

On Certificates, Expected Runtimes, and Termination in Probabilistic Pushdown Automata

Tobias Winkler and Joost-Pieter Katoen
 Software Modeling and Verification Group
 RWTH Aachen University, Aachen, Germany
 Email: {tobias.winkler,katoen}@cs.rwth-aachen.de

Abstract—Probabilistic pushdown automata (pPDA) are a natural operational model for a variety of recursive discrete stochastic processes. In this paper, we study *certificates* – succinct and easily verifiable proofs – for upper and lower bounds on various quantitative properties of a given pPDA. We reveal an intimate, yet surprisingly simple connection between the existence of such certificates and the expected time to termination of the pPDA at hand. This is established by showing that certain intrinsic properties, like the spectral radius of the Jacobian of the pPDA’s underlying polynomial equation system, are directly related to expected runtimes. As a consequence, we obtain that there always exist easy-to-check proofs for positive almost-sure termination: does a pPDA terminate in finite expected time?

are these proofs short?

I. INTRODUCTION

Probabilistic pushdown automata (pPDA) [1] are known as a universal model for various kinds of recursive discrete stochastic processes. This includes stochastic context-free grammars, probabilistic programs with procedures and recursion [2], as well as many random walks on \mathbb{N} . pPDA are equivalent to recursive Markov chains [3], [4].

The pPDA considered in this paper are similar to standard pushdown automata. The difference is that a pPDA does not read symbols from an input word, but rather determines its next configuration stochastically based on fixed distributions over the states’ outgoing transitions. A pPDA can thus be seen as a finite-state discrete-time Markov chain with an unbounded stack. See Figure 1 for an example.

A plethora of quantitative verification problems for pPDA has been studied, including termination probabilities, expectations and tail bounds of key random variables such as running time [5], [6], rewards/costs [7], memory consumption [8], satisfaction probabilities of linear-time [1], [9]–[12] and branching-time [9], [11], [13] logics, as well as probabilistic simulation [14] and bisimulation [15] (see [16] for a detailed overview).

A central issue in most of the above problems is to compute a certain set of probabilities, denoted $[pZ|q]$ for states p, q and stack symbol Z . $[pZ|q]$ stands for the probability that the pPDA eventually reaches q with empty stack assuming it is initiated in p with stack content Z [1]. It is well-known that the

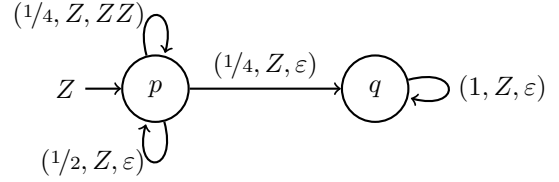


Fig. 1. An example probabilistic pushdown automaton (pPDA) [22] with states $Q = \{p, q\}$ and stack alphabet $\Gamma = \{Z\}$. A transition (a, Z, α) means that with probability a , Z is popped from the stack and α is pushed.

numbers $[pZ|q]$ are characterized as a *least fixed point solution* μf of a *positive polynomial system* (PPS) f [1], [3], [17]; and it is thus not surprising that in general, $[pZ|q]$ may be irrational and cannot easily be represented in a closed form using radical expressions [4]. On the other hand, many problems mentioned in the previous paragraph remain *decidable* by encoding $[pZ|q]$ in the *existential first-order theory of real arithmetic*, which is in *PSPACE*.

To cope with pPDA in practice, and to obtain better bounds than *PSPACE*, lots of research has focused on computing the probabilities $[pZ|q]$ *approximately*. The resulting algorithms employ variants of *Newton iteration* for the systems generated by pPDA and subclasses thereof [3], [4], [18], as well as for arbitrary polynomial systems with non-negative coefficients [19]–[22].

Our Contributions: In this paper, we are concerned with the following fundamental question: Does there exist a *simple proof* – also called *certificate* – that a claimed bound on a given quantitative pPDA property indeed holds? By *simple* we mean that the proof should be machine-checkable by a verifier with limited resources; in particular, it should be checkable in *PTIME*. Moreover, the check should be conceptually simple such that the verifier is straightforward to implement, and ideally amenable to formal verification. We primarily restrict attention to bounds on the probabilities $[pZ|q]$ as those are at the heart of solving many other problems, see above.

Our motivation for studying certificates is twofold. First, certificates can help increasing the trust one can put in the outputs of a (possibly very complex) analysis engine or probabilistic model checker. Indeed, a model checker that provides certificates for its results along with a formally verified certificate checker is as trustworthy as a fully formally verified model checker – the latter is usually not in reach for

This work has been supported by the DFG RTG 2236 (UnRAVeL), the EU’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 101008233 (MISSION), and the ERC Advanced Grant No 787914 (FRAPPANT).

tools relevant in practice. This is exactly the idea of *certifying algorithms* [23]. In fact, the authors of [23] go one step further and argue that

“ [...] *certifying algorithms are much superior to non-certifying algorithms and [...] for complex algorithmic tasks only certifying algorithms are satisfactory.* ”

Our second, more theoretical motivation is that the study of certificates may of course also spawn new algorithms for pPDA, because an algorithm for finding a certificate (or proving absence thereof) for a given hypothesis is in particular an algorithm deciding that hypothesis.

To the best of our knowledge, certificates for pPDA have only been considered explicitly in the very recent paper [22]. There, the basic idea is to exploit the fairly general rule

$$\mathbf{f}(\mathbf{u}) \leq \mathbf{u} \implies \mu \mathbf{f} \leq \mathbf{u} \quad (1)$$

which is true for all *monotonic* functions $\mathbf{f}: L \rightarrow L$ with least fixed point $\mu \mathbf{f}$, and (L, \leq) a complete lattice. (1) follows from the Knaster-Tarski fixed point theorem and is in particular applicable to the PPS \mathbf{f} arising from pPDA where the condition $\mathbf{f}(\mathbf{u}) \leq \mathbf{u}$ is reasonably simply to check. Applying (1) in practice, however, is not trivial because it is not clear if suitable, effectively representable (i.e., rational-valued) \mathbf{u} even exist. The key contributions of this paper are to extend the ideas of [22] in the following ways (also see Figure 2):

Contribution 1: We characterize the existence of rational-valued certificates for arbitrarily tight upper bounds on $[pZ|q]$ in terms of the expected runtime¹ of the given pPDA Δ . In particular, we show that such certificates exist if the expected runtime conditioned on termination is finite, a property called *cPAST* in this paper. We also obtain similar though slightly more complex results for *lower* bounds. Moreover, we show that the bit-complexity of the rational numbers occurring in the certificates can be bounded in terms of the expected runtime and the minimum non-zero probability $[pZ|q]$.

Our second contribution concerns the termination problem of pPDA: Does a given pPDA eventually reach the empty stack with probability one? This is crucial for numerous applications, e.g. model checking temporal logics [12], or ensuring that a probabilistic program has a well-defined posterior distribution [24, Sec. 5.3].

Contribution 2: We prove that *positive almost-sure termination* (PAST; termination with probability 1 in finite expected time) can *always* be certified (if it holds). This relies on Contribution 1 and a novel characterization of the expected runtimes in general pPDA. The main insight, which is of independent interest beyond the topic

of certificates, is that PAST can be proved using only sufficiently tight upper bounds on the probabilities $[pZ|q]$. This allows deciding termination of a large class of pPDA (all but those that are AST and not PAST) by practically feasible, iterative approximation algorithms.

Proof techniques: As a primary technical vehicle, we prove that the spectral radius ρ of the underlying system’s Jacobi matrix $\mathbf{f}'(\mu \mathbf{f})$ is closely related to the (conditional) expected runtimes (assuming \mathbf{f} has been previously cleaned-up by removing “unnecessary” variables). This relation is expressed via the simple (tight) inequality $\rho \leq 1 - C^{-1}$, where C is the maximal conditional expected runtime occurring in the pPDA at hand ($C = \infty$ is possible). We prove this using results about general PPS from [21] as well as standard facts about non-negative matrices. Another important ingredient are *weighted PDA* over semirings [17].

Related work: Expected runtimes and, more generally, expected rewards in pPDA are studied in [5]–[7], [16]. However, these works focus mostly on expectations *conditioned* on reaching the empty stack while being in a certain state. In our paper we need such conditional expectations as well; but additionally, we also consider the *unconditional* expected runtime of pPDA, which is crucial for proving PAST. More specifically, our Theorem 7 has – to the best of our knowledge – not appeared in previous works. A pPDA is PAST iff the equation system from Theorem 7 has a non-negative solution. This is in contrast to e.g. the equation system from [1, Ex. 3.3] which has a solution iff the expected runtime *conditioned on termination* is finite. However, this can be the case even if the termination probability is < 1 . Theorem 7 is actually more in the spirit of [25], [26, Ch. 4], where *1-exit recursive simple stochastic games* with positive rewards are considered, a model that corresponds to a *stateless* variant of PDA controlled by two players and a probabilistic environment. As we consider general pPDA with states, our result does not easily follow from theirs.

The expected *memory consumption* of a pPDA [8] is also related to its runtime as finite runtime implies finite memory usage. Apart from this, there seems to be no other known connection between the two.

It is shown in [4] that the spectral radius of the Jacobi matrix $\mathbf{f}'(\mathbf{1})$ is related to the termination probability in the case of 1-exit RMC, which are equivalent to stateless pPDA. However, expected runtimes are not considered in [4]. In fact, the connection between expected runtimes and spectral properties of the Jacobi matrix seems to have been ignored entirely in previous works.

Apart from [22], certificates for various other probabilistic and quantitative systems are studied, including Markov decision processes [27], [28], and timed automata [29]. A major technical difference to these works is that they do not need to consider non-linear arithmetic.

[2] introduces a deductive approach to prove bounds on, amongst other properties, expected runtimes of (possibly infinite-state) pPDA represented as probabilistic recursive pro-

¹Also called running time or time to termination. “Expected runtime” is a term inspired by [2].

What?	Certificate data	Verification condition(s)	Certified property	Remarks
Upper bounds Theorem 4	$\mathbf{u} \in \mathbb{Q}_{\geq 0}^{Q \times \Gamma \times Q}$	$\mathbf{f}(\mathbf{u}) \leq \mathbf{u}$	$\forall pZq \in Q \times \Gamma \times Q:$ $[pZ q] \leq \mathbf{u}_{pZq}$	Certificate exists if Δ is cPAST
Lower bounds Theorem 5	$\mathbf{l}, \mathbf{u} \in \mathbb{Q}_{\geq 0}^{Q \times \Gamma \times Q}$	$\mathbf{l} \leq \mathbf{f}(\mathbf{l}), \mathbf{f}(\mathbf{u}) \prec \mathbf{u}$ $\mathbf{l} \leq \mathbf{u}$	$\forall pZq \in Q \times \Gamma \times Q:$ $\mathbf{l}_{pZq} \leq [pZ q] < \mathbf{u}_{pZq}$	Certificate exists if Δ is cPAST
PAST Theorem 8	$\mathbf{u} \in \mathbb{Q}_{\geq 0}^{Q \times \Gamma \times Q}$ $\mathbf{r} \in \mathbb{Q}_{\geq 1}^{Q \times \Gamma}$	$\mathbf{f}(\mathbf{u}) \leq \mathbf{u}$ $M_{\Delta}(\mathbf{u})\mathbf{r} + \mathbf{1} \leq \mathbf{r}$	$\forall pZ \in Q \times \Gamma:$ $\text{ert}[pZ] \leq \mathbf{r}_{pZ}$	Certificate exists if and only if Δ is PAST

Fig. 2. $\Delta = (Q, \Gamma, \delta)$ is a pPDA with fundamental PPS \mathbf{f} . $M_{\Delta}(\mathbf{u})$ is a matrix whose coefficients can be read off Δ and \mathbf{u} . The first two rows apply more generally to least fixed point of *arbitrary* PPS (existence is then guaranteed if the system \mathbf{f} is non-singular).

grams. Our certificates for PAST are loosely inspired by theirs, and we use some of their terminology.

There also exist generalizations of pPDA for incorporating non-determinism [30] or higher-order recursion [31]. We leave certification of these models for future work.

The recent paper [32] studies certificates for (non)-reachability in non-probabilistic PDA and related problems.

Paper outline: The upcoming Section II presents a small example in detail. We then summarize necessary preliminary results and definitions in Section III. Section IV and Section V present our results for certificates for upper and lower bounds, respectively. Section VI establishes the fundamental connection between the Jacobi matrix and the expected runtimes. Certificates for PAST are introduced in Section VII. Bounds on the size of certificates are given in Section VIII. We conclude in Section IX.

II. ILLUSTRATIVE EXAMPLE

Consider the pPDA Δ in Figure 1 on page 1. To simplify the presentation, we write $x_p = [pZ|p]$ and $x_q = [pZ|q]$, i.e., x_p and x_q are the probabilities that Δ terminates in state p and q , respectively, assuming Δ is initiated in p with stack content Z (such *configurations* are denoted pZ in the following). As mentioned earlier, x_p, x_q are the component-wise least solution of a polynomial equation system, called *fundamental system* of Δ in this paper (see Theorem 1 on page 6 for the general equations). In our example:

$$x_p = \frac{1}{4}x_p^2 + \frac{1}{2} \quad (2)$$

$$x_q = \frac{1}{4}x_px_q + \frac{1}{4}x_q + \frac{1}{4} \quad (3)$$

The exact x_p and x_q are the *irrational* values $x_p = 2 - \sqrt{2} \approx 0.59$, $x_q = \sqrt{2} - 1 \approx 0.41$. We stress that in general, it is impossible to express these numbers with radicals (square roots, cubic roots, etc) [4] – in the example this only works because the polynomials have small degree. Therefore one is usually interested in proving (rational) bounds on these probabilities.

Certificates for upper bounds (Section IV): To prove that e.g. $u_p = \frac{3}{5}, u_q = \frac{1}{2}$ are valid upper bounds for x_p and x_q , it is sufficient (but not necessary) that (2) and (3) hold with *inequality* “ \geq ” when replacing x_p by u_p and x_q by u_q . This is

indeed true for our example values, e.g. $\frac{3}{5} \geq \frac{1}{4} \cdot (\frac{3}{5})^2 + \frac{1}{2}$. Note that these checks can be done in polynomial time (in the size of Δ and the bit-complexity of the rational numbers u_p, u_q) and are easy to implement – thinking of them as a certificate is thus justified. However, it is not clear if *rational-valued* bounds satisfying the constraints can always be found. In our paper we characterize a large class of pPDA where rational certificates are guaranteed to exist.

Certificates for lower bounds (Section V): The above idea can be extended to lower bounds as well, but this is slightly more involved. We will prove that $l_p = \frac{4}{7}$ and $l_q = \frac{2}{5}$ are valid lower bounds. Our theory only allows proving lower bounds in conjunction with upper bounds, so we will just reuse the upper bound certificate $u_p = \frac{3}{5}, u_q = \frac{1}{2}$.

