FI MU

Faculty of Informatics
Masaryk University Brno

On the Memory Consumption of Probabilistic Pushdown Automata

by

Tomáš Brázdil
Javier Esparza
Stefan Kiefer

FI MU Report Series **FIMU-RS-2009-07**

 October 2009

Publications in the FI MU Report Series are in general accessible via WWW:

`http://www.fi.muni.cz/reports/`

Further information can be obtained by contacting:

Faculty of Informatics
Masaryk University
Botanická 68a
602 00 Brno
Czech Republic

On the Memory Consumption of Probabilistic Pushdown Automata

Tomáš Brázdil*
Faculty of Informatics,
Masaryk University,
Czech Republic
xbrazdil@fi.muni.cz

Javier Esparza
Institut für Informatik,
Technische Universität München,
Germany
esparza@in.tum.de

Stefan Kiefer
Institut für Informatik,
Technische Universität München,
Germany
kiefer@in.tum.de

October 2, 2009

Abstract

We investigate the problem of evaluating memory consumption for systems modelled by probabilistic pushdown automata (pPDA). The space needed by a run of a pPDA is the maximal height reached by the stack during the run. The problem is motivated by the investigation of depth-first computations that play an important role for space-efficient schedulings of multithreaded programs.

We study the computation of both the distribution of the memory consumption and its expectation. For the distribution, we show that a naive method incurs an exponential blow-up, and that it can be avoided using linear equation systems. We also suggest a possibly even faster approximation method. Given $\varepsilon > 0$, these methods allow to compute bounds on the memory consumption that are exceeded with a probability of at most ε.

*Supported by the research center Institute for Theoretical Computer Science (ITI), project No. 1M0545.

For the expected memory consumption, we show that whether it is infinite can be decided in polynomial time for stateless pPDA (pBPA) and in polynomial space for pPDA. We also provide an iterative method for approximating the expectation. We show how to compute error bounds of our approximation method and analyze its convergence speed. We prove that our method converges linearly, i.e., the number of accurate bits of the approximation is a linear function of the number of iterations.

1 Introduction

The verification of probabilistic programs with possibly recursive procedures has been intensely studied in the last years. The Markov chains or Markov Decision Processes underlying these systems may have infinitely many states. Despite this fact, which prevents the direct application of the rich theory of finite Markov Chains, many positive results have been obtained. Model-checking algorithms have been proposed for both linear and branching temporal logics [11, 16, 24, 20, 35], the long-run behavior of the systems has been analyzed [10, 17], and algorithms deciding properties of several kinds of games have been described [8, 9, 19, 21, 22, 23].

In all these papers programs are modelled as probabilistic pushdown automata (pPDA) or as recursive Markov chains; the two models are very close, and nearly all results obtained for one of them can be easily translated to the other [14]. In this paper we consider pPDA. Loosely speaking, a pPDA is a pushdown automaton whose transitions carry probabilities. The *configurations* of a pPDA are pairs containing the current control state and the current stack content. A *run* is a sequence of configurations, each one obtained from its predecessor by applying a transition, which may modify the control state and modify the top of the stack. If a run reaches a configuration with empty stack, it stays in this configuration forever. We say "it terminates".

The memory consumption of a pPDA is modelled by the random variable M that assigns to a run the *maximal* stack height of the configurations visited along it (which may be infinite). We study the distribution and the expected value of M. The execution time and memory consumption of pPDA were studied in [17], but the results about the latter were much weaker. More precisely, all it was shown in [17] was that $\mathcal{P}(M = \infty)$ can be compared with 0 or 1 in polynomial space and with other rationals in exponential time.

A probabilistic recursive program whose variables have finite range can be modelled by a pPDA, and in this case M models the amount of memory needed for the recursion stack. But M is also an important parameter for the problem of scheduling multithreaded computations (see [29, 5, 3, 1] among other papers). When a multithreaded program is executed by one processor, a scheduler decides which thread to execute next, and the current states of all other active threads are stored. When threads are lightweight, this makes the memory requirements of the program heavily depend on the thread scheduler [29]. Under the usual assumption that a thread can terminate only after all threads spawned by it terminate, the space-optimal scheduler is the one that, when A spawns B, interrupts the execution of A and continues with B; this is called the *depth-first scheduler* in [29, 5]. The depth-first scheduler can be modelled by a pushdown automaton. To give an example, consider a multithreaded system with two types of threads, X and Y. Imagine that through statistical sampling we know that a thread of type X spawns either a thread of type Y or no new threads, both with probability $1/2$; a thread of type Y spawns another thread of type Y with probability $1/3$, or no new thread with probability $2/3$. The depth-first execution of this multithreaded program is modelled by a pPDA with one control state, stack symbols X, Y, and rules $X \overset{1/2}{\to} YX$, $X \overset{1/2}{\to} \varepsilon$, $Y \overset{1/3}{\to} YY$, $Y \overset{2/3}{\to} \varepsilon$. Notice that the rule $X \overset{1/2}{\to} YX$ indeed models the depth-first idea: the new thread Y is executed before the thread X.

In this simple model, pPDA for multithreaded systems have one single control state. Stack symbols represent currently active threads and pushdown rules model whether a thread dies or spawns a new thread. On the other hand, pPDA with more than one control state can model global variables with finite range (the possible values of the global store are encoded into the control states of the pPDA) [6]. For these reasons we study arbitrary pPDA in this paper, but also specialize our results (and in particular the complexity of algorithms) to so-called pBPA, which are pPDA with a single control state. As we shall see, while some algorithms are polynomial for pBPA, this is unlikely to be the case for pPDA.

Our contribution. We specifically address the problem of computing $\mathcal{P}(M \geq n)$, or at least an upper bound, for a given n. This allows us to determine the size that the stack (or the store for threads awaiting execution) must have in order to guarantee that the probability of a memory overflow does not exceed a given bound. In Section 3 we obtain a system of recurrence equations for $\mathcal{P}(M \geq n)$, and show that for a pPDA with set Q of control states and set Γ of stack symbols, $\mathcal{P}(M \geq n)$ can be computed in time

$\mathcal{O}\left(n \cdot (|Q|^2 \cdot |\Gamma|)^3\right)$ (time $\mathcal{O}\left(n \cdot |\Gamma|^3\right)$ for pBPAs) in the Blum-Shub-Smale model, the computation model in which an arithmetic operation takes one time unit, independently of the size of the operands. However, this result does not provide any information on the asymptotic behavior of $\mathcal{P}(M \geq n)$ when n grows, and moreover the algorithm is computationally inefficient for large values of n. We address these problems for pPDA in which the expected value of M is finite. We show in Section 3.2 that in this case $\mathcal{P}(M \geq n) \in \Theta(\rho^n)$, where $\rho < 1$ is the spectral radius of a certain matrix. This power law etermines the exact asymptotic behavior up to a constant, and also leads to an algorithm for computing a bound on $\mathcal{P}(M \geq n)$ with logarithmic runtime in n.

Then we turn to computing the expectation of M. In Section 3.3 we provide an algorithm that approximates the expectation, give an error bound and analyze its convergence speed. Finally, in Section 4 we study the problem of deciding whether the expected value of M is finite. We show that the problem is polynomial for pBPAs. For arbitrary pPDA we show that the problem is in PSPACE and at least as hard as the SQRT-SUM and PosSLP problems. Notice that already the problem of deciding if the termination probability of a pPDA exceeds a given bound has this same complexity.

Related work. Much work has been done also on the analysis of other well-structured infinite-state Markov chains, such as quasi-birth-death processes [31], Jackson queueing networks [32] and probabilistic lossy channel systems [34]. However, none of these classes contain pPDA or even pBPA. (The model of quasi-birth-death processes is close to ours, it roughly corresponds to pPDAs with one single stack symbol.) There is also work on general infinite-state (continuous-time) Markov chains which analyzes the behavior of the chain up to a finite depth [36, 25]. This method is very general, but it is inefficient for pPDA, because it has not been designed to exploit the pushdown structure. Our analysis techniques are strongly based on linear algebra and matrix theory, in particular Perron-Frobenius theory [4]. The closest work to ours is [24] which also uses Perron-Frobenius theory for assessing the termination probability of recursive Markov chains.

2 Preliminaries

In the rest of this paper, $\mathbb{N}$ and $\mathbb{R}$ denote the set of positive integers and real numbers, respectively. 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. Given two sets $K \subseteq \Sigma^*$ and $L \subseteq \Sigma^* \cup \Sigma^\omega$, we use $K \cdot L$ (or just KL) to denote the concatenation of K and L, i.e., $KL = \{ww' \mid w \in K, w' \in L\}$. 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 $w(0), w(1), \ldots$

Vectors and Matrices. Given a set S, we regard the elements of $\mathbb{R}^S$ as *vectors*. We use bold letters, like $\mathbf{u}$, for vectors. The vector whose components are all 0 (resp. all 1) is denoted by $\mathbf{0}$ (resp. $\mathbf{1}$). We write $\mathbf{u} = \mathbf{v}$ (resp. $\mathbf{u} \leq \mathbf{v}$) if $\mathbf{u}(s) = \mathbf{v}(s)$ (resp. $\mathbf{u}(s) \leq \mathbf{v}(s)$) holds for all $s \in S$. If $S' \subseteq S$, we write $\mathbf{u}_{|S'}$ for the restriction of $\mathbf{u}$ on S', i.e., the vector $\mathbf{v} \in \mathbb{R}^{S'}$ with $\mathbf{v}(s) = \mathbf{u}(s)$ for all $s \in S'$.

