Journal of Computer and System Sciences 81 (2015) 288–310

Contents lists available at ScienceDirect

Journal of Computer and System Sciences

www.elsevier.com/locate/jcss

Runtime analysis of probabilistic programs with unbounded recursion ☆

Tomáš Brázdil [a,1], Stefan Kiefer [b,*,2], Antonín Kučera [a,1], Ivana Hutařová Vařeková [a]

[a] Faculty of Informatics, Masaryk University, Czech Republic
[b] Department of Computer Science, University of Oxford, United Kingdom

ARTICLE INFO

Article history:
Received 21 November 2012
Received in revised form 18 February 2014
Accepted 26 May 2014
Available online 27 June 2014

Keywords:
Probabilistic pushdown automata
Recursive Markov chains
Termination time

ABSTRACT

We study the runtime in probabilistic programs with unbounded recursion. As underlying formal model for such programs we use *probabilistic pushdown automata (pPDAs)* which exactly correspond to recursive Markov chains. We show that every pPDA can be transformed into a stateless pPDA (called "pBPA") whose runtime and further properties are closely related to those of the original pPDA. This result substantially simplifies the analysis of runtime and other pPDA properties. We prove that for every pPDA the probability of performing a long run decreases *exponentially* in the length of the run, if and only if the expected runtime in the pPDA is *finite*. If the expectation is infinite, then the probability decreases "polynomially". We show that these bounds are asymptotically tight. Our tail bounds on the runtime are *generic*, i.e., applicable to any probabilistic program with unbounded recursion.

1. Introduction

We study the termination time in programs with unbounded recursion, which are either randomized or operate on statistically quantified inputs. As underlying formal model for such programs we use *probabilistic pushdown automata (pPDAs)* [4,7,14,15] which are equivalent to recursive Markov chains [17–19]. Since pushdown automata are a standard and well-established model for programs with recursive procedure calls, our abstract results imply *generic* and *tight* tail bounds for termination time, the main performance characteristic of probabilistic recursive programs.

A pPDA consists of a finite set of *control states*, a finite *stack alphabet*, and a finite set of *rules* of the form $pX \overset{x}{\hookrightarrow} q\alpha$, where p, q are control states, X is a stack symbol, α is a finite sequence of stack symbols (possibly empty), and $x \in (0,1]$ is the (rational) probability of the rule. We require that for each pX, the sum of the probabilities of all rules of the form $pX \overset{x}{\hookrightarrow} q\alpha$ is equal to 1. Each pPDA Δ induces an infinite-state Markov chain M_Δ, where the states are configurations

☆ This work has been published without proofs as a preliminary version in the *Proceedings of the 38th International Colloquium on Automata, Languages and Programming (ICALP)*, volume 6756 of LNCS, pages 319–331, 2011 at Springer. The presentation has been improved since, and the general lower tail bound in Theorem 4.1 (3) has been tightened from $\Omega(1/n)$ to $\Omega(1/\sqrt{n})$.

* Corresponding author.
E-mail addresses: brazdil@fi.muni.cz (T. Brázdil), stefan.kiefer@cs.ox.ac.uk (S. Kiefer), kucera@fi.muni.cz (A. Kučera), ivarekova@centrum.cz (I. Hutařová Vařeková).

1 Tomáš Brázdil and Antonín Kučera are supported by the Czech Science Foundation, grant No. P202/12/G061.
2 Stefan Kiefer is supported by a Royal Society University Research Fellowship.

http://dx.doi.org/10.1016/j.jcss.2014.06.005
0022-0000/

```
function And(node)                 function Or(node)
 if node.leaf then                  if node.leaf then
  return node.value                  return node.value
 else                               else
  v := Or(node.left)                 v := And(node.left)
  if v = 0 then                      if v = 1 then
   return 0                           return 1
  else                               else
   return Or(node.right)              return And(node.right)
```

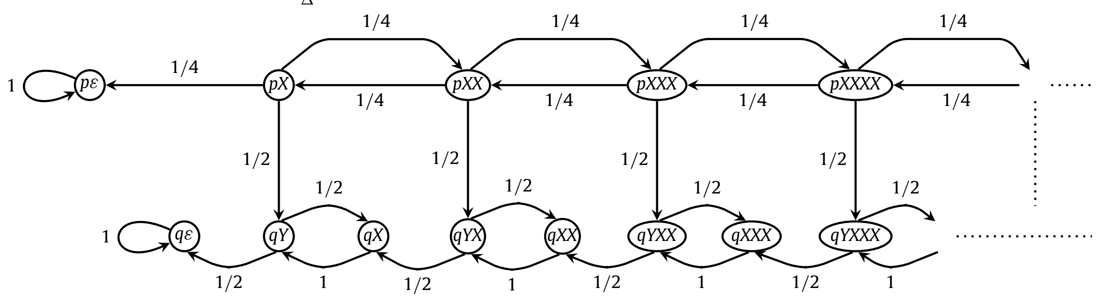
$$qA \overset{1/4}{\hookrightarrow} r_1\varepsilon \qquad qO \overset{1/4}{\hookrightarrow} r_1\varepsilon$$
$$qA \overset{1/4}{\hookrightarrow} r_0\varepsilon \qquad qO \overset{1/4}{\hookrightarrow} r_0\varepsilon$$
$$qA \overset{1/2}{\hookrightarrow} qOA \qquad qO \overset{1/2}{\hookrightarrow} qAO$$
$$r_0A \overset{1}{\hookrightarrow} r_0\varepsilon \qquad r_1O \overset{1}{\hookrightarrow} r_1\varepsilon$$
$$r_1A \overset{1}{\hookrightarrow} qO \qquad r_0O \overset{1}{\hookrightarrow} qA$$

Fig. 1. The program *Tree* and its pPDA model Δ_{Tree}. In Δ_{Tree} the control state q is the default control state, and the control states r_0, r_1 model the return values 0 and 1. The stack symbols A and O represent (invocations of) the procedures And and Or.

of the form $p\alpha$ (p is the current control state and α is the current stack content), and $pX\beta \xrightarrow{x} q\alpha\beta$ is a transition of M_Δ iff $pX \overset{x}{\hookrightarrow} q\alpha$ is a rule of Δ. We also stipulate that $p\varepsilon \xrightarrow{1} p\varepsilon$ for every control state p, where ε denotes the empty stack. For example, consider the pPDA $\hat{\Delta}$ with two control states p, q, two stack symbols X, Y, and the rules

$$pX \overset{1/4}{\hookrightarrow} p\varepsilon, \quad pX \overset{1/4}{\hookrightarrow} pXX, \quad pX \overset{1/2}{\hookrightarrow} qY, \quad pY \overset{1}{\hookrightarrow} pY, \quad qY \overset{1/2}{\hookrightarrow} qX, \quad qY \overset{1/2}{\hookrightarrow} q\varepsilon, \quad qX \overset{1}{\hookrightarrow} qY.$$

The structure of Markov chain $M_{\hat{\Delta}}$ is indicated below.



pPDAs can model programs that use unbounded "stack-like" data structures such as stacks, counters, or even queues. For instance, if the exact ordering of items stored in a queue is irrelevant, the queue can be safely replaced with a stack. Transition probabilities may reflect the random choices of the program (such as "coin flips" in randomized algorithms) or some statistical assumptions about the input data. In particular, pPDAs model *recursive* programs. The global data of such a program are stored in the finite control, and the individual procedures and functions together with their local data correspond to the stack symbols (a function call/return is modeled by pushing/popping the associated stack symbol onto/from the stack). As a simple example, consider the recursive program *Tree* of Fig. 1, which computes the value of an *And/Or-tree*, i.e., a tree such that (i) every node has either zero or two children, (ii) every inner node is either an And-node or an Or-node, and (iii) on any path from the root to a leaf And- and Or-nodes alternate. We further assume that the root is either a leaf or an And-node. *Tree* starts by invoking the function And on the root of a given And/Or-tree. Observe that the program evaluates subtrees only if necessary. Now assume that the input are random And/Or trees following the distribution of a Galton–Watson process: a node of the tree has two children with probability 1/2, and no children with probability 1/2. Furthermore, the conditional probabilities that a childless node evaluates to 0 and 1 are also both equal to 1/2. On inputs with this distribution, the algorithm corresponds to the pPDA Δ_{Tree} of Fig. 1.

We study the *termination time* of runs in a given pPDA Δ. For every pair of control states p, q and every stack symbol X of Δ, let $Run(pXq)$ be the set of all runs (infinite paths) in M_Δ initiated in pX which visit $q\varepsilon$. The termination time is modeled by the random variable $\mathbf{T}_{pX}$, which to every run w assigns either the number of steps needed to reach a configuration with empty stack, or ∞ if there is no such configuration. The conditional expected value $\mathbb{E}[\mathbf{T}_{pX} \mid Run(pXq)]$, denoted just by $E[pXq]$ for short, then corresponds to the average number of steps needed to reach $q\varepsilon$ from pX, computed only for those runs initiated in pX which terminate at $q\varepsilon$. For example, using the results of [14,15,19], one can show that the functions And and Or of the program *Tree* terminate with probability one, and the expected termination times can be computed by solving a system of linear equations. Thus, we obtain the following:

$$\begin{array}{lll} E[qAr_0] = 7.155113 & E[qAr_1] = 7.172218 & \\ E[qOr_0] = 7.172218 & E[qOr_1] = 7.155113 & \\ E[r_0Ar_0] = 1.000000 & E[r_1Ar_0] = 8.172218 & E[r_1Ar_1] = 8.155113 \\ E[r_1Or_1] = 1.000000 & E[r_0Or_1] = 8.172218 & E[r_0Or_0] = 8.155113 \end{array}$$

However, the mere expectation of the termination time does not provide much information about its distribution until we analyze the associated *tail bound*, i.e., the probability that the termination time deviates from its expected value by a given amount. That is, we are interested in bounds for the conditional probability $P(\mathbf{T}_{pX} \geq n \mid Run(pXq))$. (Note this probability makes sense regardless of whether $E[pXq]$ is finite or infinite.) Assuming that the (conditional) expectation and variance of $\mathbf{T}_{pX}$ are finite, one can apply Markov's and Chebyshev's inequalities and thus yield bounds of the form $P(\mathbf{T}_{pX} \geq n \mid Run(pXq)) \leq c/n$ and $P(\mathbf{T}_{pX} \geq n \mid Run(pXq)) \leq c/n^2$, respectively, where c is a constant depending only on the underlying pPDA. However, these bounds are asymptotically always worse than our exponential bound (see below). If $E[pXq]$ is infinite, these inequalities cannot be used at all.

Our contribution The main contributions of this paper are the following:

- We show that every pPDA can be effectively transformed into a *stateless* pPDA (called "pBPA") so that all important quantitative characteristics of runs are preserved. This simple (but fundamental) observation was overlooked in previous works on pPDAs and related models [4,7,14,15,17–19], although it simplifies virtually all of these results. Hence, we can w.l.o.g. concentrate just on the study of pBPAs. Moreover, for the runtime analysis, the transformation yields a pBPA all of whose symbols terminate with probability one, which further simplifies the analysis.
- We provide tail bounds for $\mathbf{T}_{pX}$ which are *asymptotically optimal for every pPDA* and are applicable also in the case when $E[pXq]$ is infinite. More precisely, we show that for every pair of control states p, q and every stack symbol X, there are essentially three possibilities:
  – There is a "small" k such that $P(\mathbf{T}_{pX} \geq n \mid Run(pXq)) = 0$ for all $n \geq k$.
  – $E[pXq]$ is finite and $P(\mathbf{T}_{pX} \geq n \mid Run(pXq))$ decreases exponentially in n.
  – $E[pXq]$ is infinite and $P(\mathbf{T}_{pX} \geq n \mid Run(pXq))$ decreases "polynomially" in n.
  The exact formulation of this result, including the explanation of what is meant by a "polynomial" decrease, is given in Theorem 4.1 (technically, Theorem 4.1 is formulated for pBPAs which terminate with probability one, which is no restriction as explained above). Observe that a direct consequence of the above theorem is that *all* conditional moments $\mathbb{E}[\mathbf{T}^k_{pX} \mid Run(pXq)]$ are simultaneously either finite or infinite. Clearly, if $E[pXq] = \mathbb{E}[\mathbf{T}_{pX} \mid Run(pXq)]$ is infinite, then so is $\mathbb{E}[\mathbf{T}^k_{pX} \mid Run(pXq)]$ for every $k \geq 1$. If $E[pXq]$ is finite, then $P(\mathbf{T}_{pX} \geq n \mid Run(pXq))$ decreases exponentially in n, which means that there exist $n_0 \geq 1$ and $0 < a < 1$ such that for all $n \geq n_0$ we have that $P(\mathbf{T}_{pX} \geq n \mid Run(pXq)) \leq a^n$. From this we easily obtain $\mathbb{E}[\mathbf{T}^k_{pX} \mid Run(pXq)] < \infty$. In particular, if $E[pXq]$ is finite, then so is the conditional variance of $\mathbf{T}_{pX}$.

The characterization given in Theorem 4.1 is effective. In particular, it is decidable in polynomial space whether $E[pXq]$ is finite or infinite by using the results of [14,15,19], and if $E[pXq]$ is finite, we can compute concrete bounds on the probabilities. Our results vastly improve on what was previously known on the termination time $\mathbf{T}_{pX}$. Previous work, in particular [3,15], has focused on computing expectations and variances for a class of random variables on pPDA runs, a class that includes $\mathbf{T}_{pX}$ as prime example. Note that our exponential bound given in Theorem 4.1 depends, like Markov's inequality, only on expectations, which can be efficiently approximated by the methods of [13,15].

An intuitive interpretation of our results is that pPDAs with finite (conditional) expected termination time are well-behaved in the sense that the termination time is exponentially unlikely to deviate from its expectation. Of course, a detailed analysis of a concrete pPDA may lead to better bounds, but these bounds will be *asymptotically equivalent* to our generic bounds. Also note that the conditional expected termination time can be finite even for pPDAs that do not terminate with probability one. Hence, for every $\varepsilon > 0$ we can compute a tight threshold k such that if a given pPDA terminates at all, it terminates after at most k steps with probability $1 - \varepsilon$ (this is useful for interrupting programs that are supposed but not guaranteed to terminate).

Proof techniques The main mathematical tool for establishing our results on runtime is (basic) martingale theory and its tools such as the optional stopping theorem and Azuma's inequality (see Section 4). More precisely, we construct two different martingales corresponding to the cases when the expected termination time is finite resp. infinite. In combination with our reduction to pBPAs this establishes a powerful link between pBPAs, pPDAs, and martingale theory.

Our analysis of termination time in the case when the expected termination time is infinite builds on Perron–Frobenius theory for nonnegative matrices as well as on recent results from [13,19]. We also use some of the observations presented in [7,14,15].

Related work The application of Azuma's inequality in the analysis of particular randomized algorithms is also known as the *method of bounded differences*; see, e.g., [11,25] and the references therein. In contrast, we apply martingale methods not to particular algorithms, but to the pPDA model as a whole.

Analyzing the distribution of termination time is closely related to the analysis of multitype branching processes (MT-BPs) [20]. An MT-BP is very much like a pBPA (see above). The stack symbols in pBPAs correspond to species in MT-BPs. An ε-rule corresponds to the death of an individual, whereas a rule with two or more symbols on the right hand side corresponds to reproduction. Since in MT-BPs the symbols on the right hand side of rules evolve concurrently, termination time in pBPAs does *not* correspond to extinction time in MT-BPs, but to the size of the *total progeny* of an individual, i.e., the number of direct or indirect descendants of an individual. The distribution of the total progeny of a MT-BP has been studied mainly for the case of a single species, see, e.g., [20,26,27] and the references therein, but to the best of our knowledge, no tail bounds for MT-BPs have been given. Hence, Theorem 4.1 can also be seen as a contribution to MT-BP theory.

Stochastic context-free grammars (SCFGs) [24] are also closely related to pBPAs. The termination time in pBPAs corresponds to the number of nodes in a derivation tree of a SCFG, so our analysis of pBPAs immediately applies to SCFGs. Quasi-Birth–Death processes (QBDs) can also be seen as a special case of pPDAs. A QBD is a generalization of a birth–death process studied in queuing theory and applied probability (see, e.g., [2,16,23]). Intuitively, a QBD describes an unbounded queue, using a counter to count the number of jobs in the queue, where the queue can be in one of finitely many distinct "modes". Hence, a (discrete-time) QBD can be equivalently defined by a pPDA with one stack symbol used to emulate the counter. These special pPDAs are also known as *probabilistic one-counter automata (pOCs)* [5,6,16]. Recently, it has been shown in [8] that every pOC induces a martingale apt for studying the properties of both terminating and nonterminating runs in pOCs. However, the paper [8] focuses on approximating (non-)termination probabilities and the expected termination time, and does not study the distribution of the termination time, as we do in this paper. The constructions used in [8] are based on ideas specific to pOCs that are unrelated to the ones presented in this paper.