- First, we check that (2) and (3) are satisfied with inequality “ \leq ” (reverse direction as for upper bounds) when replacing x_p by l_p and x_q by l_q . This holds.
- Next we check that $l_p \leq u_p$ and $l_q \leq u_q$. This also holds.
- Finally, we verify that (2) and (3) hold with *strict* inequality “ $>$ ” when replacing x_p by u_p and x_q by u_q .

We conclude (correctly) that $\frac{4}{7} \leq 2 - \sqrt{2}$ and $\frac{2}{5} \leq \sqrt{2} - 1$.

Positive almost-sure termination (Section VII): It is intuitively clear that our example pPDA terminates (reaches the empty stack) with probability 1, but this property is difficult to prove in general. With the lower bound certificates seen above we can only conclude that the termination probability is at least $\frac{4}{7} + \frac{2}{5} \approx 0.97$. However, our theory also admits certificates for PAST, which implies termination with probability *exactly* 1. Similar to the lower bound certificates, a certificate for PAST consists of an upper bound certificate u_p, u_q as well as numbers r_p, r_q which are upper bounds on the *expected runtime* starting in pZ and qZ , respectively. In our example, let $u_p = \frac{3}{5}, u_q = \frac{1}{2}$ as before, and let $r_p = \frac{45}{14}$, and $r_q = 1$. To *verify* the certificate we just have to check the inequalities

$$r_p \geq 1 + \frac{1}{4}(r_p + u_p r_p) + \frac{1}{4}(r_p + u_q r_q) \quad (4)$$

$$r_q \geq 1 \quad (5)$$

These inequalities can be easily read off from the transitions of Δ , just like the fundamental system (2) and (3). We find that the inequalities are satisfied and conclude that Δ is PAST. In fact, the expected time to termination starting in pZ is at most $r_p = \frac{45}{14} \approx 3.21$.

This is all that's necessary to prove positive almost-sure termination – and thus in particular *termination with probability 1* – and we show in this paper that this method is *complete*, i.e., it works for arbitrary pPDA that are PAST, even if only rational-valued bounds are allowed.

III. PRELIMINARIES

A. General Notation and Markov Chains

$\overline{\mathbb{R}}_{\geq 0} = \mathbb{R}_{\geq 0} \cup \{\infty\}$ denotes the set of *extended* non-negative reals. Arithmetic and comparison with the ∞ -symbol is done as usual: $\forall a \in \overline{\mathbb{R}}_{\geq 0}: a + \infty = \infty + a = \infty$, $0 \cdot \infty = \infty \cdot 0 = 0$, $\forall a \in \overline{\mathbb{R}}_{\geq 0} \setminus \{0\}: a \cdot \infty = \infty \cdot a = \infty$, and $\forall a \in \overline{\mathbb{R}}_{\geq 0}: a < \infty$. With these conventions, $(\overline{\mathbb{R}}_{\geq 0}, +, \cdot, 0, 1)$ is a commutative semiring. Throughout the paper, we often encounter diverging monotonic sequences $a_0 \leq a_1 \leq \dots$ in $\overline{\mathbb{R}}_{\geq 0}$; in such cases, we define $\lim_{i \rightarrow \infty} a_i = \infty$ (and the resulting ∞ symbol may well be subject to subsequent arithmetic operations obeying the above conventions).

We denote vectors by boldface letters such as \mathbf{v} . All vectors in this paper can be thought of as column vectors. A vector with all entries equal to some $a \in \overline{\mathbb{R}}_{\geq 0}$ is written \mathbf{a} . For instance, $\mathbf{0}$ and $\mathbf{1}$ denote the all-zeros and all-ones vectors, respectively. We always index a vector's components by the elements of a countable, but otherwise arbitrary index set J , and we write e.g. $\mathbf{v} \in \overline{\mathbb{R}}_{\geq 0}^J$ to indicate that \mathbf{v} is a non-negative real vector with $|J|$ components. The individual components of a vector \mathbf{v} are referred to as v_i , $i \in J$. Given vectors $\mathbf{u}, \mathbf{v} \in \overline{\mathbb{R}}_{\geq 0}^J$, we define the following *comparison operators*: $\mathbf{u} \leq \mathbf{v}$ if $\forall i \in J: u_i \leq v_i$, $\mathbf{u} < \mathbf{v}$ if $\mathbf{u} \leq \mathbf{v}$ and $\mathbf{u} \neq \mathbf{v}$, and $\mathbf{u} \prec \mathbf{v}$ if $\forall i \in J: u_i < v_i$. Note that “ \leq ” is a partial order on $\overline{\mathbb{R}}_{\geq 0}^J$.

For matrices we adopt the same index set convention as for vectors: A (square) matrix $A \in \overline{\mathbb{R}}^{J \times J}$ is indexed by pairs of elements from J . For $(i, j) \in J \times J$, $A_{i,j}$ denotes the entry in the row and column indexed by i and j , resp.

In this paper, a *discrete-time Markov chain* (DTMC) $\mathfrak{M} = (S, M)$ consists of a finite or countably infinite set S of states and a row-stochastic transition probability matrix $M \in [0, 1]^{S \times S}$. Intuitively, for $s, t \in S$, $M_{s,t}$ is the probability that \mathfrak{M} moves in one step from state s to state t . To make \mathfrak{M} a proper stochastic process, it is necessary to fix an initial state $s_{init} \in S$; however, for the sake of flexibility, we deliberately do not make s_{init} an explicit part of \mathfrak{M} . We often consider DTMC with a distinguished target (or final) set $F \subseteq S$. We then write M_F for the matrix resulting from M by setting all rows with indices in F to zero.

B. Runtime and Termination: Different Flavours

1) *Termination Moments*: Let $\mathfrak{M} = (S, M)$ be a DTMC with $F \subseteq S$ a distinguished set of target states. Assign probability $p(\pi)$ to each non-empty path $\pi = s_0 \dots s_l \in S^+ = S^* \setminus \{\varepsilon\}$ in the natural way: $p(s_0 \dots s_l) = \prod_{i=0}^{l-1} M_{s_i, s_{i+1}}$. The *length* of path $\pi = s_0 \dots s_l$ is defined as $|\pi| = l$. By convention, a path of length zero is assigned probability 1. For a given initial state $s \in S \setminus F$, let $Paths_{s \rightarrow F}^{\mathfrak{M}} = \{s\} \cdot (S \setminus F)^* \cdot F$ be the set of finite paths from s to F that visit F exactly once

(namely at the end; for $s \in F$ we define $Paths_{s \rightarrow F}^{\mathfrak{M}} = \{s\}$). Such paths are called *proper paths between s and F* in the following. It is easy to see that $p: Paths_{s \rightarrow F}^{\mathfrak{M}} \rightarrow [0, 1]$ is a *sub-probability* distribution, i.e., $\sum_{\pi \in Paths_{s \rightarrow F}^{\mathfrak{M}}} p(\pi) \leq 1$ (strict inequality is possible because there might be a positive probability to never reach F from s).

Definition 1 (Termination Moments): Let $\mathfrak{M} = (S, M)$ be a DTMC. For initial state $s \in S$, target set $F \subseteq S$ and integer $k \geq 0$, we define the *kth termination moment* $E^k[s|F]_{\mathfrak{M}}$ as the expected value of $|\pi|^k$ measured wrt. the *sub-distribution* $p: Paths_{s \rightarrow F}^{\mathfrak{M}} \rightarrow [0, 1]$. Formally:

$$E^k[s|F]_{\mathfrak{M}} := \sum_{\pi \in Paths_{s \rightarrow F}^{\mathfrak{M}}} p(\pi) \cdot |\pi|^k \in \overline{\mathbb{R}}_{\geq 0} \quad (6)$$

with the convention that $0^0 = 1$. We write just $E^k[s|F]$ if \mathfrak{M} is clear from the context. For $f \in S$, we also write $E^k[s|f]$ instead of $E^k[s|\{f\}]$.

For readers familiar with the *weakest pre-expectation calculus* [33] we mention that $E^k[s|F]_{\mathfrak{M}}$ can be expressed as $\text{wp}[P_{\mathfrak{M}}](c^k)(s)$ where $P_{\mathfrak{M}}$ is a suitable probabilistic program encoding the DTMC \mathfrak{M} , and c is a special “counter variable” tracking the number of computation steps in $P_{\mathfrak{M}}$.

Definition 1 involves summing over all (countably many) proper paths between a given initial state s and target set F . A standard observation in the theory of DTMC (e.g. [34]) is that the sum of the probabilities of all proper paths of *length* $k \geq 0$ between s and $\{f\}$ is given by $(M_f^k)_{s,f}$, and thus the sum of the probabilities of all proper paths of *any* finite length between s and $\{f\}$ is equal to $\sum_{k=0}^{\infty} (M_f^k)_{s,f}$. The following lemma allows to express such infinite sums as the solution of a linear equation system; it is an instance of Kleene iteration.

Lemma 1 (Kleene; e.g. [17]): Let J be a countable index set, and $A \in \overline{\mathbb{R}}_{\geq 0}^{J \times J}$, $\mathbf{b} \in \overline{\mathbb{R}}_{\geq 0}^J$. Then the vector

$$\sum_{k=0}^{\infty} A^k \mathbf{b} = \left(\sum_{k=0}^{\infty} A^k \right) \mathbf{b} \in \overline{\mathbb{R}}_{\geq 0}^J \quad (7)$$

is the least $\overline{\mathbb{R}}_{\geq 0}$ -valued solution of the linear equation system $\mathbf{x} = A\mathbf{x} + \mathbf{b}$.

Let $\mathfrak{M} = (S, M)$ be a DTMC with target $F \subseteq S$. In this paper, we only need the first two termination moments:

- The 0th termination moment $E^0[s|F] = [s|F]$ is just the probability to reach F from a given initial state $s \in S$. By Lemma 1, the numbers $\{[s|F] \mid s \in S\}$ constitute the least solution $\boldsymbol{\mu}^0 \in \overline{\mathbb{R}}_{\geq 0}^S$ of the linear equation system

$$\boldsymbol{\mu}^0 = M_F \boldsymbol{\mu}^0 + \boldsymbol{\chi}(F), \quad (8)$$

where $\boldsymbol{\chi}(F)$ is the characteristic $\{0, 1\}$ -vector of F , i.e., $\boldsymbol{\chi}(F)_s = 1$ iff $s \in F$.

- Similarly, it can be shown that the 1st termination moments $E^1[s|F] = E[s|F]$ are the least solution $\boldsymbol{\mu}^1 \in \overline{\mathbb{R}}_{\geq 0}^S$ of the linear equation system

$$\boldsymbol{\mu}^1 = M_F \boldsymbol{\mu}^1 + (\boldsymbol{\mu}^0 - \boldsymbol{\chi}(F)). \quad (9)$$

2) *Conditional Expected Runtimes*: Let $\mathfrak{M} = (S, M)$ be a DTMC with target $F \subseteq S$. If for a given state $s \in S$ we have that $[s|F] > 0$, then we call the quantity $\frac{E[s|F]}{[s|F]}$ the *conditional expected runtime* of \mathfrak{M} wrt. to target F and initial state s . Intuitively, $\frac{E[s|F]}{[s|F]}$ is the number of states visited until reaching F from s , *conditioned* on the fact that F is eventually reached.

3) *(Unconditional) Expected Runtimes*: Intuitively, the expected runtime of a DTMC wrt. to initial state s and target set F is the total expected number of states visited until reaching F from s . In particular, if F is not reached with probability 1, then the expected runtime is ∞ . This is the key difference to the 1st termination moment, i.e., $E[s|F]$ may be finite even if $[s|F] < 1$.

It is possible to define the expected runtimes formally as the expectation of a random variable on the probability space of *infinite* paths induced by \mathfrak{M} , but for our purposes, the following simpler characterization suffices:

Definition 2 (Expected runtimes): Let $\mathfrak{M} = (S, M)$ be a DTMC with a target set $F \subseteq S$. The *expected runtime* of \mathfrak{M} wrt. target F and initial state $s \in S$ is defined as

$$\text{ert}[s|F]_{\mathfrak{M}} = \left(\sum_{k=0}^{\infty} M_F^k (1 - \chi(F)) \right)_s \in \overline{\mathbb{R}}_{\geq 0}. \quad (10)$$

We write just $\text{ert}[s|F]$ if \mathfrak{M} is clear from the context.

Note that by Lemma 1, the values $\text{ert}[s|F]$ constitute the least solution $\mathbf{r} \in \overline{\mathbb{R}}_{\geq 0}^S$ of the linear equation system

$$\mathbf{r} = M_F \mathbf{r} + (1 - \chi(F)). \quad (11)$$

It follows from (9) that if $\mu^0 = 1$ (i.e., \mathfrak{M} is *almost-surely terminating*, see below), then $\text{ert}[s|F] = E[s|F]$. In general, however, we may have $\text{ert}[s|F] > E[s|F]$.

Another observation that will be crucial in the proof of Theorem 7 is that $\text{ert}[s|F]$ is the sum of the probabilities of *all* finite paths starting in s that never visit F , i.e., paths of the form $\{s\} \cdot (S \setminus F)^*$. This follows from (10) and the discussion preceding Lemma 1.

4) Probabilistic Termination:

Definition 3 (AST, PAST, and cPAST): Given a DTMC $\mathfrak{M} = (S, M)$ with target $F \subseteq S$, we define:

- \mathfrak{M} is *almost-surely terminating* (AST) wrt. F if $\forall s \in S$, we have $[s|F] = 1$, i.e., F is reached from every state s with probability 1.
- \mathfrak{M} is *positively AST* (PAST) wrt. F if $\forall s \in S$, $\text{ert}[s|F] < \infty$.
- \mathfrak{M} is *conditionally PAST* (cPAST) wrt. F if $\forall s \in S$ such that $[s|F] > 0$, it holds that $\frac{E[s|F]}{[s|F]} < \infty$.

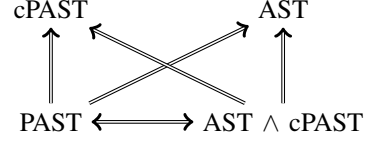
Note that, since $[s|F] = 0$ implies $E[s|F] = 0$, a DTMC is cPAST iff the 1st termination moments are all finite.

The notion of cPAST, although less well known than AST and PAST, has been considered in [6], [7] (the term “cPAST” has not been used before though). Intuitively, cPAST means that termination either occurs “fast” (in finite expected time), or not at all.

