, p_k : \mathcal{A}(t_k)\}$ . Moreover, at least one  $\mathcal{A}(r)$  with  $r \in \text{Supp}(\mu)$  must be strictly smaller than  $\mathcal{A}(\ell)$ , see [42].

For ADPs, we do not use the interpretation  $\mathcal{A}(r)$  of a term  $r$  directly, but we take the occurring annotations into account. Since only the annotated subterms may lead to non-AST sequences, we only consider the interpretations of these subterms. Hence, we compare the  $\sharp$ -sums<sup>11</sup> of the terms  $r$  in ADPs, i.e.,  $\mathcal{A}_{\text{sum}}^\sharp(r) = \sum_{r \leq_{\mathcal{A}} r'} \mathcal{A}(r')$  sums up the values of  $r$ 's annotated subterms. So for example, we have  $\mathcal{A}_{\text{sum}}^\sharp(\text{D}(\text{G})) = \mathcal{A}(\text{D}(\text{g})) + \mathcal{A}(\text{G})$ .

If in each rewrite step  $s \hookrightarrow_{\mathcal{P}} \mu$  of a chain tree  $\mathfrak{T}$ , the  $\sharp$ -sum is weakly decreasing in expectation ( $\mathcal{A}_{\text{sum}}^\sharp(s) \succcurlyeq \mathbb{E}(\mathcal{A}_{\text{sum}}^\sharp(\mu))$ ), and additionally, at least one  $\mathcal{A}_{\text{sum}}^\sharp(r)$  with  $r \in \text{Supp}(\mu)$  is strictly smaller than  $\mathcal{A}_{\text{sum}}^\sharp(s)$ , then we can conclude that  $\mathfrak{T}$  converges with probability 1. Again,  $\mathcal{A}_{\text{sum}}^\sharp$  is extended to multi-distributions by defining  $\mathcal{A}_{\text{sum}}^\sharp(\{p_1 : t_1, \dots, p_k : t_k\}) = \{p_1 : \mathcal{A}_{\text{sum}}^\sharp(t_1), \dots, p_k : \mathcal{A}_{\text{sum}}^\sharp(t_k)\}$ . So for a rewrite step of the form  $\text{G} \hookrightarrow_{\mathcal{P}} \{3/4 : \text{D}(\text{G}), 1/4 : 0\}$  one would for example obtain the constraints  $\mathcal{A}_{\text{sum}}^\sharp(\text{G}) = \mathcal{A}(\text{G}) \succcurlyeq \mathbb{E}(\{3/4 : \mathcal{A}_{\text{sum}}^\sharp(\text{D}(\text{G})), 1/4 : \mathcal{A}_{\text{sum}}^\sharp(0)\}) = 3/4 \cdot \mathcal{A}_{\text{sum}}^\sharp(\text{D}(\text{G})) + 1/4 \cdot \mathcal{A}_{\text{sum}}^\sharp(0) = 3/4 \cdot (\mathcal{A}(\text{D}(\text{g})) + \mathcal{A}(\text{G})) + 1/4 \cdot 0 = 3/4 \cdot (\mathcal{A}(\text{D}(\text{g})) + \mathcal{A}(\text{G}))$  and  $\mathcal{A}_{\text{sum}}^\sharp(\text{G}) = \mathcal{A}(\text{G}) > \mathcal{A}_{\text{sum}}^\sharp(0) = 0$ .

**Lemma 3** (Proving AST of Chain Trees with Algebras). Let  $\mathfrak{T}$  be a  $\mathcal{P}$ -CT and let  $(\mathcal{A}, >)$  be an  $\mathbb{N}$ -collapsible, barycentric  $\Sigma^\sharp$ -algebra such that for every rewrite step  $s \hookrightarrow_{\mathcal{P}} \mu$  in  $\mathfrak{T}$  we have:  $(A) : \mathcal{A}_{\text{sum}}^\sharp(s) \succcurlyeq \mathbb{E}(\mathcal{A}_{\text{sum}}^\sharp(\mu))$  and  $(B) : \exists r \in \text{Supp}(\mu) : \mathcal{A}_{\text{sum}}^\sharp(s) > \mathcal{A}_{\text{sum}}^\sharp(r)$ . Then we have  $|\mathfrak{T}| = 1$ , i.e.,  $\mathfrak{T}$  converges with probability 1.

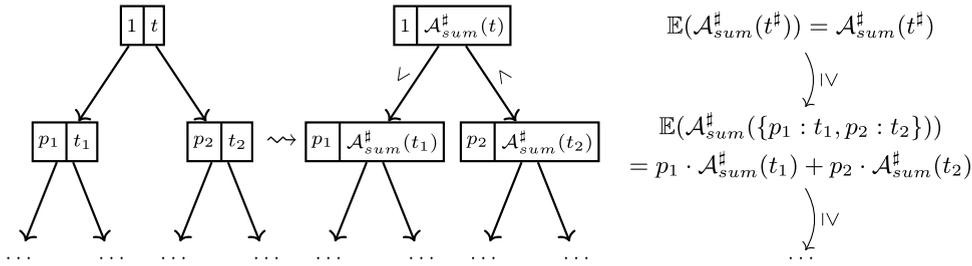
Fig. 6 visualizes the interpretation of terms in an example chain tree. As required by (A), the expected  $\sharp$ -sum is weakly decreasing in each step from a node to its children. Moreover, the step at the root of the example CT satisfies (B), since the left successor  $t_1$  has a strictly smaller  $\sharp$ -sum than  $t$ . The  $\sharp$ -sum of the right successor  $t_2$  may even be larger than  $t$ , but the expected  $\sharp$ -sum needs to be weakly decreasing.

Instead of imposing (usually infinitely many) constraints for all possible rewrite steps  $s \hookrightarrow_{\mathcal{P}} \mu$  as in Lemma 3, we only want to impose finitely many constraints which correspond to the finitely many ADPs in the ADP problem  $\mathcal{P}$ . If these inequality constraints hold for the ADPs, then they also hold for arbitrary rewrite steps if  $(\mathcal{A}, >)$  is *weakly monotonic*. So as in the non-probabilistic reduction pair processor of Theorem 3, due to the usage of  $\sharp$ -sums, weak instead of strict monotonicity is sufficient. However, as in [41], the requirement of concavity of all functions  $f_A$  is needed to ensure that we also have weak monotonicity w.r.t. expected values. So  $\mathcal{A}(t) \succcurlyeq \mathbb{E}(\mathcal{A}(\mu))$  implies

$$\begin{aligned} \mathcal{A}(f(\dots, t, \dots)) &= f_{\mathcal{A}}(\dots, \mathcal{A}(t), \dots) \\ &\succcurlyeq f_{\mathcal{A}}(\dots, \mathbb{E}(\mathcal{A}(\mu)), \dots) && \text{by weak monotonicity} \\ &\succcurlyeq \mathbb{E}(f_{\mathcal{A}}(\dots, \mathcal{A}(\mu), \dots)) && \text{by concavity} \end{aligned}$$

<sup>10</sup> A relation  $>$  is *compatible* with the semimodule structure of  $(A, +, \cdot, 0_A)$  if  $x > y$  implies  $x + z > y + z$  and  $r \cdot x > r \cdot y$  for all  $x, y, z \in A$  and  $r \in \mathbb{R}_{\geq 0}$ . Addition is *monotonic* w.r.t.  $>$  if for all  $x, y \in A$  we have  $x + y \succcurlyeq x$  and  $x + y > x$  if  $y \neq 0_A$ .

<sup>11</sup> This is the reason why we require  $\mathbb{R}_{\geq 0}$ -semimodules instead of simply requiring the existence of a barycentric operation  $\mathbb{E}$  as in [41].



**Fig. 6.** Each term  $r$  in the chain tree (left) is mapped to the corresponding  $\#$ -sum  $\mathcal{A}_{sum}^\#(r)$  (middle). The expected  $\#$ -sum (right) is always weakly decreasing.

Similar to [Theorem 3](#), we do not need to find an algebra that orients all ADPs strictly (i.e., where all rewrite steps satisfy property (B) of [Lemma 3](#)). This still allows us to simplify the problem, because then it suffices to consider only those CTs where the non-strictly oriented ADPs are applied infinitely often on annotated positions.

So to ensure the properties of [Lemma 3](#) for each reduction step, we impose the following constraints for all ADPs (which adapt the constraints (1) - (3) from [Theorem 3](#) to the probabilistic setting): First, (1) we require a weak decrease when comparing the annotated left-hand side with the expected value of the  $\#$ -sums in the right-hand sides (this guarantees that (A) holds for every a-rewrite step performed with this ADP). The processor then removes the annotations from those ADPs where (2a) there is at least one right-hand side  $r$  whose  $\#$ -sum is strictly decreasing. If  $m = \text{true}$ , then in addition, (2b) the corresponding non-annotated right-hand side  $b(r)$  must be at least weakly decreasing to ensure that nested annotations behave monotonically.<sup>12</sup> So (2) guarantees that (B) holds for every a-rewrite step performed with this ADP. Finally, (3) for every rule with the flag true (which can therefore be used for steps below annotations), the expected value must be weakly decreasing when removing all annotations (this guarantees that (A) holds for every n-rewrite step performed with this ADP).

**Theorem 12** (Reduction Pair Processor for AST & iAST). *Let  $(\mathcal{A}, >)$  be a weakly monotonic,  $\mathbb{N}$ -collapsible, barycentric  $\Sigma^\#$ -algebra. Let  $\mathcal{P} = \mathcal{P}_{\geq} \uplus \mathcal{P}_{>}$  with  $\mathcal{P}_{>} \neq \emptyset$  where:*

- |     |   |   |  |
|-----|---|---|--|
| (1) | $\forall \ell \rightarrow \mu^m \in \mathcal{P}$  | : | $\mathcal{A}_{sum}^\#(\ell^\#) \geq \mathbb{E}(\mathcal{A}_{sum}^\#(\mu))$ |
| (2) | (a) $\forall \ell \rightarrow \mu^m \in \mathcal{P}_{>} : \exists r \in \text{Supp}(\mu)$ | : | $\mathcal{A}_{sum}^\#(\ell^\#) > \mathcal{A}_{sum}^\#(r)$                  |
|     | (b) If $m = \text{true}$ , then we additionally have                                      | : | $\mathcal{A}(\ell) \geq \mathcal{A}(b(r))$                                 |
| (3) | $\forall \ell \rightarrow \mu^{\text{true}} \in \mathcal{P}$                              | : | $\mathcal{A}(\ell) \geq \mathbb{E}(\mathcal{A}(b(\mu)))$                   |

Then  $\text{Proc}_{\text{RP}}(\mathcal{P}) = \text{Proc}_{\text{RP}}^{\text{i}}(\mathcal{P}) = \{\mathcal{P}_{\geq} \cup b(\mathcal{P}_{>})\}$  is sound and complete for (i)AST.

*Proof Sketch. Soundness:* One first lifts the properties (1) - (3) from ADPs to the properties (A) and (B) of [Lemma 3](#) for rewrite steps with  $\hookrightarrow_p$ . Then, by adapting proof ideas of [23] to the setting of term rewriting, one can show that  $|\mathfrak{X}| = 1$  holds for any  $\mathcal{P}$ -CT  $\mathfrak{X}$  where every infinite path uses infinitely many steps with ADPs from  $\mathcal{P}_{>}$  at annotated positions. Thus, by the  $\mathcal{P}$ -Partition Lemma ([Lemma 1](#)) it suffices to show that  $\mathcal{P}_{\geq} \cup b(\mathcal{P}_{>})$  is AST.

*Completeness:* By [Theorem 6](#).  $\square$

Note that Requirement (2) only ensures a strict decrease (i.e., Property (B)) for rewrite steps of type (at) or (af). This is the reason why we require an infinite number of such steps within each infinite path of a chain tree, see [Definition 6](#). Moreover,  $\mathbb{N}$ -collapsibility implies that in addition to well-foundedness of  $>$ , for every  $a_0 \in A$ , the length of every sequence  $a_0 > a_1 > \dots$  is bounded. This is needed in order to show that the properties (A) and (B) are indeed sufficient for AST. Note that we only need a strict decrease on terms but not on their expected values. Therefore, in contrast to [41], we can restrict ourselves to functions  $g$  which map to the natural instead of the real numbers.

**Example 19.** To conclude iAST for  $\mathcal{R}_2$  we have to remove all remaining annotations in the ADP problem  $\{(\#9''), (\#11')\}$  with

$$\begin{aligned} G &\rightarrow \{3/4 : d(G), 1/4 : 0\}^{\text{false}} && (\#9'') \\ D(x) &\rightarrow \{1 : c(x, x)\}^{\text{false}} && (\#11') \end{aligned}$$

from [Example 16](#). Then another application of the dependency graph processor yields the empty set of ADP problems. Here, we can use the reduction pair processor with the polynomial interpretation  $(\mathcal{A}, >) = (\text{Pol}, >)$  that maps  $G$  to 1, and all other symbols to 0. Then  $(\#11')$  is weakly decreasing, and  $(\#9'')$  is strictly decreasing, since (1)  $\text{Pol}_{sum}^\#(G) = \text{Pol}(G) = 1 \geq 3/4 \cdot \text{Pol}_{sum}^\#(d(G)) + 1/4 \cdot \text{Pol}_{sum}^\#(0) = 3/4 \cdot \text{Pol}(G) = 3/4$  and (2)  $\text{Pol}_{sum}^\#(G) = \text{Pol}(G) = 1 > \text{Pol}_{sum}^\#(0) = 0$ . Thus, the annotation of  $G$  in  $(\#9'')$  is deleted.

Note that this polynomial interpretation would also satisfy the constraints for  $\text{DP}(\mathcal{R}_2) = \{(\#9), (\#11)\}$  from [Example 6](#), i.e., it would allow us to remove the annotations from the canonical ADP directly. Hence, if we extended our approach for AST to GVRFs that can

<sup>12</sup> For example, it ensures that  $\mathcal{A}(G) > \mathcal{A}(H)$  also implies that the  $\#$ -sum of  $F(G)$  is greater than the  $\#$ -sum of  $F(H)$ , i.e.,  $\mathcal{A}(G) > \mathcal{A}(H)$  must imply that  $\mathcal{A}_{sum}^\#(F(G)) = \mathcal{A}(F(G)) + \mathcal{A}(G) > \mathcal{A}(F(H)) + \mathcal{A}(H) = \mathcal{A}_{sum}^\#(F(H))$ , which is ensured by  $\mathcal{A}(g) \geq \mathcal{A}(h)$ .

duplicate annotations, then the reduction pair processor would be unsound, as it would allow us to falsely “prove” AST of  $DP(\mathcal{R}_2)$ . The problem is that we compare terms with annotations via their  $\sharp$ -sum, but for duplicating ADPs like (#11),  $\text{Pol}(d(x)) \geq \text{Pol}(c(x, x))$  does not imply  $\text{Pol}_{sum}^{\sharp}(d(G)) \geq \text{Pol}_{sum}^{\sharp}(c(G, G))$  since  $\text{Pol}_{sum}^{\sharp}(d(G)) = \text{Pol}(G)$  and  $\text{Pol}_{sum}^{\sharp}(c(G, G)) = \text{Pol}(G) + \text{Pol}(G)$ .

**Example 20.** To prove AST for  $\mathcal{R}_{alg}$ , we also have to remove all annotations from all remaining sub-problems. For instance, for the sub-problem  $\mathcal{P}_{(\sharp14'), (\sharp15')} = \{(\sharp14'), (\sharp15'), b(\sharp16) - b(\sharp20)\}$  from Example 14, we can use the reduction pair processor with the polynomial interpretation  $(A, >) = (\text{Pol}, >)$  that maps  $s(x)$  to  $x + 1$ ,  $\text{double}(x)$  to  $2x$ ,  $\text{triple}(x)$  to  $3x$ ,  $L1(x)$  to  $1$ , and all other symbols to  $0$ . Then  $(\sharp15')$  is strictly decreasing, since (1)  $\text{Pol}_{sum}^{\sharp}(L1(y)) = \text{Pol}(L1(y)) = 1 \geq 1/3 \cdot \text{Pol}_{sum}^{\sharp}(L1(\text{triple}(y))) + 2/3 \cdot \text{Pol}_{sum}^{\sharp}(\text{loop2}(\text{triple}(y))) = 1/3$  and (2)  $\text{Pol}_{sum}^{\sharp}(L1(y)) = \text{Pol}(L1(y)) = 1 > \text{Pol}_{sum}^{\sharp}(\text{loop2}(\text{triple}(y))) = 0$ . Similarly,  $(\sharp14')$  is also strictly decreasing, and we can remove all annotations from this ADP problem. One can find similar interpretations to delete the remaining annotations also from the other remaining sub-problems. This proves AST for  $DP(\mathcal{R}_{alg})$ , and hence for  $\mathcal{R}_{alg}$ .

If we only want to prove iAST for  $\mathcal{R}_{alg}$ , then the application of the reduction pair processor becomes significantly easier. Except for  $b(\sharp17) - b(\sharp20)$  in the ADP problem  $\mathcal{P}_{(\sharp14'), (\sharp15')}$  above, all other ADPs have the flag false after the usable rules processor, see Example 17. Therefore, (3) and (2b) from Theorem 12 do not apply, and thus, we obtain significantly fewer constraints which makes the search for interpretations easier for the SMT solver. For example, when proving AST instead of iAST, all ADPs in the ADP problem  $\{(\sharp19), b(\sharp14) - b(\sharp18), b(\sharp20)\}$  have the flag true and thus, here we have to find a polynomial interpretation which also makes the ADPs  $b(\sharp14) - b(\sharp16)$  weakly decreasing, which we do not require when proving iAST.

#### 4.5. Probability removal processor

In proofs with the ADP framework, one may obtain ADP problems  $\mathcal{P}$  with a non-probabilistic structure, i.e., where every ADP has the form  $\ell \rightarrow \{1 : r\}^m$ . Then the *probability removal processor* allows us to switch to ordinary (non-probabilistic) DPs.

**Theorem 13** (Probability Removal Processor). *Let  $\mathcal{P}$  be an ADP problem where every ADP in  $\mathcal{P}$  has the form  $\ell \rightarrow \{1 : r\}^m$ . Let  $\text{dp}(\mathcal{P}) = \{\ell^{\sharp} \rightarrow t^{\sharp} \mid \ell \rightarrow \{1 : r\}^m \in \mathcal{P}, t \preceq_{\sharp} r\}$ . Then  $\mathcal{P}$  is (i)AST if the non-probabilistic DP problem  $(\text{dp}(\mathcal{P}), \text{np}(\mathcal{P}))$  is (innermost) terminating. So the processor  $\text{Proc}_{PR}(\mathcal{P}) = \text{Proc}_{PR}^i(\mathcal{P}) = \emptyset$  is sound and complete for (i)AST if  $(\text{dp}(\mathcal{P}), \text{np}(\mathcal{P}))$  is (i)SN.*

*Proof Sketch.* If: Due to the trivial probabilities in each ADP  $\ell \rightarrow \{1 : r\}^m$ , a  $\mathcal{P}$ -CT  $\mathfrak{Z}$  with  $|\mathfrak{Z}| < 1$  must be an infinite path, where we use infinitely many rewrite steps at annotated positions. This path gives rise to a  $(\text{dp}(\mathcal{P}), \text{np}(\mathcal{P}))$ -chain following the construction of the non-probabilistic chain criterion.

*Only if:* Every infinite  $(\text{dp}(\mathcal{P}), \text{np}(\mathcal{P}))$ -chain gives rise to an infinite  $\mathcal{P}$ -CT  $\mathfrak{Z}$ , and because of the trivial probabilities in  $\mathcal{P}$ , we must have  $|\mathfrak{Z}| = 0$ .  $\square$

**Remark 3.** One may wonder how we can end up in an ADP problem where every ADP has the form  $\ell \rightarrow \{1 : r\}^m$ , if this does not already hold for the initial ADP problem. The reason is that the processors of our ADP framework do not remove ADPs. However, sometimes it is possible to get rid of certain ADPs with non-trivial probabilities. Note that the right-hand sides of ADPs like  $f(x) \rightarrow \{1/2 : g(x), 1/2 : h(x)\}^{\text{false}}$  without annotations and the flag false are completely irrelevant. We only keep such ADPs in our ADP problem to ensure that applying ADP processors does not change the set of normal forms, such that the innermost rewrite strategy is not affected. Thus, it is sound and complete to replace such ADPs  $\ell \rightarrow \mu^{\text{false}}$  by a new ADP  $\ell \rightarrow \{1 : \text{stop}\}^{\text{false}}$  for a fresh constant stop. In this way, we keep their original left-hand sides but get rid of any non-trivial probabilities in their right-hand side.

#### 4.6. Subterm criterion

Next, we consider the *subterm criterion* of [12] which is a purely syntactical way of removing loops from the dependency graph. This processor has not yet been adapted to the probabilistic setting in our previous conference papers, and we explain the problems that one has to solve for such an adaption. In the non-probabilistic setting, the subterm criterion searches for a *simple projection* for every annotated symbol  $f^{\sharp}$ , i.e., a projection *proj* which maps  $f^{\sharp}$  to one of its argument positions  $i$  and replaces every term  $f^{\sharp}(\dots)$  by its  $i$ -th argument. If the projection of the right-hand side  $\text{proj}(r^{\sharp})$  is a subterm of the projected left-hand side  $\text{proj}(\ell^{\sharp})$  for every dependency pair  $\ell^{\sharp} \rightarrow r^{\sharp} \in \mathcal{P}$ , then we can remove those dependency pairs from  $\mathcal{P}$  where  $\text{proj}(r^{\sharp})$  is a proper subterm of  $\text{proj}(\ell^{\sharp})$ .

**Definition 12** (Simple Projection). *A simple projection*<sup>12</sup> *proj* that assigns to every  $n$ -ary annotated symbol  $f^{\sharp} \in \mathcal{D}^{\sharp}$  with  $n \geq 1$  an argument position  $i \in \{1, \dots, n\}$ . We extend *proj* to terms with annotated root symbol as follows: For  $t_1, \dots, t_n \in \mathcal{T}$  with  $n \geq 1$  and  $\text{proj}(f^{\sharp}) = i$ , we define  $\text{proj}(f^{\sharp}(t_1, \dots, t_n)) = t_i$ . If  $f^{\sharp}$  has arity 0, then we define  $\text{proj}(f^{\sharp}) = f^{\sharp}$ .

##### 4.6.1. Subterm criterion processor for AST

In the non-probabilistic setting, if  $\text{proj}(\ell^{\sharp}) \supseteq \text{proj}(r^{\sharp})$  holds for all  $\ell^{\sharp} \rightarrow r^{\sharp}$  in a set of DPs  $\mathcal{P}$ , then one can remove all DPs from  $\mathcal{P}$  where we have  $\text{proj}(\ell^{\sharp}) \supset \text{proj}(r^{\sharp})$ . So the DP processor which transforms  $(\mathcal{P}, \mathcal{R})$  into  $(\mathcal{P} \setminus \mathcal{P}_{\supset}, \mathcal{R})$  is sound and complete for termination, where  $\mathcal{P}_{\supset} = \{\ell^{\sharp} \rightarrow r^{\sharp} \in \mathcal{P} \mid \text{proj}(\ell^{\sharp}) \supset \text{proj}(r^{\sharp})\}$ . The advantage is that in contrast to the reduction pair processor, the criterion of this processor is purely syntactical (i.e., one does not have to search for interpretations or well-founded relations), and one can remove DPs without considering the rewrite rules  $\mathcal{R}$  at all. However, the soundness of this processor for full rewriting relies on the minimality property, which does not hold in the probabilistic setting, see Remark 2. Thus, such a processor would be unsound for AST as shown by the following example.

<sup>12</sup> Projections are often denoted by  $\pi$  in the literature, but we use  $\pi$  for positions.

**Example 21.** The ADP problem with  $f(s(x)) \rightarrow \{1 : F(x)\}^{\text{true}}$  and  $g \rightarrow \{1 : s(g)\}^{\text{true}}$  is not AST as we can rewrite  $F(s(g))$  to itself by first applying the  $f$ -ADP in an (at)-step, followed by an (nt)-step with the  $g$ -ADP. However, consider the simple projection  $proj$  which maps each annotated symbol to its only argument. Then, the projected right-hand side  $proj(F(x)) = x$  of the  $f$ -ADP is a proper subterm of the projected annotated left-hand side  $proj(F(s(x))) = s(x)$ , while the  $g$ -ADP does not contain any annotations, so we can ignore it. Hence, we would falsely assume that we can remove the annotation of the  $f$ -ADP, leading to an ADP problem which is AST as it does not contain any annotations anymore. The problem here is the non-terminating  $g$ -rule that can generate infinitely many  $s$ -symbols.

#### 4.6.2. Subterm criterion processor for $\text{iAST}$

For non-probabilistic innermost rewriting, the soundness of the subterm criterion does not rely on the minimality criterion. Thus, we now show how to adapt the subterm criterion to the probabilistic setting in order to prove  $\text{iAST}$ . For innermost rewriting, [Example 21](#) is no longer a counterexample, because the ADP problem from [Example 21](#) is  $\text{iAST}$ . The reason is that terms without the symbol  $g$  only have finite reductions, and if the start term contains the symbol  $g$ , then we have to rewrite it due to innermost strategy and we cannot rewrite any  $F$ -symbols on positions above  $g$ .

However, there is another problem when adapting the subterm criterion to the probabilistic setting. Since we consider ADPs instead of ordinary DPs, we may have terms with several annotations in the right-hand side. For the reduction pair processor we solved this problem by considering the  $\sharp$ -sum. However, for the subterm criterion it is not clear how to combine several annotated subterms. Therefore, we restrict ourselves to ADP problems where all terms in the right-hand sides contain at most one annotation. This leads to the following processor.

**Theorem 14** (Subterm Criterion for  $\text{iAST}$ ). *Let  $\mathcal{P}$  be an ADP problem where all occurring terms contain at most one annotation. Let  $proj$  be a simple projection such that for all  $\ell \rightarrow \mu^m \in \mathcal{P}$  and all  $r \in \text{Supp}(\mu)$  where there exists a  $t \trianglelefteq_{\sharp} r$ , we have  $proj(\ell^{\sharp}) \trianglerighteq proj(t^{\sharp})$ . Moreover, let  $\mathcal{P}_{\triangleright} \subseteq \mathcal{P}$  be the set of ADPs where for all  $\ell \rightarrow \mu^m \in \mathcal{P}_{\triangleright}$ , there exist an  $r \in \text{Supp}(\mu)$  and a term  $t \trianglelefteq_{\sharp} r$  such that  $proj(\ell^{\sharp}) \triangleright proj(t^{\sharp})$ . Then the processor  $\text{Proc}_{\text{SC}}^{\sharp}(\mathcal{P}) = \{(\mathcal{P} \setminus \mathcal{P}_{\triangleright}) \cup \mathfrak{b}(\mathcal{P}_{\triangleright})\}$  is sound and complete for  $\text{iAST}$ .*

*Proof Sketch. Soundness:* When choosing  $\mathcal{P}_{\triangleright}$  to be  $\mathcal{P}_{\triangleright}$ , then the conditions (1) and (2a) of the reduction pair processor that consider the  $\sharp$ -sums are satisfied with the polynomial interpretation that maps each  $f(x_1, \dots, x_n)$  with  $f \in \Sigma$  to  $x_1 + \dots + x_n + 1$ , and each  $f^{\sharp}(x_1, \dots, x_n)$  with  $f^{\sharp} \in \Sigma^{\sharp}$  to  $x_{proj(f^{\sharp})}$ , i.e., we map each annotated symbol to its projected argument by the simple projection  $proj$ , and all other symbols are used to calculate the number of symbols in a term. Due to the restriction to innermost rewriting and the restriction that ADPs only contain at most a single annotated symbol in each term on the right-hand side, one can show that the conditions (2b) and (3) are irrelevant in this setting, since we cannot rewrite below the projected position before the next (at)- or (af)-step at the annotated position.

*Completeness:* By [Theorem 6](#).  $\square$

The idea of this processor is similar to the reduction pair processor. Since in all ADPs, the “value” (which is the term structure in this case) cannot increase, we can remove those ADPs where there is a chance to decrease the “value”. But in contrast to the reduction pair processor, we do not have to impose any constraints resulting from disregarding the annotations of ADPs, i.e., we do not obtain constraints of the form (2b) or (3) in [\[Theorem\]12](#).

As noted in [\[53\]](#), for non-probabilistic innermost rewriting, the subterm criterion is subsumed by the reduction pair processor when integrating projections (or “argument filterings”) and an appropriate improved notion of usable rules into the reduction pair processor [\[10\]](#). Such an improved reduction pair processor which subsumes the subterm criterion would most likely also be possible in the probabilistic setting, but the reduction pair processor of [Theorem 12](#) does not subsume the subterm criterion, as demonstrated by the following example.

**Example 22.** Consider the ADP problem  $\{\{\#16\}, \mathfrak{b}(\#14), \mathfrak{b}(\#15), \mathfrak{b}(\#17) - \mathfrak{b}(\#20)\}$  that results from the innermost dependency graph processor when proving  $\text{iAST}$  of  $\mathcal{R}_{\text{alg}}$ , see [Example 11](#). Here, we can use the simple projection  $proj$  that maps  $L2$  to its only argument. Then for the only ADP with annotations  $\{\#16\}$ , the projection of the annotated subterm of the right-hand side  $proj(L2(y)) = y$  is a proper subterm of the projected annotated left-hand side  $proj(L2(s(y))) = s(y)$ . Hence, by applying [Theorem 14](#) we result in  $\{\mathfrak{b}(\#14) - \mathfrak{b}(\#20)\}$ , which has no annotations anymore. Thus, we have directly shown that this ADP problem is  $\text{iAST}$ . So in contrast to the reduction pair processor of [Theorem 12](#), here we can ignore the other ADPs  $\mathfrak{b}(\#14), \mathfrak{b}(\#15), \mathfrak{b}(\#17) - \mathfrak{b}(\#20)$  completely even though their flags are still true at this point.

#### 4.7. Switching from full to innermost AST

In the non-probabilistic DP framework, there is a processor to switch from full to innermost rewriting if the DP problem satisfies certain conditions [\[9, Theorem 32\]](#), e.g., if the rules and DPs are non-overlapping. This is useful as the DP framework for innermost termination is more powerful than the one for full termination and in this way, one can switch to the innermost case for certain sub-problems, even if the whole TRS does not belong to any class where innermost termination implies termination [\[49\]](#). As we showed in [\[43\]](#), to ensure that  $\text{iAST}$  implies AST for PTRSs, in addition to non-overlappingness, one has to require linearity of the rules. However, in the non-probabilistic setting, the soundness of this processor relies again on the minimality property which does not hold in the probabilistic setting ([Remark 2](#)). Indeed, the following example which corresponds to [\[53, Example 3.15\]](#) shows that a similar processor in the ADP framework would be unsound, even if we require that all ADPs are non-overlapping and linear. Non-overlappingness of ADPs is defined as for TRSs, as both contain only a single left-hand side.

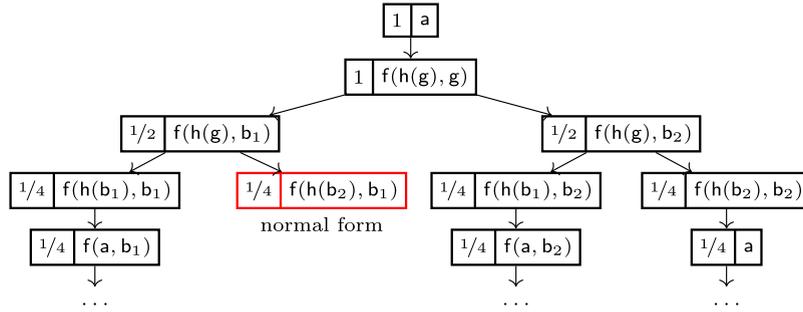


Fig. 7.  $\mathcal{R}_{\text{incpl}}$ -RST.

**Example 23.** The ADP problem containing  $f(x) \rightarrow \{1 : F(a)\}^{\text{true}}$  and  $a \rightarrow \{1 : a\}^{\text{true}}$  is not AST as we can rewrite  $F(a)$  to itself with the  $f$ -ADP. However, it is  $\text{iAST}$  since in innermost evaluations, we have to rewrite the inner  $a$ , which does not terminate but does not use any annotations, i.e., any (at) or (af) steps. Note that the ADPs are non-overlapping and linear. Thus, a processor to switch from full to innermost AST cannot be applied on the level of ADP problems.

Hence, to prove AST of PTRSs that belong to the classes of [43] where  $\text{iAST}$  implies AST, one should apply the ADP framework for  $\text{iAST}$ , because its processors are more powerful. But otherwise, one has to use the ADP framework for AST.

### 5. Transforming ADPs

Compared to our first adaption of DPs to PTRSs via dependency tuples (DTs) in [42], the ADP framework is not only easier, more elegant, and yields a complete chain criterion, but it also has important practical advantages, because every processor that performs a rewrite step benefits from our novel definition of rewriting with ADPs (whereas the rewrite relation with DTs in [42] was an “incomplete over-approximation” of the rewrite relation of the original TRS). To illustrate this, we use the following example.

**Example 24.** Consider the PTRS  $\mathcal{R}_{\text{incpl}}$  with the rules

$$a \rightarrow \{1 : f(h(g), g)\} \tag{24}$$

$$g \rightarrow \{1/2 : b_1, 1/2 : b_2\} \tag{25}$$

$$h(b_1) \rightarrow \{1 : a\} \tag{26}$$

$$f(x, b_2) \rightarrow \{1 : a\} \tag{27}$$

and the  $\mathcal{R}_{\text{incpl}}$ -RST in Fig. 7. The term  $a$  can be rewritten to the normal form  $f(h(b_2), b_1)$  with probability  $1/4$  and to the terms  $f(a, b_1)$  and  $a$  that contain the redex  $a$  with a probability of  $1/4 + 1/4 = 1/2$ . In the term  $f(a, b_2)$ , one can rewrite the subterm  $a$ , and if that ends in a normal form, one can still rewrite the outer  $f$  which will yield  $a$  again. So to over-approximate the probability of non-termination, one could consider the term  $f(a, b_2)$  as if one had two occurrences of  $a$ . Then this would correspond to a random walk where the number of  $a$  symbols is decreased by 1 with probability  $1/4$ , increased by 1 with probability  $1/4$ , and kept the same with probability  $1/2$ . Such a random walk is AST, and since a similar observation holds for all  $\mathcal{R}_{\text{incpl}}$ -RSTs,  $\mathcal{R}_{\text{incpl}}$  is AST (we will prove  $\text{iAST}$  of  $\mathcal{R}_{\text{incpl}}$  with our new transformational processors in this section). In contrast, the DT framework from [42] fails on this example, see [44] for details.

In the non-probabilistic DP framework, there exist several *transformational* processors that increase the power of the framework substantially [10]. To benefit from such transformations in the probabilistic setting as well, in Sections 5.1–5.4, we introduce the *rewriting*, *instantiation*, *forward instantiation*, and *rule overlap instantiation* processors for our ADP framework. The latter corresponds to a weaker version of the *narrowing* processor from [10], which is unsound in the probabilistic setting in general, as we will see in Example 30. These transformational processors had not been adapted in the probabilistic DT framework of [42]. While one could also adapt these processors to the setting of [42], the rewriting processor would be substantially weaker and it would fail in proving  $\text{iAST}$  of  $\mathcal{R}_{\text{incpl}}$ .

#### 5.1. Rewriting processor

We start by considering the *rewriting* processor. Here, we restrict ourselves to innermost rewriting, as a corresponding processor for full rewriting is already unsound in the non-probabilistic setting. For that reason, our novel processor can only be used to prove  $\text{iAST}$  (but not AST) of  $\mathcal{R}_{\text{incpl}}$ .

In the non-probabilistic setting [10], the rewriting processor may rewrite a redex in the right-hand side of a DP if this does not affect the construction of chains. To ensure that, the usable rules for this redex must be non-overlapping (NO). If the DP occurs in an innermost chain, then this redex is weakly innermost normalizing (or “weakly innermost terminating”), hence by NO also terminating and confluent, and thus, it has a unique normal form [49].

In the probabilistic setting, to ensure that the probabilities for the normal forms stay the same, in addition to NO we require that the ADP used for the rewrite step is linear (L) (i.e., every variable occurs at most once in the left-hand side and in each term of the multi-distribution  $\mu$  on the right-hand side) and non-erasing (NE) (i.e., each variable of the left-hand side occurs in each term of  $\text{Supp}(\mu)$ ).

**Definition 13** (Rewriting Processor). Let  $\mathcal{P}$  be an ADP problem with  $\mathcal{P} = \mathcal{P}' \uplus \{\ell \rightarrow \{p_1 : r_1, \dots, p_k : r_k\}^m\}$ . Let  $\tau \in \text{Pos}_D(r_j)$  for some  $1 \leq j \leq k$  such that  $r_j|_\tau \in \mathcal{T}$ , i.e., there is no annotation below or at the position  $\tau$ . If  $r_j \hookrightarrow_{\mathcal{P}, \tau, \text{true}} \{q_1 : e_1, \dots, q_h : e_h\}$ , where  $\hookrightarrow_{\mathcal{P}, \tau, \text{true}}$  denotes a rewrite step with  $\hookrightarrow_{\mathcal{P}}$  at position  $\tau$  and the applied ADP from  $\mathcal{P}$  must have the flag true, then we define

$$\text{Proc}_{\tau}^{\downarrow}(\mathcal{P}) = \left\{ \mathcal{P}' \cup \{\ell \rightarrow \{p_1 : b(r_1), \dots, p_k : b(r_k)\}^m, \right. \\ \left. \ell \rightarrow (\{p_1 : r_1, \dots, p_k : r_k\} \setminus \{p_j : r_j\} \cup \{p_j \cdot q_1 : e_1, \dots, p_j \cdot q_h : e_h\})^m \right\}$$

In the non-probabilistic DP framework, when applying the rewriting processor to a DP problem  $(\mathcal{P}, \mathcal{R})$ , one only transforms the DPs  $\mathcal{P}$  by rewriting, but the rules  $\mathcal{R}$  are left unchanged. But since our ADPs represent both DPs and rules, when rewriting an ADP, we add a copy of the original ADP without any annotations (i.e., this corresponds to the original rule). Another change to the rewriting processor in the non-probabilistic DP framework is the requirement that there exists no annotation below  $\tau$ . Otherwise, rewriting would potentially remove annotations from  $r_j$ . For the soundness of our processor, we have to ensure that this cannot happen. Hence, the choice of the used VRF for this rewrite step is irrelevant.

**Theorem 15** (Soundness of the Rewriting Processor).  $\text{Proc}_{\tau}^{\downarrow}$  as in Definition 13 is sound<sup>13</sup> for  $\lambda$ AST if one of the following cases holds:

1.  $\mathcal{U}_{\mathcal{P}}(r_j|_\tau)$  is NO, and the rule used for rewriting  $r_j|_\tau$  is L and NE.
2.  $\mathcal{U}_{\mathcal{P}}(r_j|_\tau)$  is NO, and all its rules have the form  $\ell' \rightarrow \{1 : r'\}^{\text{true}}$ .
3.  $\mathcal{U}_{\mathcal{P}}(r_j|_\tau)$  is NO,  $r_j \hookrightarrow_{\mathcal{P}, \tau, \text{true}} \{q_1 : e_1, \dots, q_h : e_h\}$  is an innermost step, and  $r_j|_\tau$  is a ground term.

*Proof Sketch.* Let  $\overline{\mathcal{P}}' = \text{Proc}_{\tau}^{\downarrow}(\mathcal{P})$  and  $\overline{\mathcal{P}} = \overline{\mathcal{P}}' \cup \{\ell \rightarrow \{p_1 : r_1, \dots, p_k : r_k\}^m\}$ . We call  $\ell \rightarrow \{p_1 : r_1, \dots, p_k : r_k\}^m$  the old ADP, we call  $\ell \rightarrow (\{p_1 : r_1, \dots, p_k : r_k\} \setminus \{p_j : r_j\} \cup \{p_j \cdot q_1 : e_1, \dots, p_j \cdot q_h : e_h\})^m$  the new ADP, and  $\ell \rightarrow \{p_1 : b(r_1), \dots, p_k : b(r_k)\}^m$  is called the non-annotated old ADP.

For every  $\mathcal{P}$ -CT  $\mathfrak{Z}$  we create a  $\overline{\mathcal{P}}$ -CT such that  $|\mathfrak{Z}'| \leq |\mathfrak{Z}|$ . To this end, we iteratively remove usages of the old ADP using a transformation  $\Phi$  on CTs. The limit  $\mathfrak{Z}^{(\infty)}$  of this iteration is a  $\overline{\mathcal{P}}$ -CT that converges with probability at most  $|\mathfrak{Z}|$ . The transformation  $\Phi$  always removes a “topmost” usage of the old ADP. There are two different cases, depending on whether the old ADP is applied at an annotated position:

1. If a  $\overline{\mathcal{P}}$ -CT  $\mathfrak{Z}$  uses the old ADP in a node  $v$  at a position  $\pi \in \text{Pos}_{D^{\#}}(t_v)$ , then  $\Phi(\mathfrak{Z})$  is a corresponding  $\overline{\mathcal{P}}$ -CT that uses the new ADP in node  $v$ . In the first case of Theorem 15, the non-erasingness and non-overlappingness of  $\mathcal{P}$  guarantee that the rewrite step that we performed to obtain the new ADP would be performed later on in  $\mathfrak{Z}$  anyway. Hence, due to the linearity we can move this rewrite step to an earlier node without affecting the termination probability. The second case of Theorem 15 corresponds to the original rewrite processor where all usable rules of  $r_j|_\tau$  are non-probabilistic. The third case of Theorem 15 is already an innermost rewrite step because of the restriction to ground terms, and moreover, it is the only possible rewrite step at this position.
2. If a  $\overline{\mathcal{P}}$ -CT  $\mathfrak{Z}$  uses the old ADP in a node  $v$  at a position  $\pi \notin \text{Pos}_{D^{\#}}(t_v)$ , then  $\Phi(\mathfrak{Z})$  is a corresponding  $\overline{\mathcal{P}}$ -CT that uses the non-annotated old ADP at the root.  $\square$

In the last case of Theorem 15, for any instantiation only a single innermost rewrite step is possible for  $r_j|_\tau$ . The restriction to innermost rewrite steps is only useful if  $r_j|_\tau$  is ground. Otherwise, an innermost step on  $r_j|_\tau$  might become a non-innermost step when instantiating  $r_j|_\tau$ 's variables.

Next, we show why the rewriting processor needs the new requirement L in the first case that was not imposed in the non-probabilistic setting. More precisely, we give counterexamples for soundness if the used rule is not left linear, i.e., a variable occurs more than once in the left-hand side, and if the used rule is not right linear, i.e., a variable occurs more than once in a term on the right-hand side. The other new requirement NE is currently used in the soundness proof, but we were unable to find a counterexample to soundness if the used rule is not NE. In fact, we conjecture that one can omit this requirement, but then one needs a much more complicated construction and estimation of the resulting termination probability in the soundness proof. The reason is that with only L we can guarantee that performing this rewrite step at a (possibly) non-innermost redex can only increase the probability of innermost termination for the rewritten subterm but not decrease it. Increasing the probability of termination for a proper subterm without any annotations means that we have a higher probability to apply a rewrite step at the position of an annotated symbol above it. Recall that a CT requires that on each infinite path, redexes with annotated root symbol have to be rewritten infinitely often (they correspond to the rewrite steps with Case (at) or (af) of Definition 6). Thus, a higher probability of termination of the proper subterm leads to a lower probability of the leaves in the CT as the probability to rewrite redexes with annotated root is higher. However, proving this requires a much more involved approximation of the probability for termination than our current proof, where we additionally require NE.

<sup>13</sup> For completeness in the non-probabilistic setting [10], one uses a different definition of “non-terminating” (or “infinite”) DP problems. In future work, we will examine if such a definition would also yield completeness of  $\text{Proc}_{\tau}^{\downarrow}$  in the probabilistic case.

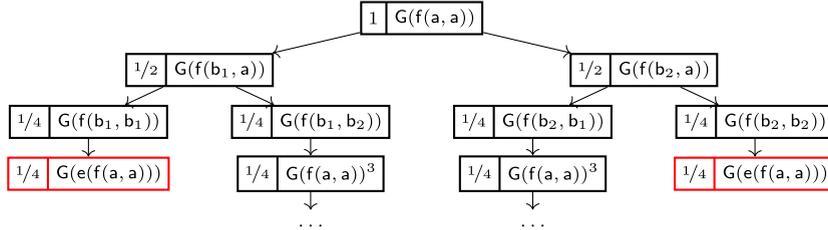
**Example 25** (Left-Linearity is Required for Soundness). To see why left-linearity is required for soundness of the rewriting processor in the probabilistic setting, consider the ADP problem  $\mathcal{P}_{\parallel}$  with

$$g(f(x, y)) \rightarrow \{1 : d(G(f(a, a)), G(f(a, a)), G(f(a, a)))\}^{\text{false}} \quad (28)$$

$$f(x, x) \rightarrow \{1 : e(f(a, a))\}^{\text{true}} \quad (29)$$

$$a \rightarrow \{1/2 : b_1, 1/2 : b_2\}^{\text{true}} \quad (29)$$

This example could also be made non-erasing by adding  $x$  as an additional argument to  $e$  and by instantiating  $x$  and  $y$  by all possible values from  $\{b_1, b_2\}$  in the  $g$ -ADP. The ADP problem  $\mathcal{P}_{\parallel}$  is not  $\text{iAST}$ , as it allows for the following CT whose leaves have a probability  $< 1$ .



Here,  $G(f(a, a))^3$  denotes the term  $d(G(f(a, a)), G(f(a, a)), G(f(a, a)))$ . Chain trees starting in  $G(e(f(a, a)))$  can never use a rewrite step with the  $g$ -ADP anymore and therefore, they converge with probability 1. So we can rewrite a single  $G$ -term to a leaf with a probability of  $1/2$  and to three copies of itself with a probability of  $1/2$ . Hence, the full CT corresponds to a random walk that terminates with probability  $< 1$ . But without the restriction to left-linearity, we could apply the rewriting processor and replace the  $g$ -ADP by  $g(f(x, y)) \rightarrow \{1 : d(G(e(f(a, a))), G(e(f(a, a))), G(e(f(a, a))))\}^{\text{false}}$  and several  $g$ -ADPs without annotations, since the usable rules  $\mathcal{U}_{\mathcal{P}_{\parallel}}(f(a, a)) = \{(28), (29)\}$  are NO. Now in every path of the CT this ADP can be used at most once and hence, the resulting ADP problem is  $\text{iAST}$ , which shows unsoundness of the rewriting processor without left-linearity.

