

Certificates for Probabilistic Pushdown Automata via Optimistic Value Iteration

Tobias Winkler and Joost-Pieter Katoen

RWTH Aachen University, Germany

Abstract. Probabilistic pushdown automata (pPDA) are a standard model for discrete probabilistic programs with procedures and recursion. In pPDA, many quantitative properties are characterized as least fixpoints of polynomial equation systems. In this paper, we study the problem of *certifying* that these quantities lie within certain bounds. To this end, we first characterize the polynomial systems that admit easy-to-check certificates for validating bounds on their least fixpoint. Second, we present a sound and complete Optimistic Value Iteration algorithm for computing such certificates. Third, we show how certificates for polynomial systems can be transferred to certificates for various quantitative pPDA properties. Experiments demonstrate that our algorithm computes succinct certificates for several intricate example programs as well as stochastic context-free grammars with $> 10^4$ production rules.

Keywords: Probabilistic Pushdown Automata · Probabilistic Model Checking · Certified Algorithms · Probabilistic Recursive Programs.

1 Introduction

Complex software is likely to contain bugs. This applies in particular to model checking tools. This is a serious problem, as the possibility of such bugs compromises the trust one can put in the verification results, rendering the process of formal modeling and analysis less useful. Ideally, the implementation of a model checker should be formally verified itself [15]. However, due to the great complexity of these tools, this is often out of reach in practice. *Certifying algorithms* [31] mitigate this problem by providing an *easy-to-check certificate* along with their regular output. This means that there exists a *verifier* that, given the input problem, the output, and the certificate, constructs a formal proof that the output is indeed correct. The idea is that the verifier is much simpler than the algorithm, and thus likely to be bug-free or even amenable to formal verification.

This paper extends the recent line of research on probabilistic certification [19,23,24,40] to *probabilistic pushdown automata* [13,30] (pPDA). pPDA and related models have applications in, amongst others, pattern recognition [38], computational biology [28], and speech recognition [25]. They are moreover a natural operational model for programs with procedures, recursion, and (discrete) probabilistic constructs such as the ability to flip coins. With the advent

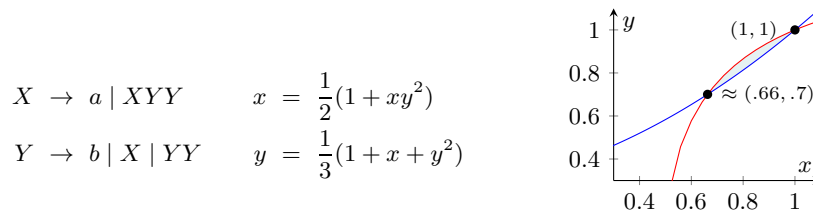


Fig. 1: Left: A stochastic context-free grammar (SCFG) and the associated positive polynomial system (PPS) which encodes the termination probabilities of each non-terminal, assuming production rules are taken uniformly at random. Right: The curves defined by the two equations. The least fixpoint (lfp) is $\approx (0.66, 0.70)$. The thin colored area to the top right of the lfp is the set of inductive, i.e., self-certifying upper bounds on the lfp.

of *probabilistic programming* [32] as a paradigm for model-based machine learning [6], such programs have received lots of attention recently. Moreover, several efficient algorithms such as Hoare’s quicksort with randomized pivot selection (e.g. [26]) are readily encoded as probabilistic recursive programs.

A pPDA can be seen as a purely probabilistic variant of a standard pushdown automaton: Instead of reading an input word, it takes its transitions randomly based on fixed probability distributions over successor states. Quantities of interest in pPDA include reachability probabilities [13], expected runtimes [8], variances [14], satisfaction probabilities of temporal logic formulas [44,41], and others (see [7] for an overview). pPDA are equivalent to *recursive Markov chains* [17].

One of the difficulties of pPDA is that they induce *infinite* Markov chains. Despite this fact, many interesting quantitative properties are decidable, albeit with rather high complexity. Therefore, in the past two decades there have been significant research efforts on efficient *approximative* algorithms for pPDA, especially a decomposed variant of *Newton iteration* [16,27,11,17,12,10,39] which provides guaranteed lower, and occasionally upper [10,12] bounds on key quantities. However, even though implementations might be complex [43], these algorithms do not produce certificates for their results.

Our technique for certificate generation is an adaption of *Optimistic Value Iteration* [22] (OVI) to the pPDA setting. In a nutshell, OVI computes *some* lower bound \vec{l} on the solution—which can be done using an approximative iterative algorithm—and then *optimistically guesses* an upper bound $\vec{u} = \vec{l} + \vec{\epsilon}$ and verifies that the guess was correct. Originally, OVI was formulated for Markov Decision Processes (MDP) where it is used to compute lower and upper bounds on minimal or maximal reachability probabilities and expected rewards. The upper bounds computed by OVI have a special property: They are *self-certifying* (also called *inductive* in this paper). This means that, given the MDP and the upper bounds, one can check that the bounds are correct without the need for an additional certificate; and this check is conceptually and practically easier than finding the bounds in the first place.

The analysis of pPDA, however, is more involved than that of MDP. In MDP, many quantitative properties are characterized as *least fixpoints* (lfp) of piece-wise *linear* equation systems and can be computed in PTIME via, e.g., LP solving. In pPDA, on the other hand, the equation systems for the same properties may contain non-linear *polynomials*, and the best known complexity bounds are usually as high as PSPACE. An example of such a non-linear system is illustrated in Figure 1 which shows the translation of a *stochastic context-free grammar* (SCFG; special case of pPDA with a single state) to a polynomial equation system encoding termination probabilities. An important observation is that the polynomials arising in this context only have positive coefficients. Such systems are called *positive polynomial systems* (PPS) in this paper.

Applications of PPS beyond the analysis of pPDA include the recent *factor graph grammars* [9] as well as obtaining approximate counting formulas for many classes of trees in the framework of *analytic combinatorics* [18].

Contributions. In summary, this paper makes the following contributions:

- We present an optimistic algorithm for computing inductive, *self-certifying* upper bounds of any desired precision $\varepsilon > 0$ on the lp of a positive polynomial system. Compared to OVI from [22], the key innovation of our algorithm is to compute a certain *direction* \vec{v} in which to guess, i.e., the guess is $\vec{u} = \vec{l} + \varepsilon\vec{v}$ rather than $\vec{u} = \vec{l} + \vec{\varepsilon}$. This is to ensure that we eventually hit an inductive bound, even if the latter lie in a very “thin strip” as in Figure 1.
- We prove that our algorithm is sound *and complete* in the sense that *if* a (non-trivial) inductive upper bound exists, then such a bound will be found.
- We show how inductive bounds on the lp of PPS can be used to certify various quantities of interest in pPDA and SCFG, such as non-termination or bounds on expected rewards/costs.
- We implement our algorithm in the software tool PRAY and compare the new technique to an out-of-the-box approach based on SMT solving.

Related Work. Certification of pPDA has not been addressed explicitly in the literature, but some existing technical results go in this direction. We mention [17, Prop. 8.7] which yields certificates for non almost-sure termination of SCFG. However, checking such certificates is not straightforward as it requires an SCC decomposition. The tool PREMO [43] implements iterative algorithms for lower bounds, but it supports neither certificates nor upper bounds.

Beyond pPDA, OVI was recently generalized to stochastic games [1]. *Farkas certificates* for MDP [19] are verified by checking a set of linear constraints, which is in spirit similar to our certificates that requires checking a set of polynomial constraints. A deductive approach for verifying probabilistic recursive programs on the syntax level was studied in [35]. The same paper also includes inductive proof rules for verifying upper bounds just like we do. Recently, a higher-order generalization of pPDA called PHORS was introduced in [29], and an algorithm for finding upper bounds inspired by the Finite Elements method was proposed.

Paper Outline. We review the relevant background information on PPS in Section 2. Section 3 presents our theoretical results on inductive upper bounds in PPS as well as the new Optimistic Value Iteration algorithm. In Section 4 we explain how inductive bounds in PPS are used to certify quantitative properties of pPPA. The experimental evaluation is in Section 5. We conclude in Section 6.

2 Preliminaries

Notation for Vectors. All vectors in this paper are *column* vectors and are written in boldface, e.g., $\vec{u} = (u_1, \dots, u_n)^T$. For vectors \vec{u}, \vec{u}' , we write $\vec{u} \leq \vec{u}'$ if \vec{u} is *component-wise* less than or equal to \vec{u}' . Moreover, we write $\vec{u} < \vec{u}'$ if $\vec{u} \leq \vec{u}'$ and $\vec{u} \neq \vec{u}'$, and $\vec{u} \prec \vec{u}'$ if \vec{u} is component-wise *strictly* smaller than \vec{u}' . The zero vector is denoted $\vec{0}$. The *max norm* of a vector \vec{u} is $\|\vec{u}\|_\infty = \max_{1 \leq i \leq n} |u_i|$. We say that \vec{u} is *normalized* if $\|\vec{u}\|_\infty = 1$.

Positive Polynomial Systems (PPS). Let $n \geq 1$ and $\vec{x} = (x_1, \dots, x_n)^T$ be a vector of variables. An n -dimensional PPS is an equation system of the form

$$x_1 = f_1(x_1, \dots, x_n) \quad \dots \quad x_n = f_n(x_1, \dots, x_n)$$

where for all $1 \leq i \leq n$, the function f_i is a *polynomial with non-negative real coefficients*. An example PPS is the system $x = \frac{1}{2}(1+xy^2), y = \frac{1}{3}(1+x+y^2)$ from Figure 1. We also use vector notation for PPS: $\vec{x} = \vec{f}(\vec{x}) = (f_1(\vec{x}), \dots, f_n(\vec{x}))^T$.

We write $\overline{\mathbb{R}}_{\geq 0} = \mathbb{R}_{\geq 0} \cup \{\infty\}$ for the *extended non-negative reals*. By convention, for all $a \in \overline{\mathbb{R}}_{\geq 0}$, $a \leq \infty$, $a + \infty = \infty + a = \infty$, and $a \cdot \infty = \infty \cdot a$ equals 0 if $a = 0$ and ∞ otherwise. For $n \geq 1$, the partial order $(\overline{\mathbb{R}}_{\geq 0}^n, \leq)$ is a *complete lattice*, i.e., all subsets of $\overline{\mathbb{R}}_{\geq 0}^n$ have an infimum and a supremum. In particular, there exists a least element $\vec{0}$ and a greatest element $\vec{\infty} = (\infty, \dots, \infty)^T$. Every PPS induces a *monotone* function $\vec{f}: \overline{\mathbb{R}}_{\geq 0}^n \rightarrow \overline{\mathbb{R}}_{\geq 0}^n$, i.e., $\vec{u} \leq \vec{v} \implies \vec{f}(\vec{u}) \leq \vec{f}(\vec{v})$. By the Knaster-Tarski fixpoint theorem, the set of fixpoints of \vec{f} is also a complete lattice, and thus there exists a *least fixpoint* (lfp) denoted by $\mu \vec{f}$.

In general, the lfp $\mu \vec{f}$ is a vector which may contain ∞ as an entry. For instance, this happens in the PPS $x = x + 1$. A PPS \vec{f} is called *feasible* if $\mu \vec{f} \prec \vec{\infty}$ (or equivalently, $\mu \vec{f} \in \mathbb{R}_{\geq 0}^n$), i.e., the lfp is a vector of *real* numbers. Besides existence of the lfp, the Knaster-Tarski theorem also implies the following:

Lemma 1 (Inductive upper bounds). *For all $\vec{u} \in \overline{\mathbb{R}}_{\geq 0}^n$ it holds that*

$$\vec{f}(\vec{u}) \leq \vec{u} \quad \text{implies} \quad \mu \vec{f} \leq \vec{u}.$$

Such a vector \vec{u} with $\vec{u} \prec \vec{\infty}$ is called inductive upper bound.

Problem statement of this paper

Given a feasible PPS \vec{f} , find an *inductive* upper bound $\vec{u} \geq \mu \vec{f}$.

If \vec{f} is feasible, then $\mu\vec{f}$ is obviously an inductive upper bound. In Section 3 we show under which conditions there exist more useful inductive upper bounds.

A PPS is called *clean* if $\mu\vec{f} \succ \vec{0}$. Every PPS can be cleaned in linear time by identifying and removing the variables that are assigned 0 in the lfp [17,12].

Given a PPS \vec{f} and a point $\vec{u} \in \mathbb{R}_{\geq 0}^n$, we define the *Jacobi matrix* of \vec{f} at \vec{u} as the $n \times n$ -matrix $\partial\vec{f}(\vec{u})$ with coefficients $\partial\vec{f}(\vec{u})_{1 \leq i, j \leq n} = \frac{\partial}{\partial x_j} f_i(\vec{u})$.