Previous work on pPDAs and the equivalent model of recursive Markov chains includes [4,7,14,15,17–19]. In this paper we use many of the results presented in these papers, which is explicitly acknowledged at appropriate places.

Organization of the paper We present our results after some preliminaries in Section 2. In Section 3 we show how to transform a given pPDA into an equivalent pBPA, and in Section 4 we design the promised martingales and derive tight tail bounds for the termination time. We conclude in Section 5.

2. Preliminaries

In the rest of this paper, $\mathbb{N}$, $\mathbb{N}_0$, and $\mathbb{R}$ denote the set of positive integers, nonnegative integers, and real numbers, respectively. The tuples of $A_1 \times A_2 \times \cdots \times A_n$ are often written simply as $a_1 a_2 \ldots a_n$. The set of all finite words over a given alphabet Σ is denoted by Σ^*, and the set of all infinite words over Σ is denoted by Σ^ω. We write ε for the empty word. The length of a given $w \in \Sigma^* \cup \Sigma^\omega$ is denoted by $|w|$, where the length of an infinite word is ∞. Given a word (finite or infinite) over Σ, the individual letters of w are denoted by $w(0), w(1), \ldots$ For $X \in \Sigma$ and $w \in \Sigma^*$, we denote by $\#(X)(w)$ the number of occurrences of X in w.

Definition 2.1 (*Markov Chains*). A *Markov chain* is a triple $M = (S, \to, Prob)$ where S is a finite or countably infinite set of *states*, $\to \subseteq S \times S$ is a *transition relation*, and *Prob* is a function which to each transition $(s,t) \in \to$ assigns its probability $Prob((s,t)) > 0$ so that for every $s \in S$ we have $\sum_{s \to t} Prob((s,t)) = 1$. We write $s \xrightarrow{x} t$ to indicate that $s \to t$ and $Prob((s,t)) = x$.

A *path* in M is a finite or infinite word $w \in S^+ \cup S^\omega$ such that $w(i-1) \to w(i)$ for every $1 \le i < |w|$. For a state s, we use $FPath(s)$ to denote the set of all finite paths initiated in s. A *run* in M is an infinite path in M. We denote by $Run[M]$ the set of all runs in M. The set of all runs that start with a given finite path w is denoted by $Run[M](w)$. When M is understood, we write just Run and $Run(w)$ instead of $Run[M]$ and $Run[M](w)$, respectively. For a given run w and $i \in \mathbb{N}_0$, we use w_i to denote the run $w(i), w(i{+}1), w(i{+}2), \ldots$. Given $s \in S$ and $A \subseteq S$, we say *A is reachable from s* if there is a run w such that $w(0) = s$ and $w(i) \in A$ for some $i \ge 0$.

To every $s \in S$ we associate the probability space $(Run(s), \mathcal{F}, \mathcal{P})$ where $\mathcal{F}$ is the σ-field generated by all *basic cylinders* $Run(w)$ where w is a finite path starting with s, and $\mathcal{P} : \mathcal{F} \to [0,1]$ is the unique probability measure such that $\mathcal{P}(Run(w)) = \prod_{i=1}^{|w|-1} x_i$ where $w(i-1) \xrightarrow{x_i} w(i)$ for every $1 \le i < |w|$. If $|w| = 1$, we put $\mathcal{P}(Run(w)) = 1$. Note that only certain subsets of $Run(s)$ are $\mathcal{P}$-measurable, but in this paper we only deal with "safe" subsets that are guaranteed to be in $\mathcal{F}$. We remark that, technically, $\mathcal{F}$ and $\mathcal{P}$ depend on the start state s, but we suppress this dependence, as long as s is clear from the context.

Definition 2.2 (*Probabilistic PDA*). A *probabilistic pushdown automaton (pPDA)* is a tuple $\Delta = (Q, \Gamma, \hookrightarrow, Prob)$ where Q is a finite set of *control states*, Γ is a finite *stack alphabet*, $\hookrightarrow \subseteq (Q \times \Gamma) \times (Q \times \Gamma^{\le 2})$ is a *transition relation* (where $\Gamma^{\le 2} = \{\alpha \in \Gamma^* \mid |\alpha| \le 2\}$), and *Prob* is a function which to each transition $(pX, q\alpha) \in \hookrightarrow$ assigns its probability $Prob((pX, q\alpha)) > 0$

so that for all $p \in Q$ and $X \in \Gamma$ we have that $\sum_{pX \hookrightarrow q\alpha} Prob((pX, q\alpha)) = 1$. We write $pX \overset{x}{\hookrightarrow} q\alpha$ to indicate that $pX \hookrightarrow q\alpha$ and $Prob((pX, q\alpha)) = x$.

Elements of $Q \times \Gamma^*$ are called *configurations* of Δ. A pPDA with just one control state is called pBPA.[3] In what follows, configurations of pBPAs are usually written without the (only) control state p (i.e., we write just α instead of $p\alpha$). We define the *size*[4] of a pPDA Δ as $|\Delta| = |Q| + |\Gamma| + |{\hookrightarrow}| + |Prob|$, where $|Prob|$ is the sum of sizes of binary representations of values taken by *Prob*. To Δ we associate the Markov chain M_Δ with $Q \times \Gamma^*$ as the set of states and transitions defined as follows:

- $p\varepsilon \overset{1}{\to} p\varepsilon$ for each $p \in Q$;
- $pX\beta \overset{x}{\to} q\alpha\beta$ is a transition of M_Δ iff $pX \overset{x}{\hookrightarrow} q\alpha$ is a transition of Δ.

For all $pXq \in Q \times \Gamma \times Q$ and $rY \in Q \times \Gamma$, we define

- $Run(pXq) = \{w \in Run(pX) \mid w(i) = q\varepsilon \text{ for some } i \in \mathbb{N}\}$
- $Run(rY{\uparrow}) = Run(rY) \setminus \bigcup_{s \in Q} Run(rYs)$.

Further, we put $[pXq] = \mathcal{P}(Run(pXq))$ and $[pX{\uparrow}] = \mathcal{P}(Run(pX{\uparrow}))$. If Δ is a pBPA, we write $[X]$ and $[X{\uparrow}]$ instead of $[pXp]$ and $[pX{\uparrow}]$, where p is the only control state of Δ.

Let $p\alpha \in Q \times \Gamma^*$. We denote by $\mathbf{T}_{p\alpha}$ a random variable over $Run(p\alpha)$ where $\mathbf{T}_{p\alpha}(w)$ is either the least $n \in \mathbb{N}_0$ such that $w(n) = q\varepsilon$ for some $q \in Q$, or ∞ if there is no such n. Intuitively, $\mathbf{T}_{p\alpha}(w)$ is the number of steps ("the time") in which the run w initiated in $p\alpha$ terminates. We write $E[p\alpha] := \mathbb{E}[\mathbf{T}_{p\alpha}]$ for the expected termination time (usually omitting the control state p for pBPAs).

3. Transforming pPDAs into pBPAs

In this section we show how to transform a given pPDA Δ into an "equivalent" pBPA $\Delta_\bullet$ such that all stack symbols of $\Delta_\bullet$ terminate either with probability 0 or 1. This transformation preserves virtually all interesting properties and it is to some extent effective. However, the transition probabilities in $\Delta_\bullet$ may take irrational values.

Let $\Delta = (Q, \Gamma, \hookrightarrow, Prob)$ be a pPDA. The construction of the associated pBPA $\Delta_\bullet$ is a relatively straightforward modification of the standard method for transforming a PDA into an equivalent context-free grammar (see, e.g., [21]), but has so far been overlooked in the existing literature on probabilistic PDA. The stack alphabet $\Gamma_\bullet$ of $\Delta_\bullet$ is defined as follows: For every $pX \in Q \times \Gamma$ such that $[pX{\uparrow}] > 0$ we add a fresh stack symbol $\langle pX{\uparrow}\rangle$, and for every $pXq \in Q \times \Gamma \times Q$ such that $[pXq] > 0$ we add a fresh stack symbol $\langle pXq\rangle$. Note that $\Gamma_\bullet$ is effectively constructible in polynomial space by applying the results of [14,19]. Now we construct the rules $\hookrightarrow_\bullet$ of $\Delta_\bullet$. For all $\langle pXq\rangle \in \Gamma_\bullet$ we have the following rules:

- if $pX \overset{x}{\hookrightarrow} rYZ$ in Δ, then for all $s \in Q$ such that $y = x \cdot [rYs] \cdot [sZq] > 0$ we put $\langle pXq\rangle \overset{y/[pXq]}{\hookrightarrow_\bullet} \langle rYs\rangle\langle sZq\rangle$;
- if $pX \overset{x}{\hookrightarrow} rY$ in Δ, where $y = x \cdot [rYq] > 0$, we put $\langle pXq\rangle \overset{y/[pXq]}{\hookrightarrow_\bullet} \langle rYq\rangle$;
- if $pX \overset{x}{\hookrightarrow} q\varepsilon$ in Δ, we put $\langle pXq\rangle \overset{x/[pXq]}{\hookrightarrow_\bullet} \varepsilon$.

For all $\langle pX{\uparrow}\rangle \in \Gamma_\bullet$ we have the following rules:

- if $pX \overset{x}{\hookrightarrow} rYZ$ in Δ, then for every $s \in Q$ where $y = x \cdot [rYs] \cdot [sZ{\uparrow}] > 0$ we add $\langle pX{\uparrow}\rangle \overset{y/[pX\uparrow]}{\hookrightarrow_\bullet} \langle rYs\rangle\langle sZ{\uparrow}\rangle$;
- for all $qY \in Q \times \Gamma$ where $y = [qY{\uparrow}] \cdot \sum_{pX \overset{z}{\hookrightarrow} qY\beta} z > 0$, we add $\langle pX{\uparrow}\rangle \overset{y/[pX\uparrow]}{\hookrightarrow_\bullet} \langle qY{\uparrow}\rangle$.

Observe that all symbols of the form $\langle pX{\uparrow}\rangle$ terminate with probability 0, and we show that all symbols of the form $\langle pXq\rangle$ terminate with probability 1. Also note that the transition probabilities of $\Delta_\bullet$ may take irrational values. Still, the construction of $\Delta_\bullet$ is to some extent "effective" due to the following proposition:

Proposition 3.1. *(See [14,19].) Let $\Delta = (Q, \Gamma, \hookrightarrow, Prob)$ be a pPDA. Let $pXq \in Q \times \Gamma \times Q$. There is a formula $\Phi(x)$ of ExTh($\mathbb{R}$) (the existential theory of the reals) with one free variable x such that the length of $\Phi(x)$ is polynomial in $|\Delta|$ and $\Phi(x/r)$ is valid iff $r = [pXq]$.*

[3] The "BPA" acronym stands for "Basic Process Algebra" and is used mainly for historical reasons. pBPAs are closely related to stochastic context-free grammars and are also called *1-exit recursive Markov chains* (see, e.g., [19]).

[4] When a pPDA Δ is used as an input of some algorithm, we assume that all transition probabilities of Δ are *rational* and represented as fractions of binary integers.

Using Proposition 3.1, one can compute formulae of $ExTh(\mathbb{R})$ that "encode" transition probabilities of $\Delta_\bullet$. Moreover, these probabilities can be effectively approximated up to an arbitrarily small positive error by employing either the decision procedure for $ExTh(\mathbb{R})$ [9] or by using Newton's method [12,13,22].

Example 3.2. Consider a pPDA Δ with two control states p, q, one stack symbol X, and the following transitions, where $a > 1/2$:

$$pX \overset{a}{\hookrightarrow} qXX, \qquad pX \overset{1-a}{\hookrightarrow} q\varepsilon, \qquad qX \overset{a}{\hookrightarrow} pXX, \qquad qX \overset{1-a}{\hookrightarrow} p\varepsilon,$$

Clearly, $[pXp] = [qXq] = 0$. Using the results of [14], one can easily verify that $[pXq] = [qXp] = (1-a)/a$. Hence, $[pX\uparrow] = [qX\uparrow] = (2a-1)/a$. Consequently, the stack symbols of $\Delta_\bullet$ are $\langle pXq\rangle$, $\langle qXp\rangle$, $\langle pX\uparrow\rangle$, and $\langle qX\uparrow\rangle$, and the transitions of $\Delta_\bullet$ are the following:

$$\begin{array}{ll}
\langle pXq\rangle \overset{1-a}{\hookrightarrow}_\bullet \langle qXp\rangle\langle pXq\rangle & \langle qXp\rangle \overset{1-a}{\hookrightarrow}_\bullet \langle pXq\rangle\langle qXp\rangle \\
\langle pXq\rangle \overset{a}{\hookrightarrow}_\bullet \varepsilon & \langle qXp\rangle \overset{a}{\hookrightarrow}_\bullet \varepsilon \\
\langle pX\uparrow\rangle \overset{1-a}{\hookrightarrow}_\bullet \langle qXp\rangle\langle pX\uparrow\rangle & \langle qX\uparrow\rangle \overset{1-a}{\hookrightarrow}_\bullet \langle pXq\rangle\langle qX\uparrow\rangle \\
\langle pX\uparrow\rangle \overset{a}{\hookrightarrow}_\bullet \langle qX\uparrow\rangle & \langle qX\uparrow\rangle \overset{a}{\hookrightarrow}_\bullet \langle pX\uparrow\rangle
\end{array}$$

As $a > 1/2$, the resulting pBPA has a tendency to decrease the stack height. Hence, both $\langle pXq\rangle$ and $\langle qXp\rangle$ terminate with probability 1.

Every run of M_Δ initiated in pX that reaches $q\varepsilon$ can be "mimicked" by the associated run of $M_{\Delta_\bullet}$ initiated in $\langle pXq\rangle$. Similarly, *almost every*[5] run of M_Δ initiated in pX that does *not* visit a configuration with empty stack corresponds to some run of $M_{\Delta_\bullet}$ initiated in $\langle pX\uparrow\rangle$.

Example 3.3. Let Δ be a pPDA with two control states p, q, one stack symbol X, and the following transitions:

$$pX \overset{0.5}{\hookrightarrow} pXX, \qquad pX \overset{0.5}{\hookrightarrow} q\varepsilon, \qquad qX \overset{1}{\hookrightarrow} q\varepsilon.$$

Then $[pXq] = 1$ and $[qXq] = 1$, which means that $\Delta_\bullet$ has just two stack symbols $\langle pXq\rangle$, $\langle qXq\rangle$, and the transitions

$$\langle pXq\rangle \overset{0.5}{\hookrightarrow}_\bullet \langle pXq\rangle\langle qXq\rangle, \qquad \langle pXq\rangle \overset{0.5}{\hookrightarrow}_\bullet \varepsilon, \qquad \langle qXq\rangle \overset{1}{\hookrightarrow}_\bullet \varepsilon.$$

The infinite run $pX, pXX, pXXX, \ldots$ does not correspond to any run in $M_{\Delta_\bullet}$ (note that $\langle pX\uparrow\rangle \notin \Gamma_\bullet$), but since the total probability of all infinite runs initiated in pX is zero, we still have that almost all (but not *all*) of these runs correspond to some run in $M_{\Delta_\bullet}$.

The correspondence between the runs of M_Δ and $M_{\Delta_\bullet}$ is formally captured by a finite family of functions $\Upsilon_\odot$ where $\odot \in Q \cup \{\uparrow\}$. For every run $w \in Run(pX)$ in M_Δ, the function $\Upsilon_\odot$ returns an infinite sequence $w_\odot$ such that $w_\odot(i) \in \Gamma_\bullet^* \cup \{\times\}$ for every $i \in \mathbb{N}_0$. The sequence $w_\odot$ is either a run of $M_{\Delta_\bullet}$ initiated in $\langle pX\odot\rangle$, or an *invalid sequence*. As we shall see, all invalid sequences have an infinite suffix of "$\times$" symbols and correspond to those runs of $Run(pX)$ that cannot be mimicked by a run of $Run(\langle pX\odot\rangle)$.