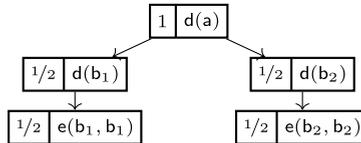
**Example 26** (Right-Linearity is Required for Soundness). To show the need for right-linearity, consider the ADP problem  $\mathcal{P}_{\parallel}$  with

$$f(e(b_1, b_1)) \rightarrow \{1 : h(F(d(a)), F(d(a)), F(d(a)), F(d(a)))\}^{\text{false}} \quad (30)$$

$$d(x) \rightarrow \{1 : e(x, x)\}^{\text{true}} \quad (31)$$

$$a \rightarrow \{1/2 : b_1, 1/2 : b_2\}^{\text{true}} \quad (31)$$

Note that for the term  $d(a)$  we have the following innermost  $\mathcal{P}_{\parallel}$ -CT.



Hence, the term  $F(d(a))$  can rewrite to  $F(e(b_1, b_1))$  and then to four occurrences of  $F(d(a))$  again with probability  $1/2$ , or it reaches a normal form with probability  $1/2$ . This is again a random walk that terminates with probability  $< 1$ , so that the ADP problem  $\mathcal{P}_{\parallel}$  is not  $\text{iAST}$ . But without the restriction to right-linearity of the used rule, it would be possible to apply the rewriting processor and replace the  $f$ -ADP by  $f(e(b_1, b_1)) \rightarrow \{1 : h(F(e(a, a)), F(e(a, a)), F(e(a, a)), F(e(a, a))))\}^{\text{false}}$  and several  $f$ -ADPs without annotations, as  $\mathcal{U}_{\mathcal{P}_{\parallel}}(d(a)) = \{(30), (31)\}$  is NO. The term  $e(a, a)$  can now be rewritten to the term  $e(b_1, b_1)$  with probability  $1/4$ , whereas one obtains a term of the form  $e(b_i, b_j)$  with  $i \neq 1$  or  $j \neq 1$  with probability  $3/4$ . Hence, now  $F(e(a, a))$  can rewrite to a term with four subterms  $F(e(a, a))$  only with a probability of  $1/4$ , and it reaches a leaf with probability  $3/4$ . This is now a random walk that terminates with probability 1 and the same happens for all possible CTs. Hence, the resulting ADP problem is  $\text{iAST}$ , which shows unsoundness of the rewriting processor without right-linearity.

As mentioned, while a rewriting processor could also be defined in the probabilistic DT framework of [42], the rewriting processor benefits from our ADP framework, because it applies the rewrite relation  $\hookrightarrow_{\mathcal{P}}$  on ADPs. In contrast, a rewriting processor in the DT framework of [42] would have to replace a DT by *multiple* new DTs, because the rewrite relation on DTs in [42] corresponds to an over-approximation of the actual rewrite relation on PTRSS. Such a rewriting processor would fail for  $\mathcal{R}_{\text{incompl}}$  whereas with the processor of Theorem 15 we can now prove that  $\mathcal{R}_{\text{incompl}}$  is  $\text{iAST}$ .

**Example 27.** After applying the usable terms and the usable rules processor to  $DP(\mathcal{R}_{\text{incompl}})$ , we obtain:

$$a \rightarrow \{1 : F(H(g), g)\}^{\text{true}} \quad (32)$$

$$g \rightarrow \{1/2 : b_1, 1/2 : b_2\}^{\text{true}} \quad (33)$$

$$h(b_1) \rightarrow \{1 : A\}^{\text{true}} \quad (34)$$

$$f(x, b_2) \rightarrow \{1 : A\}^{\text{false}} \quad (35)$$

Here, the usable terms processor has been used to replace  $F(H(G), G)$  by  $F(H(g), g)$  in the right-hand side of (32). The reason is that  $G$  is not usable as the  $g$ -ADP (33) does not contain any annotations. Moreover, the usable rules processor has been used to set the flag of (35) to false, because neither  $f$  nor  $F$  occur below annotated symbols on right-hand sides.

Next we can apply the rewriting processor on (32) repeatedly until all  $g$ -symbols are rewritten, and replace it by the ADP  $a \rightarrow \{1/4 : F(H(b_1), b_1), 1/4 : F(H(b_2), b_1), 1/4 : F(H(b_1), b_2), 1/4 : F(H(b_2), b_2)\}^{\text{true}}$  as well as several ADPs  $a \rightarrow \dots$  without annotations in the right-hand side. Now in the subterms  $F(\dots, b_1)$  and  $H(b_2)$ , the annotations are removed from the roots by the usable terms processor, as these subterms cannot rewrite to annotated instances of left-hand sides of ADPs. So the  $a$ -ADP is changed to  $a \rightarrow \{1/4 : f(H(b_1), b_1), 1/4 : f(h(b_2), b_1), 1/4 : F(H(b_1), b_2), 1/4 : F(h(b_2), b_2)\}^{\text{true}}$  (32'). This ADP corresponds to the observations that explain why  $\mathcal{R}_{\text{incompl}}$  is  $\text{iAST}$  in Example 24: We have two terms in the right-hand side that correspond to one  $A$  each, both with probability  $1/4$ , one term that corresponds to a normal form with probability  $1/4$ , and one that corresponds to two  $A$ 's with probability  $1/4$ . So again this corresponds to a random walk where the number of  $A$ -symbols is decreased by 1 with probability  $1/4$ , increased by 1 with probability  $1/4$ , and kept the same with probability  $1/2$ .

Indeed, we can now use the reduction pair processor with the polynomial interpretation  $\text{Pol}$  that maps  $A, F$ , and  $H$  to 1 and all other symbols to 0, to remove all annotations from the  $a$ -ADP (32'), because it contains the right-hand side  $f(h(b_2), b_1)$  without annotations and thus,  $\text{Pol}_{\text{sum}}^{\#}(A) = \text{Pol}(A) = 1 > \text{Pol}_{\text{sum}}^{\#}(f(h(b_2), b_1)) = 0$ . Another application of the usable terms processor removes the remaining  $A$ -annotations from (34) and (35). Since there are no more annotations left, this proves  $\text{iAST}$  of  $\mathcal{R}_{\text{incompl}}$ .

## 5.2. Instantiation processor

In the non-probabilistic setting [10], the idea of the *instantiation* processor is to consider all possible *predecessors*  $s^{\#} \rightarrow t^{\#}$  of a dependency pair  $u^{\#} \rightarrow v^{\#}$  in a chain and to compute the skeleton  $\text{Ren}(\text{Cap}_{\mathcal{R}}(t^{\#}))$  of  $t^{\#}$  (see Section 2.1) that remains unchanged when we reduce  $t^{\#}\sigma$  to  $u^{\#}\sigma$  for some instantiation  $\sigma$ , i.e., when going from one DP in a chain to the next. Then  $\text{Ren}(\text{Cap}_{\mathcal{R}}(t^{\#}))$  and  $u^{\#}$  must unify with some most general unifier (*mgu*)  $\delta$ , and  $t^{\#}\sigma \rightarrow_{\mathcal{R}}^* u^{\#}\sigma$  implies that  $\sigma$  is an instance of  $\delta$ . Hence, the instantiation processor replaces the DP  $u^{\#} \rightarrow v^{\#}$  by  $u^{\#}\delta \rightarrow v^{\#}\delta$ .

To adapt the instantiation processor to ADPs and AST, we consider all terms in the multi-distributions on the right-hand sides of all predecessors. In the non-probabilistic DP framework, the instantiation processor only instantiates the DPs, but the rules are left unchanged. Due to the duality of ADPs, similar to the rewriting processor, when instantiating an ADP, we add a copy of the original ADP without any annotations again (i.e., this corresponds to the original non-instantiated rule). In the following,  $\text{vr}(\mathcal{P})$  is a variable-renamed copy of  $\mathcal{P}$  where all variables are replaced by fresh ones, and we use  $\delta(\ell \rightarrow \{p_1 : r_1, \dots, p_k : r_k\}^m) = \ell\delta \rightarrow \{p_1 : r_1\delta, \dots, p_k : r_k\delta\}^m$  for any substitution  $\delta$  and ADP  $\ell \rightarrow \{p_1 : r_1, \dots, p_k : r_k\}^m$ .

**Theorem 16** (Instantiation Processor for AST). *Let  $\mathcal{P}$  be an ADP problem with  $\mathcal{P} = \mathcal{P}' \uplus \{\ell \rightarrow \mu^m\}$ . Then  $\text{Proc}_{\text{ins}}$  is sound and complete for AST, where  $\text{Proc}_{\text{ins}}(\mathcal{P}) = \{\mathcal{P}' \cup N \cup \{\ell \rightarrow b(\mu)^m\}\}$  with*

$$N = \left\{ \delta(\ell \rightarrow \mu^m) \left| \begin{array}{l} \ell' \rightarrow v^{m'} \in \text{vr}(\mathcal{P}), r' \in \text{Supp}(v), \\ t \not\leq_{\#}^{\#} r', \delta = \text{mgu}(\text{Ren}(\text{Cap}_{\mathcal{P}}(t^{\#}), \ell^{\#}), \ell^{\#}) \end{array} \right. \right\}$$

*Proof Sketch.* Let  $\bar{\mathcal{P}} = \mathcal{P}' \cup N \cup \{\ell \rightarrow b(\mu)^m\}$ . We again use the Starting Lemma (Lemma 2) to restrict the form of the CTs that have to be regarded.

*Soundness:* Every  $\mathcal{P}$ -CT gives rise to a  $\bar{\mathcal{P}} \cup \{\ell \rightarrow \mu^m\}$ -CT that converges with the same probability and uses the ADP  $\ell \rightarrow \mu^m$  at most at the root. The core idea is that every rewrite step with  $\ell \rightarrow \mu^m$  at a node below the root can also be done with a rule from  $N$ , or we can use  $\ell \rightarrow b(\mu)^m$  if the annotations do not matter, e.g., we rewrite at a position that is not annotated.

If there is a  $\bar{\mathcal{P}} \cup \{\ell \rightarrow \mu^m\}$ -CT which only uses  $\ell \rightarrow \mu^m$  at the root and converges with a probability  $< 1$ , then this also holds for one of its direct subtrees, which is a  $\bar{\mathcal{P}}$ -CT.

*Completeness:* Every  $\bar{\mathcal{P}}$ -CT gives rise to a  $\mathcal{P}$ -CT that converges with the same probability. The core idea is that every rewrite step with an ADP from  $N$  or the ADP  $\ell \rightarrow b(\mu)^m$  is also possible with the more general ADP  $\ell \rightarrow \mu^m$  that may also contain more annotations.  $\square$

For innermost rewriting, as in the estimation of the dependency graph, we do not need  $\text{Ren}$ , and only consider those *mgu*'s  $\delta$  where the instantiated left-hand sides are in argument normal form.

**Theorem 17** (Instantiation Processor for  $\text{iAST}$ ). *Let  $\mathcal{P}$  be an ADP problem with  $\mathcal{P} = \mathcal{P}' \uplus \{\ell \rightarrow \mu^m\}$ . Then  $\text{Proc}_{\text{ins}}^{\text{i}}$  is sound and complete for  $\text{iAST}$ , where  $\text{Proc}_{\text{ins}}^{\text{i}}(\mathcal{P}) = \{\mathcal{P}' \cup N \cup \{\ell \rightarrow b(\mu)^m\}\}$  with*

$$N = \left\{ \delta(\ell \rightarrow \mu^m) \left| \begin{array}{l} \ell' \rightarrow v^{m'} \in \text{vr}(\mathcal{P}), r' \in \text{Supp}(v), \\ t \not\leq_{\#}^{\#} r', \delta = \text{mgu}(\text{Cap}_{\mathcal{P}}(t^{\#}), \ell^{\#}), \{\ell'\delta, \ell\delta\} \subseteq \text{ANF}_{\mathcal{P}} \end{array} \right. \right\}$$

*Proof Sketch.* Similar to the proof of Theorem 16, just considering innermost rewrite steps.  $\square$

**Example 28.** Consider the PTRS  $\mathcal{R}_{\text{ins}}$  with the rules

$$\begin{aligned} c &\rightarrow \{1 : a\} & c &\rightarrow \{1 : b\} \\ f(x, y, z) &\rightarrow \{1 : g(x, y, z)\} \\ g(a, b, u) &\rightarrow \{1/2 : f(u, u, u), 1/2 : g(a, b, u)\} \end{aligned}$$

$\mathcal{R}_{\text{ins}}$  is not AST, since  $(1 : g(a, b, c))$  starts an RST without leaves. However,  $\mathcal{R}_{\text{ins}}$  is  $\text{iAST}$ . Its canonical ADPs  $DP(\mathcal{R}_{\text{ins}})$  are

$$\begin{aligned} c &\rightarrow \{1 : a\}^{\text{true}} & c &\rightarrow \{1 : b\}^{\text{true}} \\ f(x, y, z) &\rightarrow \{1 : G(x, y, z)\}^{\text{true}} \\ g(a, b, u) &\rightarrow \{1/2 : F(u, u, u), 1/2 : G(a, b, u)\}^{\text{true}} \end{aligned}$$

Using only the processors of Section 4, we cannot prove that  $\mathcal{R}_{\text{ins}}$  is  $\text{iAST}$ . However, we can apply the instantiation processor on the f-ADP.

As the term  $t^\sharp = F(u, u, u)$  in the right-hand side of the g-ADP does not have any proper subterms with (possibly annotated) defined symbols, we have  $Cap_{DP(\mathcal{R}_{\text{ins}})}(t^\sharp) = t^\sharp$ . For the left-hand side  $\ell^\sharp = F(x, y, z)$ ,  $\delta = mgu(t^\sharp, \ell^\sharp)$  instantiates  $x, y,$  and  $z$  by  $u$ . So the instantiation processor replaces the f-ADP by  $f(u, u, u) \rightarrow \{1 : G(u, u, u)\}^{\text{true}}$  (and moreover, we add  $f(x, y, z) \rightarrow \{1 : g(x, y, z)\}^{\text{true}}$ ). We can now remove the annotation in the transformed f-ADP by the dependency graph processor, and afterwards remove the annotation in the g-ADP by applying the reduction pair processor with the polynomial interpretation that maps  $G$  to the constant 1 and every other function symbol to 0. As we removed all annotations,  $DP(\mathcal{R}_{\text{ins}})$  and hence  $\mathcal{R}_{\text{ins}}$  must be  $\text{iAST}$ .

### 5.3. Forward instantiation processor

Next we adapt the *forward instantiation* processor. For non-probabilistic DPs, the idea of the forward instantiation processor [10] is to consider all possible *successors*  $u^\sharp \rightarrow v^\sharp$  of a DP  $s^\sharp \rightarrow t^\sharp$  in a chain, again, in order to find the skeleton that remains unchanged when rewriting  $t^\sharp\sigma$  to  $u^\sharp\sigma$  for some substitution  $\sigma$ . To find this skeleton, one reverses the rules of the TRS and then proceeds as for the instantiation processor. Note that these reversed rules might violate the variable conditions of TRSs, i.e., the right-hand side of a reversed rule may contain variables that do not occur in the left-hand side or the left-hand side may be a variable, i.e., we ignore these restrictions here. We say that an ADP problem  $\mathcal{P}$  is *collapsing* if it contains an ADP  $\ell \rightarrow \mu^m$  such that  $\text{Supp}(\mu) \cap \mathcal{V} \neq \emptyset$ , i.e., one of the right-hand sides of an ADP is a variable. For a non-collapsing ADP problem  $\mathcal{P}$  and a term  $\ell' \in \mathcal{T}$ , let  $Cap_{\mathcal{P}}^{-1}(\ell'^\sharp)$  result from  $\ell'^\sharp$  by replacing all its subterms  $f(\dots)$  by fresh variables where  $f$  occurs on the root position of some right-hand side in  $b(\mathcal{P})$ . For a collapsing ADP problem  $\mathcal{P}$ ,  $Cap_{\mathcal{P}}^{-1}(\ell'^\sharp)$  is simply a fresh variable.

**Theorem 18** (Forward Instantiation Processor for AST). *Let  $\mathcal{P}$  be an ADP problem with  $\mathcal{P} = \mathcal{P}' \uplus \{\ell \rightarrow \mu^m\}$ . Then  $\text{Proc}_{\text{fins}}$  is sound and complete for AST, where  $\text{Proc}_{\text{fins}}(\mathcal{P}) = \{\mathcal{P}' \cup N \cup \{\ell \rightarrow b(\mu)^m\}\}$ . Here,*

$$N = \left\{ \delta(\ell \rightarrow \mu^m) \mid \begin{array}{l} \ell' \rightarrow v^{m'} \in \text{vr}(\mathcal{P}), r \in \text{Supp}(\mu), t \leq_\sharp r, \\ \delta = mgu(t^\sharp, \text{Ren}(Cap_{\mathcal{P}}^{-1}(\ell'^\sharp))) \end{array} \right\}$$

*Proof Sketch.* Similar to the proof of Theorem 16.  $\square$

As in the non-probabilistic setting, the variant of this processor for innermost rewriting still has to use the renaming of all occurrences of variables via *Ren* since reversing the rules does not necessarily preserve the innermost evaluation strategy. However, instead of considering all ADPs from  $\mathcal{P}$  in  $Cap_{\mathcal{P}}^{-1}$  we can restrict ourselves to the usable rules of  $t^\sharp$ .

**Theorem 19** (Forward Instantiation Processor for  $\text{iAST}$ ). *Let  $\mathcal{P}$  be an ADP problem with  $\mathcal{P} = \mathcal{P}' \uplus \{\ell \rightarrow \mu^m\}$ . Then  $\text{Proc}_{\text{fins}}^{\text{i}}$  is sound and complete for  $\text{iAST}$ , where  $\text{Proc}_{\text{fins}}^{\text{i}}(\mathcal{P}) = \{\mathcal{P}' \cup N \cup \{\ell \rightarrow b(\mu)^m\}\}$ . Here,*

$$N = \left\{ \delta(\ell \rightarrow \mu^m) \mid \begin{array}{l} \ell' \rightarrow v^{m'} \in \text{vr}(\mathcal{P}), r \in \text{Supp}(\mu), t \leq_\sharp r, \\ \delta = mgu(t^\sharp, \text{Ren}(Cap_{\mathcal{V}_{\mathcal{P}}(t^\sharp)}^{-1}(\ell'^\sharp))), \{\ell\delta, \ell'\delta\} \subseteq \text{ANF}_{\mathcal{P}} \end{array} \right\}$$

*Proof Sketch.* Similar to the proof of Theorem 18, just considering innermost rewrite steps.  $\square$

**Example 29.** Consider the PTRS  $\mathcal{R}_{\text{fins}}$  with the rules

$$\begin{aligned} f(x) &\rightarrow \{1/2 : g(x), 1/2 : h(x)\} \\ g(a) &\rightarrow \{1 : f(q(a))\} & h(b) &\rightarrow \{1 : f(q(b))\} \\ q(a) &\rightarrow \{1 : a\} & q(b) &\rightarrow \{1 : b\} \end{aligned}$$

When trying to prove  $\text{iAST}$ , by the usable rules and the usable terms processor, we obtain the ADP problem  $\mathcal{P}_{\text{fins}}$ :

$$\begin{aligned} f(x) &\rightarrow \{1/2 : G(x), 1/2 : H(x)\}^{\text{false}} \\ g(a) &\rightarrow \{1 : F(q(a))\}^{\text{false}} & h(b) &\rightarrow \{1 : F(q(b))\}^{\text{false}} \\ q(a) &\rightarrow \{1 : a\}^{\text{true}} & q(b) &\rightarrow \{1 : b\}^{\text{true}} \end{aligned}$$

Here, the instantiation processor is useless, as both the mgu of  $Cap_{\mathcal{P}_{\text{fins}}}(F(q(a))) = F(y)$  and  $F(x)$ , and the mgu of  $Cap_{\mathcal{P}_{\text{fins}}}(F(q(b))) = F(z)$  and  $F(x)$  do not modify  $F(x)$ . However, the forward instantiation processor can replace the original f-ADP with  $f(a) \rightarrow \{1/2 : G(a), 1/2 : H(a)\}^{\text{false}}$  and  $f(b) \rightarrow \{1/2 : G(b), 1/2 : H(b)\}^{\text{false}}$ . The reason is that the only possible successors of the f-ADP are the g- and the h-ADP. For the g-ADP, we have  $\mathcal{V}_{\mathcal{P}_{\text{fins}}}(G(x)) = \emptyset$  and thus,  $mgu(G(x), \text{Ren}(Cap_{\mathcal{V}_{\mathcal{P}_{\text{fins}}}(G(x))}^{-1}(G(a)))) = mgu(G(x), G(a))$  instantiates  $x$  with  $a$ . Similarly, for the h-ADP, the mgu instantiates  $x$  with  $b$ . (Note that here the restriction to the usable rules is crucial in the innermost forward instantiation processor, because  $Cap_{\mathcal{P}_{\text{fins}}}^{-1}$  would replace the symbol  $a$  by a fresh variable, since it occurs on the root position in the right-hand side of an ADP from  $\mathcal{P}_{\text{fins}}$ . Hence,  $mgu(G(x), \text{Ren}(Cap_{\mathcal{P}_{\text{fins}}}^{-1}(G(a)))) = mgu(G(x), G(y))$  would not modify  $x$ .)

We can now remove the annotations of the normal forms  $G(b)$  and  $H(a)$  from the new instantiated  $f$ -ADPs with the usable terms processor, and apply the reduction pair processor with the polynomial interpretation that maps every function symbol to the constant 1, in order to remove all annotations of both new  $f$ -ADPs. Finally, we can remove all remaining annotations and prove  $\text{iAST}$  using the dependency graph processor. Since the PTRS is non-overlapping and linear, this also implies  $\text{AST}$ , see [43]. Note that without the forward instantiation processor, we would have to find a polynomial interpretation that is at least linear.

Example 29 shows that the processors that instantiate a given ADP are not only sometimes needed for a successful termination proof (as in Example 28), but sometimes they can also ease the search for algebras (e.g., polynomial or matrix interpretations) in the reduction pair processor, the most time-consuming part of the ADP framework.

#### 5.4. Rule overlap instantiation processor

The *narrowing* processor [8,10] can only be used in a weaker version in the probabilistic setting, and only for innermost rewriting. Let  $\mathcal{P} = \mathcal{P}' \uplus \{\ell \rightarrow \{p_1 : r_1, \dots, p_k : r_k\}^m\}$  be an ADP problem. For each  $1 \leq j \leq k$  and each  $t \not\leq_{\#} r_j$ , we define its *narrowing substitutions* and its *narrowing results* (as in [50], where narrowing was adapted to dependency tuples for complexity analysis of ordinary term rewriting). If we have to perform rewrite steps on (an instance of)  $t$  in order to enable the next application of an ADP at an annotated position, then the idea of the narrowing processor is to perform the first step of this reduction already on the ADP  $\ell \rightarrow \{p_1 : r_1, \dots, p_k : r_k\}^m$ . So whenever there is a  $t \not\leq_{\#} r_j$  and a non-variable position  $\tau$  in  $t$  such that  $t|_{\tau}$  unifies with the left-hand side  $\ell'$  of some ADP  $\ell' \rightarrow \{p'_1 : r'_1, \dots, p'_k : r'_k\}^m \in \text{vr}(\mathcal{P})$  using an mgu  $\delta$  (such that  $\ell\delta, \ell'\delta \in \text{ANF}_{\mathcal{P}}$  if we consider innermost rewriting), then  $\delta$  is a *narrowing substitution* of  $t$ . While the corresponding *narrowing result* could also be defined for ADPs with non-trivial probabilities, to simplify the presentation let us assume for the moment that the ADP just has the form  $\ell' \rightarrow \{1 : r'\}^{\text{true}}$ . Then the corresponding *narrowing result* is  $s = t|_{\tau}\delta$  if we rewrite at a position of  $t$  that used to be annotated in  $r_j$ , and  $s = t[b(r')]\delta$  otherwise.

If  $\delta_1, \dots, \delta_d$  are all narrowing substitutions of  $t$  with the corresponding narrowing results  $s_1, \dots, s_d$ , then one would like to define a narrowing processor that replaces  $\ell \rightarrow \{p_1 : r_1, \dots, p_k : r_k\}^m$  by  $\ell\delta_e \rightarrow \{p_1 : r_1\delta_e, \dots, p_j : s_e, \dots, p_k : r_k\delta_e\}$ , for all  $1 \leq e \leq d$ . However, the main idea of the narrowing processor, i.e., performing the first rewrite step directly on the ADPs, is unsound for probabilistic ADP problems, as shown by the following example.

**Example 30.** Consider the following ADP problem:

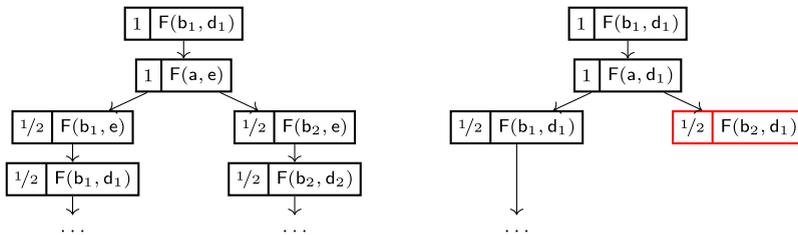
$$\begin{aligned} f(b_1, d_1) &\rightarrow \{1 : F(a, e)\}^{\text{false}} & f(b_2, d_2) &\rightarrow \{1 : F(a, e)\}^{\text{false}} \\ e &\rightarrow \{1 : d_1\}^{\text{true}} & e &\rightarrow \{1 : d_2\}^{\text{true}} \\ a &\rightarrow \{1/2 : b_1, 1/2 : b_2\}^{\text{true}} \end{aligned}$$

This ADP problem is not  $\text{iAST}$ , because it allows for the CT below on the left without any leaves. Note that all occurring terms are ground terms, hence all narrowing substitutions are just the identity function.

But if we apply the narrowing processor to the ADPs in order to rewrite  $e$ , then we replace the two  $f$ -ADPs by the following four new ADPs.

$$\begin{aligned} f(b_1, d_1) &\rightarrow \{1 : F(a, d_1)\}^{\text{false}} & f(b_1, d_1) &\rightarrow \{1 : F(a, d_2)\}^{\text{false}} \\ f(b_2, d_2) &\rightarrow \{1 : F(a, d_1)\}^{\text{false}} & f(b_2, d_2) &\rightarrow \{1 : F(a, d_2)\}^{\text{false}} \end{aligned}$$

This new ADP problem is  $\text{iAST}$ , as we reach a normal form with probability  $1/2$  after each application of an  $f$ -ADP. For example, if we use the first ADP, then we get the CT below on the right. There we reach the normal form  $F(b_2, d_1)$  with probability  $1/2$ .



The difference is that when narrowing the ADP, we have to decide how to rewrite  $e$  before we “split” the term into two different successors with a certain probability, i.e., before we rewrite  $a$  to  $\{1/2 : b_1, 1/2 : b_2\}$ .

Thus, in the probabilistic setting, we can only transform the ADP by applying a rewrite rule if we can ensure the same conditions as for the rewriting processor in Section 5.1. So in the probabilistic setting, the narrowing processor can only instantiate the ADP by the narrowing substitutions, but it must not perform any rewrite step. Instead, the rewrite steps have to be done via the rewriting processor afterwards. Thus, instead of calling it *narrowing processor*, we call it *rule overlap instantiation processor* as it only instantiates the ADPs but does not perform any rewrite steps.

#### 5.4.1. Rule overlap instantiation processor for $\text{iAST}$

Unfortunately, considering only the narrowing substitutions  $\delta_1, \dots, \delta_d$  of  $t$  is not sufficient. There could be another subterm  $t' \trianglelefloor_{\#} r_j$  (with  $t' \neq t$ ) which was involved in a CT (i.e.,  $t^{\#} \sigma \rightarrow_{\text{np}(\mathcal{P})}^* \tilde{\ell} \tilde{\sigma}$  for some substitutions  $\sigma, \tilde{\sigma}$  and a left-hand side  $\tilde{\ell}$  of an ADP, where in the innermost case we require  $\xrightarrow{\text{np}(\mathcal{P})}^*$  instead of  $\rightarrow_{\text{np}(\mathcal{P})}^*$ ), but this CT is no longer possible when instantiating  $t'$  to  $t' \delta_1, \dots, t' \delta_d$ . We say that  $t'$  is *covered*<sup>14</sup> by  $\delta_1, \dots, \delta_d$  if for each narrowing substitution  $\rho$  of  $t'$ , there is a  $\delta_e$  with  $1 \leq e \leq d$  such that  $\delta_e$  is more general than  $\rho$ , i.e.,  $\rho = \delta_e \rho'$  for some substitution  $\rho'$ . So the narrowing processor has to add another ADP  $\ell \rightarrow \{p_1 : \#_{\text{cov}_1(\delta_1, \dots, \delta_d)}(r_1), \dots, p_k : \#_{\text{cov}_k(\delta_1, \dots, \delta_d)}(r_k)\}^m$ , where  $\text{cov}_k(\delta_1, \dots, \delta_d)$  contains all positions of subterms  $t' \trianglelefloor_{\#} r_j$  which are not covered by  $\delta_1, \dots, \delta_d$ . (Therefore, in contrast to instantiation and forward instantiation, here we do not have to add another copy of the original rule without annotations, because the original rule is represented by this additional ADP.) Thus, for innermost rewriting, we obtain the following processor.

**Theorem 20** (Rule Overlap Instantiation Processor). *Let  $\mathcal{P}$  be an ADP problem with  $\mathcal{P} = \mathcal{P}' \uplus \{\ell \rightarrow \{p_1 : r_1, \dots, p_k : r_k\}^m\}$ , let  $1 \leq j \leq k$ , and let  $t \trianglelefloor_{\#} r_j$ . Let  $\delta_1, \dots, \delta_d$  be all narrowing substitutions of  $t$ , where  $d \geq 0$ . Then  $\text{Proc}_{\text{roi}}^{\dagger}(\mathcal{P}) = \{\mathcal{P}' \cup N\}$  is sound and complete for  $\text{iAST}$ , where*

$$N = \left\{ \ell \delta_e \rightarrow \{p_1 : r_1 \delta_e, \dots, p_k : r_k \delta_e\}^m \mid 1 \leq e \leq d \right\} \cup \left\{ \ell \rightarrow \{p_1 : \#_{\text{cov}_1(\delta_1, \dots, \delta_d)}(r_1), \dots, p_k : \#_{\text{cov}_k(\delta_1, \dots, \delta_d)}(r_k)\}^m \right\}$$

*Proof Sketch.* Again, we use the Starting Lemma (Lemma 2) to restrict the form of the CTs that have to be regarded.

*Soundness:* Every  $\mathcal{P}$ -CT  $\mathfrak{A}$  gives rise to a  $(\mathcal{P}' \cup N)$ -CT  $\mathfrak{A}'$  with  $|\mathfrak{A}'| = |\mathfrak{A}|$ . The core idea of this construction is that every rewrite step with  $\ell \rightarrow \{p_1 : r_1, \dots, p_k : r_k\}^m$  can also be done with a rule from  $N$ . If we use  $\ell \rightarrow \{p_1 : \#_{\text{cov}_1(\delta_1, \dots, \delta_d)}(r_1), \dots, p_k : \#_{\text{cov}_k(\delta_1, \dots, \delta_d)}(r_k)\}^m \in N$ , we may create fewer annotations than with the old ADP  $\ell \rightarrow \{p_1 : r_1, \dots, p_k : r_k\}^m$ . However, in the  $\mathcal{P}$ -CT  $\mathfrak{A}$  we never rewrite at the position of those annotations that do not get created in the CT  $\mathfrak{A}'$ , hence we can ignore them.

*Completeness:* Similar to the completeness proof of Theorem 16. We can replace each ADP  $\ell \delta_e \rightarrow \{p_1 : r_1 \delta_e, \dots, p_k : r_k \delta_e\}^m$  with the more general ADP  $\ell \rightarrow \{p_1 : r_1, \dots, p_k : r_k\}^m$ , and each ADP  $\ell \rightarrow \{p_1 : \#_{\text{cov}_1(\delta_1, \dots, \delta_d)}(r_1), \dots, p_k : \#_{\text{cov}_k(\delta_1, \dots, \delta_d)}(r_k)\}^m$  can be replaced by  $\ell \rightarrow \{p_1 : r_1, \dots, p_k : r_k\}^m$  as well, which may lead to more annotations than before.  $\square$

**Example 31.** Consider the PTRS  $\mathcal{R}_{\text{roi}}$ .

$$\begin{aligned} f(d(x)) &\rightarrow \{3/4 : e(f(g(x)), f(h(x))), 1/4 : a\} \\ g(a) &\rightarrow \{1 : d(a)\} \\ h(b) &\rightarrow \{1 : d(b)\} \end{aligned}$$

The usable terms processor transforms the canonical ADP problem into the following ADP problem  $\mathcal{P}_{\text{roi}}$ .

$$\begin{aligned} f(d(x)) &\rightarrow \{3/4 : e(F(g(x)), F(h(x))), 1/4 : a\}^{\text{true}} \\ g(a) &\rightarrow \{1 : d(a)\}^{\text{true}} \\ h(b) &\rightarrow \{1 : d(b)\}^{\text{true}} \end{aligned}$$

The ADP problem  $\mathcal{P}_{\text{roi}}$  (and thus also the original PTRS  $\mathcal{R}_{\text{roi}}$ ) is AST and hence  $\text{iAST}$ , because for every instantiation, at most one of the two “recursive F-calls” in the right-hand side of the  $f$ -ADP can be applied. The reason is that we can either use the  $g$ -rule if the variable  $x$  is instantiated with  $a$ , or we can apply the  $h$ -rule if the variable is instantiated with  $b$ , but not both. We apply  $\text{Proc}_{\text{roi}}^{\dagger}$  using the term  $t^{\#} = F(g(x))$ , where the only narrowing substitution of  $t = f(g(x))$  is  $\delta = \{x/a\}$ . For the other subterm  $F(h(x))$  with annotated root, its flattened version  $f(h(x))$  is not covered by this substitution, because the substitution  $\delta$  is not more general than the narrowing substitution  $\delta' = \{x/b\}$  of  $f(h(x))$ . Hence, we have to generate an additional ADP where this second subterm is annotated. Thus, we replace the former  $f$ -ADP by the following two new ADPs.

$$\begin{aligned} f(d(a)) &\rightarrow \{3/4 : e(F(g(a)), F(h(a))), 1/4 : a\}^{\text{true}} \\ f(d(x)) &\rightarrow \{3/4 : e(f(g(x)), F(h(x))), 1/4 : a\}^{\text{true}} \end{aligned}$$

Now one can remove the annotation of  $F(h(a))$  from the first ADP by the usable terms processor and then apply the reduction pair processor with the polynomial interpretation that maps  $F$  to 1 and all other symbols to 0 to remove all annotations, which proves  $\text{iAST}$ . Again, proving  $\text{iAST}$  with such a simple polynomial interpretation would not be possible without the rule overlap instantiation processor.

#### 5.4.2. Rule overlap instantiation processor for AST

For full rewriting, even a processor like the one of Theorem 20 which only applies the narrowing substitutions but does not perform any rewrite steps is unsound, as shown by the following example.

<sup>14</sup> This was called “captured” in [50].

**Example 32.** Reconsider the ADP problem from [Example 18](#) with the ADPs

$$\begin{aligned} a &\rightarrow \{5/8 : c(a, a), 3/8 : 0\}^{\text{true}} \\ f(c(x_1, x_2)) &\rightarrow \{1 : c(F(x_1), F(x_2))\}^{\text{true}} \end{aligned}$$

As shown in [Example 18](#), this ADP problem is not AST.

The only narrowing substitution for terms of the form  $f(x_i)$  (i.e., for the two corresponding terms in the right-hand side of the f-ADP) is  $\delta$  with  $\delta(x_i) = c(x_1^i, x_2^i)$ , since  $\delta$  is the mgu of  $f(x_i)$  and the (variable-renamed) left-hand side  $f(c(x_1^i, x_2^i))$ . Applying the rule overlap instantiation processor for each  $f(x_i) \triangleq_{\#} c(F(x_1), F(x_2))$  results in the following ADP.

$$f(c(c(x_1^1, x_2^1), c(x_1^2, x_2^2))) \rightarrow \{1 : c(F(c(x_1^1, x_2^1), F(c(x_1^2, x_2^2))))\}^{\text{true}}$$

With the ADP problem of [Example 18](#), we could rewrite  $F(a)$  to  $F(c(a, a))$  with probability  $5/8$ , and afterwards create two copies of  $F(a)$  using the original f-ADP, leading to a random walk biased towards non-termination.

With the new ADP problem, we can instead rewrite  $F(c(a, a))$  to two copies of itself. However, now this is not done with probability  $5/8$ , but only with probability  $(5/8)^2 = 25/64$ . The reason is that we first have to rewrite each of the two different  $a$  subterms to  $c(a, a)$ , each with probability  $5/8$ . Thus, we can only rewrite  $F(c(a, a))$  to the term  $F(c(c(a, a), c(a, a)))$  with probability  $25/64$ . Now we can indeed create two copies of  $F(c(a, a))$  using the new f-ADP. However, this is now a random walk biased towards termination, since  $25/64 < 1/2$ .

Indeed, the new ADP problem created by the rule overlap instantiation processor is AST, while the original ADP problem was not AST, i.e.,  $\text{Proc}_{\text{ro1}}$  is unsound for AST.

To see why  $\text{Proc}_{\text{ro1}}$  is unsound for AST, note that after applying an ADP, the next rewrite step can be “completely inside” the used substitution if one does not impose an innermost rewriting strategy. So for the “non-AST” reduction in [Example 32](#), after using the (original) f-ADP to rewrite  $F(c(a, a))$  to  $c(F(a), F(a))$ , the next “inner” rewrite step rewrites the  $a$ -subterms which did not occur in the right-hand side of the f-ADP but which were “completely inside” the substitution used for  $x_1$  and  $x_2$ . However, if we only consider the right-hand side of the f-ADP, as it does not contain any  $a$ -symbol, we only note that after one application of the f-ADP we eventually have to use the f-ADP again, leading to the narrowing substitution  $\delta$  in [Example 32](#). But when instantiating the f-ADP with this substitution, in order to simulate the previous reduction with the new f-ADP, we now need two instead of one “inner” rewrite step for the  $a$ -subterms. In the probabilistic setting, this is a problem if the critical rewrite step (e.g., from  $a$  to  $c(a, a)$ ) is only chosen with a probability  $< 1$ , because the requirement of choosing the critical step two times has a lower probability than only choosing it once.

## 6. Conclusion and evaluation

In this work, we developed the ADP framework to prove AST and  $\text{iAST}$  of PTRSs completely automatically. By using annotated DPs instead of dependency tuples as in [\[42\]](#), we obtain *complete* criteria for both  $\text{iAST}$  and AST, which also simplifies the framework substantially. As explained at the end of [Section 1](#), this paper is based on our conference papers [\[44,45\]](#) and it extends them by unifying the two separate frameworks for AST and  $\text{iAST}$  and by several new ADP processors, explanations, and remarks.

Compared to the non-probabilistic DP framework for termination of TRSs [\[8–12\]](#), analyzing  $\text{(i)AST}$  automatically is significantly more difficult due to the lack of a “minimality property” in the probabilistic setting, which would allow several further processors. Moreover, the ADP framework for PTRSs is restricted to reduction pairs based on barycentric,  $\mathbb{N}$ -collapsible algebras (e.g., to multilinear polynomial interpretations). The soundness proofs for our processors are much more involved than in the non-probabilistic setting, due to the more complex structure of chain trees. However, the processors themselves are analogous to their non-probabilistic counterparts, and thus, existing implementations of the processors can easily be adapted to their new probabilistic versions.

Our novel ADP framework for full AST is particularly useful when analyzing overlapping PTRSs, as for such PTRSs we cannot use the criteria of [\[43\]](#) for classes of PTRSs where  $\text{iAST}$  implies AST. The frameworks for AST and  $\text{iAST}$  differ in power, since some processors are less powerful for AST, and some are not even applicable for AST. [Table 1](#) illustrates the differences between the ADP frameworks for AST and  $\text{iAST}$ . The parts highlighted in italics show the differences to the non-probabilistic DP framework. Here, “(–)S” and “(–)C” stand for “(not) sound” and “(not) complete”.

We implemented our new contributions in our termination prover AProVE [\[1\]](#) and compared the new version of AProVE based on the probabilistic ADP framework including all new processors and improvements presented in this paper (journal) with a previous version of AProVE which only contains the processors as presented in our corresponding conference papers [\[44,45\]](#) (conf). Moreover, we also compare with a version of AProVE implementing our earlier framework from [\[42\]](#) which used dependency tuples instead of ADPs (dt). As explained at the end of [Section 1](#), compared to the conference papers of conf, in journal we added the subterm criterion processor, the instantiation, forward instantiation, and rule overlap instantiation processors, and we extended the probabilistic reduction pair processors to algebras, which allows the use of, e.g., matrix orders. Similarly, matrix orders can now also be used in our corresponding criterion of [\[42\]](#) for a direct application of orders to prove AST (outside of the ADP framework). This direct application of orders was only used with polynomial interpretations in dt and conf, but in journal, we also use it with matrix orders. Furthermore, note that dt is restricted to the analysis of innermost rewriting, and cannot handle arbitrary rewrite strategies.

The main goal for probabilistic termination analysis is to become as powerful as termination analysis in the non-probabilistic setting. Therefore, in our first experiment, we considered the non-probabilistic TRSs of the *TPDB* [\[54\]](#) (the benchmark set used in the annual *Termination and Complexity Competition (TermComp)* [\[55\]](#)) and compared journal, conf, and dt with AProVE’s techniques for ordinary non-probabilistic TRSs (AProVE-NP), because at the current *TermComp*, AProVE-NP was the most powerful tool for termination of ordinary non-probabilistic TRSs. AProVE-NP includes many additional DP processors and benefits from using

**Table 1**  
Differences between the ADP framework for AST and  $\mathfrak{i}$ AST.

Processor	ADPs for AST	ADPs for $\mathfrak{i}$ AST
Chain Crit.	S & C for non-duplicating	S & C
Dep. Graph	S & C	S & C
Usable Terms	S & C	S & C
Usable Rules	$\neg$ S even with $C_r$ -Rules see <a href="#">Example 18</a>	S & C
Reduction Pairs	S & C barycentric, $\mathbb{N}$ -collapsible algebras	S & C barycentric, $\mathbb{N}$ -collapsible algebras
Probability Removal	S & C	S & C
Subterm Criterion	$\neg$ S see <a href="#">Example 21</a>	S & C at most 1 annotation
From Full to Innermost Rewriting	$\neg$ S see <a href="#">Example 23</a>	—
Rewriting Processor	$\neg$ S	S & $\neg$ C see <a href="#">Footnote 13</a>
Instantiation Processor	S & C	S & C
Forward Inst. Processor	S & C	S & C
Narrowing Processor	$\neg$ S see <a href="#">Example 30</a>	$\neg$ S see <a href="#">Example 30</a>
Rule Overlap Inst. Processor	$\neg$ S see <a href="#">Example 32</a>	S & C

**Table 2**  
Evaluation results.

TRS-Innermost	#-Examples	AProVE-NP	journal	conf	dt
	366	293	170	148	134
TRS-Standard	#-Examples	AProVE-NP	journal	conf	dt
(Innermost)	1512	1114	791	667	612
(Full)	1512	1031	241	201	—
PTRS-Standard	#-Examples	AProVE-NP	journal	conf	dt
(Innermost)	128	—	107	88	77
(Full)	128	—	58	56	—

separate dependency pairs for each occurrence of a defined symbol in the right-hand side of a rule instead of ADPs. Clearly, a TRS can be represented as a PTRS with trivial probabilities, and then (innermost) AST is the same as (innermost) termination. While all three variants journal, conf, and dt have a probability removal processor to switch to the classical DP framework for such problems, we disabled that processor in this experiment. As in *TermComp*, we used a timeout of 300 seconds for each example. The “TRS Innermost” and “TRS Standard” categories contain 366 and 1512 benchmarks and the results regarding innermost rewriting are shown in the first two rows of [Table 2](#). One can see that the transformations are very important for automatic termination proofs as we get around 5% closer to AProVE-NP’s results when using only the rewriting processor (in conf), and 12–16% closer when also using the remaining transformational processors that use instantiations (in journal).

Furthermore, we also tried to prove termination w.r.t. an arbitrary evaluation strategy for the examples from the “TRS Standard” category. Here, AProVE-NP can prove termination of 1031 examples, while journal and conf (using the ADP framework for full rewriting) can only show termination of 241 and 201 examples, respectively, see Row 3 of [Table 2](#). Recall that dt cannot handle arbitrary evaluation strategies. Again, our new transformations clearly increase power, while it becomes clear that proving AST is significantly harder than proving  $\mathfrak{i}$ AST.

In the next experiment, we used the PTRS benchmark set of the *TPDB*, containing 128 typical probabilistic programs, including examples with complicated probabilistic structure and probabilistic algorithms on lists and trees. For instance, it contains the following PTRS  $\mathcal{R}_{\text{qsr}}$  for probabilistic quicksort. Here, lists are represented via the constructors nil (for the empty list) and cons, where, e.g., cons(s(0), cons(s(0), cons(0, nil))) represents the list [1, 1, 0]. To ease readability, we write  $r$  instead of  $\{1 : r\}$ .

```

rotate(cons(x, xs)) → {1/2 : cons(x, xs), 1/2 : rotate(app(xs, cons(x, nil)))}
qsr(xs) → if(empty(xs), low(hd(xs), tl(xs)), hd(xs), high(hd(xs), tl(xs)))
empty(nil) → true
empty(cons(x, xs)) → false
hd(cons(x, xs)) → x
tl(cons(x, xs)) → xs
if(true, xs, x, ys) → nil
if(false, xs, x, ys) → app(qsr(rotate(xs)), cons(x, qsr(rotate(ys))))

```

The rotate-rules rotate a list randomly often (they are AST, but not terminating). Thus, by choosing the first element of the resulting list, one obtains random pivot elements for the recursive calls of qsr in the second if-rule. In addition to the rules above,  $\mathcal{R}_{\text{qsr}}$  contains

rules for list concatenation (`app`), and rules such that `low(x, xs)` (`high(x, xs)`) returns all elements of the list `xs` that are smaller (greater or equal) than `x`, see [Appendix A.2](#). Proving  $\text{iAST}$  of the above rules requires transformational processors to instantiate and rewrite the empty-, `hd`-, and `tl`-subterms in the right-hand side of the `qsrt`-rule. Thus, `dt` and `conf` fail for this example, but `journal` can prove  $\text{iAST}$  of  $\mathcal{R}_{\text{qsrt}}$ . (Our ADP framework for AST is not applicable here, since the PTRS is duplicating.)