Example 1. Consider the example PPS \vec{f}_{ex} with variables $\vec{x} = (x, y)^T$:

$$x = f_1(x, y) = y + 0.1 \quad y = f_2(x, y) = 0.2x^2 + 0.8xy + 0.1.$$

The line and the hyperbola defined by these equations are depicted in Figure 2 on Page 7. The fixpoints of \vec{f}_{ex} are the intersections of these geometric objects; in this case there are two. In particular, \vec{f}_{ex} is feasible and its lfp is

$$\mu\vec{f}_{ex} = ((27 - \sqrt{229})/50, (22 - \sqrt{229})/50)^T \approx (0.237, 0.137)^T.$$

Therefore, \vec{f}_{ex} is clean as $\mu\vec{f}_{ex} \succ \vec{0}$. The Jacobi matrix of \vec{f}_{ex} is

$$\partial\vec{f}_{ex}(x, y) = \begin{pmatrix} \frac{\partial}{\partial x} f_1 & \frac{\partial}{\partial y} f_1 \\ \frac{\partial}{\partial x} f_2 & \frac{\partial}{\partial y} f_2 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 0.4x + 0.8y & 0.8x \end{pmatrix}.$$

Note that the lfp $\mu\vec{f}_{ex}$ contains irrational numbers. However, we can still give exact expressions for these numbers (involving square roots) because the fixpoints of \vec{f}_{ex} are the zeros of a quadratic polynomial. However, there are PPS whose lfp *cannot* be expressed using radicals, i.e., square roots, cubic roots, etc. [16]. This means that in general, there is no easy way to compute least fixpoints exactly. It is thus desirable to provide bounds, which we do in this paper. \triangle

Matrices and Eigenvectors. Let M be a real $n \times n$ -matrix. We say that M is *non-negative* (in symbols: $M \geq 0$) if it has no negative entries. M is called *irreducible* if for all $1 \leq i, j \leq n$ there exists $0 \leq k < n$ such that $(M^k)_{i,j} \neq 0$. It is easy to show that M is irreducible iff the directed graph $G_M = (\{1, \dots, n\}, E)$ with $(i, j) \in E$ iff $M_{i,j} \neq 0$ is strongly connected. A *maximal irreducible submatrix* of M is a square submatrix induced by a strongly connected component of G_M . The *period* of a strongly connected M is the length of the shortest cycle in G_M . It is instructive to note that PPS $\vec{x} = \vec{f}(\vec{x})$ are generalizations of linear equation systems of the form $\vec{x} = M\vec{x} + \vec{c}$, with $M \geq 0$ and $\vec{c} \geq \vec{0}$. Moreover, note that for any PPS \vec{f} it holds that $\partial\vec{f}(\vec{u}) \geq 0$ for all $\vec{u} \succ \vec{0}$.

An *eigenvector* of an $n \times n$ -matrix M with eigenvalue $\lambda \in \mathbb{C}$ is a (complex) vector $\vec{v} \neq \vec{0}$ satisfying $M\vec{v} = \lambda\vec{v}$. There are at most n different eigenvalues. The *spectral radius* $\rho(M) \in \mathbb{R}_{\geq 0}$ is the largest absolute value of the eigenvalues of M . The following is a fundamental theorem about non-negative matrices:

Theorem 1 (Perron-Frobenius). *Let $M \geq 0$ be irreducible.*

- (1) M has a strictly positive eigenvector $\vec{v} \succ \vec{0}$ with eigenvalue $\rho(M)$, the spectral radius of M , and all other eigenvectors $\vec{v}' \succ \vec{0}$ are scalar multiples of \vec{v} .
- (2) The eigenvalues of M with absolute value $\rho(M)$ are exactly the h numbers $\rho(M), \xi\rho(M), \dots, \xi^{h-1}\rho(M)$, where ξ is a primitive h th root of unity.

The *unique* eigenvector $\vec{v} \succ \vec{0}$ with $\|\vec{v}\|_\infty = 1$ of an irreducible non-negative matrix M is called the *Perron-Frobenius* eigenvector of M .

Strongly Connected Components. To each PPS \vec{f} we associate a finite directed graph $G_{\vec{f}} = (\{x_1, \dots, x_n\}, E)$, which, intuitively speaking, captures the dependency structure among the variables. Formally, $(x_i, x_j) \in E$ if the polynomial f_i depends on x_j , i.e., x_j appears in at least one term of f_i with a non-zero coefficient. This is equivalent to saying that the *partial derivative* $\frac{\partial}{\partial x_j} f_i$ is not the zero polynomial. We say that \vec{f} is *strongly connected* if $G_{\vec{f}}$ is strongly connected, i.e., for each pair (x_i, x_j) of variables, there exists a path from x_i to x_j in $G_{\vec{f}}$. For instance, \vec{f}_{ex} from Example 1 is strongly connected because the dependency graph has the edges $E = \{(x, y), (y, x), (y, y)\}$. Strong connectivity of PPS is a generalization of irreducibility of matrices; indeed, a matrix M is irreducible iff the PPS $\vec{x} = M\vec{x}$ is strongly connected. We often use the fact that $\partial\vec{f}(\vec{u})$ for $\vec{u} \succ \vec{0}$ is irreducible iff \vec{f} is strongly connected.

PPS are usually analyzed in a decomposed fashion by considering the subsystems induced by the *strongly connected components* (SCCs) of $G_{\vec{f}}$ in bottom-up order [16]. Here we also follow this approach and therefore focus on strongly connected PPS. The following was proved in [17, Lem. 6.5] and later generalized in [12, Thm. 4.1] (also see remark below [12, Prop. 5.4] and [17, Lem. 8.2]):

Theorem 2 ([17,12]). *If \vec{f} is feasible, strongly connected and clean, then for all $\vec{u} < \mu\vec{f}$, we have $\rho(\partial\vec{f}(\vec{u})) < 1$. As a consequence, $\rho(\partial\vec{f}(\mu\vec{f})) \leq 1$.*

Theorem 2 partitions all PPS \vec{f} which satisfy its precondition into two classes: Either (1) $\rho(\partial\vec{f}(\mu\vec{f})) < 1$, or (2) $\rho(\partial\vec{f}(\mu\vec{f})) = 1$. In the next section we show that \vec{f} admits non-trivial inductive upper bounds iff it is in class (1).

Example 2. Reconsider the PPS \vec{f}_{ex} from Example 1. It can be shown that $\vec{v} = (1, \lambda_1)^T$ where $\lambda_1 \approx 0.557$ is an eigenvector of $\partial\vec{f}_{ex}(\mu\vec{f}_{ex})$ with eigenvalue λ_1 . Thus by the Perron-Frobenius Theorem, $\rho(\partial\vec{f}_{ex}(\mu\vec{f}_{ex})) = \lambda_1 < 1$. As promised, there exist inductive upper bounds as can be seen in Figure 2. \triangle

3 Finding Inductive Upper Bounds in PPS

In this section, we are concerned with the following problem: Given a feasible, clean, and strongly connected PPS \vec{f} , find a vector $\vec{0} \prec \vec{u} \prec \infty$ such that $\vec{f}(\vec{u}) \leq \vec{u}$, i.e., an inductive upper bound on the lfp of \vec{f} (see Lemma 1).

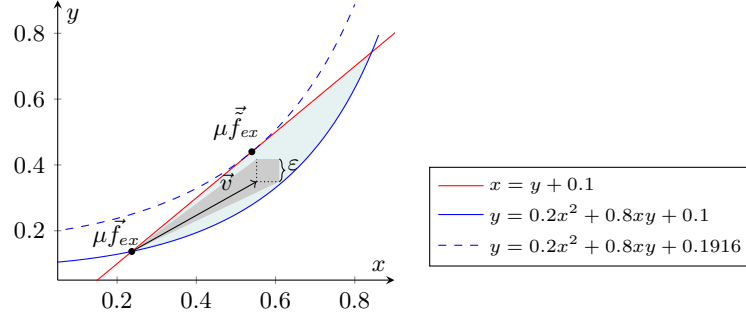


Fig. 2: The PPS \vec{f}_{ex} corresponds to the solid red line and the solid blue curve. Its inductive upper bounds form the shaded area above the lfp $\mu\vec{f}_{ex}$. Lemma 2(4) ensures that one can fit the gray “cone” pointing in direction of the Perron-Frobenius eigenvector \vec{v} inside the inductive region. The PPS \vec{f}_{ex} which comprises the dashed curve and the solid line does not have any non-trivial inductive upper bounds. Note that the tangent lines at $\mu\vec{f}_{ex}$ are parallel to each other.

3.1 Existence of Inductive Upper Bounds

An important first observation is that *inductive upper bounds other than the exact lfp do not necessarily exist*. As a simple counter-example consider the 1-dimensional PPS $x = \frac{1}{2}x^2 + \frac{1}{2}$. If u is an inductive upper bound, then

$$\frac{1}{2}u^2 + \frac{1}{2} \leq u \implies u^2 - 2u + 1 \leq 0 \implies (u - 1)^2 \leq 0 \implies u = 1,$$

and thus the only inductive upper bound is the exact lfp $u = 1$. Another example is the PPS \vec{f}_{ex} from Figure 2. What these examples have in common is the following property: Their derivative evaluated *at the lfp* is not invertible. Indeed, we have $\frac{\partial}{\partial x}(\frac{1}{2}x^2 + \frac{1}{2} - x) = x - 1$, and inserting the lfp $x = 1$ yields zero. The higher dimensional generalization of this property to arbitrary PPS \vec{f} is that the Jacobi matrix of the function $\vec{f} - \vec{x}$ evaluated at $\mu\vec{f}$ is singular; note that this is precisely the matrix $\partial\vec{f}(\mu\vec{f}) - I$. Geometrically, this means that the tangent lines at $\mu\vec{f}$ are parallel, as can be seen in Figure 2 for the example PPS \vec{f}_{ex} . It should be intuitively clear from the figure that *inductive upper bounds only exist if the tangent lines are not parallel*. The next lemma makes this more precise:

Lemma 2 (Existence of inductive upper bounds). *Let \vec{f} be a feasible, clean, and strongly connected PPS. Then the following are equivalent:*

- (1) *The matrix $I - \partial\vec{f}(\mu\vec{f})$ is non-singular.*
- (2) *The spectral radius of $\partial\vec{f}(\mu\vec{f})$ satisfies $\rho(\partial\vec{f}(\mu\vec{f})) < 1$.*
- (3) *There exists $\vec{0} \prec \vec{u} \prec \vec{\infty}$ s.t. $\vec{f}(\vec{u}) < \vec{u}$ (i.e. \vec{u} is inductive but not a fixpoint).*

- (4) The matrix $\partial \vec{f}(\mu \vec{f})$ has a unique (normalized) eigenvector $\vec{v} \succ \vec{0}$ and there exist numbers $\delta_{max} > 0$ and $\varepsilon > 0$ s.t.

$$\vec{f}(\mu \vec{f} + \delta \cdot \vec{v}) \prec \mu \vec{f} + \delta \cdot \vec{v}$$

holds for all $0 < \delta \leq \delta_{max}$ and vectors $\vec{v} \geq \vec{v}$ with $\|\vec{v} - \vec{v}\|_\infty \leq \varepsilon$.

The proof of Lemma 2 (see appendix) relies on a linear approximation of \vec{f} via Taylor’s familiar theorem as well as Theorems 1 and 2. Condition (4) of Lemma 2 means that there exists a “truncated cone”

$$C(\mu \vec{f}, \vec{v}, \varepsilon, \delta_{max}) = \{ \mu \vec{f} + \delta \vec{v} \mid 0 \leq \delta \leq \delta_{max}, \vec{v} \geq \vec{v}, \|\vec{v} - \vec{v}\|_\infty \leq \varepsilon \}$$

which is entirely contained in the inductive region. This cone is located at the lfp $\mu \vec{f}$ and points in the direction of the Perron-Frobenius eigenvector \vec{v} , as illustrated in Figure 2 (assuming $\delta_{max} = 1$ for simplicity). The length δ_{max} and the radius ε of the cone depend quantitatively on $\rho(\partial \vec{f}(\mu \vec{f}))$, but for our purposes it suffices that they are non-zero. The idea of our Optimistic Value Iteration is to *construct a sequence of guesses that eventually hits this cone*.

3.2 The Optimistic Value Iteration Algorithm

The basic idea of Optimistic Value Iteration (OVI) can be applied to monotone functions of the form $\vec{\phi}: \mathbb{R}_{\geq 0}^n \rightarrow \mathbb{R}_{\geq 0}^n$ (in [22], $\vec{\phi}$ is the Bellman operator of an MDP). Kleene’s fixpoint theorem suggests a simple method for approximating the lfp $\mu \vec{\phi}$ from below: Simply iterate $\vec{\phi}$ starting at $\vec{0}$, i.e., compute the sequence $\vec{l}_0 = \vec{0}$, $\vec{l}_1 = \vec{\phi}(\vec{l}_0)$, $\vec{l}_2 = \vec{\phi}(\vec{l}_1)$, etc.¹ In the context of MDP, this iterative scheme is known as *Value Iteration* (VI). VI is easy to implement, but it is difficult to decide when to stop the iteration. In particular, standard stopping criteria such as small absolute difference of consecutive approximations are formally unsound [20]. OVI and other algorithms [3,36] cope with this problem by computing not only a lower but also an *upper* bound on $\mu \vec{\phi}$. In the case of OVI, an upper bound with absolute error $\leq \varepsilon$ is obtained as follows (we omit some details):

- (1) Compute $\vec{l}_k \leq \mu \vec{\phi}$ such that $\|\vec{l}_k - \vec{l}_{k-1}\|_\infty \leq \tau$, for some (small) $\tau > 0$.
- (2) Guess a candidate upper bound $\vec{u} = \vec{l}_k + \vec{\varepsilon}$.
 - (a) If $\vec{\phi}(\vec{u}) \leq \vec{u}$ holds, i.e., \vec{u} is inductive, then return \vec{u} .
 - (b) If not, *refine* \vec{u} (see [22] for details). If the refined \vec{u} is still not inductive, then go back to step (1) and try again with $0 < \tau' < \tau$.