Note that we consider *universal* (“for all initial states”) definitions of AST, PAST, and cPAST. This is mostly to

simplify the presentation, as otherwise we would have to treat uninteresting special cases such as unconnected DTMC separately. It is possible to extend our results to the setting of a distinguished initial state of interest by conducting an appropriate reachability analysis in advance.

Lemma 2: The relationship between AST, PAST, and cPAST in a DTMC with a distinguished target set is as follows:



The omitted implications do not hold in general.

Proof: We give a self-contained proof for the implication $\text{PAST} \implies \text{AST}$ as this is central to our work. See [2], [35] for alternative proofs and Appendix A for the other (straightforward) implications.

To see that PAST implies AST, first note that for $n \geq 0$,

$$M_F^n (1 - \chi(F)) + \sum_{k=0}^n M_F^k \chi(F) = 1 \quad (12)$$

which follows by induction using that $M_F \mathbf{1} + \chi(F) = \mathbf{1}$. Now suppose that \mathfrak{M} is not AST, i.e., $\mu^0 < 1$. By Lemma 1, μ^0 can be written explicitly as $\sum_{k=0}^{\infty} M_F^k \chi(F)$. From (12) and $\mu^0 < 1$, it follows that $\lim_{n \rightarrow \infty} M_F^n (1 - \chi(F)) > 0$. As the expected runtimes are defined as $\mathbf{r} = \sum_{k=0}^{\infty} M_F^k (1 - \chi(F))$ (see 10), it follows that at least one component of \mathbf{r} is ∞ . Thus, \mathfrak{M} is not PAST. ■

C. Weighted and Probabilistic PDA

We now define pushdown automata (PDA) with non-negative *rational* transition weights, an instance of PDA with semiring-valued weights as described e.g. in [17].

A weighted pushdown automaton (weighted PDA) is a triple $\Delta = (Q, \Gamma, \delta)$, where Q is a *finite* set of *states*, Γ is a *finite stack alphabet*, and $\delta \subseteq Q \times \Gamma \times \mathbb{Q}_{\geq 0} \times Q \times \Gamma^{\leq 2}$ is a *weighted transition relation* (here, $\Gamma^{\leq 2} = \{\varepsilon\} \cup \Gamma \cup \Gamma^2$). Intuitively, a transition $(p, Z, a, q, \alpha) \in \delta$ means that if Δ is in state p and the stack content is $Z\beta$ for some $\beta \in \Gamma^*$ (i.e., Z is on top of the stack), then there exists a transition with weight a to the configuration where Δ is in state q and the stack content is $\alpha\beta$ (i.e., Z is popped and α is pushed on the stack). Instead of $(p, Z, a, q, \alpha) \in \delta$ we also write $pZ \xrightarrow{a} q\alpha$ if δ is clear from the context. In the following, we always write tuples from $Q \times \Gamma^*$ without parentheses and separating commata, e.g., we write pZ instead of (p, Z) .

A weighted PDA Δ is a *probabilistic* PDA (pPDA) if $\sum_{pZ \xrightarrow{a} q\alpha} a = 1$ holds for all $pZ \in Q \times \Gamma$ [1]. We emphasize that these notions of weighted (probabilistic) PDA do *not* recognize weighted (probabilistic) languages; they are just *generative* models assigning certain weights (probabilities) to finite computation traces.

More formally, we define the *semantics* of a weighted PDA Δ in terms of a countably infinite weighted transition system $\mathfrak{T}_{\Delta} = (Q \times \Gamma^*, T_{\Delta})$, i.e., T_{Δ} is a $\mathbb{Q}_{\geq 0}$ -valued square martix indexed by the *configurations* $Q \times \Gamma^*$ of Δ (which are the states

of \mathfrak{T}_Δ). T_Δ is defined as follows: For all $pZ \xrightarrow{a} q\alpha$ and $\beta \in \Gamma^*$, $T_\Delta(pZ\beta, q\alpha\beta) = a$, and all other entries of T_Δ are zero. Note that if Δ is a pPDA, then \mathfrak{T}_Δ can be seen as a DTMC with target set $F = \{p\varepsilon \mid p \in Q\}$ because the states of the form $p\varepsilon$ have no outgoing transitions in T_Δ .

We extend the notion of termination moments (Def. 1) to the weighted transition systems \mathfrak{T}_Δ induced by a weighted PDA Δ in the natural way. For reasons explained below, we mostly consider 0th moments of the form $[pZ|q\varepsilon]_\Delta$, i.e., the sum of the weights of all finite paths from pZ to $q\varepsilon$. In the following, we omit ε and Δ from the notation whenever this causes no confusion and write just $[pZ|q]$. In a pPDA, $[pZ|q]$ is the probability to reach $q\varepsilon$ from initial configuration pZ . In the context of pPDA, the probabilities $[pZ|q]$ are also called *return probabilities* [12].

We will also consider 1st termination moments of the form $E[pZ|q\varepsilon]_\Delta$ in pPDA (but never in general weighted PDA). For the sake of brevity, those are denoted as $E[pZ|q]$.

Theorem 1 (Fundamental system of PDA [17], [1]): Let $\Delta = (Q, \Gamma, \delta)$ be a weighted PDA. Consider $\mathbb{R}_{\geq 0}$ -valued variable symbols of the form $\langle pZ|q \rangle$, $pZq \in Q \times \Gamma \times Q$. Then the quadratic system of equations

$$\begin{aligned} \langle pZ|q \rangle = & \sum_{pZ \xrightarrow{a} q\varepsilon} a + \sum_{pZ \xrightarrow{a} rY} a \cdot \langle rY|q \rangle \\ & + \sum_{pZ \xrightarrow{a} rYX} a \cdot \sum_{t \in Q} \langle rY|t \rangle \cdot \langle tX|q \rangle \end{aligned} \quad (13)$$

has the (point-wise) least solution $\langle pZ|q \rangle = [pZ|q]_\Delta$.

We refer to [11, Sec. 3] for an intuitive explanation of (13).

Finally, we extend the various notions of termination from Def. 3 to pPDA $\Delta = (Q, \Gamma, \delta)$ and $F = Q \times \{\varepsilon\}$:

- Δ is *AST* if the Markov chain \mathfrak{T}_Δ is AST wrt. F , which is the case iff for all $pZ \in Q \times \Gamma$, $[pZ|F]_\Delta = \sum_{t \in Q} [pZ|t]_\Delta = 1$.
- Δ is *PAST* if \mathfrak{T}_Δ is PAST wrt. to F which is the case iff for all $pZ \in Q \times \Gamma$, $\text{ert}[pZ|F] < \infty$. In the context of pPDA, we simply write $\text{ert}[pZ]$ instead of $\text{ert}[pZ|F]$ because we will only consider expected runtimes wrt. to F as defined above.
- Δ is *cPAST* if \mathfrak{T}_Δ is cPAST wrt. to all $t\varepsilon \in F$, i.e., for all $pZt \in Q \times \Gamma \times Q$, $\frac{E[pZ|t]}{[pZ|t]}$ is either undefined or finite.

D. Positive Polynomial Systems (PPS)

Let J be a finite index set and consider a vector $\mathbf{x} = (x_j)_{j \in J}$ of variable symbols. A *positive polynomial system* (PPS) over the variables \mathbf{x} is a vector $\mathbf{f} = (f_j)_{j \in J}$ of multivariate polynomials f_j with variables in \mathbf{x} and *non-negative coefficients* in $\mathbb{R}_{\geq 0}$.

A PPS \mathbf{f} induces an endofunction $\mathbf{f}: \mathbb{R}_{\geq 0}^J \rightarrow \mathbb{R}_{\geq 0}^J$, $\mathbf{u} \mapsto \mathbf{f}(\mathbf{u}) = (f_j(\mathbf{u}))_{j \in J}$. As the coefficients of the polynomials in \mathbf{f} are non-negative, the function \mathbf{f} is *monotone*, i.e., $\mathbf{u} \leq \mathbf{v} \implies \mathbf{f}(\mathbf{u}) \leq \mathbf{f}(\mathbf{v})$, where \leq is the element-wise order on $\mathbb{R}_{\geq 0}^J$. Since the domain $(\mathbb{R}_{\geq 0}^J, \leq)$ of \mathbf{f} is a *complete lattice*, the Knaster-Tarski fixed point theorem implies that \mathbf{f}

has a *least fixed point* (lfp) $\mu\mathbf{f} \in \mathbb{R}_{\geq 0}^J$. In other words, the polynomial equation system $\mathbf{x} = \mathbf{f}(\mathbf{x})$ is guaranteed to have a *least solution* $\mu\mathbf{f} \in \mathbb{R}_{\geq 0}^J$, wrt. to \leq . A PPS \mathbf{f} is called *feasible* if $\mu\mathbf{f} \prec \infty$, i.e., $\mu\mathbf{f} \in \mathbb{R}_{\geq 0}^J$.

Lemma 3 (Inductive upper bounds): Let \mathbf{f} be a PPS and let $\mathbf{u} \geq \mathbf{0}$. The following statements hold:

- $\mathbf{f}(\mathbf{u}) \leq \mathbf{u} \implies \mu\mathbf{f} \leq \mathbf{u}$
- $\mathbf{f}(\mathbf{u}) \prec \mathbf{u} \implies \mu\mathbf{f} \prec \mathbf{u}$

Proof: The first item is an immediate consequence of the Knaster-Tarski fixed point theorem. For the second item, note that $\mathbf{u} \succ \mathbf{f}(\mathbf{u}) \geq \mathbf{f}^1(\mathbf{u}) \geq \mathbf{f}^2(\mathbf{u}) \dots$ is a monotonic and bounded sequence, hence $\lim_{i \rightarrow \infty} \mathbf{f}^i(\mathbf{u})$ exists. Since \mathbf{f} is a continuous function, $\mathbf{f}(\lim_{i \rightarrow \infty} \mathbf{f}^i(\mathbf{u})) = \lim_{i \rightarrow \infty} \mathbf{f}^{i+1}(\mathbf{u}) = \lim_{i \rightarrow \infty} \mathbf{f}^i(\mathbf{u})$, so the latter term is a fixed point of \mathbf{f} . Thus $\mu\mathbf{f} \leq \lim_{i \rightarrow \infty} \mathbf{f}^i(\mathbf{u}) \prec \mathbf{u}$. ■

We often denote PPS in the form of an equation system $\mathbf{x} = \mathbf{f}(\mathbf{x})$. In this paper we are mostly concerned with PPS of type (13), as well as with *linear* PPS that can be written in the form $\mathbf{x} = A\mathbf{x} + \mathbf{b}$, for some matrix $A \in \mathbb{R}_{\geq 0}^{J \times J}$ and vector $\mathbf{b} \in \mathbb{R}_{\geq 0}^J$ (cf. Lemma 1).

A PPS \mathbf{f} is called *clean* if $\mu\mathbf{f} \succ \mathbf{0}$. General PPS may be not clean, e.g., $x = 2x$ is not clean. However, it is possible to compute, in linear time [4], [21], a *cleaned-up* variant \mathbb{f} of \mathbf{f} by identifying and removing variables that are assigned zero in the lfp. Formally, the cleaned-up system \mathbb{f} satisfies the property that (i) \mathbb{f} is clean, (ii) \mathbb{f} is indexed by a subset $J' \subseteq J$, (iii) for all $j \in J'$, we have $(\mu\mathbb{f})_j = (\mu\mathbf{f})_j$, and (iv) for all $j \in J \setminus J'$, $(\mu\mathbf{f})_j = 0$. For example, the cleaned-up version of $(x, y) = (2x, 1 + x + y)$ is just $y = 1 + y$ (whose lfp is ∞ , i.e., it is not feasible). *Throughout the paper we consistently use the symbol \mathbf{f} for general, not necessarily clean PPS, and we reserve the symbol \mathbb{f} for clean PPS.*

Let \mathbf{f} be a PPS indexed by J . The *Jacobi matrix* \mathbf{f}' of \mathbf{f} is a $J \times J$ -matrix whose entries are the partial derivatives $(\mathbf{f}')_{i,j} = \frac{\partial}{\partial x_j} f_i$ for $i, j \in J$. The Jacobi matrix with all entries *evaluated* at point $\mathbf{u} \in \mathbb{R}_{\geq 0}^J$ is denoted $\mathbf{f}'(\mathbf{u})$.

For a finite matrix $A \in \mathbb{R}^{J \times J}$, we recall that $\lambda \in \mathbb{C}$ is an *eigenvalue* if there exists $\mathbf{v} \neq \mathbf{0}$ such that $A\mathbf{v} = \lambda\mathbf{v}$. The *spectral radius* $\rho(A)$ is the largest absolute value of the at most $|J|$ eigenvalues of A .

The next theorem plays a central role in the analysis of Newton's method for PPS (see also [4, Lem. 6.5]), and turns out very useful for our purposes as well.

Theorem 2 (See [21, Theorem 4.1 - 2]): For any clean and feasible PPS \mathbb{f} it holds that $\forall \mathbf{x} \prec \mu\mathbb{f}: \rho(\mathbb{f}'(\mathbf{x})) < 1$, and $\rho(\mathbb{f}'(\mu\mathbb{f})) \leq 1$.

The second statement follows from the first as the eigenvalues of any matrix depend continuously on its entries.

The following variant of Taylor's theorem will be crucial in several proofs throughout the paper:

Lemma 4 (Taylor's theorem; see [21, Lemma 2.3]): Let \mathbf{f} be a PPS. For all $\mathbf{u}, \mathbf{v} \geq \mathbf{0}$,

$$\mathbf{f}(\mathbf{u}) + \mathbf{f}'(\mathbf{u})\mathbf{v} \leq \mathbf{f}(\mathbf{u} + \mathbf{v}) \leq \mathbf{f}(\mathbf{u}) + \mathbf{f}'(\mathbf{u} + \mathbf{v})\mathbf{v}.$$

Another variant of Lemma 4 that we will use is this: For $\mathbf{u}, \mathbf{v} \geq \mathbf{0}$ such that $\mathbf{u} - \mathbf{v} \geq \mathbf{0}$,

$$\mathbf{f}(\mathbf{u}) - \mathbf{f}'(\mathbf{u})\mathbf{v} \leq \mathbf{f}(\mathbf{u} - \mathbf{v}) \leq \mathbf{f}(\mathbf{u}) - \mathbf{f}'(\mathbf{u} - \mathbf{v})\mathbf{v}.$$

See Appendix B for a proof.

Non-negative matrices: We also need some results from the theory of matrices with non-negative entries. We will mostly apply these results to the Jacobi matrix $\mathbf{f}'(\mathbf{u})$ of a PPS \mathbf{f} , which indeed has non-negative entries for $\mathbf{u} \geq \mathbf{0}$.