The AProVE variant `dt` can prove  $\text{iAST}$  for 77 of the 128 PTRSs from the corresponding collection of the *TPDB*, and `conf` can prove it for 88. Adding the new and improved processors in `journal` increases this number to 107, which demonstrates their power for PTRSs with non-trivial probabilities. When regarding full rewriting, `conf` can prove AST for 56 examples and `journal` succeeds on 58 PTRSs, see the last row of [Table 2](#). For details on our experiments and for instructions on how to run our implementation in AProVE via its *web interface* or locally, see: <https://aprove-developers.github.io/ProbabilisticADPs/>

On this website, we also performed experiments where we disabled individual transformational processors of the ADP framework, which shows the usefulness of each processor.

In future work, we will try to adapt the ADP framework in order to analyze stronger properties (like PAST) and we want to develop techniques (and processors for the ADP framework) in order to *disprove* AST.

## CRediT authorship contribution statement

**Jan-Christoph Kassing:** Writing – original draft; **Jürgen Giesl:** Writing – original draft.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgement

Funded by the [Deutsche Forschungsgemeinschaft](#) (DFG) Research Training Group 2236 UnRAVeL.

## Appendix A. Examples

In this appendix, we present several examples to illustrate specific strengths and weaknesses of the ADP framework and the new transformational processors. They show that in contrast to most other techniques for analyzing AST, due to probabilistic term rewriting, our approach is also suitable for the analysis of algorithms on algebraic data structures other than numbers.

### A.1. Creating and using lists

Similar to [Algorithm 1](#), the following algorithm first creates a random list, filled with random numbers, and afterwards uses the list for further computations. In general, our ADP framework is well suited to analyze algorithms that access or modify randomly generated lists.

The algorithm below computes the sum of all numbers in the generated list. Here, the function `createL(xs)` adds a prefix of arbitrary length filled with natural numbers according to a geometrical distribution in front of the list `xs`. Moreover, `app(xs, ys)` concatenates the two lists `xs` and `ys`. Finally, for a non-empty list `xs` of numbers, `sum(xs)` computes a singleton list whose only element is the sum of all numbers in `xs`. So `sum(cons(s(0), cons(s(0), cons(0, nil))))` evaluates to `cons(s(s(0)), nil)`.

$$\begin{aligned} \text{init} &\rightarrow \{1 : \text{sum}(\text{createL}(\text{nil}))\} \\ \text{addNum}(x, xs) &\rightarrow \{1/2 : \text{cons}(x, xs), 1/2 : \text{addNum}(s(x), xs)\} \\ \text{createL}(xs) &\rightarrow \{1/2 : \text{addNum}(0, xs), 1/2 : \text{createL}(\text{addNum}(0, xs))\} \\ \text{plus}(0, y) &\rightarrow \{1 : y\} \\ \text{plus}(s(x), y) &\rightarrow \{1 : s(\text{plus}(x, y))\} \\ \text{sum}(\text{cons}(x, \text{nil})) &\rightarrow \{1 : \text{cons}(x, \text{nil})\} \\ \text{sum}(\text{cons}(x, \text{cons}(y, ys))) &\rightarrow \{1 : \text{sum}(\text{cons}(\text{plus}(x, y), ys))\} \\ \text{sum}(\text{app}(xs, \text{cons}(x, \text{cons}(y, ys)))) &\rightarrow \{1 : \text{sum}(\text{app}(xs, \text{sum}(\text{cons}(x, \text{cons}(y, ys)))))\} \\ \text{app}(\text{cons}(x, xs), ys) &\rightarrow \{1 : \text{cons}(x, \text{app}(xs, ys))\} \\ \text{app}(\text{nil}, ys) &\rightarrow \{1 : ys\} \\ \text{app}(xs, \text{nil}) &\rightarrow \{1 : xs\} \end{aligned}$$

Note that the left-hand sides of the two rules `app(nil, ys) → {1 : ys}` and `app(xs, nil) → {1 : xs}` overlap and moreover, the last `sum`-rule overlaps with the first `app`-rule. Hence, this PTRS does not belong to any known class where  $\text{iAST}$  implies full AST [\[43\]](#).

Furthermore, there exists no polynomial order that proves AST for this example directly (i.e., without the use of DPs), because the left-hand side of the last sum-rule is embedded in its right-hand side. With our ADP framework for full AST, AProVE can prove AST (hence also  $\text{iAST}$ ) of this example automatically.

### A.2. Probabilistic quicksort

In [Section 6](#) we already presented the main rules of the PTRS  $\mathcal{R}_{\text{qsrt}}$  implementing a probabilistic quicksort algorithm, where  $\text{iAST}$  can only be proved if we use the new transformational processors. The following PTRS is the full implementation of this probabilistic quicksort algorithm.

$$\begin{aligned}
 & \text{rotate}(\text{nil}) \rightarrow \{1 : \text{nil}\} \\
 & \text{rotate}(\text{cons}(x, xs)) \rightarrow \{1/2 : \text{cons}(x, xs), 1/2 : \text{rotate}(\text{app}(xs, \text{cons}(x, \text{nil})))\} \\
 & \text{empty}(\text{nil}) \rightarrow \{1 : \text{true}\} \\
 & \text{empty}(\text{cons}(x, xs)) \rightarrow \{1 : \text{false}\} \\
 & \text{qsrt}(xs) \rightarrow \{1 : \text{if}(\text{empty}(xs), \text{low}(\text{hd}(xs), \text{tl}(xs)), \text{hd}(xs), \text{high}(\text{hd}(xs), \text{tl}(xs)))\} \\
 & \text{if}(\text{true}, xs, x, ys) \rightarrow \{1 : \text{nil}\} \\
 & \text{if}(\text{false}, xs, x, ys) \rightarrow \{1 : \text{app}(\text{qsrt}(\text{rotate}(xs)), \text{cons}(x, \text{qsrt}(\text{rotate}(ys))))\} \\
 & \text{hd}(\text{cons}(x, xs)) \rightarrow \{1 : x\} \\
 & \text{tl}(\text{cons}(x, xs)) \rightarrow \{1 : xs\} \\
 & \text{low}(x, \text{nil}) \rightarrow \{1 : \text{nil}\} \\
 & \text{low}(x, \text{cons}(y, ys)) \rightarrow \{1 : \text{ifLow}(\text{leq}(x, y), x, \text{cons}(y, ys))\} \\
 & \text{ifLow}(\text{true}, x, \text{cons}(y, ys)) \rightarrow \{1 : \text{low}(x, ys)\} \\
 & \text{ifLow}(\text{false}, x, \text{cons}(y, ys)) \rightarrow \{1 : \text{cons}(y, \text{low}(x, ys))\} \\
 & \text{high}(x, \text{nil}) \rightarrow \{1 : \text{nil}\} \\
 & \text{high}(x, \text{cons}(y, ys)) \rightarrow \{1 : \text{ifHigh}(\text{leq}(x, y), x, \text{cons}(y, ys))\} \\
 & \text{ifHigh}(\text{true}, x, \text{cons}(y, ys)) \rightarrow \{1 : \text{cons}(y, \text{high}(x, ys))\} \\
 & \text{ifHigh}(\text{false}, x, \text{cons}(y, ys)) \rightarrow \{1 : \text{high}(x, ys)\} \\
 & \text{leq}(0, x) \rightarrow \{1 : \text{true}\} \\
 & \text{leq}(s(x), 0) \rightarrow \{1 : \text{false}\} \\
 & \text{leq}(s(x), s(y)) \rightarrow \{1 : \text{leq}(x, y)\} \\
 & \text{app}(\text{nil}, ys) \rightarrow \{1 : ys\} \\
 & \text{app}(\text{cons}(x, xs), ys) \rightarrow \{1 : \text{cons}(x, \text{app}(xs, ys))\}
 \end{aligned}$$

As already briefly explained in [Section 6](#), this quicksort algorithm searches for a random pivot element using the first two rotate rules. Here, we rotate the list and always move the head element to the end of the list. With probability  $1/2$  we stop this iteration and use the current head element as the next pivot element. The remaining rules implement the classical quicksort algorithm without any probabilities. As mentioned in [Section 6](#), app computes list concatenation, low( $x, xs$ ) returns all elements of the list  $xs$  that are smaller than  $x$ , and high works analogously. Furthermore, hd returns the head element of a list and tl returns the rest of the list without the head. Finally, empty checks whether the list is empty.

Using our ADP framework, AProVE can automatically prove that this PTRS is  $\text{iAST}$ . (As mentioned in [Section 6](#), our ADP framework for AST is not applicable here, since the PTRS is duplicating.) This example illustrates the need for the new transformational processors: The rewriting processor ([Theorem 15](#)) is required to evaluate the functions hd, tl, and empty, and the rule overlap instantiation processor ([Theorem 20](#)) is required to determine all possible terms that these functions can actually be applied on. So for example, we need to detect that if we have a term of the form empty( $x$ ) for a variable  $x$ , then in order to apply any rewrite step, the variable  $x$  must be instantiated with either nil or a term of the form cons( $y, ys$ ) for some new variables  $y$  and  $ys$ .

### A.3. Moving elements in lists probabilistically

Another interesting probabilistic algorithm that deals with lists is the following: We are given two lists  $L_1$  and  $L_2$ . If one of the two lists is empty, then the algorithm terminates. Otherwise, we either move the head of list  $L_1$  to  $L_2$  or vice versa, both with probability

$1/2$ , and then we repeat this procedure. This algorithm is represented by the following PTRS.

```

or(false, false) → {1 : false}
or(true, x) → {1 : true}
or(x, true) → {1 : true}
moveElements(xs, ys) → {1 : if(or(empty(xs), empty(ys)), xs, ys)}
if(true, xs, ys) → {1 : xs}
if(false, xs, ys) → {1/2 : moveElements(tl(xs), cons(hd(xs), ys)), 1/2 : moveElements(cons(hd(ys), xs), tl(ys))}
empty(nil) → {1 : true}
empty(cons(x, xs)) → {1 : false}
hd(cons(x, xs)) → {1 : x}
tl(cons(x, xs)) → {1 : xs}

```

This algorithm is  $i$ AST (and even AST) because we can view this as a classical random walk on the number of elements in the first list that is both bounded from below by 0 and from above by the sum of the lengths of both lists. In order to prove this automatically, we again have to use some kind of instantiation processor, e.g., the rule overlap instantiation processor, to find all possible terms that the functions `hd`, `tl`, and `empty` can actually be applied on. In addition, we also need the rewriting processor again to evaluate these functions afterwards. Once the `moveElements`-ADP contains the symbols `cons` and `nil` in the left-hand side, we can detect the structure of the random walk using an application of the reduction pair processor that removes all annotations of this ADP. After that, there is no SCC in the dependency graph left, and we have proven  $i$ AST. Again, our ADP framework for AST is not applicable, since the PTRS is duplicating.

#### A.4. Trees

This example illustrates the use of probabilistic term rewriting for algorithms on trees. In the following algorithm (adapted from [56]), we consider binary trees represented via `leaf` and `tree(x, y)`, where `concat(x, y)` replaces the rightmost leaf of the tree  $x$  by  $y$ . The algorithm first creates two random trees and then checks whether the first tree has less leaves than the second one.

```

init → {1 : lessleaves(createT(leaf), createT(leaf))}
concat(leaf, y) → {1 : y}
concat(tree(u, v), y) → {1 : tree(u, concat(v, y))}
lessleaves(x, leaf) → {1 : false}
lessleaves(leaf, tree(x, y)) → {1 : true}
lessleaves(tree(u, v), tree(x, y)) → {1 : lessleaves(concat(u, v), concat(x, y))}
createT(xs) → {1 : xs}
createT(xs) → {1/3 : xs, 1/3 : createT(tree(xs, leaf)), 1/3 : createT(tree(leaf, xs))}

```

Note that the last two rules are overlapping and thus, there is no known criterion [43] which would allow us to conclude AST from a proof of  $i$ AST. In contrast, our ADP framework is able to prove full AST (hence  $i$ AST) for this example, while the direct application of polynomial interpretations fails.

#### A.5. Conditions on numbers

Another important task of termination analysis is to handle conditions on numbers. Such conditions occur in almost every program and often have an impact on the termination behavior. This is also true for probabilistic programs. For a successful proof of  $i$ AST without transformations, the rules of the PTRS have to express these conditions via suitable “patterns” in their left-hand sides, i.e., if one wants to check whether a number is zero or not, then one needs two different rules  $f(0) \rightarrow \dots$  and  $f(s(x)) \rightarrow \dots$  to perform this case analysis. However, in programs, conditions are usually implemented via an `if`-construct, where one could use, e.g., an additional function `gt` to check whether a number is greater than another. The same is true for conditions on other data structures than numbers, as we have seen in the previous examples. If one wants to check whether a list is empty, then without transformations, one needs two rules (for `nil` and `cons`), whereas with transformations, one can use conditions and auxiliary functions like `empty`.

The following PTRS implements the classical random walk, but we check the condition  $x > 0$  not directly in the left-hand side of the rule but with an additional function `gt(x, y)` which determines whether  $x > y$  holds. In addition, to decrease the value of a number

by one, we use the predecessor function  $p(x) = x - 1$ .

$$\begin{aligned} \text{gt}(0, 0) &\rightarrow \{1 : \text{false}\} \\ \text{gt}(s(x), 0) &\rightarrow \{1 : \text{true}\} \\ \text{gt}(0, s(y)) &\rightarrow \{1 : \text{false}\} \\ \text{gt}(s(x), s(y)) &\rightarrow \{1 : \text{gt}(x, y)\} \\ p(0) &\rightarrow \{1 : 0\} \\ p(s(x)) &\rightarrow \{1 : x\} \\ \text{loop}(x) &\rightarrow \{1 : \text{if}(\text{gt}(x, 0), x)\} \\ \text{if}(\text{false}, x) &\rightarrow \{1 : \text{stop}\} \\ \text{if}(\text{true}, x) &\rightarrow \{1/2 : \text{loop}(p(x)), 1/2 : \text{loop}(s(x))\} \end{aligned}$$

In this case, we need the rewriting processor to evaluate the functions  $\text{gt}$  and  $p$  and again we need the rule overlap instantiation processor to check for all possible terms that these functions can actually be applied on. Thus, the ADP framework proves  $\text{iAST}$  for this example. Since the PTRS is duplicating, the ADP framework for AST cannot be applied.

#### A.6. Limits of the instantiation processors

Whenever we have an ADP where the left-hand side of the rule is also contained in the support of the right-hand side, then instantiations become useless, because we will always have at least the same ADP again after applying the processor. For example, we need to apply one of the instantiation processors in order to prove innermost termination for the TRS with the rules  $c \rightarrow a$ ,  $c \rightarrow b$ ,  $f(x, y, z) \rightarrow g(x, y, z)$ , and  $g(a, b, z) \rightarrow f(z, z, z)$ . If we make the  $f$ - and  $g$ -rules probabilistic by adding the possibility to do nothing in a rewrite step with the probability  $1/2$ , then we result in the following PTRS.

$$\begin{aligned} c &\rightarrow \{1 : a\} \quad c \rightarrow \{1 : b\} \\ f(x, y, z) &\rightarrow \{1/2 : g(x, y, z), 1/2 : f(x, y, z)\} \\ g(a, b, z) &\rightarrow \{1/2 : f(z, z, z), 1/2 : g(a, b, z)\} \end{aligned}$$

This PTRS is  $\text{iAST}$  (but not AST, since the original TRS was also just innermost terminating but not terminating), but we are unable to show this using the instantiation processor, because if one tries to instantiate any of the rules, this will result in at least the same rule after the processor. In contrast, in [Example 28](#) we had almost the same PTRS but the  $f$ -rule remained non-probabilistic. There, we were able to apply the instantiation processor and prove  $\text{iAST}$  using the ADP framework. (However, while instantiation does not help in our example above, we can prove  $\text{iAST}$  using the rule overlap instantiation processor.)

#### A.7. Transformations do not suffice for inductive reasoning

Transformational processors are useful to perform a case analysis, but they do not suffice for PTRSs where one needs inductive reasoning for the termination analysis. For example, we cannot show AST (not even  $\text{iAST}$ ) of the following PTRS [\[57\]](#) even though the if-structure seems similar to the one of the probabilistic quicksort example.

$$\begin{aligned} \text{even}(0) &\rightarrow \{1 : \text{true}\} & \text{even}(s(0)) &\rightarrow \{1 : \text{false}\} \\ \text{even}(s(s(x))) &\rightarrow \{1 : \text{even}(x)\} & \text{loop}(x) &\rightarrow \{1 : \text{if}(\text{even}(x), x)\} \\ \text{if}(\text{false}, x) &\rightarrow \{1 : \text{stop}\} & \text{if}(\text{true}, x) &\rightarrow \{1/2 : \text{loop}(x), 1/2 : \text{loop}(s(x))\} \end{aligned}$$

The idea here is that the recursion of  $\text{loop}$  stops if its argument is an even number. If it is not even, then we either increase the value by 1 or use the same value again. Here, a simple case analysis does not suffice, but we actually have to show (inductively), that if a number  $x$  is odd, then  $x + 1$  is even. This is not possible with the new transformational processors but needs other types of processors for inductive reasoning (e.g., as in [\[58\]](#) for the non-probabilistic DP framework).

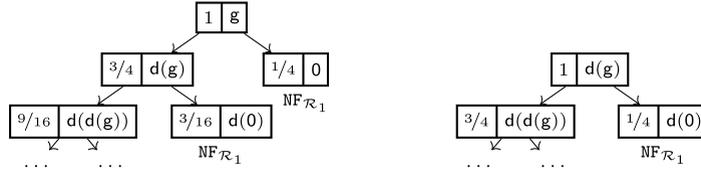
## Appendix B. Proofs

In this section, we prove the chain criterion and the soundness and completeness of our processors of the ADP framework. To this end, we first give some additional definitions and notations, and then prove the lemmas about ADP problems mentioned in [Section 4](#) that we will need throughout the rest of this appendix. We start by defining the notion of a *subtree*.

**Definition 14** (Subtree). Let  $\mathcal{P}$  be an ADP problem and let  $\mathfrak{T} = (V, E, L)$  be a tree that satisfies Conditions (1)-(3) of a  $\mathcal{P}$ -CT. Let  $W \subseteq V$  be non-empty, weakly connected, and for all  $v \in W$  we have  $vE \cap W = \emptyset$  or  $vE \cap W = vE$ . Then, we define the *subtree*  $\mathfrak{T}[W]$  by  $\mathfrak{T}[W] = (W, E \cap (W \times W), L^W)$ . Let  $w \in W$  be the root of  $G^{\mathfrak{T}[W]}$ . To ensure that the root of our subtree has the probability 1 again, we use the labeling  $L^W(v) = (\frac{p_v^{\mathfrak{T}}}{p_v^W} : t_v^{\mathfrak{T}})$  for all nodes  $v \in W$ . If  $W$  contains the root of  $(V, E)$ , then we call the subtree *grounded*.

The property of being non-empty and weakly connected ensures that the resulting graph  $G^{\mathfrak{X}[W]}$  is a tree again. The property that we have  $vE \cap W = \emptyset$  or  $vE \cap W = vE$  ensures that the sum of probabilities for the successors of a node  $v$  is equal to the probability for the node  $v$  itself.

**Example 33.** Reconsider the PTRS  $\mathcal{R}_1$  with the rule  $g \rightarrow \{3/4 : d(g), 1/4 : 0\}$ . Below one can see the  $\mathcal{R}_1$ -RST from Section 2.2 (on the left), and the subtree that starts at the node of the term  $d(g)$  (on the right). Note that the probabilities are normalized such that the root has the probability 1 again.



We divide the inner nodes of a  $\mathcal{P}$ -CT into  $V \setminus \text{Leaf} = \Gamma \uplus \bar{\Gamma}$  such that the rewrite step at each node from  $\Gamma$  is done with Case (at) or (af), and the rewrite step at each node from  $\bar{\Gamma}$  is done with Case (nt) or (nf). Note that  $\Gamma$  contains the ‘‘important’’ nodes that count for the global Case (4) of a  $\mathcal{P}$ -CT, i.e., every infinite path must contain infinitely many nodes from  $\Gamma$ .

Now we start by proving the  $\Gamma$ -Partition Lemma. Compared to the  $\mathcal{P}$ -Partition Lemma (Lemma 1 from Section 4), the lemma presented (and proven) here is a bit more detailed and technical. Instead of simply partitioning the set of ADPs, we partition the set of nodes  $\Gamma$  within a chain tree. In this way, we obtain an even more versatile lemma that we can use throughout our proofs. Furthermore, note that this lemma does not rely on the structure of ADPs or rewriting, but instead, it works for arbitrary infinite trees labeled by probabilities.

**Lemma 4** ( $\Gamma$ -Partition Lemma). *Let  $\mathcal{P}$  be an ADP problem and let  $\mathfrak{X} = (V, E, L)$  be a  $\mathcal{P}$ -CT that converges with probability  $< 1$ . Assume that we can partition  $\Gamma = \Gamma_1 \uplus \Gamma_2$  such that every subtree that only contains inner nodes from  $\bar{\Gamma} \cup \Gamma_1$  converges with probability 1. Then there is a grounded subtree  $\mathfrak{X}'$  that converges with probability  $< 1$  such that every infinite path has an infinite number of nodes from  $\Gamma_2$ .*

**Proof.** Let  $\mathfrak{X} = (V, E, L)$  be a  $\mathcal{P}$ -CT with  $|\mathfrak{X}| = c < 1$  for some  $c \in \mathbb{R}$ . Since we have  $0 \leq c < 1$ , there is an  $\varepsilon > 0$  such that  $c + \varepsilon < 1$ . Remember that the formula for the geometric series is

$$\sum_{n=1}^{\infty} \left(\frac{1}{d}\right)^n = \frac{1}{d-1}, \text{ for all } d \in \mathbb{R} \text{ such that } \frac{1}{|d|} < 1.$$

Let  $d = \frac{1}{\varepsilon} + 2$ . Now, we have  $\frac{1}{d} = \frac{1}{\frac{1}{\varepsilon} + 2} < 1$  and

$$\frac{1}{\varepsilon} + 1 < \frac{1}{\varepsilon} + 2 \Leftrightarrow \frac{1}{\varepsilon} + 1 < d \Leftrightarrow \frac{1}{\varepsilon} < d - 1 \Leftrightarrow \frac{1}{d-1} < \varepsilon \Leftrightarrow \sum_{n=1}^{\infty} \left(\frac{1}{d}\right)^n < \varepsilon. \tag{B.1}$$

We will now construct a subtree  $\mathfrak{X}' = (V', E', L')$  such that every infinite path has an infinite number of  $\Gamma_2$  nodes and such that

$$|\mathfrak{X}'| \leq |\mathfrak{X}| + \sum_{n=1}^{\infty} \left(\frac{1}{d}\right)^n \tag{B.2}$$

and then, we finally have

$$|\mathfrak{X}'| \stackrel{\text{(B.2)}}{\leq} |\mathfrak{X}| + \sum_{n=1}^{\infty} \left(\frac{1}{d}\right)^n = c + \sum_{n=1}^{\infty} \left(\frac{1}{d}\right)^n \stackrel{\text{(B.1)}}{<} c + \varepsilon < 1.$$

The idea of this construction is that we cut infinite subtrees of pure  $\bar{\Gamma} \cup \Gamma_1$  nodes as soon as the probability for normal forms is high enough. In this way, one obtains paths where after finitely many  $\bar{\Gamma} \cup \Gamma_1$  nodes, there is either a node from  $\Gamma_2$  or a leaf.

For any node  $v \in V$ , let  $\mathcal{L}(v)$  be the number of  $\Gamma_2$  nodes in the path from the root to  $v$ . Furthermore, for any set  $W \subseteq V$  and  $k \in \mathbb{N}$ , let  $\mathfrak{X}(W, k) = \{v \in W \mid \mathcal{L}(v) \leq k \vee (v \in \Gamma_2 \wedge \mathcal{L}(v) \leq k + 1)\}$  be the set of all nodes in  $W$  that have at most  $k$  nodes from  $\Gamma_2$  in the path from the root to its predecessor. So if  $v \in \mathfrak{X}(W, k)$  is not in  $\Gamma_2$ , then we have at most  $k$  nodes from  $\Gamma_2$  in the path from the root to  $v$  and if  $v \in \mathfrak{X}(W, k)$  is in  $\Gamma_2$ , then we have at most  $k + 1$  nodes from  $\Gamma_2$  in the path from the root to  $v$ . We will inductively define a set  $U_k \subseteq V$  such that  $U_k \subseteq \mathfrak{X}(V, k)$  and then define the subtree as  $\mathfrak{X}' = \mathfrak{X}[\bigcup_{k \in \mathbb{N}} U_k]$ .

We start by considering the subtree  $\mathfrak{X}_0 = \mathfrak{X}[\mathfrak{X}(V, 0)]$ . This tree only contains inner nodes from  $\bar{\Gamma} \cup \Gamma_1$ . While the node set  $\mathfrak{X}(V, 0)$  itself may contain nodes from  $\Gamma_2$ , they can only occur at the leaves of  $\mathfrak{X}_0$ . Therefore, we get  $|\mathfrak{X}_0| = 1$  by the prerequisite of the lemma. In Fig. B.8 one can see the different possibilities for  $\mathfrak{X}_0$ . Either  $\mathfrak{X}_0$  is finite or  $\mathfrak{X}_0$  is infinite. In the first case, we can add all the nodes to  $U_0$  since there is no infinite path of pure  $\bar{\Gamma} \cup \Gamma_1$  nodes. Hence, we define  $U_0 = \mathfrak{X}(V, 0)$ . In the second case, we have to cut the tree at a specific depth once the probability of leaves is high enough. Let  $d_0(w)$  be the depth of the node  $w$  in the tree  $\mathfrak{X}_0$ . Moreover, let  $D_0(k) = \{v \in \mathfrak{X}(V, 0) \mid d_0(w) \leq k\}$  be the set of nodes in  $\mathfrak{X}_0$  that have a depth of at most  $k$ . Since  $|\mathfrak{X}_0| = 1$  and  $|\cdot|$  is monotonic w.r.t.

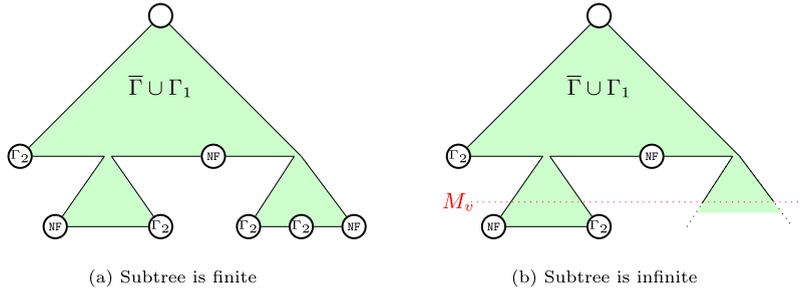


Fig. B.8. Possibilities for  $\mathfrak{T}_0$  and  $\mathfrak{T}_v$ .

the depth of the tree  $\mathfrak{T}_0$ , we can find an  $M_0 \in \mathbb{N}$  such that

$$\sum_{v \in \text{Leaf}^{\mathfrak{T}_0}, d_0(v) \leq M_0} p_v^{\mathfrak{T}_0} \geq 1 - \frac{1}{d}$$

Here,  $\text{Leaf}^{\mathfrak{T}}$  and  $p_v^{\mathfrak{T}}$  denote the set of leaves and the probability of the node  $v$  in the tree  $\mathfrak{T}$ , respectively.

We include all nodes from  $D_0(M_0)$  in  $U_0$  and delete every other node of  $\mathfrak{T}_0$ . In other words, we cut the tree after depth  $M_0$ . This cut can be seen in Fig. B.8, indicated by the dotted line. We now know that this cut may increase the probability of leaves by at most  $\frac{1}{d}$ . Therefore, we define  $U_0 = D_0(M_0)$ .

For the induction step, assume that we have already defined a subset  $U_i \subseteq \mathfrak{Z}(V, i)$ . Let  $H_i = \{v \in U_i \mid v \in \Gamma_2, \mathcal{L}(v) = i + 1\}$  be the set of leaves in  $\mathfrak{Z}[U_i]$  that are in  $\Gamma_2$ . For each  $v \in H_i$ , we consider the subtree that starts at  $v$  until we reach the next node from  $\Gamma_2$ , including the node itself. Everything below such a node will be cut. To be precise, we regard the tree  $\mathfrak{T}_v = (V_v, E_v, L_v) = \mathfrak{Z}[\mathfrak{Z}(vE^*, i + 1)]$ . Here,  $vE^*$  is the set of all nodes that are reachable from  $v$ .

First, we show that  $|\mathfrak{T}_v| = 1$ . For every direct successor  $w$  of  $v$ , the subtree  $\mathfrak{T}_w = \mathfrak{T}_v[wE_v^*]$  of  $\mathfrak{T}_v$  that starts at  $w$  does not contain any inner nodes from  $\Gamma_2$ . Hence, we have  $|\mathfrak{T}_w| = 1$  by the prerequisite of the lemma again, and hence

$$|\mathfrak{T}_v| = \sum_{w \in vE} p_w \cdot |\mathfrak{T}_w| = \sum_{w \in vE} p_w \cdot 1 = \sum_{w \in vE} p_w = 1.$$

For the construction of  $U_{i+1}$ , we have the same cases as before, see Fig. B.8. Either  $\mathfrak{T}_v$  is finite or  $\mathfrak{T}_v$  is infinite. Let  $Z_v$  be the set of nodes that we want to add to our node set  $U_{i+1}$  from the tree  $\mathfrak{T}_v$ . In the first case we can add all the nodes again and set  $Z_v = V_v$ . In the second case, we once again cut the tree at a specific depth once the probability for leaves is high enough. Let  $d_v(z)$  be the depth of the node  $z$  in the tree  $\mathfrak{T}_v$ . Moreover, let  $D_v(k) = \{v \in V_v \mid d_v(z) \leq k\}$  be the set of nodes in  $\mathfrak{T}_v$  that have a depth of at most  $k$ . Since  $|\mathfrak{T}_v| = 1$  and  $|\cdot|$  is monotonic w.r.t. the depth of the tree  $\mathfrak{T}_v$ , we can find an  $M_v \in \mathbb{N}$  such that

$$\sum_{w \in \text{Leaf}^{\mathfrak{T}_v}, d_v(w) \leq M_v} p_w^{\mathfrak{T}_v} \geq 1 - \left(\frac{1}{d}\right)^{i+1} \cdot \frac{1}{|H_i|}$$

We include all nodes from  $D_v(M_v)$  in  $U_{i+1}$  and delete every other node of  $\mathfrak{T}_v$ . In other words, we cut the tree after depth  $M_v$ . We now know that this cut may increase the probability of leaves by at most  $\left(\frac{1}{d}\right)^{i+1} \cdot \frac{1}{|H_i|}$ . Therefore, we set  $Z_v = D_v(M_v)$ .

We do this for each  $v \in H_i$  and in the end, we set  $U_{i+1} = U_i \cup \bigcup_{v \in H_i} Z_v$ .

It is straightforward to see that  $\bigcup_{k \in \mathbb{N}} U_k$  satisfies the conditions of Definition 14, as we only cut after certain nodes in our construction. Hence,  $\bigcup_{k \in \mathbb{N}} U_k$  is non-empty and weakly connected, and for each of its nodes, it either contains no or all successors. Furthermore,  $\mathfrak{T}' = \mathfrak{Z}[\bigcup_{k \in \mathbb{N}} U_k]$  is a subtree which does not contain an infinite path of pure  $\bar{\Gamma} \cup \Gamma_1$  nodes as we cut every such path after a finite depth.

It remains to prove that  $|\mathfrak{T}'| \leq |\mathfrak{Z}| + \sum_{n=1}^{\infty} \left(\frac{1}{d}\right)^n$  holds. During the  $i$ -th iteration of the construction, we may increase the value of  $|\mathfrak{Z}|$  by the sum of all probabilities corresponding to the new leaves resulting from the cuts. As we cut at most  $|H_i|$  trees in the  $i$ -th iteration and for each such tree, we added at most a total probability of  $\left(\frac{1}{d}\right)^{i+1} \cdot \frac{1}{|H_i|}$  for the new leaves, the value of  $|\mathfrak{Z}|$  might increase by

$$|H_i| \cdot \left(\frac{1}{d}\right)^{i+1} \cdot \frac{1}{|H_i|} = \left(\frac{1}{d}\right)^{i+1}$$

in the  $i$ -th iteration, and hence in total, we then get

$$|\mathfrak{T}'| \leq |\mathfrak{Z}| + \sum_{n=1}^{\infty} \left(\frac{1}{d}\right)^n,$$

as desired (see (B.2)).  $\square$

Next, we prove the *Starting Lemma*. The proof nicely illustrates how to use the  $\Gamma$ -Partition Lemma.

**Proof.** (*Proof of Lemma 2*) Assume that every  $\mathcal{P}$ -CT  $\mathfrak{X}$  converges with probability 1 if it starts with  $(1 : t)$  and  $\text{Pos}_{D^\#}(t) = \{\varepsilon\}$ . We now prove that then also every  $\mathcal{P}$ -CT  $\mathfrak{X} = (V, E, L)$  that starts with  $(1 : t)$  for some arbitrary annotated term  $t$  converges with probability 1, and thus  $\mathcal{P}$  is AST. We prove the claim by induction on the number of annotations in the initial term  $t$ .

If  $t$  contains no annotation, then  $\mathfrak{X}$  is trivially finite (it cannot contain an infinite path, since there are no nodes in  $\Gamma$ ) and hence, it converges with probability 1. Next, if  $t$  contains exactly one annotation at position  $\pi$ , then we can ignore everything above the annotation, as we will never use an (at) or (af) step above the annotated position, and we cannot duplicate or change annotations by rewriting above them, since we use VRFs and not GVRFs. For  $t|_\pi$  with  $\text{Pos}_{D^\#}(t|_\pi) = \{\varepsilon\}$ , we know by our assumption that such a CT converges with probability 1.

Next, we regard the induction step, and assume for a contradiction that for a term  $t$  with  $n > 1$  annotations, there is a CT  $\mathfrak{X} = (V, E, L)$  that converges with probability  $< 1$ . Here, our induction hypothesis is that every  $\mathcal{P}$ -CT  $\mathfrak{X}$  that starts with  $(1 : t')$ , where  $t'$  contains  $m$  annotations for some  $1 \leq m < n$  converges with probability 1. Let  $\Pi_1 = \{\tau\}$  and  $\Pi_2 = \{\chi \in \text{Pos}_{D^\#}(t) \mid \chi \neq \tau\}$  for some  $\tau \in \text{Pos}_{D^\#}(t)$  and consider the two terms  $\#_{\Pi_1}(t)$  and  $\#_{\Pi_2}(t)$ , which contain both strictly less than  $n$  annotations. By our induction hypothesis, we know that every  $\mathcal{P}$ -CT that starts with  $(1 : \#_{\Pi_1}(t))$  or  $(1 : \#_{\Pi_2}(t))$  converges with probability 1. Let  $\mathfrak{X}_1 = (V, E, L_1)$  be the tree that starts with  $(1 : \#_{\Pi_1}(t))$  and uses the same rules, the same positions, and the same VRFs as in  $\mathfrak{X}$ .

We can partition the  $\Gamma^\mathfrak{X}$ -nodes (i.e., the  $\Gamma$ -nodes of  $\mathfrak{X}$ ) into  $\Gamma_1^\mathfrak{X} = \Gamma^{\mathfrak{X}_1}$  and  $\Gamma_2^\mathfrak{X} = \Gamma^\mathfrak{X} \setminus \Gamma_1^\mathfrak{X}$  (i.e., the  $\Gamma$ -nodes of  $\mathfrak{X}_1$ , and the rest). Note that  $\mathfrak{X}_1$  itself may not be a  $\mathcal{P}$ -CT again, since there might exist paths without an infinite number of  $\Gamma_1^\mathfrak{X}$ -nodes, but obviously every subtree  $\mathfrak{X}'_1$  of  $\mathfrak{X}_1$  such that every infinite path has an infinite number of  $\Gamma_1^\mathfrak{X}$ -nodes is a  $\mathcal{P}$ -CT again. Moreover, by extending such a subtree to be grounded, i.e., adding the initial path from the root of  $\mathfrak{X}_1$  to  $\mathfrak{X}'_1$ , we created a  $\mathcal{P}$ -CT that starts with  $\#_{\Pi_1}(t)$ , and hence by our induction hypothesis, converges with probability 1. Thus, this also holds for  $\mathfrak{X}'_1$ .

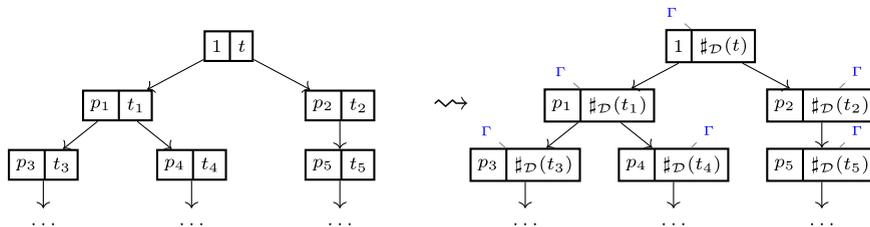
We want to use the  $\Gamma$ -Partition Lemma (Lemma 4) for the tree  $\mathfrak{X}$ . For this, we have to show that every subtree  $\mathfrak{X}'_1$  of  $\mathfrak{X}$  that only contains inner nodes from  $\bar{\Gamma}^\mathfrak{X} \cup \Gamma_1^\mathfrak{X}$  converges with probability 1. But since  $\mathfrak{X}'_1$  only contains inner nodes from  $\bar{\Gamma}^\mathfrak{X} \cup \Gamma_1^\mathfrak{X}$  it must either contain infinitely many  $\Gamma_1^\mathfrak{X}$ -nodes (and by the previous paragraph it converges with probability 1), or it contains only finitely many  $\Gamma_1^\mathfrak{X}$ -nodes, hence must be finite itself, and converges with probability 1.

We have shown that the conditions for the  $\Gamma$ -Partition Lemma (Lemma 4) are satisfied. Thus, we can apply the  $\Gamma$ -Partition Lemma to obtain a grounded subtree  $\mathfrak{X}'_2$  of  $\mathfrak{X}$  with  $|\mathfrak{X}'_2| < 1$  such that on every infinite path, we have an infinite number of  $\Gamma_2^\mathfrak{X}$  nodes. Let  $\mathfrak{X}_2$  be the tree that starts with  $\#_{\Pi_2}(t)$  and uses the same rules, the same positions, and the same VRFs as in  $\mathfrak{X}'_2$ . Again, all local properties for a  $\mathcal{P}$ -CT are satisfied for  $\mathfrak{X}_2$ . Additionally, this time we know that every infinite path has an infinite number of nodes from  $\Gamma_2^\mathfrak{X}$  in  $\mathfrak{X}'_2$ , hence we also know that the global property of a CT is satisfied. This means that  $\mathfrak{X}_2$  is a  $\mathcal{P}$ -CT that starts with  $\#_{\Pi_2}(t)$  and with  $|\mathfrak{X}_2| < 1$ . This is our desired contradiction, which proves the induction step.

So we have proven that we can restrict ourselves to CTs that start with  $(1 : t)$  for a term  $t \in \mathcal{T}^\#$  where  $\text{Pos}_{D^\#}(t) = \{\varepsilon\}$  when analyzing AST and iAST. Finally, since we eventually have to rewrite at the root of our term  $t$  (that only contains an annotation at the root), we can assume that  $b(t) = s\theta$  for a substitution  $\theta$  and an ADP  $s \rightarrow \mu^m \in \mathcal{P}$ . Note that for iAST, we can only perform such a rewrite step if  $b(t) = s\theta \in \text{ANF}_{\mathcal{P}}$ .  $\square$

Next, we prove the chain criterion for full rewriting. In the following, we use the shorthand notations  $\#_D(t) = \#_{\text{Pos}_D(t)}(t)$  and  $\#_\varepsilon(t) = \#_{\{\varepsilon\}}(t)$ .

**Proof.** (*Proof of Theorem 5 Soundness*: Assume that  $\mathcal{R}$  is not AST. Then, there exists an  $\mathcal{R}$ -RST  $\mathfrak{X} = (V, E, L)$  whose root is labeled with  $(1 : t)$  for some term  $t \in \mathcal{T}$  that converges with probability  $< 1$ . We will construct a  $DP(\mathcal{R})$ -CT  $\mathfrak{X}' = (V, E, L')$  with the same underlying tree structure and an adjusted labeling such that  $p_v^\mathfrak{X} = p_v^{\mathfrak{X}'}$  for all  $v \in V$ , where all the inner nodes are in  $\Gamma$ . Since the tree structure and the probabilities are the same, we then get  $|\mathfrak{X}| = |\mathfrak{X}'|$ . So there exists a  $DP(\mathcal{R})$ -CT  $\mathfrak{X}'$  that converges with probability  $< 1$  and  $DP(\mathcal{R})$  is not AST either.

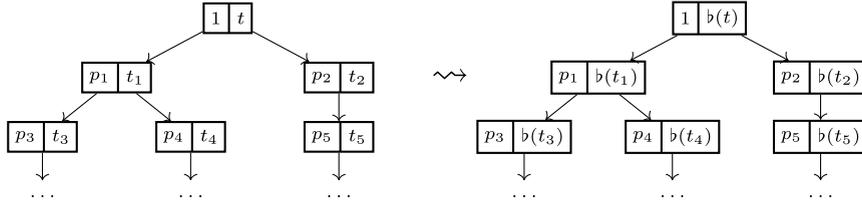


We label all nodes  $v \in V$  in  $\mathfrak{X}'$  with  $\#_D(t_v)$ , where  $t_v$  is the term for the node  $v$  in  $\mathfrak{X}$ . The annotations ensure that we rewrite with Case (at) so that the node  $v$  is contained in  $\Gamma$ . We only have to show that the edge relation represents valid rewrite steps with  $\hookrightarrow_{DP(\mathcal{R})}$ . Let  $v \in V \setminus \text{Leaf}$  and  $vE = \{w_1, \dots, w_k\}$  be the set of its successors. Since  $v$  is not a leaf, we have  $t_v \rightarrow_{\mathcal{R}} \{ \frac{p_{w_1}}{p_v} : t_{w_1}, \dots, \frac{p_{w_k}}{p_v} : t_{w_k} \}$ . This means that there is a rule  $\ell \rightarrow \{p_1 : r_1, \dots, p_k : r_k\} \in \mathcal{R}$ , a position  $\pi$ , and a substitution  $\sigma$  such that  $t_v|_\pi = \ell \sigma$ . Furthermore, we have  $t_{w_j} = t_v[r_j \sigma]_\pi$  for all  $1 \leq j \leq k$ .

The corresponding ADP for the rule is  $\ell \rightarrow \{p_1 : \#_D(r_1), \dots, p_k : \#_D(r_k)\}^{\text{true}}$ . Furthermore,  $\pi \in \text{Pos}_{D^\#}(\#_D(t_v))$  as all defined symbols are annotated in  $\#_D(t_v)$ . Hence, we can rewrite  $\#_D(t_v)$  with  $\ell \rightarrow \{p_1 : \#_D(r_1), \dots, p_k : \#_D(r_k)\}^{\text{true}}$ , using the position  $\pi$ , the substitution  $\sigma$ , and Case (at) applies. Furthermore, we take some VRF  $(\varphi_j)_{1 \leq j \leq k}$  that is surjective on the positions of the variables in the right-

hand side, i.e., for all  $1 \leq j \leq k$  and all positions  $\tau \in \text{Pos}_V(r_j)$  there exists a  $\tau' \in \text{Pos}_V(\ell)$  such that  $\varphi_j(\tau') = \tau$ . Such a VRF must exist, since  $\mathcal{R}$  is non-duplicating. We have  $\#_D(t_v) \xrightarrow{DP(\mathcal{R})} \{p_1 : \#_D(t_{w_1}), \dots, p_k : \#_D(t_{w_k})\}$  since by rewriting with Case (at) we get  $\#_D(t_v)[\#_{\text{Pos}_{D^\#}(r_j) \cup \Psi_j(r_j \sigma)}]_\pi = \#_D(t_{w_j})$  with  $\Psi_j$  defined as in Definition 5. Note that since the used VRF is surjective on the variable positions of the right-hand side, we do not remove any annotation in the substitution. Furthermore, we annotated all defined symbols in  $r_j$ . Thus, we result in  $\#_D(t_{w_j})$  where all defined symbols are annotated again.

**Completeness:** Assume that  $DP(\mathcal{R})$  is not AST. Then, there exists a  $DP(\mathcal{R})$ -CT  $\mathfrak{Z} = (V, E, L)$  whose root is labeled with  $(1 : t)$  for some annotated term  $t \in \mathcal{T}^\#$  that converges with probability  $< 1$ . We will construct an  $\mathcal{R}$ -RST  $\mathfrak{Z}' = (V, E, L')$  with the same underlying tree structure and an adjusted labeling such that  $p_v^\mathfrak{Z} = p_v^{\mathfrak{Z}'}$  for all  $v \in V$ . Since the tree structure and the probabilities are the same, we then get  $|\mathfrak{Z}'| = |\mathfrak{Z}|$ . Therefore, there exists an  $\mathcal{R}$ -RST  $\mathfrak{Z}'$  that converges with probability  $< 1$ . Hence,  $\mathcal{R}$  is not AST either.



We label all nodes  $v \in V$  in  $\mathfrak{Z}'$  with  $b(t_v)$ , where  $t_v$  is the term for the node  $v$  in  $\mathfrak{Z}$ , i.e., we remove all annotations. We only have to show that  $\mathfrak{Z}'$  is indeed a valid RST, i.e., that the edge relation represents valid rewrite steps with  $\rightarrow_{\mathcal{R}}$ , but this follows directly from the fact that if we remove all annotations in Definition 5, then we get the ordinary probabilistic term rewriting relation again.  $\square$

To prove the innermost chain criterion, we define the set  $\text{Pos}_{D \wedge \text{NF}_{\mathcal{R}}}(t)$  of all positions of subterms of  $t$  that may be used as a redex now or in future rewrite steps, because the subterm has a defined root symbol and is not in normal form w.r.t.  $\mathcal{R}$ . This set is needed, as we cannot simply annotate every defined symbol in the soundness proof of the innermost chain criterion. Instead, we can (and it suffices to) only guarantee that all the symbols in  $\text{Pos}_{D \wedge \text{NF}_{\mathcal{R}}}(t)$  are annotated.