We present our variant of OVI for PPS as Algorithm 1. The main differences to the above scheme are that (i) we do not insist on Kleene iteration for obtaining the lower bounds \vec{l} , and (ii) we approximate the eigenvector \vec{v} from condition (4) of Lemma 2 and compute the “more informed” guesses $\vec{u} = \vec{l} + \varepsilon \vec{v}$, for various ε . *Refining the guesses* as original OVI does *is not necessary* (but see our remarks in Section 3.3 regarding floating point computations).

¹ In order for the Kleene sequence to converge to the lfp, i.e., $\lim_{k \rightarrow \infty} \vec{l}_k = \mu \vec{\phi}$, it suffices that $\vec{\phi}$ is ω -continuous. This already implies monotonicity.

Algorithm 1: Optimistic Value Iteration (OVI) for PPS

```

input      : strongly connected clean PPS  $\vec{f}$ ; maximum abs. error  $\varepsilon > 0$ 
output     : a pair  $(\vec{l}, \vec{u})$  of real vectors s.t.  $\vec{l} \leq \mu \vec{f}$ ,  $\vec{f}(\vec{u}) \leq \vec{u}$  (hence
                $\mu \vec{f} \leq \vec{u}$ ), and  $\|\vec{l} - \vec{u}\|_\infty \leq \varepsilon$ 
termination : guaranteed if  $\vec{f}$  is feasible and  $I - \partial \vec{f}(\mu \vec{f})$  is non-singular
1  $\vec{l} \leftarrow \vec{0}$ ;  $N \leftarrow 0$ ;
2  $\tau \leftarrow \varepsilon$ ;                                /*  $\tau$  is the current tolerance */
3 while true do
4    $\vec{l}' \leftarrow \text{improveLowerBound}(\vec{f}, \vec{l})$ ; /* e.g. Kleene or Newton update */
   /* guess and verify phase starts here */
5   if  $\|\vec{l}' - \vec{l}\|_\infty \leq \tau$  then
6      $\vec{v} \leftarrow \text{approxEigenvect}(\partial \vec{f}(\vec{l}), \tau)$ ; /* recall  $\vec{v}$  is normalized */
7     for  $k$  from 0 to  $N$  do
8        $\vec{u} \leftarrow \vec{l}' + d^k \varepsilon \cdot \vec{v}$ ; /* optimistic guess,  $d \in (0, 1)$  */
9       if  $\vec{f}(\vec{u}) \leq \vec{u}$  then
10         $\text{return } (\vec{l}', \vec{u})$ ; /* guess was successful */
11       $N \leftarrow N + 1$ ;
12     $\tau \leftarrow c \cdot \tau$ ; /* decrease tolerance for next guess,  $c \in (0, 1)$  */
13   $\vec{l} \leftarrow \vec{l}'$ ;
    
```

The functions `improveLowerBound` and `approxEigenvect` used in Algorithm 1 must satisfy the following contracts:

- The sequence $\vec{l}_0 = \vec{0}$, $\vec{l}_{i+1} = \text{improveLowerBound}(\vec{f}, \vec{l}_i)$ is a monotonically increasing sequence converging to the lfp $\mu \vec{f}$.
- `approxEigenvect` must satisfy the following: Let $M \geq 0$ be an irreducible matrix with (normalized) Perron-Frobenius eigenvector $\vec{v} \succ \vec{0}$. Then for all $\varepsilon > 0$, we require that there exists $\tau > 0$ such that $\|\text{approxEigenvect}(M, \tau) - \vec{v}\|_\infty \leq \varepsilon$. In words, `approxEigenvect` approximates \vec{v} up to arbitrarily small absolute error if the tolerance τ is chosen sufficiently small.

In practice, both the Kleene and the Newton [16,17,12] update operator can be used to implement `improveLowerBound`. We outline a possible implementation of `approxEigenvect` further below in Section 3.3.

Example 3. Consider the following PPS \vec{f} : $x = \frac{1}{4}x^2 + \frac{1}{8}$, $y = \frac{1}{4}xy + \frac{1}{4}y + \frac{1}{4}$. The table illustrates the execution of Algorithm 1 on \vec{f} with $\varepsilon = 0.1$ and $c = 0.5$:

| # | N | τ | \vec{l} | \vec{l}' | $\ \vec{l}' - \vec{l}\ _\infty$ | \vec{v} | \vec{u} | $\vec{f}(\vec{u}) \leq \vec{u}$ |
|---|-----|--------|------------|--------------|---------------------------------|------------|-------------|---------------------------------|
| 1 | 0 | 0.1 | (0, 0) | (0.4, 0.3) | 0.4 | | | |
| 2 | 0 | 0.1 | (0.4, 0.3) | (0.5, 0.4) | 0.1 | (1.0, 0.8) | (0.5, 0.38) | ✗ |
| 3 | 1 | 0.05 | (0.5, 0.4) | (0.55, 0.41) | 0.05 | (1.0, 0.9) | (0.6, 0.49) | ✓ |

The algorithm has to improve the lower bound 3 times (corresponding to the 3 lines of the table). After the second improvement, the difference between the current lower bound \vec{l}_2 and the new bound \vec{l}'_2 does not exceed the current tolerance $\tau_2 = 0.1$ and the algorithm enters the optimistic guessing stage. The first guess \vec{u}_2 is not successful. The tolerance is then decreased to $\tau_3 = c \cdot \tau_2 = 0.05$ and the lower bound is improved to \vec{l}'_3 . The next guess \vec{u}_3 is inductive. \triangle

Theorem 3. *Algorithm 1 is correct: when invoked with a strongly connected clean PPS \vec{f} and $\varepsilon > 0$, then (if it terminates) it outputs a pair (\vec{l}, \vec{u}) s.t. $\vec{l} \leq \mu\vec{f}$, $\vec{f}(\vec{u}) \leq \vec{u}$ (and thus $\mu\vec{f} \leq \vec{u}$), and $\|\vec{l} - \vec{u}\|_\infty \leq \varepsilon$. Moreover, if \vec{f} is feasible and $I - \partial\vec{f}(\mu\vec{f})$ is non-singular, then the algorithm terminates.*

The proof of Theorem 3 (see appendix) crucially relies on condition (4) of Lemma 2 that assures the existence of a “truncated cone” of inductive bounds centered around the Perron-Frobenius eigenvector of $\partial\vec{f}(\mu\vec{f})$ (see Figure 2 for an illustration). Intuitively, since the lower bounds \vec{l} computed by the algorithm approach the lfp $\mu\vec{f}$, the eigenvectors of $\partial\vec{f}(\vec{l})$ approach those of $\partial\vec{f}(\mu\vec{f})$. As a consequence, it is guaranteed that the algorithm eventually finds an eigenvector that intersects the cone. The inner loop starting on line 7 is needed because the “length” of the cone is a priori unknown; the purpose of the loop is to scale the eigenvector down so that it is ultimately small enough to fit inside the cone.

3.3 Considerations for Implementing OVI

As mentioned above, there are at least two options for `improveLowerBound`: Kleene or Newton iteration. We now show that `approxEigenvector` can be effectively implemented as well. Further below we make some remarks on floating point arithmetic.

Approximating the Eigenvector. A possible implementation of `approxEigenvector` relies on the *power iteration* method (e.g. [37, Thm. 4.1]). Given a square matrix M and an initial vector \vec{v}_0 with $M\vec{v}_0 \neq \vec{0}$, power iteration computes the sequence $(\vec{v}_i)_{i \geq 0}$ such that for $i > 0$, $\vec{v}_i = M\vec{v}_{i-1} / \|M\vec{v}_{i-1}\|_\infty$.

Lemma 3. *Let $M \geq 0$ be irreducible. Then power iteration applied to $M + I$ and any $\vec{v}_0 > \vec{0}$ converges to the Perron-Frobenius eigenvector $\vec{v} \succ \vec{0}$ of M .*

The convergence rate of power iteration is determined by the ratio $|\lambda_2|/|\lambda_1|$ where λ_1 and λ_2 are eigenvalues of largest and second largest absolute value, respectively. Each time `approxEigenvector` is called in Algorithm 1, the result of the previous call to `approxEigenvector` (if available) may be used as an initial approximation \vec{v}_0 .

Exact vs Floating Point Arithmetic. So far we have assumed exact arithmetic for the computations in Algorithm 1, but an actual implementation should use floating point arithmetic for efficiency. However, this may (and actually does) lead to unsound results. More specifically, the condition $\vec{f}(\vec{u}) \leq \vec{u}$ may hold in floating point arithmetic even though it is actually violated. As a remedy, we propose to nevertheless run the algorithm with floats, but then verify its output \vec{u} with exact arbitrary-precision rational arithmetic. That is, we compute a rational number approximation $\vec{u}_{\mathbb{Q}}$ of \vec{u} and check $\vec{f}(\vec{u}_{\mathbb{Q}}) \leq \vec{u}_{\mathbb{Q}}$ with exact arithmetic. If the check fails, we resort to the following refinement scheme which is an instance of the general k -induction principle for complete lattices from [5]: We iteratively check the conditions

$$\vec{f}(\vec{u}_{\mathbb{Q}}) \sqcap \vec{f}(\vec{u}_{\mathbb{Q}}) \leq \vec{u}_{\mathbb{Q}}, \quad \vec{f}(\vec{u}_{\mathbb{Q}} \sqcap \vec{f}(\vec{u}_{\mathbb{Q}} \sqcap \vec{f}(\vec{u}_{\mathbb{Q}}))) \leq \vec{u}_{\mathbb{Q}}, \quad \text{and so on,}$$

where \sqcap denotes pointwise minimum. If one of the checks is satisfied, then $\mu \vec{f} \leq \vec{u}_{\mathbb{Q}}$ [5]. This scheme often works well in practice (see Section 5). The original OVI from [22] uses a similar technique to refine its guesses.

4 Certificates for Probabilistic Pushdown Automata

This section shows how the results from Section 3 can be applied to pPDA. We introduce some additional notation. For finite sets A , $\mathcal{D}(A)$ denotes the set of *probability distributions* on A . We often denote tuples or triples without parentheses and separating commata when this causes no confusion, e.g., we may write ab rather than (a, b) .

Definition 1 (pPDA [13]). *A probabilistic pushdown automaton (pPDA) is a triple $\Delta = (Q, \Gamma, P)$ where $Q \neq \emptyset$ is a finite set of states, $\Gamma \neq \emptyset$ is a finite stack alphabet, and $P: Q \times \Gamma \rightarrow \mathcal{D}(Q \times \Gamma^{\leq 2})$ is a probabilistic transition function.*

In the following, we often write $qZ \xrightarrow{p} r\alpha$ instead of $P(qZ)(r\alpha) = p$ [13]. Intuitively, $qZ \xrightarrow{p} r\alpha$ means that if the pPDA is in state q and Z is on top of the stack, then with probability p , the pPDA moves to state r , pops Z and pushes α on the stack. More formally, the semantics of a pPDA $\Delta = (Q, \Gamma, P)$ is a countably infinite Markov chain with state space $Q \times \Gamma^*$ and transition probability matrix M such that for all $q, r \in Q$, $Z \in \Gamma$, $\alpha \in \Gamma^{\leq 2}$, $\gamma \in \Gamma^*$, we have

$$M(qZ\gamma, r\alpha\gamma) = P(qZ)(r\alpha), \quad M(q\varepsilon, q\varepsilon) = 1,$$

and all other transition probabilities are zero. This Markov chain, where the initial state is fixed to qZ , is denoted $\mathcal{M}_{\Delta}^{qZ}$ (see Figure 3 for an example). As usual, one can formally define a probability measure \mathbb{P}_{Δ}^{qZ} on the infinite runs of $\mathcal{M}_{\Delta}^{qZ}$ via the standard cylinder construction (e.g., [2, Sec. 10]).

Consider a triple $qZr \in Q \times \Gamma \times Q$. We define the *return probability*² $[qZr]$ as the probability of reaching $r\varepsilon$ in the Markov chain $\mathcal{M}_{\Delta}^{qZ}$, i.e., $[qZr] = \mathbb{P}_{\Delta}^{qZ}(\diamond\{r\varepsilon\})$, where $\diamond\{r\varepsilon\}$ is the set of infinite runs of $\mathcal{M}_{\Delta}^{qZ}$ that eventually hit state $r\varepsilon$.

² When modeling procedural programs with pPDA, $[qZr]$ is the probability that a given procedure *returns* a specific value to its calling context. These probabilities

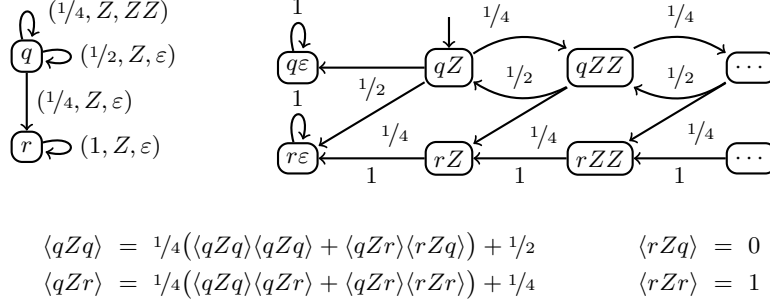


Fig. 3: Top left: The pPDA $\Delta_{ex} = (\{q, r\}, \{Z\}, P)$ where P comprises the transitions $qZ \xrightarrow{1/4} qZZ$, $qZ \xrightarrow{1/2} q\epsilon$, $qZ \xrightarrow{1/4} r\epsilon$, $rZ \xrightarrow{1} r\epsilon$. Top right: A fragment of the *infinite* underlying Markov chain, assuming initial configuration qZ . Bottom: The associated equation system from Theorem 4.