Theorem 3 (See [37, Section 10.3, Fact 12e]): Let $A \geq 0$ be a finite-dimensional non-negative square matrix. The following are equivalent:

- 1) $\lim_{k \rightarrow \infty} A^k = 0$
- 2) $\rho(A) < 1$
- 3) $(I - A)^{-1}$ exists and is non-negative
- 4) $\exists \mathbf{v} \succ \mathbf{0}: A\mathbf{v} \prec \mathbf{v}$

Moreover, under either one of the above conditions, $\sum_{k=0}^{\infty} A^k = (I - A)^{-1}$.

We will moreover use the following results concerning the spectral radius of a non-negative matrix:

- Lemma 5:* Let $A \geq 0$ be a non-negative finite square matrix.
- The spectral radius $\rho(A)$ is an eigenvalue of A [37, Section 10.3, Fact 1a].
 - If $0 \leq B \leq A$, then $\rho(B) \leq \rho(A)$ [37, Section 10.3, Fact 7].
 - Let $c \in \mathbb{R}_{\geq 0}$. Then $(\exists \mathbf{u} \succ \mathbf{0}: A\mathbf{u} \leq c\mathbf{u}) \implies \rho(A) \leq c$ [37, Section 10.3 Fact 6a].

IV. CERTIFYING UPPER BOUNDS

In this section, we study certificates for upper bounds on the lfp of a general PPS. We do not (yet) consider pPDA.

Definition 4 (Non-singular PPS): A feasible PPS \mathbf{f} is called *non-singular* if the matrix $I - \mathbf{f}'(\mu\mathbf{f})$ has an inverse.

Lemma 6: Let \mathbf{f} be feasible, non-singular, and clean. Then the conditions from Theorem 3 apply to $\mathbf{f}'(\mu\mathbf{f}) \geq \mathbf{0}$.

Proof: Since $I - \mathbf{f}'(\mu\mathbf{f})$ is non-singular, 1 is *not* an eigenvalue of $\mathbf{f}'(\mu\mathbf{f})$. By Theorem 2, $\rho(\mathbf{f}'(\mu\mathbf{f})) \leq 1$. Since the spectral radius of a non-negative matrix is an eigenvalue (Lemma 5), it follows that $\rho(\mathbf{f}'(\mu\mathbf{f})) < 1$, which is the second condition of Theorem 3. ■

In Lemma 6 it is essential that \mathbf{f} is clean: Consider e.g. $f: x = 2x$. Then $1 - f'(\mu f) = 1 - 2 = -1$ which has inverse $-1 < 0$. This is the reason why we often consider clean PPS throughout the paper.

The following generalizes a part of [22, Lem. 2].

Theorem 4 (Certificates for upper bounds on the lfp): Let \mathbf{f} be a feasible and clean PPS indexed by J . Then \mathbf{f} is non-singular if and only if $\exists \mathbf{u} \in \mathbb{R}_{>0}^J: \mathbf{f}(\mathbf{u}) \prec \mathbf{u}$ (and thus $\mu\mathbf{f} \prec \mathbf{u}$ by Lemma 3.)

In this case, there exists a *rational* \mathbf{u} arbitrarily close to the lfp $\mu\mathbf{f}$, i.e., $\forall \varepsilon > 0: \exists \mathbf{u} \in \mathbb{Q}_{>0}^J: \mathbf{f}(\mathbf{u}) \prec \mathbf{u} \wedge \|\mu\mathbf{f} - \mathbf{u}\|_{\infty} < \varepsilon$.

Proof of Theorem 4: Assume \mathbf{f} is non-singular. By the right inequality of Lemma 4, $\mathbf{f}(\mu\mathbf{f} + \mathbf{v}) \leq \mu\mathbf{f} + \mathbf{f}'(\mu\mathbf{f} + \mathbf{v})\mathbf{v}$

for all $\mathbf{v} \geq \mathbf{0}$. By Lemma 6 and Theorem 3, $\mathbf{f}'(\mu\mathbf{f})\mathbf{v} \prec \mathbf{v}$ for some $\mathbf{v} \succ \mathbf{0}$. Since the function $\mathbf{x} \mapsto \mathbf{f}'(\mathbf{x})$ is continuous, we can choose $\delta > 0$ s.t. $\mathbf{f}'(\mu\mathbf{f} + \delta\mathbf{v})\mathbf{v} \prec \mathbf{v}$ still holds. Now

$$\mathbf{f}(\mu\mathbf{f} + \delta\mathbf{v}) \leq \mu\mathbf{f} + \delta\mathbf{f}'(\mu\mathbf{f} + \delta\mathbf{v})\mathbf{v} \prec \mu\mathbf{f} + \delta\mathbf{v}, \quad (14)$$

so the claim holds with $\mathbf{u} := \mu\mathbf{f} + \delta\mathbf{v} \succ \mathbf{0}$. Furthermore, note that we can choose $\delta > 0$ sufficiently small s.t. $\|\mathbf{u} - \mu\mathbf{f}\|_{\infty} < \varepsilon$ for any given $\varepsilon > 0$. Since \mathbf{f} is continuous, $\mathbf{f}(\mathbf{u}') \prec \mathbf{u}'$ holds for all \mathbf{u}' in an open ball around \mathbf{u} ; such a ball contains rational vectors.

For the other direction assume that there exists $\mathbf{u} \succ \mathbf{0}$ with $\mathbf{f}(\mathbf{u}) \prec \mathbf{u}$. By Lemma 3, $\mu\mathbf{f} \prec \mathbf{u}$, and so $\mathbf{u} = \mu\mathbf{f} + \mathbf{v}$ for some $\mathbf{v} \succ \mathbf{0}$. The left inequality of Lemma 4 yields:

$$\mu\mathbf{f} + \mathbf{f}'(\mu\mathbf{f})\mathbf{v} \leq \mathbf{f}(\mu\mathbf{f} + \mathbf{v}) \prec \mu\mathbf{f} + \mathbf{v}. \quad (15)$$

As \mathbf{f} is feasible, $\mu\mathbf{f} \prec \infty$, and thus $\mathbf{f}'(\mu\mathbf{f})\mathbf{v} \prec \mathbf{v}$ by (15). The latter statement is the last condition from Theorem 3, so $I - \mathbf{f}'(\mu\mathbf{f})$ has an inverse (which is also non-negative). ■

We show in Section VI that in the case of PPS associated with pPDA, non-singularity is equivalent to cPAST.

V. CERTIFYING LOWER BOUNDS

In this section we show that it is also possible to certify lower bounds in general PPS. The contents of this section are not crucial for understanding the subsequent sections.

Recall that the simple induction principle from Lemma 3 does *not* work for lower bounds on the lfp. Indeed, if $f: L \rightarrow L$ is a monotonic function and L a complete lattice, then we only have the “co-induction” rule

$$\forall l \in L: \quad l \leq f(l) \implies l \leq \nu f,$$

where νf is the *greatest* fixed point of f . As a consequence, the implication $l \leq f(l) \implies l \leq \mu f$ holds for all $l \in L$ if and only if f has a *unique* fixed point. However, a PPS \mathbf{f} may have multiple fixed points in general. For example, the one-variable PPS $x = f(x) = x^2 + \frac{1}{5}$ satisfies $1 \leq f(1) = \frac{6}{5}$, but $\mu f = \frac{1}{2} - \frac{1}{2\sqrt{5}} < 1$.

We circumvent this problem by providing a co-inductive point \mathbf{l} along with a (strict) inductive upper bound \mathbf{u} :

Theorem 5 (Certificates for lower bounds on the lfp): Let \mathbf{f} be a PPS indexed by J . Then for all $\mathbf{l}, \mathbf{u} \in \mathbb{R}_{\geq 0}^J$,

$$\mathbf{l} \leq \mathbf{f}(\mathbf{l}) \wedge \mathbf{l} \leq \mathbf{u} \wedge \mathbf{f}(\mathbf{u}) \prec \mathbf{u} \implies \mathbf{l} \leq \mu\mathbf{f} \prec \mathbf{u}.$$

Moreover, if \mathbf{f} is feasible, non-singular and clean, then for all $\varepsilon > 0$ there exist $\mathbf{l}, \mathbf{u} \in \mathbb{Q}_{\geq 0}^J$ satisfying the above conditions such that $\|\mathbf{l} - \mathbf{u}\|_{\infty} \leq \varepsilon$.

Proof: Let \mathbf{l}, \mathbf{u} be such that $\mathbf{l} \leq \mathbf{f}(\mathbf{l})$, $\mathbf{l} \leq \mathbf{u}$, and $\mathbf{f}(\mathbf{u}) \leq \mathbf{u}$. The first step of the proof is to construct a fixed point $\tilde{\mathbf{u}} \prec \mathbf{u}$. Let $\tilde{\mathbf{u}} = \lim_{n \rightarrow \infty} \mathbf{f}^n(\mathbf{u}) \prec \mathbf{u}$. By monotonicity, we have $\mathbf{l} \leq \mathbf{f}^n(\mathbf{l}) \leq \mathbf{f}^n(\mathbf{u})$ for all $n \geq 0$, so $\mathbf{l} \leq \tilde{\mathbf{u}}$. Moreover,

\tilde{u} is a fixed point of f since the latter is continuous, so by the variant of Taylor's theorem (below Lemma 4),

$$\tilde{u} = f(\tilde{u}) = f(u - (u - \tilde{u})) \leq f(u) - f'(\tilde{u})(u - \tilde{u}) \quad (16)$$

$$\implies f'(\tilde{u})(u - \tilde{u}) \leq f(u) - \tilde{u} \prec u - \tilde{u} . \quad (17)$$

The last inequality implies that $\rho(f'(\tilde{u})) < 1$ by Theorem 3. We will need this further below.

We now show that $\mu f = \tilde{u}$. Towards contradiction assume that $\tilde{u} > \mu f$, i.e., $\tilde{u} - \mu f > \mathbf{0}$. By the right inequality of Taylor's theorem,

$$f(\tilde{u}) = f(\mu f + (\tilde{u} - \mu f)) \leq \mu f + f'(\tilde{u})(\tilde{u} - \mu f). \quad (18)$$

From (18) and $\tilde{u} = f(\tilde{u})$ it follows that

$$\tilde{u} - \mu f = f(\tilde{u}) - \mu f \leq f'(\tilde{u})(\tilde{u} - \mu f) . \quad (19)$$

Using (19) and the fact that $f'(\tilde{u})$ is non-negative, it follows that $\tilde{u} - \mu f \leq f'(\tilde{u})^n(\tilde{u} - \mu f)$ for all $n \geq 0$. However, since $\rho(f'(\tilde{u})) < 1$, we have by Theorem 3 that $\lim_{n \rightarrow \infty} f'(\tilde{u})^n = 0$, so there exists n such that $f'(\tilde{u})^n(\tilde{u} - \mu f) < \tilde{u} - \mu f$, contradiction. It follows that $\mu f = \tilde{u}$. As $l \leq \tilde{u}$ we have shown that l is indeed a lower bound on the lfp of f .

We now prove the second part of the theorem. Let $f = \mathbb{f}$ be feasible, non-singular and clean. A rational u with $\mathbb{f}(u) \prec u$ arbitrarily close to $\mu \mathbb{f}$ exists by Theorem 4. It remains to show existence of a rational co-inductive $l \leq \mathbb{f}(l)$. The argument is analogous to the proof of Theorem 4, but this time for lower bounds: Since \mathbb{f} feasible, non-singular and clean, Lemma 6 and Theorem 3 yield that there exists $v \succ \mathbf{0}$ such that $\mathbb{f}'(\mu \mathbb{f})v \prec v$. Choose such a $v \leq \mu \mathbb{f}$. By the variant of Taylor's theorem,

$$\mu \mathbb{f} - v \prec \mu \mathbb{f} - \mathbb{f}'(\mu \mathbb{f})v \leq \mathbb{f}(\mu \mathbb{f} - v) \quad (20)$$

With $l = \mu \mathbb{f} - v$ it follows that $l \prec \mathbb{f}(l)$. Moreover, $v \succ \mathbf{0}$ can be chosen arbitrarily small, and hence l arbitrarily close to $\mu \mathbb{f}$. Finally, by continuity of \mathbb{f} , the inequality $l' \prec \mathbb{f}(l')$ holds for all l' in an open ball around l ; such a ball contains rational vectors. ■

VI. CFAST AND NON-SINGULAR PPS

In this section we characterize non-singularity in terms of expected runtimes. In a nutshell, the main result is that the (cleaned-up) fundamental PPS of a pPDA Δ is non-singular iff Δ is cFAST. We begin by characterizing the 1st termination moments as the solution of a finite linear equation system.

Lemma 7 (System of termination moments): Let $\Delta = (Q, \Gamma, \delta)$ be a pPDA and let $(E\langle pZ|q \rangle)_{pZq \in Q \times \Gamma \times Q}$ be $\mathbb{R}_{\geq 0}$ -valued variables. Then the 1st termination moments $E[pZ|q]$ constitute the least solution of the following linear PPS:

$$\begin{aligned} E\langle pZ|q \rangle &= [pZ|q] + \sum_{pZ \xrightarrow{\alpha} rY} a \cdot E\langle rY|q \rangle \\ &+ \sum_{pZ \xrightarrow{\alpha} rYX} a \cdot (E\langle rY|t \rangle \cdot [tX|q] + [rY|t] \cdot E\langle tX|q \rangle) \end{aligned} \quad (21)$$

where $[pZ|q]$, etc., are the return probabilities of Δ .

Proof sketch: We sketch a proof based on a specific type of formal power series known as *probability generating functions*, e.g. [34, App. F], [38]. We will use the fact that Theorem 1 actually holds in a more general setting where the transition weights come from a *semiring* with suitable closure properties [17]. The *semiring* $\mathbb{R}_{\geq 0}[[z]]$ of formal power series with coefficients in $\mathbb{R}_{\geq 0}$ satisfies these properties. Here, z is a *formal variable*, and the elements in $\mathbb{R}_{\geq 0}[[z]]$ are *formal series* of the form $\sum_{k=0}^{\infty} a_k z^k$, with coefficients $a_k \in \mathbb{R}_{\geq 0}$. See [17] for details.