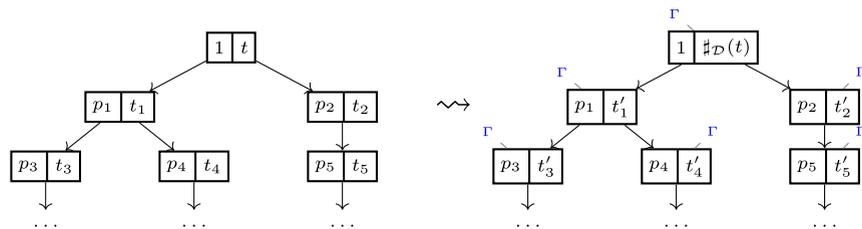
**Definition 15** ( $\text{Pos}_{D \wedge \text{NF}_{\mathcal{R}}}$ ). Let  $\mathcal{R}$  be a PTRS. For a term  $t \in \mathcal{T}$  we define  $\text{Pos}_{D \wedge \text{NF}_{\mathcal{R}}}(t) = \{\pi \mid \pi \in \text{Pos}_D(t), t|_\pi \notin \text{NF}_{\mathcal{R}}\}$ .

**Example 34.** Consider the following PTRS  $\mathcal{R}$  over a signature with  $D = \{f, g\}$  and  $C = \{a, b\}$  with the rules  $f(x, a) \rightarrow \{1 : b\}$  and  $g(a) \rightarrow \{1 : a\}$ . For the term  $t = f(g(a), g(b))$  we have  $\text{Pos}_D(t) = \{\varepsilon, 1, 2\}$  and  $\text{Pos}_{D \wedge \text{NF}_{\mathcal{R}}}(t) = \{\varepsilon, 1\}$ .

Finally, for two (possibly annotated) terms  $s, t$  we define  $s \stackrel{b}{=} t$  if  $b(s) = b(t)$ . Remember that for innermost rewriting, we do not use any VRFs, hence, we can ignore them in the proofs for innermost rewriting.

**Proof.** (Proof of Theorem 4) In the following, we will often implicitly use that for an annotated term  $t \in \mathcal{T}^\#$ , we have  $b(t) \in \text{ANF}_{\mathcal{R}}$  iff  $t \in \text{ANF}_{DP(\mathcal{R})}$  since a rewrite rule and its corresponding canonical annotated dependency pair have the same left-hand side.

**Soundness:** Assume that  $\mathcal{R}$  is not iAST. Then, there exists an innermost  $\mathcal{R}$ -RST  $\mathfrak{Z} = (V, E, L)$  whose root is labeled with  $(1 : t)$  for some term  $t \in \mathcal{T}$  that converges with probability  $< 1$ . We use almost the same construction as in the proof of soundness for AST, however, instead of annotating all defined symbols, we only ensure that all defined symbols from  $\text{Pos}_{D \wedge \text{NF}_{\mathcal{R}}}(t)$  are annotated.



We construct the new labeling  $L'$  for the  $DP(\mathcal{R})$ -CT inductively such that for all inner nodes  $v \in V \setminus \text{Leaf}$  with children nodes  $vE = \{w_1, \dots, w_k\}$  we have  $t'_v \xrightarrow{DP(\mathcal{R})} \{\frac{p_{w_1}}{p_v} : t'_{w_1}, \dots, \frac{p_{w_k}}{p_v} : t'_{w_k}\}$  and use Case (at) so that all nodes are in  $\Gamma$ . Let  $W \subseteq V$  be the set of nodes  $v$  where we have already defined the labeling  $L'(v)$ . During our construction, we ensure that for all  $v \in W$  we have

$$t_v \stackrel{b}{=} t'_v \text{ and } \text{Pos}_{D \wedge \text{NF}_{\mathcal{R}}}(t_v) \subseteq \text{Pos}_{D^\#}(t'_v).$$

This means that the corresponding term  $t_v$  for the node  $v$  in  $\mathfrak{Z}$  has the same structure as the term  $t'_v$  in  $\mathfrak{Z}'$ , and additionally, all the possible redexes in  $t_v$  are annotated in  $t'_v$ . We label the root of  $\mathfrak{Z}'$  with  $\#_D(t)$ . Here, we have  $t \stackrel{b}{=} \#_D(t)$  and  $\text{Pos}_{D \wedge \text{NF}_{\mathcal{R}}}(t) \subseteq \text{Pos}_D(t) = \text{Pos}_{D^\#}(\#_D(t))$ . As long as there is still an inner node  $v \in W$  such that its successors are not contained in  $W$ , we do the following. Let  $vE = \{w_1, \dots, w_k\}$  be the set of its successors. We need to define the corresponding terms  $t'_{w_1}, \dots, t'_{w_k}$  for the nodes  $w_1, \dots, w_k$ . Since  $v$  is not a leaf, we have  $t_v \xrightarrow{i} \{\frac{p_{w_1}}{p_v} : t_{w_1}, \dots, \frac{p_{w_k}}{p_v} : t_{w_k}\}$ . This means that there is a rule  $\ell \rightarrow \{p_1 : r_1, \dots, p_k : r_k\} \in \mathcal{R}$ , a position  $\pi$ , and a substitution  $\sigma$  such that  $t_v|_\pi = \ell \sigma \in \text{ANF}_{\mathcal{R}}$ . Furthermore, we have  $t_{w_j} = t_v[r_j \sigma]_\pi$  for all  $1 \leq j \leq k$ . So the labeling of the successor  $w_j$  in  $\mathfrak{Z}$  is  $L(w_j) = (p_v \cdot p_j : t_v[r_j \sigma]_\pi)$  for all  $1 \leq j \leq k$ .

The corresponding ADP for the rule is  $\ell \rightarrow \{p_1 : \#_D(r_1), \dots, p_k : \#_D(r_k)\}^{\text{true}}$ . Furthermore,  $\pi \in \text{Pos}_{D \wedge \neg \text{NF}_R}(t_v) \subseteq_{(IH)} \text{Pos}_{D^\#}(t'_v)$  and  $t_v \stackrel{b}{=}_{(IH)} t'_v$ . Hence, we can rewrite  $t'_v$  with  $\ell \rightarrow \{p_1 : \#_D(r_1), \dots, p_k : \#_D(r_k)\}^{\text{true}}$ , using the position  $\pi$  and the substitution  $\sigma$ , and Case (at) applies. We get  $t'_v \stackrel{c}{\rightarrow}_{DP(R)} \{p_1 : t'_{w_1}, \dots, p_k : t'_{w_k}\}$  with  $t'_{w_j} = t'_v[\#_D(r_j)\sigma]_\pi$ . This means that we have  $t_{w_j} \stackrel{b}{=} t'_{w_j}$ . It remains to prove  $\text{Pos}_{D \wedge \neg \text{NF}_R}(t_{w_j}) \subseteq \text{Pos}_{D^\#}(t'_{w_j})$  for all  $1 \leq j \leq k$ . For all positions  $\tau \in \text{Pos}_{D \wedge \neg \text{NF}_R}(t_{w_j}) = \text{Pos}_{D \wedge \neg \text{NF}_R}(t_v[r_j\sigma]_\pi)$  that are orthogonal or above  $\pi$ , we have  $\tau \in \text{Pos}_{D \wedge \neg \text{NF}_R}(t_v, \mathcal{R}) \subseteq_{(IH)} \text{Pos}_{D^\#}(t'_v)$ , and all annotations orthogonal or above  $\pi$  remain in  $t'_{w_j}$  as they were in  $t'_v$ . For all positions  $\tau \in \text{Pos}_{D \wedge \neg \text{NF}_R}(t_{w_j}) = \text{Pos}_{D \wedge \neg \text{NF}_R}(t_v[r_j\sigma]_\pi)$  that are below  $\pi$ , we know that, due to innermost evaluation, at least the defined root symbol of a term that is not in normal form must be inside  $r_j$ , and thus  $\tau \in \text{Pos}_{D^\#}(t'_{w_j})$ , as all defined symbols of  $r_j$  are annotated in  $t'_{w_j} = t'_v[\#_D(r_j)\sigma]_\pi$ .

Completeness: Analogous to completeness for AST.  $\square$

Next, we prove soundness and completeness of the ADP processors. Recall that most processors are complete by [Theorem 6](#).

**Proof.** (Proof of [Theorem 7](#)) Let  $\bar{X} = X \cup b(\mathcal{P} \setminus X)$  for  $X \subseteq \mathcal{P}$ .

Completeness: By [Theorem 6](#).

Soundness: Let  $\mathfrak{G}$  be the  $\mathcal{P}$ -dependency graph. Suppose that every  $\bar{\mathcal{P}}_i$ -CT converges with probability 1 for all  $1 \leq i \leq n$ . We prove that then also every  $\mathcal{P}$ -CT converges with probability 1. Let  $\mathfrak{B} = \{\mathcal{P}_1, \dots, \mathcal{P}_n\} \cup \{\{\alpha\} \subseteq \mathcal{P} \mid \alpha \text{ is not in an SCC of } \mathfrak{G}\}$  be the set of all SCCs and all singleton sets of nodes that do not belong to any SCC. For two  $X_1, X_2 \in \mathfrak{B}$  we say that  $X_2$  is a *direct successor* of  $X_1$  (denoted  $X_1 >_{\mathfrak{G}} X_2$ ) if there exist nodes  $\alpha \in X_1$  and  $\beta \in X_2$  such that there is an edge from  $\alpha$  to  $\beta$  in  $\mathfrak{G}$ . The core steps of this proof are the following:

1. We show that every ADP problem  $\bar{X}$  with  $X \in \mathfrak{B}$  is AST.
2. We show that composing SCCs maintains the AST property.
3. We show that for every  $X \in \mathfrak{B}$ , the ADP problem  $\bigcup_{X >_{\mathfrak{G}}^* Y} Y$  is AST by induction on  $>_{\mathfrak{G}}$ .
4. We conclude that  $\mathcal{P}$  must be AST.

**1. Every ADP problem  $\bar{X}$  with  $X \in \mathfrak{B}$  is AST:** We start by proving the following:

Every ADP problem  $\bar{X}$  with  $X \in \mathfrak{B}$  is AST. (B.3)

To prove (B.3), note that if  $X$  is an SCC, then it follows from our assumption that  $\bar{X}$  is AST. If  $X$  is a singleton set of a node that does not belong to any SCC, then assume for a contradiction that  $\bar{X}$  is not AST. By [Lemma 2](#) there exists an  $\bar{X}$ -CT  $\mathfrak{Z} = (V, E, L)$  that converges with probability  $< 1$  and starts with  $(1 : t)$  where  $\text{Pos}_{D^\#}(t) = \{\varepsilon\}$  and  $b(t) = s\theta$  for a substitution  $\theta$  and some ADP  $\alpha = s \rightarrow \{p_1 : r_1, \dots, p_k : r_k\}^m \in \bar{X}$ . If  $\alpha \notin X$ , then the resulting terms after the first rewrite step contain no annotations anymore and this cannot start a CT that converges with probability  $< 1$ . Hence, we have  $\alpha \in X$  and thus,  $X = \{\alpha\}$ , since  $X$  is a singleton set. Assume for a contradiction that there exists a node  $v \in \Gamma$  in  $\mathfrak{Z}$  that is not the root and introduces new annotations. W.l.o.g., let  $v$  be reachable from the root without traversing any other node that introduces new annotations. This means that for the corresponding term  $t_v$  for node  $v$  there is a  $t' \preceq_{\#} t_v$  at position  $\tau$  such that  $t' = s\sigma'$  for some substitution  $\sigma'$  (since  $s \rightarrow \dots$  is the only ADP in  $\bar{X}$  that contains any annotations in the right-hand side). Let  $(z_0, \dots, z_m)$  with  $z_m = v$  be the path from the root to  $v$  in  $\mathfrak{Z}$ . The first rewrite step at the root must be  $s\theta \xrightarrow{X} \{p_1 : r_1\theta, \dots, p_k : r_k\theta\}$ . After that, we only use ADPs with the flag true below the annotated position that will be used for the rewrite step at node  $v$ , as otherwise, the position  $\tau$  would not be annotated in  $t_v$ . Therefore, we must have an  $1 \leq j \leq k$  and a  $t'' \preceq_{\#} r_j$  such that  $t''\theta \xrightarrow{*}_{\text{np}(\mathcal{P})} s^\# \sigma'$ , which means that there must be a self-loop for the only ADP in  $X$ , which is a contradiction to our assumption that  $X$  is a singleton consisting of an ADP that is not in any SCC of  $\mathfrak{G}$ .

Now, we have proven that the  $\bar{X}$ -CT  $\mathfrak{Z}$  does not introduce new annotations. By definition of a  $\mathcal{P}$ -CT, every infinite path must contain an infinite number of nodes in  $\Gamma$ , i.e., nodes where we rewrite at an annotation. Thus, every path in  $\mathfrak{Z}$  must be finite, which means that  $\mathfrak{Z}$  is finite itself, as the tree is finitely branching. But every finite CT converges with probability 1, which is a contradiction to our assumption that  $\mathfrak{Z}$  converges with probability  $< 1$ .

**2. Composing SCCs maintains the AST property:** In the second step, we show that composing SCCs maintains the AST property. More precisely, we prove the following:

Let  $\hat{X} \subseteq \mathfrak{B}$  and  $\hat{Y} \subseteq \mathfrak{B}$  such that there are no  $X_1, X_2 \in \hat{X}$  and  $Y \in \hat{Y}$  which satisfy both  $X_1 >_{\mathfrak{G}}^* Y >_{\mathfrak{G}}^* X_2$  and  $Y \notin \hat{X}$ , and such that there are no  $Y_1, Y_2 \in \hat{Y}$  and  $X \in \hat{X}$  which satisfy both  $Y_1 >_{\mathfrak{G}}^* X >_{\mathfrak{G}}^* Y_2$  and  $X \notin \hat{Y}$ . (B.4)

If both  $\overline{\bigcup_{X \in \hat{X}} X}$  and  $\overline{\bigcup_{Y \in \hat{Y}} Y}$  are AST, then  $\overline{\bigcup_{X \in \hat{X}} X \cup \bigcup_{Y \in \hat{Y}} Y}$  is AST.

To show (B.4), we assume that both  $\overline{\bigcup_{X \in \hat{X}} X}$  and  $\overline{\bigcup_{Y \in \hat{Y}} Y}$  are AST. Let  $\bar{Z} = \overline{\bigcup_{X \in \hat{X}} X \cup \bigcup_{Y \in \hat{Y}} Y}$ . The property in (B.4) for  $\hat{X}$  and  $\hat{Y}$  says that a path between two nodes from  $\bigcup_{X \in \hat{X}} X$  that only traverses nodes from  $Z$  must also be a path that only traverses nodes from  $\bigcup_{X \in \hat{X}} X$ , so that  $\bigcup_{Y \in \hat{Y}} Y$  cannot be used to “create” new paths between two nodes from  $\bigcup_{X \in \hat{X}} X$ , and vice versa. Assume for a contradiction that  $\bar{Z}$  is not AST. By [Lemma 2](#) there exists a  $\bar{Z}$ -CT  $\mathfrak{Z} = (V, E, L)$  that converges with probability  $< 1$  and starts with  $(1 : t)$  where  $\text{Pos}_{D^\#}(t) = \{\varepsilon\}$  and  $b(t) = s\theta$  for a substitution  $\theta$  and an ADP  $s \rightarrow \dots \in \bar{Z}$ .

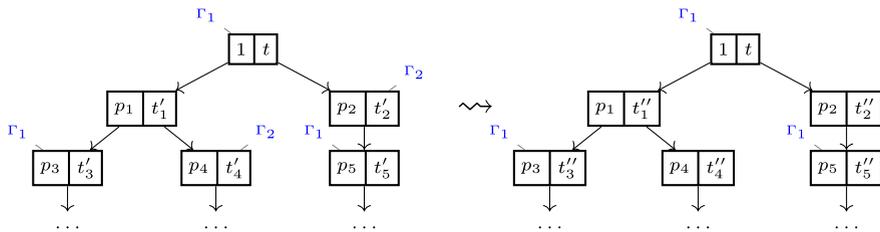
If  $s \rightarrow \dots \notin \bigcup_{X \in \hat{X}} X \cup \bigcup_{Y \in \hat{Y}} Y$ , then the resulting terms contain no annotations anymore and this cannot start a CT that converges with probability  $< 1$ . W.l.o.g., we may assume that the ADP that is used for the rewrite step at the root is in  $\bigcup_{X \in \hat{X}} X$ . Otherwise, we simply swap  $\bigcup_{X \in \hat{X}} X$  with  $\bigcup_{Y \in \hat{Y}} Y$  in the following.

We can partition the set  $\Gamma$  of our  $\overline{Z}$ -CT  $\mathfrak{Z}$  into the sets

- $\Gamma_1 := \{v \in \Gamma \mid v \text{ together with the labeling and its successors represents a step with an ADP from } \bigcup_{X \in \hat{X}} X\}$
- $\Gamma_2 := \Gamma \setminus \Gamma_1$

Note that the labeling represents a step with an ADP from  $\mathcal{P} \setminus \bigcup_{X \in \hat{X}} X$  at nodes  $v \in \Gamma_2$ . Every  $\overline{\bigcup_{Y \in \hat{Y}} Y}$ -CT converges with probability 1, since  $\overline{\bigcup_{Y \in \hat{Y}} Y}$  is AST. Thus, also every  $\overline{\bigcup_{Y \in \hat{Y}} Y \setminus \bigcup_{X \in \hat{X}} X}$ -CT converges with probability 1 (as it contains fewer annotations than  $\overline{\bigcup_{Y \in \hat{Y}} Y}$ ). Furthermore, we have  $|\mathfrak{Z}'| < 1$  by our assumption. By the  $\Gamma$ -Partition Lemma (Lemma 4) we can find a grounded sub  $\overline{Z}$ -CT  $\mathfrak{Z}' = (V', E', L')$  with  $|\mathfrak{Z}'| < 1$  such that every infinite path has an infinite number of  $\Gamma_1$ -nodes.

We now construct a  $\overline{\bigcup_{X \in \hat{X}} X}$ -CT  $\mathfrak{Z}'' = (V'', E'', L'')$  with  $\Gamma_1 \cap \Gamma^{\mathfrak{Z}''} \subseteq \Gamma^{\mathfrak{Z}'}$  that has the same underlying tree structure and adjusted labeling such that all nodes get the same probabilities as in  $\mathfrak{Z}'$ . Hence, we then obtain  $|\mathfrak{Z}''| = |\mathfrak{Z}'|$ . Moreover, every infinite path in  $\mathfrak{Z}''$  contains infinitely many  $\Gamma^{\mathfrak{Z}''}$ -nodes, since every path in  $\mathfrak{Z}'$  contains infinitely many  $\Gamma_1$ -nodes and we have  $\Gamma_1 \cap \Gamma^{\mathfrak{Z}''} \subseteq \Gamma^{\mathfrak{Z}'}$ . But this implies that  $\mathfrak{Z}''$  is indeed a  $\overline{\bigcup_{X \in \hat{X}} X}$ -CT with  $|\mathfrak{Z}''| < 1$ , which is a contradiction to our assumption that  $\overline{\bigcup_{X \in \hat{X}} X}$  is AST.



The core idea of this construction is that annotations introduced by rewrite steps at a node  $v \in \Gamma_2$  are not important for our computation. The reason is that if annotations are introduced using an ADP from  $\bigcup_{Y \in \hat{Y}} Y$  that is not contained in  $\bigcup_{X \in \hat{X}} X$ , then by the prerequisite of (B.4), we know that such an ADP has no path in the dependency graph to an ADP in  $\bigcup_{X \in \hat{X}} X$ . Hence, by definition of the dependency graph, we are never able to use these terms for a rewrite step with an ADP from  $\bigcup_{X \in \hat{X}} X$  to introduce new annotations. We can therefore apply the non-annotated ADP from  $\bigcup_{Y \in \hat{Y}} Y$  instead.

We now construct the new labeling  $L''$  for the  $\overline{\bigcup_{X \in \hat{X}} X}$ -CT  $\mathfrak{Z}''$  recursively. Let  $W \subseteq V$  be the set of nodes where we have already defined the labeling  $L''$ . Furthermore, for any term  $t'_v$ , let  $\text{Junk}_{\hat{X}}(t'_v)$  denote the positions of all annotated subterms  $s \leq_{\#} t'_v$  that can never be used for a rewrite step with an ADP from  $\hat{X}$ , as indicated by the dependency graph. To be precise, we define  $\pi \in \text{Junk}_{\hat{X}}(t'_v) \Leftrightarrow$  there is no  $B \in \mathfrak{B}$  with  $B >_{\#}^* X$  for some  $X \in \hat{X}$  such that there is an ADP  $\ell \rightarrow \{p_1 : r_1, \dots, p_k : r_k\}^m \in B$ , and a substitution  $\sigma$  with  $\#_{\varepsilon}(t'_v|_{\pi}) \rightarrow_{\text{np}(\mathcal{P})}^* \ell^{\#} \sigma$ . During our construction, we ensure that for every  $v \in W$  we have

$$t'_v \stackrel{b}{=} t''_v \text{ and } \text{Pos}_{\mathcal{D}^{\#}}(t'_v) \setminus \text{Junk}_{\hat{X}}(t'_v) \subseteq \text{Pos}_{\mathcal{D}^{\#}}(t''_v). \quad (\text{B.5})$$

We start by setting  $t''_v = t'_v = t$  for the root  $v$  of  $\mathfrak{Z}'$ . Here, our property (B.5) is clearly satisfied. As long as there is still an inner node  $v \in W$  such that its successors are not contained in  $W$ , we do the following. Let  $vE = \{w_1, \dots, w_k\}$  be the set of its successors. We need to define the corresponding terms for the nodes  $w_1, \dots, w_k$  in  $\mathfrak{Z}''$ . Since  $v$  is not a leaf and  $\mathfrak{Z}'$  is a  $\overline{Z}$ -CT, we have  $t'_v \hookrightarrow_{\overline{Z}} \{\frac{p_{w_1}}{p_v} : t'_{w_1}, \dots, \frac{p_{w_k}}{p_v} : t'_{w_k}\}$ , and hence, we have to deal with the following two cases:

1. If we use an ADP from  $\bigcup_{X \in \hat{X}} X$  in  $\mathfrak{Z}'$ , then we perform the rewrite step with the same ADP, the same VRF  $(\varphi_j)_{1 \leq j \leq k}$ , the same position  $\pi$ , and the same substitution in  $\mathfrak{Z}''$ . Since we have  $t'_v \stackrel{b}{=}_{(IH)} t''_v$ , we also get  $t'_{w_j} \stackrel{b}{=}_{w_j} t''_{w_j}$  for all  $1 \leq j \leq k$ . Furthermore, since we rewrite at position  $\pi$  it cannot be in  $\text{Junk}_{\hat{X}}(t'_v)$ , and hence, if  $\pi \in \text{Pos}_{\mathcal{D}^{\#}}(t'_v)$ , then also  $\pi \in \text{Pos}_{\mathcal{D}^{\#}}(t''_v)$  by (B.5). Thus, whenever we create annotations in the rewrite step in  $\mathfrak{Z}'$  (a step with (af) or (at)), then we do the same in  $\mathfrak{Z}''$  (the step is also an (af) or (at) step, respectively), and whenever we remove annotations in the rewrite step in  $\mathfrak{Z}'$  (a step with (nf) or (nt)), then the same happened in  $\mathfrak{Z}''$  (the step is also an (nf) or (nt) step). Therefore, we also get  $\text{Pos}_{\mathcal{D}^{\#}}(t'_{w_j}) \setminus \text{Junk}_{\hat{X}}(t'_{w_j}) \subseteq \text{Pos}_{\mathcal{D}^{\#}}(t''_{w_j})$  for all  $1 \leq j \leq k$  and (B.5) is again satisfied.
2. If we use an ADP from  $\mathcal{P} \setminus \bigcup_{X \in \hat{X}} X$  in  $\mathfrak{Z}'$ , and we use the ADP  $\ell \rightarrow \{p_1 : r_1, \dots, p_k : r_k\}^m$ , then we can use  $\ell \rightarrow \{p_1 : b(r_1), \dots, p_k : b(r_k)\}^m$  instead, with the same VRF  $(\varphi_j)_{1 \leq j \leq k}$ , the same position  $\pi$ , and the same substitution. Note that if  $\pi \in \text{Pos}_{\mathcal{D}^{\#}}(t'_v)$ , then all positions of annotations introduced by the ADP are in  $\text{Junk}_{\hat{X}}(t'_{w_j})$  for all  $1 \leq j \leq k$ , since the used ADP is not in  $\bigcup_{X \in \hat{X}} X$  and by (B.4) we cannot use another ADP from  $\bigcup_{Y \in \hat{Y}} Y \setminus \bigcup_{X \in \hat{X}} X$  to create a path in the dependency graph to a node in  $\bigcup_{X \in \hat{X}} X$  again. Otherwise, we remove the annotations during the application of the rule anyway. Again, (B.5) is satisfied.

We have now shown that (B.4) holds.

**3. For every  $X \in \mathfrak{B}$ , the ADP problem  $\overline{\bigcup_{X >_{\mathfrak{G}}^* Y} Y}$  is AST:** Using (B.3) and (B.4), by induction on  $>_{\mathfrak{G}}$  we now prove that

$$\text{for every } X \in \mathfrak{B}, \text{ the ADP problem } \overline{\bigcup_{X >_{\mathfrak{G}}^* Y} Y} \text{ is AST.} \quad (\text{B.6})$$

Note that  $>_{\mathfrak{G}}$  is well founded, since  $\mathfrak{G}$  is finite.

For the base case, we consider an  $X \in \mathfrak{B}$  that is minimal w.r.t.  $>_{\mathfrak{G}}$ . Hence, we have  $\bigcup_{X >_{\mathfrak{G}}^* Y} Y = X$ . By (B.3),  $\overline{X}$  is AST.

For the induction step, we consider an  $X \in \mathfrak{B}$  and assume that  $\overline{\bigcup_{Y >_{\mathfrak{G}}^* Z} Z}$  is AST for every  $Y \in \mathfrak{B}$  with  $X >_{\mathfrak{G}}^+ Y$ . Let  $\text{Succ}(X) = \{Y \in \mathfrak{B} \mid X >_{\mathfrak{G}} Y\} = \{Y_1, \dots, Y_m\}$  be the set of all direct successors of  $X$ . The induction hypothesis states that  $\overline{\bigcup_{Y_i >_{\mathfrak{G}}^* Z} Z}$  is AST for all  $1 \leq i \leq m$ . We first prove by induction that for all  $1 \leq u \leq m$ ,  $\overline{\bigcup_{1 \leq i \leq u} \bigcup_{Y_i >_{\mathfrak{G}}^* Z} Z}$  is AST.

In the inner induction base, we have  $u = 1$  and hence  $\overline{\bigcup_{1 \leq i \leq u} \bigcup_{Y_i >_{\mathfrak{G}}^* Z} Z} = \overline{\bigcup_{Y_1 >_{\mathfrak{G}}^* Z} Z}$ . By our outer induction hypothesis we know that  $\overline{\bigcup_{Y_1 >_{\mathfrak{G}}^* Z} Z}$  is AST.

In the inner induction step, assume that the claim holds for some  $1 \leq u < m$ . Then  $\overline{\bigcup_{Y_{u+1} >_{\mathfrak{G}}^* Z} Z}$  is AST by our outer induction hypothesis and  $\overline{\bigcup_{1 \leq i \leq u} \bigcup_{Y_i >_{\mathfrak{G}}^* Z} Z}$  is AST by our inner induction hypothesis. By (B.4), we know that then  $\overline{\bigcup_{1 \leq i \leq u+1} \bigcup_{Y_i >_{\mathfrak{G}}^* Z} Z}$  is AST as well. The conditions for (B.4) are clearly satisfied, as we use the reflexive, transitive closure  $>_{\mathfrak{G}}^*$  of the direct successor relation in both  $\overline{\bigcup_{1 \leq i \leq u} \bigcup_{Y_i >_{\mathfrak{G}}^* Z} Z}$  and  $\overline{\bigcup_{Y_{u+1} >_{\mathfrak{G}}^* Z} Z}$ .

Now we have shown that  $\overline{\bigcup_{1 \leq i \leq m} \bigcup_{Y_i >_{\mathfrak{G}}^* Z} Z}$  is AST. We know that  $\overline{X}$  is AST by our assumption and that  $\overline{\bigcup_{1 \leq i \leq m} \bigcup_{Y_i >_{\mathfrak{G}}^* Z} Z}$  is AST. Hence, by (B.4) we obtain that  $\overline{\bigcup_{X >_{\mathfrak{G}}^* Y} Y}$  is AST. Again, the conditions of (B.4) are satisfied, since  $X$  is strictly greater w.r.t.  $>_{\mathfrak{G}}^+$  than all  $Z$  with  $Y_i >_{\mathfrak{G}}^* Z$  for some  $1 \leq i \leq m$ .

**4.  $\mathcal{P}$  is AST:** In (B.6) we have shown that  $\overline{\bigcup_{X >_{\mathfrak{G}}^* Y} Y}$  for every  $X \in \mathfrak{B}$  is AST. Let  $X_1, \dots, X_m \in \mathfrak{B}$  be the maximal elements of  $\mathfrak{B}$  w.r.t.  $>_{\mathfrak{G}}$ . By induction, one can prove that  $\overline{\bigcup_{1 \leq i \leq u} \bigcup_{X_i >_{\mathfrak{G}}^* Y} Y}$  is AST for all  $1 \leq u \leq m$  by (B.4), analogous to the previous induction. Again, the conditions of (B.4) are satisfied as we use the reflexive, transitive closure of  $>_{\mathfrak{G}}$ . In the end, we know that  $\overline{\bigcup_{1 \leq i \leq m} \bigcup_{X_i >_{\mathfrak{G}}^* Y} Y} = \mathcal{P}$  is AST and this ends the proof.  $\square$

**Proof.** (Proof of Theorem 8) Let  $\overline{X} = X \cup b(\mathcal{P} \setminus X)$  for  $X \subseteq \mathcal{P}$ .

Completeness: By Theorem 6.

Soundness: The proof is analogous to the proof for AST but uses the innermost dependency graph  $\mathfrak{G}$  instead. Only Step 1. and 2. change slightly due to innermost rewriting.

Step 1. Reusing the notation from the proof for AST, existence of a node  $v \in \Gamma$  in  $\mathfrak{X}$  that is not the root and where we introduce new annotations leads to a term  $t' \triangleleft_{\#} t_v$  such that  $t' = s\sigma' \in \text{ANF}_{\mathcal{P}}$  for some substitution  $\sigma'$  and the ADP  $s \rightarrow \dots \in X$  where  $s\sigma' \in \text{ANF}_{\mathcal{P}}$ . The first rewrite step at the root is  $s\theta \xrightarrow{\text{c}_{\overline{X}}} \{p_1 : r_1\theta, \dots, p_k : r_k\theta\}$ . Therefore, we must have a  $1 \leq j \leq k$  and a  $t'' \triangleleft_{\#} r_j$  such that  $t''\theta \xrightarrow{\text{c}_{\text{np}(\mathcal{P})}} s^{\#}\sigma'$  and  $s\theta, s\sigma' \in \text{ANF}_{\mathcal{P}}$ , which means that there must be a self-loop for the only ADP in  $X$ , which is the desired contradiction.

Step 2. The construction in this step now considers innermost CTs. Hence, the definition of  $\text{Junk}_{\overline{X}}(t'_v)$  uses the innermost dependency graph and innermost rewriting. To be precise, we define  $\pi \in \text{Junk}_{\overline{X}}(t'_v) \Leftrightarrow$  there is no  $B \in \mathfrak{B}$  with  $B >_{\mathfrak{G}}^* X$  for some  $X \in \overline{X}$  such that there is an ADP  $\ell \rightarrow \{p_1 : r_1, \dots, p_k : r_k\}^m \in B$ , and a substitution  $\sigma$  with  $\#_{\varepsilon}(t'_v |_{\pi}) \xrightarrow{\text{c}_{\text{np}(\mathcal{P})}} \ell^{\#}\sigma$  and  $\ell^{\#}\sigma \in \text{ANF}_{\text{np}(\mathcal{P})}$ . Due to the definition of the innermost dependency graph, we can still perform each rewrite step from  $\mathfrak{X}'$  analogously in  $\mathfrak{X}''$  with  $\Gamma_1 \cap \Gamma^{\mathfrak{X}'} \subseteq \Gamma^{\mathfrak{X}''}$ .  $\square$

**Proof.** (Proof of Theorem 9)

Completeness: By Theorem 6.

Soundness: Let  $\mathcal{P}$  be not AST. Then by Lemma 2 there exists a  $\mathcal{P}$ -CT  $\mathfrak{Z} = (V, E, L)$  that converges with probability  $< 1$  whose root is labeled with  $(1 : t)$  and  $\text{Pos}_{D^{\#}}(t) = \{\varepsilon\}$ . We will now create a  $\mathcal{T}_{\text{UT}}(\mathcal{P})$ -CT  $\mathfrak{Z}' = (V, E, L')$ , with the same underlying tree structure, and an adjusted labeling such that  $p_v^{\mathfrak{Z}'} = p_v^{\mathfrak{Z}}$  for all  $v \in V$ . Hence, we get  $|\mathfrak{Z}'| = |\mathfrak{Z}| < 1$ , and thus  $\mathcal{T}_{\text{UT}}(\mathcal{P})$  is not AST either.

We construct the new labeling  $L'$  for the  $\mathcal{T}_{\text{UT}}(\mathcal{P})$ -CT  $\mathfrak{Z}'$  recursively. Let  $W \subseteq V$  be the set of nodes where we have already defined the labeling  $L'$ . During our construction, we ensure that for every node  $v \in W$  we have

$$t_v \stackrel{b}{=} t'_v \text{ and } \text{Pos}_{D^{\#}}(t_v) \setminus \text{Junk}(t_v) \subseteq \text{Pos}_{D^{\#}}(t'_v). \quad (\text{B.7})$$

Here, for any term  $t_v$ , let  $\text{Junk}(t_v)$  be the set of positions that can never be used for a rewrite step with an ADP that contains annotations. To be precise, we define  $\pi \in \text{Junk}(t_v) \Leftrightarrow$  there is no ADP  $\ell \rightarrow \{p_1 : r_1, \dots, p_k : r_k\}^m \in \mathcal{P}$  with annotations and substitution  $\sigma$  such that  $\#_{\varepsilon}(t_v |_{\pi}) \xrightarrow{\text{c}_{\text{np}(\mathcal{P})}} \ell^{\#}\sigma$ .

We start with the same term  $t$  at the root. Here, our property (B.7) is clearly satisfied. As long as there is still an inner node  $v \in W$  such that its successors are not contained in  $W$ , we do the following. Let  $vE = \{w_1, \dots, w_k\}$  be the set of its successors. We need to define the terms for the nodes  $w_1, \dots, w_k$  in  $\mathfrak{Z}'$ . Since  $v$  is not a leaf and  $\mathfrak{Z}$  is a  $\mathcal{P}$ -CT, we have  $t_v \hookrightarrow_{\mathcal{P}} \{\frac{p_{w_1}}{p_v} : t_{w_1}, \dots, \frac{p_{w_k}}{p_v} : t_{w_k}\}$ .

If we performed a step with  $\hookrightarrow_p$  using the ADP  $\ell \rightarrow \{p_1 : r_1, \dots, p_k : r_k\}^m$ , the VRF  $(\varphi_j)_{1 \leq j \leq k}$ , the position  $\pi$ , and the substitution  $\sigma$  in  $\mathfrak{Z}$ , then we can use the ADP  $\ell \rightarrow \{p_1 : \#_{\Delta_P(r_1)}(r_1), \dots, p_k : \#_{\Delta_P(r_k)}(r_k)\}^m$  with the same VRF  $(\varphi_j)_{1 \leq j \leq k}$ , the same position  $\pi$ , and the same substitution  $\sigma$ . Now, we directly get  $t_{w_j} = t'_{w_j}$  for all  $1 \leq j \leq k$ . To prove  $\text{Pos}_{D^\#}(t_{w_j}) \setminus \text{Junk}(t_{w_j}) \subseteq \text{Pos}_{D^\#}(t'_{w_j})$ , note that if  $\pi \in \text{Pos}_{D^\#}(t_v) \cap \text{Junk}(t_v)$ , then  $\ell \rightarrow \{p_1 : r_1, \dots, p_k : r_k\}^m$  contains no annotations by definition of  $\text{Junk}(t_v)$ . Therefore, it does not matter whether we rewrite with case **(at)** or **(nt)** (**(af)** or **(nf)**). Otherwise, if  $\pi \in \text{Pos}_{D^\#}(t_v) \setminus \text{Junk}(t_v)$ , then the original rule contains the same terms with possibly more annotations, but all missing annotations are in  $\text{Junk}(t_v)$  by definition of  $\#_{\Delta_P(r_j)}(r_j)$ . Thus, we get  $\text{Pos}_{D^\#}(t_{w_j}) \setminus \text{Junk}(t_{w_j}) \subseteq \text{Pos}_{D^\#}(t'_{w_j})$  for all  $1 \leq j \leq k$ .

Finally, in order to see that  $\mathfrak{Z}'$  is indeed a  $\mathcal{T}_{\text{UT}}(\mathcal{P})$ -CT, we have to make sure that on every infinite path there are still an infinite number of rewrite steps with Case **(at)** or **(af)**. Assume otherwise, i.e., that there exists an infinite path with no such rewrite steps. The same path exists in  $\mathfrak{Z}$ , which is a  $\mathcal{P}$ -CT, so there are infinitely many rewrite steps with Case **(at)** or **(af)** in  $\mathfrak{Z}$ . This means that every rewrite step with Case **(at)** or **(af)** at a node  $v$  in this path is performed at a position in  $\text{Junk}(t_{w_j})$ . But since such rewrite steps decrease the number of annotations in the term, we cannot have infinitely many of them without rewriting at a position  $\text{Pos}_{D^\#}(t_v) \setminus \text{Junk}(t_v)$ , which is a contradiction.  $\square$

**Proof.** (Proof of Theorem 10)

Completeness: By Theorem 6.

Soundness: We do the same construction as for AST but the definition of  $\text{Junk}(t_v)$  changes according to innermost rewriting. Here, we define  $\pi \in \text{Junk}(t_v) \Leftrightarrow$  there is no ADP  $\ell \rightarrow \{p_1 : r_1, \dots, p_k : r_k\}^m \in \mathcal{P}$  with annotations and no substitution  $\sigma$  such that  $\#_{\varepsilon}(t_v |_{\pi}) \xrightarrow[\text{np}(\mathcal{P})]{\text{I}^*} \ell^\# \sigma$  and  $\ell^\# \sigma \in \text{ANF}_{\mathcal{P}}$ .  $\square$

**Proof.** (Proof of Theorem 11) Let  $\overline{\mathcal{P}} = \mathcal{U}(\mathcal{P}) \cup \{\ell \rightarrow \mu^{\text{false}} \mid \ell \rightarrow \mu^m \in \mathcal{P} \setminus \mathcal{U}(\mathcal{P})\}$ .

Completeness: By Theorem 6.

Soundness: Assume that  $\mathcal{P}$  is not iAST. Then by Lemma 2 there exists a  $\mathcal{P}$ -CT  $\mathfrak{Z} = (V, E, L)$  that converges with probability  $< 1$  whose root is labeled with  $(1 : t)$  and  $\flat(t) = s\theta \in \text{ANF}_{\mathcal{P}}$  for a substitution  $\theta$  and an ADP  $s \rightarrow \dots \in \mathcal{P}$ , and  $\text{Pos}_{D^\#}(t) = \{\varepsilon\}$ . By the definition of usable rules, as in the non-probabilistic case, rules  $\ell \rightarrow \mu \in \mathcal{P}$  that are not usable (i.e.,  $\ell \rightarrow \mu \notin \overline{\mathcal{P}}$ ) will never be used below an annotated symbol in such a  $\mathcal{P}$ -CT. Hence, we can also view  $\mathfrak{Z}$  as a  $\overline{\mathcal{P}}$ -CT that converges with probability  $< 1$  and thus  $\overline{\mathcal{P}}$  is not iAST.  $\square$

The proof for the reduction pair processor is split into two parts. First, we consider the technical part of the proof, and then we conclude the soundness and completeness of the processor. We use a slightly more complicated lemma here instead of Lemma 3 presented in Section 4. While Lemma 3 imposed conditions on the rewrite relation  $\hookrightarrow_p$ , we only impose conditions on the ADPs and lift them to  $\hookrightarrow_p$ -steps in the proof of the following lemma. Moreover, in Lemma 3, we considered the case where one term in the support decreases strictly in every rewrite step. Instead, we now want to regard CTs where such “strictly decreasing” ADPs are used infinitely often in each infinite path of the CT.

Again, let  $\mathcal{A}_{\text{sum}}^\#(r) = \sum_{t \leq_\# r} \mathcal{A}(t^\#)$ . Moreover, we use the *prefix order* ( $\pi \leq \tau \Leftrightarrow$  there exists  $\chi \in \mathbb{N}^*$  such that  $\pi \cdot \chi = \tau$ ) to compare positions.

**Lemma 5** (Proving AST and iAST on CTs with Algebras). *Let  $(\mathcal{A}, >)$  be a weakly monotonic,  $\mathbb{N}$ -collapsible, barycentric  $\Sigma^\#$ -algebra. Let  $\mathcal{P} = \mathcal{P}_{\geq} \uplus \mathcal{P}_{>}$  with  $\mathcal{P}_{>} \neq \emptyset$  where:*

- |     |   |   |  |
|-----|---|---|--|
| (1) | $\forall \ell \rightarrow \mu^m \in \mathcal{P}$  | : | $\mathcal{A}_{\text{sum}}^\#(\ell^\#) \geq \mathbb{E}(\mathcal{A}_{\text{sum}}^\#(\mu))$ |
| (2) | (a) $\forall \ell \rightarrow \mu^m \in \mathcal{P}_{>} : \exists r \in \text{Supp}(\mu)$ | : | $\mathcal{A}_{\text{sum}}^\#(\ell^\#) > \mathcal{A}_{\text{sum}}^\#(r)$                  |
|     | (b) If $m = \text{true}$ , then we additionally have                                      | : | $\mathcal{A}(\ell) \geq \mathcal{A}(\flat(r))$   |
| (3) | $\forall \ell \rightarrow \mu^{\text{true}} \in \mathcal{P}$                              | : | $\mathcal{A}(\ell) \geq \mathbb{E}(\mathcal{A}(\flat(\mu)))$                             |

Let  $\mathfrak{Z} = (V, E, L)$  be a  $\mathcal{P}$ -CT. We can partition  $\Gamma = \Gamma_{\geq} \uplus \Gamma_{>}$  according to  $\mathcal{P} = \mathcal{P}_{\geq} \uplus \mathcal{P}_{>}$ . If

(+) every infinite path in  $\mathfrak{Z}$  has an infinite number of  $\Gamma_{>}$  nodes,

then  $|\mathfrak{Z}| = 1$ .

**Proof.** This proof uses the proof idea for AST from [23]. For a  $\mathcal{P}$ -CT  $\mathfrak{Z}$  that satisfies (+), the core steps of the proof are the following:

- I. We extend the conditions (1), (2), and (3) to rewrite steps instead of just rules (and thus, to edges of a CT).
- II. We create a CT  $\mathfrak{Z}^{\leq N}$  for any  $N \in \mathbb{N}$ .
- III. We prove that  $|\mathfrak{Z}^{\leq N}| \geq p_{\text{min}}^N$  for any  $N \in \mathbb{N}$ .
- IV. We prove that  $|\mathfrak{Z}^{\leq N}| = 1$  for any  $N \in \mathbb{N}$ .
- V. Finally, we prove that  $|\mathfrak{Z}| = 1$ .

Here, let  $p_{\text{min}} > 0$  be the minimal probability that occurs in the ADPs from  $\mathcal{P}$ . As  $\mathcal{P}$  consists of only finitely many ADPs and all occurring multi-distributions are finite as well, this minimum is well defined. We have  $p_{\text{min}}^N = \underbrace{p_{\text{min}} \cdot \dots \cdot p_{\text{min}}}_{N \text{ times}}$ .

**I. We extend the conditions to rewrite steps instead of just rules:** We show that the conditions (1), (2), and (3) of the lemma extend to rewrite steps instead of just rules:

- (a) If  $s \rightarrow \{p_1 : t_1, \dots, p_k : t_k\}$  using a rewrite rule  $\ell \rightarrow \{p_1 : r_1, \dots, p_k : r_k\}$  with  $\mathcal{A}(\ell) \geq \mathcal{A}(r_j)$  for some  $1 \leq j \leq k$ , then we have  $\mathcal{A}(s) \geq \mathcal{A}(t_j)$ .
- (b) If  $a \hookrightarrow_{\mathcal{P}} \{p_1 : b_1, \dots, p_k : b_k\}$  using  $\ell \rightarrow \{p_1 : r_1, \dots, p_k : r_k\}^m \in \mathcal{P}_{>}$  at a position  $\pi \in \text{Pos}_{D^{\#}}(a)$ , then  $\mathcal{A}_{sum}^{\#}(a) > \mathcal{A}_{sum}^{\#}(b_j)$  for some  $1 \leq j \leq k$ .
- (c) If  $s \rightarrow \{p_1 : t_1, \dots, p_k : t_k\}$  using a rewrite rule  $\ell \rightarrow \{p_1 : r_1, \dots, p_k : r_k\}$  with  $\mathcal{A}(\ell) \geq \sum_{1 \leq j \leq k} p_j \cdot \mathcal{A}(r_j)$ , then  $\mathcal{A}(s) \geq \sum_{1 \leq j \leq k} p_j \cdot \mathcal{A}(t_j)$ .
- (d) If  $a \hookrightarrow_{\mathcal{P}} \{p_1 : b_1, \dots, p_k : b_k\}$  using  $\ell \rightarrow \{p_1 : r_1, \dots, p_k : r_k\}^m \in \mathcal{P}$ , then  $\mathcal{A}_{sum}^{\#}(a) \geq \sum_{1 \leq j \leq k} p_j \cdot \mathcal{A}_{sum}^{\#}(b_j)$ .

**Proofs:**

- (a) In this case, there exist a rule  $\ell \rightarrow \{p_1 : r_1, \dots, p_k : r_k\}$  with  $\mathcal{A}(\ell) \geq \mathcal{A}(r_j)$  for some  $1 \leq j \leq k$ , a substitution  $\sigma$ , and a position  $\pi$  of  $s$  such that  $s|_{\pi} = \ell \sigma$  and  $t_h = s[r_h \sigma]_{\pi}$  for all  $1 \leq h \leq k$ .