Theorem 4 (The PPS of return probabilities [13]). *Let $\Delta = (Q, \Gamma, P)$ be a pPDA and $(\langle qZr \rangle)_{qZr \in Q \times \Gamma \times Q}$ be variables. For each $\langle qZr \rangle$, define*

$$\langle qZr \rangle = \sum_{qZ \xrightarrow{p} sYX} p \cdot \sum_{t \in Q} \langle sYt \rangle \cdot \langle tXr \rangle + \sum_{qZ \xrightarrow{p} sY} p \cdot \langle sYr \rangle + \sum_{qZ \xrightarrow{p} r\epsilon} p$$

and call the resulting PPS \vec{f}_Δ . Then $\mu \vec{f}_\Delta = ([qZr])_{qZr \in Q \times \Gamma \times Q}$.

We refer to [30, Sec. 3] for an intuitive explanation of the equations in \vec{f}_Δ .

Example 4. Figure 3 shows a pPDA Δ_{ex} and the associated PPS $\vec{f}_{\Delta_{ex}}$. The least non-negative solution is $\langle qZq \rangle = 2 - \sqrt{2} \approx 0.586$ and $\langle qZr \rangle = \sqrt{2} - 1 \approx 0.414$ (and, of course, $\langle rZq \rangle = 0$, $\langle rZr \rangle = 1$). Thus by Theorem 4, the return probabilities are $[qZq] = 2 - \sqrt{2}$ and $[qZr] = \sqrt{2} - 1$. \triangle

The PPS \vec{f}_Δ is always feasible (because $\mu \vec{f}_\Delta \leq \vec{1}$). \vec{f}_Δ is neither necessarily strongly connected nor clean. Let \tilde{f}_Δ denote the cleaned up version of \vec{f}_Δ .

Proposition 1 (Basic Certificates for pPDA). *A basic certificate for $\Delta = (Q, \Gamma, P)$ is a rational inductive upper bound $\vec{u} \in \mathbb{Q}_{\geq 0}^{Q \times \Gamma \times Q}$ on the lfp of the return probabilities system \vec{f}_Δ (see Thm. 4). They have the following properties:*

- (Existence) $\forall \varepsilon > 0$ there exists a basic certificate \vec{u} with $\|\mu \vec{f}_\Delta - \vec{u}\|_\infty \leq \varepsilon$ if all maximal irreducible submatrices M of $\partial \tilde{f}_\Delta(\mu \vec{f}_\Delta)$ satisfy $\rho(M) < 1$.

were called *termination probabilities* in previous works [12,7] but we believe this term is more appropriate for the numbers $[qZ\downarrow] = \sum_r [qZr]$, i.e., the probability to eventually reach the empty stack from initial configuration qZ .

- (Complexity) Let β be the maximum number of bits used to encode any of the numerators and denominators of the fractions occurring in $\vec{u} \in \mathbb{Q}_{\geq 0}^{Q \times \Gamma \times Q}$. Then checking $\vec{f}_\Delta(\vec{u}) \leq \vec{u}$, i.e., whether \vec{u} is basic certificate for Δ , can be done in time polynomial in β and the size of Δ .

Existence of basic certificates follows from Lemma 2 applied to each SCC of the cleaned-up version of \vec{f}_Δ individually. However, note that in order to merely *check* the certificate, i.e., verify the inequality $\vec{f}(\vec{u}) \leq \vec{u}$, neither do SCCs need to be computed nor does the system has to be cleaned up.

Example 5. Reconsider the example pPDA and its associated (non-strongly connected) system of return probabilities from Figure 3. We verify that $\vec{u}_{qZq} = 3/5$ and $\vec{u}_{qZr} = 1/2$ (as well as $\vec{u}_{rZq} = 0, \vec{u}_{rZr} = 1$) is a basic certificate:

$$\frac{1}{4} \left(\frac{3}{5} \cdot \frac{3}{5} + \frac{1}{2} \cdot 0 \right) + \frac{1}{2} = \frac{59}{100} \leq \frac{3}{5} \quad , \quad \frac{1}{4} \left(\frac{3}{5} \cdot \frac{1}{2} + \frac{1}{2} \cdot 1 \right) + \frac{1}{4} = \frac{45}{100} \leq \frac{1}{2} .$$

Note that $[qZq] \approx 0.586 \leq 3/5 = 0.6$ and $[qZr] \approx 0.414 \leq 1/2 = 0.5$. \triangle

In the following we outline how a variety of key quantities associated to pPDA can be verified using basic certificates. More details are in the appendix.

Upper Bounds on Temporal Properties. We may use basic certificates to verify that a bad state r_{bad} is reached with low probability, e.g., at most $p = 0.01$. To this end, we remove the outgoing transitions of r_{bad} and add the transitions $r_{bad}Z \xrightarrow{1} r_{bad}\varepsilon$ for all $Z \in \Gamma$. Clearly, r_{bad} is reached with probability at most p from initial configuration qZ iff $[qZr_{bad}] \leq p$. The results of [13] imply that this idea can be generalized to *until*-properties of the form $\mathcal{C}_1 \mathcal{U} \mathcal{C}_2$, where \mathcal{C}_1 and \mathcal{C}_2 are *regular* sets of configurations. (This requires a small extension of the basic certificates, but the overall idea stays the same).

Certificates for the Output Distribution. Once a pPDA reaches the empty stack, we say that it has *terminated*. When modeling procedural programs, this corresponds to returning from a program's main procedure. Assuming initial configuration qZ , the probability sub-distribution over the possible return values is then given by the return probabilities $\{[qZr] \mid r \in Q\}$. Missing probability mass models the probability of non-termination. A basic certificate can thus be used immediately to verify a point-wise upper bound on the output distribution as well as to certify that a program is not *almost-surely terminating* (AST). If a pPDA Δ is already known to be AST, then we can also certify a lower bound on the output distribution: Suppose that \vec{u} is a basic certificate for Δ and assume that Δ is AST from initial configuration qZ . Define $\varepsilon = \sum_{r \in Q} \vec{u}_{qZr} - 1$. Then for all $r \in Q$, we have $\vec{u}_{qZr} - \varepsilon \leq [qZr] \leq \vec{u}_{qZr}$.

Example 6. The pPDA Δ_{ex} from Figure 3 is AST from initial configuration qZ , as the transition $qZ \xrightarrow{1/4} r\varepsilon$ is eventually taken with probability 1, and the stack is emptied certainly once r is reached. Using the basic certificate from Example 5 we can thus (correctly) certify that $0.5 \leq [qZq] \leq 0.6$ and $0.4 \leq [qZr] \leq 0.5$.

Certificates for Expected Rewards or Costs. Suppose we have equipped a pPDA with a state-based reward (or cost) function $Q \rightarrow \mathbb{R}_{\geq 0}$. It was shown in [14] that the expected total reward accumulated during the run of a pPDA is the solution of a **linear** equation system where the return probabilities $[qZr]$ appear as coefficients. Given a basic certificate \vec{u} , we can replace each coefficient $[qZr]$ by \vec{u}_{qZr} and thus obtain an equation system whose solution is an over-approximation of the true expected reward. We may extend the basic certificate \vec{u} by the solution of this linear system to make verification straightforward. Note that a program's *expected runtime* [8,35] is a special case of total expected reward.

5 Implementation and Experiments

Our Tool: PRAY. We implemented our algorithm in the prototypical Java-tool PRAY (Probabilistic Recursion ANALyzer). It supports two input formats: (i) Recursive probabilistic programs in a Java-like syntax (e.g. Figure 4); these programs are automatically translated to pPDA. (ii) Explicit PPS in the same syntax used by the tool PREMO [43]. The output of PRAY is a rational *inductive* upper bound on the lfp of the return probability PPS of the input program's pPDA model (a basic certificate), or on the lfp of the explicitly given PPS. The absolute precision ε is configurable. The implementation works as follows:

- (1) It parses the input and, if the latter was a program, constructs a pPDA model and the associated PPS of return probabilities.
- (2) It computes an SCC decomposition of the PPS under consideration using standard algorithms implemented in the JGRAPH library [33].
- (3) It applies Algorithm 1 to the individual SCC in reverse topological order using floating point arithmetic. Algorithm 1 is instantiated with Kleene iteration³, the power iteration for approximating eigenvectors as outlined in Section 3.3, and constants $c = 0.1$, $d = 0.5$. We allow ≤ 10 guesses per SCC.
- (4) If stage (3) is successful, the tool verifies the resulting floating point certificate using exact rational number arithmetic as described in Section 3.3.

Baselines. To the best of our knowledge, no alternative techniques for finding *inductive* upper bounds in PPS have been described explicitly in the literature. However, there is an (almost) out-of-the-box approach using an SMT solver: Given a PPS $\vec{x} = \vec{f}(\vec{x})$, compute some lower bound $\vec{l} \leq \mu\vec{f}$ using an iterative technique. Then query the SMT solver for a model (variable assignment) of the quantifier-free first-order logic formula $\varphi_{\vec{f}}(\vec{x}) = \bigwedge_{i=1}^n f_i(\vec{x}) \leq x_i \wedge \vec{l}_i \leq x_i \leq \vec{l}_i + \varepsilon$ in the (decidable) theory of polynomial real arithmetic with inequality (aka QF_NRA in the SMT community). If such a model \vec{u} exists, then clearly $\mu\vec{f} \leq \vec{u}$ and $\|\vec{l} - \vec{u}\|_\infty \leq \varepsilon$. If no model exists, then improve \vec{l} and try again. We have

³ In fact, we use the slightly optimized Gauss-Seidel iteration (see [42, Sec. 5.2]) which provides a good trade-off between ease of implementation and efficiency [42].

```

bool and() {
    prob {
        1//2: return
        (1//2: true | 1//2: false);
        1//2: {
            if(!or()) return false;
            else return or(); } } }

bool or() {
    prob {
        1//2: return
        (1//2: true | 1//2: false);
        1//2: {
            if(and()) return true;
            else return and(); } } }
    
```

Fig. 4: Program evaluating a random and-or tree [8]. The **prob**-blocks execute the contained statements with the respective probabilities (syntax inspired by Java’s **switch**). Our tool automatically translates this program to a pPDA and computes a basic certificate (Proposition 1) witnessing that calling **and()** returns **true** and **false** with probability $\leq 382/657 \approx 0.58$ and $391/933 \approx 0.42$, resp.

implemented this approach using the state-of-the-art SMT solvers CVC5 [4] and Z3 [34], the winners of the 2022 SMT-COMP in the category QF_NRA⁴.

As yet another baseline, we have also implemented a variant of OVI for PPS which is closer to the original MDP algorithm from [22]. In this variant, called “standard OVI” from now on, we compute the candidate \vec{u} based on the “relative” update rule $\vec{u} = (1 + \varepsilon)\vec{l}$, where \vec{l} is the current lower bound [22].

Research Questions. We aim to shed some light on the following questions: (A) How well does our algorithm scale? (B) Is the algorithm suitable for PPS with different characteristics, e.g., dense or sparse? (C) Is the requirement $\rho(\partial \vec{f}(\mu \vec{f})) < 1$ restrictive in practice? (D) How does our OVI compare to the baselines?

Benchmarks. To answer the above questions we run our implementation on two sets of benchmarks (Table 3 and Table 2, respectively). The first set consists of various example programs from the literature as well as a few new programs, which are automatically translated to pPDA. This translation is standard and usually takes not more than a few seconds. The programs **golden**, **and-or** (see Figure 4), **virus**, **gen-fun** are adapted from [35,8,41] and [32, Program 5.6], respectively. The source code of all considered programs is in the appendix. We have selected only programs with possibly unbounded recursion depth which induce *infinite* Markov chains. The second benchmark set comprises explicitly given PPS⁵. The instances **brown**, **lemonde**, **negra**, **swbd**, **tiger**, **tuebadz**, and **wsj** all encode SCFG from the area of language processing (see [43] for details). **random** is the return probability system of a randomly generated pPDA.