Given two vector spaces $\mathbb{R}^S, \mathbb{R}^T$ we identify a linear function $A : \mathbb{R}^S \to \mathbb{R}^T$ with its corresponding matrix $A \in \mathbb{R}^{T \times S}$. We use capital letters for matrices and I for the identity matrix. We call a matrix nonnegative if all its entries are nonnegative. For nonnegative square matrices $A \in \mathbb{R}^{S \times S}$ we define the matrix sum $A^* = \sum_{i=0}^{\infty} A^i = I + A + AA + \cdots$. It is a well-known fact (see e.g. [26]) that A^* converges (or "exists") iff $\rho(A) < 1$, where $\rho(A)$ denotes the spectral radius of A, i.e., the largest absolute value of the eigenvalues of A. Perron-Frobenius theory for nonnegative matrices (see e.g. [4]) states that $\rho(A)$ is an eigenvalue of A. If A^* exists, then $A^* = (I - A)^{-1}$.

Markov Chains. Our models of interest induce (infinite-state) Markov chains.

Definition 2.1. *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* $s \to t$ *of* M *assigns its probability* $Prob(s \to t) > 0$ *so that for every* $s \in S$ *we have* $\sum_{s \to t} Prob(s \to t) = 1$ *(as usual, we write* $s \xrightarrow{x} t$ *instead of* $Prob(s \to t) = x$*).*

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 $1 \leq i < |w|$. 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 Run (or $Run(w)$) instead of $Run[M]$ (or $Run[M](w)$, resp.).

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 $\mathcal{P}(Run(w)) = \Pi_{i=1}^{|w|-1} x_i$ where $w(i-1) \xrightarrow{x_i} w(i)$ for every $1 \leq i < |w|$. If $|w| = 1$, we put $\mathcal{P}(Run(w)) = 1$. 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 $\mathcal{F}$. Given $s \in S$ and $A \subseteq S$, we say A *is reachable from* s if $\mathcal{P}(\{w \in Run(s) \mid \exists i \geq 0 : w(i) \in A\}) > 0$.

Probabilistic Pushdown Automata (pPDA).

Definition 2.2. *A* probabilistic pushdown automaton (pPDA) *is a tuple* $\Delta = (Q, \Gamma, \delta, Prob)$ *where* Q *is a finite set of* control states, Γ *is a finite* stack alphabet, $\delta \subseteq Q \times \Gamma \times Q \times \Gamma^{\leq 2}$ *(where* $\Gamma^{\leq 2} = \{\alpha \in \Gamma^*, |\alpha| \leq 2\}$*) is a* transition relation*, and Prob is a function which to each transition* $pX \to q\alpha$ *assigns a rational probability* $Prob(pX \to q\alpha) > 0$ *so that for all* $p \in Q$ *and* $X \in \Gamma$ *we have that* $\sum_{pX \to q\alpha} Prob(pX \to q\alpha) = 1$ *(as usual, we write* $pX \stackrel{x}{\to} q\alpha$ *instead of* $Prob(pX \to q\alpha) = x$*).*

Elements of $Q \times \Gamma^*$ are called *configurations* of Δ. A pPDA with just one control state is called pBPA (pBPAs correspond to 1-exit recursive Markov chains defined in [24]). In what follows, configurations of pBPAs are usually written without the control state (i.e., we write only α instead of $p\alpha$).

Example 2.3. *As a running example we choose the pBPA* $\Delta = (\{p\}, \{X, Y, Z, W\}, \delta, Prob)$ *with* δ *and Prob given as follows.*

$$\begin{array}{llll} X \stackrel{1/4}{\to} \varepsilon & X \stackrel{1/4}{\to} Y & Y \stackrel{2/3}{\to} \varepsilon & Z \stackrel{1}{\to} Z \\ X \stackrel{1/4}{\to} XX & X \stackrel{1/4}{\to} Z & Y \stackrel{1/3}{\to} YY & W \stackrel{1}{\to} YW \end{array}$$

We can interpret this example as a model of a multithreaded system with four kinds of threads. Notice that threads of type Z *and* W *do not terminate (our results also deal with this possibility). We are interested in the minimal number of threads* n *such that the probability that the execution of* X *requires to store more than* n *threads is at most* 10^{-5}*.*

We define the size $|\Delta|$ of a pPDA Δ as follows: $|\Delta| = |Q| + |\Gamma| + |\delta| + |Prob|$ where $|Prob|$ equals the sum of sizes of binary representations of values of *Prob*. To Δ we associate the Markov chain M_Δ with $Q \times \Gamma^*$ as set of states and transitions defined as follows:

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

Given $p, q \in Q$ and $X \in \Gamma$, we often write pXq to denote (p, X, q). Given pXq we define

$$Run(pXq) = \{w \in Run(pX) \mid \exists i \geq 0 : w(i) = q\varepsilon\} \quad \text{and} \quad [pXq] = \mathcal{P}(Run(pXq)) .$$

Maximal Stack Height. Given $p\alpha \in Q \times \Gamma^*$, we denote by $\mathit{height}(p\alpha) = |\alpha|$ the stack height of $p\alpha$. Given $pX \in Q \times \Gamma$, the maximal stack height of a run is defined by setting

$$M_{pX}(w) = \sup\{\mathit{height}(w(i)) \mid i \geq 0\} \quad \text{for all runs } w \in \mathit{Run}(pX).$$

It is easy to show that for all $n \in \mathbb{N} \cup \{\infty\}$ the set $M_{pX}^{-1}(n) = \{w \in \mathit{Run}(pX) \mid M_{pX}(w) = n\}$ is measurable. Hence the expectation EM_{pX} of M_{pX} exists and we have

$$EM_{pX} = \sum_{n \in \mathbb{N} \cup \{\infty\}} n \cdot \mathcal{P}(M_{pX}^{-1}(n)) .$$

For what follows, we fix a pPDA $\Delta = (Q, \Gamma, \delta, \mathit{Prob})$ with initial configuration $p_0X_0 \in Q \times \Gamma$. We are interested in the random variable $M_{p_0X_0}$ modelling the memory consumption of Δ. More concretely, we wish to compute or approximate the distribution of $M_{p_0X_0}$ and its expectation.

3 Computing the Memory Consumption

The problem of computing the distribution of the maximal stack height is the problem of computing the probability of reaching a given height. So, for every $n \geq 1$ we define a vector $\mathbf{P}[n] \in \mathbb{R}^{Q \times \Gamma}$ with

$$\mathbf{P}[n](pX) = \mathcal{P}(\{w \in \mathit{Run}(pX) \mid M_{pX}(w) \geq n\}) \quad \text{for every } pX \in Q \times \Gamma ,$$

i.e., $\mathbf{P}[n](pX)$ is the probability that the maximal stack height is $\geq n$ in a run of $\mathit{Run}(pX)$.

There is a "naive" method to compute $\mathbf{P}[n](p_0X_0)$. (Recall that M_Δ is the Markov chain associated with Δ.) First, compute the Markov chain M_Δ^{n+1} obtained from M_Δ by restricting it to the states with a height of at most $n+1$. Note that M_Δ^{n+1} has finitely many states. Then compute $\mathbf{P}[n](p_0X_0)$ by computing the probability of reaching a state of height $n+1$ starting from p_0X_0. This can be done as usual by solving a linear equation system. The problem with this approach is that the number of states in M_Δ^{n+1} is $\Theta(|Q| \cdot |\Gamma|^n)$, i.e., exponential in n, and the linear equation system has equally many equations.