Let $z\Delta$ be the $\mathbb{R}_{\geq 0}[[z]]$ -weighted PDA obtained from Δ by multiplying each transition probability by z . The 0th termination moment is then a power series $[pZ|q]_{z\Delta} = \sum_{k=0}^{\infty} a_k z^k$ where for all $k \geq 0$, the coefficient a_k equals the sum of the probabilities of the proper paths of length k between pZ and $q\epsilon$. Consider the usual differentiation operator ∂ for formal power series. We have

$$\partial[pZ|q]_{z\Delta} = \sum_{k=1}^{\infty} k \cdot a_k z^{k-1} \quad (22)$$

and thus by definition of the termination moments,

$$\partial[pZ|q]_{z\Delta}|_{z=1} = E[pZ|q], \text{ and also} \quad (23)$$

$$[pZ|q]_{z\Delta}|_{z=1} = [pZ|q], \quad (24)$$

where $\cdot|_{z=1} : \mathbb{R}_{\geq 0}[[z]] \rightarrow \mathbb{R}_{\geq 0}$ is the function that substitutes the formal variable z for 1 and takes the resulting limit (which may be ∞). By (the extended) Theorem 1, the triples $[pZ|q]_{z\Delta}$ are the least solution in $\mathbb{R}_{\geq 0}[[z]]$ of the following equation system with variables $\langle pZ|q \rangle$:

$$\begin{aligned} \langle pZ|q \rangle &= z \cdot \left(\sum_{pZ \xrightarrow{\alpha} q\epsilon} a + \sum_{pZ \xrightarrow{\alpha} rY} a \cdot \langle rY|q \rangle \right. \\ &\quad \left. + \sum_{pZ \xrightarrow{\alpha} rYX} a \cdot \sum_{t \in Q} \langle rY|t \rangle \cdot \langle tX|q \rangle \right) . \end{aligned} \quad (25)$$

Differentiating all equations in this system using the product rule yields the following *additional* equations:

$$\begin{aligned} \partial \langle pZ|q \rangle &= \langle pZ|q \rangle + z \cdot \left(\sum_{pZ \xrightarrow{\alpha} rY} a \cdot \partial \langle rY|q \rangle \right. \\ &\quad \left. + \sum_{pZ \xrightarrow{\alpha} rYX} a \cdot \sum_{t \in Q} (\partial \langle rY|t \rangle \cdot \langle tX|q \rangle + \langle rY|t \rangle \cdot \partial \langle tX|q \rangle) \right) \end{aligned} \quad (26)$$

where $\partial \langle pZ|q \rangle$, $\partial \langle rY|t \rangle$, etc., are fresh variable symbols. It can be shown that the power series $\partial[pZ|q]_{z\Delta}$ and $[pZ|q]_{z\Delta}$ constitute the least solution of the overall system (25) \wedge (26). Finally, since the substitution operation $\cdot|_{z=1}$ is a semiring-homomorphism we may apply it to the equations in (26), and use (23) and (24) to obtain (21). ■

Remark 1: It is also possible to prove Lemma 7 using the equation system from [7] for the conditional expected runtimes $\frac{E[pZ|q]}{[pZ|q]}$ (see Appendix C). However, we believe our proof using generating functions is of interest as it allows finding equation systems for *any* higher termination moment by repeatedly differentiating (25).

Lemma 8 (System of termination moments): Let Δ be a pPDA. The linear PPS (21) can be written in matrix form as follows:

$$\mathbf{x} = \mu\mathbf{f} + \mathbf{f}'(\mu\mathbf{f})\mathbf{x}, \quad (27)$$

where \mathbf{f} is the fundamental PPS of Δ and the variable vector \mathbf{x} comprises the symbols $E\langle pZ|q \rangle$.

Proof: For a general PPS \mathbf{g} indexed by J and $i \in J$,

$$(\mathbf{g}'(\mu\mathbf{g})\mathbf{x})_i = \sum_{j \in J} \left(\frac{\partial}{\partial_j} \mathbf{g}_i \right) (\mu\mathbf{g}) \cdot \mathbf{x}_j. \quad (28)$$

Using (28) and the fact that (27) is indexed by $Q \times \Gamma \times Q$, we can write the equation in (27) corresponding to $E\langle pZ|q \rangle$ as follows:

$$\begin{aligned} E\langle pZ|q \rangle &= [pZ|q] + \sum_{p'Z'q' \in Q \times \Gamma \times Q} \frac{\partial}{\partial_{E\langle p'Z'|q' \rangle}} \left(\right. \\ &\quad \left. \sum_{pZ \xrightarrow{a} q\varepsilon} a + \sum_{pZ \xrightarrow{a} rY} a \cdot E\langle rY|q \rangle + \sum_{\substack{pZ \xrightarrow{a} rYX \\ t \in Q}} a \cdot E\langle rY|t \rangle \cdot E\langle tX|q \rangle \right) (\mu\mathbf{f}) \cdot E\langle p'Z'|q' \rangle \end{aligned} \quad (29)$$

$$\begin{aligned} &= [pZ|q] + \sum_{pZ \xrightarrow{a} rY} a \cdot E\langle rY|q \rangle \\ &+ \sum_{pZ \xrightarrow{a} rYX} a \cdot \sum_{t \in Q} (E\langle rY|t \rangle \cdot [tX|q] + [rY|t] \cdot E\langle tX|q \rangle) \end{aligned} \quad (30)$$

To transform (29) into (30), we have used (among similar arguments) that for all $rYt \in Q \times \Gamma \times Q$,

$$\begin{aligned} &\left(\frac{\partial}{\partial_{E\langle rY|t \rangle}} a \cdot E\langle rY|t \rangle \cdot E\langle tX|q \rangle \right) (\mu\mathbf{f}) \cdot E\langle rY|t \rangle \\ &= \left(a \cdot E\langle tX|q \rangle \right) (\mu\mathbf{f}) \cdot E\langle rY|t \rangle \\ &= a \cdot E\langle rY|t \rangle \cdot [tX|q]. \quad (\text{as } \mu\mathbf{f}_{tXq} = [tX|q]) \end{aligned}$$

(30) is identical to Equation (21). \blacksquare

One can interpret (27) as the pPDA variant of (9). In fact, (9) can be proved using the same generating function trick as in the proof of Lemma 7.

We need one additional lemma before showing the main result of this section.

Lemma 9: Let \mathbb{f} be a feasible and clean linear PPS of the form $\mathbf{x} = A\mathbf{x} + \mathbf{b}$. Then the equivalent conditions from Theorem 3 apply to the matrix A . In particular, \mathbb{f} is non-singular.

Proof: Note that \mathbb{f}' is a constant matrix. Since \mathbb{f} is feasible and clean, we have $\rho(\mathbb{f}') < 1$ by Theorem 2, which is the second condition from Theorem 3. \blacksquare

Theorem 6 (Non-singular iff cPAST): Let $\Delta = (Q, \Gamma, \delta)$ be a pPDA with cleaned-up fundamental PPS \mathbb{f} . Then \mathbb{f} is non-singular if and only if Δ is cPAST.

Proof: The cleaned-up variant of (27) is

$$\mathbf{x} = \mu\mathbb{f} + \mathbb{f}'(\mu\mathbb{f})\mathbf{x}. \quad (31)$$

This is because $[pZ|q] = 0$ iff $E[pZ|q] = 0$.

Now suppose that \mathbb{f} is non-singular, i.e., $(I - \mathbb{f}'(\mu\mathbb{f}))^{-1}$ exists. Then (31) has the unique *real-valued* solution $(I - \mathbb{f}'(\mu\mathbb{f}))^{-1} \mu\mathbb{f}$. In particular, all the termination moments $E[pZ|q]$ differ from ∞ .

On the other hand, if Δ is cPAST, then (31) is feasible. As (31) is clean, $\mathbb{f}'(\mu\mathbb{f})$ is non-singular by Lemma 9. \blacksquare

As claimed in Figure 2, Theorem 4 and Theorem 6 together imply that arbitrarily tight rational-valued inductive upper bounds on the lfp of the fundamental system \mathbf{f} of a pPDA Δ exist if Δ is cPAST.

VII. CERTIFYING PAST

In this section we are concerned with certificates for positive almost-sure termination (PAST; see Def. 3). We begin by characterizing the expected runtimes in terms of a linear equation system.

Theorem 7 (System of expected runtimes): Let $\Delta = (Q, \Gamma, \delta)$ be a pPDA and let $(\text{ert}\langle pZ \rangle)_{pZ \in Q \times \Gamma}$ be $\mathbb{R}_{\geq 0}$ -valued variables. The expected runtimes $\text{ert}[pZ]$ constitute the least solution of the following linear PPS:

$$\begin{aligned} \text{ert}\langle pZ \rangle &= 1 + \sum_{pZ \xrightarrow{a} rY} a \cdot \text{ert}\langle rY \rangle \\ &+ \sum_{pZ \xrightarrow{a} rYX} a \cdot (\text{ert}\langle rY \rangle + \sum_{t \in Q} [rY|t] \cdot \text{ert}\langle tX \rangle) \end{aligned} \quad (32)$$

Proof: We give an automata theoretic proof. The idea is to construct a *non-probabilistic* weighted PDA Δ' such that for all $pZ \in Q \times \Gamma$, $\text{ert}[pZ] = [pZ|q^*]_{\Delta'}$, where q^* is a distinguished state in Δ' . In other words, we express the expected runtimes of Δ in terms of the 0th termination moments of a new automaton.

Given a pPDA $\Delta = (Q, \Gamma, \delta)$, we construct the weighted PDA $\Delta' = (Q', \Gamma, \delta')$ as follows: $Q' = Q \cup \{q^*\}$, and

$$\begin{aligned} \delta' &= \delta \cup \{pZ \xrightarrow{1} q^*Z \mid pZ \in Q \times \Gamma\} \\ &\cup \{q^*Z \xrightarrow{1} q^*\varepsilon \mid Z \in \Gamma\}. \end{aligned} \quad (33)$$

Intuitively, Δ' is obtained from Δ by adding a new state q^* and weight-1 transitions from every state of Δ to q^* (these new transitions are independent of the current topmost stack symbol). Moreover, once q^* is reached, Δ' simply empties its stack. Note that Δ' is not probabilistic anymore.

Fix an initial configuration $pZ \in Q \times \Gamma$. Consider the 0th termination moment $[pZ|q^*]_{\Delta'} \in \mathbb{R}_{\geq 0}$ which is the sum of the weights of all proper finite paths

$$\{pZ\} \cdot (Q \times \Gamma^+)^* \cdot (\{q^*\} \times \Gamma^+) \cdot \{q^*\varepsilon\} \quad (34)$$

between pZ and $q^*\varepsilon$ in $\mathfrak{T}_{\Delta'}$. It follows immediately from the construction of Δ' that there is a weight-preserving bijection between the paths (34) and the paths in \mathfrak{T}_{Δ} that have not yet reached a configuration with empty stack, i.e., the paths

$$\{pZ\} \cdot (Q \times \Gamma^+)^* \cdot \quad (35)$$

As explained below Def. 2, the sum of the probabilities of the paths (35) is exactly $\text{ert}[pZ]_\Delta$. Thus, $[pZ|q^*]_{\Delta'} = \text{ert}[pZ]_\Delta$.

To finish the proof we use the fact that we have access to the 0th moments $[pZ|q^*]_{\Delta'}$ via the fundamental PPS $\mathbf{f}_{\Delta'}$ of Δ' . By construction of Δ' , $\mathbf{f}_{\Delta'}$ contains the equations from \mathbf{f}_Δ , as well as for each $pZ \in Q \times \Gamma$ the additional equations

$$\begin{aligned} \langle pZ|q^* \rangle &= \langle q^*Z|q^* \rangle + \sum_{pZ \xrightarrow{a} rY} a \cdot \langle rY|q^* \rangle \\ &+ \sum_{pZ \xrightarrow{a} rYX} a \cdot \left(\sum_{t \in Q} \langle rY|t \rangle \cdot \langle tX|q^* \rangle + \langle rY|q^* \rangle \cdot \langle q^*X|q^* \rangle \right) \end{aligned} \quad (36)$$

$$\langle q^*Z|q^* \rangle = 1, \quad (37)$$

where the transitions under the summation symbols refer to those of Δ (not to those of Δ'). Simplifying and identifying $\langle pZ|q^* \rangle$ with $\text{ert}[pZ]$ yields (32). ■

Remark 2: If one assumes upfront that the pPDA Δ is AST, then it is not difficult to derive equation system (32) from [7, Thm 3.1/Ex. 3.3] (see Appendix D for details). However, Theorem 7 also holds if Δ is not AST. That is, if $\sum_{q \in Q} [pZ|q] < 1$ for some $pZ \in Q \times \Gamma$, then $\text{ert}[pZ] = \infty$ in the least solution of (32). In other words, system (32) has a non-negative real-valued solution (i.e., it is feasible) iff Δ is PAST. The direction from left to right does not easily follow from [7]. However, this direction is crucial for proving/certifying PAST without explicitly proving AST beforehand (see Corollary 1 and Theorem 8 further below). The fact that PAST can be proved directly without proving AST first is a major insight of this paper.

System (32) can also be written in matrix form as

$$\mathbf{r} = M_\Delta(\mu\mathbf{f})\mathbf{r} + \mathbf{1}, \quad (38)$$

where $\mathbf{r} = (\text{ert}[pZ])_{pZ \in Q \times \Gamma}$, and $M_\Delta(\mu\mathbf{f})$ is a matrix that depends on the transitions of Δ and the return probabilities $\mu\mathbf{f}$; the exact entries of $M_\Delta(\mu\mathbf{f})$ can be read off from (32).

Corollary 1 (Proving PAST): Let Δ be a pPDA with fundamental PPS \mathbf{f} and suppose that $\mu\mathbf{f} \leq \mathbf{u}$. If $M_\Delta(\mathbf{u})\mathbf{r} + \mathbf{1} \leq \mathbf{r}$ has a solution $\mathbf{r} \in \mathbb{R}_{\geq 1}^{Q \times \Gamma}$, then Δ is PAST. In this case, $\text{ert}[pZ] \leq r_{pZ}$ for all $pZ \in Q \times \Gamma$.

Proof: We have $M_\Delta(\mathbf{u}) \geq M_\Delta(\mu\mathbf{f})$ entry-wise. Thus any $\mathbf{r} \geq \mathbf{0}$ with $M_\Delta(\mathbf{u})\mathbf{r} + \mathbf{1} \leq \mathbf{r}$ is a component-wise upper bound on the least solution \mathbf{r}' of $M_\Delta(\mu\mathbf{f})\mathbf{r}' + \mathbf{1} = \mathbf{r}'$. But by Theorem 7, the components of \mathbf{r}' are exactly the expected runtimes. ■

Corollary 1 allows turning any iterative scheme (e.g. [21], [22]) for computing a monotonically converging sequence

$$\mathbf{u}^{(0)} \geq \mathbf{u}^{(1)} \geq \dots \geq \mu\mathbf{f}, \quad \lim_{i \rightarrow \infty} \mathbf{u}^{(i)} = \mu\mathbf{f}, \quad (39)$$

of upper bounds on the lfp of \mathbf{f} into a semi-algorithm that detects in finite time if a given pPDA Δ is PAST or non-AST, and that may diverge only if Δ is AST but not PAST. For $i = 0, 1, 2, \dots$, after computing $\mathbf{u}^{(i)}$, the semi-algorithm simply checks if

- 1) $M_\Delta(\mathbf{u}^{(i)})\mathbf{r} + \mathbf{1} \leq \mathbf{r}$ is feasible in $\mathbb{R}_{\geq 0}$ in which case it reports “PAST”, or else

- 2) if $\sum_t [pZ|t] < 1$ for some $pZ \in Q \times \Gamma$ in which case it reports “non-AST”.