So, let $\odot \in Q \cup \{\uparrow\}$, and let w be a run of M_Δ initiated in pX. We define an infinite sequence $w_\odot$ over $\Gamma_\bullet^* \cup \{\times\}$ inductively as follows:

- $w_\odot(0)$ is either $\langle pX\odot\rangle$ or $\times$, depending on whether $\langle pX\odot\rangle \in \Gamma_\bullet$ or not, respectively.
- If $w_\odot(i) = \times$ or $w_\odot(i) = \varepsilon$, then $w_\odot(i{+}1) = w_\odot(i)$. Otherwise, we have that $w_\odot(i) = \langle pX\dagger\rangle\alpha$, where $\dagger \in Q \cup \{\uparrow\}$, and $w(i) = pX\gamma$ for some $\gamma \in \Gamma^*$. Let $pX \hookrightarrow r\beta$ be the rule of Δ used to derive the transition $w(i) \to w(i{+}1)$. We put

$$w_\odot(i{+}1) = \begin{cases} \alpha & \text{if } \beta = \varepsilon \text{ and } \dagger = r; \\ \langle rY\dagger\rangle\alpha & \text{if } \beta = Y \text{ and } [rY\dagger] > 0; \\ \langle rYs\rangle\langle sZ\dagger\rangle\alpha & \text{if } \beta = YZ,\ [sZ\dagger] > 0,\ \text{and there is } k > i \text{ such that } w(k) = sZ\gamma \\ & \text{and the stack length in all } w(j),\ \text{where } i < j < k,\ \text{is strictly larger} \\ & \text{than the stack length in } w(i); \\ \langle rY\uparrow\rangle\alpha & \text{if } \beta = YZ,\ \dagger = \uparrow [rY\uparrow] > 0,\ \text{and the stack length in every } w(j),\ \text{where } j > i, \\ & \text{is strictly larger that the stack length in } w(i); \\ \times & \text{otherwise.} \end{cases}$$

[5] Here "almost every" is meant in the usual probabilistic sense, i.e., the probability of the remaining runs is zero.

We say that $w \in Run(pX)$ is $\odot$-*invalid* if $w_\odot(i) = \times$ for some $i \in \mathbb{N}_0$. Otherwise, w is $\odot$-*valid*. It is easy to check that w is $\odot$-valid iff $w_\odot \in Run(\langle pX\odot\rangle)$. Hence, $\Upsilon_\odot$ can be seen as a partial function from $Run(pX)$ to $Run(\langle pX\odot\rangle)$ which is defined only for $\odot$-valid runs. Further, for every $\odot$-valid $w \in Run(pX)$ and every $i \in \mathbb{N}_0$ we have that

- $w(i) = rY\beta$ iff $w_\odot(i) = \langle rY\dagger\rangle\gamma$ for some $\dagger \in Q \cup \{\uparrow\}$ and $\gamma \in \Gamma_\bullet^*$,
- $w(i) = r\varepsilon$ iff $w_\odot(i) = \varepsilon$ and $\odot = r$.

Hence, $\Upsilon_\odot$ preserves all properties of runs that depend just on the heads of visited configurations. Further, $\Upsilon_\odot$ preserves the probability of all measurable subsets of $Run(pX)$ with respect to a probability measure $P_\odot$ defined as follows. Let F be the standard σ-field over $Run(pX)$ generated by all basic cylinders (see Section 2). The function $P_\odot$ is the unique probability function over F such that for every $w \in FPath(pX)$ we have that

$$P_\odot\big(Run(w)\big) = \frac{P(Run(w) \cap Run(pX\odot))}{[pX\odot]}$$

where P is the standard probability function introduced in Section 2. Note that $P_\odot(R) = P(R \cap Run(pX\odot))/[pX\odot]$ for every $R \in F$.

Now we can state the main proposition, which says that $\Upsilon_\odot$ is a probability preserving measurable function.

Proposition 3.4. *Let $\Delta = (Q, \Gamma, \hookrightarrow, Prob)$ be a pPDA, $p \in Q$, $X \in \Gamma$, and $\odot \in \Gamma \cup \{\uparrow\}$ such that $[pX\odot] > 0$. Then for every measurable subset $R \subseteq Run(\langle pX\odot\rangle)$ we have that $\Upsilon_\odot^{-1}(R) \subseteq Run(pX)$ is measurable and $P(R) = P_\odot(\Upsilon_\odot^{-1}(R))$. Here $\Upsilon_\odot^{-1}(R)$ is the set of all $w \in Run(pX)$ such that $w_\odot \in R$.*

Proof. Since the probability space $(Run(\langle pX\odot\rangle), F, P)$ is generated by all $Run(v)$ where $v \in FPath(\langle pX\odot\rangle)$, it suffices to show that $\Upsilon_\odot^{-1}(Run(v))$ is measurable and $P(Run(v)) = P_\odot(\Upsilon_\odot^{-1}(Run(v)))$ for all $v \in FPath(\langle pX\odot\rangle)$.

Let us start with some auxiliary observations. Note that every configuration γ reachable from $\langle pX\odot\rangle$ in $M_{\Delta_\bullet}$ is of the form $\gamma = \langle p_1X_1p_2\rangle \cdots \langle p_{k-1}X_{k-1}p_k\rangle\langle p_kX_k\odot\rangle$ where $k \ge 0$ (if $k = 0$, then $\gamma = \varepsilon$). We put

$$P[\gamma] = [p_1X_1p_2] \cdots [p_{k-1}X_{k-1}p_k] \cdot [p_kX_k\odot]$$

Further, we say that a configuration $p\alpha$ of Δ is *compatible with* γ if $p = p_1$ and $\alpha = X_1 \cdots X_k\beta$, where $\beta \in \Gamma^*$. If $\odot \ne \uparrow$, we also require that $\beta = \varepsilon$. A run w of M_Δ *models* γ, written $w \models \gamma$, if the following conditions are satisfied:

- w is initiated in a configuration $p_1X_1 \cdots X_k\beta$ compatible with γ;
- w starts with a finite prefix of the form

  $$p_1X_1 \cdots X_k\beta \to^* p_2X_2 \cdots X_k\beta \to^* \cdots \to^* p_kX_k\beta$$

  where for all $1 \le i < k$, the stack length of all intermediate configurations visited along the subpath $p_iX_i \cdots X_k\beta \to^* p_{i+1}X_{i+1} \cdots X_k\beta$ is at least $|X_i \cdots X_k\beta|$. Further, if $\odot = \uparrow$, then the stack length in all configurations visited after $p_kX_k\beta$ is at least $|X_k\beta|$; and if $\odot = q$ for some $q \in Q$, then the above prefix is followed by a path from p_kX_k to $q\varepsilon$ (recall that $\beta = \varepsilon$ if $\odot = q$).

A straightforward induction on k reveals that for every configuration $p_1X_1 \cdots X_k\beta$ compatible with γ we have that

$$P\big(\big\{w \in Run(p_1X_1 \cdots X_k\beta) \bigm| w \models \gamma\big\}\big) = P[\gamma] \tag{1}$$

Now we can continue with the main proof. For a finite path u in M_Δ ending in a configuration $p\delta$ and a set of runs $R \subseteq Run(p\delta)$, we write $u \oplus R$ to denote the set of all runs obtained by concatenating u and w_1 for some $w \in R$. Let $v = \alpha_0, \ldots, \alpha_n$ be a finite path in $M_{\Delta_\bullet}$ initiated in $\langle pX\odot\rangle$. We say that a finite path $p_0\delta_0, \ldots, p_n\delta_n$ in M_Δ initiated in pX is *compatible with* v if $p_i\delta_i$ is compatible with α_i for every $0 \le i \le n$. Let $C(v)$ be the set of all finite paths compatible with v. It is easy to check (by induction on n) that

$$\Upsilon_\odot^{-1}\big(Run(\alpha_0, \ldots, \alpha_n)\big) = \bigcup_{p_0\delta_0, \ldots, p_n\delta_n \in C(\alpha_0, \ldots, \alpha_n)} p_0\delta_0, \ldots, p_n\delta_n \oplus \big\{w \in Run(p_n\delta_n) \bigm| w \models \alpha_n\big\} \tag{2}$$

Hence, $\Upsilon_\odot^{-1}(Run(\alpha_0, \ldots, \alpha_n))$ is measurable. By combining (1) and (2) we further obtain

$$P_\odot\big(\Upsilon_\odot^{-1}\big(Run(\alpha_0, \ldots, \alpha_n)\big)\big) = \frac{P[\alpha_n]}{[pX\odot]} \cdot \sum_{p_0\delta_0, \ldots, p_n\delta_n \in C(\alpha_0, \ldots, \alpha_n)} P\big(Run(p_0\delta_0, \ldots, p_n\delta_n)\big) \tag{3}$$

It remains to prove that $P_\odot(\Upsilon_\odot^{-1}(Run(\alpha_0, \ldots, \alpha_n))) = P(Run(\alpha_0, \ldots, \alpha_n))$. We proceed by induction on n. In what follows, the symbols P and $P_\odot$ are slightly overloaded since we need to consider sets of runs initiated in various configurations.

In the base case, when $n = 0$ and $\alpha_0 = \langle pX\odot\rangle$, we have that $P(Run(\langle pX\odot\rangle)) = 1$ and

$$P_\odot\big(\Upsilon_\odot^{-1}\big(Run(\langle pX\odot\rangle)\big)\big) = \frac{P[\langle pX\odot\rangle]}{[pX\odot]} \cdot \sum_{p_0\delta_0 \in C(\langle pX\odot\rangle)} P\big(Run(p_0\delta_0)\big) = 1$$

by applying (3) (note that $P[\langle pX\odot\rangle] = [pX\odot]$ and the only path initiated in pX compatible with $\langle pX\odot\rangle$ is pX).

For the inductive step, let us denote the finite path $\alpha_0, \dots, \alpha_{n-1}$ by $\bar{v}$ (i.e., $v = \alpha_0, \dots, \alpha_n = \bar{v}, \alpha_n$), and let x be the probability of the transition $\alpha_{n-1} \to \alpha_n$ in $M_{\Delta_\bullet}$. By induction hypothesis, $P_\odot(\Upsilon_\odot^{-1}(Run(\bar{v}))) = P(Run(\bar{v}))$, and by applying (3) we get

$$P\big(Run(\bar{v})\big) = \frac{P[\alpha_{n-1}]}{[pX\odot]} \cdot \sum_{p_0\delta_0,\dots,p_{n-1}\delta_{n-1} \in C(\bar{v})} P\big(Run(p_0\delta_0, \dots, p_{n-1}\delta_{n-1})\big) \tag{4}$$

Further, for every configuration $p_{n-1}\delta_{n-1}$ which is compatible with α_{n-1} we have that

$$P\big(\big\{w \in Run(p_{n-1}\delta_{n-1}) \bigm| w \models \alpha_{n-1}, w_1 \models \alpha_n\big\}\big) = x \cdot P\big(\big\{w \in Run(p_{n-1}\delta_{n-1}) \bigm| w \models \alpha_{n-1}\big\}\big) \tag{5}$$

Equality (5) follows easily by considering possible forms of the rule that induces the transition $\alpha_{n-1} \xrightarrow{x} \alpha_n$. From now on we use u to range over $C(v)$, and $\bar{u}$ to range over $C(\bar{v})$. That is, u abbreviates $p_0\delta_0, \dots, p_n\delta_n$, and $\bar{u}$ abbreviates $p_0\delta_0, \dots, p_{n-1}\delta_{n-1}$. However, we keep writing $p_n\delta_n$ instead of $u(n)$, and $p_{n-1}\delta_{n-1}$ instead of $\bar{u}(n-1)$, because we find these symbols more suggestive. Using this notation, we finally obtain that $P_\odot(\Upsilon_\odot^{-1}(Run(v)))$ is equal to

$$\begin{aligned}
& P_\odot\Big(\bigcup_{u \in C(v)} u \oplus \big\{w \in Run(p_n\delta_n) \bigm| w \models \alpha_n\big\}\Big) && \text{(by (2))}\\
&= \frac{1}{[pX\odot]} \cdot P\Big(\bigcup_{u \in C(v)} u \oplus \big\{w \in Run(p_n\delta_n) \bigm| w \models \alpha_n\big\}\Big) && \text{(defn. of } P_\odot\text{)}\\
&= \frac{1}{[pX\odot]} \cdot P\Big(\bigcup_{\bar{u} \in C(\bar{v})} \bar{u} \oplus \big\{w \in Run(p_{n-1}\delta_{n-1}) \bigm| w \models \alpha_{n-1}, w_1 \models \alpha_n\big\}\Big) &&\\
&= \frac{1}{[pX\odot]} \cdot \sum_{\bar{u} \in C(\bar{v})} P\big(\bar{u} \oplus \big\{w \in Run(p_{n-1}\delta_{n-1}) \bigm| w \models \alpha_{n-1}, w_1 \models \alpha_n\big\}\big) &&\\
&= \frac{1}{[pX\odot]} \cdot \sum_{\bar{u} \in C(\bar{v})} P\big(Run(\bar{u})\big) \cdot x \cdot P\big(\big\{w \in Run(p_{n-1}\delta_{n-1}) \bigm| w \models \alpha_{n-1}\big\}\big) && \text{(by (5))}\\
&= \frac{x \cdot P[\alpha_{n-1}]}{[pX\odot]} \cdot \sum_{\bar{u} \in C(\bar{v})} P\big(Run(\bar{u})\big) && \text{(by (1))}\\
&= x \cdot P\big(Run(\bar{v})\big) = P\big(Run(v)\big) && \text{(by (4))} \quad \square
\end{aligned}$$

In particular, Proposition 3.4 implies that all symbols of the form $\langle pXq\rangle$ which belong to $\Gamma_\bullet$ terminate with probability 1. To see this, let R be the set of all terminating runs initiated in $\langle pXq\rangle$. By Theorem 3.4, we obtain $P(R) = P_q(\Upsilon_q^{-1}(R)) = P_q(Run(pXq)) = 1$.

It is worth noting that all configurations reachable from a nonterminating configuration $\langle pX\uparrow\rangle \in \Gamma_\bullet$ take the form $\alpha\langle qY\uparrow\rangle$, where α terminates almost surely and $\langle qY\uparrow\rangle$ never terminates. It follows that $\Delta_\bullet$ can be transformed into a *finite-state* Markov chain whose states are the nonterminating symbols of $\Gamma_\bullet$, and transitions correspond to finite paths between two consecutive visits to nonterminating symbols. This finite-state Markov chain is very useful when investigating the properties of nonterminating runs, and many of the existing results about pPDAs can be substantially simplified using this approach.

Another consequence of Proposition 3.4 is the following:

Proposition 3.5. *Let $pXq \in Q \times \Gamma \times Q$ and $[pXq] > 0$. Then for all $n \in \mathbb{N}$ we have that*

$$P\big(\mathbf{T}_{pX} = n \bigm| Run(pXq)\big) = P(\mathbf{T}_{\langle pXq\rangle} = n).$$

Here, $\mathbf{T}_{pX} : Run(pX) \to \mathbb{N}_0$ and $\mathbf{T}_{\langle pXq\rangle} : Run(\langle pXq\rangle) \to \mathbb{N}_0$ are the random variables introduced in Section 2.

Proof. Let R be the set of all $w \in Run(\langle pXq\rangle)$ such that $\mathbf{T}_{\langle pXq\rangle}(w) = n$. Observe that $\Upsilon_\odot^{-1}(R)$ is the set $\hat{R}$ of all $\hat{w} \in Run(pXq)$ such that $\mathbf{T}_{pX}(\hat{w}) = n$. Hence,

$$P(\mathbf{T}_{\langle pXq\rangle} = n) = P(R) = P_q\big(\Upsilon_q^{-1}(R)\big) = P_q(\hat{R}) = P(\hat{R})/[pXq] = P\big(\mathbf{T}_{pX} = n \bigm| Run(pXq)\big) \qquad \square$$

4. Analysis of pBPAs

In this section we establish the promised tight tail bounds for the termination time. By virtue of Proposition 3.5, it suffices to analyze *almost surely terminating* pBPAs, i.e., only pBPAs such that all stack symbols terminate with probability 1. In what follows we assume that Δ is such a pBPA, and we also fix an initial stack symbol X_0. For $X, Y \in \Gamma$, we say that *X depends directly on Y*, if there is a rule $X \hookrightarrow \alpha$ such that Y occurs in α. Further, we say that *X depends on Y*, if either X depends directly on Y, or X depends directly on a symbol $Z \in \Gamma$ which depends on Y. One can compute, in linear time, the directed acyclic graph (DAG) of strongly connected components (SCCs) of the dependence relation. The *height* of this DAG, denoted by h, is defined as the longest distance between a top SCC and a bottom SCC plus 1 (i.e., $h = 1$ if there is only one SCC). We can safely assume that all symbols on which X_0 does not depend were removed from Δ. We abbreviate $P(\mathbf{T}_{X_0} \geq n \mid Run(X_0))$ to $P(\mathbf{T}_{X_0} \geq n)$, and we use p_{min} to denote $\min\{p \mid X \stackrel{p}{\hookrightarrow} \alpha \text{ in } \Delta\}$. Here is our main result:

Theorem 4.1. *Let Δ be an almost surely terminating pBPA with stack alphabet Γ. Assume that $X_0 \in \Gamma$ depends on all $X \in \Gamma \setminus \{X_0\}$, and let $p_{min} = \min\{p \mid X \stackrel{p}{\hookrightarrow} \alpha \text{ in } \Delta\}$. Then one of the following is true:*

(1) $P(\mathbf{T}_{X_0} \geq 2^{|\Gamma|}) = 0$.