We perform structural induction on  $\pi$ . So in the induction base, let  $\pi = \varepsilon$ . Hence, we have  $s = \ell \sigma \rightarrow \{p_1 : r_1 \sigma, \dots, p_k : r_k \sigma\}$ . By assumption, we have  $\mathcal{A}(\ell) \geq \mathcal{A}(r_j)$  for some  $1 \leq j \leq k$ . As these inequations hold for all instantiations of the occurring variables, for  $t_j = r_j \sigma$  we have

$$\mathcal{A}(s) = \mathcal{A}(\ell \sigma) \geq \mathcal{A}(r_j \sigma) = \mathcal{A}(t_j).$$

In the induction step, we have  $\pi = i.\pi'$ ,  $s = f(s_1, \dots, s_i, \dots, s_n)$ ,  $f \in \Sigma^{\#}$ ,  $s_i \rightarrow \{p_1 : t_{i,1}, \dots, p_k : t_{i,k}\}$ , and  $t_j = f(s_1, \dots, t_{i,j}, \dots, s_n)$  with  $t_{i,j} = s_i[r_j \sigma]_{\pi'}$  for all  $1 \leq j \leq k$ . Then by the induction hypothesis we have  $\mathcal{A}(s_i) \geq \mathcal{A}(t_{i,j})$ . For  $t_j = f(s_1, \dots, t_{i,j}, \dots, s_n)$  we obtain

$$\begin{aligned} \mathcal{A}(s) &= \mathcal{A}(f(s_1, \dots, s_i, \dots, s_n)) \\ &= f_{\mathcal{A}}(\mathcal{A}(s_1), \dots, \mathcal{A}(s_i), \dots, \mathcal{A}(s_n)) \\ &\geq f_{\mathcal{A}}(\mathcal{A}(s_1), \dots, \mathcal{A}(t_{i,j}), \dots, \mathcal{A}(s_n)) \\ &\quad \text{(by weak monotonicity of } f_{\mathcal{A}} \text{ and } \mathcal{A}(s_i) \geq \mathcal{A}(t_{i,j})) \\ &= \mathcal{A}(f(s_1, \dots, t_{i,j}, \dots, s_n)) \\ &= \mathcal{A}(t_j). \end{aligned}$$

- (b) In this case, there exist an ADP  $\ell \rightarrow \{p_1 : r_1, \dots, p_k : r_k\}^m \in \mathcal{P}_{>}$ , a VRF  $(\varphi_j)_{1 \leq j \leq k}$ , a substitution  $\sigma$ , and position  $\pi \in \text{Pos}_{D^{\#}}(a)$  with  $b(a|_{\pi}) = \ell \sigma$  and  $b_j = a[r_j \sigma]_{\pi}$ . First, assume that  $m = \text{true}$ . Let  $I_1 = \{\tau \in \text{Pos}_{D^{\#}}(a) \mid \tau < \pi\}$  be the set of positions of all annotations strictly above  $\pi$ ,  $I_2 = \{\tau \in \text{Pos}_{D^{\#}}(a) \mid \gamma \in \text{Pos}_{\mathcal{V}}(\ell), \pi < \tau \leq \pi.\gamma\}$  be the set of positions of all annotations inside the left-hand side  $\ell$  of the used redex  $\ell \sigma$  (but not on the root of  $\ell$ ),  $I_3 = \{\tau \in \text{Pos}_{D^{\#}}(a) \mid \gamma \in \text{Pos}_{\mathcal{V}}(\ell), \pi.\gamma < \tau\}$  be the set of positions of all annotations inside the substitution, and let  $I_4 = \{\tau \in \text{Pos}_{D^{\#}}(a) \mid \tau \perp \pi\}$  be the set of positions of all annotations orthogonal to  $\pi$ . Furthermore, for each  $\tau \in I_1$  let  $\kappa_{\tau}$  be the position such that  $\tau.\kappa_{\tau} = \pi$ , and for each  $\tau \in I_3$  let  $\gamma_{\tau}$  and  $\rho_{\tau}$  be the positions such that  $\gamma_{\tau} \in \text{Pos}_{\mathcal{V}}(\ell)$  and  $\pi.\gamma_{\tau}.\rho_{\tau} = \tau$ . By Requirement (2), there exists a  $1 \leq j \leq k$  with  $\mathcal{A}(\ell^{\#}) = \mathcal{A}_{sum}^{\#}(\ell^{\#}) > \mathcal{A}_{sum}^{\#}(r_j) = \sum_{t \leq_{\#} r_j} \mathcal{A}(t^{\#})$  and, additionally,  $\mathcal{A}(\ell) \geq \mathcal{A}(b(r_j))$  since  $m = \text{true}$ . As these inequations hold for all instantiations of the occurring variables, we have

$$\begin{aligned} &\mathcal{A}_{sum}^{\#}(a) \\ &= \sum_{t \leq_{\#} a} \mathcal{A}(t^{\#}) \\ &= \mathcal{A}(\#_{\varepsilon}(a|_{\pi})) + \sum_{\tau \in I_1} \mathcal{A}(\#_{\varepsilon}(a|_{\tau})) + \sum_{\tau \in I_2} \mathcal{A}(\#_{\varepsilon}(a|_{\tau})) + \sum_{\tau \in I_3} \mathcal{A}(\#_{\varepsilon}(a|_{\tau})) + \sum_{\tau \in I_4} \mathcal{A}(\#_{\varepsilon}(a|_{\tau})) \\ &\geq \mathcal{A}(\#_{\varepsilon}(a|_{\pi})) + \sum_{\tau \in I_1} \mathcal{A}(\#_{\varepsilon}(a|_{\tau})) + \sum_{\tau \in I_3} \mathcal{A}(\#_{\varepsilon}(a|_{\tau})) + \sum_{\tau \in I_4} \mathcal{A}(\#_{\varepsilon}(a|_{\tau})) \\ &\quad \text{(removed } I_2) \\ &= \mathcal{A}(\#_{\varepsilon}(\ell)\sigma) + \sum_{\tau \in I_1} \mathcal{A}(\#_{\varepsilon}(a|_{\tau})) + \sum_{\tau \in I_3} \mathcal{A}(\#_{\varepsilon}(a|_{\tau})) + \sum_{\tau \in I_4} \mathcal{A}(\#_{\varepsilon}(a|_{\tau})) \\ &\quad \text{(as } \#_{\varepsilon}(a|_{\pi}) = \#_{\varepsilon}(\ell)\sigma) \\ &> \sum_{t \leq_{\#} r_j} \mathcal{A}(\#_{\varepsilon}(t)\sigma) + \sum_{\tau \in I_1} \mathcal{A}(\#_{\varepsilon}(a|_{\tau})) + \sum_{\tau \in I_3} \mathcal{A}(\#_{\varepsilon}(a|_{\tau})) + \sum_{\tau \in I_4} \mathcal{A}(\#_{\varepsilon}(a|_{\tau})) \\ &\quad \text{(as } \mathcal{A}(\#_{\varepsilon}(\ell)) > \sum_{t \leq_{\#} r_j} \mathcal{A}(\#_{\varepsilon}(t)), \text{ hence } \mathcal{A}(\#_{\varepsilon}(\ell)\sigma) > \sum_{t \leq_{\#} r_j} \mathcal{A}(\#_{\varepsilon}(t)\sigma)) \\ &\geq \sum_{t \leq_{\#} r_j \sigma} \mathcal{A}(\#_{\varepsilon}(t)) + \sum_{\tau \in I_1} \mathcal{A}(\#_{\varepsilon}(a|_{\tau}[r_j \sigma]_{\kappa_{\tau}})) + \sum_{\tau \in I_3} \mathcal{A}(\#_{\varepsilon}(a|_{\tau})) + \sum_{\tau \in I_4} \mathcal{A}(\#_{\varepsilon}(a|_{\tau})) \\ &\quad \text{(by } \mathcal{A}(\ell) \geq \mathcal{A}(r_j) \text{ and (a))} \\ &\geq \sum_{t \leq_{\#} r_j \sigma} \mathcal{A}(\#_{\varepsilon}(t)) + \sum_{\tau \in I_1} \mathcal{A}(\#_{\varepsilon}(a|_{\tau}[r_j \sigma]_{\kappa_{\tau}})) + \sum_{\tau \in I_3, \varphi_j(\gamma_{\tau}, \rho_{\tau}) \neq \perp} \mathcal{A}(\#_{\varepsilon}(b_j|_{\pi.\varphi_j(\gamma_{\tau}, \rho_{\tau})})) + \sum_{\tau \in I_4} \mathcal{A}(\#_{\varepsilon}(a|_{\tau})) \\ &\quad \text{(moving } \tau = \pi.\gamma_{\tau}.\rho_{\tau} \in I_3 \text{ via the VRF)} \end{aligned}$$

$$\begin{aligned}
&= \sum_{t \sqsubseteq_{\#}^b b_j} \mathcal{A}(t^{\#}) \\
&= \mathcal{A}_{sum}^{\#}(b_j)
\end{aligned}$$

In the case  $m = \text{false}$ , we additionally remove  $\sum_{\tau \in I_1} \mathcal{A}(\#_{\epsilon}(a|_{\tau}[r_j\sigma]_{\kappa_{\tau}}))$ , and the inequation remains correct.

- (c) In this case, there exist a rule  $\ell \rightarrow \{p_1 : r_1, \dots, p_k : r_k\}$  with  $\mathcal{A}(\ell) \geq \sum_{1 \leq j \leq k} p_j \cdot \mathcal{A}(r_j)$ , a substitution  $\sigma$ , and a position  $\pi$  of  $s$  such that  $s|_{\pi} = \ell\sigma$ , and  $t_j = s[r_j\sigma]_{\pi}$  for all  $1 \leq j \leq k$ .

We perform structural induction on  $\pi$ . So in the induction base  $\pi = \epsilon$  we have  $s = \ell\sigma \rightarrow \{p_1 : r_1\sigma, \dots, p_k : r_k\sigma\}$ . As  $\mathcal{A}(\ell) \geq \sum_{1 \leq j \leq k} p_j \cdot \mathcal{A}(r_j)$  holds for all instantiations of the occurring variables, for  $t_j = r_j\sigma$  we obtain

$$\mathcal{A}(s) = \mathcal{A}(\ell\sigma) \geq \sum_{1 \leq j \leq k} p_j \cdot \mathcal{A}(r_j\sigma) = \sum_{1 \leq j \leq k} p_j \cdot \mathcal{A}(t_j).$$

In the induction step, we have  $\pi = i.\pi'$ ,  $s = f(s_1, \dots, s_i, \dots, s_n)$ ,  $s_i \rightarrow \{p_1 : t_{i,1}, \dots, p_k : t_{i,k}\}$ , and  $t_j = f(s_1, \dots, t_{i,j}, \dots, s_n)$  with  $t_{i,j} = s_i[r_j\sigma]_{\pi'}$  for all  $1 \leq j \leq k$ . Then by the induction hypothesis we have  $\mathcal{A}(s_i) \geq \sum_{1 \leq j \leq k} p_j \cdot \mathcal{A}(t_{i,j})$ . Thus, we have

$$\begin{aligned}
\mathcal{A}(s) &= \mathcal{A}(f(s_1, \dots, s_i, \dots, s_n)) \\
&= f_{\mathcal{A}}(\mathcal{A}(s_1), \dots, \mathcal{A}(s_i), \dots, \mathcal{A}(s_n)) \\
&\geq f_{\mathcal{A}}(\mathcal{A}(s_1), \dots, \sum_{1 \leq j \leq k} p_j \cdot \mathcal{A}(t_{i,j}), \dots, \mathcal{A}(s_n)) \quad (\text{by weak monotonicity of } f_{\mathcal{A}} \text{ and } \mathcal{A}(s_i) \geq \sum_{1 \leq j \leq k} p_j \cdot \mathcal{A}(t_{i,j})) \\
&\geq \sum_{1 \leq j \leq k} p_j \cdot f_{\mathcal{A}}(\mathcal{A}(s_1), \dots, \mathcal{A}(t_{i,j}), \dots, \mathcal{A}(s_n)) \quad (\text{as } f_{\mathcal{A}} \text{ is concave}) \\
&= \sum_{1 \leq j \leq k} p_j \cdot \mathcal{A}(f(s_1, \dots, t_{i,j}, \dots, s_n)) \\
&= \sum_{1 \leq j \leq k} p_j \cdot \mathcal{A}(t_j).
\end{aligned}$$

- (d) In this case, there exist an ADP  $\ell \rightarrow \{p_1 : r_1, \dots, p_k : r_k\}^m \in \mathcal{P}$ , a substitution  $\sigma$ , and position  $\pi$  with  $b(a|_{\pi}) = \ell\sigma$  and  $b_j = a[r_j\sigma]_{\pi}$ . First, assume that  $m = \text{true}$  and  $\pi \in \text{Pos}_{D^{\#}}(a)$ . Let  $I_1 = \{\tau \in \text{Pos}_{D^{\#}}(a) \mid \tau < \pi\}$  be the set of positions of all annotations strictly above  $\pi$ ,  $I_2 = \{\tau \in \text{Pos}_{D^{\#}}(a) \mid \gamma \in \text{Pos}_{\mathcal{V}}(\ell), \pi < \tau \leq \pi.\gamma\}$  be the set of positions of all annotations inside the left-hand side  $\ell$  of the used redex  $\ell\sigma$  (but not on the root of  $\ell$ ),  $I_3 = \{\tau \in \text{Pos}_{D^{\#}}(a) \mid \gamma \in \text{Pos}_{\mathcal{V}}(\ell), \pi.\gamma < \tau\}$  be the set of positions of all annotations inside the substitution, and let  $I_4 = \{\tau \in \text{Pos}_{D^{\#}}(a) \mid \tau \perp \pi\}$  be the set of positions of all annotations orthogonal to  $\pi$ . Furthermore, for each  $\tau \in I_1$  let  $\kappa_{\tau}$  be the position such that  $\tau.\kappa_{\tau} = \pi$ , and for each  $\tau \in I_3$  let  $\gamma_{\tau}$  and  $\rho_{\tau}$  be the positions such that  $\gamma_{\tau} \in \text{Pos}_{\mathcal{V}}(\ell)$  and  $\pi.\gamma_{\tau}.\rho_{\tau} = \tau$ . By Requirement (1), we have  $\mathcal{A}(\#_{\epsilon}(\ell)) = \mathcal{A}_{sum}^{\#}(\#_{\epsilon}(\ell)) \geq \sum_{1 \leq j \leq k} p_j \cdot \sum_{t \sqsubseteq_{\#}^b r_j} \mathcal{A}(\#_{\epsilon}(t))$  and by (3) we have  $\mathcal{A}(\ell) \geq \sum_{1 \leq j \leq k} p_j \cdot \mathcal{A}(b(r_j))$ . As these inequations hold for all instantiations of the occurring variables, we have

$$\begin{aligned}
&\mathcal{A}_{sum}^{\#}(a) \\
&= \sum_{t \sqsubseteq_{\#}^b a} \mathcal{A}(t^{\#}) \\
&= \mathcal{A}(\#_{\epsilon}(a|_{\pi})) + \sum_{\tau \in I_1} \mathcal{A}(\#_{\epsilon}(a|_{\tau})) + \sum_{\tau \in I_2} \mathcal{A}(\#_{\epsilon}(a|_{\tau})) + \sum_{\tau \in I_3} \mathcal{A}(\#_{\epsilon}(a|_{\tau})) + \sum_{\tau \in I_4} \mathcal{A}(\#_{\epsilon}(a|_{\tau})) \\
&\geq \mathcal{A}(\#_{\epsilon}(a|_{\pi})) + \sum_{\tau \in I_1} \mathcal{A}(\#_{\epsilon}(a|_{\tau})) + \sum_{\tau \in I_3} \mathcal{A}(\#_{\epsilon}(a|_{\tau})) + \sum_{\tau \in I_4} \mathcal{A}(\#_{\epsilon}(a|_{\tau})) \\
&\quad (\text{removed } I_2) \\
&= \mathcal{A}(\#_{\epsilon}(\ell)\sigma) + \sum_{\tau \in I_1} \mathcal{A}(\#_{\epsilon}(a|_{\tau})) + \sum_{\tau \in I_3} \mathcal{A}(\#_{\epsilon}(a|_{\tau})) + \sum_{\tau \in I_4} \mathcal{A}(\#_{\epsilon}(a|_{\tau})) \\
&\quad (\text{as } a|_{\pi} = \#_{\epsilon}(\ell)\sigma) \\
&\geq \sum_{1 \leq j \leq k} p_j \cdot \sum_{t \sqsubseteq_{\#}^b r_j\sigma} \mathcal{A}(\#_{\epsilon}(t)) + \sum_{\tau \in I_1} \mathcal{A}(\#_{\epsilon}(a|_{\tau})) + \sum_{\tau \in I_3} \mathcal{A}(\#_{\epsilon}(a|_{\tau})) + \sum_{\tau \in I_4} \mathcal{A}(\#_{\epsilon}(a|_{\tau})) \\
&\quad (\text{by } \mathcal{A}(\#_{\epsilon}(\ell)) \geq \sum_{1 \leq j \leq k} p_j \cdot \sum_{t \sqsubseteq_{\#}^b r_j} \mathcal{A}(\#_{\epsilon}(t)), \text{ hence } \mathcal{A}(\#_{\epsilon}(\ell)\sigma) \geq \sum_{1 \leq j \leq k} p_j \cdot \sum_{t \sqsubseteq_{\#}^b r_j\sigma} \mathcal{A}(\#_{\epsilon}(t)))
\end{aligned}$$

$$\begin{aligned}
&\geq \sum_{1 \leq j \leq k} p_j \cdot \sum_{t \sqsubseteq_{\#} r_j \sigma} \mathcal{A}(\#_{\epsilon}(t)) + \sum_{\tau \in I_1} \sum_{1 \leq j \leq k} p_j \cdot \mathcal{A}(\#_{\epsilon}(a|_{\tau}[r_j \sigma]_{\kappa_{\tau}})) + \sum_{\tau \in I_3} \mathcal{A}(\#_{\epsilon}(a|_{\tau})) + \sum_{\tau \in I_4} \mathcal{A}(\#_{\epsilon}(a|_{\tau})) \\
&\quad \text{(by } \mathcal{A}(\ell) \geq \sum_{1 \leq j \leq k} p_j \cdot \mathcal{A}(r_j) \text{ and (c))} \\
&= \sum_{1 \leq j \leq k} p_j \cdot \sum_{t \sqsubseteq_{\#} r_j \sigma} \mathcal{A}(\#_{\epsilon}(t)) + \sum_{1 \leq j \leq k} \sum_{\tau \in I_1} p_j \cdot \mathcal{A}(\#_{\epsilon}(a|_{\tau}[r_j \sigma]_{\kappa_{\tau}})) + \sum_{\tau \in I_3} \mathcal{A}(\#_{\epsilon}(a|_{\tau})) + \sum_{\tau \in I_4} \mathcal{A}(\#_{\epsilon}(a|_{\tau})) \\
&= \sum_{1 \leq j \leq k} p_j \cdot \sum_{t \sqsubseteq_{\#} r_j \sigma} \mathcal{A}(\#_{\epsilon}(t)) + \sum_{1 \leq j \leq k} p_j \cdot \sum_{\tau \in I_1} \mathcal{A}(\#_{\epsilon}(a|_{\tau}[r_j \sigma]_{\kappa_{\tau}})) + \sum_{\tau \in I_3} \mathcal{A}(\#_{\epsilon}(a|_{\tau})) + \sum_{\tau \in I_4} \mathcal{A}(\#_{\epsilon}(a|_{\tau})) \\
&= \sum_{1 \leq j \leq k} p_j \cdot \sum_{t \sqsubseteq_{\#} r_j \sigma} \mathcal{A}(\#_{\epsilon}(t)) + \sum_{1 \leq j \leq k} p_j \cdot \sum_{\tau \in I_1} \mathcal{A}(\#_{\epsilon}(a|_{\tau}[r_j \sigma]_{\kappa_{\tau}})) + \sum_{1 \leq j \leq k} p_j \cdot \sum_{\tau \in I_3} \mathcal{A}(\#_{\epsilon}(a|_{\tau})) + \sum_{1 \leq j \leq k} p_j \cdot \sum_{\tau \in I_4} \mathcal{A}(\#_{\epsilon}(a|_{\tau})) \\
&= \sum_{1 \leq j \leq k} p_j \cdot \left( \sum_{t \sqsubseteq_{\#} r_j \sigma} \mathcal{A}(\#_{\epsilon}(t)) + \sum_{\tau \in I_1} \mathcal{A}(\#_{\epsilon}(a|_{\tau}[r_j \sigma]_{\kappa_{\tau}})) + \sum_{\tau \in I_3} \mathcal{A}(\#_{\epsilon}(a|_{\tau})) + \sum_{\tau \in I_4} \mathcal{A}(\#_{\epsilon}(a|_{\tau})) \right) \\
&\geq \sum_{1 \leq j \leq k} p_j \cdot \left( \sum_{t \sqsubseteq_{\#} r_j \sigma} \mathcal{A}(\#_{\epsilon}(t)) + \sum_{\tau \in I_1} \mathcal{A}(\#_{\epsilon}(a|_{\tau}[r_j \sigma]_{\kappa_{\tau}})) + \sum_{\tau \in I_3, \varphi_j(\gamma_{\tau}) \neq \perp} \mathcal{A}(\#_{\epsilon}(b_j|_{\pi \cdot \varphi_j(\gamma_{\tau}) \cdot \rho_{\tau}})) + \sum_{\tau \in I_4} \mathcal{A}(\#_{\epsilon}(a|_{\tau})) \right) \\
&\quad \text{(moving } \tau = \pi \cdot \gamma_{\tau} \cdot \rho_{\tau} \in I_3 \text{ via the VRF)} \\
&= \sum_{1 \leq j \leq k} p_j \cdot \sum_{t \sqsubseteq_{\#} b_j} \mathcal{A}(t^{\#}) \\
&= \mathcal{A}_{sum}^{\#}(b_j)
\end{aligned}$$

In the case  $\pi \notin \text{Pos}_{D^{\#}}(a)$ , we need to remove  $\mathcal{A}(\#_{\epsilon}(\ell)\sigma)$  as this annotated subterm does not exist in  $a$ , and therefore also  $\sum_{t \sqsubseteq_{\#} r_j \sigma} \mathcal{A}(t^{\#})$  in the end, leading to the same result. In the case  $m = \text{false}$ , we additionally remove  $\sum_{\tau \in I_1} \mathcal{A}(\#_{\epsilon}(a|_{\tau}[r_j \sigma]_{\kappa_{\tau}}))$  in the end.

**II. We create a CT  $\mathfrak{T}^{\leq N}$  for any  $N \in \mathbb{N}$ :** Let  $\mathfrak{T} = (V, E, L)$  be a  $\mathcal{P}$ -CT that satisfies (+). Since  $\mathcal{A}$  is  $\mathbb{N}$ -collapsible, there exists a concave embedding  $g_{\mathcal{A} \rightarrow \mathbb{N}} : \mathcal{A} \rightarrow \mathbb{N}$ . We define the *tree value* of any node  $v \in V$  in our CT by

$$\mathcal{Val}_{\mathfrak{T}} : V \rightarrow \mathbb{N}, \quad v \mapsto \begin{cases} 0, & \text{if } v \in \text{Leaf} \\ g_{\mathcal{A} \rightarrow \mathbb{N}}(\mathcal{A}_{sum}^{\#}(t_v \delta_a)) + 1, & \text{otherwise} \end{cases}$$

Remember that  $t_v$  denotes the term in the labeling of the node  $v$ . The substitution  $\delta_a$  maps every occurring variable to some constant  $a$ . By definition of  $>$ , we know that  $\mathcal{A}_{sum}^{\#}(t) > \mathcal{A}_{sum}^{\#}(s)$  implies  $\mathcal{A}_{sum}^{\#}(t \delta_a) > \mathcal{A}_{sum}^{\#}(s \delta_a)$ . The substitution ensures that we are only working with ground terms, so that  $\mathcal{Val}_{\mathfrak{T}}$  really maps each term to a natural number. Furthermore, note that we have  $\mathcal{Val}_{\mathfrak{T}}(v) = 0$  if and only if  $v$  is a leaf in  $\mathfrak{T}$ .

For any  $N \in \mathbb{N}$ , we create a modified tree  $\mathfrak{T}^{\leq N}$ , where we cut everything below a node  $v$  of the tree with  $\mathcal{Val}_{\mathfrak{T}}(v) \geq N + 1$ . Let  $C = \text{Leaf}^{\mathfrak{T}^{\leq N}} \setminus \text{Leaf}^{\mathfrak{T}}$  be the set of all new leaves in  $\mathfrak{T}^{\leq N}$  due to the cut. So for all  $v \in C$  we have  $\mathcal{Val}_{\mathfrak{T}}(v) \geq N + 1$ .

Our goal is to prove that we have  $|\mathfrak{T}| = 1$ . First, we prove that  $|\mathfrak{T}^{\leq N}| = 1$  for any  $N \in \mathbb{N}$ . However, this does not yet prove that  $|\mathfrak{T}| = 1$ , which we will show afterwards.

**III. We prove that  $|\mathfrak{T}^{\leq N}| \geq p_{min}^N$  for any  $N \in \mathbb{N}$ :** Note that in a rewrite step  $a \hookrightarrow_p \{p_1 : b_1, \dots, p_k : b_k\}$  with Case (nt) or (nf) or in a rewrite step with an ADP from  $\mathcal{P}_{\geq}$ , we cannot guarantee that there exists a  $b_j$  that is strictly decreasing in value. It is also possible that all values stay the same. Thus, there does not have to be a single witness path of length  $N$  in  $\mathfrak{T}^{\leq N}$  that shows termination with a probability of at least  $p_{min}^N$ . Instead, we have to use multiple witness paths of finite length to ensure that we have  $|\mathfrak{T}^{\leq N}| \geq p_{min}^N$ .<sup>15</sup> We first explain how to find such a set of witness paths in a finite sub  $\mathcal{P}$ -CT that starts with a node from  $\mathcal{P}_{>}$ , and then we prove by induction that we have  $|\mathfrak{T}^{\leq N}| \geq p_{min}^N$  in a second step.

**III.1. We find witnesses in subtrees of  $\mathfrak{T}^{\leq N}$ :** In this part, we prove a first observation regarding the existence of certain witnesses that shows a guaranteed decrease of values. Let  $\Gamma$  now refer to  $\mathfrak{T}^{\leq N}$ , i.e.,  $\Gamma$  only contains inner nodes of  $\mathfrak{T}^{\leq N}$  and thus,  $\Gamma \cap C = \emptyset$ . Moreover, let  $\Gamma_{>}$  and  $\Gamma_{\geq}$  correspond to the rewrite steps at nodes in  $\Gamma$  with  $\mathcal{P}_{>}$  and  $\mathcal{P}_{\geq}$ , respectively. For every  $v \in \Gamma_{>}$ , let  $\mathfrak{T}_v$  be the subtree that starts at  $v$  and where we cut every edge after the second node from  $\Gamma_{>}$ . This is illustrated in Fig. B.9, where an empty node stands for a leaf that already existed in  $\mathfrak{T}^{\leq N}$  (i.e., it is in  $\text{Leaf}^{\mathfrak{T}}$  or in  $C$ ). Since  $\mathfrak{T}$  satisfies (+) and is finitely branching, we know that  $\mathfrak{T}_v$  must

be finite. We want to prove that for such a tree  $\mathfrak{T}_v$ , we have a set of leaves (a set of witnesses) that show a certain decrease of values compared to the root value  $\mathcal{Val}_{\mathfrak{T}}(v)$ . To be precise, we want to prove that there exists a set  $W^v \subseteq \text{Leaf}^{\mathfrak{T}_v}$  of leaves in  $\mathfrak{T}_v$  with the following two properties:

(W-1) For all  $w \in W^v$  we have  $\overline{\mathcal{Val}_{\mathfrak{T}}}(w) \leq \overline{\mathcal{Val}_{\mathfrak{T}}}(v) - 1$ .

(W-2)  $\sum_{w \in W^v} p_w^{\mathfrak{T}_v} \cdot p_{min}^{\overline{\mathcal{Val}_{\mathfrak{T}}}(w)} \geq p_{min}^{\overline{\mathcal{Val}_{\mathfrak{T}}}(v)}$ .

<sup>15</sup>  $\mathbb{N}$ -collapsibility ensures that  $p_{min}^N > 0$  is a lower bound on the termination probability. If one wants to adapt this theorem to more complex algebras (that are not  $\mathbb{N}$ -collapsible), then one needs to find different ways to obtain a constant lower bound  $> 0$  for every  $N \in \mathbb{N}$ .

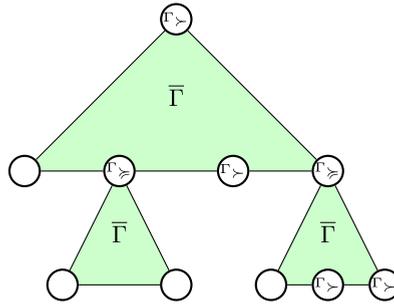


Fig. B.9. Subtree  $\mathfrak{T}_v$ .

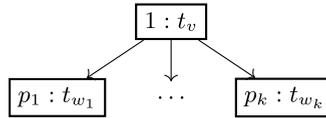


Fig. B.10. Nodes in the induction base.

Here, we use an *adjusted tree value function*  $\overline{\mathcal{V}al}_{\mathfrak{T}} : V^{\mathfrak{T} \leq N} \rightarrow \mathbb{N}$ , such that for every node  $v \in V^{\mathfrak{T} \leq N}$  we have

$$\overline{\mathcal{V}al}_{\mathfrak{T}}(v) = \begin{cases} 0, & \text{if } v \text{ is a leaf in } \mathfrak{T}^{\leq N} \\ \mathcal{V}al_{\mathfrak{T}}(v), & \text{otherwise} \end{cases}$$

This is the same tree value function as before, except for the nodes in  $C$ . For all  $v \in C$  we have  $\mathcal{V}al_{\mathfrak{T}}(v) \geq N + 1$  and  $\overline{\mathcal{V}al}_{\mathfrak{T}}(v) = 0$ . Again, we have  $p_{min}^n = \underbrace{p_{min} \cdot \dots \cdot p_{min}}_{n \text{ times}}$  for any natural number  $n \in \mathbb{N}$ , e.g.,  $\overline{\mathcal{V}al}_{\mathfrak{T}}(v) \in \mathbb{N}$ . The first property (W-1) states that all of our

witnesses have a strictly smaller value than the root. Furthermore, we have to be careful that the probabilities for our witnesses are not too low. The second property (W-2) states that the sum of all probabilities for the witnesses is still big enough. The additional  $p_{min}^{\overline{\mathcal{V}al}_{\mathfrak{T}}(w)}$  is used to allow smaller probabilities for our witnesses if the value decrease is high enough.

In order to show the existence of such a set  $W^v$ , we prove by induction on the height  $H$  of  $\mathfrak{T}_v$  that there exists a set of nodes  $W_i^v$  such that for all  $1 \leq i \leq H$  we have

(W-1!) For all  $w \in W_i^v$  we have  $\overline{\mathcal{V}al}_{\mathfrak{T}}(w) \leq \overline{\mathcal{V}al}_{\mathfrak{T}}(v) - 1$ .

(W-2!)  $\sum_{w \in W_i^v} p_w^{\mathfrak{T}_v} \cdot p_{min}^{\overline{\mathcal{V}al}_{\mathfrak{T}}(w)} \geq p_{min}^{\overline{\mathcal{V}al}_{\mathfrak{T}}(v)}$ .

(W-3!) Every  $w \in W_i^v$  is a leaf or an inner node at the  $i$ -th depth of  $\mathfrak{T}_v$ .

Then in the end if  $i = H$  is the height of the tree  $\mathfrak{T}_v$ , we get a set  $W^v = W_H^v$  such that every node in  $W_H^v$  is a leaf in  $\mathfrak{T}_v$  (i.e.,  $W_H^v \subseteq \text{Leaf}^{\mathfrak{T}_v}$ ) and both (W-1) and (W-2) are satisfied.

In the induction base, we consider depth  $i = 1$  and look at the rewrite step at the root, see Fig. B.10. The first edge represents a rewrite step with  $\mathcal{P}_{\succ}$  at a position  $\pi \in \text{Pos}_{D^\#}(t_v)$ . Let  $vE = \{w_1, \dots, w_k\}$  be the set of the successors of  $v$ . We have  $t_v \hookrightarrow_p \{p_1 : t_{w_1}, \dots, p_k : t_{w_k}\}$  using a rule from  $\mathcal{P}_{\succ}$ .

Due to (b) there is a  $1 \leq j \leq k$  with  $\mathcal{A}_{sum}^\#(t_v) > \mathcal{A}_{sum}^\#(t_{w_j})$ . This means  $g_{\mathcal{A} \rightarrow \mathbb{N}}(\mathcal{A}_{sum}^\#(t_v, \delta_a)) \geq g_{\mathcal{A} \rightarrow \mathbb{N}}(\mathcal{A}_{sum}^\#(t_{w_j}, \delta_a)) + 1$ , and hence  $\mathcal{V}al_{\mathfrak{T}}(v) \geq \mathcal{V}al_{\mathfrak{T}}(w_j) + 1$ , which also implies that  $\overline{\mathcal{V}al}_{\mathfrak{T}}(v) \geq \overline{\mathcal{V}al}_{\mathfrak{T}}(w_j) + 1$ . Since  $0 < p_{min} \leq 1$ , we therefore have  $p_{min}^{\overline{\mathcal{V}al}_{\mathfrak{T}}(v)} \leq p_{min}^{\overline{\mathcal{V}al}_{\mathfrak{T}}(w_j)+1}$ . Thus we can set  $W_0^v = \{w_j\}$  and have (W-1!) satisfied, since  $\overline{\mathcal{V}al}_{\mathfrak{T}}(v) \geq \overline{\mathcal{V}al}_{\mathfrak{T}}(w_j) + 1$  so  $\overline{\mathcal{V}al}_{\mathfrak{T}}(w_j) \leq \overline{\mathcal{V}al}_{\mathfrak{T}}(v) - 1$ . Property (W-3!) is clearly satisfied and (W-2!) holds as well since we have

$$\sum_{w \in W_1^v} p_w^{\mathfrak{T}_v} \cdot p_{min}^{\overline{\mathcal{V}al}_{\mathfrak{T}}(w)} = p_j \cdot p_{min}^{\overline{\mathcal{V}al}_{\mathfrak{T}}(w_j)} \begin{pmatrix} p_j \geq p_{min} \wedge \\ \overline{\mathcal{V}al}_{\mathfrak{T}}(w_j) \leq \overline{\mathcal{V}al}_{\mathfrak{T}}(v) - 1 \\ \geq \end{pmatrix} p_{min} \cdot p_{min}^{\overline{\mathcal{V}al}_{\mathfrak{T}}(v)-1} = p_{min}^{\overline{\mathcal{V}al}_{\mathfrak{T}}(v)}.$$

In the induction step, we consider depth  $i > 1$ , see Fig. B.11. Due to the induction hypothesis, there is a set  $W_{i-1}^v$  that satisfies (W-1!), (W-2!), and (W-3!). For every node  $w \in W_{i-1}^v$  that is not a leaf of  $\mathfrak{T}_v$ , we proceed as follows: Let  $wE = \{w_1, \dots, w_k\}$  be the set of its successors. We rewrite  $t_w$  either with  $\hookrightarrow_p$  and an ADP from  $\mathcal{P}_{\geq}$  or with  $\hookrightarrow_p$  and a position  $\pi \notin \text{Pos}_{D^\#}(t_w)$ , which rewrites  $t_w$  to a multi-distribution  $\{p_1 : t_{w_1}, \dots, p_k : t_{w_k}\}$ .

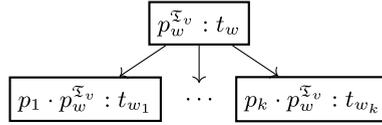


Fig. B.11. Induction step.

Due to (d), we have  $\mathcal{A}_{sum}^\#(t_w) \geq \sum_{1 \leq j \leq k} p_j \cdot \mathcal{A}_{sum}^\#(t_{w_j})$ . Since  $g_{\mathcal{A} \rightarrow \mathbb{N}}$  is a concave embedding we get

$$\begin{aligned} g_{\mathcal{A} \rightarrow \mathbb{N}}(\mathcal{A}_{sum}^\#(t_w \delta_a)) &\geq g_{\mathcal{A} \rightarrow \mathbb{N}}\left(\sum_{1 \leq j \leq k} p_j \cdot \mathcal{A}_{sum}^\#(t_{w_j} \delta_a)\right) \\ &\geq \sum_{1 \leq j \leq k} p_j \cdot g_{\mathcal{A} \rightarrow \mathbb{N}}(\mathcal{A}_{sum}^\#(t_{w_j} \delta_a)) \\ \iff \overline{\text{val}}_{\mathfrak{X}}(w) &\geq \sum_{1 \leq j \leq k} p_j \cdot \overline{\text{val}}_{\mathfrak{X}}(w_j) \end{aligned}$$

Hence, we either have  $\overline{\text{val}}_{\mathfrak{X}}(w) = \overline{\text{val}}_{\mathfrak{X}}(w_j)$  for all  $1 \leq j \leq k$  or there exists at least one  $1 \leq j \leq k$  with  $\overline{\text{val}}_{\mathfrak{X}}(w) > \overline{\text{val}}_{\mathfrak{X}}(w_j)$ .

We partition the set  $W_{i-1}^v$  into the disjoint subsets  $W_{i-1}^{v(1)}$ ,  $W_{i-1}^{v(2)}$ , and  $W_{i-1}^{v(3)}$ , where

- $w \in W_{i-1}^{v(1)} \iff w \in W_{i-1}^v$  and  $w$  is a leaf of  $\mathfrak{X}_v$ .
- $w \in W_{i-1}^{v(2)} \iff w \in W_{i-1}^v$  and  $\overline{\text{val}}_{\mathfrak{X}}(w) = \overline{\text{val}}_{\mathfrak{X}}(w_j)$  for all  $1 \leq j \leq k$ .
- $w \in W_{i-1}^{v(3)} \iff w \in W_{i-1}^v$  and there exists a  $1 \leq j \leq k$  with  $\overline{\text{val}}_{\mathfrak{X}}(w) > \overline{\text{val}}_{\mathfrak{X}}(w_j)$ . We denote this node  $w_j$  by  $w^+$ .

In the first case,  $w$  is already contained in  $W_{i-1}^v$ . In the second case, we get  $\overline{\text{val}}_{\mathfrak{X}}(w) = \overline{\text{val}}_{\mathfrak{X}}(w_j)$  for all  $1 \leq j \leq k$ . And in the third case, we have  $\overline{\text{val}}_{\mathfrak{X}}(w) > \overline{\text{val}}_{\mathfrak{X}}(w^+)$ , hence  $\overline{\text{val}}_{\mathfrak{X}}(w^+) \leq \overline{\text{val}}_{\mathfrak{X}}(w) - 1$ . So for all of these nodes, we can be sure that the value does not increase. We can now define  $W_i^v$  as:

$$\begin{aligned} W_i^v &= W_{i-1}^{v(1)} \\ &\cup \bigcup_{w \in W_{i-1}^{v(2)}} wE \\ &\cup \{w^+ \mid w \in W_{i-1}^{v(3)}\} \end{aligned}$$

Intuitively, this means that every leaf remains inside the set of witnesses ( $W_{i-1}^{v(1)}$ ) and for every inner node  $w$  we have two cases. If there exists a successor  $w^+$  with a strictly smaller value, then we replace the node  $w$  by this successor  $w^+$  in our set of witnesses ( $\{w^+ \mid w \in W_{i-1}^{v(3)}\}$ ). Otherwise, all the successors  $w_1, \dots, w_k$  of the node  $w$  have the same value as the node  $w$  itself, so that we have to replace  $w$  with all of its successors in our set of witnesses as there is no single node with a guaranteed value decrease ( $\bigcup_{w \in W_{i-1}^{v(2)}} wE$ ).

It remains to show that our properties (W-1!), (W-2!), and (W-3!) are still satisfied for  $W_i^v$ . In order to see that (W-1!) is satisfied, note that we have  $\overline{\text{val}}_{\mathfrak{X}}(w) \leq \overline{\text{val}}_{\mathfrak{X}}(v) - 1$ , for all  $w \in W_{i-1}^v$  by our induction hypothesis. Thus, for all  $w' \in W_{i-1}^{v(1)} \subseteq W_{i-1}^v$ , we also obtain  $\overline{\text{val}}_{\mathfrak{X}}(w') \leq \overline{\text{val}}_{\mathfrak{X}}(v) - 1$ . For all  $w' \in \bigcup_{w \in W_{i-1}^{v(2)}} wE$ , we have  $\overline{\text{val}}_{\mathfrak{X}}(w') = \overline{\text{val}}_{\mathfrak{X}}(w)$  for some  $w \in W_{i-1}^v$  and thus  $\overline{\text{val}}_{\mathfrak{X}}(w') = \overline{\text{val}}_{\mathfrak{X}}(w) \leq \overline{\text{val}}_{\mathfrak{X}}(v) - 1$ . Finally, for all  $w' \in \{w^+ \mid w \in W_{i-1}^{v(3)}\}$ , we have  $\overline{\text{val}}_{\mathfrak{X}}(w') \leq \overline{\text{val}}_{\mathfrak{X}}(w) - 1$  for some  $w \in W_{i-1}^v$  and thus  $\overline{\text{val}}_{\mathfrak{X}}(w') \leq \overline{\text{val}}_{\mathfrak{X}}(w) - 1 \leq \overline{\text{val}}_{\mathfrak{X}}(v) - 2 \leq \overline{\text{val}}_{\mathfrak{X}}(v) - 1$ .

Property (W-3!) holds as well, since every node  $w \in W_{i-1}$  that is not a leaf of  $\mathfrak{X}_v$  is at depth  $i - 1$  of the tree  $\mathfrak{X}_v$  by our induction hypothesis. We exchange each such node with one or all of its successors. The leaves in  $W_{i-1}$  are at a depth of at most  $i$  by induction hypothesis and remain in  $W_i$ . Hence, all the nodes of  $W_i$  are at a depth of at most  $i$ , and the nodes that are no leaves are at a depth of precisely  $i$ .