A better algorithm is obtained by observing that the Markov chain induced by a pPDA has a certain regular structure. We exploit this to get rid of the state explosion in the "naive" method. (This has also been observed in the analysis of other structured infinite-state systems, see e.g. [31].) In the following we describe the improved method, which is based on linear recurrences. We are mainly interested in the probabilities $\mathbf{P}[n]$ to reach height n, but as an auxiliary quantity we use the probability of not exceeding height n in terminating runs. Formally, for every $n \geq 0$ we define a vector $\mathbf{T}[n] \in \mathbb{R}^{Q\times\Gamma\times Q}$ such that

$$\mathbf{T}[n](pXq) = \mathcal{P}(\{w \in Run(pXq) \mid M_{pX}(w) \leq n\}) \quad \text{for every } pXq \in Q \times \Gamma \times Q\,,$$

i.e., $\mathbf{T}[n](pXq)$ is the probability of all runs of $Run(pX)$ that terminate at q and do not exceed the height n. To every $pXq \in Q \times \Gamma \times Q$ we associate a variable $\mathbf{T}\langle n\rangle(pXq)$. Consider the following equation system: If $\mathbf{T}[n](pXq) = 0$, then we put $\mathbf{T}\langle n\rangle(pXq) = 0$. Otherwise, we put

$$\mathbf{T}\langle n\rangle(pXq) = \sum_{pX \xrightarrow{y} q\varepsilon} y + \sum_{pX \xrightarrow{y} rY} y\mathbf{T}\langle n\rangle(rYq) + \sum_{pX \xrightarrow{y} rYZ} \sum_{s\in Q} y\mathbf{T}[n-1](rYs)\mathbf{T}\langle n\rangle(sZq)\,.$$

Proposition 3.1. *For every $n \geq 0$, the vector $\mathbf{T}[n]$ is the* unique *solution of that equation system.*

The values $\mathbf{T}[n]$ can be used to set up an equation system for $\mathbf{P}[n]$. To every $pX \in Q \times \Gamma$ we associate a variable $\mathbf{P}\langle n\rangle(pX)$. Consider the following equation system: We put $\mathbf{P}\langle 1\rangle(pX) = 1$. If $\mathbf{P}[n](pX) = 0$, then we put $\mathbf{P}\langle n\rangle(pX) = 0$. Otherwise, we put

$$\mathbf{P}\langle n\rangle(pX) = \sum_{pX \xrightarrow{y} qY} y\mathbf{P}\langle n\rangle(qY) + \sum_{pX \xrightarrow{y} qYZ} y\mathbf{P}[n-1](qY) + \sum_{pX \xrightarrow{y} qYZ} \sum_{r\in Q} y\mathbf{T}[n-2](qYr)\mathbf{P}\langle n\rangle(rZ)\,.$$

Proposition 3.2. *For every $n \geq 1$, the vector $\mathbf{P}[n]$ is the* unique *solution of that equation system.*

Example 3.3. *In our example we have for $n \geq 1$*

$$\begin{aligned}
\mathbf{T}[n](X) &= 1/4 + 1/4\,\mathbf{T}[n](Y) + 1/4\,\mathbf{T}[n](Z) + 1/4\,\mathbf{T}[n-1](X)\mathbf{T}[n](X)\\
\mathbf{T}[n](Y) &= 2/3 + 1/3\,\mathbf{T}[n-1](Y)\mathbf{T}[n](Y)\\
\mathbf{T}[n](Z) &= 0\\
\mathbf{T}[n](W) &= 0
\end{aligned}$$

and for $n \geq 2$

$$\begin{aligned}
\mathbf{P}[n](X) &= 1/4\,\mathbf{P}[n](Y) + 1/4\,\mathbf{P}[n](Z) + 1/4\,\mathbf{P}[n-1](X) + 1/4\,\mathbf{T}[n-2](X)\mathbf{P}[n](X)\\
\mathbf{P}[n](Y) &= 1/3\,\mathbf{P}[n-1](Y) + 1/3\,\mathbf{T}[n-2](Y)\mathbf{P}[n](Y)\\
\mathbf{P}[n](Z) &= 0\\
\mathbf{P}[n](W) &= \mathbf{P}[n-1](Y) + \mathbf{T}[n-2](Y)\mathbf{P}[n](W)\,.
\end{aligned}$$

Solving those systems successively for increasing n *shows that* $n = 17$ *is the smallest number* n *such that* $\mathbf{P}[n](X) \leq 10^{-5}$*. In the interpretation as a multithreaded system this means that the probability that* 17 *or more threads need to be stored is at most* 10^{-5}*.*

Using the above equation systems, we can compute $\mathbf{T}[n]$ and $\mathbf{P}[n]$ iteratively for increasing n by plugging in the values obtained in earlier iterations. The cost of each iteration is dominated by solving the equation system for $\mathbf{T}[n]$, which can be done, using Gaussian elimination, in time $\mathcal{O}\left((|Q|^2 \cdot |\Gamma|)^3\right)$ in the Blum-Shub-Smale model. So the total time to compute $\mathbf{P}[n]$ is linear in n.

Proposition 3.4. *The value* $\mathbf{P}[n]$ *can be computed by setting up and solving the equation systems of Propositions 3.1 and 3.2 in time* $\mathcal{O}\left(n \cdot (|Q|^2 \cdot |\Gamma|)^3\right)$ *in the Blum-Shub-Smale model.*

The values $\mathbf{P}[n]$ that can be computed by Proposition 3.4 also allow to approximate the expectation $\mathrm{E}M_{p_0X_0}$: Since $\mathrm{E}Y = \sum_{n=1}^{\infty} \mathcal{P}(Y \geq n)$ holds for any random variable Y with values in $\mathbb{N}$, we have $\mathrm{E}M_{p_0X_0} = \sum_{n=1}^{\infty} \mathbf{P}[n](p_0X_0)$, so one can approximate $\mathrm{E}M_{p_0X_0}$ by computing $\sum_{n=1}^{k} \mathbf{P}[n](p_0X_0)$ for some finite k.

Proposition 3.4 is simple and effective, but not fully satisfying for several reasons. First, it does not indicate how fast $\mathbf{P}[n](p_0X_0)$ decreases (if at all) for increasing n. Second, although computing $\mathbf{P}[n]$ using Proposition 3.4 is more efficient than using the "naive" method, it may still be too costly for large n, especially if Q or Γ are large. Instead, one may prefer an upper bound on $\mathbf{P}[n]$ if it is fast to compute. Finally, we wish for an approximation method for $\mathrm{E}M_{p_0X_0}$ that comes with an error bound.

In the following we achieve these goals for pPDAs in which the expected memory consumption is finite. So we assume the following on the pPDA Δ for the rest of the section.

ASSUMPTION: The expectation $\mathrm{E}M_{p_0X_0}$ is finite.

Notice that from the practical point of view this is a mild assumption: systems with infinite expected memory consumption also have infinite expected running time, and are unlikely to be considered suitable in reasonable scenarios. In Section 4 we discuss the assumption in more detail. In particular, we show that whether $\mathrm{E}M_{p_0X_0}$ is finite can be decided in polynomial time for pBPA, but also that this problem is unlikely to be decidable in polynomial time for general pPDA.

3.1 The Matrix A

This subsection leads to a matrix A which is crucial for our analysis. It is useful to get rid of certain irregularities in the equation systems of Propositions 3.1 and 3.2. The following lemma shows that the variables in the equation systems do not change from 0 to positive (or from positive to 0) if n is sufficiently large. (Recall that, by definition, $\mathbf{T}[n] \le \mathbf{T}[n+1]$ and $\mathbf{P}[n] \ge \mathbf{P}[n+1]$ for all $n \ge 1$.)

Lemma 3.5.

1. $\mathbf{T}[|Q|^2|\Gamma|+1](pXq) > 0 \iff$ *for all* $n \ge |Q|^2|\Gamma|+1 : \mathbf{T}[n](pXq) > 0 \iff [pXq] > 0$*;*
2. $\mathbf{P}[|Q||\Gamma|+1](pX) > 0 \iff$ *for all* $n \ge 1 : \mathbf{P}[n](pX) > 0$*.*

Another irregularity can be removed by restricting $\mathbf{T}[n]$ and $\mathbf{P}[n]$ to their "interesting" components; in particular, we filter out entries of $\mathbf{P}[n]$ that cannot create large stacks. Let $\mathcal{T} \subseteq Q \times \Gamma \times Q$ denote the set of all pXq such that $pX\Gamma^*$ is reachable from p_0X_0, and $[pXq] > 0$. Let $\mathcal{H} \subseteq Q \times \Gamma$ denote the set of all pX such that $pX\Gamma^*$ is reachable from p_0X_0, and $\mathbf{P}[n](pX) > 0$ for all $n \ge 1$.

Lemma 3.6. *The sets $\mathcal{T}$ and $\mathcal{H}$ are computable in polynomial time.*

Example 3.7. *For our running example, we fix X as the initial configuration. Then $W\Gamma^*$ is not reachable and $\mathbf{P}[n](Z) = 0$ for $n \ge 2$, hence $\mathcal{H} = \{X, Y\}$. Furthermore, $\mathcal{T} = \{X, Y\}$.*

We define $\mathbf{t}[n] \in \mathbb{R}^{\mathcal{T}}$ by $\mathbf{t}[n] := \mathbf{T}[n]_{|\mathcal{T}}$, i.e., $\mathbf{t}[n] \in \mathbb{R}^{\mathcal{T}}$ is the restriction of $\mathbf{T}[n]$ to $\mathcal{T}$. Similarly, we define $\mathbf{p}[n] := \mathbf{P}[n]_{|\mathcal{H}}$. Now we bring the equation systems for $\mathbf{t}[n]$ and $\mathbf{p}[n]$ from Propositions 3.1 and 3.2 in a compact matrix form.