Summary of Experimental Results. We ran the experiments on a standard notebook. The approach based on CVC5 turns out to be not competitive (see Appendix D). We thus focus on Z3 in the following. Both PRAY and the Z3 approach could handle most of the programs from Table 3 within a 10 minute time limit. The considered programs induce sparse PPS with 38 - 26,367 variables, and most

⁴ <https://smt-comp.github.io/2022/results>

⁵ These examples come with PReMo: <https://cgi.csc.liv.ac.uk/~dominik/premo/>

Table 1: Experiments with PPS obtained from recursive probabilistic programs. Columns *vars* and *terms* display the number of variables and terms in the PPS. Columns *sccs* and *scc_{max}* indicate the number of non-trivial SCC and the size of the largest SCC. *G* is total number of guesses made by OVI (at least one guess per SCC). *t_{tot}* is the total runtime excluding the time for model construction. *t_Q* is the percentage of *t_{tot}* spent on exact rational arithmetic. *D* is the average number of decimal digits of the rational numbers in the certificate. The timeout (TO) was set to 10 minutes. Timings are in ms. The absolute precision is $\varepsilon = 10^{-3}$.

| benchmark | Q | P | I | vars | terms | sccs | scc _{max} | cert | G | D | t _Q | t _{tot} | cert _{z3} | D _{z3} | t _{z3} | cert _{std} | G _{std} | D _{std} | t _{std} |
|----------------|------|------|------|-------|-------|------|--------------------|------|----|---|----------------|------------------|--------------------|-----------------|-----------------|---------------------|------------------|------------------|------------------|
| rw-0.499 | 18 | 29 | 5 | 38 | 45 | 1 | 12 | ✓ | 5 | 5 | 17% | 163 | ✓ | 2 | 11 | ✓ | 4 | 5 | 59 |
| rw-0.500 | 18 | 29 | 5 | 38 | 45 | 1 | 12 | ✗ | 10 | - | - | 7327 | ✓ | 2 | 10 | ✗ | 10 | - | 8083 |
| rw-0.501 | 18 | 29 | 5 | 38 | 45 | 1 | 12 | ✓ | 5 | 4 | 6% | 36 | ✓ | 13 | 12 | ✓ | 4 | 5 | 23 |
| geom-offspring | 24 | 40 | 5 | 52 | 80 | 4 | 24 | ✓ | 8 | 6 | 13% | 15 | ✓ | 9 | 16 | ✓ | 8 | 6 | 14 |
| golden | 27 | 49 | 6 | 81 | 94 | 1 | 36 | ✓ | 1 | 5 | 30% | 10 | ✓ | 7 | 14 | ✓ | 2 | 4 | 12 |
| and-or | 50 | 90 | 7 | 149 | 182 | 1 | 48 | ✓ | 2 | 4 | 26% | 19 | ✓ | 12 | 15260 | ✓ | 2 | 4 | 19 |
| gen-fun | 85 | 219 | 7 | 202 | 327 | 1 | 16 | ✓ | 2 | 3 | 32% | 22 | ✓ | 15 | 141 | ✓ | 2 | 3 | 21 |
| virus | 68 | 149 | 27 | 341 | 551 | 1 | 220 | ✓ | 1 | 5 | 38% | 40 | ✓ | 7 | 139 | ✓ | 1 | 6 | 59 |
| escape10 | 109 | 174 | 23 | 220 | 263 | 1 | 122 | ✓ | 1 | 4 | 5% | 56 | ✓ | 7 | 48 | ✓ | 1 | 8 | 71 |
| escape25 | 258 | 413 | 53 | 518 | 621 | 1 | 300 | ✓ | 1 | 5 | 17% | 245 | ✓ | 7 | 15958 | ✓ | 1 | 9 | 172 |
| escape50 | 508 | 813 | 103 | 1018 | 1221 | 1 | 600 | ✓ | 1 | 7 | 23% | 653 | ✓ | 7 | 410 | ✗ | 1 | - | 400 |
| escape75 | 760 | 1215 | 153 | 1522 | 1825 | 1 | 904 | ✓ | 2 | 9 | 10% | 3803 | ✗ | - | TO | ✗ | 1 | - | 635 |
| escape100 | 1009 | 1614 | 203 | 2020 | 2423 | 1 | 1202 | ✗ | 5 | - | - | 29027 | ✓ | 6 | 939 | ✗ | 1 | - | 901 |
| escape200 | 2008 | 3213 | 403 | 4018 | 4821 | 1 | 2400 | ✗ | 6 | - | - | 83781 | ✗ | - | TO | ✗ | 1 | - | 2206 |
| sequential5 | 230 | 490 | 39 | 1017 | 1200 | 10 | 12 | ✓ | 15 | 4 | 26% | 103 | ✓ | 8 | 1074 | ✓ | 15 | 5 | 204 |
| sequential7 | 572 | 1354 | 137 | 3349 | 3856 | 14 | 12 | ✓ | 21 | 5 | 27% | 1049 | ✓ | 8 | 12822 | ✓ | 20 | 5 | 1042 |
| sequential10 | 3341 | 8666 | 1036 | 26367 | 29616 | 20 | 12 | ✓ | 30 | 5 | 2% | 100613 | ✓ | 8 | 453718 | ✓ | 30 | 6 | 101554 |
| mod5 | 44 | 103 | 10 | 296 | 425 | 1 | 86 | ✓ | 1 | 5 | 39% | 28 | ✓ | 9 | 34150 | ✗ | 2 | - | 178 |
| mod7 | 64 | 159 | 14 | 680 | 1017 | 1 | 222 | ✓ | 1 | 6 | 69% | 172 | ✓ | 7 | 443 | ✗ | 2 | - | 624 |
| mod10 | 95 | 244 | 20 | 1574 | 2403 | 1 | 557 | ✗ | 1 | - | - | 675 | ✓ | 7 | 1245 | ✗ | 2 | - | 882 |

of them have just a single SCC. Notably, the examples with greatest maximum SCC size were only solved by z3. PRAY and z3 need at most 95 and 31 seconds, respectively, for the instances where they succeed. In many cases (e.g., rw-5.01, golden, virus, brown, swbd), the resulting certificates formally disprove AST. For the explicit PPS in Table 2, PRAY solves all instances whereas z3 only solves 3/8 within the time limit, and only finds the trivial solution $\vec{1}$. Most of these benchmarks contain dense high-degree polynomials and our tool spends most time on performing exact arithmetic. PRAY never needs more than 6 guesses per SCC if it succeeds.

Evaluation of Research Questions. (A) Scalability: Our algorithm succeeded on instances with maximum SCC size of up to 8,000 and number of terms over 50,000. PRAY could solve all instances with a maximum SCC size of $\leq 1,000$ in less than 2 minutes per instance. For the examples where our algorithm does not succeed (e.g., escape100) it is mostly because it fails converting a floating point to a rational certificate. (B) PPS with different flavors: The problems in Table 3 (low degree and sparse, i.e., few terms per polynomials) and Table 2 (higher degree and dense) are quite different. A comparison to the SMT approach suggests that our technique might be especially well suited for dense problems with higher degrees. (C) Non-singularity: The only instance where our algorithm fails because of the non-singularity condition is the symmetric random walk rw-

Table 2: Experiments with explicitly given PPS (setup as in Table 3).

| benchmark | vars | terms | scs | scs _{max} | cert | G | D | t _Q | t _{tot} | cert _{z3} | D _{z3} | t _{z3} | cert _{std} | G _{std} | D _{std} | t _{std} |
|-----------|-------|-------|-----|--------------------|------|---|---|----------------|------------------|--------------------|-----------------|-----------------|---------------------|------------------|------------------|------------------|
| brown | 37 | 22866 | 1 | 22 | ✓ | 2 | 6 | 74% | 3212 | ✗ | - | TO | ✓ | 2 | 8 | 9065 |
| lemonde | 121 | 32885 | 1 | 48 | ✓ | 2 | 5 | 97% | 40738 | ✗ | - | TO | ✓ | 2 | 5 | 38107 |
| negra | 256 | 29297 | 1 | 149 | ✓ | 2 | 7 | 89% | 10174 | ✓ | 1 | 37248 | ✓ | 1 | 7 | 8873 |
| swbd | 309 | 47578 | 1 | 243 | ✓ | 1 | 7 | 93% | 18989 | ✗ | - | TO | ✓ | 1 | 8 | 67314 |
| tiger | 318 | 52184 | 1 | 214 | ✓ | 2 | 8 | 98% | 94490 | ✓ | 1 | 17454 | ✓ | 1 | 8 | 90801 |
| tuebadz | 196 | 8932 | 2 | 168 | ✓ | 4 | 9 | 85% | 2666 | ✓ | 1 | 15323 | ✓ | 3 | 9 | 2700 |
| wsj | 240 | 31170 | 1 | 194 | ✓ | 2 | 9 | 96% | 30275 | ✗ | - | TO | ✓ | 2 | 9 | 29038 |
| random | 10000 | 20129 | 1 | 8072 | ✓ | 3 | 7 | 5% | 17585 | ✗ | - | TO | ✓ | 4 | 8 | 16357 |

0.500. We therefore conjecture that this condition is often satisfied in practice. (D) Comparison to SMT: There is no clear winner. Some instances can only be solved by one tool or the other (e.g. `escape100` and `brown`). However, PRAY often delivers more succinct certificates, i.e., the rational numbers have less digits. Overall, z3 behaves less predictably than PRAY.

6 Conclusion and Future Work

We have proposed using inductive bounds as certificates for various properties in probabilistic recursive models. Moreover, we have presented the first dedicated algorithm for computing inductive upper bounds. While our algorithm already scales to non-trivial problems, the main bottleneck is the generation of an exact rational bound from a floating point approximation. This might be improved using appropriate rounding modes as in [21]. Additional future work includes further certificates for pPDA, especially for *lower* bounds and termination.

References

1. Azeem, M., Evangelidis, A., Kretínský, J., Slivinskiy, A., Weininger, M.: Optimistic and Topological Value Iteration for Simple Stochastic Games. CoRR **abs/2207.14417** (2022)
2. Baier, C., Katoen, J.: Principles of model checking. MIT Press (2008)
3. Baier, C., Klein, J., Leuschner, L., Parker, D., Wunderlich, S.: Ensuring the Reliability of Your Model Checker: Interval Iteration for Markov Decision Processes. In: CAV (1). Lecture Notes in Computer Science, vol. 10426, pp. 160–180. Springer (2017)
4. Barbosa, H., Barrett, C.W., Brain, M., Kremer, G., Lachnitt, H., Mann, M., Mohamed, A., Mohamed, M., Niemetz, A., Nötzli, A., Ozdemir, A., Preiner, M., Reynolds, A., Sheng, Y., Tinelli, C., Zohar, Y.: cvc5: A versatile and industrial-strength SMT solver. In: TACAS (1). Lecture Notes in Computer Science, vol. 13243, pp. 415–442. Springer (2022)
5. Batz, K., Chen, M., Kaminski, B.L., Katoen, J., Matheja, C., Schröer, P.: Latticed k-induction with an application to probabilistic programs. In: CAV (2). Lecture Notes in Computer Science, vol. 12760, pp. 524–549. Springer (2021)

6. Bishop, C.M.: Model-based machine learning. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **371**(1984), 20120222 (2013)
7. Brázdil, T., Esparza, J., Kiefer, S., Kucera, A.: Analyzing probabilistic pushdown automata. *Formal Methods Syst. Des.* **43**(2), 124–163 (2013)
8. Brázdil, T., Kiefer, S., Kucera, A., Vareková, I.H.: Runtime analysis of probabilistic programs with unbounded recursion. *J. Comput. Syst. Sci.* **81**(1), 288–310 (2015)
9. Chiang, D., Riley, D.: Factor Graph Grammars. In: *NeurIPS* (2020)
10. Esparza, J., Gaiser, A., Kiefer, S.: Computing Least Fixed Points of Probabilistic Systems of Polynomials. In: *STACS. LIPIcs*, vol. 5, pp. 359–370. Schloss Dagstuhl - Leibniz-Zentrum für Informatik (2010)
11. Esparza, J., Kiefer, S., Luttenberger, M.: Convergence Thresholds of Newton’s Method for Monotone Polynomial Equations. In: *STACS. LIPIcs*, vol. 1, pp. 289–300. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, Germany (2008)
12. Esparza, J., Kiefer, S., Luttenberger, M.: Computing the Least Fixed Point of Positive Polynomial Systems. *SIAM J. Comput.* **39**(6), 2282–2335 (2010)
13. Esparza, J., Kucera, A., Mayr, R.: Model Checking Probabilistic Pushdown Automata. In: *LICS*. pp. 12–21. IEEE Computer Society (2004)
14. Esparza, J., Kucera, A., Mayr, R.: Quantitative Analysis of Probabilistic Pushdown Automata: Expectations and Variances. In: *LICS*. pp. 117–126. IEEE Computer Society (2005)
15. Esparza, J., Lammich, P., Neumann, R., Nipkow, T., Schimpf, A., Smaus, J.: A fully verified executable LTL model checker. In: *CAV. Lecture Notes in Computer Science*, vol. 8044, pp. 463–478. Springer (2013)
16. Etessami, K., Yannakakis, M.: Recursive Markov Chains, Stochastic Grammars, and Monotone Systems of Nonlinear Equations. In: *STACS. Lecture Notes in Computer Science*, vol. 3404, pp. 340–352. Springer (2005)
17. Etessami, K., Yannakakis, M.: Recursive Markov chains, stochastic grammars, and monotone systems of nonlinear equations. *J. ACM* **56**(1), 1:1–1:66 (2009)
18. Flajolet, P., Sedgewick, R.: *Analytic Combinatorics*. Cambridge University Press (2009)
19. Funke, F., Jantsch, S., Baier, C.: Farkas Certificates and Minimal Witnesses for Probabilistic Reachability Constraints. In: *TACAS (1). Lecture Notes in Computer Science*, vol. 12078, pp. 324–345. Springer (2020)
20. Haddad, S., Monmege, B.: Reachability in mdps: Refining convergence of value iteration. In: *RP. Lecture Notes in Computer Science*, vol. 8762, pp. 125–137. Springer (2014)
21. Hartmanns, A.: Correct Probabilistic Model Checking with Floating-Point Arithmetic. In: *TACAS (2). Lecture Notes in Computer Science*, vol. 13244, pp. 41–59. Springer (2022)
22. Hartmanns, A., Kaminski, B.L.: Optimistic value iteration. In: *CAV (2). Lecture Notes in Computer Science*, vol. 12225, pp. 488–511. Springer (2020)
23. Jantsch, S.: *Certificates and Witnesses for Probabilistic Model Checking*. Ph.D. thesis, Dresden University of Technology, Germany (2022)
24. Jantsch, S., Funke, F., Baier, C.: Minimal Witnesses for Probabilistic Timed Automata. In: *ATVA. Lecture Notes in Computer Science*, vol. 12302, pp. 501–517. Springer (2020)
25. Jurafsky, D., Wooters, C., Segal, J., Stolcke, A., Fosler, E., Tajchman, G.N., Morgan, N.: Using a stochastic context-free grammar as a language model for speech recognition. In: *ICASSP*. pp. 189–192. IEEE Computer Society (1995)