Finally, we regard (W-2!). For  $\bigcup_{w \in W_{i-1}^{v(2)}} wE$  we have:

$$\begin{aligned} &\sum_{w' \in \bigcup_{w \in W_{i-1}^{v(2)}} wE} p_{w'}^{\mathfrak{X}_v} \cdot \overline{\text{val}}_{\mathfrak{X}}(w') \\ &= \sum_{w \in W_{i-1}^{v(2)}} \sum_{w' \in wE} p_{w'}^{\mathfrak{X}_v} \cdot \overline{\text{val}}_{\mathfrak{X}}(w') \\ &= \sum_{w \in W_{i-1}^{v(2)}, wE = \{w_1, \dots, w_k\}} \sum_{1 \leq j \leq k} p_{w_j}^{\mathfrak{X}_v} \cdot \overline{\text{val}}_{\mathfrak{X}}(w_j) \\ &= \sum_{w \in W_{i-1}^{v(2)}, wE = \{w_1, \dots, w_k\}} \sum_{1 \leq j \leq k} p_{w_j}^{\mathfrak{X}_v} \cdot \overline{\text{val}}_{\mathfrak{X}}(w) \quad (\text{as } \overline{\text{val}}_{\mathfrak{X}}(w_j) = \overline{\text{val}}_{\mathfrak{X}}(w) \text{ for all } 1 \leq j \leq k) \\ &= \sum_{w \in W_{i-1}^{v(2)}, wE = \{w_1, \dots, w_k\}} \sum_{1 \leq j \leq k} p_j \cdot p_w^{\mathfrak{X}_v} \cdot \overline{\text{val}}_{\mathfrak{X}}(w) \quad (\text{as } p_{w_j}^{\mathfrak{X}_v} = p_j \cdot p_w^{\mathfrak{X}_v} \text{ for all } 1 \leq j \leq k) \end{aligned}$$

$$\begin{aligned}
 &= \sum_{w \in W_{i-1}^{v(2)}, wE = \{w_1, \dots, w_k\}} p_w^{\mathfrak{Z}_v} \cdot p_{\min}^{\overline{\text{Val}}_{\mathfrak{Z}}(w)} \cdot \sum_{1 \leq j \leq k} p_j \\
 &= \sum_{w \in W_{i-1}^{v(2)}} p_w^{\mathfrak{Z}_v} \cdot p_{\min}^{\overline{\text{Val}}_{\mathfrak{Z}}(w)} \cdot 1 \quad (\text{as } \sum_{1 \leq j \leq k} p_j = 1) \\
 &= \sum_{w \in W_{i-1}^{v(2)}} p_w^{\mathfrak{Z}_v} \cdot p_{\min}^{\overline{\text{Val}}_{\mathfrak{Z}}(w)}.
 \end{aligned}$$

For  $\{w^+ \mid w \in W_{i-1}^{v(3)}\}$  we have:

$$\begin{aligned}
 &\sum_{w' \in \{w^+ \mid w \in W_{i-1}^{v(3)}\}} p_{w'}^{\mathfrak{Z}_v} \cdot p_{\min}^{\overline{\text{Val}}_{\mathfrak{Z}}(w')} \\
 &= \sum_{w \in W_{i-1}^{v(3)}} p_w^{\mathfrak{Z}_v} \cdot p_{\min}^{\overline{\text{Val}}_{\mathfrak{Z}}(w^+)} \\
 &\geq \sum_{w \in W_{i-1}^{v(3)}} p_{\min} \cdot p_w^{\mathfrak{Z}_v} \cdot p_{\min}^{\overline{\text{Val}}_{\mathfrak{Z}}(w^+)} \quad (\text{as } p_w^{\mathfrak{Z}_v} \geq p_{\min} \cdot p_w^{\mathfrak{Z}_v}) \\
 &\geq \sum_{w \in W_{i-1}^{v(3)}} p_{\min} \cdot p_w^{\mathfrak{Z}_v} \cdot p_{\min}^{\overline{\text{Val}}_{\mathfrak{Z}}(w)-1} \quad (\text{as } \overline{\text{Val}}_{\mathfrak{Z}}(w^+) \leq \overline{\text{Val}}_{\mathfrak{Z}}(w) - 1) \\
 &= \sum_{w \in W_{i-1}^{v(3)}} p_w^{\mathfrak{Z}_v} \cdot p_{\min}^{\overline{\text{Val}}_{\mathfrak{Z}}(w)}.
 \end{aligned}$$

To summarize, we have

$$\begin{aligned}
 &\sum_{w' \in W_i^v} p_{w'}^{\mathfrak{Z}_v} \cdot p_{\min}^{\overline{\text{Val}}_{\mathfrak{Z}}(w')} \\
 &= \sum_{w' \in W_{i-1}^{v(1)}} p_{w'}^{\mathfrak{Z}_v} \cdot p_{\min}^{\overline{\text{Val}}_{\mathfrak{Z}}(w')} + \sum_{w' \in \bigcup_{w \in W_{i-1}^{v(2)}} wE} p_{w'}^{\mathfrak{Z}_v} \cdot p_{\min}^{\overline{\text{Val}}_{\mathfrak{Z}}(w')} + \sum_{w' \in \{w^+ \mid w \in W_{i-1}^{v(3)}\}} p_{w'}^{\mathfrak{Z}_v} \cdot p_{\min}^{\overline{\text{Val}}_{\mathfrak{Z}}(w')} \\
 &\geq \sum_{w \in W_{i-1}^{v(1)}} p_w^{\mathfrak{Z}_v} \cdot p_{\min}^{\overline{\text{Val}}_{\mathfrak{Z}}(w)} + \sum_{w \in W_{i-1}^{v(2)}} p_w^{\mathfrak{Z}_v} \cdot p_{\min}^{\overline{\text{Val}}_{\mathfrak{Z}}(w)} + \sum_{w \in W_{i-1}^{v(3)}} p_w^{\mathfrak{Z}_v} \cdot p_{\min}^{\overline{\text{Val}}_{\mathfrak{Z}}(w)} \\
 &\stackrel{(IH)}{\geq} p_{\min}^{\overline{\text{Val}}_{\mathfrak{Z}}(v)}.
 \end{aligned}$$

**III.2. We prove that  $|\mathfrak{Z}^{\leq N}| \geq p_{\min}^N$  for any  $N \in \mathbb{N}$  by induction:** We now prove that  $|\mathfrak{Z}^{\leq N}| \geq p_{\min}^N$  holds for any  $N \in \mathbb{N}$ . Let  $Z_k$  denote the set of nodes in  $\mathfrak{Z}^{\leq N}$  from  $\Gamma_{>}$  that have precisely  $k - 1$  nodes from  $\Gamma_{>}$  above them (so they are themselves the  $k$ -th node from  $\Gamma_{>}$ ). Here,  $\Gamma$ ,  $\Gamma_{\geq}$ , and  $\Gamma_{>}$  again refer to  $\mathfrak{Z}^{\leq N}$ , i.e., we again have  $\Gamma \cap C = \emptyset$ . Let  $\text{Leaf}_k$  denote the set of all leaves in  $\mathfrak{Z}^{\leq N}$  that are reachable by a path that uses less than  $k$  nodes from  $\Gamma_{>}$ . We show by induction that for all  $1 \leq k \leq N + 1$ , we have

$$\sum_{v \in Z_k \cup \text{Leaf}_k, 0 \leq \overline{\text{Val}}_{\mathfrak{Z}}(v) \leq N+1-k} p_v^{\mathfrak{Z}^{\leq N}} \cdot p_{\min}^{\overline{\text{Val}}_{\mathfrak{Z}}(v)} \geq p_{\min}^N$$

Then, for  $k = N + 1$  we finally have:

$$\begin{aligned}
 &p_{\min}^N \\
 &\leq \sum_{v \in Z_{N+1} \cup \text{Leaf}_{N+1}, 0 \leq \overline{\text{Val}}_{\mathfrak{Z}}(v) \leq N+1-(N+1)} p_v^{\mathfrak{Z}^{\leq N}} \cdot p_{\min}^{\overline{\text{Val}}_{\mathfrak{Z}}(v)} \\
 &= \sum_{v \in Z_{N+1} \cup \text{Leaf}_{N+1}, 0 \leq \overline{\text{Val}}_{\mathfrak{Z}}(v) \leq 0} p_v^{\mathfrak{Z}^{\leq N}} \cdot p_{\min}^{\overline{\text{Val}}_{\mathfrak{Z}}(v)} \\
 &= \sum_{v \in Z_{N+1} \cup \text{Leaf}_{N+1}, \overline{\text{Val}}_{\mathfrak{Z}}(v) = 0} p_v^{\mathfrak{Z}^{\leq N}} \cdot p_{\min}^{\overline{\text{Val}}_{\mathfrak{Z}}(v)} \\
 &= \sum_{v \in Z_{N+1} \cup \text{Leaf}_{N+1}, \overline{\text{Val}}_{\mathfrak{Z}}(v) = 0} p_v^{\mathfrak{Z}^{\leq N}} \cdot 1
 \end{aligned}$$

$$\begin{aligned}
 &= \sum_{v \in Z_{N+1} \cup \text{Leaf}_{N+1}, \overline{\text{val}}_{\mathfrak{T}}(v)=0} p_v^{\mathfrak{T} \leq N} \\
 &= \sum_{v \in \text{Leaf}_{N+1}} p_v^{\mathfrak{T} \leq N} \quad (\text{as } \overline{\text{val}}_{\mathfrak{T}}(v) = 0 \text{ iff } v \text{ is a leaf of } \mathfrak{T}^{\leq N}) \\
 &\leq \sum_{v \in \text{Leaf}^{\mathfrak{T} \leq N}} p_v^{\mathfrak{T} \leq N} \\
 &= |\mathfrak{T}^{\leq N}|
 \end{aligned}$$

In the induction base, we have  $k = 1$ , and thus

$$\begin{aligned}
 &\sum_{v \in Z_1 \cup \text{Leaf}_1, 0 \leq \overline{\text{val}}_{\mathfrak{T}}(v) \leq N+1-1} p_v^{\mathfrak{T} \leq N} \cdot p_{\min}^{\overline{\text{val}}_{\mathfrak{T}}(v)} \\
 &= \sum_{v \in Z_1 \cup \text{Leaf}_1, 0 \leq \overline{\text{val}}_{\mathfrak{T}}(v) \leq N} p_v^{\mathfrak{T} \leq N} \cdot p_{\min}^{\overline{\text{val}}_{\mathfrak{T}}(v)} \\
 &\geq \sum_{v \in Z_1 \cup \text{Leaf}_1, 0 \leq \overline{\text{val}}_{\mathfrak{T}}(v) \leq N} p_v^{\mathfrak{T} \leq N} \cdot p_{\min}^N \quad (\text{since } \overline{\text{val}}_{\mathfrak{T}}(v) \leq N) \\
 &= p_{\min}^N \cdot \sum_{v \in Z_1 \cup \text{Leaf}_1, 0 \leq \overline{\text{val}}_{\mathfrak{T}}(v) \leq N} p_v^{\mathfrak{T} \leq N} \\
 &= p_{\min}^N \cdot \sum_{v \in Z_1 \cup \text{Leaf}_1} p_v^{\mathfrak{T} \leq N} \\
 &= p_{\min}^N \cdot 1 \quad (\text{since } \sum_{v \in Z_1 \cup \text{Leaf}_1} p_v^{\mathfrak{T} \leq N} = 1) \\
 &= p_{\min}^N
 \end{aligned}$$

Here, we have  $\sum_{v \in Z_1 \cup \text{Leaf}_1} p_v^{\mathfrak{T} \leq N} = 1$ , since  $Z_1 \cup \text{Leaf}_1$  are the leaves of the finite subtree where we cut everything below the first node of  $\mathcal{P}_{>}$  (i.e., we cut directly after the nodes in  $Z_1$ ). This tree is finite, because by (+) there is no infinite path without  $\Gamma_{>}$  nodes. All finite CTs converge with probability 1.

In the induction step, we assume that the statement holds for some  $1 \leq k \leq N$ . Then we have

$$\begin{aligned}
 &p_{\min}^N \\
 (IH) \leq &\sum_{v \in Z_k \cup \text{Leaf}_k, 0 \leq \overline{\text{val}}_{\mathfrak{T}}(v) \leq N+1-k} p_v^{\mathfrak{T} \leq N} \cdot p_{\min}^{\overline{\text{val}}_{\mathfrak{T}}(v)} \\
 &= \sum_{v \in \text{Leaf}_k} p_v^{\mathfrak{T} \leq N} \cdot p_{\min}^{\overline{\text{val}}_{\mathfrak{T}}(v)} + \sum_{v \in Z_k, 1 \leq \overline{\text{val}}_{\mathfrak{T}}(v) \leq N+1-k} p_v^{\mathfrak{T} \leq N} \cdot p_{\min}^{\overline{\text{val}}_{\mathfrak{T}}(v)} \\
 &\leq \sum_{v \in \text{Leaf}_k} p_v^{\mathfrak{T} \leq N} \cdot p_{\min}^{\overline{\text{val}}_{\mathfrak{T}}(v)} + \sum_{v \in Z_k, 1 \leq \overline{\text{val}}_{\mathfrak{T}}(v) \leq N+1-k} p_v^{\mathfrak{T} \leq N} \cdot \sum_{w \in W^v} p_w^{\mathfrak{T}} \cdot p_{\min}^{\overline{\text{val}}_{\mathfrak{T}}(w)} \\
 &\quad (\text{existence of the set } W^v \text{ by the previous Step 3.1 and (W-2)}) \\
 &= \sum_{v \in \text{Leaf}_k} p_v^{\mathfrak{T} \leq N} \cdot p_{\min}^{\overline{\text{val}}_{\mathfrak{T}}(v)} + \sum_{v \in Z_k, 1 \leq \overline{\text{val}}_{\mathfrak{T}}(v) \leq N+1-k} \sum_{w \in W^v} p_v^{\mathfrak{T} \leq N} \cdot p_w^{\mathfrak{T}} \cdot p_{\min}^{\overline{\text{val}}_{\mathfrak{T}}(w)} \\
 &= \sum_{v \in \text{Leaf}_k} p_v^{\mathfrak{T} \leq N} \cdot p_{\min}^{\overline{\text{val}}_{\mathfrak{T}}(v)} + \sum_{v \in Z_k, 1 \leq \overline{\text{val}}_{\mathfrak{T}}(v) \leq N+1-k} \sum_{w \in W^v} p_w^{\mathfrak{T} \leq N} \cdot p_{\min}^{\overline{\text{val}}_{\mathfrak{T}}(w)} \\
 &\quad (\text{as } p_v^{\mathfrak{T} \leq N} \cdot p_w^{\mathfrak{T}} = p_w^{\mathfrak{T} \leq N}).
 \end{aligned}$$

Every node in  $W^v$  is either contained in  $\text{Leaf}_{k+1}$  or contained in  $Z_{k+1}$ . The reason for that is that  $W^v$  only contains leaves of  $\mathfrak{v}$  and a leaf in  $\mathfrak{v}$  is either also a leaf in  $\mathfrak{T}^{\leq N}$  so that it is contained in  $\text{Leaf}_{k+1}$ , or contained in  $\mathcal{P}_{>}$  and thus in  $Z_{k+1}$ , since  $v$  is contained in  $Z_k$  and there is no other inner node from  $\mathcal{P}_{>}$  in  $\mathfrak{v}$ . Furthermore, we know that  $\overline{\text{val}}_{\mathfrak{T}}(w) \leq \overline{\text{val}}_{\mathfrak{T}}(v) - 1$  for all  $w \in W^v$  by (W-1). Moreover, we have  $\text{Leaf}_k \subseteq \text{Leaf}_{k+1}$ . Thus, we get

$$\begin{aligned}
 &\sum_{v \in \text{Leaf}_k} p_v^{\mathfrak{T} \leq N} \cdot p_{\min}^{\overline{\text{val}}_{\mathfrak{T}}(v)} + \sum_{v \in Z_k, 1 \leq \overline{\text{val}}_{\mathfrak{T}}(v) \leq N+1-k} \sum_{w \in W^v} p_w^{\mathfrak{T} \leq N} \cdot p_{\min}^{\overline{\text{val}}_{\mathfrak{T}}(w)} \\
 &\leq \sum_{w \in Z_{k+1} \cup \text{Leaf}_{k+1}, 0 \leq \overline{\text{val}}_{\mathfrak{T}}(w) \leq N+1-(k+1)} p_w^{\mathfrak{T} \leq N} \cdot p_{\min}^{\overline{\text{val}}_{\mathfrak{T}}(w)}.
 \end{aligned}$$

To summarize, we have shown that  $|\mathfrak{T}^{\leq N}| \geq p_{\min}^N$ .

**IV. We prove that  $|\mathfrak{T}^{\leq N}| = 1$  for any  $N \in \mathbb{N}$ :** We have proven that  $|\mathfrak{T}^{\leq N}| \geq p_{\min}^N$  holds for all  $\mathcal{P}$ -CTs  $\mathfrak{T}$  that satisfy (+). Hence, for any  $N \in \mathbb{N}$ , we have

$$p_{\min}^* := \inf_{\mathfrak{T} \text{ is a } \mathcal{P}\text{-CT satisfying (+)}} (|\mathfrak{T}^{\leq N}|) \geq p_{\min}^N > 0. \tag{B.8}$$

We now prove by contradiction that this is enough to ensure  $p_N^* = 1$ . So assume that  $p_N^* < 1$ . Then we define  $\epsilon := \frac{p_N^* \cdot (1 - p_N^*)}{2} > 0$ . By definition of the infimum,  $p_N^* + \epsilon$  is not a lower bound of  $|\mathfrak{T}^{\leq N}|$  for all  $\mathcal{P}$ -CTs  $\mathfrak{T}$  that satisfy (+). Hence, there must exist a  $\mathcal{P}$ -CT  $\mathfrak{T}$  satisfying (+) such that

$$p_N^* \leq |\mathfrak{T}^{\leq N}| < p_N^* + \epsilon. \tag{B.9}$$

For readability, let  $Z = \text{Leaf}^{\mathfrak{T}^{\leq N}}$  be the set of leaves of the tree  $\mathfrak{T}^{\leq N}$  and let  $\bar{Z} = V^{\mathfrak{T}^{\leq N}} \setminus \text{Leaf}^{\mathfrak{T}^{\leq N}}$  be the set of inner nodes of the tree  $\mathfrak{T}^{\leq N}$ . Since  $p_v^{\mathfrak{T}^{\leq N}} = p_v^{\mathfrak{T}}$  for all  $v \in V^{\mathfrak{T}^{\leq N}}$ , in the following we just write  $p_v$ . By the monotonicity of  $|\cdot|$  w.r.t. the depth of the tree  $\mathfrak{T}^{\leq N}$ , there must exist a natural number  $m^* \in \mathbb{N}$  such that

$$\sum_{v \in Z, d(v) \leq m^*} p_v > \frac{p_N^*}{2}. \tag{B.10}$$

Here,  $d(v)$  denotes the depth of node  $v$  in  $\mathfrak{T}^{\leq N}$ . For every  $v \in V^{\mathfrak{T}^{\leq N}}$  with  $d(v) = m^*$  and  $v \notin Z$ , we define the sub  $\mathcal{P}$ -CT of  $\mathfrak{T}^{\leq N}$  starting at node  $v$  by  $\mathfrak{T}^{\leq N}(v) = \mathfrak{T}^{\leq N}[v(E^{\mathfrak{T}^{\leq N}})^*]$ . Then we have

$$|\mathfrak{T}^{\leq N}| = \sum_{v \in Z, d(v) \leq m^*} p_v + \sum_{v \in \bar{Z}, d(v) = m^*} p_v \cdot |\mathfrak{T}^{\leq N}(v)|. \tag{B.11}$$

Furthermore, we have

$$\sum_{v \in \bar{Z}, d(v) = m^*} p_v = 1 - \sum_{v \in Z, d(v) \leq m^*} p_v, \tag{B.12}$$

since  $\sum_{v \in \bar{Z}, d(v) = m^*} p_v + \sum_{v \in Z, d(v) \leq m^*} p_v = 1$ , as the nodes  $v \in \bar{Z}$  with  $d(v) = m^*$  are the leaves of the finite, grounded subtree of  $\mathfrak{T}^{\leq N}$  where we cut every edge after the nodes of depth  $m^*$ . We obtain

$$\begin{aligned} & p_N^* + \epsilon \\ & > |\mathfrak{T}^{\leq N}| \quad (\text{by (B.9)}) \\ & = \sum_{v \in Z, d(v) \leq m^*} p_v + \sum_{v \in \bar{Z}, d(v) = m^*} p_v \cdot \underbrace{|\mathfrak{T}^{\leq N}(v)|}_{\geq p_N^*} \quad (\text{by (B.11) and (B.8)}) \\ & \geq \sum_{v \in Z, d(v) \leq m^*} p_v + \sum_{v \in \bar{Z}, d(v) = m^*} p_v \cdot p_N^* \\ & = \sum_{v \in Z, d(v) \leq m^*} p_v + p_N^* \cdot \sum_{v \in \bar{Z}, d(v) = m^*} p_v \\ & = \sum_{v \in Z, d(v) \leq m^*} p_v + p_N^* \cdot (1 - \sum_{v \in Z, d(v) \leq m^*} p_v) \quad (\text{by (B.12)}) \\ & = p_N^* + \sum_{v \in Z, d(v) \leq m^*} p_v - p_N^* \cdot \sum_{v \in Z, d(v) \leq m^*} p_v \\ & = p_N^* + \sum_{v \in Z, d(v) \leq m^*} p_v \cdot (1 - p_N^*) \\ & > p_N^* + (1 - p_N^*) \cdot \frac{p_N^*}{2} \quad (\text{by (B.10)}) \\ & = p_N^* + \epsilon, \end{aligned}$$

which is a contradiction. So  $p_N^* = 1$ . In particular, this means that for every  $N \in \mathbb{N}$  and every  $\mathcal{P}$ -CT  $\mathfrak{T}$  satisfying (+), we have

$$|\mathfrak{T}^{\leq N}| = 1. \tag{B.13}$$

**V. Finally, we prove that  $|\mathfrak{T}| = 1$ :** We adjust the tree value function for this new tree  $\mathfrak{T}^{\leq N}$  once again and define:

$$\widehat{\mathcal{V}al}_{\mathfrak{T}} : V^{\mathfrak{T}^{\leq N}} \rightarrow \mathbb{N}, \quad v \mapsto \begin{cases} N + 1, & \text{if } v \in C0 \\ g_{\mathcal{A} \rightarrow \mathbb{N}}(\mathcal{A}_{sum}^{\#}(t_v, \delta_a)) + 1, & \text{if } v \in \text{Leaf}^{\mathfrak{T}} \\ & \text{otherwise} \end{cases}$$

Now for a node  $v \in V^{\mathfrak{T}^{\leq N}}$  we have  $0 \leq \widehat{\mathcal{V}al}_{\mathfrak{T}}(v) \leq N + 1$ . Furthermore, we have  $\mathcal{V}al_{\mathfrak{T}}(v) \geq \widehat{\mathcal{V}al}_{\mathfrak{T}}(v)$  for all  $v \in V^{\mathfrak{T}^{\leq N}}$ . Let  $|\mathfrak{T}^{\leq N}|_{\widehat{\mathcal{V}al}_{\mathfrak{T}}} = \sum_{v \in \text{Leaf}^{\mathfrak{T}^{\leq N}}} p_v \cdot \widehat{\mathcal{V}al}_{\mathfrak{T}}(v)$ . This (possibly infinite) sum is well defined (i.e., it is a convergent series), because all addends are non-negative and the sum is bounded from above by  $N + 1$ , since

$$\sum_{v \in \text{Leaf}^{\mathfrak{T}^{\leq N}}} p_v \cdot \widehat{\mathcal{V}al}_{\mathfrak{T}}(v) \leq \sum_{v \in \text{Leaf}^{\mathfrak{T}^{\leq N}}} p_v \cdot (N + 1) = (N + 1) \cdot \sum_{v \in \text{Leaf}^{\mathfrak{T}^{\leq N}}} p_v \leq (N + 1) \cdot 1 = N + 1.$$

For the root  $r$  of  $\mathfrak{T}$  and  $\mathfrak{T}^{\leq N}$ , we have  $\widehat{\mathcal{V}al}_{\mathfrak{T}}(r) \geq |\mathfrak{T}^{\leq N}|_{\widehat{\mathcal{V}al}_{\mathfrak{T}}}$  because due to (d), we know that the (expected) value is non-increasing.

Now we fix  $N \in \mathbb{N}$  and a CT  $\mathfrak{T}$  satisfying (+), and obtain the corresponding transformed tree  $\mathfrak{T}^{\leq N}$ . Note that by (B.13) we have  $|\mathfrak{T}^{\leq N}| = 1 = q_N + q'_N$ , where  $q_N := \sum_{v \in \text{Leaf}^{\mathfrak{T}^{\leq N}}, \widehat{\mathcal{V}al}_{\mathfrak{T}}(v) = 0} p_v$  and  $q'_N = \sum_{v \in \text{Leaf}^{\mathfrak{T}^{\leq N}}, \widehat{\mathcal{V}al}_{\mathfrak{T}}(v) > 0} p_v = \sum_{v \in C} p_v$ . Now we can determine

$|\mathfrak{Z}^{\leq N}|_{\widehat{val}_{\mathfrak{Z}}}$ . The probabilities of zero-valued nodes (i.e., leaves in the original tree) add up to  $q_N$ , while the probabilities of new leaves due to a cut add up to probability  $q'_N = 1 - q_N$ . So  $|\mathfrak{Z}^{\leq N}|_{\widehat{val}_{\mathfrak{Z}}} = q_N \cdot 0 + (1 - q_N) \cdot (N + 1) = (1 - q_N) \cdot (N + 1)$ . Thus,

$$val_{\mathfrak{Z}}(\mathbf{r}) \geq \widehat{val}_{\mathfrak{Z}}(\mathbf{r}) \geq |\mathfrak{Z}^{\leq N}|_{\widehat{val}_{\mathfrak{Z}}} \geq (1 - q_N) \cdot (N + 1),$$

which implies  $q_N \geq 1 - \frac{val_{\mathfrak{Z}}(\mathbf{r})}{N+1}$ . Note that  $q_N$  is weakly monotonically increasing and bounded from above by 1 for  $N \rightarrow \infty$ . Hence,  $q := \lim_{N \rightarrow \infty} q_N$  exists and  $1 \geq q \geq \lim_{N \rightarrow \infty} (1 - \frac{val_{\mathfrak{Z}}(\mathbf{r}^{\mathfrak{Z}})}{N+1}) = 1$ , i.e.,  $q = 1$ . Hence, we obtain  $|\mathfrak{Z}| = \lim_{N \rightarrow \infty} q_N = q = 1$ .  $\square$

By Lemma 5, we have finally proven the main technical part for the reduction pair processor in the probabilistic setting. Next, we want to prove the soundness and completeness of the actual processor, which is fairly easy using Lemma 5 in addition to the  $\Gamma$ -Partition Lemma.

**Proof.** (Proof of Theorem 12) Let  $\overline{\mathcal{P}} = \mathcal{P}_{\geq} \cup b(\mathcal{P}_{>})$ .

Completeness: By Theorem 6.

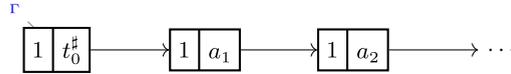
Soundness: Suppose that every (innermost)  $\overline{\mathcal{P}}$ -CT converges with probability 1. Assume for a contradiction that there exists an (innermost)  $\mathcal{P}$ -CT  $\mathfrak{Z} = (V, E, L)$  that converges with probability  $< 1$ . Let us partition  $\Gamma = \Gamma_{\geq} \uplus \Gamma_{>}$  according to the used ADPs for the rewrite steps. Since every  $\overline{\mathcal{P}}$ -CT converges with probability 1, so does every subtree of  $\mathfrak{Z}$  that only contains  $\Gamma$ -nodes from  $\Gamma_{\geq}$ . By the  $\Gamma$ -Partition Lemma (Lemma 4), there must be an (innermost)  $\mathcal{P}$ -CT  $\mathfrak{Z}'$  with  $|\mathfrak{Z}'| < 1$  such that every infinite path has an infinite number of edges corresponding to  $\mathcal{P}_{>}$  steps. But this is a contradiction to Lemma 5.  $\square$

**Proof.** (Proof of Theorem 13) Let  $\mathcal{P}$  be an ADP problem such that every ADP in  $\mathcal{P}$  has the form  $\ell \rightarrow \{1 : r\}^m$ . Note that every  $\mathcal{P}$ -CT is a single (not necessarily finite) path. For such a CT  $\mathfrak{Z}$  that is only a single path, we have only two possibilities for  $|\mathfrak{Z}|$ . If the path is finite, then  $|\mathfrak{Z}| = 1$ , since we have a single leaf in this tree with probability 1. Otherwise, we have an infinite path, which means that there is no leaf at all and hence  $|\mathfrak{Z}| = 0$ . We prove both the “if” and the “only if” direction, where for innermost rewriting, one has to use  $np_i(\mathcal{P})$  instead of  $np(\mathcal{P})$ .

“only if”: We first consider AST. Assume that  $(dp(\mathcal{P}), np(\mathcal{P}))$  is not terminating. Then there exists an infinite  $(dp(\mathcal{P}), np(\mathcal{P}))$ -chain

$$t_0^{\#} \xrightarrow{dp(\mathcal{P})} \circ \xrightarrow{np(\mathcal{P})}^* t_1^{\#} \xrightarrow{dp(\mathcal{P})} \circ \xrightarrow{np(\mathcal{P})}^* t_2^{\#} \xrightarrow{dp(\mathcal{P})} \circ \xrightarrow{np(\mathcal{P})}^* \dots$$

such that for all  $i \in \mathbb{N}$  we have  $t_i^{\#} = \ell_i^{\#} \sigma_i$  for some  $\ell_i^{\#} \rightarrow r_i^{\#} \in dp(\mathcal{P})$  and some substitution  $\sigma_i$ . From this infinite  $(dp(\mathcal{P}), np(\mathcal{P}))$ -chain, we will now construct an infinite  $\mathcal{P}$ -CT  $\mathfrak{Z} = (V, E, L)$ . As explained above, we then know that this infinite  $\mathcal{P}$ -CT must be an infinite path, and thus  $|\mathfrak{Z}| = 0$ . This means that  $\mathcal{P}$  is not AST, and thus, a processor with  $Proc_{PR}(\mathcal{P}) = \emptyset$  would be unsound.



We start our CT with  $(1 : t_0^{\#})$ . In the non-probabilistic rewrite sequence, we have  $t_0^{\#} \xrightarrow{dp(\mathcal{P})} \circ \xrightarrow{np(\mathcal{P})}^* t_1^{\#}$ , so there exists a natural number  $k \geq 1$  such that

$$t_0^{\#} = \ell_0^{\#} \sigma_0 \rightarrow_{dp(\mathcal{P})} r_0^{\#} \sigma_0 = s_1^{\#} \rightarrow_{np(\mathcal{P})} s_2^{\#} \rightarrow_{np(\mathcal{P})} \dots \rightarrow_{np(\mathcal{P})} s_k^{\#} = t_1^{\#} = \ell_1^{\#} \sigma_1$$

Performing the same rewrite steps with  $\mathcal{P}$  yields terms  $a_1, \dots, a_k$  such that  $s_i \leq_{\#} a_i$  for all  $1 \leq i \leq k$ . Here, one needs that all ADPs that yield the rules in  $np(\mathcal{P})$  have the flag true and thus, the annotations above the redex are not removed. For all  $1 \leq i \leq k$ , we now construct  $a_i$  inductively.

In the induction base ( $i = 1$ ), let  $\ell_0 \rightarrow \{1 : r'_0\}^m \in \mathcal{P}$  be the ADP that was used to create the dependency pair  $\ell_0^{\#} \rightarrow r_0^{\#}$  in  $dp(\mathcal{P})$ . This means that we have  $r_0 \leq_{\#} r'_0$ . Since we have  $t_0^{\#} = \ell_0^{\#} \sigma_0$ , we can also rewrite  $t_0^{\#}$  with the ADP  $\ell_0 \rightarrow \{1 : r'_0\}^m \in \mathcal{P}$  and the substitution  $\sigma_0$  at the root position. We result in  $a_1 = r'_0 \sigma_0$  and thus we have  $s_1 = r_0 \sigma_0 \leq_{\#} r'_0 \sigma_0 = a_1$ .

In the induction step, we assume that  $s_i \leq_{\#} a_i$  for some  $1 \leq i < k$ . Let  $\pi$  be the annotated position of  $a_i$  where  $s_i = b(a_i)|_{\pi}$ . In our non-probabilistic rewrite sequence we have  $s_i^{\#} \rightarrow_{np(\mathcal{P})} s_{i+1}^{\#}$  using a rule  $\ell' \rightarrow b(r') \in np(\mathcal{P})$  and substitution  $\delta_i$  at a position  $\tau \in \mathbb{N}^+$  such that  $s_i^{\#}|_{\tau} = \ell' \delta_i$  and  $s_{i+1}^{\#} = s_i^{\#}[b(r')\delta_i]_{\tau}$ . We can mirror this rewrite step with the ADP  $\ell' \rightarrow \{1 : r'\}^{true} \in \mathcal{P}$ , since by construction we have  $s_i \leq_{\#} a_i$  and  $s_i = b(a_i)|_{\pi}$ . We obtain  $a_i \xrightarrow{\mathcal{P}} a_{i+1} = a_i[\#_{\Phi}(r'\delta_i)]_{\pi, \tau}$  with  $\Phi = Pos_{D^{\#}}(r_j) \cup \Psi_j$  (step with (at)) or  $\Phi = \Psi_1$  (step with (nt)) by rewriting the subterm of  $a_i$  at position  $\pi, \tau$ , which implies  $s_{i+1} = s_i[b(r')\delta_i]_{\tau} \leq_{\#} a_i[\#_{\Phi}(r'\delta_i)]_{\pi, \tau} = a_{i+1}$ .

At the end of this induction, we result in  $a_k$ . Next, we can mirror the step  $t_1^{\#} \xrightarrow{dp(\mathcal{P})} \circ \xrightarrow{np(\mathcal{P})}^* t_2^{\#}$  from our non-probabilistic rewrite sequence with the same construction, etc. This results in an infinite  $\mathcal{P}$ -CT. To see that this is indeed a  $\mathcal{P}$ -CT, note that all the local properties are satisfied since every edge represents a rewrite step with  $\xrightarrow{\mathcal{P}}$ . The global property is also satisfied since, in an infinite  $(dp(\mathcal{P}), np(\mathcal{P}))$ -chain, we use an infinite number of steps with  $\rightarrow_{dp(\mathcal{P})}$  so that our resulting CT has an infinite number of  $\Gamma$ -nodes.

For  $\mathfrak{i}AST$  we just have to consider innermost rewriting, while the construction remains the same. More precisely, we start with an innermost  $(dp(\mathcal{P}), np_i(\mathcal{P}))$ -chain

$$t_0^{\#} \xrightarrow{dp(\mathcal{P})} \circ \xrightarrow{np_i(\mathcal{P})}^* t_1^{\#} \xrightarrow{dp(\mathcal{P})} \circ \xrightarrow{np_i(\mathcal{P})}^* t_2^{\#} \xrightarrow{dp(\mathcal{P})} \circ \xrightarrow{np_i(\mathcal{P})}^* \dots$$

and construct an innermost  $\mathcal{P}$ -CT. But since all the steps in the chain are innermost and since  $\text{np}_i(\mathcal{P})$  has the same left-hand sides as  $\mathcal{P}$ , the steps in the CT are innermost as well.

“if $^{\sharp}$ ”: Assume that  $\mathcal{P}$  is not AST, i.e., that the processor  $\text{Proc}_{\text{PR}}(\mathcal{P}) = \emptyset$  is unsound. By Lemma 2, there exists a  $\mathcal{P}$ -CT  $\mathfrak{Z} = (V, E, L)$  that converges with probability  $< 1$  and starts with  $(1 : t^{\sharp})$  such that  $t^{\sharp} = \ell^{\sharp}\sigma_0$  for some substitution  $\sigma_0$  and an ADP  $\ell \rightarrow \{1 : r\}^m \in \mathcal{P}$ . As explained above, this tree must be an infinite path.

From  $\mathfrak{Z}$ , we will now construct an infinite  $(\text{dp}(\mathcal{P}), \text{np}(\mathcal{P}))$ -chain, which shows that  $(\text{dp}(\mathcal{P}), \text{np}(\mathcal{P}))$  is not terminating. We start our infinite chain with the term  $t_0^{\sharp} = t^{\sharp}$ . We have  $t^{\sharp} \hookrightarrow_{\mathcal{P}} \{1 : a_1\}$ , where  $a_1 = \sharp_{\Phi_1}(r\sigma_0) = r\sigma_0$ .

There must be a term  $r_0 \leq_{\sharp} r$  (i.e.,  $r_0\sigma_0 \leq r\sigma_0$ ) such that if we replace  $a_1 = r\sigma_0$  with  $r_0^{\sharp}\sigma_0$  and obtain the same  $\mathcal{P}$ -CT except when we would rewrite terms that do not exist anymore (i.e., we ignore these rewrite steps), then we still end up in an infinite number of nodes in  $\Gamma$  (otherwise,  $\mathfrak{Z}$  would not have an infinite number of nodes in  $\Gamma$ ).

Let  $\pi$  be the annotated position of  $r\sigma_0$  where  $r_0\sigma_0 = b(r\sigma_0)|_{\pi}$ . We can rewrite the term  $t_0^{\sharp}$  with the dependency pair  $\ell^{\sharp} \rightarrow r_0^{\sharp} \in \text{dp}(\mathcal{P})$ , using the substitution  $\sigma_0$  since  $t_0 = t = \ell\sigma_0$ . Hence, we result in  $t_0^{\sharp} \rightarrow_{\text{dp}(\mathcal{P})} r_0^{\sharp}\sigma_0 = s_1^{\sharp}$ . Next, we mirror the rewrite steps from the  $\mathcal{P}$ -CT that are performed strictly below the root of  $r_0^{\sharp}\sigma_0$  with  $\text{np}(\mathcal{P})$  until we would rewrite at the root of  $r_0^{\sharp}\sigma_0$ . With this construction, we ensure that each the term  $s_i^{\sharp}$  in our  $(\text{dp}(\mathcal{P}), \text{np}(\mathcal{P}))$ -chain satisfies  $s_i \leq_{\sharp} a_i$  and  $s_i = b(a_i)|_{\pi}$ . A rewrite step at position  $\pi.\tau$  in  $a_i$  with  $\mathcal{P}$  is then mirrored with  $\text{np}(\mathcal{P})$  in  $s_i^{\sharp}$  at position  $\tau$ . Note that we only use a finite number of  $\text{np}(\mathcal{P})$  steps until we rewrite at the root.

So eventually, we result in a term  $t_1^{\sharp} = s_k^{\sharp}$  with  $s_k = b(a_k)|_{\pi}$ , and we rewrite at position  $\pi$  in the  $\mathcal{P}$ -CT. We mirror the step  $a_k \hookrightarrow_{\mathcal{P}} a_{k+1}$  with  $\text{dp}(\mathcal{P})$  and then use the same construction again until we reach term  $t_2^{\sharp}$ , etc. This construction creates a sequence  $t_0^{\sharp}, t_1^{\sharp}, \dots$  of terms such that

$$t_0^{\sharp} \rightarrow_{\text{dp}(\mathcal{P})} r_0^{\sharp}\sigma_0 \xrightarrow{*}_{\text{np}(\mathcal{P})} t_1^{\sharp} \rightarrow_{\text{dp}(\mathcal{P})} r_1^{\sharp}\sigma_1 \xrightarrow{*}_{\text{np}(\mathcal{P})} \dots$$

Therefore,  $(\text{dp}(\mathcal{P}), \text{np}(\mathcal{P}))$  is not terminating.

Again, the proof works analogously for  $\text{iAST}$  just considering innermost rewriting.  $\square$

**Proof.** (Proof of Theorem 14) Consider the polynomial interpretation that maps each  $f(x_1, \dots, x_n)$  with  $f \in \Sigma$  to  $x_1 + \dots + x_n + 1$ , and each  $f^{\sharp}(x_1, \dots, x_n)$  with  $f^{\sharp} \in \Sigma^{\sharp}$  to  $x_{\text{proj}(f^{\sharp})}$ , i.e., we map each annotated symbol to its projected argument by the simple projection  $\text{proj}$ , and all other symbols are used to calculate the number of symbols in a term.

When choosing  $\mathcal{P}_{\geq}$  to be  $\mathcal{P}_{\leq}$ , then the conditions of the reduction pair processor that consider the  $\sharp$ -sums (Case (1) and (2a)) are both satisfied with this interpretation. On the other hand, the conditions on the flattened ADPs without annotations (Case (2b) and (3)) may not be satisfied. However, due to the restriction to innermost rewriting and the restriction that ADPs only contain at most a single annotated symbol in each term on the right-hand side, those cases are irrelevant, because of the following observation:

If we perform an innermost rewrite step with an ADP  $\ell \rightarrow \mu^m \in \mathcal{P}$  at an annotated position  $\tau$  with the substitution  $\sigma$ , then  $\ell^{\sharp}\sigma \in \text{ANF}_{\mathcal{P}}$ . Let  $r \in \text{Supp}(\mu)$  and  $t \leq_{\sharp} r$  if  $r$  contains an annotated symbol. Note that there exists at most one such  $t$  for each  $r$ , as there is at most one annotation in  $r$ . Since  $\text{proj}(\ell^{\sharp}) \geq \text{proj}(t^{\sharp})$  by definition of the subterm criterion, and all proper subterms of  $\ell^{\sharp}\sigma$  are in  $\text{NF}_{\mathcal{P}}$ , we have  $\text{proj}(t^{\sharp}) \in \text{NF}_{\mathcal{P}}$ . Hence, we cannot rewrite below the projected position before the next (at)- or (af)-step at the annotated position. More precisely, this means that the  $\sharp$ -sum of the terms cannot change by rewrite steps at non-annotated positions. Thus, in this case the reduction pair processor works without requiring Case (2b) and (3), as these requirements are never used in its proof if the rules have no impact on the  $\sharp$ -sum of a term.  $\square$

**Proof.** (Proof of Theorem 15) Let  $\overline{\mathcal{P}'} = \text{Proc}_{\tau}^{\dagger}(\mathcal{P})$  and  $\overline{\mathcal{P}} = \overline{\mathcal{P}'} \cup \{\ell \rightarrow \{p_1 : r_1, \dots, p_k : r_k\}^m\}$ . We call  $\ell \rightarrow \{p_1 : r_1, \dots, p_k : r_k\}^m$  the old ADP, we call  $\ell \rightarrow (\{p_1 : r_1, \dots, p_k : r_k\} \setminus \{p_j : r_j\} \cup \{p_j \cdot q_1 : e_1, \dots, p_j \cdot q_h : e_h\})^m$  the new ADP, and  $\ell \rightarrow \{p_1 : b(r_1), \dots, p_k : b(r_k)\}^m$  is called the non-annotated old ADP.

**First Case:** We start with the case where  $\mathcal{U}_{\mathcal{P}}(r_j|_{\tau})$  is NO and the used rule  $\ell \rightarrow \{\hat{p}_1 : \hat{r}_1, \dots, \hat{p}_h : \hat{r}_h\}$  is L and NE. First, note that the rule used for rewriting  $r_j|_{\tau}$  in the rewrite processor is contained in  $\mathcal{U}_{\mathcal{P}}(r_j|_{\tau})$ .<sup>16</sup> Since  $\mathcal{U}_{\mathcal{P}}(r_j|_{\tau})$  is NO, we can therefore be sure that there is only a single rule applicable at position  $\tau$ .

Let  $\mathcal{P}$  be not  $\text{iAST}$ . Then there exists a  $\mathcal{P}$ -CT  $\mathfrak{Z} = (V, E, L)$  that converges with probability  $c < 1$ . We will now create a  $\overline{\mathcal{P}'}$ -CT  $\mathfrak{Z}' = (V', E', L')$  such that  $|\mathfrak{Z}'| \leq |\mathfrak{Z}| < 1$ , and hence  $\overline{\mathcal{P}'}$  is not  $\text{iAST}$  either.

The core steps of this construction are the following:

1. We iteratively remove usages of the old ADP using a transformation  $\Phi$  on CTs. The limit  $\mathfrak{Z}^{(\infty)}$  of this iteration is a  $\overline{\mathcal{P}'}$ -CT that converges with probability at most  $|\mathfrak{Z}'|$ . The transformation  $\Phi$  always removes a “topmost” usage of the old ADP at some node  $x \in V$ . There are two different cases, depending on whether the old ADP is applied at an annotated position:
  - 1.1. For a  $\mathcal{P}$ -CT  $\mathfrak{Z}_x$  that uses the old ADP at the root  $x$  at a position  $\pi \in \text{Pos}_{\mathcal{D}^{\sharp}}(t_x)$ , we create a new  $\overline{\mathcal{P}'}$ -CT  $\Phi(\mathfrak{Z}_x)$  that uses the new ADP at the root.
  - 1.2. For a  $\mathcal{P}$ -CT  $\mathfrak{Z}_x$  that uses the old ADP at the root  $x$  at a position  $\pi \notin \text{Pos}_{\mathcal{D}^{\sharp}}(t_x)$ , we create a new  $\overline{\mathcal{P}'}$ -CT  $\Phi(\mathfrak{Z}_x)$  that uses the non-annotated old ADP at the root.

<sup>16</sup> The rules that are applicable at position  $\tau$  do not have to be applicable in an innermost RST, see [59, Section 6] and [53, Example 5.14].

This proof structure will also be used in the other cases for soundness.

**1. We iteratively remove usages of the old ADP:** W.l.o.g., in  $\mathfrak{Z}$  there exists at least one rewrite step performed at some node  $x$  with the old ADP (otherwise,  $\mathfrak{Z}$  would already be a  $\overline{\mathcal{P}'}$ -CT). Furthermore, we can assume that this is the first such rewrite step in the path from the root to the node  $x$  and that  $x$  is a node of minimum depth with this property. We will now replace this rewrite step with a rewrite step using the new ADP such that we result in a CT  $\mathfrak{Z}^{(1)}$  with the following connections between  $\mathfrak{Z}$  and  $\mathfrak{Z}^{(1)}$ . Here, for an infinite path  $p = v_1, v_2, \dots$  in a CT  $\mathfrak{Z} = (V, E, L)$ , by  $C(p) = w_1, w_2, \dots$  we denote the sequence of  $\{\Gamma, \overline{\Gamma}\}$ -labels for these nodes, i.e., we have  $w_i = \Gamma$  if  $v_i \in \Gamma$  or  $w_i = \overline{\Gamma}$  otherwise for all  $i$ .

- (a)  $|\mathfrak{Z}^{(1)}| \leq |\mathfrak{Z}| = c$ , and
- (b) for every infinite path  $p$  in  $\mathfrak{Z}^{(1)}$  we can find an infinite path  $p'$  in  $\mathfrak{Z}$  such that  $C(p)$  and  $C(p')$  only differ in a finite number of  $\overline{\Gamma}$ -nodes, i.e., we can remove and add a finite number of  $\overline{\Gamma}$ -nodes from  $C(p')$  to get  $C(p)$ .

This construction only works because  $\mathcal{U}_p(r_j|_x)$  is NO and there exists no annotation below  $\tau$ , and the resulting CT has only at most the same probability of termination as the original one due to the fact that  $\hat{\ell} \rightarrow \{\hat{p}_1 : \hat{r}_1, \dots, \hat{p}_h : \hat{r}_h\}^{m'}$  is also L and NE. Let  $\mathfrak{Z}_x$  be the induced subtree that starts at node  $x$ , i.e.  $\mathfrak{Z}_x = \mathfrak{Z}[xE^*]$ . The construction defines a new tree  $\Phi(\mathfrak{Z}_x)$  such that (a) and (b) w.r.t.  $\mathfrak{Z}_x$  and  $\Phi(\mathfrak{Z}_x)$  holds, and where we use the new ADP at the root node  $x$  instead of the old one (i.e., we pushed the first use of the old ADP deeper into the tree). Then, by replacing the subtree  $\mathfrak{Z}_x$  with the new tree  $\Phi(\mathfrak{Z}_x)$  in  $\mathfrak{Z}$ , we get a  $\mathcal{P}$ -CT  $\mathfrak{Z}^{(1)}$ , with (a) and (b) w.r.t.  $\mathfrak{Z}$  and  $\mathfrak{Z}^{(1)}$ , and where we use the new ADP at node  $x$  instead of the old one. We can then do this replacement iteratively for every use of the old ADP, i.e., we again replace the first use of the old ADP in  $\mathfrak{Z}^{(1)}$  to get  $\mathfrak{Z}^{(2)}$  with (a) and (b) w.r.t.  $\mathfrak{Z}^{(1)}$  and  $\mathfrak{Z}^{(2)}$ , and so on. The limit of all these CTs  $\lim_{i \rightarrow \infty} \mathfrak{Z}^{(i)}$  is a  $\overline{\mathcal{P}'}$ -CT, that we denote by  $\mathfrak{Z}^{(\infty)}$  and that converges with probability at most  $c < 1$ , and hence,  $\overline{\mathcal{P}'}$  is not iAST.

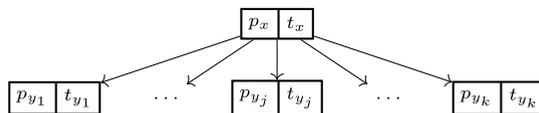
To see that  $\mathfrak{Z}^{(\infty)}$  is indeed a valid  $\overline{\mathcal{P}'}$ -CT, note that in every iteration of the construction we turn a use of the old ADP at minimum depth into a use of the new one. Hence, for every depth  $H$  of the tree, we eventually turned every use of the old ADP up to depth  $H$  into a use of the new one so that the construction will not change the tree above depth  $H$  anymore, i.e., there exists an  $m_H$  such that  $\mathfrak{Z}^{(\infty)}$  and  $\mathfrak{Z}^{(m_H)}$  are the same trees up to depth  $H$ . This means that the sequence  $\lim_{i \rightarrow \infty} \mathfrak{Z}^{(i)}$  really converges into a tree that satisfies the first three conditions of a  $\overline{\mathcal{P}'}$ -CT. We only have to show that the last condition of a CT, namely that every infinite path in  $\lim_{i \rightarrow \infty} \mathfrak{Z}^{(i)}$  contains infinitely many nodes from  $\Gamma$ , holds as well. First, by induction on  $n$  one can prove that all trees  $\mathfrak{Z}^{(i)}$  for all  $1 \leq i \leq n$  satisfy Condition (4), because due to (b) we can find for each infinite path  $p \in \mathfrak{Z}^{(i)}$  an infinite path  $p'$  in  $\mathfrak{Z}$  such that  $C(p)$  and  $C(p')$  only differ in a finite number of  $\overline{\Gamma}$ -labels. Hence, if  $p$  contains no nodes from  $\Gamma$ , then  $p'$  contains no nodes from  $\Gamma$ , which is a contradiction to  $\mathfrak{Z}$  satisfying Condition (4). Moreover, we only replace subtrees after a node in  $\Gamma$ , the node itself remains in  $\Gamma$ , and after we replaced the subtree at a node  $v$ , we will never replace a predecessor of  $v$  anymore. This means that the subsequence between the  $i$ -th and  $(i + 1)$ -th occurrence of  $\Gamma$  in  $p$  and the subsequence between the  $i$ -th and  $(i + 1)$ -th occurrence of  $\Gamma$  in  $p'$  only differ in a finite number of  $\overline{\Gamma}$  nodes again, for every  $i$ . Now, let  $p = v_1, v_2, \dots$  be an infinite path that starts at the root in  $\mathfrak{Z}^{(\infty)}$  and only contains finitely many nodes from  $\Gamma$ , i.e., there exists an  $1 \leq i$  such that  $v_i, v_{i+1}, \dots$  contains no node from  $\Gamma$ . Again, let  $m_H \in \mathbb{N}$  such that  $\mathfrak{Z}^{(\infty)}$  and  $\mathfrak{Z}^{(m_H)}$  are the same trees up to depth  $H$ . The path  $v_1, \dots, v_H$  must be a path in  $\mathfrak{Z}^{(m_H)}$  as well. For two different  $H_1, H_2$  with  $i < H_1 < H_2$  we know that since the path  $v_i, \dots, v_{H_2}$  contains no node from  $\Gamma$ , the path must also exist in  $\mathfrak{Z}^{(m_{H_1})}$ . We can now construct an infinite path in  $\mathfrak{Z}^{(m_i)}$  that contains no nodes from  $\Gamma$ , which is a contradiction.