For $\mathbf{t}[n]$, we define a vector $\mathbf{c} \in \mathbb{R}^{\mathcal{T}}$, a linear function $\tilde{L}$ on $\mathbb{R}^{\mathcal{T}}$, and a bilinear function $\tilde{Q} : \mathbb{R}^{\mathcal{T}} \times \mathbb{R}^{\mathcal{T}} \to \mathbb{R}^{\mathcal{T}}$ as follows:

$$(\mathbf{c})(pXq) = \sum_{pX \xrightarrow{y} q\varepsilon} y \qquad\qquad (\tilde{L}\mathbf{v})(pXq) = \sum_{\substack{pX \xrightarrow{y} rY \\ rYq \in \mathcal{T}}} y\mathbf{v}(rYq)$$

$$(\tilde{Q}(\mathbf{u},\mathbf{v}))(pXq) = \sum_{pX \xrightarrow{y} rYZ} \sum_{\substack{s \in Q \\ rYs \in \mathcal{T} \\ sZq \in \mathcal{T}}} y\mathbf{u}(rYs)\mathbf{v}(sZq)$$

By $\tilde{Q}(\mathbf{u},\cdot)$ we denote a linear function satisfying $\tilde{Q}(\mathbf{u},\cdot)(\mathbf{v}) = \tilde{Q}(\mathbf{u},\mathbf{v})$.

For $\mathbf{p}[n]$, we define linear functions L and L' on $\mathbb{R}^{\mathcal{H}}$, and a bilinear function $Q : \mathbb{R}^{\mathcal{T}} \times \mathbb{R}^{\mathcal{H}} \to \mathbb{R}^{\mathcal{H}}$ as follows:

$$(L\mathbf{v})(pX) = \sum_{\substack{pX \xrightarrow{y} qY \\ qY \in \mathcal{H}}} y\mathbf{v}(qY) \qquad (L'\mathbf{v})(pX) = \sum_{\substack{pX \xrightarrow{y} qYZ \\ qY \in \mathcal{H}}} y\mathbf{v}(qY)$$

$$(Q(\mathbf{u}, \mathbf{v}))(pX) = \sum_{pX \xrightarrow{y} qYZ} \sum_{\substack{r \in Q \\ qYr \in \mathcal{T} \\ rZ \in \mathcal{H}}} y\mathbf{u}(qYr)\mathbf{v}(rZ)$$

By $Q(\mathbf{u}, \cdot)$ we denote a linear function satisfying $Q(\mathbf{u}, \cdot)(\mathbf{v}) = Q(\mathbf{u}, \mathbf{v})$.
Using Propositions 3.1 and 3.2 we obtain for $n \geq |Q|^2|\Gamma| + 3$ (recall Lemma 3.5):

Proposition 3.8. *The following equations hold for all* $n \geq |Q|^2|\Gamma| + 3$*:*

$$\mathbf{t}[n] = \mathbf{c} + \tilde{L}\mathbf{t}[n] + \tilde{Q}(\mathbf{t}[n-1], \mathbf{t}[n]) \quad \textit{and} \quad \mathbf{p}[n] = L\mathbf{p}[n] + L'\mathbf{p}[n-1] + Q(\mathbf{t}[n-2], \mathbf{p}[n])$$

Example 3.9. *In our example we have for* $n \geq 1$

$$\mathbf{t}[n] = \overbrace{\begin{pmatrix} 1/4\ \mathbf{t}[n-1](X) & 1/4 \\ 0 & 1/3\ \mathbf{t}[n-1](Y) \end{pmatrix}}^{\tilde{L}+\tilde{Q}(\mathbf{t}[n-1],\cdot)} \mathbf{t}[n] + \overbrace{\begin{pmatrix} 1/4 \\ 2/3 \end{pmatrix}}^{\mathbf{c}}$$

and for $n \geq 2$

$$\mathbf{p}[n] = \overbrace{\begin{pmatrix} 1/4\ \mathbf{t}[n-2](X) & 1/4 \\ 0 & 1/3\ \mathbf{t}[n-2](Y) \end{pmatrix}}^{L+Q(\mathbf{t}[n-2],\cdot)} \mathbf{p}[n] + \overbrace{\begin{pmatrix} 1/4 & 0 \\ 0 & 1/3 \end{pmatrix}}^{L'} \mathbf{p}[n-1] \ .$$

Unlike $\mathbf{P}[n]$, the vector $\mathbf{p}[n]$ can be expressed in the form $A_n\mathbf{p}[n-1]$ for a suitable matrix A_n:

Proposition 3.10. *Let* $A_n := (L + Q(\mathbf{t}[n-2], \cdot))^* L'$*. Then for every* $n \geq |Q|^2|\Gamma| + 3$ *the matrix* A_n *exists and* $\mathbf{p}[n] = A_n\mathbf{p}[n-1]$.

The key of our further analysis is to replace the matrix A_n by $A = \lim_{n\to\infty} A_n$. Since $A_n = (L + Q(\mathbf{t}[n-2], \cdot))^* L'$, we have

$$A := (L + Q(\mathbf{t}, \cdot))^* L'$$

where we define $\mathbf{t} = \lim_{n\to\infty} \mathbf{t}[n]$. (Observe that $\mathbf{t}(pXq) = [pXq]$.) It is not immediate from Proposition 3.10 that A exists, but it can be proved:

Proposition 3.11. *The matrix* A *exists and its spectral radius* ρ *satisfies* $\rho < 1$.

Proposition 3.11 is the technical core of this paper. Its proof is quite involved and relies on Perron-Frobenius theory [4]. We give a proof sketch and a full proof in Appendix B.7.

Example 3.12. *The termination probabilities* $\mathbf{t}$ *can be computed as the least solution of a non-linear equation system [16, 24]. Applied to our example we obtain* $\mathbf{t}(X) = 2 - \sqrt{2} \approx 0.586$ *and* $\mathbf{t}(Y) = 1$*. Basic computations yield the following matrix* A *whose spectral radius is* $\rho = 1/2$*.*

$$A = \begin{pmatrix} 1/(2+\sqrt{2}) & 1/(4+2\sqrt{2}) \\ 0 & 1/2 \end{pmatrix}$$

3.2 Approximating the Distribution and a Tail Bound

We can assume $p_0X_0 \in \mathcal{H}$ in the following, because otherwise, by Lemma 3.5, we would have $\mathbf{P}[n](p_0X_0) = 0$ for $n \geq |Q|^2|\Gamma| + 3$, removing any need for further analysis.
The following theorem suggests an efficient approximation algorithm.

Theorem 3.13. *Let* $n_\perp := |Q|^2|\Gamma|+3$ *and* $\widehat{\mathbf{p}}[n] := \mathbf{p}[n]$ *for* $n < n_\perp$ *and* $\widehat{\mathbf{p}}[n_\perp + n] := A^n\mathbf{p}[n_\perp]$ *for* $n \geq 0$*. Then* $\mathbf{p}[n] \leq \widehat{\mathbf{p}}[n]$ *holds for all* $n \geq 1$*. Moreover, there exists* d *with* $0 < d \leq 1$ *and*

$$d \cdot \widehat{\mathbf{p}}[n](p_0X_0) \leq \mathbf{p}[n](p_0X_0) \leq \widehat{\mathbf{p}}[n](p_0X_0) \ .$$

The proposition shows that $\mathbf{p}[n](p_0X_0)$ and the approximation $\widehat{\mathbf{p}}[n](p_0X_0)$ differ at most by a constant factor. Given A, the matrix powers A^n can be computed by repeated squaring, which allows to compute this upper bound in time $\mathcal{O}\left((|Q| \cdot |\Gamma|)^3 \cdot \log n\right)$ in the Blum-Shub-Smale model. To compute $A = (L + Q(\mathbf{t}, \cdot))^*L'$ itself, we can compute the matrix star via the matrix inverse, as stated in the preliminaries. Computing the vector $\mathbf{t}$ of termination probabilities requires a more detailed discussion. The vector is the least solution of a nonlinear equation system, and its components may be irrational and even non-expressible by radicals [16, 24]. However, there are several ways to compute at least upper bounds on $\mathbf{t}$ (which suffices to obtain upper bounds on $\mathbf{p}[n]$, as A depends monotonically on $\mathbf{t}$), or lower-bound approximations sufficiently accurate for all practical purposes. First, $\mathbf{t}$ can be approximated in polynomial space using binary search [24]. Second, one may use Newton's method and the results of [28, 15] on its convergence speed. These papers show that Newton's method converges at least linearly, and exponentially in many cases, i.e., the number of accurate bits grows exponentially in the number of iterations. Finally, one of the conditions systems must often satisfy is termination with probability 1. For pBPAs this condition can be checked in polynomial time [24]. If this test is passed, we know $\mathbf{t} = \mathbf{1}$, and the problem of computing $\mathbf{t}$ is solved.

Theorem 3.13 provides a tail bound for $\mathbf{p}[n](p_0X_0)$:

Corollary 3.14. *We have* $\mathbf{p}[n](p_0X_0) \in \Theta(\rho^n)$.