(2) $E[X_0]$ *is* **finite** *and for all* $n \in \mathbb{N}$ *with* $n \geq 2E[X_0]$ *we have that*

$$p_{min}^n \leq P(\mathbf{T}_{X_0} \geq n) \leq \exp\left(1 - \frac{n}{8E_{max}^2}\right)$$

where $E_{max} = \max_{X \in \Gamma} E[X]$.

(3) $E[X_0]$ *is* **infinite** *and there is* $n_0 \in \mathbb{N}$ *such that for all* $n \geq n_0$ *we have that*

$$c/n^{1/2} \leq P(\mathbf{T}_{X_0} \geq n) \leq d_1/n^{d_2}$$

where $d_1 = 18h|\Gamma|/p_{min}^{3|\Gamma|}$, *and* $d_2 = 1/(2^{h+1} - 2)$. *Here, h is the height of the DAG of SCCs of the dependence relation, and c is a suitable positive constant depending on* Δ.

More colloquially, Theorem 4.1 states that Δ satisfies either (1) or (2) or (3), where (1) is when Δ does not have any long terminating runs; and (2) resp. (3) is when the expected termination time is finite (resp. infinite) and the probability of performing a terminating run of length n decreases exponentially (resp. polynomially) in n.

One can effectively distinguish between the three cases set out in Theorem 4.1. More precisely, case (1) can be recognized in polynomial time by looking only at the structure of the pBPA, i.e., disregarding the probabilities. Determining whether $E[X_0]$ is finite or infinite can be done in polynomial space by employing the results of [3,15]. This holds even if the transition probabilities of Δ are represented just symbolically by formulae of $ExTh(\mathbb{R})$ (see Proposition 3.1).

The proof of Theorem 4.1 is based on designing suitable martingales that are used to analyze the concentration of the termination time. Recall that a *martingale* is an infinite sequence of random variables $m^{(0)}, m^{(1)}, \ldots$ such that, for all $i \in \mathbb{N}$, $\mathbb{E}[|m^{(i)}|] < \infty$, and $\mathbb{E}[m^{(i+1)} \mid m^{(1)}, \ldots, m^{(i)}] = m^{(i)}$ almost surely. If $|m^{(i)} - m^{(i-1)}| < c_i$ for all $i \in \mathbb{N}$, then we have the following *Azuma's inequality* (see, e.g., [28]):

$$P\left(m^{(n)} - m^{(0)} \geq t\right) \leq \exp\left(\frac{-t^2}{2\sum_{k=1}^{n} c_k^2}\right)$$

We split the proof of Theorem 4.1 into four propositions (namely Propositions 4.2–4.5 below), which together imply Theorem 4.1.

The following proposition establishes the lower bound from Theorem 4.1 (2):

Proposition 4.2. *Let Δ be an almost surely terminating pBPA with stack alphabet Γ. Let $p_{min} = \min\{p \mid X \stackrel{p}{\hookrightarrow} \alpha \text{ in } \Delta\}$. Assume that $P(\mathbf{T}_{X_0} \geq 2^{|\Gamma|}) > 0$. Then we have*

$$p_{min}^n \leq P(\mathbf{T}_{X_0} \geq n) \quad \text{for all } n \in \mathbb{N}.$$

Proof. Let $\mathbf{T}_{X_0}(w) \geq n$ for some $n \in \mathbb{N}$ and some $w \in Run(X_0)$. It follows from the definition of the probability space of a pPDA that the set of all runs starting with $w(0), w(1), \ldots, w(n)$ has a probability of at least p_{min}^n. Therefore, in order to complete the proof, it suffices to show that $P(\mathbf{T}_{X_0} \geq 2^{|\Gamma|}) > 0$ implies $P(\mathbf{T}_{X_0} \geq n) > 0$ for all $n \in \mathbb{N}$.

To this end, we use a form of the pumping lemma for context-free languages. Notice that a pBPA can be regarded as a context-free grammar with probabilities (a stochastic context-free grammar) with an empty set of terminal symbols and Γ as the set of nonterminal symbols. Each finite run $w \in Run(X_0)$ corresponds to a derivation tree with root X_0 that derives the word ε. The termination time $\mathbf{T}_{X_0}$ is the number of (internal) nodes in the tree. In the rest of the proof we use this correspondence.

Let $P(\mathbf{T}_{X_0} \geq 2^{|\Gamma|}) > 0$. Then there is a run $w \in Run(X_0)$ with $\mathbf{T}_{X_0}(w) \geq 2^{|\Gamma|}$. This run w corresponds to a derivation tree with at least $2^{|\Gamma|}$ (internal) nodes. In this tree there is a path from the root (labeled with X_0) to a leaf such that on this path there are two different nodes, both labeled with the same symbol. Let us call those nodes n_1 and n_2, where n_1 is the node closer to the root. By replacing the subtree rooted at n_2 with the subtree rooted at n_1 we obtain a larger derivation tree. This completes the proof. □

The following proposition establishes the upper bound of Theorem 4.1 (2):

Proposition 4.3. *Let Δ be an almost surely terminating pBPA with stack alphabet Γ. Assume that X_0 depends on all $X \in \Gamma \setminus \{X_0\}$. Define*

$$E_{max} := \max_{X \in \Gamma} E[X] \quad \text{and} \quad B := \max_{X \hookrightarrow \alpha} \left| 1 - E[X] + \sum_{Y \in \Gamma} \#(Y)(\alpha) \cdot E[Y] \right|.$$

Then for all $n \in \mathbb{N}$ with $n \geq 2E[X_0]$ we have

$$P(\mathbf{T}_{X_0} \geq n) \leq \exp \frac{2E[X_0] - n}{2B^2} \leq \exp\left(1 - \frac{n}{8E_{max}^2}\right).$$

Proof. Let $w \in Run(X_0)$. We denote by $I(w)$ the maximal number $j \geq 0$ such that $w(j-1) \neq \varepsilon$. Given $i \geq 0$, we define $m^{(i)}(w) := E[w(i)] + \min\{i, I(w)\}$. We prove that $E(m^{(i+1)} \mid m^{(i)}) = m^{(i)}$, i.e., $m^{(0)}, m^{(1)}, \ldots$ forms a martingale. It has been shown in [15] that

$$\begin{aligned} E[X] &= \sum_{X \overset{x}{\hookrightarrow} \varepsilon} x + \sum_{X \overset{x}{\hookrightarrow} Y} x \cdot \big(1 + E[Y]\big) + \sum_{X \overset{x}{\hookrightarrow} YZ} x \cdot \big(1 + E[Y] + E[Z]\big) \\ &= 1 + \sum_{X \overset{x}{\hookrightarrow} Y} x \cdot E[Y] + \sum_{X \overset{x}{\hookrightarrow} YZ} x \cdot \big(E[Y] + E[Z]\big). \end{aligned}$$

On the other hand, let us fix a path $u \in FPath(X_0)$ of length $i+1$ and let w be an arbitrary run of $Run(u)$. First assume that $u(i) = X\alpha \in \Gamma\Gamma^*$. Then we have:

$$\begin{aligned} &\mathbb{E}\big[m^{(i+1)} \bigm| Run(u)\big] \\ &\quad = \mathbb{E}\big[E\big[w(i+1)\big] + i + 1 \bigm| Run(u)\big] \\ &\quad = i + 1 + \mathbb{E}\big[E\big[w(i+1)\big] \bigm| Run(u)\big] \\ &\quad = i + 1 + E[\alpha] + \sum_{X \overset{x}{\hookrightarrow} Y} x \cdot E[Y] + \sum_{X \overset{x}{\hookrightarrow} YZ} x \cdot \big(E[Y] + E[Z]\big) \\ &\quad = E[X] + E[\alpha] + i = E[X\alpha] + i = m^{(i)}(w) \end{aligned}$$

If $u(i) = \varepsilon$, then for every $w \in Run(u)$ we have $m^{(i+1)}(w) = I(w) = m^{(i)}(w)$. This proves that $m^{(0)}, m^{(1)}, \ldots$ is a martingale.

By Azuma's inequality (see [28]), we have

$$P\big(m^{(n)} - E[X_0] \geq n - E[X_0]\big) \leq \exp\left(\frac{-(n - E[X_0])^2}{2\sum_{k=1}^{n} B^2}\right) \leq \exp\left(\frac{2E[X_0] - n}{2B^2}\right).$$

For every $w \in Run(X_0)$ we have that $w(n) \neq \varepsilon$ implies $m^{(n)} \geq n$. It follows:

$$P(\mathbf{T}_{X_0} \geq n) \leq P\big(m^{(n)} \geq n\big) \leq \exp\left(\frac{2E[X_0] - n}{2B^2}\right) \leq \exp\left(1 - \frac{n}{8E_{max}^2}\right),$$

where the final inequality follows from the inequality $B \leq 2E_{max}$. □

The following proposition establishes the upper bound of Theorem 4.1 (3):

Proposition 4.4. *Let Δ be an almost surely terminating pBPA with stack alphabet Γ. Assume that X_0 depends on all $X \in \Gamma \setminus \{X_0\}$. Let $p_{min} = \min\{p \mid X \overset{p}{\hookrightarrow} \alpha \text{ in } \Delta\}$. Let h denote the height of the DAG of SCCs. Then there is $n_0 \in \mathbb{N}$ such that*

$$P(\mathbf{T}_{X_0} \geq n) \leq \frac{18h|\Gamma|/p_{min}^{3|\Gamma|}}{n^{1/(2^{h+1}-2)}} \quad \text{for all } n \geq n_0.$$

Proof sketch. A full proof is given in Section 4.1. Assume that $E[X_0]$ is infinite. To give some idea of the (quite involved) proof, let us first consider a simple pBPA Δ with $\Gamma = \{X\}$ and the rules $X \overset{1/2}{\hookrightarrow} XX$ and $X \overset{1/2}{\hookrightarrow} \varepsilon$. In fact, Δ is closely related to a simple random walk starting at 1, for which the time until it hits 0 can be exactly analyzed (see, e.g., [28, Chapter 10.12]). Clearly, we have $h = |\Gamma| = 1$ and $p_{min} = 1/2$. Theorem 4.1 (3) implies $P(\mathbf{T}_X \geq n) \in O(1/\sqrt{n})$. Let us sketch why this upper bound holds.

Let $\theta > 0$, define $g(\theta) := \frac{1}{2} \cdot \exp(-\theta \cdot (-1)) + \frac{1}{2} \cdot \exp(-\theta \cdot (+1))$, and define for a run $w \in Run(X)$ the sequence

$$m_\theta^{(i)}(w) = \begin{cases} \exp(-\theta \cdot |w(i)|)/g(\theta)^i & \text{if } i = 0 \text{ or } w(i-1) \neq \varepsilon \\ m_\theta^{(i-1)}(w) & \text{otherwise.} \end{cases}$$

One can show (cf. [28, Chapter 10.12]) that $m_\theta^{(0)}, m_\theta^{(1)}, \ldots$ is a martingale, i.e., $\mathbb{E}[m_\theta^{(i)} \mid m_\theta^{(i-1)}] = m_\theta^{(i-1)}$ for all $\theta > 0$. Our proof crucially depends on some analytic properties of the function $g : \mathbb{R} \to \mathbb{R}$: It is easy to verify that $1 = g(0) < g(\theta)$ for all $\theta > 0$, and $0 = g'(0)$, and $1 = g''(0)$. One can show that Doob's Optional-Stopping Theorem (see Theorem 10.10 (ii) of [28]) applies, which implies $m_\theta^{(0)} = \mathbb{E}[m_\theta^{(\mathbf{T}_X)}]$. It follows that for all $n \in \mathbb{N}$ and $\theta > 0$ we have that

$$\begin{aligned} \exp(-\theta) = m_\theta^{(0)} = \mathbb{E}\big[m_\theta^{(\mathbf{T}_X)}\big] = \mathbb{E}\big[g(\theta)^{-\mathbf{T}_X}\big] = \sum_{i=0}^{\infty} P(\mathbf{T}_X = i) \cdot g(\theta)^{-i} \\ \leq \sum_{i=0}^{n-1} P(\mathbf{T}_X = i) \cdot 1 + \sum_{i=n}^{\infty} P(\mathbf{T}_X = i) \cdot g(\theta)^{-n} \\ = 1 - P(\mathbf{T}_X \geq n) + P(\mathbf{T}_X \geq n) \cdot g(\theta)^{-n} \end{aligned} \tag{6}$$

Rearranging this inequality yields $P(\mathbf{T}_X \geq n) \leq \frac{1-\exp(-\theta)}{1-g(\theta)^{-n}}$, from which one obtains, setting $\theta := 1/\sqrt{n}$, and using the mentioned properties of g and several applications of l'Hopital's rule, that $P(\mathbf{T}_X \geq n) \in O(1/\sqrt{n})$.

Next we sketch how we generalize this proof to pBPAs that consist of only one SCC, but have more than one stack symbol. In this case, the term $|w(i)|$ in the definition of $m_\theta^{(i)}(w)$ needs to be replaced by the sum of *weights* of the symbols in $w(i)$. Each $Y \in \Gamma$ has a weight which is drawn from the dominant eigenvector of a certain matrix, which is characteristic for Δ. Perron–Frobenius theory guarantees the existence of a suitable weight vector $\vec{u} \in \mathbb{R}_+^\Gamma$. The function g consequently needs to be replaced by a function g_Y for each $Y \in \Gamma$. We need to keep the property that $g_Y''(0) > 0$. Intuitively, this means that Δ must have, for each $Y \in \Gamma$, a rule $Y \hookrightarrow \alpha$ such that Y and α have different weights. This can be accomplished by transforming Δ into a certain normal form.

Finally, we sketch how the proof is generalized to pBPAs with more than one SCC. For simplicity, assume that Δ has only two stack symbols, say X and Y, where X depends on Y, but Y does not depend on X. Let us change the execution order of pBPAs as follows: whenever a rule with $\alpha \in \Gamma^*$ on the right hand side fires, then all X-symbols in α are added on top of the stack, but all Y-symbols are added at the *bottom* of the stack. This change does not influence the termination time of pBPAs, but it allows to decompose runs into two phases: an X-phase where X-rules are executed which may produce Y-symbols or further X-symbols; and a Y-phase where Y-rules are executed which may produce further Y-symbols but no X-symbols, because Y does not depend on X. Arguing only qualitatively, assume that $\mathbf{T}_X$ is "large". Then either (a) the X-phase is "long" or (b) the X-phase is "short", but the Y-phase is "long". For the probability of event (a) one can give an upper bound using the bound for one SCC, because the produced Y-symbols can be ignored. For event (b), observe that if the X-phase is short, then only few Y-symbols can be created during the X-phase. For a bound on the probability of event (b) we need a bound on the probability that a pBPA with one SCC and a "short" initial configuration takes a "long" time to terminate. The previously sketched proof for an initial configuration with a single stack symbol can be suitably generalized to handle other "short" configurations. All details are given in Section 4.1. □

The following proposition establishes the lower bound of Theorem 4.1 (3):

Proposition 4.5. *Let Δ be an almost surely terminating pBPA with stack alphabet Γ. Assume that X_0 depends on all $X \in \Gamma \setminus \{X_0\}$. Assume $E[X_0] = \infty$. Then there is $c > 0$ such that*

$$\frac{c}{\sqrt{n}} \leq P(\mathbf{T}_{X_0} \geq n) \quad \textit{for all } n \in \mathbb{N}.$$

The proof of Proposition 4.5 follows the lines of the previous proof sketch, but with an additional trick: To obtain the desired bound, one needs to take the derivative with respect to θ on both sides of Eq. (6). The full proof is given in Section 4.2.