26. Karp, R.M.: An introduction to randomized algorithms. *Discret. Appl. Math.* **34**(1-3), 165–201 (1991)
27. Kiefer, S., Luttenberger, M., Esparza, J.: On the convergence of newton’s method for monotone systems of polynomial equations. In: *STOC*. pp. 217–226. ACM (2007)
28. Knudsen, B., Hein, J.: Pfold: RNA secondary structure prediction using stochastic context-free grammars. *Nucleic Acids Res.* **31**(13), 3423–3428 (2003)
29. Kobayashi, N., Lago, U.D., Grellois, C.: On the Termination Problem for Probabilistic Higher-Order Recursive Programs. *Log. Methods Comput. Sci.* **16**(4) (2020)
30. Kucera, A., Esparza, J., Mayr, R.: Model checking probabilistic pushdown automata. *Log. Methods Comput. Sci.* **2**(1) (2006)
31. McConnell, R.M., Mehlhorn, K., Näher, S., Schweitzer, P.: Certifying algorithms. *Comput. Sci. Rev.* **5**(2), 119–161 (2011)
32. van de Meent, J., Paige, B., Yang, H., Wood, F.: An Introduction to Probabilistic Programming. *CoRR* **abs/1809.10756** (2018)
33. Michail, D., Kinable, J., Naveh, B., Sichi, J.V.: Jgrapht - A java library for graph data structures and algorithms. *ACM Trans. Math. Softw.* **46**(2), 16:1–16:29 (2020)
34. de Moura, L.M., Bjørner, N.S.: Z3: an efficient SMT solver. In: *TACAS. Lecture Notes in Computer Science*, vol. 4963, pp. 337–340. Springer (2008)
35. Olmedo, F., Kaminski, B.L., Katoen, J., Matheja, C.: Reasoning about Recursive Probabilistic Programs. In: *LICS*. pp. 672–681. ACM (2016)
36. Quatmann, T., Katoen, J.: Sound Value Iteration. In: *CAV* (1). *Lecture Notes in Computer Science*, vol. 10981, pp. 643–661. Springer (2018)
37. Saad, Y.: Numerical methods for large eigenvalue problems: revised edition. *SIAM* (2011)
38. Simistira, F., Katsouros, V., Carayannis, G.: Recognition of online handwritten mathematical formulas using probabilistic SVMs and stochastic context free grammars. *Pattern Recognit. Lett.* **53**, 85–92 (2015)
39. Stewart, A., Etessami, K., Yannakakis, M.: Upper Bounds for Newton’s Method on Monotone Polynomial Systems, and P-Time Model Checking of Probabilistic One-Counter Automata. *J. ACM* **62**(4), 30:1–30:33 (2015)
40. Wimmer, S., von Mutius, J.: Verified Certification of Reachability Checking for Timed Automata. In: *TACAS* (1). *Lecture Notes in Computer Science*, vol. 12078, pp. 425–443. Springer (2020)
41. Winkler, T., Gehnen, C., Katoen, J.: Model Checking Temporal Properties of Recursive Probabilistic Programs. In: *FoSSaCS. Lecture Notes in Computer Science*, vol. 13242, pp. 449–469. Springer (2022)
42. Wojtczak, D.: Recursive probabilistic models : efficient analysis and implementation. Ph.D. thesis, University of Edinburgh, UK (2009)
43. Wojtczak, D., Etessami, K.: PReMo : An Analyzer for Probabilistic Recursive Models. In: *TACAS. Lecture Notes in Computer Science*, vol. 4424, pp. 66–71. Springer (2007)
44. Yannakakis, M., Etessami, K.: Checking LTL properties of recursive markov chains. In: *QEST*. pp. 155–165. IEEE Computer Society (2005)

A Full Proofs

A.1 Proof of Lemma 2

Lemma 2 (Existence of inductive upper bounds). *Let \vec{f} be a feasible, clean, and strongly connected PPS. Then the following are equivalent:*

- (1) *The matrix $I - \partial\vec{f}(\mu\vec{f})$ is non-singular.*
- (2) *The spectral radius of $\partial\vec{f}(\mu\vec{f})$ satisfies $\rho(\partial\vec{f}(\mu\vec{f})) < 1$.*
- (3) *There exists $\vec{0} \prec \vec{u} \prec \vec{\infty}$ s.t. $\vec{f}(\vec{u}) < \vec{u}$ (i.e. \vec{u} is inductive but not a fixpoint).*
- (4) *The matrix $\partial\vec{f}(\mu\vec{f})$ has a unique (normalized) eigenvector $\vec{v} \succ \vec{0}$ and there exist numbers $\delta_{max} > 0$ and $\varepsilon > 0$ s.t.*

$$\vec{f}(\mu\vec{f} + \delta \cdot \vec{v}) \prec \mu\vec{f} + \delta \cdot \vec{v}$$

holds for all $0 < \delta \leq \delta_{max}$ and vectors $\vec{v} \geq \vec{v}$ with $\|\vec{v} - \vec{v}\|_\infty \leq \varepsilon$.

We now explain the proof of Lemma 2. The proof heavily relies on a *linear approximation* of \vec{f} around the lfp $\mu\vec{f}$. Intuitively, this is where the Jacobi matrix $\partial\vec{f}(\mu\vec{f})$ comes into play. This is formalized via Taylor's familiar theorem.

Lemma 4 (Taylor's Theorem; cf. [12, Lem. 2.3]). *Let \vec{f} be a feasible PPS. Then for all vectors $\vec{u} \geq \vec{0}$, we have*

$$\vec{f}(\mu\vec{f} + \vec{u}) = \mu\vec{f} + \partial\vec{f}(\mu\vec{f})\vec{u} + R_{\vec{u}}\vec{u}$$

where $R_{\vec{u}}$ is a matrix that depends on \vec{u} such that $\lim_{\vec{u} \rightarrow \vec{0}} R_{\vec{u}} = 0$. More specifically, it holds that $\vec{0} \leq R_{\vec{u}}\vec{u} \leq (\partial\vec{f}(\mu\vec{f} + \vec{u}) - \partial\vec{f}(\mu\vec{f}))\vec{u}$.

Proof (Proof of Lemma 2). “(1) \implies (2)”: By Theorem 2 we have $\rho(\partial\vec{f}(\mu\vec{f})) \leq 1$. Towards contradiction assume that $\rho(\partial\vec{f}(\mu\vec{f})) = 1$. By the Perron-Frobenius Theorem, 1 is an eigenvalue of $\partial\vec{f}(\mu\vec{f})$, which means that there exists $\vec{u} \neq \vec{0}$ such that $\partial\vec{f}(\mu\vec{f})\vec{u} = \vec{u}$. This \vec{u} is in the kernel of $I - \partial\vec{f}(\mu\vec{f})$, which contradicts the assumption that $I - \partial\vec{f}(\mu\vec{f})$ is non-singular.

“(2) \implies (1)”: It is a well-known result that for an arbitrary real matrix M the series $\sum_{k=0}^{\infty} M^k$ converges iff $\rho(M) < 1$. The limit of the series is the inverse of $I - M$ because

$$(I - M) \sum_{k=0}^{\infty} M^k = \sum_{k=0}^{\infty} M^k - \sum_{k=1}^{\infty} M^k = M^0 = I.$$

“(2) \implies (4)”: Let $\rho(\partial\vec{f}(\mu\vec{f})) =: \lambda < 1$. By the Perron-Frobenius Theorem, the Jacobi matrix $\partial\vec{f}(\mu\vec{f})$ has a unique normalized eigenvector $\vec{v} \succ \vec{0}$ wrt. eigenvalue λ :

$$\partial\vec{f}(\mu\vec{f})\vec{v} = \lambda\vec{v} \prec \vec{v}. \tag{1}$$

Our goal is to define the values ε and δ_{max} whose existence we claimed in Lemma 2(4). Let $c_{min} > 0$ be the smallest component of $(1 - \lambda)\vec{v} \succ \vec{0}$. We define

$$\varepsilon := \frac{c_{min}}{3\|\partial\vec{f}(\mu\vec{f})\|_\infty}, \quad (2)$$

where $\|\partial\vec{f}(\mu\vec{f})\|_\infty = \max_{\|\vec{y}\|_\infty=1} \|\partial\vec{f}(\mu\vec{f})\vec{y}\|_\infty$ is the maximum row sum of $\partial\vec{f}(\mu\vec{f})$. Note that $\|\cdot\|_\infty$ is the operator norm induced by the maximum norm. Then it holds for all $\vec{\varepsilon}$ with $\|\vec{\varepsilon}\|_\infty \leq \varepsilon$ that

$$\|\partial\vec{f}(\mu\vec{f})\vec{\varepsilon}\|_\infty \leq \|\partial\vec{f}(\mu\vec{f})\|_\infty \|\vec{\varepsilon}\|_\infty \leq \|\partial\vec{f}(\mu\vec{f})\|_\infty \frac{c_{min}}{3\|\partial\vec{f}(\mu\vec{f})\|_\infty} = \frac{1}{3}c_{min}. \quad (3)$$

The first inequality in (3) is a property of operator norms (which is straightforward in the case of the maximum norm). Since c_{min} was the smallest component of $(1 - \lambda)\vec{v}$, (3) implies

$$\partial\vec{f}(\mu\vec{f})\vec{\varepsilon} \leq \frac{1}{3}(1 - \lambda)\vec{v}. \quad (4)$$

We now define δ_{max} as follows:

$$\delta_{max} := \sup \{ \delta > 0 \mid \forall \vec{\varepsilon} \geq \vec{0} \text{ s.t. } \|\vec{\varepsilon}\|_\infty \leq \varepsilon: R_{\delta(\vec{v}+\vec{\varepsilon})}(\vec{v}+\vec{\varepsilon}) \leq \frac{1}{2}(1 - \lambda)\vec{v} \}, \quad (5)$$

where $R_{\delta(\vec{v}+\vec{\varepsilon})}$ is the matrix from Lemma 4 which satisfies

$$\vec{f}(\mu\vec{f} + \delta(\vec{v} + \vec{\varepsilon})) = \mu\vec{f} + \delta\partial\vec{f}(\mu\vec{f})(\vec{v} + \vec{\varepsilon}) + \delta R_{\delta(\vec{v}+\vec{\varepsilon})}(\vec{v} + \vec{\varepsilon}).$$

We now argue that $\delta_{max} > 0$. This is not immediately obvious because of the \forall -quantification in (5). Let $\delta > 0$ be arbitrary. Further, let $\vec{\varepsilon} \geq \vec{0}$ be such that $\|\vec{\varepsilon}\|_\infty \leq \varepsilon$. In the following, we write $\vec{\varepsilon}' = (\varepsilon \dots \varepsilon)$. We have

$$\begin{aligned} & R_{\delta(\vec{v}+\vec{\varepsilon})}(\vec{v} + \vec{\varepsilon}) \\ &= \frac{1}{\delta} R_{\delta(\vec{v}+\vec{\varepsilon})} \delta(\vec{v} + \vec{\varepsilon}) \\ &\leq \frac{1}{\delta} (\partial\vec{f}(\mu\vec{f} + \delta(\vec{v} + \vec{\varepsilon})) - \partial\vec{f}(\mu\vec{f}))(\vec{v} + \vec{\varepsilon}) \quad (\text{Lemma 4}) \\ &= (\partial\vec{f}(\mu\vec{f} + \delta(\vec{v} + \vec{\varepsilon})) - \partial\vec{f}(\mu\vec{f}))(\vec{v} + \vec{\varepsilon}) \\ &\leq (\partial\vec{f}(\mu\vec{f} + \delta(\vec{v} + \vec{\varepsilon}')) - \partial\vec{f}(\mu\vec{f}))(\vec{v} + \vec{\varepsilon}') \quad (\text{Jacobi matrix is monotonic}) \\ &=: M_\delta(\vec{v} + \vec{\varepsilon}') \end{aligned}$$

Note that M_δ does not depend on $\vec{\varepsilon}$ and $\lim_{\delta \rightarrow 0} M_\delta = 0$. We can therefore find a specific $\delta^* > 0$ such that $M_{\delta^*}(\vec{v} + \vec{\varepsilon}') \leq \frac{1}{2}(1 - \lambda)\vec{v}$. On the other hand, we have just

shown for all $\vec{\varepsilon} \geq \vec{0}$ with $\|\vec{\varepsilon}\|_\infty \leq \varepsilon$ and all $\delta > 0$ that $R_{\delta(\vec{v}+\vec{\varepsilon})}(\vec{v}+\vec{\varepsilon}) \leq M_\delta(\vec{v}+\vec{\varepsilon})$. So we have in particular for all $\vec{\varepsilon} \geq \vec{0}$ with $\|\vec{\varepsilon}\|_\infty \leq \varepsilon$ that

$$R_{\delta^*(\vec{v}+\vec{\varepsilon})}(\vec{v}+\vec{\varepsilon}) \leq M_{\delta^*}(\vec{v}+\vec{\varepsilon}) \leq \frac{1}{2}(1-\lambda)\vec{v}.$$

Hence $\delta_{max} \geq \delta^* > 0$.