Example 3.15. *Since in our example Proposition 3.10 holds already for* $n \geq 2$*, we have* $\widehat{\mathbf{p}}[n] = A^{n-1}\mathbf{1}$ *for* $n \geq 1$*. With the matrix* A *from Example 3.12 and using* $\mathbf{p}[n] \leq \widehat{\mathbf{p}}[n]$ *we obtain:*

$$\mathbf{p}[2] \leq 0.5 \cdot \mathbf{1}\,, \quad \mathbf{p}[5] \leq 0.07 \cdot \mathbf{1}\,, \quad \mathbf{p}[17] \leq 10^{-4} \cdot \mathbf{1}\,, \quad \mathbf{p}[65] \leq 10^{-19} \cdot \mathbf{1}\,, \quad \ldots$$

Binary search can be used to determine that $n = 18$ *is the least number* n *for which* $\mathbf{p}[n] \leq \widehat{\mathbf{p}}[n] \leq 10^{-5} \cdot \mathbf{1}$ *holds, so the comparison with Example 3.3 shows that the overapproximation is quite tight here. As* $\rho = 1/2$*, Corollary 3.14 yields* $\mathbf{p}[n](p_0X_0) \in \Theta(1/2^n)$*.*

3.3 Approximating the Expectation

We define an approximation method for the expectation $EM_{p_0X_0}$, and bound its error. As mentioned below Proposition 3.4, we have $EM_{p_0X_0} = \sum_{n=1}^{\infty} \mathbf{p}[n](p_0X_0)$, which can be (under-) approximated by the partial sums $\sum_{n=1}^{k} \mathbf{p}[n](p_0X_0)$. The values $\mathbf{p}[n](p_0X_0)$ can be computed using Proposition 3.8.

The following theorem gives error bounds on this approximation method and shows that it converges linearly, i.e., the number of accurate bits (as defined in [28]) is a linear function of the number of iterations. (Recall for the following statement that for a vector $\mathbf{v} \in \mathbb{R}^{\mathcal{H}}$ its 1-norm $\|\mathbf{v}\|_1$ is defined as $\sum_{h \in \mathcal{H}} |\mathbf{v}(h)|$, and that for a matrix B its 1-norm $\|B\|_1$ is the maximal 1-norm of its columns.)

Theorem 3.16. *Let* $UM_{p_0X_0}(k) := \sum_{n=1}^{k} \mathbf{p}[n](p_0X_0)$*. For all* $k \geq |Q|^2|\Gamma| + 3$

$$EM_{p_0X_0} - UM_{p_0X_0}(k) \leq \|A^*\|_1 \, \|\mathbf{p}[k]\|_1 \leq ab^k$$

where $a > 0$ *and* $0 < b < 1$ *are computable rational numbers. Hence, the sequence* $(UM_{p_0X_0}(k))_k$ *converges linearly to* $EM_{p_0X_0}$*.*

The computation procedure of the constants a and b from Theorem 3.16 is somewhat involved(see Appendix B.9), but the first inequality of Theorem 3.16 gives concrete error bounds as well:

Example 3.17. *Using Proposition 3.8 we compute* $\sum_{n=1}^{12} \mathbf{p}[n](X) = 1.5731\ldots$ *and furthermore* $\|\mathbf{p}[12]\|_1 \approx 0.00042$. *We have* $\|A^*\|_1 = 1 + \sqrt{2} \approx 2.4$. *Theorem 3.16 yields*

$$1.57 < \mathrm{E}M_X \leq 1.5731\ldots + \|A^*\|_1 \cdot \|\mathbf{p}[12]\|_1 < 1.58\,.$$

4 Finiteness of the Expected Memory Consumption

In this section we study the complexity of the finite-expectation problem that asks whether the expectation of the memory consumption is finite.

4.1 Expected Memory Consumption of pPDA

For pPDA we can show the following theorem.

Theorem 4.1. *The problem whether* $\mathrm{E}M_{p_0X_0}$ *is finite is decidable in polynomial space.*

The proof is based on the following proposition which strengthens Proposition 3.11 from the previous section which stated that, under the assumption that $\mathrm{E}M_{p_0X_0}$ is finite, the spectral radius ρ of A satisfies $\rho < 1$.

Proposition 4.2. *Suppose* $\mathcal{P}(M_{p_0X_0} < \infty) = 1$. *Then the matrix* A *exists. Moreover, its spectral radius* ρ *satisfies* $\rho < 1$ *if and only if* $\mathrm{E}M_{p_0X_0}$ *is finite.*

The condition $\mathcal{P}(M_{p_0X_0} < \infty) = 1$ can be checked in polynomial space [17]. If it does not hold, then clearly $\mathrm{E}M_{p_0X_0} = \infty$. Otherwise one checks $\rho \geq 1$. Roughly speaking, this can be done in polynomial space because the matrix A is given in terms of the termination probabilities $\mathbf{t}$ which can be expressed in the existential theory of the reals, which is decidable in polynomial space [12, 33].

We can also show that this upper complexity bound from Theorem 4.1 cannot be significantly lowered without a major breakthrough on long-standing and fundamental problems on numerical computations, namely the SQRT-SUM and the PosSLP problems (see [2, 24] or Appendix C.2 for the definition and a discussion of these problems):

Theorem 4.3. *The* PosSLP *problem is P-time many-one reducible to the decision problem whether the expected maximal height of a pPDA is finite.*

It follows that SQRT-SUM is (Turing) reducible to the finite-stack problem, because SQRT-SUM is (Turing) reducible to PosSLP [2, 24].

4.2 Expected Memory Consumption of pBPA

Now we show that for pBPA the finite-expectation problem can be decided in polynomial time. Let us fix a pBPA $\Delta = (\{p\}, \Gamma, \delta, \mathit{Prob})$, and fix an initial configuration $X_0 \in \Gamma$. Let Γ_0 denote the set of all symbols $Y \in \Gamma$ such that $Y\Gamma^*$ is reachable from X_0.

We divide the set Γ_0 into two groups as follows.et *Term* be the set of all symbols $X \in \Gamma_0$ such that $\mathbf{t}(X) = 1$, i.e., a run from a *Term*-symbol terminates almost surely. We define $\mathit{NTerm} = \Gamma_0 \setminus \mathit{Term}$. The following proposition follows from [24] (see also [7]).

Proposition 4.4. *The sets Term and NTerm can be computed in polynomial time.*

For symbols in *Term* we have the following proposition.

Proposition 4.5. *If $X_0 \in \mathit{Term}$, the problem whether $\mathrm{EM}_{X_0} < \infty$ is decidable in polynomial time.*

The decision procedure of Proposition 4.5 is similar to the one of Theorem 4.1, but it runs in polynomial time. Roughly speaking, this is because $X_0 \in \mathit{Term}$ implies $\mathbf{t} = \mathbf{1}$, which makes the matrix A computable in polynomial time. For the case $X_0 \in \mathit{NTerm}$, we use the following proposition.

Proposition 4.6. *If $X_0 \in \mathit{NTerm}$, then the problem whether $\mathcal{P}(M_{X_0} < \infty) = 1$ is decidable in polynomial time.*

ALGORITHM DECIDING WHETHER EM_{X_0} IS FINITE:

1. Compute the sets *Term* and *NTerm* (using Proposition 4.4).
2. Decide whether all $Y \in \mathit{NTerm}$ satisfy $\mathcal{P}(M_Y < \infty) = 1$ (using Proposition 4.6). If no, then stop and return 'no'.
3. Decide whether all $Y \in \mathit{Term}$ satisfy $\mathrm{EM}_Y < \infty$ (using Proposition 4.5). If no, then return 'no'. Otherwise return 'yes'.

Example 4.7. *Consider our running example for $X_0 = X$. In this case $\Gamma_0 = \{X, Y, Z\}$, $\mathit{NTerm} = \{X, Z\}$ and $\mathit{Term} = \{Y\}$. The algorithm decides the finiteness of EM_X as follows. In the first step, the algorithm computes the sets Term and NTerm using Proposition 4.4 (this involves invocation of nontrivial procedures of [24]). Since $\mathcal{P}(M_X < \infty) = \mathcal{P}(M_Z < \infty) = 1$, the algorithm does not stop in step 2. Consequently, the algorithm applies Proposition 4.5 with $X_0 = Y$ which gives that $\mathrm{EM}_Y < \infty$, and hence the algorithm returns 'yes'.*

In some more detail, the decision procedure of Proposition 4.5 proceeds by checking the spectral radius ρ *of the matrix* A*: First, the sets* $\mathcal{H} = \{Y\}$ *and* $\mathcal{T} = \{Y\}$ *are computed. Then the matrix (i.e. the number)* $A = (L + Q(\mathbf{t}, \cdot))^* L' = 1/2$ *is computed. Finally, the decision procedure of Proposition 4.5 decides whether* $\rho < 1$*. As* $\rho = 1/2$*, it finds* $EM_Y < \infty$*.*

Theorem 4.8. *The above algorithm returns 'yes' iff* EM_{X_0} *is finite. It runs in polynomial time.*

A proof sketch and a full proof of this theorem can be found in Appendix C.5.