- 3) If neither (a) nor (b) apply, then continue with $i + 1$.

The correctness of this algorithm is straightforward with Corollary 1. Termination in the PAST case follows because $\mathbf{u}^{(i)}$ is eventually sufficiently close to $\mu\mathbf{f}$ such that $M_\Delta(\mathbf{u}^{(i)})\mathbf{r} + \mathbf{1} \leq \mathbf{r}$ has a solution in $\mathbb{R}_{\geq 0}$.

Theorem 8 (Certifying PAST): Let $\Delta = (Q, \Gamma, \delta)$ be a pPDA with fundamental PPS \mathbf{f} . Δ is PAST if and only if $\exists \mathbf{u} \in \mathbb{Q}_{\geq 0}^{Q \times \Gamma \times Q}, \mathbf{r} \in \mathbb{Q}_{\geq 1}^{Q \times \Gamma} : \mathbf{f}(\mathbf{u}) \leq \mathbf{u}$ and $M_\Delta(\mathbf{u})\mathbf{r} + \mathbf{1} \leq \mathbf{r}$, where M_Δ is the matrix from (38).

Proof: Assume that Δ is PAST. Recall from Lemma 2 that PAST implies cPAST. By Theorem 6, \mathbb{f} , the cleaned-up version of \mathbf{f} , is non-singular. By Theorem 4, there exists $\mathbf{u}' \in \mathbb{Q}_{> 0}^{J'}$, where $J' \subseteq Q \times \Gamma \times Q$ is the index set of \mathbb{f} , such that $\mathbb{f}(\mathbf{u}') < \mathbf{u}'$. Moreover, such a \mathbf{u}' can be chosen arbitrarily close to $\mu\mathbb{f}$. We can extend \mathbf{u}' to a vector indexed by \mathbf{u} in $Q \times \Gamma \times Q$ by setting the components that belong to variables not present in \mathbb{f} to zero; we then still have $\mathbf{f}(\mathbf{u}) \leq \mathbf{u}$. Since Δ is PAST, $M_\Delta(\mu\mathbf{f})\mathbf{r} + \mathbf{1} = \mathbf{r}$ is feasible by Theorem 7, i.e., it has a solution in $\mathbb{R}_{\geq 1}^{Q \times \Gamma}$. The entries of $M_\Delta(\bullet)$ depend continuously on the argument \bullet , so we can choose \mathbf{u} sufficiently close to $\mu\mathbf{f}$ such that $M_\Delta(\mathbf{u})\mathbf{r} + \mathbf{1} = \mathbf{r}$ is still feasible. As $M_\Delta(\mathbf{u})\mathbf{r} + \mathbf{1} = \mathbf{r}$ is obviously clean, $M_\Delta(\mathbf{u})$ is non-singular by Lemma 9. Since $M_\Delta(\mathbf{u})$ has rational entries, the unique solution of $\mathbf{r} = (\mathbf{I} - M_\Delta(\mathbf{u}))^{-1}\mathbf{1}$ is thus also rational.

The other direction follows from Corollary 1 using $\mathbf{f}(\mathbf{u}) \leq \mathbf{u} \implies \mu\mathbf{f} \leq \mathbf{u}$. ■

Our results also admit a simple proof of the following result which is already known in the literature [26, Chap. 4], [25, Thm. 3]. A pBPA is a pPDA with $|Q| = 1$.

Corollary 2 (PAST in pBPA): Let Δ be a pBPA with fundamental PPS \mathbf{f} . Then Δ is PAST iff $\mathbf{r} = \mathbf{1} + \mathbf{f}'(\mathbf{1})\mathbf{r}$ has a unique solution $\mathbf{r} \in \mathbb{Q}_{\geq 1}^\Gamma$. In this case, $\text{ert}[Z] = r_Z$ for all $Z \in \Gamma$. As a consequence, deciding PAST for pBPA can be done in PTIME, and the bit complexity of the expected runtimes is polynomial in the encoding size of Δ .

Proof: First note that since Δ has a single state, the termination moments and expected runtimes can be written as $E^k[Z]$ and $\text{ert}[Z]$, respectively, where $Z \in \Gamma$. Moreover, the system (21) of 1st moments collapses to

$$E\langle Z \rangle = [Z] + \sum_{Z \xrightarrow{a} Y} a \cdot E\langle Y \rangle + \sum_{Z \xrightarrow{a} YX} a \cdot (E\langle Y \rangle \cdot [X] + [Y] \cdot E\langle X \rangle) \quad (40)$$

and the system (32) of expected runtimes collapses to

$$\text{ert}\langle Z \rangle = 1 + \sum_{Z \xrightarrow{a} Y} a \cdot \text{ert}\langle Y \rangle + \sum_{Z \xrightarrow{a} YX} a \cdot (\text{ert}\langle Y \rangle + [Y] \cdot \text{ert}\langle X \rangle). \quad (41)$$

Now suppose Δ is PAST. Then Δ is also AST, so for all $Z \in \Gamma$, $[Z] = 1$, i.e., $\mu\mathbf{f} = \mathbf{1}$. Then (40) and (41) are just the same equation system, so $E[Z] = \text{ert}[Z]$ for all $Z \in \Gamma$.

By Lemma 8, (40) can be written in matrix form as $\mathbf{x} = \mathbf{1} + \mathbf{f}'(\mathbf{1})\mathbf{x}$. The latter is a linear, clean and feasible PPS, so by Lemma 9, it has the unique solution $\mathbf{x} = (\mathbf{I} - \mathbf{f}'(\mathbf{1}))^{-1}\mathbf{1}$. Since $\mathbf{f}'(\mathbf{1})$ has rational entries, it follows from standard linear algebra that the solution is rational and can be computed in polynomial time. In particular, the rational numbers in the solution can be encoded with polynomially many bits.

For the other direction suppose that $\mathbf{r} = \mathbf{1} + \mathbf{f}'(\mathbf{1})$ is feasible. But then (41) is also feasible, so Δ is PAST. ■

VIII. COMPLEXITY OF CERTIFICATES

Our results so far have been *qualitative*, i.e., we have shown existence of certificates but we have not yet discussed how large (the rational numbers in) these certificates are. This is addressed in the present section.

From Theorem 6 it follows that $\rho(\mathbf{f}'_\Delta(\mu\mathbf{f}_\Delta)) < 1$ iff Δ is cPAST. The next lemma is a quantitative generalization:

Lemma 10 (Bound on spectrum): Let Δ be a pPDA with cleaned-up fundamental PPS \mathbf{f} , and suppose that Δ is cPAST. Let $C = \max_{pZq} \frac{\mathbb{E}[pZ|q]}{\mathbb{E}[pZ|q]} \geq 1$ be the maximal conditional expected runtime of Δ , and $\mu\mathbf{f}_{\min} = \min_{pZq} [pZ|q]$ be the minimum non-zero return probability. Then there exists $\mathbf{v} \succ \mathbf{0}$, $\|\mathbf{v}\|_1 = 1$, s.t. $\mathbf{f}'(\mu\mathbf{f})\mathbf{v} \leq (1 - C^{-1})\mathbf{v}$, and $\mathbf{v}_{\min} \geq \frac{\mu\mathbf{f}_{\min}}{C|Q|^2|\Gamma|}$.

As a consequence, $\rho(\mathbf{f}'(\mu\mathbf{f})) \leq (1 - C^{-1})$.

Proof: Let \mathbf{v} be the vector of positive termination moments, i.e., the unique solution of $\mathbf{v} = \mu\mathbf{f} + \mathbf{f}'(\mu\mathbf{f})\mathbf{v} \succ \mathbf{0}$. Then

$$\mathbf{f}'(\mu\mathbf{f})\mathbf{v} = \mathbf{v} - \mu\mathbf{f} \leq c\mathbf{v} \quad (42)$$

where

$$\begin{aligned} c &= \max_{pZq} \frac{\mathbb{E}[pZ|q] - [pZ|q]}{\mathbb{E}[pZ|q]} = 1 - \min_{pZq} \frac{[pZ|q]}{\mathbb{E}[pZ|q]} \\ &= 1 - \left(\max_{pZq} \frac{\mathbb{E}[pZ|q]}{[pZ|q]} \right)^{-1} = 1 - C^{-1}. \end{aligned}$$

By Lemma 5, (42) implies that $c = 1 - C^{-1}$ is an upper bound on $\rho(\mathbf{f}'(\mu\mathbf{f}))$. We can normalize \mathbf{v} (in the $\|\cdot\|_1$ norm) as follows: $\mathbf{v}' = \frac{\mathbf{v}}{\sum_{pZq} \mathbb{E}[pZ|q]}$, $\|\mathbf{v}'\|_1 = 1$. Then

$$\begin{aligned} \mathbf{v}'_{\min} &\geq \frac{\min_{pZq} \mathbb{E}[pZ|q]}{|Q|^2|\Gamma| \max_{pZq} \mathbb{E}[pZ|q]} \\ &\geq \frac{\min_{pZq} [pZ|q]}{|Q|^2|\Gamma| \max_{pZq} \frac{\mathbb{E}[pZ|q]}{[pZ|q]}} = \frac{\mu\mathbf{f}_{\min}}{|Q|^2|\Gamma|C}. \end{aligned}$$

The bound from Lemma 10 on the spectral radius $\rho(\mathbf{f}'(\mu\mathbf{f}))$ is tight: Consider e.g. the pPDA Δ_a with a single state and stack symbol and the transitions $pZ \xrightarrow{a} \varepsilon$, $pZ \xrightarrow{1-a} pZZ$. The fundamental PPS of Δ_a is $\langle Z \rangle = (1-p)\langle Z \rangle^2 + p$. For $a > \frac{1}{2}$, Δ_a is PAST with expected runtime $C = \frac{1}{2a-1}$, and the spectral radius is indeed $2(1-a) = 1 - C^{-1}$. ■

The next lemma allows bounding the side lengths of a (hyper)cube fully contained in the inductive region $\{\mathbf{u} \geq \mathbf{0} \mid \mathbf{f}(\mathbf{u}) \leq \mathbf{u}\}$ (see Appendix E for the proof).

Lemma 11: Let $\Delta = (Q, \Gamma, \delta)$ be a pPDA with cleaned-up fundamental PPS \mathbf{f} . If Δ is cPAST, then there exists a vector $\mathbf{v} \succ \mathbf{0}$, $\|\mathbf{v}\|_1 = 1$, such that the inequality

$$\mathbf{f}(\mu\mathbf{f} + d(\mathbf{v} + \mathbf{a})) \leq \mu\mathbf{f} + d(\mathbf{v} + \mathbf{a}) \quad (43)$$

holds for all

$$0 \leq d \leq \frac{\mu\mathbf{f}_{\min}}{C} \frac{1}{4|Q|^2|\Gamma|(|Q|^2|\Gamma| + 1)^2} \quad (44)$$

and $\mathbf{a} \geq \mathbf{0}$ with $\|\mathbf{a}\|_\infty \leq e$, where

$$e = \frac{\mu\mathbf{f}_{\min}}{C^2} \frac{1}{4|Q|^4|\Gamma|^2} \quad (45)$$

where $C = \max_{pZq} \frac{\mathbb{E}[pZ|q]}{[pZ|q]} < \infty$ is the maximal conditional expected runtime in Δ , and $\mu\mathbf{f}_{\min} > 0$ is the minimal entry of $\mu\mathbf{f}$.

Theorem 9 (Complexity of inductive upper bounds): Let $\Delta = (Q, \Gamma, \delta)$ be a pPDA with fundamental PPS \mathbf{f} . Suppose that Δ is cPAST with maximal conditional expected runtime C and minimal non-zero return probability $\mu\mathbf{f}_{\min}$. Then for every $\varepsilon > 0$ there exists $\mathbf{u} \in \mathbb{Q}_{\geq 0}^{Q \times \Gamma \times Q}$, with $\mathbf{f}(\mathbf{u}) \leq \mathbf{u}$ and $\|\mu\mathbf{f} - \mathbf{u}\|_\infty \leq \varepsilon$, such that the rational numbers in \mathbf{u} can be encoded as pairs of integers with bit-complexity at most:

$$\mathcal{O}\left(\log C + \log \frac{1}{\mu\mathbf{f}_{\min}} + \log |Q| + \log |\Gamma| + \log \frac{1}{\varepsilon}\right) \quad (46)$$

Proof: Note that a (hyper)cube in \mathbb{R}^n located within distance $\mathcal{O}(1)$ to the origin and side length $a > 0$ contains a rational vector whose entries can be encoded as pairs of integers with at most $\mathcal{O}(\log a)$ bits. (46) follows from Lemma 11 by multiplying the bounds (44) and (45), taking logarithms, and omitting constant factors. ε can be taken into account by scaling d in (43) appropriately. ■

In the remainder of this section we discuss some bounds on the quantities occurring in (46). It is already known that $1/\mu\mathbf{f}_{\min}$ may be doubly-exponential in the size of Δ (this means that $\log(1/\mu\mathbf{f}_{\min})$ is exponential) [3]. This can be seen in the following example.

Example 1: Observe that the stochastic context-free grammar (SCFG)

$$Z_1 \rightarrow X^{2^n} \quad (47)$$

$$X \xrightarrow{1/2} a \quad X \xrightarrow{1/2} b \quad (48)$$

produces the terminal string a^{2^n} with probability 2^{-2^n} . It is straightforward to simulate sampling this SCFG using a pPDA (the rule (47) can be realized with n stack symbols using a repeated squaring trick [3]). The example can be extended to see that the expected runtimes, 1st termination moments, and conditional expected runtimes can all be doubly-exponential in magnitude in the size of Δ as well; this works by sampling the grammar repeatedly until the outcome a^{2^n} occurs for the first time, which takes 2^{2^n} trials on average. See Figure 3 for a concrete pPDA implementing this behavior.

Given the close relationship between $\mu\mathbf{f}$ and the 1st termination moments (and hence conditional expected runtimes)

established in Lemma 8, it is natural to ask whether very large runtimes only occur in pPDA with very small non-zero return probabilities (i.e., where $1/\mu_{\min}^f$ is very large; this is indeed the case in Example 1). Perhaps surprisingly, the answer is *no*. To see this, we exhibit an example family where the large runtime is “caused” by a different phenomenon than in Figure 3.