Tightness of the bounds in the case of infinite expectation If $E[X_0]$ is infinite, the lower and upper bounds of Theorem 4.1 (3) asymptotically coincide in the "strongly connected" case (i.e., where $h = 1$ holds for the height of the DAG of the SCCs of the dependence relation). In other words, in the strongly connected case we must have $P(\mathbf{T} \geq n) \in \Theta(1/\sqrt{n})$. Otherwise (i.e., for larger h) the upper bound in Theorem 4.1 (3) cannot be substantially tightened. This follows from the following proposition:

Proposition 4.6. *Let Δ_h be the pBPA with $\Gamma_h = \{X_1, \ldots, X_h\}$ and the following rules:*

$$X_h \stackrel{1/2}{\hookrightarrow} X_h X_h, \quad X_h \stackrel{1/2}{\hookrightarrow} X_{h-1}, \quad \ldots, \quad X_2 \stackrel{1/2}{\hookrightarrow} X_2 X_2, \quad X_2 \stackrel{1/2}{\hookrightarrow} X_1, \quad X_1 \stackrel{1/2}{\hookrightarrow} X_1 X_1, \quad X_1 \stackrel{1/2}{\hookrightarrow} \varepsilon$$

Then $[X_h] = 1$, $E[X_h] = \infty$, and there is $c_h > 0$ with

$$\frac{c_h}{n^{1/2^h}} \leq P(\mathbf{T}_{X_h} \geq n) \quad \text{for all } n \in \mathbb{N}.$$

Proposition 4.6 is proved in Section 4.3.

4.1. Proof of Proposition 4.4

In this subsection we prove Proposition 4.4. Given a finite set Γ, we regard the elements of $\mathbb{R}^\Gamma$ as vectors. Given two vectors $\vec{u}, \vec{v} \in \mathbb{R}^\Gamma$, we define a scalar product by setting $\vec{u} \bullet \vec{v} := \sum_{X \in \Gamma} \vec{u}(X) \cdot \vec{v}(X)$. Further, elements of $\mathbb{R}^{\Gamma \times \Gamma}$ are regarded as matrices, with the usual matrix–vector multiplication.

It will be convenient for the proof to measure the termination time of pBPAs starting in an arbitrary initial configuration $\alpha_0 \in \Gamma\Gamma^*$, not just with a single initial symbol $X_0 \in \Gamma$. To this end we generalize $\mathbf{T}_{X_0}$, $Run(X_0)$, etc. to $\mathbf{T}_{\alpha_0}$, $Run(\alpha_0)$, etc. in the straightforward way.

It will also be convenient to allow "pBPAs" that have transition rules with more than two stack symbols on the right-hand side. We call them *relaxed pBPAs*. All concepts associated to a pBPA, e.g., the induced Markov chain, termination time, etc., are defined analogously for relaxed pBPAs.

A relaxed pBPA is called *strongly connected*, if the DAG of the dependence relation on its stack alphabet consists of a single SCC.

For any $\alpha \in \Gamma^*$, define $\#(\alpha)$ as the Parikh image of α, i.e., the vector of $\mathbb{N}^\Gamma$ such that $\#(\alpha)(Y)$ is the number of occurrences of Y in α. Given a relaxed pBPA Δ, let $A_\Delta \in \mathbb{R}^{\Gamma \times \Gamma}$ be the matrix with

$$A_\Delta(X, Y) = \sum_{X \stackrel{p}{\hookrightarrow} \alpha} p \cdot \#(\alpha)(Y).$$

We drop the subscript of A_Δ if Δ is clear from the context. Intuitively, $A(X, Y)$ is the expected number of Y-symbols pushed on the stack when executing a rule with X on the left hand side. For instance, if $X \stackrel{1/5}{\hookrightarrow} XX$ and $X \stackrel{4/5}{\hookrightarrow} \varepsilon$, then $A(X, X) = 2/5$. Note that A is nonnegative. The matrix A plays a crucial role in the analysis of pPDAs and related models (see e.g. [19]) and in the theory of branching processes [20]. We have the following lemma:

Lemma 4.7. *Let Δ be an almost surely terminating, strongly connected pBPA. Then there is a positive vector $\vec{u} \in \mathbb{R}^\Gamma_+$ such that $A \cdot \vec{u} \leq \vec{u}$, where $\leq$ is meant componentwise. All such vectors $\vec{u}$ satisfy $\frac{\vec{u}_{min}}{\vec{u}_{max}} \geq p_{min}^{|\Gamma|}$, where p_{min} denotes the least rule probability in Δ, and $\vec{u}_{min}$ and $\vec{u}_{max}$ denote the least and the greatest component of $\vec{u}$, respectively.*

Proof. Let $X, Y \in \Gamma$. Since Δ is strongly connected, there is a sequence $X = X_1, X_2, \ldots, X_n = Y$ with $n \geq 1$ such that X_i depends directly on X_{i+1} for all $1 \leq i \leq n-1$. A straightforward induction on n shows that $A^n(X, Y) \neq 0$; i.e., A is *irreducible*. The assumption that Δ is almost surely terminating implies that the spectral radius of A is less than or equal to one, see, e.g., Section 8.1 of [19]. Perron–Frobenius theory (see, e.g., [1]) then implies that there is a positive vector $\vec{u} \in \mathbb{R}^\Gamma_+$ such that $A \cdot \vec{u} \leq \vec{u}$; e.g., one can take for $\vec{u}$ the dominant eigenvector of A.

Let $A \cdot \vec{u} \leq \vec{u}$. It remains to show that $\frac{\vec{u}_{min}}{\vec{u}_{max}} \geq p_{min}^{|\Gamma|}$. The proof is essentially given in [13], we repeat it for convenience. W.l.o.g. let $\Gamma = \{X_1, \ldots, X_{|\Gamma|}\}$. We write $\vec{u}_i$ for $\vec{u}(X_i)$. W.l.o.g. let $\vec{u}_1 = \vec{u}_{max}$ and $\vec{u}_{|\Gamma|} = \vec{u}_{min}$. Since Δ is strongly connected, there is a sequence $1 = r_1, r_2, \ldots, r_q = |\Gamma|$ with $q \leq |\Gamma|$ such that X_{r_j} depends on $X_{r_{j+1}}$ for all j. We have

$$\frac{\vec{u}_{min}}{\vec{u}_{max}} = \frac{\vec{u}_{|\Gamma|}}{\vec{u}_1} = \frac{\vec{u}_{r_q}}{\vec{u}_{r_{q-1}}} \cdot \ldots \cdot \frac{\vec{u}_{r_2}}{\vec{u}_{r_1}}.$$

By picking the smallest factor in the product, we find j with $2 \leq j \leq q$ such that

$$\frac{\vec{u}_{min}}{\vec{u}_{max}} \geq \left(\frac{\vec{u}_s}{\vec{u}_t}\right)^{q-1} \geq \left(\frac{\vec{u}_s}{\vec{u}_t}\right)^{|\Gamma|} \quad \text{where } s := r_j \text{ and } t := r_{j-1}. \tag{7}$$

We have $A \cdot \vec{u} \le \vec{u}$, which implies $A(X_s, X_t) \cdot \vec{u}_t \le \vec{u}_s$ and so $A(X_s, X_t) \le \vec{u}_s / \vec{u}_t$. On the other hand, since X_s depends on X_t, we clearly have $p_{min} \le A(X_s, X_t)$. Combining those inequalities with (7) yields $\frac{\vec{u}_{min}}{\vec{u}_{max}} \ge (A(X_s, X_t))^{|\Gamma|} \ge p_{min}^{|\Gamma|}$. $\square$

Given a relaxed pBPA Δ and vector $\vec{u} \in \mathbb{R}_+^\Gamma$, we say that Δ is *$\vec{u}$-progressive*, if Δ has, for all $X \in \Gamma$, a rule $X \hookrightarrow \alpha$ such that $|\vec{u}(X) - \#(\alpha) \bullet \vec{u}| \ge \vec{u}_{min}/2$. The following lemma states that, intuitively, any pBPA can be transformed into a $\vec{u}$-progressive relaxed pBPA that is at least as fast but no more than $|\Gamma|$ times faster.

Lemma 4.8. *Let Δ be an almost surely terminating pBPA with stack alphabet Γ. Let p_{min} denote the least rule probability in Δ, and let $\vec{u} \in \mathbb{R}_+^\Gamma$ with $A_\Delta \cdot \vec{u} \le \vec{u}$. Then one can construct a $\vec{u}$-progressive, almost surely terminating relaxed pBPA Δ' with stack alphabet Γ such that for all $\alpha_0 \in \Gamma^*$ and for all $a \ge 0$*

$$P'(\mathbf{T}_{\alpha_0} \ge a) \le P(\mathbf{T}_{\alpha_0} \ge a) \le P'\big(\mathbf{T}_{\alpha_0} \ge a/|\Gamma|\big),$$

where P and P' are the probability measures associated with Δ and Δ', respectively. Furthermore, the least rule probability in Δ' is at least $p_{min}^{|\Gamma|}$, and $A_{\Delta'} \cdot \vec{u} \le \vec{u}$. Finally, if $A_\Delta \cdot \vec{u} = \vec{u}$, then $A_{\Delta'} \cdot \vec{u} = \vec{u}$.

Proof. A sequence of transitions $X_1 \hookrightarrow \alpha_1, \dots, X_n \hookrightarrow \alpha_n$ is called *derivation sequence from X_1 to α_n*, if for all $i \in \{2, \dots, n\}$ the symbol $X_i \in \Gamma$ occurs in α_{i-1}. The *word induced* by a derivation sequence $X_1 \hookrightarrow \alpha_1, \dots, X_n \hookrightarrow \alpha_n$ is obtained by taking α_1, replacing an occurrence of X_2 by α_2, then replacing an occurrence of X_3 by α_3, etc., and finally replacing an occurrence of X_n by α_n.

Given a pBPA Δ and a derivation sequence $s = (X_1 \overset{p_1}{\hookrightarrow} \alpha_1^1 X_2 \alpha_1^2, X_2 \overset{p_2}{\hookrightarrow} \alpha_2, \dots, X_n \overset{p_n}{\hookrightarrow} \alpha_n)$ with $X_i \ne X_j$ for all $1 \le i < j \le n$, we define the *contraction $Con(s)$* of s, a set of X_1-transitions with possibly more than two symbols on the right hand side. The contraction $Con(s)$ will include a rule $X_1 \hookrightarrow \gamma$, where γ is the word induced by s. We define $Con(s)$ inductively over the length n of s. If $n = 1$, then $Con(s) = \{X_1 \overset{p_1}{\hookrightarrow} \alpha_1^1 X_2 \alpha_1^2\}$. If $n \ge 2$, let $s' = (X_2 \overset{p_2}{\hookrightarrow} \alpha_2, \dots, X_n \overset{p_n}{\hookrightarrow} \alpha_n)$ and define

$$\delta_2 := \{X_2 \hookrightarrow \beta \mid X_2 \hookrightarrow \beta \text{ is a rule in } \Delta\} - \big\{X_2 \overset{p2}{\hookrightarrow} \alpha_2\big\} \cup Con\big(s'\big); \tag{8}$$

i.e., δ_2 is the set of X_2-transitions in Δ with $X_2 \overset{p2}{\hookrightarrow} \alpha_2$ replaced by $Con(s')$. W.l.o.g. assume $\delta_2 = \{X_2 \overset{q_1}{\hookrightarrow} \beta_1, \dots, X_2 \overset{q_k}{\hookrightarrow} \beta_k\}$. Then we define

$$Con(s) := \big\{X_1 \overset{p_1 q_1}{\hookrightarrow} \alpha_1^1 \beta_1 \alpha_1^2, \dots, X_1 \overset{p_1 q_k}{\hookrightarrow} \alpha_1^1 \beta_k \alpha_1^2\big\}.$$

The following properties are easy to show by induction on n:

(a) $Con(s)$ contains $X_1 \hookrightarrow \gamma$, where γ is the word induced by s.
(b) The rule probabilities are at least p_{min}^n.
(c) Let Δ' be the relaxed pBPA obtained from Δ by replacing $X_1 \overset{p_1}{\hookrightarrow} \alpha_1^1 X_2 \alpha_1^2$ with $Con(s)$. Then each path in $M_{\Delta'}$ corresponds in a straightforward way to a path in M_Δ, namely to the path obtained by "re-expanding" the contractions. The corresponding path in M_Δ has the same probability and is not shorter but at most $|\Gamma|$ times longer than the one in $M_{\Delta'}$.
(d) Let Δ' be as in (c). Then $A_{\Delta'} \cdot \vec{u} \le \vec{u}$. Let us prove that explicitly. The induction hypothesis $n = 1$ is trivial. For the induction step, using the definition for δ_2 in (8) and $\delta_2 = \{X_2 \overset{q_1}{\hookrightarrow} \beta_1, \dots, X_2 \overset{q_k}{\hookrightarrow} \beta_k\}$, we know by the induction hypothesis that $\sum_{i=1}^k q_i \cdot \#(\beta_i) \bullet \vec{u} \le \vec{u}(X_2)$. This implies

$$\sum_{i=1}^k p_1 q_i \cdot \#\big(\alpha_1^1 \beta_i \alpha_1^2\big) \bullet \vec{u} \le p_1 \cdot \#\big(\alpha_1^1 X_2 \alpha_1^2\big) \bullet \vec{u}, \quad \text{and hence}$$

$$(A_{\Delta'} \cdot \vec{u})(X_1) \le (A_\Delta \cdot \vec{u})(X_1) \le \vec{u}(X_1).$$

Since A_Δ and $A_{\Delta'}$ may differ only in the X_1-row, we have $A_{\Delta'} \cdot \vec{u} \le \vec{u}$.
(e) Let Δ' be as in (c) and (d). If $A_\Delta \cdot \vec{u} = \vec{u}$, then $A_{\Delta'} \cdot \vec{u} = \vec{u}$. This follows as in (d), with the inequality signs replaced by equality.

Associate to each symbol $X_1 \in \Gamma$ a shortest derivation sequence

$$c(X_1) = (X_1 \hookrightarrow \alpha_1, \dots, X_{n-1} \hookrightarrow \alpha_{n-1}, X_n \hookrightarrow \varepsilon)$$

from X_1 to ε. Since Δ is almost surely terminating, the length of $c(X_1)$ is at most $|\Gamma|$ for all $X_1 \in \Gamma$. Let $X_1 \in \Gamma$, and let γ_1 denote the word induced by $c(X_1)$, and let γ_2 denote the word induced by the derivation sequence $c_2(X_1) := (X_1 \hookrightarrow \alpha_1, \dots, X_{n-1} \hookrightarrow \alpha_{n-1})$. We have $\#(\gamma_2) \bullet \vec{u} = \#(\gamma_1) \bullet \vec{u} + \vec{u}(X_n) \ge \#(\gamma_1) \bullet \vec{u} + \vec{u}_{min}$, so we can choose $\gamma \in \{\gamma_1, \gamma_2\}$ such that $|\vec{u}(X_1) - \#(\gamma) \bullet \vec{u}| \ge \vec{u}_{min}/2$. Choose $\hat{c}(X_1) \in \{c(X_1), c_2(X_1)\}$ such that $\hat{c}(X_1)$ induces γ. (Of course, if $c_2(X_1)$ has length zero, take $\hat{c}(X_1) = c(X_1)$.) Note that $(X_1 \hookrightarrow \gamma) \in Con(\hat{c}(X_1))$.

The relaxed pBPA Δ' from the statement of the lemma is obtained by replacing, for all $X_1 \in \Gamma$, the first rule of $\hat{c}(X_1)$ with $Con(\hat{c}(X_1))$. The properties (a)–(e) from above imply:

(a) The relaxed pBPA Δ' is $\vec{u}$-progressive.
(b) The rule probabilities are at least $p_{min}^{|\Gamma|}$.
(c) For each finite path w' in $M_{\Delta'}$ from some $\alpha_0 \in \Gamma^*$ to ε there is a finite path w in M_Δ from α_0 to ε such that $|w'| \le |w| \le |\Gamma| \cdot |w'|$ and $P'(w') = P(w)$. Hence, $P'(\mathbf{T}_{\alpha_0} < a/|\Gamma|) \le P(\mathbf{T}_{\alpha_0} < a) \le P'(\mathbf{T}_{\alpha_0} < a)$ holds for all $a \ge 0$, which implies $P'(\mathbf{T}_{\alpha_0} \ge a) \le P(\mathbf{T}_{\alpha_0} \ge a) \le P'(\mathbf{T}_{\alpha_0} \ge a/|\Gamma|)$.
(d) We have $A_{\Delta'} \cdot \vec{u} \le \vec{u}$.
(e) If $A_\Delta \cdot \vec{u} = \vec{u}$, then $A_{\Delta'} \cdot \vec{u} = \vec{u}$.