5 Conclusions

We have investigated the memory consumption of probabilistic pushdown automata (pPDA). Technically speaking, we have studied the random variable M returning the maximal stack height of a pPDA. In [17] a PSPACE algorithm was provided for deciding whether the runs with $M = \infty$ have nonzero probability, but the distribution of M and its expectation have not been studied.

For computing the distribution of M, we have shown that the exponential blow-up of the naive method can be avoided using a system of linear equations. We have also provided an approximation method that gives upper bounds. This can be used, e.g., for providing space that suffices with a probability of, say, 99%.

Computing the expectation EM was mentioned in [17] as "harder problem" and left open. Using novel proof techniques, we have provided a rather complete solution. We have shown that whether the expected maximal stack height of a pBPA is finite can be decided in polynomial time, while for general pPDA the problem is in PSPACE. By means of a reduction to the PosSLP and SQRT-SUM problems we have furthermore shown that this complexity cannot be significantly lowered without major breakthroughs. Finally, we have defined an iterative method for approximating the expected maximal stack height, and have shown that it converges linearly.

The complexity of the decision problem $EM_{p_0 X_0} < k$ for a finite bound k is an open question. The application of our results to program models similar to those of [13] is also left for future research.

References

[1] K. Agrawal, C.E. Leiserson, Y. He, and W.J. Hsu. Adaptive work-stealing with parallelism feedback. *ACM TOCS*, 26(3), 2008.

[2] E. Allender, P. Bürgisser, J. Kjeldgaard-Pedersen, and P. B. Miltersen. On the complexity of numerical analysis. In *IEEE Conference on Computational Complexity*, pages 331–339. IEEE Computer Society, 2006.

[3] N.S. Arora, R.D. Blumofe, and C.G. Plaxton. Thread scheduling for multiprogrammed microprocessors. *Theory of Computing Systems*, 34:115–144, 2001.

[4] A. Berman and R.J. Plemmons. *Nonnegative matrices in the mathematical sciences*. Academic Press, 1979.

[5] R.D. Blumofe and C.E. Leiserson. Scheduling multithreaded computations by work stealing. *Journal of the ACM*, 46(5):720–748, 1999.

[6] A. Bouajjani and J. Esparza. Rewriting models of boolean programs. In *Proceedings of RTA 2006*, Seattle, USA, 2006.

[7] T. Brázdil. *Verification of Probabilistic Recursive Sequential Programs*. PhD thesis, Faculty of Informatics, Masaryk University, 2007. Full version at `http://www.fi.muni.cz/~xbrazdil/thesis.pdf`.

[8] T. Brázdil, V. Brožek, V. Forejt, and A. Kučera. Reachability in recursive Markov decision processes. In *CONCUR 2006*, pages 358–374, 2006.

[9] T. Brázdil, V. Brožek, A. Kučera, and J. Obdržálek. Qualitative reachability in stochastic BPA games. In *STACS 2009*, pages 207–218, 2009.

[10] T. Brázdil, J. Esparza, and A. Kučera. Analysis and prediction of the long-run behavior of probabilistic sequential programs with recursion. In *FOCS'05*, pages 521–530, 2005.

[11] T. Brázdil, A. Kučera, and O. Stražovský. On the decidability of temporal properties of probabilistic pushdown automata. In *STACS 2005*, pages 145–157, 2005.

[12] J. Canny. Some algebraic and geometric computations in PSPACE. In *STOC'88*, pages 460–467, 1988.

[13] K. Chatterjee, D. Ma, R. Majumdar, T. Zhao, T.A. Henzinger, and J. Palsberg. Stack size analysis for interrupt-driven programs. *Inform. and Computat.*, 194(2):144–174, 2004.

[14] J. Esparza and K. Etessami. Verifying probabilistic procedural programs. In *FSTTCS 2004*, pages 16–31, 2004.

[15] J. Esparza, S. Kiefer, and M. Luttenberger. Convergence thresholds of Newton's method for monotone polynomial equations. In *STACS 2008*, pages 289–300, 2008.

[16] J. Esparza, A. Kučera, and R. Mayr. Model checking probabilistic pushdown automata. In *LICS 2004*, pages 12–21. IEEE, 2004.

[17] J. Esparza, A. Kučera, and R. Mayr. Quantitative analysis of probabilistic pushdown automata: Expectations and variances. In *LICS'05*, pages 117–126. IEEE, 2005.

[18] J. Esparza, A. Kučera, and S. Schwoon. Model-checking LTL with regular valuations for pushdown systems. *Information and Computation*, 186(2):355–376, November 2003.

[19] K. Etessami, D. Wojtczak, and M. Yannakakis. Recursive stochastic games with positive rewards. In *ICALP 2008 (1)*, pages 711–723, 2008.

[20] K. Etessami and M. Yannakakis. Algorithmic verification of recursive probabilistic state machines. In *TACAS 2005*, pages 253–270, 2005.

[21] K. Etessami and M. Yannakakis. Recursive Markov decision processes and recursive stochastic games. In *ICALP 2005*, pages 891–903, 2005.

[22] K. Etessami and M. Yannakakis. Efficient qualitative analysis of classes of recursive Markov decision processes and simple stochastic games. In *STACS 2006*, pages 634–645, 2006.

[23] K. Etessami and M. Yannakakis. Recursive concurrent stochastic games. *Logical Methods in Computer Science*, 4(4), 2008.

[24] K. Etessami and M. Yannakakis. Recursive markov chains, stochastic grammars, and monotone systems of nonlinear equations. *Journal of the ACM*, 56(1):1–66, 2009.

[25] E.M. Hahn, H. Hermanns, B. Wachter, and L. Zhang. INFAMY: An infinite-state Markov model checker. In *CAV*, LNCS 5643, pages 641–647, 2009.

[26] R.A. Horn and C.A. Johnson. *Matrix Analysis*. Cambridge University Press, 1985.

[27] J. G. Kemeny and J. L. Snell. *Finite Markov Chains: With a New Appendix "Generalization of a Fundamental Matrix"*. Springer, first edition, 1983.

[28] S. Kiefer, M. Luttenberger, and J. Esparza. On the convergence of Newton's method for monotone systems of polynomial equations. In *STOC 2007*, pages 217–226, 2007.

[29] G.J. Narlikar and G.E. Belloch. Space-efficient scheduling of nested parallelism. *ACM TOPLAS*, 21(1):138–173, 1999.

[30] J. R. Norris. *Markov Chains*. Oxford University Press, first edition, 1998.

[31] A. Remke, B. Haverkort, and L. Cloth. CSL model checking algorithms for QBDs. *Theoretical Computer Science*, 382(1):24–41, 2007.

[32] A. Remke and B.R. Haverkort. CSL model checking algorithms for infinite-state structured Markov chains. In *FORMATS*, LNCS 4763, pages 336–351, 2007.

[33] J. Renegar. On the computational complexity and geometry of the first-order theory of the reals. Parts I–III. *Journal of Symbolic Computation*, 13(3):255–352, 1992.

[34] P. Schnoebelen. The verification of probabilistic lossy channel systems. In *Validation of Stochastic Systems*, LNCS 2925, pages 445–465. Springer, 2004.

[35] M. Yannakakis and K. Etessami. Checking LTL properties of recursive Markov chains. In *QEST 2005*, pages 155–165, 2005.

[36] L. Zhang, H. Hermanns, E.M. Hahn, and B. Wachter. Time-bounded model checking of infinite-state continuous-time markov chains. In *ACSD*, pages 98–107. IEEE, 2008.

APPENDIX

In the appendix we give missing proofs from the main body. In Section A we provide lemmata on nonnegative matrices that we use later on.

Additional Notation and Concepts Used in the Appendix

Markov Chains. Given any Markov chain $(S, \to, \mathit{Prob})$, a state $s \in S$ and a set $A \subseteq S$, we define

$$\mathcal{P}(s \to^* A) = \mathcal{P}(\{w \in \mathit{Run}(s) \mid \exists i \geq 0 : w(i) \in A\})$$

(Recall that A is reachable from s if $\mathcal{P}(s \to^* A) > 0$.)

$$\mathcal{P}(s \to^+ A) = \mathcal{P}(\{w \in \mathit{Run}(s) \mid \exists i \geq 1 : w(i) \in A\})$$

We sometimes write $\mathcal{P}(s \to^* t)$ and $\mathcal{P}(s \to^+ t)$ instead of $\mathcal{P}(s \to^* \{t\})$ and $\mathcal{P}(s \to^+ \{t\})$, respectively.

Probabilistic Pushdown Automata. Given a finite path v in M_Δ and a sequence $\omega = t_1, \ldots, t_{|v|-1}$ of transitions of $\delta \subseteq Q \times \Gamma \times Q \times \Gamma^{\leq 2}$, we say that ω *induces* v if for every $1 \leq i < |v|$ holds that $v(i-1) \to v(i)$ is induced by t_i (i.e., $v(i-1)$ is of the form $pX\alpha$, t_i is of the form $pX \to q\beta$, and $v(i) = q\beta\alpha$).