Theorem 10: There exists a family $\{\Delta_n \mid n \geq 1\}$ of pPDA with distinguished initial configuration pY such that $\text{ert}[pY] = \mathbb{E}[pY|p] \geq 2^{2^n}$ and the minimal non-zero return probability is $\mu_{\min}^f \geq c$, for a constant $c > 0$. The encoding size of Δ_n is linear in n .

Proof: Δ_n is depicted in Figure 4. The overall idea of Δ_n is to simulate the pBPA

$$Y \xrightarrow{a_n} \varepsilon \quad Y \xrightarrow{1-a_n} YY \quad (49)$$

with expected runtime $(2a_n - 1)^{-1}$ for probabilities a_n that approach $1/2$ very fast from above as n increases.

To simulate probabilities $a_n > 1 - a_n$ that are very close to each other, i.e., both are *almost* $1/2$, the idea is as follows: Consider a symmetric two-state DTMC where the probability to stay in the current state is $3/4$ and the probability to switch states is $1/4$. One can show that for all $k \geq 0$,

$$\begin{pmatrix} 3/4 & 1/4 \\ 1/4 & 3/4 \end{pmatrix}^k = \frac{1}{2} \begin{pmatrix} 1 + 2^{-k} & 1 - 2^{-k} \\ 1 - 2^{-k} & 1 + 2^{-k} \end{pmatrix}. \quad (50)$$

This means that if the DTMC is started with initial distribution $(1 \ 0)$, then after k steps the distribution is $(\frac{1}{2} + \frac{1}{2^{k+1}} \ \frac{1}{2} - \frac{1}{2^{k+1}})$, i.e., one bit of precision is gained per step.

The pPDA Δ_n simulates the DTMC described above for $k = 2^n$ steps using a repeated squaring technique as in Example 1 (states s and r). From (50) it follows that $a_n = [sZ_1|s] = \frac{1}{2} + \frac{1}{2^{2^n+1}}$ and $1 - a_n = [sZ_1|r] = \frac{1}{2} - \frac{1}{2^{2^n+1}}$. Thus the expected runtime of Δ_n from initial configuration pY is at least $(2a_n - 1)^{-1} = 2^{2^n}$. ■

There are also some upper bounds known on $1/\mu_{\min}^f$. For instance, [21, Lem. 5.13] proves a doubly-exponential upper bound for the case where \mathbb{f} is a strongly connected quadratic PPS. For the (conditional) expected runtimes C and the 1st termination moments we are not aware of any upper bound at all in the case of general pPDA (however, recall from Corollary 2, that in certain sub-classes such as pBPA, the expected runtimes *can* be bounded).

IX. CONCLUSIONS AND OPEN QUESTIONS

We have shown that the (large) class of pPDA that are cPAST admits easy-to-check certificates for two-sided bounds on the return probabilities $[pZ|q]$. On the equation system level, we have linked cPAST to the non-singularity of the system’s Jacobi matrix $I - \mathbb{f}'(\mu_{\min}^f)$. Moreover, we have presented a novel characterization of PAST and showed that it can always be certified (provided the pPDA at hand is in fact PAST).

Based on our results we conjecture the following: *Almost all pPDA are cPAST*, and in particular, *almost all pPDA*

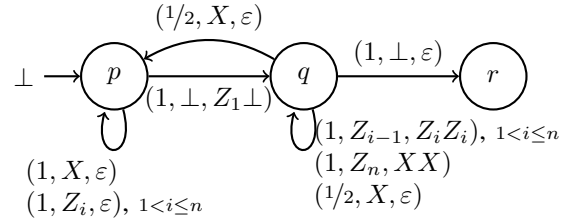


Fig. 3. A family of PAST pPDA with doubly exponential expected runtime $\text{ert}[p\perp] \in \Theta(2^{2^n})$. The minimal non-zero return probability is $\mu_{\min}^f = [qZ_1|q] = 2^{-2^n}$. Transitions not reachable from $p\perp$ are omitted.

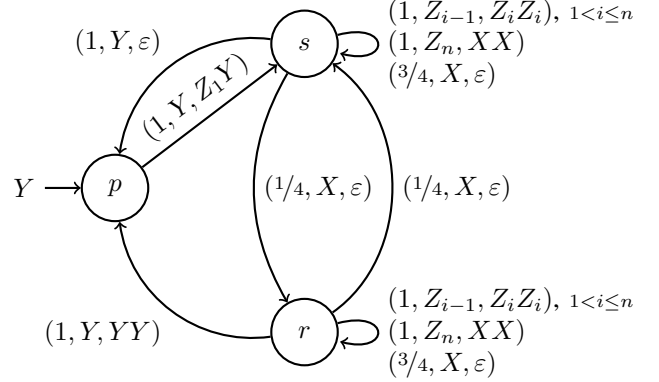


Fig. 4. A family of PAST pPDA with doubly exponential expected runtime $\text{ert}[pY] \in \Theta(2^{2^n})$ and bounded minimal return probabilities $\mu_{\min}^f \geq 1/4$. Transitions not reachable from pY are omitted.

that are AST are also PAST. This is because for a random pPDA Δ (with real-valued weights sampled from continuous intervals), the matrix $\mathbb{f}'(\mu_{\min}^f)$ is essentially a random matrix, and random matrices are almost-surely non-singular. We leave an appropriate formalization of this result for future work. [22] provides some empirical evidence that singular $I - \mathbb{f}'(\mu_{\min}^f)$ occur indeed very rarely in practice. Also note that *all* finite DTMC are cPAST, and it would be interesting to see how this evolves as one ascends to even higher levels in the generative Chomsky hierarchy [39].

Further open problems include finding a general upper bound on the expected runtimes, and on the exact complexity of computing certificates.

ACKNOWLEDGMENT

We thank the anonymous reviewers for their detailed feedback which led, among many others improvements, to a significant simplification of Theorem 5.

REFERENCES

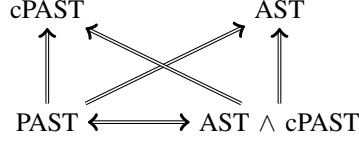
- [1] J. Esparza, A. Kucera, and R. Mayr, “Model checking probabilistic pushdown automata,” in *LICS*. IEEE Computer Society, 2004, pp. 12–21.
- [2] F. Olmedo, B. L. Kaminski, J. Katoen, and C. Matheja, “Reasoning about Recursive Probabilistic Programs,” in *LICS*. ACM, 2016, pp. 672–681.
- [3] K. Etessami and M. Yannakakis, “Recursive Markov Chains, Stochastic Grammars, and Monotone Systems of Nonlinear Equations,” in *STACS*, ser. Lecture Notes in Computer Science, vol. 3404. Springer, 2005, pp. 340–352.

- [4] —, “Recursive Markov chains, stochastic grammars, and monotone systems of nonlinear equations,” *J. ACM*, vol. 56, no. 1, pp. 1:1–1:66, 2009.
- [5] T. Brázdil, S. Kiefer, A. Kucera, and I. H. Vareková, “Runtime analysis of probabilistic programs with unbounded recursion,” in *ICALP (2)*, ser. Lecture Notes in Computer Science, vol. 6756. Springer, 2011, pp. 319–331.
- [6] —, “Runtime analysis of probabilistic programs with unbounded recursion,” *J. Comput. Syst. Sci.*, vol. 81, no. 1, pp. 288–310, 2015.
- [7] J. Esparza, A. Kucera, and R. Mayr, “Quantitative Analysis of Probabilistic Pushdown Automata: Expectations and Variances,” in *LICS*. IEEE Computer Society, 2005, pp. 117–126.
- [8] T. Brázdil, J. Esparza, and S. Kiefer, “On the memory consumption of probabilistic pushdown automata,” in *FSTTCS*, ser. LIPIcs, vol. 4. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2009, pp. 49–60.
- [9] T. Brázdil, A. Kucera, and O. Strazovský, “On the decidability of temporal properties of probabilistic pushdown automata,” in *STACS*, ser. Lecture Notes in Computer Science, vol. 3404. Springer, 2005, pp. 145–157.
- [10] M. Yannakakis and K. Etessami, “Checking LTL properties of recursive Markov chains,” in *QEST*. IEEE Computer Society, 2005, pp. 155–165.
- [11] A. Kucera, J. Esparza, and R. Mayr, “Model checking probabilistic pushdown automata,” *Log. Methods Comput. Sci.*, vol. 2, no. 1, 2006.
- [12] T. Winkler, C. Gehnen, and J. Katoen, “Model Checking Temporal Properties of Recursive Probabilistic Programs,” in *FoSSaCS*, ser. Lecture Notes in Computer Science, vol. 13242. Springer, 2022, pp. 449–469.
- [13] T. Brázdil, V. Brozek, V. Forejt, and A. Kucera, “Branching-time model-checking of probabilistic pushdown automata,” *J. Comput. Syst. Sci.*, vol. 80, no. 1, pp. 139–156, 2014.
- [14] M. Huang, H. Fu, and J. Katoen, “Deciding probabilistic simulation between probabilistic pushdown automata and finite-state systems,” *Inf. Comput.*, vol. 268, 2019.
- [15] T. Brázdil, A. Kucera, and O. Strazovský, “Deciding probabilistic bisimilarity over infinite-state probabilistic systems,” *Acta Informatica*, vol. 45, no. 2, pp. 131–154, 2008.
- [16] T. Brázdil, J. Esparza, S. Kiefer, and A. Kucera, “Analyzing probabilistic pushdown automata,” *Formal Methods Syst. Des.*, vol. 43, no. 2, pp. 124–163, 2013.
- [17] W. Kuich, “Semirings and formal power series: Their relevance to formal languages and automata,” in *Handbook of Formal Languages (1)*. Springer, 1997, pp. 609–677.
- [18] A. Stewart, K. Etessami, and M. Yannakakis, “Upper Bounds for Newton’s Method on Monotone Polynomial Systems, and P-Time Model Checking of Probabilistic One-Counter Automata,” *J. ACM*, vol. 62, no. 4, pp. 30:1–30:33, 2015.
- [19] S. Kiefer, M. Luttenberger, and J. Esparza, “On the convergence of Newton’s method for monotone systems of polynomial equations,” in *STOC*. ACM, 2007, pp. 217–226.
- [20] J. Esparza, S. Kiefer, and M. Luttenberger, “Convergence Thresholds of Newton’s Method for Monotone Polynomial Equations,” in *STACS*, ser. LIPIcs, vol. 1. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, Germany, 2008, pp. 289–300.
- [21] —, “Computing the Least Fixed Point of Positive Polynomial Systems,” *SIAM J. Comput.*, vol. 39, no. 6, pp. 2282–2335, 2010.
- [22] T. Winkler and J. Katoen, “Certificates for probabilistic pushdown automata via optimistic value iteration,” in *TACAS (2)*, ser. Lecture Notes in Computer Science, vol. 13994. Springer, 2023, pp. 391–409.
- [23] R. M. McConnell, K. Mehlhorn, S. Näher, and P. Schweitzer, “Certifying algorithms,” *Comput. Sci. Rev.*, vol. 5, no. 2, pp. 119–161, 2011.
- [24] J. van de Meent, B. Paige, H. Yang, and F. Wood, “An Introduction to Probabilistic Programming,” *CoRR*, vol. abs/1809.10756, 2018.
- [25] K. Etessami, D. Wojtczak, and M. Yannakakis, “Recursive Stochastic Games with Positive Rewards,” in *ICALP (1)*, ser. Lecture Notes in Computer Science, vol. 5125. Springer, 2008, pp. 711–723.
- [26] D. Wojtczak, “Recursive probabilistic models : efficient analysis and implementation,” Ph.D. dissertation, University of Edinburgh, UK, 2009.
- [27] A. Hartmanns and B. L. Kaminski, “Optimistic value iteration,” in *CAV (2)*, ser. Lecture Notes in Computer Science, vol. 12225. Springer, 2020, pp. 488–511.
- [28] F. Funke, S. Jantsch, and C. Baier, “Farkas Certificates and Minimal Witnesses for Probabilistic Reachability Constraints,” in *TACAS (1)*, ser. Lecture Notes in Computer Science, vol. 12078. Springer, 2020, pp. 324–345.
- [29] S. Wimmer and J. von Mutius, “Verified Certification of Reachability Checking for Timed Automata,” in *TACAS (1)*, ser. Lecture Notes in Computer Science, vol. 12078. Springer, 2020, pp. 425–443.
- [30] K. Etessami and M. Yannakakis, “Recursive Markov decision processes and recursive stochastic games,” in *ICALP*, ser. Lecture Notes in Computer Science, vol. 3580. Springer, 2005, pp. 891–903.
- [31] N. Kobayashi, U. Dal Lago, and C. Grellois, “On the Termination Problem for Probabilistic Higher-Order Recursive Programs,” *Log. Methods Comput. Sci.*, vol. 16, no. 4, 2020.
- [32] D. Chistikov, R. Majumdar, and P. Schepper, “Subcubic certificates for CFL reachability,” *Proc. ACM Program. Lang.*, vol. 6, no. POPL, pp. 1–29, 2022.
- [33] A. McIver and C. Morgan, *Abstraction, Refinement and Proof for Probabilistic Systems*, ser. Monographs in Computer Science. Springer, 2005.
- [34] V. G. Kulkarni, *Modeling and analysis of stochastic systems*. Chapman & Hall, Ltd., 1995.
- [35] U. Dal Lago, “On probabilistic λ -calculus,” *Foundations of Probabilistic Programming*, pp. 121–144, 2020.
- [36] T. Winkler and J.-P. Katoen, “On certificates, expected runtimes, and termination in probabilistic pushdown automata,” 2023. [Online]. Available: <https://arxiv.org/abs/2304.09997>
- [37] U. G. Rothblum, “Nonnegative matrices and stochastic matrices,” in *Handbook of Linear Algebra*, L. Hogben, Ed. CRC press, 2006.
- [38] H. S. Wilf, *generatingfunctionology*. CRC press, 2005.
- [39] T. F. Icard, “Calibrating generative models: The probabilistic chomsky-schützenberger hierarchy,” *Journal of Mathematical Psychology*, vol. 95, 2020.

APPENDIX

A. Proof of Lemma 2

Lemma 2: The relationship between AST, PAST, and cPAST in a DTMC with a distinguished target set is as follows:



The omitted implications do not hold in general.

Proof: Suppose that the DTMC is $\mathfrak{M} = (S, M)$ and the target is $F \subseteq S$. The only non-trivial implications are the following ones:

- $\text{PAST} \implies \text{AST}$: Shown in main text.
- $\text{PAST} \implies \text{cPAST}$. Since $\text{PAST} \implies \text{AST}$ we know that the vector of termination probabilities is $\mu^0 = \mathbf{1}$. Thus by (9) and (11), $\text{ert}[s|F] = \mathbb{E}[s|F]$ for all $s \in S$.
- $\text{AST} \wedge \text{cPAST} \implies \text{PAST}$: Again, since we have AST we know that the vector of termination probabilities is $\mu^0 = \mathbf{1}$, and so by (9) and (11), $\text{ert}[s|F] = \mathbb{E}[s|F]$ for all $s \in S$. But since we have cPAST as well, the termination moments $\mathbb{E}[s|F]$ are all finite. Therefore the expected runtimes $\text{ert}[s|F]$ are all finite as well.