This completes the proof of the lemma. □

Proposition 4.9. *Let Δ be an almost surely terminating relaxed pBPA with stack alphabet Γ. Let $\vec{u} \in \mathbb{R}_+^\Gamma$ be such that $\vec{u}_{max} = 1$ and $A_\Delta \cdot \vec{u} \le \vec{u}$ and Δ is $\vec{u}$-progressive. Let p_{min} denote the least rule probability in Δ. Let $C := 17|\Gamma|/(p_{min} \cdot \vec{u}^2_{min})$. Then for each $k \in \mathbb{N}_0$ there is $n_0 \in \mathbb{N}$ such that*

$$P\big(\mathbf{T}_{\alpha_0} \ge n^{2k+2}/\big(2|\Gamma|\big)\big) \le C/n \quad \textit{for all } n \ge n_0 \textit{ and for all } \alpha_0 \in \Gamma^* \textit{ with } 1 \le |\alpha_0| \le n^k.$$

Proof. For each $X \in \Gamma$ we define a function $g_X : \mathbb{R} \to \mathbb{R}$ by setting

$$g_X(\theta) := \sum_{X \overset{p}{\hookrightarrow} \alpha} p \cdot \exp\big(-\theta \cdot \big(-\vec{u}(X) + \#(\alpha) \bullet \vec{u}\big)\big).$$

The following lemma states important properties of g_X.

Lemma 4.10. *The following holds for all $X \in \Gamma$:*

(a) For all $\theta > 0$ we have $1 = g_X(0) < g_X(\theta)$.
(b) For all $\theta > 0$ we have $0 \le g'_X(0) < g'_X(\theta)$.
(c) For all $\theta \ge 0$ we have $0 < g''_X(\theta)$. In particular, $g''_X(0) \ge p_{min} \cdot \vec{u}^2_{min}/4$.

Proof of the lemma.

(a) Clearly, $g_X(0) = 1$. The inequality $g_X(0) < g_X(\theta)$ follows from (b).
(b) We have:

$$g_X(\theta) = \sum_{X \overset{p}{\hookrightarrow} \alpha} p \cdot \exp\big(-\theta \cdot \big(-\vec{u}(X) + \#(\alpha) \bullet \vec{u}\big)\big)$$

$$g'_X(\theta) = \sum_{X \overset{p}{\hookrightarrow} \alpha} p \cdot \big(\vec{u}(X) - \#(\alpha) \bullet \vec{u}\big) \cdot \exp\big(-\theta \cdot \big(-\vec{u}(X) + \#(\alpha) \bullet \vec{u}\big)\big)$$

Let $A(X)$ denote the X-row of A, i.e., the vector $\vec{v} \in \mathbb{R}^\Gamma$ such that $\vec{v}(Y) = A(X, Y)$. Then $A \cdot \vec{u} \le \vec{u}$ implies

$$\begin{aligned} g'_X(0) &= \sum_{X \overset{p}{\hookrightarrow} \alpha} p \cdot \big(\vec{u}(X) - \#(\alpha) \bullet \vec{u}\big) \\ &= \vec{u}(X) - \sum_{X \overset{p}{\hookrightarrow} \alpha} p \cdot \#(\alpha) \bullet \vec{u} = \vec{u}(X) - A(X) \bullet \vec{u} \\ &\ge \vec{u}(X) - \vec{u}(X) = 0. \end{aligned}$$

The inequality $g'_X(0) < g'_X(\theta)$ follows from (c).
(c) We have

$$g''_X(\theta) = \sum_{X \overset{p}{\hookrightarrow} \alpha} p \cdot \big(\vec{u}(X) - \#(\alpha) \bullet \vec{u}\big)^2 \cdot \exp\big(-\theta \cdot \big(-\vec{u}(X) + \#(\alpha) \bullet \vec{u}\big)\big) > 0.$$

Since Δ is $\vec{u}$-progressive, there is a rule $X \overset{p}{\hookrightarrow} \alpha$ with $|\vec{u}(X) - \#(\alpha) \bullet \vec{u}| \ge \vec{u}_{min}/2$. Hence, for $\theta = 0$ we have $g''_X(0) \ge p_{min} \cdot \vec{u}^2_{min}/4$.

This proves the lemma. □

We construct a martingale by generalizing the martingale from the proof sketch for Proposition 4.4. Let in the following $\theta > 0$. Given a run $w \in Run(\alpha_0)$ and $i \geq 0$, we write $X^{(i)}(w)$ for the symbol $X \in \Gamma$ for which $w(i) = X\alpha$. Define

$$m_\theta^{(i)}(w) = \begin{cases} \exp(-\theta \cdot \#(w(i)) \bullet \vec{u}) \cdot \prod_{j=0}^{i-1} \frac{1}{g_{X^{(j)}(w)}(\theta)} & \text{if } i = 0 \text{ or } w(i-1) \neq \varepsilon \\ m_\theta^{(i-1)}(w) & \text{otherwise} \end{cases}$$

Lemma 4.11. *$m_\theta^{(0)}, m_\theta^{(1)}, \ldots$ is a martingale.*

Proof of the lemma. Let us fix a path $v \in FPath(\alpha_0)$ of length $i \geq 1$ and let w be an arbitrary run of $Run(v)$. First assume that $v(i-1) = X\alpha \in \Gamma\Gamma^*$. Then we have:

$$\begin{aligned}
&\mathbb{E}\big[m_\theta^{(i)} \,\big|\, Run(v)\big] \\
&= \mathbb{E}\left[\exp\big(-\theta \cdot \#\big(w(i)\big) \bullet \vec{u}\big) \cdot \prod_{j=0}^{i-1} \frac{1}{g_{X^{(j)}(w)}(\theta)} \,\middle|\, Run(v)\right] \\
&= \sum_{X \overset{p}{\hookrightarrow} \alpha} p \cdot \exp\big(-\theta \cdot \big(\#\big(w(i-1)\big) - \vec{1}_X + \#(\alpha)\big) \bullet \vec{u}\big) \cdot \prod_{j=0}^{i-1} \frac{1}{g_{X^{(j)}(w)}(\theta)} \\
&= \sum_{X \overset{p}{\hookrightarrow} \alpha} p \cdot \exp\big(-\theta \cdot \big(\#\big(w(i-1)\big) \bullet \vec{u} - \vec{u}(X) + \#(\alpha) \bullet \vec{u}\big)\big) \cdot \prod_{j=0}^{i-1} \frac{1}{g_{X^{(j)}(w)}(\theta)} \\
&= \exp\big(-\theta \cdot \#\big(w(i-1)\big) \bullet \vec{u}\big) \cdot \sum_{X \overset{p}{\hookrightarrow} \alpha} p \cdot \exp\big(-\theta \cdot \big(-\vec{u}(X) + \#(\alpha) \bullet \vec{u}\big)\big) \cdot \prod_{j=0}^{i-1} \frac{1}{g_{X^{(j)}(w)}(\theta)} \\
&= \exp\big(-\theta \cdot \#\big(w(i-1)\big) \bullet \vec{u}\big) \cdot g_{X^{(i-1)}(w)}(\theta) \cdot \prod_{j=0}^{i-1} \frac{1}{g_{X^{(j)}(w)}(\theta)} \\
&= \exp\big(-\theta \cdot \#\big(w(i-1)\big) \bullet \vec{u}\big) \cdot \prod_{j=0}^{i-2} \frac{1}{g_{X^{(j)}(w)}(\theta)} \\
&= m_\theta^{(i-1)}(w).
\end{aligned}$$

If $v(i-1) = \varepsilon$, then for every $w \in Run(v)$ we have $m_\theta^{(i)}(w) = m_\theta^{(i-1)}(w)$. Hence, $m_\theta^{(0)}, m_\theta^{(1)}, \ldots$ is a martingale. □

Since $\theta > 0$ and since $g_{X^{(j)}(w)}(\theta) \geq 1$ by Lemma 4.10 (a), we have $0 \leq m_\theta^{(i)}(w) \leq 1$, so the martingale is bounded. Since, furthermore, $\mathbf{T}_{\alpha_0}$ (we write only $\mathbf{T}$ in the following) is finite with probability 1, it follows using Doob's Optional-Stopping Theorem (see Theorem 10.10 (ii) of [28]) that $m_\theta^{(0)} = \mathbb{E}[m_\theta^{(\mathbf{T})}]$. Let $k \in \mathbb{N}_0$. For each $n \in \mathbb{N}$ we have:

$$\begin{aligned}
&\exp\big(-\theta \cdot \vec{u}_{max} \cdot n^k\big) \\
&\leq \exp\big(-\theta \cdot \vec{u} \bullet \#(\alpha_0)\big) = m_\theta^{(0)} \\
&= \mathbb{E}\big[m_\theta^{(\mathbf{T})}\big] && \text{(by optional-stopping)} \\
&= \mathbb{E}\left[\exp(-\theta \cdot 0) \cdot \prod_{j=0}^{\mathbf{T}-1} \frac{1}{g_{X^{(j)}}(\theta)}\right] \\
&= \mathbb{E}\left[\prod_{j=0}^{\mathbf{T}-1} \frac{1}{g_{X^{(j)}}(\theta)}\right] \\
&\leq \mathbb{E}\left[\frac{1}{g_X(\theta)^{\mathbf{T}}}\right] && \text{(for some } X \in \Gamma) \\
&= \sum_{i=0}^{\infty} P(\mathbf{T} = i) \cdot \frac{1}{g_X(\theta)^i}
\end{aligned}$$

$$\leq \sum_{i=0}^{\lceil n^{2k+2}/(2|\Gamma|)\rceil-1} P(\mathbf{T}=i)\cdot 1 \qquad \text{(Lemma 4.10 (a))}$$

$$+ \sum_{i=\lceil n^{2k+2}/(2|\Gamma|)\rceil}^{\infty} P(\mathbf{T}=i)\cdot \frac{1}{g_X(\theta)^{n^{2k+2}/(2|\Gamma|)}}$$

$$= 1 - P\big(\mathbf{T} \geq n^{2k+2}/(2|\Gamma|)\big)$$

$$+ P\big(\mathbf{T} \geq n^{2k+2}/(2|\Gamma|)\big)\cdot \frac{1}{g_X(\theta)^{n^{2k+2}/(2|\Gamma|)}}$$

Rearranging the inequality, we obtain

$$P\big(\mathbf{T} \geq n^{2k+2}/(2|\Gamma|)\big) \leq \frac{1-\exp(-\theta\cdot \vec{u}_{max}\cdot n^k)}{1-g_X(\theta)^{-n^{2k+2}/(2|\Gamma|)}}. \tag{9}$$

For the following we set $\theta = n^{-(k+1)}$. We want to give an upper bound for the right hand side of (9). To this end we will show:

$$\lim_{n\to\infty} \frac{(1-\exp(-n^{-(k+1)}\cdot \vec{u}_{max}\cdot n^k))\cdot n}{1-g_X(n^{-(k+1)})^{-n^{2(k+1)}/(2|\Gamma|)}} \leq \frac{1}{1-\exp(-p_{min}\cdot \vec{u}^2_{min}/(16|\Gamma|))}. \tag{10}$$

Combining (9) with (10), we obtain

$$\begin{aligned}\limsup_{n\to\infty} n\cdot P\big(\mathbf{T}\geq n^{2k+2}/(2|\Gamma|)\big) &\leq \frac{1}{1-\exp(-p_{min}\cdot \vec{u}^2_{min}/(16|\Gamma|))}\\ &< \frac{1}{1-(1-\frac{16}{17}\cdot(p_{min}\cdot \vec{u}^2_{min}/(16|\Gamma|)))}\\ &= 17|\Gamma|/\big(p_{min}\cdot \vec{u}^2_{min}\big),\end{aligned}$$

which implies the proposition. Here the second inequality is by observing that $\exp(-x) < 1-\frac{16}{17}x$ holds for all $x \in (0, \frac{1}{16})$.

To prove (10), we compute limits for the nominator and the denominator separately. For the nominator, we use l'Hopital's rule to obtain:

$$\lim_{n\to\infty} \frac{1-\exp(-\vec{u}_{max}\cdot n^{-1})}{n^{-1}} = \lim_{n\to\infty} \frac{-\vec{u}_{max}\cdot n^{-2}\cdot \exp(-\vec{u}_{max}\cdot n^{-1})}{-n^{-2}} = \vec{u}_{max} = 1.$$

For the denominator of (10) we consider first the following limit:

$$\begin{aligned}&\lim_{n\to\infty} \frac{1}{2|\Gamma|}\cdot n^{2(k+1)}\cdot \ln g_X\big(n^{-(k+1)}\big)\\ &= \frac{1}{2|\Gamma|}\lim_{n\to\infty}\frac{\ln g_X(n^{-(k+1)})}{n^{-2(k+1)}}\\ &= \frac{1}{2|\Gamma|}\lim_{n\to\infty}\frac{g'_X(n^{-(k+1)})\cdot(-(k+1))\cdot n^{-k-2}}{g_X(n^{-(k+1)})\cdot(-2(k+1))\cdot n^{-2k-3}} && \text{(l'Hopital's rule)}\\ &= \frac{1}{4|\Gamma|}\lim_{n\to\infty}\frac{g'_X(n^{-(k+1)})}{n^{-(k+1)}} && \text{(by Lemma 4.10 (a)).}\end{aligned}$$

If $g'_X(0) > 0$, then the limit is $+\infty$. Otherwise, by Lemma 4.10 (b), we have $g'_X(0) = 0$ and hence

$$\begin{aligned}&= \frac{1}{4|\Gamma|}\lim_{n\to\infty}\frac{g''_X(n^{-(k+1)})\cdot(-(k+1))\cdot n^{-k-2}}{(-(k+1))\cdot n^{-k-2}} && \text{(l'Hopital's rule)}\\ &= \frac{1}{4|\Gamma|} g''_X(0) \geq p_{min}\cdot \vec{u}^2_{min}/(16|\Gamma|) && \text{(by Lemma 4.10 (c)).}\end{aligned}$$

This proves (10) and thus completes the proof of Proposition 4.9. □

The following lemma serves as induction base for the proof of Proposition 4.4.

Lemma 4.12. *Let Δ be an almost surely terminating pBPA with stack alphabet Γ. Assume that all SCCs of Δ are bottom SCCs. Let p_{min} denote the least rule probability in Δ. Let $D := 17|\Gamma|/p_{min}^{3|\Gamma|}$. Then for each $k \in \mathbb{N}_0$ there is $n_0 \in \mathbb{N}$ such that*

$$P\left(\mathbf{T}_{\alpha_0} \geq n^{2k+2}/2\right) \leq D/n \quad \text{for all } n \geq n_0 \text{ and for all } \alpha_0 \in \Gamma^* \text{ with } 1 \leq |\alpha_0| \leq n^k.$$

Proof. Decompose Γ into its SCCs, say $\Gamma = \Gamma_1 \cup \cdots \cup \Gamma_s$, and let the pBPA Δ_i be obtained by restricting Δ to the Γ_i-symbols. For each $i \in \{1, \ldots, s\}$, Lemma 4.7 gives a vector $\vec{u}_i \in \mathbb{R}_+^{\Gamma_i}$. W.l.o.g. we can assume for each i that the largest component of $\vec{u}_i$ is equal to 1, because $\vec{u}_i$ can be multiplied with any positive scalar without changing the properties guaranteed by Lemma 4.7. If the vectors $\vec{u}_i$ are assembled (in the obvious way) to the vector $\vec{u} \in \mathbb{R}_+^{\Gamma}$, the assertions of Lemma 4.7 carry over; i.e., we have $A_\Delta \cdot \vec{u} \leq \vec{u}$ and $\vec{u}_{max} = 1$ and $\vec{u}_{min} \geq p_{min}^{|\Gamma|}$. Let Δ' be the $\vec{u}$-progressive relaxed pBPA from Lemma 4.8, and denote by P' and p'_{min} its associated probability measure and least rule probability, respectively. Then we have:

$$\begin{aligned}
P\left(\mathbf{T}_{\alpha_0} \geq n^{2k+2}/2\right) &\leq P'\left(\mathbf{T}_{\alpha_0} \geq n^{2k+2}/(2|\Gamma|)\right) && \text{(by Lemma 4.8)} \\
&\leq 17|\Gamma|/\left(p'_{min} \cdot \vec{u}_{min}^2 \cdot n\right) && \text{(by Proposition 4.9)} \\
&\leq 17|\Gamma|/\left(p'_{min} \cdot p_{min}^{2|\Gamma|} \cdot n\right) && \text{(as argued above)} \\
&\leq 17|\Gamma|/\left(p_{min}^{3|\Gamma|} \cdot n\right) && \text{(by Lemma 4.8).} \qquad \square
\end{aligned}$$

Now we are ready to prove Proposition 4.4, which is restated here.