Given a configuration $pX\alpha$ of a pPDA, we call pX the *head* and α the *tail* of $pX\alpha$ (we also write $\mathit{head}(pX\alpha) = pX$ and $\mathit{tail}(pX\alpha) = \alpha$). The head of $p\varepsilon$ is p and the tail is ε (here ε denotes the empty stack).

Vectors. For any vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^M$, we write $\mathbf{u} \gg \mathbf{v}$ to mean $\mathbf{u}(m) > \mathbf{v}(m)$ for all $m \in M$.

ExTh($\mathbb{R}$). A formula of **ExTh**($\mathbb{R}$), the existential fragment of the first-order theory of the reals, is of the form $\exists x_1 \ldots \exists x_n R(x_1, \ldots, x_n)$ where $R(x_1, \ldots, x_n)$ is a boolean combination of comparisons of the form $p(x_1, \ldots, x_n) \sim 0$ where $p(x_1, \ldots, x_n)$ is a multivariate polynomial and $\sim \in \{<, >, \leq, \geq, =, \neq\}$. The theory **ExTh**($\mathbb{R}$) is known to be decidable in polynomial space [12, 33].

A Constant. We set $n_\perp := |Q|^2|\Gamma| + 3$ throughout the appendix.

A Lemmata on Nonnegative Matrices

Lemma A.1. *Let* $A \in \mathbb{R}^{n \times n}$ *be a nonnegative matrix.*

(a) A^* *exists iff* $\rho(A) < 1$.

(b) If A^* *exists then* $A^* = (I - A)^{-1}$.

(c) If $\rho(A) \geq 1$ *then there is a principal submatrix* B *of* A *such that* $\rho(B) \geq 1$ *and* B *is strongly connected (strongly connected means that for all* $1 \leq i, j \leq n$ *there is* $k \geq 0$ *with* $(B^k)_{ij} \neq 0$*).*

Proof. Statements (a) and (b) are standard theory on nonnegative matrices, see e.g. [4]. Statement (c) follows from (a) and the fact that as long as A is not strongly connected, $\rho(A)$ is an eigenvalue of some proper principal submatrix of A (see [4], Corollary 2.1.6). □

Lemma A.2. *Let* $A, B \in \mathbb{R}^{n \times n}$ *be nonnegative matrices such that* A^* *exists. We have* $\rho(A + B) < 1$ *iff* $\rho(A^* B) < 1$.

Proof. The direction "⇒" follows from the theory of M-matrices and regular splittings, see [4], Theorem 6.2.3 part P_{48}. For the direction "⇐" let $\rho(A^* B) < 1$. Then by Lemma A.1 the matrix $(A^* B)^*$ exists. As A^* exists, also $(A^* B)^* A^*$ exists. This matrix equals $(A + B)^*$ which is easily seen by reordering infinite (absolutely) converging sums. So, $(A + B)^*$ exists as well, hence by Lemma A.1 we have $\rho(A + B) < 1$. □

Lemma A.3. *Let* $A \in \mathbb{R}^{m \times m}$ *be a nonnegative matrix such that* A^* *exists. Let* $(\varepsilon_n)_{n \in \mathbb{N}}$ *be a sequence with* $\varepsilon_n \geq \varepsilon_{n+1} \geq 0$ *converging to* 0*. Then there exists an* n_1 *and a nonnegative matrix* K *such that for all* $n \geq n_1$

$$\big((1 - \varepsilon_n)A\big)^* \geq (I - \varepsilon_n K)A^* .$$

Proof. We can assume $\varepsilon_n \leq 1$. Let $M = (I - A)^{-1} A$. Then by Lemma A.1 and a simple computation

$$\big((1 - \varepsilon_n)A\big)^* = \big(I + \varepsilon_n M\big)^{-1} A^* .$$

Choose n_1 large enough so that $\rho(\varepsilon_n M) < 1$. Then $(\varepsilon_n M)^*$ exists and so

$$\begin{aligned} \big(I + \varepsilon_n M\big)^{-1} &= I - (\varepsilon_n M) + (\varepsilon_n M)^2 - (\varepsilon_n M)^3 + - \cdots \\ &\geq I - (\varepsilon_n M)(\varepsilon_n M)^* \\ &\geq I - \varepsilon_n M(\varepsilon_{n_1} M)^* \end{aligned}$$

Choose $K = M(\varepsilon_{n_1} M)^*$ and the claim follows. □

B Proofs of Section 3

On page 9, we made the assumption for Section 3 that $EM_{p_0X_0}$ is finite. For many proofs in this section, this assumption is not needed, or only a weaker assumption is needed, namely that $\mathcal{P}(M_{p_0X_0} < \infty) = 1$. (This assumption is weaker, as $\mathcal{P}(M_{p_0X_0} = \infty) > 0$ clearly implies that $EM_{p_0X_0}$ is infinite.) For clarity and for later reuse of results we make those assumptions explicit in this appendix, i.e., we include the assumption in the preconditions of the statement whenever it is needed. For instance, no such assumption is needed for Propositions 3.1 and 3.2, as it is not mentioned in the restatements of those propositions below.

B.1 Proof of Propositions 3.1 and 3.2

Here are restatements of Propositions 3.1 and 3.2.

PROPOSITION 3.1. *For every* $n \geq 0$, *the vector* $\mathbf{T}[n]$ *is the* unique *solution of the equation system on page 8.*

PROPOSITION 3.2. *For every* $n \geq 1$, *the vector* $\mathbf{P}[n]$ *is the* unique *solution of the equation system on page 8.*

We prove only Proposition 3.2, since the proof of Proposition 3.1 is analogous.

Proof. The proof that for every $n \geq 2$, the vector $\mathbf{P}[n]$ solves the system is straightforward. Let us prove uniqueness. Fix $n \geq 2$. Let us consider a finite state Markov chain $M = (S, \hookrightarrow, Prob')$ where $S = (Q \times \Gamma) \cup \{term\}$ and $pX \hookrightarrow qY$ iff

$$x = \sum_{pX \overset{y}{\hookrightarrow} qY} y + \sum_{pX \overset{y}{\hookrightarrow} rZY} y\mathbf{T}[n-2](rZq) > 0$$

in which case $Prob'(pX \hookrightarrow qY) = x$. Finally, we define $pX \hookrightarrow term$ iff

$$x = \sum_{pX \overset{y}{\hookrightarrow} qYZ} y\mathbf{P}[n-1](qY) > 0$$

in which case we define $Prob'(pX \hookrightarrow qY) = x$. Let us denote by f_{pX} the probability of reaching $term$ from pX in M. By [30] Theorem 1.3.2, the vector of all values f_{pX} is the least nonnegative solution (with respect to component-wise ordering) of the following system

$$x_{pX} = \begin{cases} 0 & \text{if } f_{pX} = 0 \\ \sum_{pX \overset{y}{\hookrightarrow} term} y + \sum_{pX \overset{y}{\hookrightarrow} qY} y x_{qY} & \text{otherwise.} \end{cases}$$

We first show that this system has only one solution. Obviously, it suffices to consider the restricted system obtained by omitting the equations for x_{pX} with $f_{pX} = 0$ and replacing those x_{pX} on the right hand sides with 0. Hence the least nonnegative solution of this system is positive in all components. If there existed a solution different from the least nonnegative one, then, by basic linear algebra facts, all points on the straight line defined by the two solutions would be solutions as well. In particular, there would exist a nonnegative solution with a zero component, contradicting the existence of a least nonnegative solution which is *positive*. So the system defined above has only one solution.

However, substituting the real transition probabilities of M to the above system we obtain the original system for $\mathbf{P}[n]$. It means that the vector of the values f_{pX} is also the unique solution of the system for $\mathbf{P}[n]$. □

B.2 Proof of Proposition 3.4

Here is a restatement of Proposition 3.4.