Next, we want to prove that we really have  $|\mathfrak{Z}^{(\infty)}| \leq c$ . Again, by induction on  $n$  one can prove that  $|\mathfrak{Z}^{(i)}| \leq c$  for all  $1 \leq i \leq n$ . Assume for a contradiction that  $\mathfrak{Z}^{(\infty)}$  converges with probability greater than  $c$ , i.e.  $|\mathfrak{Z}^{(\infty)}| > c$ . Then there exists an  $H \in \mathbb{N}$  for the depth such that  $\sum_{x \in \text{Leaf} \mathfrak{Z}^{(\infty)}, d(x) \leq H} p_x > c$ . Here,  $d(x)$  denotes the depth of node  $x$ . Again, let  $m_H \in \mathbb{N}$  such that  $\mathfrak{Z}^{(\infty)}$  and  $\mathfrak{Z}^{(m_H)}$  are the same trees up to depth  $H$ . But this would mean that  $|\mathfrak{Z}^{(m_H)}| \geq \sum_{x \in \text{Leaf} \mathfrak{Z}^{(m_H)}, d(x) \leq H} p_x = \sum_{x \in \text{Leaf} \mathfrak{Z}^{(\infty)}, d(x) \leq H} p_x > c$ , which is a contradiction to  $|\mathfrak{Z}^{(m_H)}| \leq c$ .

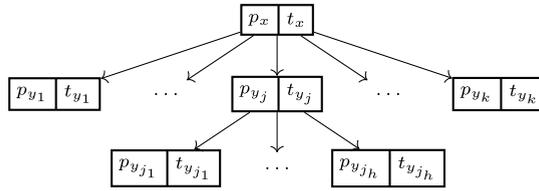
It remains to show the mentioned construction  $\Phi(\circ)$ .

**1.1. Construction of  $\Phi(\circ)$  if  $\pi \in \text{Pos}_{D^e}(t_x)$ :** Let  $\mathfrak{Z}_x$  be a  $\mathcal{P}$ -CT that uses the old ADP at the root node  $x$ , i.e.  $t_x \stackrel{c}{\leftarrow} \mathcal{P} \{p_{y_1} : t_{y_1}, \dots, p_{y_k} : t_{y_k}\}$  using the ADP  $\ell \rightarrow \{p_1 : r_1, \dots, p_k : r_k\}^m$ , the position  $\pi$  with  $\pi \in \text{Pos}_{D^e}(t_x)$ , and a substitution  $\sigma$  such that  $b(t_x|_\pi) = \ell \sigma \in \text{ANF}_{\mathcal{P}}$ . Then  $t_{y_j} = t_x[r_j \sigma]_\pi$  if  $m = \text{true}$ , or  $t_{y_j} = b_\pi^\uparrow(t_x[r_j \sigma]_\pi)$ , otherwise.

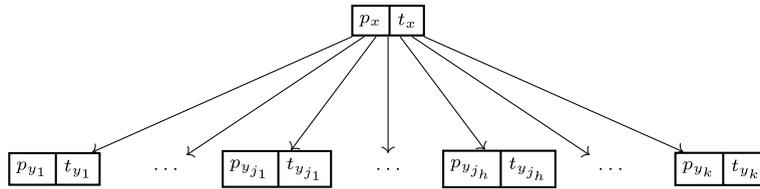
**1.1.1 General construction of  $\Phi(\mathfrak{Z}_x)$ :** Instead of applying only the old ADP at the root  $x$



where we use an arbitrary rewrite step at node  $y_j$  afterwards, we want to directly apply the rewrite rule at position  $\pi.\tau$  of the term  $t_{y_j}$  in our CT, which we performed on  $r_j$  at position  $\tau$  to transform the old into the new ADP, to get

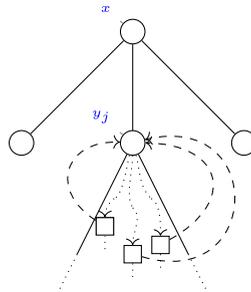


Then, we can contract the edge  $(x, y_j)$  to get



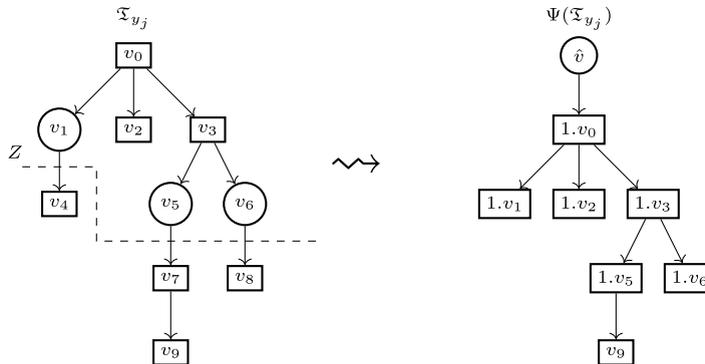
and this is equivalent to applying the new ADP. Note that the rewrite step at position  $\pi.\tau$  may not be an innermost rewrite step in the CT.

The subtrees that start at the nodes  $y_1, \dots, y_{j-1}, y_{j+1}, \dots, y_k$  remain completely the same. We only have to construct a new subtree for node  $y_j$ , i.e., the node that really changed when applying the rewriting processor. To be precise, let  $\mathfrak{T}_{y_j} = \mathfrak{T}_x[y_j E^*]$  be the subtree starting at node  $y_j$ . The construction first creates a new subtree  $\Psi(\mathfrak{T}_{y_j})$  such that (a) and (b) hold w.r.t.  $\mathfrak{T}_{y_j}$  and  $\Psi(\mathfrak{T}_{y_j})$ , and that directly performs the first rewrite step at position  $\pi.\tau$  at the root of the tree, by pushing it from the original position in the tree  $\mathfrak{T}_{y_j}$  to the root. This can be seen in the diagram below.



Again, this push only results in the exact same termination probability due to our restriction that  $\hat{\ell} \rightarrow \{\hat{p}_1 : \hat{r}_1, \dots, \hat{p}_h : \hat{r}_h\}$  is L and NE. Then, by replacing  $\mathfrak{T}_{y_j}$  by  $\Psi(\mathfrak{T}_{y_j})$  in  $\mathfrak{T}_x$  we result in a tree  $\mathfrak{T}'_x$  such that (a) and (b) hold w.r.t.  $\mathfrak{T}_x$  and  $\mathfrak{T}'_x$ , and such that we perform the desired rewrite step at node  $y_j$ . Finally, we contract the edge  $(x, y_j)$  in  $\mathfrak{T}'_x$ , in order to get  $\Phi(\mathfrak{T}_x)$ . Again, (a) and (b) hold w.r.t.  $\mathfrak{T}_x$  and  $\Phi(\mathfrak{T}_x)$ , and we use the new ADP at the root  $x$  in  $\Phi(\mathfrak{T}_x)$ . It only remains to explain the construction of  $\Psi(\mathfrak{T}_{y_j})$ .

**1.1.2. Construction of  $\Psi(\mathfrak{T}_{y_j})$ :** We will move the first rewrite step that takes place at position  $\pi.\tau$  from the original tree  $\mathfrak{T}_{y_j}$  (example on the left below) to the top of the new tree  $\Psi(\mathfrak{T}_{y_j})$  (example on the right below) and show that (a) and (b) both hold after this construction. Below, the circled nodes represent the nodes where we perform a rewrite step at position  $\pi.\tau$ .



We will define the  $\mathcal{P}$ -CT  $\Psi(\mathfrak{T}_{y_j})$  that satisfies the properties (a) and (b) w.r.t.  $\mathfrak{T}_{y_j}$  and  $\Psi(\mathfrak{T}_{y_j})$ , and that directly performs the rewrite step  $t_{y_j} \xrightarrow{\hat{\ell}}_{\mathcal{P}, \pi.\tau, \text{true}} \{\hat{p}_1 : t_{y_{j1}}, \dots, \hat{p}_h : t_{y_{jh}}\}$ , with the rule  $\hat{\ell} \rightarrow \{\hat{p}_1 : \hat{r}_1, \dots, \hat{p}_h : \hat{r}_h\}^{\text{true}} \in \mathcal{P}$ , a substitution  $\hat{\sigma}$ , and the position  $\pi.\tau$ , at the

new root  $\hat{v}$ . Here, we have  $b(t_{y_j} |_{\pi.\tau}) = b(r_j \sigma |_{\tau}) = b(r_j |_{\tau} \sigma) = \hat{\ell} \hat{\sigma}$ . Let  $Z$  be the set of all nodes  $v$  of  $\mathfrak{T}_{y_j}$  where we did not perform a rewrite step at position  $\pi.\tau$  in the path from the root  $x$  to the node  $v$ , or  $v$  is the first node in the path that performs a rewrite step at position  $\pi.\tau$ . In the example we have  $Z = \{v_0, \dots, v_6\} \setminus \{v_4\}$ . For each of these nodes  $z \in Z$  and each  $1 \leq e \leq h$ , we create a new node  $e.z \in V'$  with edges as in  $\mathfrak{T}_{y_j}$  for the nodes in  $Z$ , e.g., for the node  $1.v_3$  we create an edge to  $1.v_5$  and  $1.v_6$ . Furthermore, we add the edges from the new root  $\hat{v}$  to the nodes  $e.y_j$  for all  $1 \leq e \leq h$ . Remember that  $y_j$  was the root in the tree  $\mathfrak{T}_{y_j}$  and has to be contained in  $Z$ . For example, for the node  $\hat{v}$  we create an edge to  $1.v_0$ . For all these new nodes in  $Z$ , we show that the following holds:

- (T1)  $t_z[b(\hat{r}_e)\gamma]_{\pi.\tau} = t'_{e.z}$  for the substitution  $\gamma$  such that  $b(t_z |_{\pi.\tau}) = \hat{\ell} \gamma$
- (T2)  $\text{Pos}_{D^\#}(t_z[b(\hat{r}_e)\gamma]_{\pi.\tau}) \subseteq \text{Pos}_{D^\#}(t'_{e.z})$ .
- (T3)  $p_{e.z}^{\Psi(\mathfrak{T}_{y_j})} = p_z^{\mathfrak{T}_{y_j}} \cdot \hat{p}_e$

Note that we only regard the subtree of the  $j$ -th child of the root. By the prerequisite in the definition of the rewrite processor (there is no annotation below or at position  $\tau$ ), there are no annotations on or below  $\pi.\tau$  for all nodes in  $Z$ .

Now, for a leaf  $e.z \in V'$  either  $z \in V$  is also a leaf (e.g., node  $v_2$ ) or we rewrite the position  $\pi.\tau$  at node  $z$  in  $\mathfrak{T}_{y_j}$  (e.g., node  $v_1$ ). If we rewrite  $t_z \xrightarrow{\hat{\ell}}_{P, \pi.\tau, \text{true}} \{\hat{p}_1 : t_{w_1}, \dots, \hat{p}_h : t_{w_h}\}$ , then we have  $t_{w_e} = t_z[b(\hat{r}_e)\gamma]_{\pi.\tau}$ . Here, we get  $t_{w_e} \stackrel{b}{=} t_z[b(\hat{r}_e)\gamma]_{\pi.\tau} \stackrel{(T1)}{=} t'_{e.z}$ ,  $\text{Pos}_{D^\#}(t_{w_e}) \subseteq \text{Pos}_{D^\#}(t_z[b(\hat{r}_e)\gamma]_{\pi.\tau}) \stackrel{(T2)}{\subseteq} \text{Pos}_{D^\#}(t'_{e.z})$  and  $p_{e.z}^{\Psi(\mathfrak{T}_{y_j})} = p_z^{\mathfrak{T}_{y_j}} \cdot \hat{p}_e \stackrel{(T3)}{=} p_{w_e}^{\mathfrak{T}_{y_j}}$ , and we can again copy the rest of this subtree of  $\mathfrak{T}_{y_j}$  in our newly generated tree  $\Psi(\mathfrak{T}_{y_j})$ . In our example,  $v_1$  has the only successor  $v_4$ , hence we can copy the subtree starting at node  $v_4$ , which is only the node itself, to the node  $1.v_1$  in  $\Psi(\mathfrak{T}_{y_j})$ . For  $v_5$  we have the only successor  $v_7$ , hence we can copy the subtree starting at node  $v_7$ , which is the node itself together with its successor  $v_9$ , to the node  $1.v_5$  in  $\Psi(\mathfrak{T}_{y_j})$ . So essentially, we just have to define the part of the tree before we reach the rewrite step in  $\mathfrak{T}_{y_j}$ , and then, we have to show that (a) and (b) for  $\mathfrak{T}_{y_j}$  and  $\Psi(\mathfrak{T}_{y_j})$  are satisfied. We show the latter first, and then explain the proof that this label gives us indeed a valid  $\mathcal{P}$ -CT.

We start by showing (a) for  $\mathfrak{T}_{y_j}$  and  $\Psi(\mathfrak{T}_{y_j})$ . Let  $u$  be a leaf in  $\Psi(\mathfrak{T}_{y_j})$ . If  $u = e.v$  for some node  $v \in Z$  that is a leaf in  $\mathfrak{T}_{y_j}$  (e.g., node  $1.v_2$ ), then also  $e.v$  must be a leaf in  $\Psi(\mathfrak{T}_{y_j})$  for every  $1 \leq e \leq h$ . Here, we get  $\sum_{1 \leq e \leq h} p_{e.v}^{\Psi(\mathfrak{T}_{y_j})} \stackrel{(T3)}{=} \sum_{1 \leq e \leq h} p_v^{\mathfrak{T}_{y_j}} \cdot \hat{p}_e = p_v^{\mathfrak{T}_{y_j}} \cdot \sum_{1 \leq e \leq h} \hat{p}_e = p_v^{\mathfrak{T}_{y_j}} \cdot 1 = p_v^{\mathfrak{T}_{y_j}}$ . If  $u = e.v$  for some node  $v \in Z$  that is not a leaf in  $\mathfrak{T}_{y_j}$  (e.g., node  $1.v_1$ ), then we know by construction that all successors of  $v$  in  $\mathfrak{T}_{y_j}$  are not contained in  $Z$  and are leaves. Here, we get  $p_{e.v}^{\Psi(\mathfrak{T}_{y_j})} \stackrel{(T3)}{=} p_v^{\mathfrak{T}_{y_j}} \cdot \hat{p}_e = p_w^{\mathfrak{T}_{y_j}}$  for the (unique)  $e$ -th successor  $w$  of  $v$ . Finally, if  $u$  does not have the form  $u = e.v$ , then  $u$  is also a leaf in  $\mathfrak{T}_{y_j}$  with  $p_u^{\Psi(\mathfrak{T}_{y_j})} = p_u^{\mathfrak{T}_{y_j}}$ . Note that these cases cover no leaf of  $\mathfrak{T}_{y_j}$  twice. This implies that we have

$$\begin{aligned}
 & |\Psi(\mathfrak{T}_{y_j})| \\
 &= \sum_{v \in \text{Leaf}^{\Psi(\mathfrak{T}_{y_j})}} p_{e.v}^{\Psi(\mathfrak{T}_{y_j})} \\
 &= \sum_{\substack{e.v \in \text{Leaf}^{\Psi(\mathfrak{T}_{y_j})} \\ v \in Z}} p_{e.v}^{\Psi(\mathfrak{T}_{y_j})} + \sum_{\substack{e.v \in \text{Leaf}^{\Psi(\mathfrak{T}_{y_j})} \\ v \in wE, v \notin Z, w \in Z}} p_{e.v}^{\Psi(\mathfrak{T}_{y_j})} + \sum_{\substack{v \in \text{Leaf}^{\Psi(\mathfrak{T}_{y_j})} \\ v \in \text{Leaf}^{\mathfrak{T}_{y_j}}, v \in wE, w \notin Z}} p_v^{\Psi(\mathfrak{T}_{y_j})} \\
 &\leq \sum_{\substack{v \in \text{Leaf}^{\mathfrak{T}_{y_j}} \\ v \in Z}} \left( \sum_{1 \leq e \leq h} p_{e.v}^{\Psi(\mathfrak{T}_{y_j})} \right) + \sum_{\substack{v \in \text{Leaf}^{\mathfrak{T}_{y_j}}, 1 \leq e \leq h \\ v \in wE, v \notin Z, w \in Z}} p_{e.v}^{\Psi(\mathfrak{T}_{y_j})} + \sum_{\substack{v \in \text{Leaf}^{\mathfrak{T}_{y_j}} \\ v \in wE, w \notin Z}} p_v^{\Psi(\mathfrak{T}_{y_j})} \\
 &\leq \sum_{\substack{v \in \text{Leaf}^{\mathfrak{T}_{y_j}} \\ v \in Z}} p_v^{\mathfrak{T}_{y_j}} + \sum_{\substack{v \in \text{Leaf}^{\mathfrak{T}_{y_j}} \\ v \in wE, v \notin Z, w \in Z}} p_v^{\mathfrak{T}_{y_j}} + \sum_{\substack{v \in \text{Leaf}^{\mathfrak{T}_{y_j}} \\ v \in wE, w \notin Z}} p_v^{\mathfrak{T}_{y_j}} \\
 &= \sum_{v \in \text{Leaf}^{\mathfrak{T}_{y_j}}} p_v^{\mathfrak{T}_{y_j}} \\
 &= |\mathfrak{T}_{y_j}|
 \end{aligned}$$

Next, we show (b) for  $\mathfrak{T}_{y_j}$  and  $\Psi(\mathfrak{T}_{y_j})$ . Let  $p = u_0, u_1, \dots$  be an infinite path in  $\Psi(\mathfrak{T}_{y_j})$  that starts at the root  $\hat{v}$ . If for all  $1 \leq i$  we have  $u_i = e.v_i$  for some node  $v_i \in Z$  and  $1 \leq e \leq h$ , then  $p = v_1, \dots$  is our desired path in  $\mathfrak{T}_{y_j}$ . Otherwise, there is a maximal  $1 \leq o$  such that for all  $1 \leq i \leq o$  we have  $u_i = e.v_i$  for some node  $v_i \in Z$  and  $1 \leq e \leq h$ . Then our desired path is  $v_1, \dots, v_o, w, u_{o+1}, \dots$ . Here,  $w$  is the  $e$ -th successor of  $v_o$  in  $\mathfrak{T}_{y_j}$ . Note that in the first case, we remove one  $\bar{\Gamma}$ -label at the start of our path, while in the other case, we just move an  $\bar{\Gamma}$ -label from the first position to a later one in the path. This shows that (b) is satisfied.

Finally, we prove that  $\Psi(\mathfrak{T}_{y_j})$  is a valid  $\mathcal{P}$ -CT. We only need to prove that the construction satisfies all of our conditions for the nodes in  $Z$ . As the rest of the tree is copied, we can be sure that all Conditions (1)-(3) of a  $\mathcal{P}$ -CT are satisfied. Additionally, due to (b) w.r.t.  $\mathfrak{T}_{y_j}$  and  $\Psi(\mathfrak{T}_{y_j})$ , we get (4) as well, because  $\mathfrak{T}_{y_j}$  satisfies (4).

At the root, after applying the old ADP, we directly perform the rewrite step  $t_{y_j} \xrightarrow{\hat{\ell}}_P \{\hat{p}_1 : t'_{1.y_j}, \dots, \hat{p}_h : t'_{h.y_j}\}$ , with the rule  $\hat{\ell} \rightarrow \{\hat{r}_1 : \hat{r}_1, \dots, \hat{r}_h : \hat{r}_h\}^{\text{true}} \in \mathcal{P}$ , a substitution  $\hat{\sigma}$ , and the position  $\tau$ . Then,  $t_{e.y_j} = t_{y_j}[b(\hat{r}_e)\hat{\sigma}]_{\pi.\tau}$ . Here, the conditions (T1)-(T3) are clearly

satisfied. We now construct the rest of this subtree by mirroring the rewrite steps in the original tree (which is always possible due to the fact that  $\mathcal{U}_{\mathcal{P}}(r_j|_{\tau})$  is NO), and once we encounter the rewrite step that we moved to the top, i.e., once we use a rewrite step at position  $\pi.\tau$ , we skip this rewrite step and directly go on with the  $e$ -th successor if we are in a path that went to the  $e$ -th successor in the initial rewrite step, as described above. In the following, we distinguish between two different cases for a rewrite step at a node  $u$ :

- (A) We use a step with  $\overset{\leftarrow}{c}_p$  in  $\mathfrak{X}_{y_j}$  at a position orthogonal to  $\pi.\tau$ .
- (B) We use a step with  $\overset{\leftarrow}{c}_p$  in  $\mathfrak{X}_{y_j}$  at a position below  $\pi.\tau$ . Note that this is the more interesting case, where we need to use the properties L and NE.

Note that we cannot rewrite above  $\pi.\tau$  before rewriting at position  $\pi.\tau$  due to the innermost restriction.

**(A) If we have**  $t_u \overset{\leftarrow}{c}_p \{ \frac{p_{q_1}}{p_u} : t_{q_1}, \dots, \frac{p_{q_h}}{p_u} : t_{q_h} \}$ , then there is a rule  $\bar{\ell} \rightarrow \{ \bar{p}_1 : \bar{r}_1, \dots, \bar{p}_h : \bar{r}_h \}^m \in \mathcal{P}$ , a substitution  $\delta$ , and a position  $\zeta \in \mathbb{N}^+$  with  $b(t_u|_{\zeta}) = \bar{\ell}\delta \in \text{ANF}_{\mathcal{P}}$ . Furthermore, let  $\zeta \perp \pi.\tau$ . Then, we have  $t'_{e,u}|_{\zeta} \stackrel{b}{=}_{(T1)} t_u[\hat{r}_e\gamma]_{\pi.\tau}|_{\zeta} = t_u|_{\zeta}$  for the substitution  $\gamma$  such that  $t_u|_{\pi.\tau} = \hat{\ell}\gamma$ , and we can rewrite  $t'_{e,u}$  using the same rule, same substitution, and same position. Then (T3) is again satisfied. Furthermore, we have  $t'_{e,q_j} = t'_{e,u}[\bar{r}_e\delta]_{\zeta}$  if  $\zeta \in \text{Pos}_{D^{\#}}(t'_{e,u})$  and  $m = \text{true}$ ,  $t'_{e,q_j} = b_{\zeta}^{\uparrow}(t'_{e,u}[\bar{r}_e\delta]_{\zeta})$  if  $\zeta \in \text{Pos}_{D^{\#}}(t'_{e,u})$  and  $m = \text{false}$ ,  $t'_{e,q_j} = t'_{e,u}[b(\bar{r}_e)\delta]_{\zeta}$  if  $\zeta \notin \text{Pos}_{D^{\#}}(t'_{e,u})$  and  $m = \text{true}$ , or  $t'_{e,q_j} = b_{\zeta}^{\uparrow}(t'_{e,u}[b(\bar{r}_e)\delta]_{\zeta})$  if  $\zeta \notin \text{Pos}_{D^{\#}}(t'_{e,u})$  and  $m = \text{false}$ . In all four cases we get (T1) (using the same substitution  $\gamma$ ) and (T2) as well.

**(B) If we have**  $t_u \overset{\leftarrow}{c}_p \{ \frac{p_{q_1}}{p_u} : t_{q_1}, \dots, \frac{p_{q_h}}{p_u} : t_{q_h} \}$ , then there is a rule  $\bar{\ell} \rightarrow \{ \bar{p}_1 : \bar{r}_1, \dots, \bar{p}_h : \bar{r}_h \}^m \in \mathcal{P}$ , a substitution  $\delta$ , and a position  $\zeta \in \mathbb{N}^+$  with  $b(t_u|_{\zeta}) = \bar{\ell}\delta \in \text{ANF}_{\mathcal{P}}$ . Furthermore, let  $\zeta > \pi.\tau$ . Since  $\hat{\ell} \rightarrow \{ \hat{p}_1 : \hat{r}_1, \dots, \hat{p}_h : \hat{r}_h \}$  is L and NE, we know that  $\hat{\ell}$  contains exactly the same variables as  $\bar{r}_e$  and all of them exactly once. Furthermore, since  $\mathcal{U}_{\mathcal{P}}(r_j|_{\tau})$  is non-overlapping, we know that the rewriting must be completely inside the substitution  $\gamma$  for the substitution  $\gamma$  such that  $t_u|_{\pi.\tau} = \hat{\ell}\gamma$ , i.e., there is a position  $\eta_c$  of a variable  $c$  in  $\hat{\ell}$  and another position  $\chi_c$  with  $\pi.\tau.\eta_c.\chi_c = \zeta$ . Let  $\varphi_e(c)$  be the (unique) variable position of  $c$  in  $\hat{r}_e$ . Then, we have  $t'_{e,u}|_{\pi.\tau.\varphi_e(c).\chi_c} \stackrel{b}{=}_{(T1)} t_u[\hat{r}_e\gamma]_{\pi.\tau}|_{\pi.\tau.\varphi_e(c).\chi_c} = \hat{r}_e\gamma|_{\varphi_e(c).\chi_c} = \gamma(c)|_{\chi_c} = \hat{\ell}\delta|_{\eta_c.\chi_c} = t_u|_{\pi.\tau.\eta_c.\chi_c}$ , and we can rewrite  $t'_{e,u}$  using the same rule, same substitution, and position  $\pi.\tau.\varphi_e(c).\chi_c$ . Again, in all four cases (T1)-(T3) are satisfied.

**1.2.  $\pi \notin \text{Pos}_{D^{\#}}(t_x)$ :** If we have  $\pi \notin \text{Pos}_{D^{\#}}(t_x)$ , then we can simply use the non-annotated old ADP instead of the old one. Since we would remove all annotations in the right-hand side of the rule anyway, due to  $\pi \notin \text{Pos}_{D^{\#}}(t_x)$ , this leads to the same labels in the resulting  $\bar{\mathcal{P}}$ -CT.

**Second Case:** Next, we prove the theorem in the case where  $\mathcal{U}_{\mathcal{P}}(r_j|_{\tau})$  is NO and all rules in  $\mathcal{U}_{\mathcal{P}}(r_j|_{\tau})$  have the form  $\ell' \rightarrow \{ 1 : r' \}$  for some terms  $\ell'$  and  $r'$ .

The proof uses the same idea as in the first case but the construction of  $\Phi(\circ)$  is easier since  $\mathcal{U}_{\mathcal{P}}(r_j|_{\tau})$  is non-probabilistic. First assume that  $\mathcal{U}_{\mathcal{P}}(r_j|_{\tau})$  is not weakly innermost terminating. This means that after using the ADP  $\ell \rightarrow \{ p_1 : r_1, \dots, p_k : r_k \}^m$  as we did in the soundness proof for the first case, every subterm at a position above  $\pi.\tau$  will never be in  $\text{NF}_{\mathcal{P}}$  in the subtree starting at the  $j$ -th successor  $y_j$ . Furthermore, since all rules in  $\mathcal{U}_{\mathcal{P}}(r_j|_{\tau})$  have the form  $\ell' \rightarrow \{ 1 : r' \}$ , we can simply remove all nodes that perform a rewrite step below  $\pi.\tau$ . To be precise, if there is a node  $v$  that performs a rewrite step below position  $\pi.\tau$ , then we have  $t_v \overset{\leftarrow}{c}_p \{ 1 : t_w \}$  for the only successor  $w$  of  $v$ . Here, we have  $t_w \stackrel{b}{=} t_v[r'\sigma]_{\zeta}$  for the used substitution  $\sigma$  and position  $\zeta$  below  $\pi.\tau$ . The construction  $\Psi(\mathfrak{X}_{y_j})$  contracts all edges  $(x, y)$  where we use a rewrite step at a position below  $\pi.\tau$ . This only removes  $\bar{\Gamma}$ -nodes, as there is no annotation below or at position  $\pi.\tau$ . Furthermore, we adjust the labeling such that the subterm at position  $\pi.\tau$  remains the same for the whole CT. Finally, we exchange the rewrite step at the root  $x$  from using the old ADP to using the new ADP. Since all subterms at a position above  $\pi.\tau$  will never be reduced to normal forms in the original CT, it does not matter which subterms really occur. It is easy to see that  $\Phi(\mathfrak{X}_x)$  is a valid  $\mathcal{P}$ -CT and that (a) and (b) hold w.r.t.  $\mathfrak{X}_x$  and  $\Phi(\mathfrak{X}_x)$ , and this ends the proof if  $\mathcal{U}_{\mathcal{P}}(r_j|_{\tau})$  is not weakly innermost terminating.

If  $\mathcal{U}_{\mathcal{P}}(r_j|_{\tau})$  is weakly innermost terminating, then it follows directly that it is also confluent and terminating [60, Thm. 3.2.11]. We can use the construction  $\Psi(\circ)$  from the first case to iteratively push the next innermost rewrite step that is performed below position  $\pi.\tau$  to a higher position in the tree, until we reach the node that performs the rewrite step at position  $\pi.\tau$ . Note that we do not need the conditions L or NE for the used rule here, because only Case (A) of the construction can happen. There is no rewrite step below possible (Case (B)), since we move an innermost rewrite step further up and there is no rewrite step above possible, since before doing this construction it was a valid innermost  $\mathcal{P}$ -CT. Since  $\mathcal{U}_{\mathcal{P}}(r_j|_{\tau})$  is terminating, this construction ends after a finite number of  $\Psi(\circ)$ -applications in a tree  $\mathfrak{X}_{y_j}^{(\infty)}$ . Now  $\mathfrak{X}_{y_j}^{(\infty)}$  first rewrites below or at position  $\pi.\tau$  until it is a normal form. Since all rules in  $\mathcal{U}_{\mathcal{P}}(r_j|_{\tau})$  have the form  $\ell' \rightarrow \{ 1 : r' \}$ , these rewrite steps are a single path in  $\mathfrak{X}_{y_j}^{(\infty)}$ , and it does not matter how long the path is. Furthermore, it does not matter which rewrite strategy we use, since  $\mathcal{U}_{\mathcal{P}}(r_j|_{\tau})$  is confluent. We will always reach the same normal form at the end of this path. Hence, we can replace the steps with the old ADP and the corresponding innermost rewrite steps in the CT by a step with the new ADP (where the rewriting does not necessarily correspond to the innermost strategy), i.e., we move this non-innermost step directly to the point where the ADP is applied. We will reach the same normal form and can copy the rest of the tree again.

**Third Case:** Finally, we prove the theorem in the case where  $\mathcal{U}_{\mathcal{P}}(r_j|_{\tau})$  is NO,  $r_j|_{\tau}$  is a ground term, and we have  $D_j \overset{\leftarrow}{c}_{p,\pi,\tau,\text{true}} \{ q_1 : E_1, \dots, q_h : E_h \}$ , i.e., it is an innermost step.

Once again, we use the same idea as in the proof for the second case but the construction of  $\Phi(\circ)$  is, again, easier. Note that if  $\mathcal{U}_{\mathcal{P}}(r_j|\tau)$  is non-overlapping,  $r_j$  contains no variable below position  $\tau$ , and we perform an innermost rewrite step, then this is always an innermost rewrite step in every possible CT and this is the only possible rewrite step at this position. Hence, we can move this innermost step directly after the use of the ADP using the construction  $\Psi(\circ)$ . Again, here only Case (A) can happen. The reason is that we have to perform this rewrite step eventually if we want to rewrite above position  $\tau$ , and all other rewrite steps that we can perform in such a situation would be at orthogonal positions. So we get the same normal forms in the leaves with the same probability.  $\square$

For the proofs of the instantiation processors, we first show some basic properties of  $Cap_{\mathcal{P}}$  and  $Ren$ , similar to the known properties of the non-probabilistic versions of  $Cap_{\mathcal{P}}$  and  $Ren$ .

**Lemma 6** (Property of  $Cap_{\mathcal{P}}$  and  $Ren$ ). *Let  $t, s \in \mathcal{T}^{\#}$ .*

1. If  $t\sigma \xrightarrow{\text{np}(\mathcal{P})}^* s$  for some substitution  $\sigma$  that maps each variable to a normal form, i.e.,  $\sigma(x) \in \text{NF}_{\mathcal{P}}$  for all  $v \in \mathcal{V}$ , then  $s = Cap_{\mathcal{P}}(t)\delta$  for some substitution  $\delta$  which only differs from  $\sigma$  on the fresh variables that are introduced by  $Cap_{\mathcal{P}}$ .
2. If  $t\sigma \xrightarrow{\text{np}(\mathcal{P})}^* s$  for some substitution  $\sigma$ , then  $s = Ren(Cap_{\mathcal{P}}(t))\delta$  for some substitution  $\delta$  which only differs from  $\sigma$  on the fresh variables that are introduced by  $Ren$ .

**Proof.**

1. Let  $t\sigma \xrightarrow{\text{np}(\mathcal{P}, \rho_1)} s_1 \xrightarrow{\text{np}(\mathcal{P}, \rho_2)} \dots \xrightarrow{\text{np}(\mathcal{P}, \rho_n)} s_n = s$ , where  $\xrightarrow{\text{np}(\mathcal{P}, \rho)}$  again denotes a rewrite step at position  $\rho$ . Let  $\{\pi_1, \dots, \pi_m\}$  be the set of positions where  $Cap_{\mathcal{P}}$  replaces the subterms of  $t$  by corresponding fresh variables  $x_1, \dots, x_m$ . Note that the rewrite steps cannot be “completely inside” the used substitution  $\sigma$ , as it maps every variable to a normal form. Hence, by the definition of  $Cap_{\mathcal{P}}$ , for each  $\rho_i$  there is a higher position  $\pi_j \leq \rho_i$ . Thus,  $s$  can at most differ from  $t\sigma$  on positions below a  $\pi_j$ . We define  $\delta$  to be like  $\sigma$  but on the fresh variables  $x_1, \dots, x_m$  we define  $\delta(x_j) = s|_{\pi_j}$ . Then by construction  $Cap_{\mathcal{P}}(t)\delta = s$  and  $\delta$  and  $\sigma$  differ only on the fresh variables.
2. Let  $t\sigma \xrightarrow{\text{np}(\mathcal{P}, \rho_1)} s_1 \xrightarrow{\text{np}(\mathcal{P}, \rho_2)} \dots \xrightarrow{\text{np}(\mathcal{P}, \rho_n)} s_n = s$ . Let  $\{\pi_1, \dots, \pi_m\}$  be the set of positions where  $Ren \circ Cap_{\mathcal{P}}$  replaces the subterms of  $t$  by corresponding fresh variables  $x_1, \dots, x_m$ . By the definition of  $Ren$  and  $Cap_{\mathcal{P}}$ , for each  $\rho_i$  there is a higher position  $\pi_j \leq \rho_i$ . (If the rewrite step is “completely inside” the substitution  $\sigma$ , then  $Ren$  replaces the corresponding variable in  $t$  by a fresh one, otherwise  $Cap_{\mathcal{P}}$  replaces the subterm in  $t$  by a new variable, which is again replaced by  $Ren$  afterwards.) Hence,  $s$  can at most differ from  $t\sigma$  on positions below a  $\pi_j$ . We define  $\delta$  to be like  $\sigma$  but on the fresh variables  $x_1, \dots, x_m$  we define  $\delta(x_j) = s|_{\pi_j}$ . Then by construction  $Ren(Cap_{\mathcal{P}}(t))\delta = s$ , and  $\delta$  and  $\sigma$  differ only on the fresh variables.

$\square$

Next, we prove soundness and completeness of the instantiation processor for both innermost and full rewriting. There are only small changes in the proof due to the rewrite strategy that we highlight once they occur. The same holds for the following forward instantiation processor.

**Proof.** (Proof of both [Theorem 16](#) and [17](#)) We will use the following two observations. As in the proof of [Theorem 15](#), let  $\mu = \{p_1 : r_1, \dots, p_k : r_k\}$ ,  $\overline{\mathcal{P}}' = \mathcal{P}' \cup N \cup \{\ell \rightarrow \{p_1 : b(r_1), \dots, p_k : b(r_k)\}^m\}$ , and  $\overline{\mathcal{P}} = \overline{\mathcal{P}}' \cup \{\ell \rightarrow \{p_1 : r_1, \dots, p_k : r_k\}^m\}$ . First, note that  $\text{ANF}_{\mathcal{P}} = \text{ANF}_{\overline{\mathcal{P}}} = \text{ANF}_{\overline{\mathcal{P}'}}$ , since the left-hand sides in  $\overline{\mathcal{P}}$  and  $\overline{\mathcal{P}'}$  are either already from rules in  $\mathcal{P}$  or instantiated left-hand sides from  $\mathcal{P}$ . So it suffices to consider only  $\text{ANF}_{\mathcal{P}}$ . Second, assume that there exists a  $\overline{\mathcal{P}}$ -CT  $\mathfrak{Z}$  that converges with probability  $< 1$  whose root is labeled with  $(1 : t)$  and  $b(t) = s\theta$  for a substitution  $\theta$  and an ADP  $s \rightarrow \dots \in \overline{\mathcal{P}}$ , and  $\text{Pos}_{D^{\#}}(t) = \{\varepsilon\}$ . If in addition, the ADP  $\ell \rightarrow \{p_1 : r_1, \dots, p_k : r_k\}^m$  may only be used at the root, then we know that not only  $\overline{\mathcal{P}}$  is not AST but also that  $\overline{\mathcal{P}'}$  is not AST. The reason is that if only the root  $v$  uses the ADP  $\ell \rightarrow \{p_1 : r_1, \dots, p_k : r_k\}^m$  then all the subtrees starting at one of its direct successors  $vE = \{w_1, \dots, w_k\}$  are  $\overline{\mathcal{P}'}$ -CTs. Furthermore, since  $\mathfrak{Z}$  converges with probability  $< 1$ , there must be at least one subtree  $\mathfrak{Z}[w_i E^*]$  that starts at the node  $w_i$  for some  $1 \leq i \leq k$  and also converges with probability  $< 1$ .

**Soundness:** Let  $\mathcal{P}$  be not (i)AST. Then by [Lemma 2](#) there exists an (innermost)  $\mathcal{P}$ -CT  $\mathfrak{Z} = (V, E, L)$  that converges with probability  $< 1$  whose root is labeled with  $(1 : t)$  and  $b(t) = s\theta$  for a substitution  $\theta$  and an ADP  $s \rightarrow \dots \in \mathcal{P}$ , and  $\text{Pos}_{D^{\#}}(t) = \{\varepsilon\}$ . We will now create an (innermost)  $\overline{\mathcal{P}}$ -CT  $\mathfrak{Z}' = (V, E, L')$ , with the same underlying tree structure, and an adjusted labeling such that  $p_v^{\mathfrak{Z}'} = p_v^{\mathfrak{Z}}$  for all  $v \in V$ . Furthermore, we will at most use the ADP  $\ell \rightarrow \{p_1 : r_1, \dots, p_k : r_k\}^m$  at the root. Since the tree structure and the probabilities are the same, we then get  $|\mathfrak{Z}'| = |\mathfrak{Z}| < 1$ , and hence, using our previous discussion,  $\overline{\mathcal{P}'}$  is not (i)AST either.

The core idea is that every rewrite step with  $\ell \rightarrow \{p_1 : r_1, \dots, p_k : r_k\}^m$  at a node  $v$  that is not the root can also be done with a rule from  $N$ , or we can use  $\ell \rightarrow \{p_1 : b(r_1), \dots, p_k : b(r_k)\}^m$  if the annotations do not matter, e.g., we rewrite at a position that is not annotated.

We only consider AST in the following construction. If not mentioned otherwise, the proof for iAST works analogously just by considering innermost rewrite steps. We construct the new labeling  $L'$  for the  $\overline{\mathcal{P}}$ -CT  $\mathfrak{Z}'$  inductively such that for all nodes  $v \in V \setminus \text{Leaf}$  with children nodes  $vE = \{w_1, \dots, w_h\}$  we have  $t'_v \xrightarrow{\mathcal{P}} \{ \frac{p_{w_1}}{p_v} : t'_{w_1}, \dots, \frac{p_{w_h}}{p_v} : t'_{w_h} \}$  and for all non-root nodes in  $\Gamma$  we even have  $t'_v \xrightarrow{\mathcal{P}'}$   $\{ \frac{p_{w_1}}{p_v} : t'_{w_1}, \dots, \frac{p_{w_h}}{p_v} : t'_{w_h} \}$ . Let  $W \subseteq V$  be the set of nodes  $v$  where we have already defined the labeling  $L'(x)$ . During our construction, we ensure that for every node  $v \in W$  we have

$$t'_v \stackrel{b}{=} t'_v \text{ and } \text{Pos}_{D^{\#}}(t'_v) \subseteq \text{Pos}_{D^{\#}}(t'_v). \tag{B.14}$$

This means that the corresponding term  $t_v$  for the node  $v$  in  $\mathfrak{X}$  has the same structure as the term  $t'_v$  in  $\mathfrak{X}'$ , and additionally, every annotation in  $t_v$  also exists in  $t'_v$ . The second condition ensures that if we rewrite using Case (af) or (at) in  $\mathfrak{X}$ , we do the same in  $\mathfrak{X}'$ , i.e., the corresponding node  $v$  remains in  $\Gamma$  in  $\mathfrak{X}'$ . We label the root of  $\mathfrak{X}'$  by  $(1 : t)$ . Here, (B.14) obviously holds. As long as there is still an inner node  $v \in W$  such that its successors are not contained in  $W$ , we do the following. Let  $vE = \{w_1, \dots, w_h\}$  be the set of its successors. We need to define the corresponding terms  $t'_{w_1}, \dots, t'_{w_m}$  for the nodes  $w_1, \dots, w_h$ . Since  $v$  is not a leaf and  $\mathfrak{X}$  is a  $\mathcal{P}$ -CT, we have  $t_v \hookrightarrow_{\mathcal{P}} \{\frac{p_{w_1}}{p_v} : t_{w_1}, \dots, \frac{p_{w_h}}{p_v} : t_{w_h}\}$ . We have the following three different cases:

(A) We have  $t_v \hookrightarrow_{\mathcal{P}} \{\frac{p_{w_1}}{p_v} : t_{w_1}, \dots, \frac{p_{w_h}}{p_v} : t_{w_h}\}$  with an ADP  $\ell' \rightarrow \{p_1 : r'_1, \dots, p_h : r'_h\}^m$  that is either different to  $\ell \rightarrow \{p_1 : r_1, \dots, p_k : r_k\}^m$  or the node  $v$  is the root, using the position  $\pi$ , a VRF  $(\varphi_j)_{1 \leq j \leq k}$ , and a substitution  $\sigma$  such that  $b(t_v|_{\pi}) = \ell\sigma$ . Then,  $t_v \stackrel{b}{=}_{(IH)} t'_v$ , and hence also  $b(t'_v|_{\pi}) = \ell\sigma$ . Thus, we can rewrite the term  $t'_v$  using the same ADP, the same position, the same substitution, and the same VRF. This means that we have  $t'_v \hookrightarrow_{\overline{\mathcal{P}}} \{\frac{p_{w_1}}{p_v} : t'_{w_1}, \dots, \frac{p_{w_h}}{p_v} : t'_{w_h}\}$ . Let  $1 \leq j \leq h$ . If  $\pi \in \text{Pos}_{D^\#}(t_v)$ , then also  $\pi \in \text{Pos}_{D^\#}(t'_v)$  by (B.14). Whenever we create annotations in the rewrite step in  $\mathfrak{X}$  (a step with (af) or (at)), then we do the same in  $\mathfrak{X}'$  (the step is also an (af)- or (at)-step, respectively), and whenever we remove annotations in the rewrite step in  $\mathfrak{X}'$  (a step with (nf) or (nt)), then we do the same in  $\mathfrak{X}$  (the step is also either a (nt)- or (nf)-step). Therefore, we also get  $\text{Pos}_{D^\#}(t_{w_j}) \subseteq \text{Pos}_{D^\#}(t'_{w_j})$  for all  $1 \leq j \leq k$  and (B.14) is again satisfied.

(B) We have  $t_v \hookrightarrow_{\mathcal{P}} \{\frac{p_{w_1}}{p_v} : t_{w_1}, \dots, \frac{p_{w_k}}{p_v} : t_{w_k}\}$  using the ADP  $\ell \rightarrow \{p_1 : r_1, \dots, p_k : r_k\}^m$ , the position  $\pi$ , a VRF  $(\varphi_j)_{1 \leq j \leq k}$ , and a substitution  $\sigma$  such that  $b(t_v|_{\pi}) = \ell\sigma$ , and  $\pi \notin \text{Pos}_{D^\#}(t_v)$ . Since  $t_v \stackrel{b}{=}_{(IH)} t'_v$ , we can rewrite the term  $t'_v$  using the ADP  $\ell \rightarrow \{p_1 : b(r_1), \dots, p_k : b(r_k)\}^m$ , the same position, the same substitution, and the same VRF. This means that we have  $t'_v \hookrightarrow_{\overline{\mathcal{P}}} \{\frac{p_{w_1}}{p_v} : t'_{w_1}, \dots, \frac{p_{w_k}}{p_v} : t'_{w_k}\}$  and (B.14) is again satisfied. In order to prove this, one can do a similar analysis as above. Note that only cases (nf) and (nt) can be applied in  $\mathfrak{X}$  (since  $\pi \notin \text{Pos}_{D^\#}(t_v)$ ), and we would remove the annotations of the terms  $r_j$  anyway in those cases.

(C) Finally, we have  $t_v \hookrightarrow_{\mathcal{P}} \{\frac{p_{w_1}}{p_v} : t_{w_1}, \dots, \frac{p_{w_k}}{p_v} : t_{w_k}\}$  using the ADP  $\ell \rightarrow \{p_1 : r_1, \dots, p_k : r_k\}^m$ , the position  $\pi$ , a VRF  $(\varphi_j)_{1 \leq j \leq k}$ , and a substitution  $\sigma$  such that  $b(t_v|_{\pi}) = \ell\sigma$ ,  $\pi \in \text{Pos}_{D^\#}(t_v)$ , and  $v$  is not the root. Then  $t_{w_j} = b_{\pi}^{\dagger}(t_v[\#_{\text{Pos}_{D^\#}(r_j)} \cup \Psi_j(r_j\sigma)]_{\pi})$  if  $m = \text{false}$  and  $t_{w_j} = t_v[\#_{\text{Pos}_{D^\#}(r_j)} \cup \Psi_j(r_j\sigma)]_{\pi}$ , otherwise.