Proposition 4.4. *Let Δ be an almost surely terminating pBPA with stack alphabet Γ. Assume that X_0 depends on all $X \in \Gamma \setminus \{X_0\}$. Let $p_{min} = \min\{p \mid X \xhookrightarrow{p} \alpha \text{ in } \Delta\}$. Let h denote the height of the DAG of SCCs. Then there is $n_0 \in \mathbb{N}$ such that*

$$P(\mathbf{T}_{X_0} \geq n) \leq \frac{18h|\Gamma|/p_{min}^{3|\Gamma|}}{n^{1/(2^{h+1}-2)}} \quad \text{for all } n \geq n_0.$$

Proof. Let $D = 17|\Gamma|/p_{min}^{3|\Gamma|}$ be the D from Lemma 4.12. We will show:

$$P\left(\mathbf{T}_{X_0} \geq n^{2^{h+1}-2}\right) \leq \frac{hD}{n} \quad \text{for almost all } n \in \mathbb{N}. \tag{11}$$

Eq. (11) implies the proposition. Indeed, assume that $P(\mathbf{T}_{X_0} \geq m^{2^{h+1}-2}) \leq \frac{hD}{m}$ holds for all $m \geq m_0$ for some $m_0 \in \mathbb{N}$. Let n be large enough so that $m^{2^{h+1}-2} \leq n \leq (m+1)^{2^{h+1}-2}$ holds for some $m \geq m_0$. Then we have:

$$P(\mathbf{T}_{X_0} \geq n) \leq P\left(\mathbf{T}_{X_0} \geq m^{2^{h+1}-2}\right) \leq \frac{hD}{m} \leq \frac{hD}{n^{1/(2^{h+1}-2)} - 1} \leq \frac{18}{17} \cdot \frac{hD}{n^{1/(2^{h+1}-2)}},$$

where the last inequality holds for large enough n. Thus we have shown that (11) implies the proposition.

We prove (11) by induction on h. The case $h = 1$ (induction base) is implied by Lemma 4.12. Let $h \geq 2$. Partition Γ into $\Gamma_{high} \cup \Gamma_{low}$ such that Γ_{low} contains the variables of the SCCs of depth h in the DAG of SCCs, and Γ_{high} contains the other variables (in "higher" SCCs). If $X_0 \in \Gamma_{low}$, then we can restrict Δ to the variables that are in the same SCC as X_0, and Lemma 4.12 implies (11). So we can assume $X_0 \in \Gamma_{high}$.

Assume for a moment that $P(\mathbf{T}_{X_0} \geq n^{2^{h+1}-2})$ holds for a run $w \in Run(X_0)$; i.e., we have:

$$\begin{aligned}
n^{2^{h+1}-2} &\leq \left|\left\{i \in \mathbb{N}_0 \mid w(i) \in \Gamma\Gamma^*\right\}\right| \\
&= \left|\left\{i \in \mathbb{N}_0 \mid w(i) \in \Gamma_{high}\Gamma^*\right\}\right| + \left|\left\{i \in \mathbb{N}_0 \mid w(i) \in \Gamma_{low}\Gamma^*\right\}\right|.
\end{aligned}$$

It follows that one of the following events is true for w:

(a) At least n^{2^h-2} steps in w have a Γ_{high}-symbol on top of the stack. More formally,

$$\left|\left\{i \in \mathbb{N}_0 \mid w(i) \in \Gamma_{high}\Gamma^*\right\}\right| \geq n^{2^h-2}.$$

(b) Event (a) is not true, but at least $n^{2^{h+1}-2} - n^{2^h-2}$ steps in w have a Γ_{low}-symbol on top of the stack. More formally,

$$\left|\left\{i \in \mathbb{N}_0 \mid w(i) \in \Gamma_{high}\Gamma^*\right\}\right| < n^{2^h-2} \quad \text{and}$$
$$\left|\left\{i \in \mathbb{N}_0 \mid w(i) \in \Gamma_{low}\Gamma^*\right\}\right| \geq n^{2^{h+1}-2} - n^{2^h-2}.$$

In order to give bounds on the probabilities of events (a) and (b), it is convenient to "reshuffle" the execution order of runs in the following way: Whenever a rule $X \hookrightarrow \alpha$ is executed, we do not replace the X-symbol on top of the stack by α, but instead we push only the Γ_{high}-symbols in α on top of the stack, whereas the Γ_{low}-symbols in α are added to the *bottom* of the stack. Since Δ is a pBPA and thus does not have control states, the reshuffling of the execution order does not influence the distribution of the termination time. The advantage of this execution order is that each run can be decomposed into two phases:

(1) In the first phase, the symbol on the top of the stack is always a Γ_{high}-symbol. When rules are executed, Γ_{low}-symbols may be produced, which are added to the bottom of the stack.
(2) In the second phase, the stack consists of Γ_{low}-symbols exclusively. Notice that by definition of Γ_{low}, no new Γ_{high}-symbols can be produced.

In terms of those phases, the events (a) and (b) above can be reformulated as follows:

(a) The first phase of w consists of at least n^{2^h-2} steps. The probability of this event is equal to

$$P_{\Delta_{high}}\left(\mathbf{T}_{X_0} \geq n^{2^h-2}\right),$$

where Δ_{high} is the pBPA obtained from Δ by deleting all Γ_{low}-symbols from the right hand sides of the rules and deleting all rules with Γ_{low}-symbols on the left hand side, and $P_{\Delta_{high}}$ is its associated probability measure.

(b) The first phase of w consists of fewer than n^{2^h-2} steps (which implies that at most n^{2^h-2} Γ_{low}-symbols are produced during the first phase), and the second phase consists of at least $n^{2^{h+1}-2} - n^{2^h-2}$ steps. Therefore, the probability of the event (b) is at most

$$\max\left\{P_{\Delta_{low}}\left(\mathbf{T}_{\alpha_0} \geq n^{2^{h+1}-2} - n^{2^h-2}\right) \,\middle|\, \alpha_0 \in \Gamma_{low}^*,\ 1 \leq |\alpha_0| \leq n^{2^h-2}\right\},$$

where Δ_{low} is the pBPA Δ restricted to the Γ_{low}-symbols, and $P_{\Delta_{low}}$ is its associated probability measure. Notice that $n^{2^{h+1}-2} - n^{2^h-2} \geq n^{2^{h+1}-2}/2$ for large enough n. Furthermore, by the definition of Γ_{low}, the SCCs of Δ_{low} are all bottom SCCs. Hence, by Lemma 4.12, the above maximum is at most D/n.

Summing up, we have for almost all $n \in \mathbb{N}$:

$$\begin{aligned} P\left(\mathbf{T}_{X_0} \geq n^{2^{h+1}-2}\right) &\leq P\left(\text{event (a)}\right) + P\left(\text{event (b)}\right) \\ &\leq P_{\Delta_{high}}\left(\mathbf{T}_{X_0} \geq n^{2^h-2}\right) + D/n && \text{(as argued above)} \\ &\leq \frac{(h-1)D}{n} + \frac{D}{n} = \frac{hD}{n} && \text{(by the induction hypothesis).} \end{aligned}$$

This completes the induction proof. □

4.2. Proof of Proposition 4.5

The proof of Proposition 4.5 is similar to the proof of Proposition 4.4 from the previous subsection. Here is a restatement of Proposition 4.5.

Proposition 4.5. *Let Δ be an almost surely terminating pBPA with stack alphabet Γ. Assume that X_0 depends on all $X \in \Gamma \setminus \{X_0\}$. Assume $E[X_0] = \infty$. Then there is $c > 0$ such that*

$$\frac{c}{\sqrt{n}} \leq P(\mathbf{T}_{X_0} \geq n) \quad \text{for all } n \in \mathbb{N}.$$

Proof. For a square matrix M denote by $\rho(M)$ the spectral radius of M, i.e., the greatest absolute value of its eigenvectors. Let A_Δ be the matrix from the previous subsection. We claim:

$$\rho(A_\Delta) = 1. \tag{12}$$

The assumption that Δ is almost surely terminating implies that $\rho(A_\Delta) \leq 1$, see, e.g., Section 8.1 of [19]. Assume for a contradiction that $\rho(A_\Delta) < 1$. Using standard theory of nonnegative matrices (see, e.g., [1]), this implies that the matrix inverse $B := (I - A_\Delta)^{-1}$ (here, I denotes the identity matrix) exists; i.e., B is finite in all components. It is shown in [15] that $E[X_0] = (B \cdot \vec{1})(X_0)$ (here, $\vec{1}$ denotes the vector with $\vec{1}(X) = 1$ for all X). This is a contradiction to our assumption that $E[X_0] = \infty$. Hence, (12) is proved.

It follows from (12) and standard theory of nonnegative matrices [1] that A_Δ has a principal submatrix, say A', which is irreducible and satisfies $\rho(A') = 1$. Let Γ' be the subset of Γ such that A' is obtained from A by deleting all rows and columns which are not indexed by Γ'. Let Δ' be the pBPA with stack alphabet Γ' such that Δ' is obtained from Δ by removing all rules with symbols from $\Gamma \setminus \Gamma'$ on the left hand side and removing all symbols from $\Gamma \setminus \Gamma'$ from all right hand sides. Clearly, $A_{\Delta'} = A'$, so $\rho(A_{\Delta'}) = 1$ and $A_{\Delta'}$ is irreducible. Since Δ' is a sub-pBPA of Δ and X_0 depends on all symbols in Γ', it suffices to prove the proposition for Δ' and an arbitrary start symbol $X_0' \in \Gamma'$.

Therefore, w.l.o.g. we can assume in the following that $A_\Delta = A$ is irreducible. Then it follows, using (12) and Perron–Frobenius theory [1], that there is a positive vector $\vec{u} \in \mathbb{R}_+^\Gamma$ such that $A \cdot \vec{u} = \vec{u}$. W.l.o.g. we assume $\vec{u}(X_0) = 1$. Using Lemma 4.8 we can assume w.l.o.g. that Δ is $\vec{u}$-progressive. (The pBPA Δ may be relaxed.)

As in the proof of Proposition 4.9, for each $X \in \Gamma$ we define a function $g_X : \mathbb{R} \to \mathbb{R}$ by setting

$$g_X(\theta) := \sum_{X \overset{p}{\hookrightarrow} \alpha} p \cdot \exp\big(-\theta \cdot \big(-\vec{u}(X) + \#(\alpha)\bullet\vec{u}\big)\big).$$

The following lemma states some properties of g_X.

Lemma 4.13. *The following holds for all $X \in \Gamma$:*

(a) *For all $\theta > 0$ we have $1 = g_X(0) < g_X(\theta)$.*
(b) *For all $\theta > 0$ we have $0 = g'_X(0) < g'_X(\theta)$.*
(c) *For all $\theta \geq 0$ we have $0 < g''_X(\theta)$.*
(d) *There is $c_2 > 0$ such that for all $0 < \theta \leq 1$ we have $g'_X(\theta) \leq c_2\theta$.*
(e) *There is $c_3 > 1$ such that for all $n \in \mathbb{N}$ we have $g_X(1/\sqrt{n})^n \geq c_3$.*
(f) *There is $c_4 > 0$ such that for all $n \in \mathbb{N}$ we have $\frac{1/n}{1-1/g_X(1/\sqrt{n})} \leq c_4$.*

Proof of the lemma. The proof of items (a)–(c) follows exactly the proof of Lemma 4.10 and is therefore omitted. (For the equality $0 = g'_X(0)$ in (b) one uses $A \cdot \vec{u} = \vec{u}$.)

(d) It suffices to prove that $g'_X(\theta)/\theta$ is bounded for $\theta \to 0$. Using l'Hopital's rule we have $\lim_{\theta\to 0} g'_X(\theta)/\theta = g''_X(0) > 0$.
(e) Clearly, we have $g_X(1/\sqrt{n})^n > 1$ for all n. Furthermore, we have:

$$\begin{aligned}
\lim_{n\to\infty} \ln g_X(1/\sqrt{n})^n &= \lim_{n\to\infty} \frac{\ln g_X(n^{-1/2})}{1/n} \\
&= \frac{1}{2}\lim_{n\to\infty} \frac{g'_X(n^{-1/2})}{n^{-1/2}} && \text{(l'Hopital's rule)} \\
&= \frac{g''_X(0)}{2} && \text{(l'Hopital's rule)} \\
&> 0 && \text{(by (c))}
\end{aligned}$$

Hence the claim follows.
(f) The claim follows again from l'Hopital's rule:

$$\begin{aligned}
\lim_{n\to\infty} \frac{1/n}{1-1/g_X(n^{-1/2})} &= \lim_{n\to\infty} \frac{-1/n^2}{(1/g_X(n^{-1/2}))^2 \cdot g'_X(n^{-1/2}) \cdot (-1/2)n^{-3/2}} \\
&= \lim_{n\to\infty} \frac{2n^{-1/2}}{g'_X(n^{-1/2})} = \frac{2}{g''_X(0)} < \infty
\end{aligned}$$

This completes the proof of the lemma. □

Let in the following $\theta > 0$. As in the proof of Proposition 4.9, given a run $w \in Run(X_0)$ and $i \geq 0$, we write $X^{(i)}(w)$ for the symbol $X \in \Gamma$ for which $w(i) = X\alpha$. Define

$$m_\theta^{(i)}(w) = \begin{cases} \exp(-\theta \cdot \#(w(i))\bullet\vec{u}) \cdot \prod_{j=0}^{i-1} \frac{1}{g_{X^{(j)}(w)}(\theta)} & \text{if } i = 0 \text{ or } w(i-1) \neq \varepsilon \\ m_\theta^{(i-1)}(w) & \text{otherwise} \end{cases}$$

As in Lemma 4.11, one can show that the sequence $m_\theta^{(0)}, m_\theta^{(1)}, \ldots$ is a martingale. As in the proof of Proposition 4.9, Doob's Optional-Stopping Theorem implies $\exp(-\theta) = m_\theta^{(0)} = \mathbb{E}[m_\theta^{(\mathbf{T}_{X_0})}]$. Hence we have for each $n \in \mathbb{N}$ (writing $\mathbf{T}$ for $\mathbf{T}_{X_0}$):

$$
\begin{aligned}
\exp(-\theta) &= \mathbb{E}\left[m_\theta^{(\mathbf{T})}\right] && \text{(by optional-stopping)} \\
&= \mathbb{E}\left[\exp(-\theta\cdot 0)\cdot \prod_{j=0}^{\mathbf{T}-1}\frac{1}{g_{X^{(j)}}(\theta)}\right] \\
&= \mathbb{E}\left[\prod_{j=0}^{\mathbf{T}-1}\frac{1}{g_{X^{(j)}}(\theta)}\right]
\end{aligned}
$$

We show that by taking, on both sides, the derivative with respect to θ we get

$$
\exp(-\theta) \le \sum_{i=1}^{\infty} i\cdot P(\mathbf{T}=i)\cdot \frac{g'_{1,\theta}(\theta)}{g_{0,\theta}(\theta)^{i+1}}, \tag{13}
$$

where $g_{0,\theta} = g_X$ and $g_{1,\theta} = g_Y$ for some $X, Y \in \Gamma$ possibly depending on θ. Indeed, we have:

$$
\begin{aligned}
\exp(-\theta) &= -\frac{d}{d\theta}\mathbb{E}\left[\prod_{j=0}^{\mathbf{T}-1}\frac{1}{g_{X^{(j)}}(\theta)}\right] \\
&= -\frac{d}{d\theta}\int_{w\in Run(X_0)} \prod_{j=0}^{\mathbf{T}(w)-1}\frac{1}{g_{X^{(j)}(w)}(\theta)}\,dP \\
&= \int_{w\in Run(X_0)} \sum_{i=0}^{\mathbf{T}(w)-1}\frac{g'_{X^{(i)}(w)}(\theta)}{g_{X^{(i)}(w)}(\theta)^2}\prod_{j\in\{0,\dots,\mathbf{T}(w)-1\}\setminus\{i\}}\frac{1}{g_{X^{(j)}(w)}(\theta)}\,dP \\
&\le \int_{w\in Run(X_0)} \mathbf{T}(w)\frac{g'_{1,\theta}(\theta)}{g_{0,\theta}(\theta)^{\mathbf{T}(w)+1}}\,dP \\
&= \sum_{i=1}^{\infty} i\cdot P(\mathbf{T}=i)\cdot\frac{g'_{1,\theta}(\theta)}{g_{0,\theta}(\theta)^{i+1}}
\end{aligned}
$$

The integral in this computation could in fact be replaced by a countable sum over terminating runs, because Δ is almost surely terminating. This justifies the exchange of the derivative and the integral in the third line of the previous computation. Thus (13) follows.

The following lemma bounds an "upper" subseries of the right-hand-side of (13).