Finally, let $0 < \delta \leq \delta_{max}$ and $\vec{v} \geq \vec{v}$ with $\|\vec{v} - \vec{v}\|_\infty \leq \varepsilon$, i.e., $\vec{v} = \vec{v} + \vec{\varepsilon}$ for some $\vec{\varepsilon} \geq \vec{0}$ with $\|\vec{\varepsilon}\|_\infty \leq \varepsilon$. Then

$$\begin{aligned} & \vec{f}(\mu\vec{f} + \delta(\vec{v} + \vec{\varepsilon})) \\ &= \mu\vec{f} + \delta\partial\vec{f}(\mu\vec{f})(\vec{v} + \vec{\varepsilon}) + \delta R_{\delta(\vec{v}+\vec{\varepsilon})}(\vec{v} + \vec{\varepsilon}) \\ & \quad \text{(by Taylor's Theorem (Lemma 4))} \\ &= \mu\vec{f} + \delta\lambda\vec{v} + \delta\partial\vec{f}(\mu\vec{f})\vec{\varepsilon} + \delta R_{\delta(\vec{v}+\vec{\varepsilon})}(\vec{v} + \vec{\varepsilon}) \quad \text{(by (1))} \\ &\leq \mu\vec{f} + \delta\lambda\vec{v} + \delta\frac{1}{3}(1-\lambda)\vec{v} + \delta R_{\delta(\vec{v}+\vec{\varepsilon})}(\vec{v} + \vec{\varepsilon}) \quad \text{(by (4))} \\ &\leq \mu\vec{f} + \delta\lambda\vec{v} + \delta\frac{1}{3}(1-\lambda)\vec{v} + \delta\frac{1}{2}(1-\lambda)\vec{v} \quad \text{(by (5))} \\ &\prec \mu\vec{f} + \delta\lambda\vec{v} + \delta\frac{1}{2}(1-\lambda)\vec{v} + \delta\frac{1}{2}(1-\lambda)\vec{v} \quad \text{(because } \delta(1-\lambda)\vec{v} \succ \vec{0}\text{)} \\ &= \mu\vec{f} + \delta\vec{v} \quad \text{(simplification)} \\ &\leq \mu\vec{f} + \delta(\vec{v} + \vec{\varepsilon}) \quad \text{(because } \vec{\varepsilon} \geq \vec{0}\text{)} \end{aligned}$$

“(4) \implies (3)”: Trivial.

“(3) \implies (2)”: By (3) there exists \vec{u} such that $\vec{f}(\vec{u}) < \vec{u}$. By Lemma 1 this implies that $\mu\vec{f} < \vec{u}$, so we can write $\vec{u} = \mu\vec{f} + \vec{v}$ for some $\vec{v} > \vec{0}$.

Using Taylor's Theorem (Lemma 4), it follows that

$$\vec{f}(\mu\vec{f} + \vec{v}) = \mu\vec{f} + \partial\vec{f}(\mu\vec{f})\vec{v} + R_{\vec{v}}\vec{v} < \mu\vec{f} + \vec{v}. \quad (6)$$

Using that $R_{\vec{v}}\vec{v} \geq \vec{0}$, (6) implies that

$$\partial\vec{f}(\mu\vec{f})\vec{v} < \vec{v}. \quad (7)$$

The claim now follows by applying the following lemma to the matrix $\partial\vec{f}(\mu\vec{f})$ and the vector \vec{v} :

Lemma 5. *Let $M \geq 0$ be an irreducible $n \times n$ -matrix. If there exists $\vec{u} > \vec{0}$ such that $M\vec{u} < \vec{u}$, then $\vec{u} \succ \vec{0}$, $M^n\vec{u} \prec \vec{u}$ and $\rho(M) < 1$.*

Proof. First observe that since multiplication by M is monotone we have for all $0 \leq k_1 \leq k_2$ that

$$\vec{0} \leq M^{k_2}\vec{u} \leq M^{k_1}\vec{u} \leq \vec{u}.$$

We first show that $\vec{u} \succ \vec{0}$, which is essentially [12, Lemma 5.3]. Since $\vec{u} > \vec{0}$, there must be $1 \leq i \leq n$ such that $\vec{u}_i > 0$. Now let $1 \leq j \leq n$ be arbitrary. Since

M is irreducible there exists $0 \leq k < n$ such that $M_{j,i}^k > 0$. This implies that $(M^k \vec{u})_j > 0$. By monotonicity, $\vec{u} \geq M^k \vec{u}$, and thus $\vec{u}_j \geq (M^k \vec{u})_j > 0$. Since j was arbitrary, $\vec{u} \succ \vec{0}$.

Next we show $M^n \vec{u} \prec \vec{u}$. Since $M \vec{u} < \vec{u}$ holds by assumption, there exists $1 \leq i \leq n$ such that $(M \vec{u})_i < \vec{u}_i$. Let $1 \leq j \leq n$ be a arbitrary. Since M is irreducible, there exists $0 \leq k < n$ such that $(M^k)_{j,i} > 0$. We now show that $(M^n \vec{u})_j < u_j$ which implies that $M^n \vec{u} \prec \vec{u}$ as j was chosen arbitrarily:

$$\begin{aligned}
 & (M^n \vec{u})_j \\
 & \leq (M^k M \vec{u})_j && \text{(by monotonicity, and because } k+1 \leq n) \\
 & = (M^k)_{j,i} (M \vec{u})_i + \sum_{l \neq i} (M^k)_{j,l} (M \vec{u})_l && \text{(Def. matrix-vector product)} \\
 & < (M^k)_{j,i} \vec{u}_i + \sum_{l \neq i} (M^k)_{j,l} (M \vec{u})_l && \text{(because } (M \vec{u})_i < \vec{u}_i \text{ and } (M^k)_{j,i} > 0) \\
 & \leq (M^k)_{j,i} \vec{u}_i + \sum_{l \neq i} (M^k)_{j,l} \vec{u}_l && \text{(because } (M \vec{u})_l \leq \vec{u}_l) \\
 & = (M^k \vec{u})_j \leq \vec{u}_j
 \end{aligned}$$

It remains to show that $\rho(M) < 1$. We do this by showing that the powers of M (i.e., the sequence $(M^k)_{k \geq 0}$) converge to the zero matrix. Since $M^n \vec{u} \prec \vec{u}$, we can choose $c < 1$ such that $M^n \vec{u} \leq c \vec{u}$. Then for all $m \geq 1$ it holds that $M^{nm} \vec{u} \leq c^m \vec{u}$, so we have

$$\lim_{k \rightarrow \infty} M^k \vec{u} = \vec{0}.$$

Recall from above that we already know $\vec{u} \succ \vec{0}$. Thus $\lim_{k \rightarrow \infty} M^k \vec{u} = \vec{0}$ means that a positive linear combination of the entries of each individual row of M^k converges to zero, i.e., for all $1 \leq i \leq n$ we have $\lim_{k \rightarrow \infty} \sum_j M_{i,j}^k \vec{u}_j = 0$, and thus for all $1 \leq j \leq n$, $\lim_{k \rightarrow \infty} M_{i,j}^k = 0$. Thus $\lim_{k \rightarrow \infty} M^k = 0$, which completes the proof. \square

A.2 Proof of Theorem 3

Theorem 3. *Algorithm 1 is correct: when invoked with a strongly connected clean PPS \vec{f} and $\varepsilon > 0$, then (if it terminates) it outputs a pair (\vec{l}, \vec{u}) s.t. $\vec{l} \leq \mu \vec{f}$, $\vec{f}(\vec{u}) \leq \vec{u}$ (and thus $\mu \vec{f} \leq \vec{u}$), and $\|\vec{l} - \vec{u}\|_\infty \leq \varepsilon$. Moreover, if \vec{f} is feasible and $I - \partial \vec{f}(\mu \vec{f})$ is non-singular, then the algorithm terminates.*

Proof. Correctness is obvious, so we only show termination assuming that \vec{f} is feasible and $I - \partial \vec{f}(\mu \vec{f})$ is non-singular. Clearly, the algorithm terminates iff it eventually finds a \vec{u} in line 8 which is inductive.

Assume towards contradiction that the algorithm never terminates, i.e., it never finds an inductive \vec{u} . For all $i \geq 1$ let $\vec{l}_i, \vec{v}_i, \tau_i$ be the values of the variables \vec{l}, \vec{v} and τ at the i th time the inner loop at line 7 is reached (note that we then have $N = i - 1$). Clearly, $\lim_{i \rightarrow \infty} \tau_i = 0$. By the contract satisfied by

improveLowerBound, we have $\lim_{i \rightarrow \infty} \partial \vec{f}(\vec{l}_i) = \partial \vec{f}(\mu \vec{f})$. Since the eigenvectors of $\partial \vec{f}(\mu \vec{f})$ depend continuously on those of the matrices $\partial \vec{f}(\vec{l}_i)$, and because of the contract satisfied by **approxEigvec**, the sequence $\vec{v}_1, \vec{v}_2, \dots$ converges to the true unique normalized Perron-Frobenius eigenvector \vec{v}_{true} of $\partial \vec{f}(\mu \vec{f})$.

We now apply condition (4) of Lemma 2. The condition ensures that the cone

$$C(\mu \vec{f}, \vec{v}_{true}, \varepsilon', \delta_{max}) = \{ \mu \vec{f} + \delta \vec{v} \mid 0 \leq \delta \leq \delta_{max}, \|\vec{v} - \vec{v}_{true}\|_\infty \leq \varepsilon' \}$$

which is located at $\mu \vec{f}$, points in direction \vec{v}_{true} and has radius ε' and length δ_{max} contains only inductive points. For the sake of illustration suppose that the algorithm already knows δ_{max} and computes $\vec{u}_i = \vec{l}_i + \delta \vec{v}_i$ for some $0 < \delta < \delta_{max}$ instead of executing the loop starting at line 7. But then the sequence $(\vec{u}_i)_{i \geq 1}$ converges to $\mu \vec{f} + \delta \vec{v}_{true}$, which is a point that lies *inside the interior* of C , so there must be some $i \geq 1$ such that $\vec{u}_i \in C$, i.e., \vec{u}_i is inductive.

The remaining difficulty is that δ_{max} is of course unknown in practice. We handle this using the inner loop that starts at line 7. Eventually, the variable N is sufficiently large such that $d^k \varepsilon < \delta_{max}$ for some $k \leq N$. Termination then follows by applying the argument in the previous paragraph to $\delta = d^k \varepsilon$. \square

A.3 Proof of Lemma 3

Lemma 3. *Let $M \geq 0$ be irreducible. Then power iteration applied to $M + I$ and any $\vec{v}_0 > \vec{0}$ converges to the Perron-Frobenius eigenvector $\vec{v} \succ \vec{0}$ of M .*

Proof. Consider the following conditions for an irreducible matrix $M \geq 0$ and a vector $M\vec{v}_0$ with $M\vec{v}_0 \neq \vec{0}$:

1. M has a unique dominant eigenvalue $|\lambda_1| > |\lambda_2| \geq \dots \geq |\lambda_n|$.
2. λ_1 is *semisimple*, i.e., its algebraic multiplicity⁶ equals its geometric multiplicity⁷.
3. \vec{v}_0 is not orthogonal to the eigenspace $\{\vec{v} \mid M\vec{v} = \lambda_1 \vec{v}\}$.

It is known that if all these conditions are satisfied, then the power iteration sequence $(\vec{v}_i)_{i \in \mathbb{N}}$ converges to a (normalized) eigenvector \vec{v} with eigenvalue λ_1 (e.g. [37, Theorem 4.1]).

We now show that these conditions are satisfied for the irreducible matrix $M + I \geq 0$ and every initial vector $\vec{v}_0 > \vec{0}$. The eigenvectors of M and $M + I$ are exactly the same but the eigenvalues are all shifted by $+1$. Indeed, if \vec{v} is some eigenvector of M with eigenvalue λ , then $(M + I)\vec{v} = \lambda \vec{v} + \vec{v} = (\lambda + 1)\vec{v}$. However, unlike M , the matrix $M + I$ always has period 1, and so it has a unique dominant eigenvalue λ_1 by Theorem 1(2). Therefore the first of the above three conditions is satisfied by the matrix $M + I$.

⁶ The algebraic multiplicity is the multiplicity of a given eigenvalue as a root of the characteristic polynomial.

⁷ The geometric multiplicity is the dimension of the eigenspace associated with a particular eigenvalue.

Next, by Theorem 1(1) it holds that the geometric multiplicity of λ_1 is 1. As the algebraic multiplicity is bounded by the geometric multiplicity, it must also be 1 and thus the matrix $M + I$ satisfies the second condition as well.