We now look at the (not necessarily direct) predecessor  $z$  of  $v$  that is in  $\Gamma$ , where the ADP is applied that is used to create the annotation of the function symbol at position  $\pi$ . There is always such a node  $z$ . The reason is that  $v$  is not the first node in  $\Gamma$  and by the Starting Lemma (Lemma 2) we can assume that a step at the root of the term takes place at the root of the CT, i.e.,  $t_v$  results from right-hand sides of  $\mathcal{P}$ . Furthermore, we only use rules with the flag  $m = \text{true}$  to evaluate the arguments of the annotated function symbol as otherwise, the position  $\pi$  would not be annotated in  $t_v$ . We show that the path from  $z$  to  $v$  can also be taken when using one of the new instantiations of  $\ell \rightarrow \{p_1 : r_1, \dots, p_k : r_k\}^m$  instead.

Let this predecessor  $z \in \Gamma$  use the ADP  $\ell' \rightarrow \{p'_1 : r'_1, \dots, p'_h : r'_h\}^m$ , the position  $\pi'$ , and the substitution  $\sigma'$ . Furthermore, let  $1 \leq j \leq h$  and  $\tau$  be the position of  $r'_j$  such that  $\#_{\epsilon}(r'_j|_{\tau})\sigma' \rightarrow_{\text{np}(\mathcal{P})}^* \#_{\epsilon}(t_v|_{\pi}) = \#_{\epsilon}(\ell)\sigma$ , which exists by the previous paragraph.



The following consideration now differs for iAST and AST due to the different approximations (iAST uses only  $\text{Cap}_p$  and AST uses  $\text{Ren}o\text{Cap}_p$ ).

- **AST:** From  $\#_{\epsilon}(r'_j|_{\tau})\sigma' \rightarrow_{\text{np}(\mathcal{P})}^* \#_{\epsilon}(\ell)\sigma$  together with Lemma 6 we get  $\#_{\epsilon}(\ell)\sigma = \text{Ren}(\text{Cap}_p(\#_{\epsilon}(r'_j|_{\tau})))\delta$  for some substitution  $\delta$  that differs from  $\sigma'$  at most on the variables that are introduced by  $\text{Ren}$ . W.l.o.g., we can assume that  $\sigma'$  is equal to  $\delta$  on all these fresh variables, and since the ADPs are variable-renamed, we can also assume that  $\sigma$  is equal to  $\delta$  on all the fresh variables and all the variables from  $\ell'$ . Hence,  $\#_{\epsilon}(\ell)\delta = \text{Ren}(\text{Cap}_p(\#_{\epsilon}(r'_j|_{\tau})))\delta$  implies that we can find an mgu  $\gamma$  of  $\#_{\epsilon}(\ell)$  and  $\text{Ren}(\text{Cap}_p(\#_{\epsilon}(r'_j|_{\tau})))$  with  $\sigma = \gamma\zeta$  for some substitution  $\zeta$ . Hence, we can apply the new ADP  $\ell\gamma \rightarrow \{p_1 : r_1\gamma, \dots, p_k : r_k\gamma\}^m \in N$  with the position  $\pi$ , the same VRF, and the substitution  $\zeta$ . This means that we have  $t'_v \hookrightarrow_{\mathcal{P}} \{\frac{p_{w_1}}{p_v} : t'_{w_1}, \dots, \frac{p_{w_k}}{p_v} : t'_{w_k}\}$  with  $t'_{w_j} = t'_v[\#_{\text{Pos}_{D^\#}(r_j)} \cup \Psi_j(r_j\gamma\zeta)]_{\pi}$  if  $m = \text{true}$ , or  $t'_{w_j} = b_{\pi}^{\dagger}(t'_v[\#_{\text{Pos}_{D^\#}(r_j)} \cup \Psi_j(r_j\gamma\zeta)]_{\pi})$ , otherwise. Since  $\sigma = \gamma\zeta$  we directly get  $t_{w_j} = t'_{w_j}$  and  $\text{Pos}_{D^\#}(t_{w_j}) \subseteq \text{Pos}_{D^\#}(t'_{w_j})$  so that (B.14) is satisfied again.
- **iAST:** From  $\#_{\epsilon}(r'_j|_{\tau})\sigma' \xrightarrow{\text{np}(\mathcal{P})}^* \#_{\epsilon}(\ell)\sigma$  and the fact that  $\sigma'$  maps every variable to a normal form due to innermost rewriting, Lemma 6 implies  $\#_{\epsilon}(\ell)\sigma = \text{Cap}_p(\#_{\epsilon}(r'_j|_{\tau}))\delta$  for some substitution  $\delta$  that differs from  $\sigma'$  at most on the variables that are introduced by  $\text{Cap}_p$ . W.l.o.g., we can assume that  $\sigma'$  is equal to  $\delta$  on all these fresh variables, and since the ADPs are variable-renamed, we can also assume that  $\sigma$  is equal to  $\delta$  on all the fresh variables and all the variables from  $\ell'$ . Hence,  $\#_{\epsilon}(\ell)\delta = \text{Cap}_p(\#_{\epsilon}(r'_j|_{\tau}))\delta$  shows there is an mgu  $\gamma$  of  $\#_{\epsilon}(\ell)$  and  $\text{Cap}_p(\#_{\epsilon}(r'_j|_{\tau}))$  with  $\sigma = \gamma\zeta$  for some substitution  $\zeta$ . Moreover, the property  $\{\ell'\sigma', \ell\sigma\} \subseteq \text{ANF}_p$  must remain true when replacing  $\sigma$  and  $\sigma'$  by the more general substitution  $\gamma$ , i.e.,  $\{\ell'\gamma, \ell\gamma\} \subseteq \text{ANF}_p$ . The rest is analogous to the AST case.

**Completeness:** Let  $\overline{\mathcal{P}'}$  be not (i)AST. Then there exists an (innermost)  $\overline{\mathcal{P}'}$ -CT  $\mathfrak{X} = (V, E, L)$  that converges with probability  $< 1$ . We will now create an (innermost)  $\mathcal{P}$ -CT  $\mathfrak{X}' = (V, E, L')$ , with the same underlying tree structure, and an adjusted labeling such that

$p_v^{\mathfrak{Z}} = p_v^{\mathfrak{Z}'}$  for all  $v \in V$ . Since the tree structure and the probabilities are the same, we then get  $|\mathfrak{Z}'| = |\mathfrak{Z}|$ , and thus  $\mathcal{P}$  is not (i)AST either.

The core idea of this construction is that every rewrite step with an ADP from  $N$  or the ADP  $\ell \rightarrow \{p_1 : b(r_1), \dots, p_k : b(r_k)\}^m$  is also possible with the more general ADP  $\ell \rightarrow \{p_1 : r_1, \dots, p_k : r_k\}^m$  that may also contain more annotations. As for soundness, we focus on AST, but the proof is analogous for iAST. We construct the new labeling  $L'$  for the  $\mathcal{P}$ -CT  $\mathfrak{Z}'$  inductively such that for all inner nodes  $v \in V \setminus \text{Leaf}$  with children nodes  $vE = \{w_1, \dots, w_m\}$  we have  $t'_v \hookrightarrow_{\mathcal{P}} \{\frac{p_{w_1}}{p_v} : t'_{w_1}, \dots, \frac{p_{w_m}}{p_v} : t'_{w_m}\}$ . Let  $W \subseteq V$  be the set of nodes  $v$  where we have already defined the labeling  $L'(x)$ . During our construction, we ensure that for every  $v \in W$  (analogous to the soundness construction) we have

$$t_v \stackrel{b}{=} t'_v \text{ and } \text{Pos}_{\mathcal{D}^\#}(t_v) \subseteq \text{Pos}_{\mathcal{D}^\#}(t'_v). \tag{B.15}$$

We label the root of  $\mathfrak{Z}'$  exactly as the root of  $\mathfrak{Z}$ . Here, (B.15) obviously holds. As long as there is still an inner node  $v \in W$  such that its successors are not contained in  $W$ , we do the following. Let  $vE = \{w_1, \dots, w_h\}$  be the set of its successors. We need to define the corresponding terms  $t'_{w_1}, \dots, t'_{w_h}$  for the nodes  $w_1, \dots, w_h$ . Since  $v$  is not a leaf and  $\mathfrak{Z}$  is a  $\overline{\mathcal{P}'}$ -CT, we have  $t_v \hookrightarrow_{\overline{\mathcal{P}'}} \{\frac{p_{w_1}}{p_v} : t_{w_1}, \dots, \frac{p_{w_h}}{p_v} : t_{w_h}\}$ . We have the following three different cases:

- (A) If it is a step with  $\hookrightarrow_{\mathcal{P}}$  using an ADP from  $\mathcal{P}'$  in  $\mathfrak{Z}$ , then we perform a rewrite step with the same ADP, the same position, the same substitution, and the same VRF in  $\mathfrak{Z}'$ . This is analogous to Case (A) of the soundness proof.
- (B) If it is a step with  $\hookrightarrow_{\mathcal{P}}$  using the ADP  $\ell \rightarrow \{p_1 : b(r_1), \dots, p_k : b(r_k)\}^m$  in  $\mathfrak{Z}$ , then we use the ADP  $\ell \rightarrow \{p_1 : r_1, \dots, p_k : r_k\}^m$  that contains more annotations in  $\mathfrak{Z}'$ , at the same position, using the same substitution, and the same VRF. Since we use the same rule but with more annotations, we end up with  $t_{w_j} \stackrel{b}{=} t'_{w_j}$  and  $\text{Pos}_{\mathcal{D}^\#}(t_{w_j}) \subseteq \text{Pos}_{\mathcal{D}^\#}(t'_{w_j})$  again.
- (C) If it is a step with  $\hookrightarrow_{\mathcal{P}}$  using an ADP  $\ell\gamma \rightarrow \{p_1 : r_1\gamma, \dots, p_k : r_k\gamma\}^m$  from  $N$  in  $\mathfrak{Z}$  and a substitution  $\delta$ , then we use the more general ADP  $\ell \rightarrow \{p_1 : r_1, \dots, p_k : r_k\}^m$  in  $\mathfrak{Z}'$ , at the same position, and the same VRF. Furthermore, we use the substitution  $\gamma\delta$ .

□

**Proof.** (Proof of both Theorem 18 and 19) As in the proof of Theorem 16, let  $\mu = \{p_1 : r_1, \dots, p_k : r_k\}$ ,  $\overline{\mathcal{P}'} = \mathcal{P}' \cup N \cup \{\ell \rightarrow \{p_1 : b(r_1), \dots, p_k : b(r_k)\}^m\}$ , and  $\overline{\mathcal{P}} = \overline{\mathcal{P}'} \cup \{\ell \rightarrow \{p_1 : r_1, \dots, p_k : r_k\}^m\}$ .

**Soundness:** The core proof idea, the construction itself, and even the induction hypothesis are completely the same as for the soundness proof of Theorem 16, only the third case (C) changes.

(C) If we have  $t_v \hookrightarrow_{\mathcal{P}} \{\frac{p_{w_1}}{p_v} : t_{w_1}, \dots, \frac{p_{w_k}}{p_v} : t_{w_k}\}$  using the ADP  $\ell \rightarrow \{p_1 : r_1, \dots, p_k : r_k\}^m$ , the position  $\pi \in \text{Pos}_{\mathcal{D}^\#}(s)$ , and a substitution  $\sigma$  such that  $b(t_{v|\pi}) = \ell\sigma$  (where  $\ell\sigma \in \text{ANF}_{\mathcal{P}}$  in the innermost case), and  $v$  is not the root, then  $t_{w_j} = b_{\pi}^{\uparrow}(t_v[\#_{\text{Pos}_{\mathcal{D}^\#}(r_j)} \cup \Psi_j(r_j\sigma)]_{\pi})$  if  $m = \text{false}$  and  $t_{w_j} = t_v[\#_{\text{Pos}_{\mathcal{D}^\#}(r_j)} \cup \Psi_j(r_j\sigma)]_{\pi}$ , otherwise.

First consider the case, where there exists no successor  $z$  of  $v$  where an ADP is applied at an annotated subterm introduced by some right-hand side  $r_j$  for some  $1 \leq j \leq k$ . Then, we can use  $\ell \rightarrow \{p_1 : b(r_1), \dots, p_k : b(r_k)\}^m$  instead, because the annotations will never be used.

Otherwise, there exists a successor  $z$  of  $v$  where an ADP is applied at an annotated subterm introduced by the right-hand side  $r_j$  for some  $1 \leq j \leq k$ . We show that the path to this successor can also be taken when using one of the new instantiations of  $\ell \rightarrow \{p_1 : r_1, \dots, p_k : r_k\}^m$  instead.<sup>17</sup> At node  $z$  we use an ADP  $\ell' \rightarrow \{p'_1 : r'_1, \dots, p'_h : r'_h\}^{m'}$  and the substitution  $\sigma'$ . Moreover, let  $\tau$  be the position of  $r_j$  with  $\#_{\epsilon}(r_j|_{\tau})\sigma \rightarrow_{\text{np}(\mathcal{P})}^* \#_{\epsilon}(\ell')\sigma'$ .



Again we have two different cases:

- **AST:** Let  $Q' = \text{np}(\mathcal{P})$  and let  $Q = Q'^{-1}$ . Note that  $\#_{\epsilon}(r_j|_{\tau})\sigma \rightarrow_{Q'}^* \#_{\epsilon}(\ell')\sigma'$  and hence,  $\#_{\epsilon}(\ell')\sigma' \rightarrow_Q^* \#_{\epsilon}(r_j|_{\tau})\sigma$ . By Lemma 6 one directly obtains  $\#_{\epsilon}(r_j|_{\tau})\sigma = \text{Ren}(\text{Cap}_Q(\#_{\epsilon}(\ell')))\delta$  for some substitution  $\delta$  that differs from  $\sigma'$  at most on the variables that are introduced by  $\text{Ren}$ . W.l.o.g. we can assume that  $\sigma'$  is equal to  $\delta$  on all these fresh variables, and since the ADPs are variable-renamed, we can also assume that  $\sigma$  is equal to  $\delta$  on all the fresh variables and all the variables from  $\#_{\epsilon}(\ell')$ . Hence,  $\#_{\epsilon}(r_j|_{\tau})\sigma = \text{Ren}(\text{Cap}_Q(\#_{\epsilon}(\ell')))\sigma$  shows there is an mgu  $\gamma$  of  $\#_{\epsilon}(r_j|_{\tau})\sigma$  and  $\text{Ren}(\text{Cap}_Q(\#_{\epsilon}(\ell')))$  with  $\sigma = \gamma\zeta$  for some substitution  $\zeta$ . Hence, we can apply the new ADP  $\ell\gamma \rightarrow \{p_1 : r_1\gamma, \dots, p_k : r_k\gamma\}^m \in N$  with the position  $\pi$  and the substitution  $\zeta$ . This means that we have  $t'_v \hookrightarrow_{\mathcal{P}} \{\frac{p_{w_1}}{p_v} : t'_{w_1}, \dots, \frac{p_{w_k}}{p_v} : t'_{w_k}\}$  with  $t'_{w_j} = b_{\pi}^{\uparrow}(t'_v[\#_{\text{Pos}_{\mathcal{D}^\#}(r_j)} \cup \Psi_j(r_j\gamma\zeta)]_{\pi})$  if  $m = \text{true}$ , or  $t'_{w_j} = t'_v[\#_{\text{Pos}_{\mathcal{D}^\#}(r_j)} \cup \Psi_j(r_j\gamma\zeta)]_{\pi}$ , otherwise. Since,  $\sigma = \gamma\zeta$  we directly get  $t_{w_j} \stackrel{b}{=} t'_{w_j}$  and  $\text{Pos}_{\mathcal{D}^\#}(t_{w_j}) \subseteq \text{Pos}_{\mathcal{D}^\#}(t'_{w_j})$  so that (B.14) is satisfied again

<sup>17</sup> It suffices to consider only one of those successors  $v$  of  $v$  to find an instantiation of the ADP that could be used instead of the old ADP in order to reach all such successors. The reason is that this instantiation is equal or more general than the actual concrete instantiation of the ADP that was used to perform the ADP-step in the actual CT. Hence, every other such successor can also be used with the more general instantiation of the ADP.

- **iAST:** Let  $Q' = \text{np}(\mathcal{U}_p(\#_\varepsilon(r_j|_\tau)))$  and let  $Q = Q'^{-1}$ . Again  $\#_\varepsilon(\ell')\sigma' \rightarrow_Q^* \#_\varepsilon(r_j|_\tau)\sigma$ , since  $\#_\varepsilon(r_j|_\tau)\sigma \rightarrow_{Q'}^* \#_\varepsilon(\ell')\sigma'$ . By [Lemma 6](#) one directly obtains  $\#_\varepsilon(r_j|_\tau)\sigma = \text{Ren}(\text{Cap}_Q(\#_\varepsilon(\ell')))\delta$  for some substitution  $\delta$  that differs from  $\sigma'$  at most on the variables that are introduced by  $\text{Ren}$  and  $\text{Cap}_Q$ . W.l.o.g. we can assume that  $\sigma'$  is equal to  $\delta$  on all these fresh variables, and since the ADPs are variable-renamed, we can also assume that  $\sigma$  is equal to  $\delta$  on all the fresh variables and all the variables from  $\#_\varepsilon(\ell')$ . Hence,  $\#_\varepsilon(r_j|_\tau)\sigma = \text{Ren}(\text{Cap}_Q(\#_\varepsilon(\ell')))\sigma$  shows there is an mgu  $\gamma$  of  $\#_\varepsilon(r_j|_\tau)$  and  $\text{Ren}(\text{Cap}_Q(\#_\varepsilon(\ell')))$  with  $\sigma = \gamma\zeta$  for some substitution  $\zeta$ . Moreover, the property  $\{\ell\sigma, \ell'\sigma'\} \subseteq \text{ANF}_p$  must remain true when replacing  $\sigma$  and  $\sigma'$  by the more general substitution  $\gamma$ , i.e.,  $\{\ell\gamma, \ell'\gamma\} \subseteq \text{ANF}_p$ . Hence, we can apply the new ADP  $\ell\gamma \rightarrow \{p_1 : r_1\gamma, \dots, p_k : r_k\gamma\}^m \in N$  with the position  $\pi$  and the substitution  $\zeta$ . This means that we have  $t'_v \hookrightarrow_p \{\frac{p_{w_1}}{p_v} : t'_{w_1}, \dots, \frac{p_{w_k}}{p_v} : t'_{w_k}\}$  with  $t'_{w_j} = \text{b}^{\uparrow}_\pi(t'_v[\#_{\text{Pos}_{D^\#}(r_j)} \cup \Psi_j(r_j\gamma\zeta)]_\pi)$  if  $m = \text{true}$ , or  $t'_{w_j} = t'_v[\#_{\text{Pos}_{D^\#}(r_j)} \cup \Psi_j(r_j\gamma\zeta)]_\pi$ , otherwise. Since,  $\sigma = \gamma\zeta$  we directly get  $t_{w_j} \stackrel{\text{b}}{=} t'_{w_j}$  and  $\text{Pos}_{D^\#}(t_{w_j}) \subseteq \text{Pos}_{D^\#}(t'_{w_j})$  so that [\(B.14\)](#) is satisfied again.

**Completeness:** Completely the same as for [Theorem 16](#).  $\square$

**Proof. (Proof of [Theorem 20](#)) Soundness:** We use the same idea as in the proof of soundness for [Theorem 18](#) but at a node  $v$  that uses the position  $\pi \in \text{Pos}_{D^\#}(t_v)$  we do not look at the next node that rewrites at an annotated position below or equal to  $\pi$ , but at all such nodes that can be either annotated or not.

Let  $\mathcal{P}$  be not iAST. Then by [Lemma 2](#) there exists an innermost  $\mathcal{P}$ -CT  $\mathfrak{X} = (V, E, L)$  that converges with probability  $< 1$  that starts with  $(1 : t)$  and  $t = s\theta \in \text{ANF}_p$  for a substitution  $\theta$  and an ADP  $s \rightarrow \dots \in \mathcal{P}$ , and  $\text{Pos}_{D^\#}(t) = \{\varepsilon\}$ . Let  $\mathcal{P}' = \mathcal{P}' \cup N$ . We will now create an innermost  $\overline{\mathcal{P}'}$ -CT  $\mathfrak{X}' = (V, E, L')$ , with the same underlying tree structure, and an adjusted labeling such that  $p_v^\mathfrak{X} = p_v^{\mathfrak{X}'}$  for all  $v \in V$ . Since the tree structure and the probabilities are the same, we then get  $|\mathfrak{X}'| = |\mathfrak{X}|$ , and hence  $\overline{\mathcal{P}'}$  is not iAST either.

The core idea of this construction is that every rewrite step with  $\ell \rightarrow \{p_1 : r_1, \dots, p_k : r_k\}^m$  can also be done with a rule from  $N$ . If we use  $\ell \rightarrow \{p_1 : \#_{\text{cov}_1(\delta_1, \dots, \delta_d)}(r_1), \dots, p_k : \#_{\text{cov}_k(\delta_1, \dots, \delta_d)}(r_k)\}^m \in N$ , we may create fewer annotations than we did when using the old ADP  $\ell \rightarrow \{p_1 : r_1, \dots, p_k : r_k\}^m$ . However, in the  $\mathcal{P}$ -CT  $\mathfrak{X}$  we never rewrite at the position of those annotations that do not get created in the CT  $\mathfrak{X}'$ , hence we can ignore them. We construct the new labeling  $L'$  for the  $\overline{\mathcal{P}'}$ -CT  $\mathfrak{X}'$  inductively such that for all nodes  $v \in V \setminus \text{Leaf}$  with  $vE = \{w_1, \dots, w_m\}$  we have  $t'_v \stackrel{\text{c}}{\hookrightarrow}_{\overline{\mathcal{P}'}} \{\frac{p_{w_1}}{p_v} : t'_{w_1}, \dots, \frac{p_{w_m}}{p_v} : t'_{w_m}\}$ . Let  $W \subseteq V$  be the set of nodes  $v$  where we have already defined the labeling  $L'(x)$ . During our construction, we ensure that for every  $v \in W$  we have

$$t_v \stackrel{\text{b}}{=} t'_v \text{ and } \text{Pos}_{D^\#}(t_v) \setminus \text{Junk}(t_v) \subseteq \text{Pos}_{D^\#}(t'_v). \quad (\text{B.16})$$

In this case, we define  $\pi \in \text{Junk}(t_v) \Leftrightarrow$  there is no ADP  $\ell \rightarrow \{p_1 : r_1, \dots, p_k : r_k\}^m \in \mathcal{P}$  and no substitution  $\sigma$  such that  $\#_\varepsilon(t_v|_\pi) \rightarrow_{\text{np}(\mathcal{P})}^* \ell^\# \sigma$ .

For the construction, we start with the same term at the root. Here, [\(B.16\)](#) obviously holds. As long as there is still an inner node  $v \in W$  such that its successors are not contained in  $W$ , we do the following. Let  $vE = \{w_1, \dots, w_m\}$  be the set of its successors. We need to define the corresponding sets  $t'_{w_1}, \dots, t'_{w_m}$  for the nodes  $w_1, \dots, w_m$ . Since  $v$  is not a leaf and  $\mathfrak{X}$  is a  $\mathcal{P}$ -CT, we have  $t_v \stackrel{\text{c}}{\hookrightarrow}_p \{\frac{p_{w_1}}{p_v} : t_{w_1}, \dots, \frac{p_{w_m}}{p_v} : t_{w_m}\}$ . We have the following three cases:

- If it is a step with  $\stackrel{\text{c}}{\hookrightarrow}_p$  using an ADP that is different from  $\ell \rightarrow \{p_1 : r_1, \dots, p_k : r_k\}^m$  in  $\mathfrak{X}$ , then we perform a rewrite step with the same ADP, the same redex, and the same substitution in  $\mathfrak{X}'$ . Analogous to Case (A) of the soundness proof for [Theorem 16](#), we can show [\(B.16\)](#) for the resulting terms.
- If it is a step with  $\stackrel{\text{c}}{\hookrightarrow}_p$  using the ADP  $\ell \rightarrow \{p_1 : r_1, \dots, p_k : r_k\}^m$  at a position  $\pi \notin \text{Pos}_{D^\#}(t_v)$  in  $\mathfrak{X}$ , then we perform a rewrite step with  $\ell \rightarrow \{p_1 : \#_{\text{cov}_1(\delta_1, \dots, \delta_d)}(r_1), \dots, p_k : \#_{\text{cov}_k(\delta_1, \dots, \delta_d)}(r_k)\}^m$ , same redex, same substitution, and same position in  $\mathfrak{X}'$ . Analogous to Case (B) of the soundness proof for [Theorem 16](#), we can show [\(B.16\)](#) for the resulting terms. Note that the rule that we use contains fewer annotations than the original rule, but since  $\pi \notin \text{Pos}_{D^\#}(t_v)$ , we remove all annotations from the rule during the application of the rewrite step anyway.
- If it is a step with  $\stackrel{\text{c}}{\hookrightarrow}_p$  using the ADP  $\ell \rightarrow \{p_1 : r_1, \dots, p_k : r_k\}^m$  at a position  $\pi \in \text{Pos}_{D^\#}(t_v)$  in  $\mathfrak{X}$ , then we look at specific successors to find a substitution  $\delta$  such that  $\ell\delta \rightarrow \{p_1 : r_1\delta, \dots, p_k : r_k\delta\}^m \in N$  or we detect that we can use the ADP  $\ell \rightarrow \{p_1 : \#_{\text{cov}_1(\delta_1, \dots, \delta_d)}(r_1), \dots, p_k : \#_{\text{cov}_k(\delta_1, \dots, \delta_d)}(r_k)\}^m$  and perform a rewrite step with this new ADP, at the same position in  $\mathfrak{X}'$ .

So it remains to consider Case (C) in detail. Here, we have  $t_v \stackrel{\text{c}}{\hookrightarrow}_p \{\frac{p_{w_1}}{p_v} : t_{w_1}, \dots, \frac{p_{w_k}}{p_v} : t_{w_k}\}$  using the ADP  $\ell \rightarrow \{p_1 : r_1, \dots, p_k : r_k\}^m$ , the position  $\pi \in \text{Pos}_{D^\#}(t_v)$ , and a substitution  $\sigma$  such that  $\text{b}(t_v|_\pi) = \ell\sigma \in \text{ANF}_p$ .

We first consider the case where there is no successor  $z$  of  $v$  where an ADP is applied at an annotated position below or at  $\pi$ , or an ADP is applied on a position above  $\pi$  before reaching such a node  $v$ . Then, we can use  $\ell \rightarrow \{p_1 : \#_{\text{cov}_1(\delta_1, \dots, \delta_d)}(r_1), \dots, p_k : \#_{\text{cov}_k(\delta_1, \dots, \delta_d)}(r_k)\}^m$  instead, because the annotations will never be used, so they do not matter.

Otherwise, there exists a successor  $z$  of  $v$  where an ADP is applied at an annotated position below or at  $\pi$ , and no ADP is applied on a position above  $\pi$  before. Let  $v_1, \dots, v_n$  be all (not necessarily direct) successors that rewrite below position  $\pi$ , or rewrite at position  $\pi$ , and on the path from  $v$  to  $z$  there is no other node with this property, and no node that performs a rewrite step above  $\pi$ . Furthermore, let  $t_1, \dots, t_n$  be the used redexes and  $\rho_1, \dots, \rho_n$  be the used substitutions.

- **(C1)** If none of the redexes  $t_1, \dots, t_n$  are covered by  $t$ , then we use the ADP  $\ell \rightarrow \{p_1 : \#_{\text{cov}_1(\delta_1, \dots, \delta_d)}(r_1), \dots, p_k : \#_{\text{cov}_k(\delta_1, \dots, \delta_d)}(r_k)\}^m$  with the position  $\pi \in \text{Pos}_{D^\#}(t_v) \setminus \text{Junk}(t_v) \subseteq_{(IH)} \text{Pos}_{D^\#}(t'_v)$  and the substitution  $\sigma$ . Once again, [\(B.16\)](#) is satisfied for our resulting terms.

- **(C2)** If  $t = t_i$  for some  $1 \leq i \leq n$ , then we can find a narrowing substitution  $\delta_e$  of  $t$  that is more general than  $\sigma$ , i.e., we have  $\delta_e \gamma = \sigma$ . Now, we use the ADP  $\ell \delta_e \rightarrow \{p_1 : r_1 \delta_e, \dots, p_k : r_k \delta_e\}^m$  with the position  $\pi \in \text{Pos}_{D^\#}(t_v) \setminus \text{Junk}(t_v) \subseteq_{(IH)} \text{Pos}_{D^\#}(t'_v)$  and the substitution  $\gamma$  such that  $b(t_v|_\pi) = \ell \delta_e \gamma = \ell \sigma \in \text{ANF}_P$ . Once again, (B.16) is satisfied for our resulting terms.
- **(C3)** If  $t \neq t_i$  for all  $1 \leq i \leq n$  but there is an  $1 \leq i \leq n$  such that  $t_i$  is covered, then, since  $t_i$  is covered, there exists a narrowing substitution  $\delta_e$  of  $t$  that is more general than  $\rho_i$ , i.e., there exists a substitution  $\kappa_1$  with  $\delta_e \kappa_1 = \rho_i$ , and since we use  $\rho_i$  later on we additionally have that  $\rho_i$  is more general than  $\sigma$ , i.e., there exists a substitution  $\kappa_2$  with  $\rho_i \kappa_2 = \sigma$ . Now, we use the ADP  $\ell \delta_e \rightarrow \{p_1 : r_1 \delta_e, \dots, p_k : r_k \delta_e\}^m$  with the position  $\pi \in \text{Pos}_{D^\#}(t_v) \setminus \text{Junk}(t_v) \subseteq_{(IH)} \text{Pos}_{D^\#}(t'_v)$  and the substitution  $\kappa_1 \kappa_2$  such that  $b(t_v|_\pi) = \ell \delta_e \kappa_1 \kappa_2 = \ell \sigma \in \text{ANF}_P$ . Once again, (B.16) is satisfied for our resulting terms.

**Completeness:** The proof is analogous to the completeness proof of Theorem 16. We can replace each ADP  $\ell \delta_e \rightarrow \{p_1 : r_1 \delta_e, \dots, p_k : r_k \delta_e\}^m$  with the more general one  $\ell \rightarrow \{p_1 : r_1, \dots, p_k : r_k\}^m$ , and each ADP  $\ell \rightarrow \{p_1 : \#_{\text{cov}_1(\delta_1, \dots, \delta_d)}(r_1), \dots, p_k : \#_{\text{cov}_k(\delta_1, \dots, \delta_d)}(r_k)\}^m$  can be replaced by  $\ell \rightarrow \{p_1 : r_1, \dots, p_k : r_k\}^m$  as well, leading to more annotations than before.  $\square$

## References

- [1] J. Giesl, C. Aschermann, M. Brockschmidt, F. Emmes, F. Frohn, C. Fuhs, J. Hensel, C. Otto, M. Plücker, P. Schneider-Kamp, T. Ströder, S. Swiderski, R. Thiemann, Analyzing program termination and complexity automatically with AProVE, *J. Autom. Reason.* 58 (1) (2017) 3–31. <https://doi.org/10.1007/s10817-016-9388-y>
- [2] J. Waldmann, Matchbox: a tool for match-bounded string rewriting, in: Proc. RTA '04, LNCS 3091, 2004, pp. 85–94. [https://doi.org/10.1007/978-3-540-25979-4\\_6](https://doi.org/10.1007/978-3-540-25979-4_6)
- [3] A. Geser, D. Hofbauer, R. Waldmann, Sparse tiling through overlap closures for termination of string rewriting, in: Proc. FSCD '19, LIPIcs 131, 2019, pp. 21:1–21:21. <https://doi.org/10.4230/LIPICS.FSCD.2019.21>
- [4] R. Gutiérrez, S. Lucas, MU-TERM: verify termination properties automatically (system description), in: Proc. IJCAR '20, LNCS 12167, 2020, pp. 436–447. [https://doi.org/10.1007/978-3-030-51054-1\\_28](https://doi.org/10.1007/978-3-030-51054-1_28)
- [5] A. Yamada, K. Kusakari, T. Sakabe, Nagoya termination tool, in: Proc. RTA-TLCA '14, LNCS 8560, 2014, pp. 466–475. [https://doi.org/10.1007/978-3-319-08918-8\\_32](https://doi.org/10.1007/978-3-319-08918-8_32)
- [6] É. Payet, Non-termination in term rewriting and logic programming, *J. Autom. Reason.* 68 (1) (2024) 4. <https://doi.org/10.1007/s10817-023-09693-Z>
- [7] M. Korp, C. Sternagel, H. Zankl, A. Middeldorp, Tyrolean termination tool 2, in: Proc. RTA '09, LNCS 5595, 2009, pp. 295–304. [https://doi.org/10.1007/978-3-642-02348-4\\_21](https://doi.org/10.1007/978-3-642-02348-4_21)
- [8] T. Arts, J. Giesl, Termination of term rewriting using dependency pairs, *Theor. Comput. Sci.* 236 (1–2) (2000) 133–178. [https://doi.org/10.1016/S0304-3975\(99\)00207-8](https://doi.org/10.1016/S0304-3975(99)00207-8)
- [9] J. Giesl, R. Thiemann, P. Schneider-Kamp, The dependency pair framework: combining techniques for automated termination proofs, in: Proc. LPAR '04, LNCS 3452, 2004, pp. 301–331. [https://doi.org/10.1007/978-3-540-32275-7\\_21](https://doi.org/10.1007/978-3-540-32275-7_21)
- [10] J. Giesl, R. Thiemann, P. Schneider-Kamp, S. Falke, Mechanizing and improving dependency pairs, *J. Autom. Reason.* 37 (3) (2006) 155–203. <https://doi.org/10.1007/s10817-006-9057-7>
- [11] N. Hirokawa, A. Middeldorp, Automating the dependency pair method, *Inf. Comput.* 199 (1–2) (2005) 172–199. <https://doi.org/10.1016/j.ic.2004.10.004>
- [12] N. Hirokawa, A. Middeldorp, Tyrolean termination tool: techniques and features, *Inf. Comput.* 205 (4) (2007) 474–511. <https://doi.org/10.1016/J.IC.2006.08.010>
- [13] R. Thiemann, C. Sternagel, Certification of termination proofs using CeTA, in: Proc. TPHOLs '09, LNCS 5674, 2009, pp. 452–468. [https://doi.org/10.1007/978-3-642-03359-9\\_31](https://doi.org/10.1007/978-3-642-03359-9_31)
- [14] E. Contejean, P. Courtieu, J. Forest, O. Pons, X. Urbain, Certification of automated termination proofs, in: Proc. ProCoS '07, LNCS 4720, 2007, pp. 148–162. [https://doi.org/10.1007/978-3-540-74621-8\\_10](https://doi.org/10.1007/978-3-540-74621-8_10)
- [15] F. Blanqui, A. Koprowski, CoLoR: a Coq library on well-founded rewrite relations and its application to the automated verification of termination certificates, *Math. Struct. Comput. Sci.* 21 (4) (2011) 827–859. <https://doi.org/10.1017/S0960129511000120>
- [16] A.A. Almeida, M. Ayala-Rincón, Formalizing the dependency pair criterion for innermost termination, *Sci. Comput. Prog.* 195 (2020). <https://doi.org/10.1016/j.scico.2020.102474>
- [17] A.D. Gordon, T.A. Meisinger, A.V. Nori, S.K. Rajamani, Probabilistic programming, in: Proc. FOSE '14, 2014, pp. 167–181. <https://doi.org/10.1145/2593882.2593900>
- [18] D. Kozen, A probabilistic PDL, *J. Comput. Syst. Sci.* 30 (2) (1985) 162–178.
- [19] A. McIver, C. Morgan, *Abstraction, Refinement and Proof for Probabilistic Systems*, Springer, 2005. <https://doi.org/10.1007/B138392>
- [20] O. Bournez, C. Kirchner, Probabilistic rewrite strategies. Applications to ELAN, in: Proc. RTA '02, LNCS 2378, 2002, pp. 252–266. [https://doi.org/10.1007/3-540-45610-4\\_18](https://doi.org/10.1007/3-540-45610-4_18)
- [21] O. Bournez, F. Garnier, Proving positive almost-sure termination, in: Proc. RTA '05, LNCS 3467, 2005, pp. 323–337. [https://doi.org/10.1007/978-3-540-32033-3\\_24](https://doi.org/10.1007/978-3-540-32033-3_24)
- [22] B.L. Kaminski, J.P. Katoen, C. Matheja, F. Olmedo, Weakest precondition reasoning for expected runtimes of randomized algorithms, *J. ACM* 65 (2018) 1–68. <https://doi.org/10.1145/3208102>
- [23] A. McIver, C. Morgan, B.L. Kaminski, J.P. Katoen, A new proof rule for almost-sure termination, *Proc. ACM Program. Lang.* 2 (POPL) (2018). <https://doi.org/10.1145/3158121>
- [24] F. Meyer, M. Hark, J. Giesl, Inferring expected runtimes of probabilistic integer programs using expected sizes, in: Proc. TACAS '21, LNCS 12651, 2021, pp. 250–269. [https://doi.org/10.1007/978-3-030-72016-2\\_14](https://doi.org/10.1007/978-3-030-72016-2_14)
- [25] S. Agrawal, K. Chatterjee, P. Novotný, Lexicographic ranking supermartingales: an efficient approach to termination of probabilistic programs, *Proc. ACM Program. Lang.* 2 (POPL) (2017). <https://doi.org/10.1145/3158122>
- [26] K. Chatterjee, H. Fu, P. Novotný, Termination analysis of probabilistic programs with martingales, in: G. Barthe, J.P. Katoen, A. Silva (Eds.), *Foundations of Probabilistic Programming*, Cambridge University Press, 2020, p. 221–258. <https://doi.org/10.1017/9781108770750.008>
- [27] B.L. Kaminski, J.P. Katoen, C. Matheja, Expected runtime analysis by program verification, in: G. Barthe, J.P. Katoen, A. Silva (Eds.), *Foundations of Probabilistic Programming*, Cambridge University Press, 2020, p. 185–220. <https://doi.org/10.1017/9781108770750.007>
- [28] L.M. Ferrer Fioriti, H. Hermanns, Probabilistic termination: soundness, completeness, and compositionality, in: Proc. POPL '15, 2015, pp. 489–501. <https://doi.org/10.1145/2676726.2677001>
- [29] J. Giesl, P. Giesl, M. Hark, Computing expected runtimes for constant probability programs, in: Proc. CADE '19, LNCS 11716, 2019, pp. 269–286. [https://doi.org/10.1007/978-3-030-29436-6\\_16](https://doi.org/10.1007/978-3-030-29436-6_16)
- [30] M. Huang, H. Fu, K. Chatterjee, A.K. Goharshady, Modular verification for almost-sure termination of probabilistic programs, *Proc. ACM Program. Lang.* 3 (OOPSLA) (2019). <https://doi.org/10.1145/3360555>
- [31] M. Moosbrugger, E. Bartocci, J.P. Katoen, L. Kovács, Automated termination analysis of polynomial probabilistic programs, in: Proc. ESOP '21, LNCS 12648, 2021, pp. 491–518. [https://doi.org/10.1007/978-3-030-72019-3\\_18](https://doi.org/10.1007/978-3-030-72019-3_18)
- [32] M. Avanzini, G. Moser, M. Schaper, A modular cost analysis for probabilistic programs, *Proc. ACM Program. Lang.* 4 (OOPSLA) (2020). <https://doi.org/10.1145/3428240>

- [33] V.C. Ngo, Q. Carbonneaux, J. Hoffmann, Bounded expectations: resource analysis for probabilistic programs, in: Proc. PLDI '18, 2018, pp. 496–512. <https://doi.org/10.1145/3192366.3192394>
- [34] P. Schröder, K. Batz, B.L. Kaminski, J. Katoen, C. Matheja, A deductive verification infrastructure for probabilistic programs, Proc. ACM Program. Lang. 7 (OOPSLA2) (2023) 2052–2082. <https://doi.org/10.1145/3622870>
- [35] R. Beutner, L. Ong, On probabilistic termination of functional programs with continuous distributions, in: Proc. PLDI '21, 2021, pp. 1312–1326. <https://doi.org/10.1145/3453483.3454111>
- [36] U. Dal Lago, C. Grellois, Probabilistic termination by monadic affine sized typing, in: Proc. ESOP '17, LNCS 10201, 2017, pp. 393–419. [https://doi.org/10.1007/978-3-662-54434-1\\_15](https://doi.org/10.1007/978-3-662-54434-1_15)
- [37] U. Dal Lago, C. Faggian, S.R. Della Rocca, Intersection types and (positive) almost-sure termination, Proc. ACM Program. Lang. 5 (POPL) (2021). <https://doi.org/10.1145/3434313>
- [38] D. Wang, D.M. Kahn, J. Hoffmann, Raising expectations: automating expected cost analysis with types, Proc. ACM Program. Lang. 4 (ICFP) (2020). <https://doi.org/10.1145/3408992>
- [39] L. Leutgeb, G. Moser, F. Zuleger, Automated expected amortised cost analysis of probabilistic data structures, in: Proc. CAV '22, LNCS 13372, 2022, pp. 70–91. [https://doi.org/10.1007/978-3-031-13188-2\\_4](https://doi.org/10.1007/978-3-031-13188-2_4)
- [40] K. Batz, B.L. Kaminski, J.P. Katoen, C. Matheja, L. Verscht, A calculus for amortized expected runtimes, Proc. ACM Program. Lang. 7 (POPL) (2023). <https://doi.org/10.1145/3571260>
- [41] M. Avanzini, U. Dal Lago, A. Yamada, On probabilistic term rewriting, Sci. Comput. Program. 185 (2020). <https://doi.org/10.1016/j.scico.2019.102338>
- [42] J.C. Kassing, J. Giesl, Proving almost-sure innermost termination of probabilistic term rewriting using dependency pairs, in: Proc. CADE '23, LNCS 14132, 2023, pp. 344–364. [https://doi.org/10.1007/978-3-031-38499-8\\_20](https://doi.org/10.1007/978-3-031-38499-8_20)
- [43] J.C. Kassing, J. Giesl, From innermost to full probabilistic term rewriting: almost-sure termination, complexity, and modularity, Log. Methods Comput. Sci. 21 (4) (2025). [https://doi.org/10.46298/lmcs-21\(4:28\)2025](https://doi.org/10.46298/lmcs-21(4:28)2025).
- [44] J.C. Kassing, S. Dollase, J. Giesl, A complete dependency pair framework for almost-sure innermost termination of probabilistic term rewriting, in: Proc. FLOPS '24, LNCS 14659, 2024, pp. 62–80. [https://doi.org/10.1007/978-981-97-2300-3\\_4](https://doi.org/10.1007/978-981-97-2300-3_4)
- [45] J.C. Kassing, J. Giesl, Annotated dependency pairs for full almost-sure termination of probabilistic term rewriting, in: Principles of Verification: Cycling the Probabilistic Landscape, LNCS 15260, 2024, pp. 339–366. [https://doi.org/10.1007/978-3-031-75783-9\\_14](https://doi.org/10.1007/978-3-031-75783-9_14)
- [46] F. Baader, T. Nipkow, Term Rewriting and All That, Cambridge University Press, 1998. <https://doi.org/10.1017/CBO9781139172752>
- [47] D.S. Lankford, On Proving Term Rewriting Systems are Noetherian, Memo MTP-3, Math. Dept., Louisiana Technical University, Ruston, LA, 1979. [http://www.ens-lyon.fr/LIP/REWRITING/TERMINATION/Lankford\\_Poly\\_Term.pdf](http://www.ens-lyon.fr/LIP/REWRITING/TERMINATION/Lankford_Poly_Term.pdf).
- [48] J. Endrullis, J. Waldmann, H. Zantema, Matrix interpretations for proving termination of term rewriting, J. Autom. Reason. 40 (2–3) (2008) 195–220. <https://doi.org/10.1007/S10817-007-9087-9>
- [49] B. Gramlich, Abstract relations between restricted termination and confluence properties of rewrite systems, Fundam. Inform. 24 (1995) 2–23. <https://doi.org/10.3233/FI-1995-24121>
- [50] L. Noschinski, F. Emmes, J. Giesl, Analyzing innermost runtime complexity of term rewriting by dependency pairs, J. Autom. Reason. 51 (2013) 27–56. [https://doi.org/10.1007/978-3-642-22438-6\\_32](https://doi.org/10.1007/978-3-642-22438-6_32)
- [51] J.C. Kassing, G. Vartanyan, J. Giesl, A dependency pair framework for relative termination of term rewriting, in: Proc. IJCAR '24, LNCS 14740, 2024, pp. 360–380. [https://doi.org/10.1007/978-3-031-63501-4\\_19](https://doi.org/10.1007/978-3-031-63501-4_19)
- [52] X. Urbain, Modular and incremental automated termination proofs, J. Autom. Reason. 32 (4) (2004) 315–355. <https://doi.org/10.1007/BF03177743>
- [53] R. Thiemann, The DP Framework for Proving Termination of Term Rewriting, Ph.D. thesis, RWTH Aachen University, 2007. <https://verify.rwth-aachen.de/da/thiemann-diss.pdf>.
- [54] TPDB, Termination Problem Data Base, 2024, <https://github.com/TermCOMP/TPDB-ARI>.
- [55] J. Giesl, A. Rubio, C. Sternagel, J. Waldmann, A. Yamada, The termination and complexity competition, in: Proc. TACAS '19, LNCS 11429, 2019, pp. 156–166. [https://doi.org/10.1007/978-3-030-17502-3\\_10](https://doi.org/10.1007/978-3-030-17502-3_10)
- [56] T. Arts, J. Giesl, A Collection of Examples for Termination of Term Rewriting Using Dependency Pairs, Technical Report AIB-2001-09, RWTH Aachen University, 2001. <https://verify.rwth-aachen.de/giesl/papers/examples.pdf>.
- [57] J. Niederhauser, Personal Communication, 2023.
- [58] C. Fuhs, J. Giesl, M. Parting, P. Schneider-Kamp, S. Swiderski, Proving termination by dependency pairs and inductive theorem proving, J. Autom. Reason. 47 (2) (2011) 133–160. <https://doi.org/10.1007/s10817-010-9215-9>
- [59] J. Giesl, T. Arts, Verification of Erlang processes by dependency pairs, Appl. Algebra Eng. Commun. Comput. 12 (1/2) (2001) 39–72. <https://doi.org/10.1007/s002000100063>
- [60] B. Gramlich, Termination and Confluence: Properties of Structured Rewrite Systems, Ph.D. thesis, University of Kaiserslautern, 1996. <https://www.logic.at/staff/gramlich/papers/thesis96.pdf>.