Lemma 4.14. *For all $\varepsilon > 0$ there is $a \in \mathbb{N}$ such that for all $n \in \mathbb{N}$ and $\theta = 1/\sqrt{n}$ we have*

$$
\sum_{i=an+1}^{\infty} i\cdot P(\mathbf{T}=i)\cdot\frac{g'_{1,\theta}(\theta)}{g_{0,\theta}(\theta)^{i+1}} \le \varepsilon.
$$

Proof of the lemma. By rearranging the series we get for all $n \in \mathbb{N}$ and $\theta = 1/\sqrt{n}$:

$$
\begin{aligned}
&\sum_{i=an+1}^{\infty} i\cdot P(\mathbf{T}=i)\cdot\frac{g'_{1,\theta}(\theta)}{g_{0,\theta}(\theta)^{i+1}} \\
&\le \sum_{i=0}^{an-1}\frac{P(\mathbf{T}>an)\cdot g'_{1,\theta}(\theta)}{g_{0,\theta}(\theta)^{an+2}} + \sum_{i=an}^{\infty}\frac{P(\mathbf{T}>i)\cdot g'_{1,\theta}(\theta)}{g_{0,\theta}(\theta)^{i+2}} \\
&\le \underbrace{\frac{an\cdot P(\mathbf{T}>an)\cdot g'_{1,\theta}(\theta)}{g_{0,\theta}(\theta)^{an}}}_{=:q_1} + \underbrace{\sum_{i=an}^{\infty}\frac{P(\mathbf{T}>i)\cdot g'_{1,\theta}(\theta)}{g_{0,\theta}(\theta)^{i}}}_{=:q_2}
\end{aligned}
$$

We bound q_1 and q_2 separately. By Proposition 4.4 there is $c_1 > 0$ such that $P(\mathbf{T} > k) \le c_1/\sqrt{k}$. Hence we have, using Lemma 4.13 (d), (e):

$$
\begin{aligned}
q_1 &\le \frac{\sqrt{an}\cdot c_1\cdot c_2/\sqrt{n}}{c_3^a} = \frac{c_1c_2\sqrt{a}}{c_3^a}, && \text{and similarly,}\\
q_2 &\le \frac{c_1}{\sqrt{an}}\cdot\frac{c_2}{\sqrt{n}}\cdot\sum_{i=an}^{\infty}\frac{1}{g_{0,\theta}(\theta)^i}\\
&= \frac{c_1c_2}{\sqrt{a}\cdot n\cdot g_{0,\theta}(\theta)^{an}\cdot(1-1/g_{0,\theta}(\theta))}\\
&\le \frac{c_1c_2c_4}{\sqrt{a}\cdot c_3^a} && \text{(by Lemma 4.13 (e), (f)).}
\end{aligned}
$$

These bounds on q_1 and q_2 can be made arbitrarily small by choosing a large enough. This completes the proof of the lemma. □

This lemma implies a first lower bound on the distribution of $\mathbf{T}$:

Lemma 4.15. *For any $c > 0$ there is $s \in \mathbb{N}$ such that for all $n \in \mathbb{N}$ we have:*

$$\sum_{i=1}^{sn} i\cdot P(\mathbf{T}=i) \ge c\sqrt{n}.$$

Proof of the lemma. Let $a \in \mathbb{N}$ be the number from Lemma 4.14 for $\varepsilon = \exp(-1)/2$. For all $n \in \mathbb{N}$ and $\theta = 1/\sqrt{n}$ we have:

$$
\begin{aligned}
& g'_{1,\theta}(\theta)\cdot\sum_{i=1}^{an} i\cdot P(\mathbf{T}=i)\\
&\ge \sum_{i=1}^{an} i\cdot P(\mathbf{T}=i)\cdot\frac{g'_{1,\theta}(\theta)}{g_{0,\theta}(\theta)^{i+1}}\\
&\ge \exp(-\theta)-\varepsilon && \text{(by (13) and Lemma 4.14)}\\
&\ge \exp(-1)-\varepsilon=\varepsilon && \text{(by the choice of } \varepsilon\text{),}
\end{aligned}
$$

so, with Lemma 4.13 (d) we have for all $n \in \mathbb{N}$:

$$\sum_{i=1}^{an} i\cdot P(\mathbf{T}=i) \ge \frac{\varepsilon}{c_2}\sqrt{n}.$$

For the given number $c > 0$, choose $s := a\lceil cc_2/\varepsilon\rceil^2$. Then it follows for all $m \in \mathbb{N}$:

$$\sum_{i=1}^{sm} i\cdot P(\mathbf{T}=i) \ge c\sqrt{m},$$

which proves the lemma. □

Now we can complete the proof of the proposition. By Proposition 4.4 there is $c_1 > 0$ such that $P(\mathbf{T} > n) \le c_1/\sqrt{n}$ for all $n \in \mathbb{N}$. By Lemma 4.15, there is $s \in \mathbb{N}$ such that

$$\sum_{i=1}^{sn} i\cdot P(\mathbf{T}=i) \ge (2c_1+2)\sqrt{n} \quad \text{for all } n \in \mathbb{N}.$$

We have for all $n \in \mathbb{N}$:

$$
\begin{aligned}
\sum_{i=n}^{sn} i\cdot P(\mathbf{T}=i) &\ge \sum_{i=1}^{sn} i\cdot P(\mathbf{T}=i) - \sum_{i=1}^{n} i\cdot P(\mathbf{T}=i)\\
&\ge (2c_1+2)\sqrt{n} - \sum_{i=0}^{n} P(\mathbf{T}>i) && \text{(by the choice of } s \text{ above)}\\
&\ge (2c_1+2)\sqrt{n} - 1 - \sum_{i=1}^{n}\frac{c_1}{\sqrt{i}} && \text{(by the choice of } c_1 \text{ above)}
\end{aligned}
$$

$$\geq (2c_1+1)\sqrt{n} - \int_0^n \frac{c_1}{\sqrt{i}}\,di$$
$$= (2c_1+1)\sqrt{n} - 2c_1\sqrt{n}$$
$$= \sqrt{n}$$

It follows:

$$snP(\mathbf{T}\geq n) \geq sn\sum_{i=n}^{sn} P(\mathbf{T}=i) \geq \sum_{i=n}^{sn} i\cdot P(\mathbf{T}=i)$$
$$\geq \sqrt{n} \qquad \text{(by the computation above)}$$

Hence we have

$$P(\mathbf{T}\geq n) \geq \frac{1/s}{\sqrt{n}},$$

which completes the proof of the proposition. □

4.3. Proof of Proposition 4.6

Here is a restatement of Proposition 4.6.

Proposition 4.6. *Let Δ_h be the pBPA with $\Gamma_h = \{X_1,\ldots,X_h\}$ and the following rules:*

$$X_h \overset{1/2}{\hookrightarrow} X_hX_h,\quad X_h \overset{1/2}{\hookrightarrow} X_{h-1},\quad \ldots,\quad X_2 \overset{1/2}{\hookrightarrow} X_2X_2,\quad X_2 \overset{1/2}{\hookrightarrow} X_1,\quad X_1 \overset{1/2}{\hookrightarrow} X_1X_1,\quad X_1 \overset{1/2}{\hookrightarrow} \varepsilon$$

Then $[X_h]=1$, $E[X_h]=\infty$, and there is $c_h>0$ with

$$\frac{c_h}{n^{1/2^h}} \leq P(\mathbf{T}_{X_h}\geq n) \quad \textit{for all } n\in\mathbb{N}.$$

Proof. Observe that the third statement implies the second statement, since

$$E[X_h] = \sum_{n=1}^{\infty} P(\mathbf{T}_{X_h}\geq n) \geq \sum_{n=1}^{\infty} c_h\cdot n^{-1/2^h} \geq \sum_{n=1}^{\infty} c_h/n = \infty.$$

We proceed by induction on h. Let $h=1$. The pBPA Δ_1 is equivalent to a random walk on $\{0,1,2,\ldots\}$, started at 1, with an absorbing barrier at 0. It is well-known (see, e.g., [10]) that the probability that the random walk finally reaches 0 is 1, but that there is $c_1>0$ such that the probability that the random has not reached 0 after n steps is at least $c_1/\sqrt{n}$. Hence $[X_1]=1$ and $P(\mathbf{T}_{X_1}\geq n)\geq c_1/\sqrt{n} = c_1\cdot n^{-1/2}$.

Let $h>1$. The behavior of Δ_h can be described in terms of a random walk W_h whose states correspond to the number of X_h-symbols in the stack. Whenever an X_h-symbol is on top of the stack, the total number of X_h-symbols in the stack increases by 1 with probability 1/2, or decreases by 1 with probability 1/2, very much like the random walk equivalent to Δ_1. In the second case (i.e., the rule $X_h \overset{1/2}{\hookrightarrow} X_{h-1}$ is taken), the random walk W_h resumes only after a run of Δ_{h-1} (started with a single X_{h-1}-symbol) has terminated. By the induction hypothesis, $[X_{h-1}]=1$, so with probability 1 all spawned "sub-runs" of Δ_{h-1} terminate. Since W_h also terminates with probability 1, it follows $[X_h]=1$.

It remains to show that there is $c_h>0$ with $P(\mathbf{T}_{X_h}\geq n)\geq c_h\cdot n^{-1/2^h}$ for all $n\geq 1$. Consider, for any $n\geq 1$ and any $\ell>0$, the event A_ℓ that W_h needs at least ℓ steps to terminate (not counting the steps of the spawned sub-runs) and that at least one of the spawned sub-runs needs at least n steps to terminate. Clearly, $\mathbf{T}_{X_h}(w)\geq n$ holds for all $w\in A_\ell$, so it suffices to find $c_h>0$ so that for all $n\geq 1$ there is $\ell>0$ with $P(A_\ell)\geq c_h\cdot n^{-1/2^h}$. At least half of the steps of W_h are steps down, so whenever W_h needs at least 2ℓ steps to terminate, it spawns at least ℓ sub-runs. It follows:

$$P(A_\ell) \geq P(W_h \text{ needs at least } 2\ell \text{steps})\cdot\left(1-\left(P(\mathbf{T}_{X_{h-1}}<n)\right)^\ell\right)$$
$$\geq \frac{c_1}{\sqrt{2\ell}}\cdot\left(1-\left(1-c_{h-1}\cdot n^{-1/2^{h-1}}\right)^\ell\right) \qquad \text{(by induction hypothesis)}$$

Now we fix $\ell := n^{1/2^{h-1}}$. Then the second factor of the product above converges to $1-e^{-c_{h-1}}$ for $n\to\infty$, so for large enough n

$$P(A_\ell) \geq \frac{c_1}{2}\cdot\left(1-e^{-c_{h-1}}\right)\cdot n^{-1/2^h}.$$

Hence, we can choose $c_h < \frac{c_1}{2}\cdot(1-e^{-c_{h-1}})$ such that $P(A_\ell)\geq c_h\cdot n^{-1/2^h}$ holds for all $n\geq 1$. □

5. Conclusions and future work

We have provided a reduction from stateful to stateless pPDAs which gives new insights into the theory of pPDAs and at the same time simplifies it substantially. We have used this reduction and martingale theory to exhibit a dichotomy result that precisely characterizes the distribution of the termination time in terms of its expected value.

Although the bounds presented in this paper are asymptotically optimal, there is still space for improvements. We conjecture that our results can be extended to more general reward-based models, where each configuration is assigned a nonnegative reward and the total reward accumulated in a given service is considered instead of its length. This is particularly challenging if the rewards are unbounded (for example, the reward assigned to a given configuration may correspond to the total memory allocated by the procedures in the current call stack). Full answers to these questions would generalize some of the existing deep results about simpler models, and probably reveal an even richer underlying theory of pPDAs which is still undiscovered.

Acknowledgment

The authors thank Javier Esparza for useful suggestions.

References

[1] A. Berman, R.J. Plemmons, Nonnegative Matrices in the Mathematical Sciences, Academic Press, 1979.
[2] D. Bini, G. Latouche, B. Meini, Numerical Methods for Structured Markov Chains, Oxford University Press, 2005.
[3] T. Brázdil, Verification of probabilistic recursive sequential programs, PhD thesis, Masaryk University, Faculty of Informatics, 2007.
[4] T. Brázdil, V. Brožek, J. Holeček, A. Kučera, Discounted properties of probabilistic pushdown automata, in: Proceedings of LPAR 2008, in: Lecture Notes in Computer Science, vol. 5330, Springer, 2008, pp. 230–242.
[5] T. Brázdil, V. Brožek, K. Etessami, One-counter stochastic games, in: Proceedings of FST&TCS 2010, in: Leibniz International Proceedings in Informatics, vol. 8, Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2010, pp. 108–119.
[6] T. Brázdil, V. Brožek, K. Etessami, A. Kučera, D. Wojtczak, One-counter Markov decision processes, in: Proceedings of SODA 2010, SIAM, 2010, pp. 863–874.
[7] T. Brázdil, J. Esparza, A. Kučera, Analysis and prediction of the long-run behavior of probabilistic sequential programs with recursion, in: Proceedings of FOCS 2005, IEEE Computer Society Press, 2005, pp. 521–530.
[8] T. Brázdil, S. Kiefer, A. Kučera, Efficient analysis of probabilistic programs with an unbounded counter, in: Proceedings of CAV 2011, in: Lecture Notes in Computer Science, vol. 6806, Springer, 2011, pp. 208–224.
[9] J. Canny, Some algebraic and geometric computations in PSPACE, in: Proceedings of STOC'88, ACM Press, 1988, pp. 460–467.
[10] K.L. Chung, Markov Chains with Stationary Transition Probabilities, Springer, 1967.
[11] D.P. Dubhashi, A. Panconesi, Concentration of Measure for the Analysis of Randomized Algorithms, Cambridge University Press, 2009.
[12] J. Esparza, S. Kiefer, M. Luttenberger, Convergence thresholds of Newton's method for monotone polynomial equations, in: STACS 2008, 2008, pp. 289–300.
[13] J. Esparza, S. Kiefer, M. Luttenberger, Computing the least fixed point of positive polynomial systems, SIAM J. Comput. 39 (6) (2010) 2282–2335.
[14] J. Esparza, A. Kučera, R. Mayr, Model-checking probabilistic pushdown automata, in: Proceedings of LICS 2004, IEEE Computer Society Press, 2004, pp. 12–21.
[15] J. Esparza, A. Kučera, R. Mayr, Quantitative analysis of probabilistic pushdown automata: expectations and variances, in: Proceedings of LICS 2005, IEEE Computer Society Press, 2005, pp. 117–126.
[16] K. Etessami, D. Wojtczak, M. Yannakakis, Quasi-birth–death processes, tree-like QBDs, probabilistic 1-counter automata, and pushdown systems, in: Proceedings of 5th Int. Conf. on Quantitative Evaluation of Systems, QEST'08, IEEE Computer Society Press, 2008.
[17] K. Etessami, M. Yannakakis, Algorithmic verification of recursive probabilistic systems, in: Proceedings of TACAS 2005, in: Lecture Notes in Computer Science, vol. 3440, Springer, 2005, pp. 253–270.
[18] K. Etessami, M. Yannakakis, Checking LTL properties of recursive Markov chains, in: Proceedings of 2nd Int. Conf. on Quantitative Evaluation of Systems, QEST'05, IEEE Computer Society Press, 2005, pp. 155–165.
[19] K. Etessami, M. Yannakakis, Recursive Markov chains, stochastic grammars, and monotone systems of nonlinear equations, J. Assoc. Comput. Mach. 56 (2009).
[20] T.E. Harris, The Theory of Branching Processes, Springer, 1963.
[21] J.E. Hopcroft, J.D. Ullman, Introduction to Automata Theory, Languages, and Computation, Addison–Wesley, 1979.
[22] S. Kiefer, M. Luttenberger, J. Esparza, On the convergence of Newton's method for monotone systems of polynomial equations, in: STOC 2007, 2007, pp. 217–226.
[23] G. Latouche, V. Ramaswami, Introduction to Matrix Analytic Methods in Stochastic Modeling, ASA–SIAM Series on Statistics and Applied Probability, 1999.
[24] C. Manning, H. Schütze, Foundations of Statistical Natural Language Processing, MIT Press, 1999.
[25] R. Motwani, P. Raghavan, Randomized Algorithms, Cambridge University Press, 2006.
[26] A.G. Pakes, Some limit theorems for the total progeny of a branching process, Adv. Appl. Probab. 3 (1) (1971) 176–192.
[27] M.P. Quine, W. Szczotka, Generalisations of the Bienayme–Galton–Watson branching process via its representation as an embedded random walk, Ann. Appl. Probab. 4 (4) (1994) 1206–1222.
[28] D. Williams, Probability with Martingales, Cambridge University Press, 1991.