Finally, the third condition is satisfied for any $\vec{v}_0 > \vec{0}$ because the scalar product $\vec{v}_0 \cdot \vec{v}$ is non-zero (either strictly positive or strictly negative) for all non-zero eigenvectors \vec{v} of λ_1 by Theorem 1(1). \square

A.4 Proof of Proposition 1

Proposition 1 (Basic Certificates for pPDA). *A basic certificate for $\Delta = (Q, \Gamma, P)$ is a rational inductive upper bound $\vec{u} \in \mathbb{Q}_{\geq 0}^{Q \times \Gamma \times Q}$ on the lfp of the return probabilities system \vec{f}_Δ (see Thm. 4). They have the following properties:*

- (Existence) $\forall \varepsilon > 0$ there exists a basic certificate \vec{u} with $\|\mu \vec{f}_\Delta - \vec{u}\|_\infty \leq \varepsilon$ if all maximal irreducible submatrices M of $\partial \vec{f}_\Delta(\mu \vec{f}_\Delta)$ satisfy $\rho(M) < 1$.
- (Complexity) Let β be the maximum number of bits used to encode any of the numerators and denominators of the fractions occurring in $\vec{u} \in \mathbb{Q}_{\geq 0}^{Q \times \Gamma \times Q}$. Then checking $\vec{f}_\Delta(\vec{u}) \leq \vec{u}$, i.e., whether \vec{u} is basic certificate for Δ , can be done in time polynomial in β and the size of Δ .

Proof. This proof closely follows the general idea of decomposed analysis of PPS [16].

We first address existence. Note that \vec{f}_Δ is guaranteed to be feasible, in fact $\vec{0} \leq \mu \vec{f}_\Delta \leq \vec{1}$. For all qZr with $(\mu \vec{f}_\Delta)_{qZr} = 0$ we set $\vec{u}_{qZr} = 0$. By removing these variables from \vec{f}_Δ we obtain the clean PPS \vec{f}_Δ with $\vec{0} \prec \mu \vec{f}_\Delta$.

Now consider the decomposition of \vec{f}_Δ into the subsystems induced by the strongly connected components of the graph $G_{\vec{f}_\Delta} : \vec{f}_\Delta^1, \dots, \vec{f}_\Delta^m$. Note that in these subsystems, some variables might only appear on the right hand sides but not on the left (e.g. $x_1 = 0.5x_1 + 0.5x_2$, $x_2 = 0.5x_1 + 0.5x_3$). Since $\mu \vec{f}_\Delta \succ \vec{0}$, there is a 1 - 1 correspondence of these subsystems and the maximal irreducible submatrices M_i of $\partial \vec{f}_\Delta(\mu \vec{f}_\Delta)$. More specifically, $M_i = \partial \vec{f}_\Delta^i(\mu \vec{f}_\Delta)$ ⁸. By assumption, $\rho(M_i) < 1$ ⁹.

Now assume w.l.o.g. that \vec{f}_Δ^1 is a bottom SCC (i.e., in the dependency graph $G_{\vec{f}_\Delta}$ there is no path from the variables in \vec{f}_Δ^1 to any variable not in \vec{f}_Δ^1). Then \vec{f}_Δ^1 is a strongly connected PPS with $\partial \vec{f}_\Delta^1(\mu \vec{f}_\Delta) = \partial \vec{f}_\Delta^1(\mu \vec{f}_\Delta^1)$ and we can apply Lemma 2(4) to obtain a rational \vec{u}^1 with $\vec{f}_\Delta^1(\vec{u}^1) \leq \vec{u}^1$ and $\|\mu \vec{f}_\Delta^1 - \vec{u}^1\|_\infty \leq \varepsilon$ (in fact, we can do this for any $\varepsilon > 0$).

Suppose we have done the above for all bottom SCCs and now start traversing the DAG of SCCs bottom-up, i.e., in reverse topological order. Let \vec{u} be the

⁸ The Jacobi matrix of a sub-PPS with $n' < n$ equations is an $n' \times n'$ matrix where all variables that occur only on the right hand sides are considered constants.

⁹ The spectral radius of the zero matrix is zero.

bound we have constructed to far (i.e., \vec{u} contains \vec{u}^1 and the bounds from the other bottom SCC as subvectors and is zero elsewhere). Note that we can always make \vec{u} *smaller* while retaining the inductivity property. W.l.o.g. suppose that subsystem \vec{f}_Δ^2 is one of the first non-bottom SCCs in the reverse topological order. The idea is now to modify \vec{f}_Δ^2 to a strongly connected PPS \vec{f}_u^2 by replacing all variables that occur only in right hand sides by their value in \vec{u} . Clearly, $\lim_{\vec{u} \rightarrow \mu \vec{f}_\Delta} \partial \vec{f}_u^2(\mu \vec{f}_u^2) = \partial \vec{f}_\Delta^2(\mu \vec{f}_\Delta)$. This means we can choose \vec{u} sufficiently close to $\mu \vec{f}_\Delta$ such that the spectral radius of $\partial \vec{f}_u^2(\mu \vec{f}_u^2)$ is strictly smaller than 1. We can then apply Lemma 2(4) to \vec{f}_u^2 to obtain a rational \vec{u}^2 with $\vec{f}_u^2(\vec{u}^2) \leq \vec{u}^2$ to enlarge our current \vec{u} with.

We can repeat this scheme for all finitely many subsystems until we have constructed a rational \vec{u} with $\vec{f}_u^i(\vec{u}) \leq \vec{u}$ for all i . Clearly, this \vec{u} also satisfies $\vec{f}_\Delta(\vec{u}) \leq \vec{u}$. Finally, we may extend \vec{u} by zero entries corresponding to the variables that are assigned zero in the lfp of the (not necessarily clean) \vec{f}_Δ . This yields an inductive upper bound for \vec{f}_Δ . We stress that in order to verify this bound, we neither have to clean \vec{f}_Δ nor do we have to compute the SCCs.

For complexity observe that \vec{f}_Δ is cubic in the size of Δ and that all polynomials in \vec{f}_Δ have degree at most 2. Since multiplication and addition of rational numbers can be done in polynomial time in the number of their bits, evaluating a polynomial of fixed maximum degree can also be done in polynomial time in the size of the polynomial and the number of bits representing the rationals where the polynomial is to be evaluated. Note that this is not true for arbitrary polynomials where exponents are encoded in binary: For instance, evaluating the polynomial x^{2^n} (which can be represented with $\mathcal{O}(n)$ bits) at $x = 2$ yields 2^{2^n} , a number that needs $\mathcal{O}(2^n)$ bits. This means that in order to verify certificates efficiently with exact rational arithmetic, it is important that the polynomials in the PPS do not have very high degrees. Fortunately, this is the case for pPDA.

B Certificates for Expected Rewards

We can certify upper bounds on the expected value of *rewards* collected during the run of a pPDA. To simplify the presentation, in this section we assume w.l.o.g. that $qZ \xrightarrow{p} r\alpha$ with $p > 0$ implies $|\alpha| \in \{0, 2\}$, i.e., all transitions either decrease or increase the stack height by 1. Let $R: Q \rightarrow \mathbb{R}_{\geq 0}$ be a state-based reward function. Consider the following PPS $\vec{f}_{\Delta, R}$ with variables $\{\langle E_{qZr} \rangle \mid qZr \in Q \times \Gamma \times Q\}$:

$$\langle E_{qZr} \rangle = \sum_{qZ \xrightarrow{p} sYX} p \cdot \sum_{t \in Q} [sYt] \cdot [tXr] \cdot K_{qZ, sYX} + \sum_{qZ \xrightarrow{p} r\varepsilon} p \cdot R(r),$$

where $K_{qZ, sYX} = R(r) + \langle E_{sYt} \rangle + \langle E_{tXr} \rangle$. Note that $\vec{f}_{\Delta, R}$ is *linear* but uses the return probabilities which are themselves characterized as the lfp of the *non-linear* system \vec{f}_{Δ}^R from Theorem 4 as coefficients.

Suppose that in the lfp $\mu \vec{f}_{\Delta, R}$, each variable E_{qZr} is assigned the quantity $E_{qZr} \in \mathbb{R}_{\geq 0}$. It follows from the results of [14] that E_{qZr} equals the *expected value* of the following random variable V_R^r under the probability measure \mathbb{P}^{qZ} :

$$V_R^r(q_0\gamma_0, q_1\gamma_1, \dots) = \sum_{i > 0}^{firstHit(r\varepsilon)} R(q_i)$$

where $firstHit(r\varepsilon)$ is the minimum integer k such that $q_k\gamma_k = r\varepsilon$, or 0 if no such k exists. In words, E_{qZr} is the expected reward accumulated on the runs from qZ to $r\varepsilon$, where it is assumed that runs which never reach $r\varepsilon$ contribute zero reward. Consequently, $E(qZ) = \sum_{r \in Q} E_{qZr}$ is the expected reward accumulated on all terminating runs.

Example 7. Setting $R = 1$ we can characterize the *expected runtime* of pPDA. Reconsider Example 4. The equation system for expected runtimes becomes

$$\begin{aligned} \langle E_{qZq} \rangle &= \frac{1}{4}([qZq]^2(1+2\langle E_{qZq} \rangle) + [qZr][rZq](1+\langle E_{qZr} \rangle + \langle E_{rZq} \rangle)) + \frac{1}{2} \\ \langle E_{qZr} \rangle &= \frac{1}{4}([qZq][qZr](1+\langle E_{qZq} \rangle + \langle E_{qZr} \rangle) + [qZr][rZr](1+\langle E_{qZr} \rangle + \langle E_{rZr} \rangle)) + \frac{1}{4} \end{aligned}$$

as well as $\langle E_{rZq} \rangle = 0$ and $\langle E_{rZr} \rangle = 1$. The solution is $\langle E_{qZq} \rangle = 2063/2624 \approx 0.786$ and $\langle E_{qZr} \rangle = 59/82 \approx 0.712$, so the total expected runtime is $E(qZ) \approx 1.506$. \triangle

C Benchmark Programs

| | | |
|---|---|--|
| <pre> void f() { if flip(p) { f(); f(); } } # main block { f(); } </pre> | <pre> void f() { if flip(1//2) { f(); f(); f(); } } # main block { f(); } </pre> | <pre> void offspring() { while flip(2//5) { offspring(); while flip(3//5) { offspring(); } } } # main block { offspring(); } </pre> |
| (a) rw-p | (b) golden | (c) geom-offspring |

| | | |
|---|---|--|
| <pre> void gen_operator() { uniform(4); } void gen_expression() { prob { 4//10: uniform(10); 3//10: { } 3//10: { gen_operator(); gen_expression(); gen_expression(); } } } void gen_function() { gen_operator(); gen_expression(); gen_expression(); } # main block { gen_function(); } </pre> | <pre> void young() { int y = uniform(4); while(y > 0) { young(); y = y-1; } } void elder() { int y = uniform(2); while(y > 0) { young(); y = y-1; } } void gen_function() { int e = uniform(5); while(e > 0) { elder(); e = e-1; } } # main block { young(); } </pre> | <pre> bool f() { prob { 1//2: return flip(1//2); 1//2: if f() { return f(); } else { return false; } } } # main blocok { bool res1 = f(); ... bool resN = f(); } </pre> |
| (d) gun-fun | (e) virus | (f) sequentialN |

```

int f(int n, int m) {
  prob {
    (n+1)/(n+2) : {
      f((n + 1) % m, m);
      f((n + 1) % m, m);
      return 0;
    }
    1/(n+2) :
      return 0;
  }
}

# main block
{
  f(0, N);
}

```

(a) escapeN

```

void f(int n) {
  while(n > 0) {
    prob {
      2/3: f(n-1);
      1/3: f((n+1) % N);
    }
    n = n-1;
  }
}

# main block
{
  f(1);
}

```

(b) modN

D Z3 vs CVC5

Table 3: Comparison of the SMT-approach (see §*Baselines* in Section 5) using z3 and cvc5 on SCFG given as explicit PPS (right), and on programs automatically translated to pPDA (left).

| benchmark | $cert_{z3}$ | t_{z3} | $cert_{cvc5}$ | t_{cvc5} |
|----------------|-------------|----------|---------------|------------|
| rw-0.499 | ✓ | 11 | ✓ | 92 |
| rw-0.500 | ✓ | 10 | ✓ | 87 |
| rw-0.501 | ✓ | 12 | ✓ | 104 |
| geom-offspring | ✓ | 16 | ✓ | 4687 |
| golden | ✓ | 14 | ✓ | 1097 |
| and-or | ✓ | 15260 | ✗ | TO |
| gen-fun | ✓ | 141 | ✗ | TO |
| virus | ✓ | 139 | ✓ | 163727 |
| escape10 | ✓ | 48 | ✓ | 12031 |
| escape25 | ✓ | 15958 | ✗ | TO |
| escape50 | ✓ | 410 | ✗ | TO |
| escape75 | ✗ | TO | ✗ | TO |
| escape100 | ✓ | 939 | ✗ | TO |
| escape200 | ✗ | TO | ✗ | TO |
| sequential5 | ✓ | 1074 | ✗ | TO |
| sequential7 | ✓ | 12822 | ✗ | TO |
| sequential10 | ✓ | 453718 | ✗ | TO |
| mod5 | ✓ | 34150 | ✗ | TO |
| mod7 | ✓ | 443 | ✗ | TO |
| mod10 | ✓ | 1245 | ✗ | TO |

| benchmark | $cert_{z3}$ | t_{z3} | $cert_{cvc5}$ | t_{cvc5} |
|-----------|-------------|----------|---------------|------------|
| brown | ✗ | TO | ✗ | TO |
| lemonde | ✗ | TO | ✗ | TO |
| negra | ✓ | 37248 | ✓ | 10144 |
| swbd | ✗ | TO | ✗ | Error |
| tiger | ✓ | 17454 | ✓ | 16118 |
| tuebadz | ✓ | 15323 | ✓ | 5534 |
| wsj | ✗ | TO | ✗ | TO |
| random | ✗ | TO | ✗ | TO |