For the non-implications, consider the following family of DTMC $\mathfrak{M}_p = (S, M)$ with $p \in (0, 1)$: $S = \mathbb{N}$, $F = \{0\}$, and for all $i \geq 1$, $M_{i,i+1} = (1-p)$ and $M_{i,i-1} = p$.

- $\mathfrak{M}_{1/2}$ is AST but neither PAST nor cPAST.
- $\mathfrak{M}_{1/3}$ is cPAST but not AST (and thus also not PAST).

■

B. Proof of variant of Lemma 4

We show that for a PPS \mathbf{f} and $\mathbf{u}, \mathbf{v} \geq \mathbf{0}$ such that $\mathbf{u} - \mathbf{v} \geq \mathbf{0}$,

$$\mathbf{f}(\mathbf{u}) - \mathbf{f}'(\mathbf{u})\mathbf{v} \leq \mathbf{f}(\mathbf{u} - \mathbf{v}) \leq \mathbf{f}(\mathbf{u}) - \mathbf{f}'(\mathbf{u} - \mathbf{v})\mathbf{v} . \quad (51)$$

The following proof is a straightforward adaption of the proof of [21, Lemma 2.3]. It suffices to show the claim for a multivariate polynomial $\mathbf{f} = f$ with non-negative coefficients. Let $g(t) = f(\mathbf{u} - t\mathbf{v})$. We have

$$f(\mathbf{u} - \mathbf{v}) = g(1) = g(0) + \int_0^1 g'(s)ds = f(\mathbf{u}) + \int_0^1 -f'(\mathbf{u} - s\mathbf{v})\mathbf{v}ds \quad (52)$$

For all $s \in [0, 1]$, as f' has non-negative coefficients and $\mathbf{u} - \mathbf{v} \geq \mathbf{0}$, it holds that

$$f'(\mathbf{u}) \geq f'(\mathbf{u} - s\mathbf{v}) \geq f'(\mathbf{u} - \mathbf{v}) \quad (53)$$

and thus

$$-f'(\mathbf{u}) \leq -f'(\mathbf{u} - s\mathbf{v}) \leq -f'(\mathbf{u} - \mathbf{v}) . \quad (54)$$

The claim follows.

C. Proof of Lemma 7

Lemma 7 (System of termination moments): Let $\Delta = (Q, \Gamma, \delta)$ be a pPDA and let $(\mathbb{E}\langle pZ|q \rangle)_{pZq \in Q \times \Gamma \times Q}$ be $\overline{\mathbb{R}}_{\geq 0}$ -valued variables. Then the 1st termination moments $\mathbb{E}\langle pZ|q \rangle$ constitute the least solution of the following *linear* PPS:

$$\begin{aligned} \mathbb{E}\langle pZ|q \rangle &= [pZ|q] + \sum_{pZ \xrightarrow{\Delta} rY} a \cdot \mathbb{E}\langle rY|q \rangle \\ &+ \sum_{\substack{pZ \xrightarrow{\Delta} rYX \\ t \in Q}} a \cdot (\mathbb{E}\langle rY|t \rangle \cdot [tX|q] + [rY|t] \cdot \mathbb{E}\langle tX|q \rangle) \end{aligned} \quad (21)$$

where $[pZ|q]$, etc., are the return probabilities of Δ .

Proof: We give a straightforward proof using an equation system from [7, Theorem 3.1] (we use the formulation given in [16, Section 5.1]). In these works, it was shown that the conditional expected runtimes $\frac{\mathbb{E}[pZ|q]}{[pZ|q]}$, for $pZq \in Q \times \Gamma \times Q$ such that $[pZ|q] > 0$, satisfy the equations

$$\frac{\mathbb{E}[pZ|q]}{[pZ|q]} = 1 + \frac{1}{[pZ|q]} \left(\sum_{pZ \xrightarrow{a} rY} a \cdot [rY|q] \cdot \frac{\mathbb{E}[rY|q]}{[rY|q]} + \sum_{pZ \xrightarrow{a} rYX} a \cdot \sum_{t \in Q} [rY|t] \cdot [tX|q] \cdot \left(\frac{\mathbb{E}[rY|t]}{[rY|t]} + \frac{\mathbb{E}[tX|q]}{[tX|q]} \right) \right). \quad (55)$$

Moreover, the conditional expected runtimes constitute exactly the least solution in $\overline{\mathbb{R}}_{\geq 0}$ of the above equations when interpreting the fractions as variable symbols. Multiplying (55) by $[pZ|q]$ (and reinterpreting the symbols $\mathbb{E}[\cdot|\cdot]$ as variables) we obtain (21). Note however, that in (21) we also have equations for *all* $pZq \in Q \times \Gamma \times Q$, not just for the ones with $[pZ|q] > 0$. ■

D. Derivation of (32) from [7] in the AST case

Let $\Delta = (Q, \Gamma, \delta)$ be a pPDA. We have already explained in Appendix C how to derive the equation system for the 1st termination moments $\mathbb{E}[pZ|q]$, $pZq \in Q \times \Gamma \times Q$, from the results of [7]:

$$\mathbb{E}[pZ|q] = [pZ|q] + \sum_{pZ \xrightarrow{a} rY} a \cdot \mathbb{E}[rY|q] + \sum_{pZ \xrightarrow{a} rYX} a \cdot (\mathbb{E}[rY|t] \cdot [tX|q] + [rY|t] \cdot \mathbb{E}[tX|q]) \quad (56)$$

Now assume that Δ is AST. Then for all $pZ \in Q \times \Gamma$, we have $\text{ert}[pZ] = \sum_{q \in Q} \mathbb{E}[pZ|q]$, and $\sum_{q \in Q} [pZ|q] = 1$. Summing over all $q \in Q$ in (56) yields

$$\text{ert}[pZ] = 1 + \sum_{pZ \xrightarrow{a} rY} a \cdot \text{ert}[rY] + \sum_{pZ \xrightarrow{a} rYX} a \cdot (\text{ert}[rY] + [rY|t] \cdot \text{ert}[tX]) \quad (57)$$

which is the same as (32). However, we stress that (32) also holds in the non-AST case, which does not easily follow from [7].

E. Proof of Lemma 11

Lemma 11: Let $\Delta = (Q, \Gamma, \delta)$ be a pPDA with cleaned-up fundamental PPS \mathbb{f} . If Δ is cPAST, then there exists a vector $\mathbf{v} \succ \mathbf{0}$, $\|\mathbf{v}\|_1 = 1$, such that the inequality

$$\mathbb{f}(\mu\mathbb{f} + d(\mathbf{v} + \mathbf{a})) \leq \mu\mathbb{f} + d(\mathbf{v} + \mathbf{a}) \quad (43)$$

holds for all

$$0 \leq d \leq \frac{\mu\mathbb{f}_{\min}}{C} \frac{1}{4|Q|^2|\Gamma|(|Q|^2|\Gamma| + 1)^2} \quad (44)$$

and $\mathbf{a} \geq \mathbf{0}$ with $\|\mathbf{a}\|_\infty \leq e$, where

$$e = \frac{\mu\mathbb{f}_{\min}}{C^2} \frac{1}{4|Q|^4|\Gamma|^2} \quad (45)$$

where $C = \max_{pZq} \frac{\mathbb{E}[pZ|q]}{[pZ|q]} < \infty$ is the maximal conditional expected runtime in Δ , and $\mu\mathbb{f}_{\min} > 0$ is the minimal entry of $\mu\mathbb{f}$.

Proof: Let $\mathbf{v} \succ \mathbf{0}$ be the vector with $\|\mathbf{v}\|_1 = 1$ from Lemma 10 that satisfies

$$\mathbb{f}'(\mu\mathbb{f})\mathbf{v} \leq (1 - C^{-1})\mathbf{v}. \quad (58)$$

For the sake of brevity we write $c := 1 - C^{-1}$ in this proof. Suppose further that d and \mathbf{a} satisfy the preconditions (44) and (45). Taylor expansion (Lemma 4) yields

$$\mathbb{f}(\mu\mathbb{f} + d(\mathbf{v} + \mathbf{a})) \leq \mu\mathbb{f} + \mathbb{f}'(\mu\mathbb{f} + d(\mathbf{v} + \mathbf{a}))d(\mathbf{v} + \mathbf{a}). \quad (59)$$

For the RHS of (59) to be at most $\mu\mathbb{f} + d(\mathbf{v} + \mathbf{a})$ it suffices that

$$\mathbb{f}'(\mu\mathbb{f} + d(\mathbf{v} + \mathbf{a}))(\mathbf{v} + \mathbf{a}) \leq \mathbf{v} + \mathbf{a}. \quad (60)$$

Suppose that \mathbb{f} is indexed by $J \subseteq Q \times \Gamma \times Q$, $|J| =: D$. For $i \in J$, let $\mathbb{f}'_i(\mathbf{x})$ be the i th column of the Jacobi matrix $\mathbb{f}'(\mathbf{x})$. We now view the $\mathbf{x} = \mathbb{f}'_i(\mathbf{x})$ as a collection of individual PPS. Applying Taylor's theorem to \mathbb{f}'_i , we obtain

$$\mathbb{f}'_i(\mu\mathbb{f} + d(\mathbf{v} + \mathbf{a})) \leq \mathbb{f}'_i(\mu\mathbb{f}) + \mathbb{f}''_i(\mu\mathbb{f} + d(\mathbf{v} + \mathbf{a}))d(\mathbf{v} + \mathbf{a}) \quad (61)$$

$$\leq \mathbb{f}'_i(\mu\mathbb{f}) + 2d\mathbf{v} + 2dDe\mathbf{1} \quad (62)$$

In (61), \mathbb{f}_i'' is the Jacobian of the PPS \mathbb{f}_i' . For (62), we have used that \mathbb{f}_i'' is actually a non-negative *constant* matrix whose coefficients are bounded by 2; this is because all its components are second (partial) derivatives of polynomials with total degree at most 2 and coefficients in $[0, 1]$. We can now express the LHS of (60) as follows:

$$\mathbb{f}'(\mu\mathbb{f} + d(\mathbf{v} + \mathbf{a}))(\mathbf{v} + \mathbf{a}) \quad (63)$$

$$= \sum_{i=1}^D \mathbb{f}_i'(\mu\mathbb{f} + d(\mathbf{v} + \mathbf{a}))(\mathbf{v}_i + \mathbf{a}_i) \quad (64)$$

$$\leq \sum_{i=1}^D (\mathbb{f}_i'(\mu\mathbb{f}) + 2d\mathbf{v} + 2dDe\mathbf{1})(\mathbf{v}_i + \mathbf{a}_i) \quad (\text{by (62)})$$

$$\leq (\mathbb{f}'(\mu\mathbb{f}) + 2d + 2dDe)\mathbf{v} + \mathbb{f}'(\mu\mathbb{f})\mathbf{a} + 2dDe(1 + De)\mathbf{1} \quad (\text{rearranging and using } \mathbf{a}_i \leq e, \|\mathbf{v}\|_1 = 1)$$

$$\leq (\mathbb{f}'(\mu\mathbb{f}) + 2d + 2dDe)\mathbf{v} + 2De\mathbf{1} + 2dDe(1 + De)\mathbf{1} \quad (\text{using that } \mathbb{f}'(\mu\mathbb{f}) \leq 2 \text{ coefficient-wise})$$

$$= (\mathbb{f}'(\mu\mathbb{f}) + 2d + 2dDe)\mathbf{v} + 2De(1 + d(1 + De))\mathbf{1} \quad (\text{rearranging})$$

$$\leq (c + 2d + 2dDe)\mathbf{v} + 2De(1 + d(1 + De))\mathbf{1} \quad (65)$$

In (65) we have used (58). Recall from (60) that we need to ensure that expression (63) is at most $\mathbf{v} + \mathbf{a}$. We show that (63) is in fact at most \mathbf{v} . For the latter, it is sufficient that expression (65) is at most \mathbf{v} , which is the case iff

$$2De(1 + d(1 + De)) \leq (1 - c - 2d - 2dDe)\mathbf{v}_{\min} \quad (66)$$

where $\mathbf{v}_{\min} > 0$ is the minimal component of the vector \mathbf{v} . It can be checked² that the inequality (66) is satisfied if

$$e < \frac{\mathbf{v}_{\min}(1 - c)}{2D} \text{ and } d \leq \frac{\mathbf{v}_{\min}(1 - c) - 2De}{2(De + 1)(De + \mathbf{v}_{\min})} \quad (67)$$

To complete the proof reconsider the bounds (44) and (45). Recall that $c = 1 - C^{-1}$, $\mathbf{v}_{\min} \geq \frac{\mu\mathbb{f}_{\min}}{C|Q|^2|\Gamma|}$ and observe that

$$\frac{\mu\mathbb{f}_{\min}}{C^2} \frac{1}{4|Q|^4|\Gamma|^2} \leq \frac{\mathbf{v}_{\min}}{C} \frac{1}{4|Q|^2|\Gamma|} < \frac{\mathbf{v}_{\min}(1 - c)}{2D} \quad (68)$$

Furthermore,

$$\frac{\mu\mathbb{f}_{\min}}{C} \frac{1}{4|Q|^2|\Gamma|(|Q|^2|\Gamma| + 1)^2} \quad (69)$$

$$\leq \frac{\mu\mathbb{f}_{\min}}{C} \frac{\mathbf{v}_{\min}(1 - c) - 2De}{4|Q|^2|\Gamma|(|Q|^2|\Gamma| + 1)^2} \quad (70)$$

$$\leq \mathbf{v}_{\min} \frac{\mathbf{v}_{\min}(1 - c) - 2De}{4(|Q|^2|\Gamma| + 1)^2} \quad (71)$$

$$\leq \frac{\mathbf{v}_{\min}(1 - c) - 2De}{2(De + 1)(De + \mathbf{v}_{\min})} \quad (\text{as } \mathbf{v}_{\min} \leq 1, |Q|^2|\Gamma| \geq D, \text{ and } e \leq 1)$$

so the inequalities (67), and hence (66), (60), and finally (43) are also satisfied. ■

²https://www.wolframalpha.com/input?i=%7B2*D*x%281%2B%281%2BD*x%29%29%3C%3D%281-C-2y-2*D*y*x%29*V%2C0%3CC%3C1%2Cx%3E0%2Cy%3E0%2C1%3E%3DV%3E0%2CD%3E0%7D