PROPOSITION 3.4. *The value $\mathbf{P}[n]$ can be computed by setting up and solving the equation systems of Propositions 3.1 and 3.2 in time $\mathcal{O}\left(n \cdot (|Q|^2 \cdot |\Gamma|)^3\right)$ in the Blum-Shub-Smale model.*

Proof. To set up the equation system for $\mathbf{T}[n]$ we have to determine, for each $pXq \in Q \times \Gamma \times Q$, whether $\mathbf{T}[n](pXq) = 0$. To this end, we set up an equation system for $\widetilde{\mathbf{T}}[n]$, where $\widetilde{\mathbf{T}}[n](pXq)$ equals, for each $pXq \in Q \times \Gamma \times Q$, the boolean value **true** if $\mathbf{T}[n](pXq) > 0$, and **false** if $\mathbf{T}[n](pXq) = 0$. Following [18] and defining an order **false** $<$ **true**, the boolean vector $\widetilde{\mathbf{T}}[n](pXq)$ is the (componentwise) least solution of the following boolean equation system over the variable vector $\widetilde{\mathbf{T}}\langle n\rangle$.

$$\widetilde{\mathbf{T}}\langle n\rangle(pXq) = \bigvee_{pX \overset{y}{\hookrightarrow} q\varepsilon} \mathbf{true} \vee \bigvee_{pX \overset{y}{\hookrightarrow} rY} \widetilde{\mathbf{T}}\langle n\rangle(rYq) \vee \bigvee_{pX \overset{y}{\hookrightarrow} rYZ} \bigvee_{s \in Q} \left(\widetilde{\mathbf{T}}[n-1](rYs) \wedge \widetilde{\mathbf{T}}\langle n\rangle(sZq)\right) .$$

Its least solution, and hence $\widetilde{\mathbf{T}}[n]$, can be computed using fixed-point iteration of that equation system. By implementing it with a worklist algorithm [18], this costs

$$\mathcal{O}(|\delta| \cdot |Q|^2) \leq \mathcal{O}(|Q|^2 \cdot |\Gamma|^3 \cdot |Q|^2) = \mathcal{O}(|Q|^4 \cdot |\Gamma|^3)$$

operations. Once $\widetilde{\mathbf{T}}[n]$ is computed, the equation system for $\mathbf{T}[n]$ from Proposition 3.1 can be set up in time $\mathcal{O}(|Q|^4 \cdot |\Gamma|^3)$ and solved in time $\mathcal{O}\left((|Q|^2 \cdot |\Gamma|)^3\right)$ using Gaussian elimination.

Since all this needs to be done for all n' with $1 \leq n' \leq n$, computing $\mathbf{T}[n]$ requires

$$\mathcal{O}\left(n \cdot (|Q|^2 \cdot |\Gamma|)^3\right)$$

operations.

The values $\mathbf{T}[n]$ can be used to compute $\mathbf{P}[n]$ using the equation system from Proposition 3.2. It can be set up and solved similarly as shown above for $\mathbf{T}[n]$. The costs for $\mathbf{P}[n]$ are at most the costs for $\mathbf{T}[n]$. (Some costs decrease by a factor of $|Q|$.) So the total time for determining $\mathbf{P}[n]$ in the Blum-Shub-Smale model is $\mathcal{O}\left(n \cdot (|Q|^2 \cdot |\Gamma|)^3\right)$. $\square$

B.3 Proof of Lemma 3.5

Here is a restatement of Lemma 3.5.

LEMMA 3.5.

1. $\mathbf{T}[|Q|^2|\Gamma|+1](pXq) > 0 \Longleftrightarrow$ *for all* $n \geq |Q|^2|\Gamma|+1 : \mathbf{T}[n](pXq) > 0 \Longleftrightarrow [pXq] > 0$;
2. $\mathbf{P}[|Q||\Gamma|+1](pX) > 0 \Longleftrightarrow$ *for all* $n \geq 1 : \mathbf{P}[n](pX) > 0$.

Proof. The proof of this lemma is based on arguments similar to the pumping lemma from language theory.

ad 1. The first "$\Longleftrightarrow$" follows from the fact that, by definition, $\mathbf{T}[n](pXq)$ is monotonically increasing with n. Since $\mathbf{T}[|Q|^2|\Gamma|+1](pXq) > 0$ implies $[pXq] > 0$, it remains to show that $[pXq] > 0$ implies $\mathbf{T}[|Q|^2|\Gamma|+1](pXq) > 0$. Let us define a sequence of sets $A_0, A_1, A_2, \ldots \subseteq Q \times \Gamma \times Q$ as follows: We define $A_0 = \emptyset$ and

$$\begin{aligned} A_{i+1} &= A_i \\ &\cup \{pXq \mid pX \hookrightarrow q\varepsilon\} \\ &\cup \{pXq \mid pX \hookrightarrow rY, rYq \in A_i\} \\ &\cup \{pXq \mid pX \hookrightarrow rYZ; rYs, sZq \in A_i)\}\,. \end{aligned}$$

It is easy to see that $[pXq] > 0$ implies that $pXq \in A_n$ for some $n \geq 0$. As $A_i \subseteq A_{i+1}$ and A_{i+1} depends only on A_i, we can choose $n \leq |Q \times \Gamma \times Q| = |Q|^2|\Gamma|$. We show, by induction, that if $pXq \in A_n$, then there is a path from pX to $q\varepsilon$ with maximal stack height at most $n+1$. This trivially holds for $n = 0$. Let us consider $pXq \in A_{n+1}$. If either $pXq \in A_n$, or $pX \hookrightarrow q\varepsilon$, then we are done. Also, if $pX \hookrightarrow rY$ where $rYq \in A_n$, we are done. Assume that $pX \to rYZ$ and assume that $rYs, sZq \in A_n$ for some $s \in Q$. By induction, there is a path u from rY to $s\varepsilon$ with maximal stack height at most n, and also a path v from sZ to $q\varepsilon$ with maximal stack height at most n. Assume that u is induced by a sequence $t_1, \dots, t_k$ of transitions of δ and v is induced by a sequence $t_{k+1}, \dots, t_\ell$. Now a sequence of transitions $pX \to rYZ, t_1, \dots, t_\ell$ induces a path from pX to $q\varepsilon$ with maximal stack height at most $n+1$.

ad 2. Let us assume that

$$\mathcal{P}(\{w \in Run(pX) \mid M_{pX}(w) \geq |Q||\Gamma| + 1\}) > 0$$

There is a path $v = v(0) \dots v(m)$ with $v(0) = pX$ and $|v(m)| \geq |Q||\Gamma| + 1$. By the pigeonhole principle there are $0 \leq j < k \leq m$ such that $head(v(j)) = head(v(k))$ and $|v(j)| < |v(k)|$ and $|v(j)| \leq |v(t)|$ for all $j \leq t \leq k$.

Let us assume that v is induced by the sequence $t_1, t_2, \dots t_m$ of transitions of δ (for the meaning of *induce* see the beginning of Appendix). Now for every $\ell \geq 1$ we define a path v_ℓ induced by the following sequence of transitions of δ:

$$t_1, t_2, \dots, t_j, (t_{j+1}, \dots, t_k)^\ell$$

It is easy to see that v_ℓ reaches a height of at least ℓ. Hence

$$\mathcal{P}(\{w \mid M_{pX}(w) \geq \ell\}) \geq \mathcal{P}(Run(v_\ell)) > 0\,.$$

□

B.4 Proof of Lemma 3.6

We prove the following stronger version of Lemma 3.6.

LEMMA 3.6 (stronger version).

(1) The set $\mathcal{T}$ is computable in polynomial time. If $pXq \in \mathcal{T}$ and $\mathbf{T}\langle n\rangle(rYs)$ occurs in the equation for $\mathbf{T}\langle n\rangle(pXq)$ with a nonzero coefficient, then either $rYs \in \mathcal{T}$, or for all $n \geq 0$ holds $\mathbf{T}[n](rYs) = 0$.

(2) The set $\mathcal{H}$ is computable in polynomial time. If $pX \in \mathcal{H}$ and $\mathbf{P}\langle n\rangle(rY)$ occurs in the equation for $\mathbf{P}\langle n\rangle(pX)$ with a nonzero coefficient, then either $rY \in \mathcal{H}$, or for all $n \geq |Q||\Gamma| + 1$ holds $\mathbf{P}[n](rY) = 0$.

Proof. (1) We start by showing that the problem whether $pXq \in \mathcal{T}$ is decidable in polynomial time. Whether $pX\Gamma^*$ is reachable from p_0X_0 can be decided using algorithms of [18] in polynomial time. Note that $[pXq] > 0$ iff $q\varepsilon$ is reachable from pX. The latter can also be decided in polynomial time using results of [18].

Assume that for some $pXq \in \mathcal{T}$, a variable $\mathbf{T}\langle n\rangle(rYs)$ occurs in the equation for $\mathbf{T}\langle n\rangle(pXq)$ with a nonzero coefficient. Then $rY\Gamma^*$ is reachable from pX, and hence also from p_0X_0. Thus if $rYs \notin \mathcal{T}$, then $[rYs] = 0$, and hence for all $n \geq 0$ holds $\mathbf{T}[n](rYs) = 0$.

(2) We start by showing that the problem whether $pX \in \mathcal{H}$ is decidable in polynomial time. Whether $pX\Gamma^*$ is reachable from p_0X_0 can be decided using algorithms of [18] in polynomial time. By Lemma 3.5, $\mathbf{P}[n](pX) > 0$ for all $n \geq 0$ iff $\mathbf{P}[|Q||\Gamma| + 1](pX) > 0$. The latter can also be decided in polynomial time using results of [18].

Assume that for some $pX \in \mathcal{H}$, a variable $\mathbf{P}\langle n\rangle(rY)$ occurs with a nonzero coefficient in the equation for $\mathbf{P}\langle n\rangle(pX)$. Then clearly $rY\Gamma^*$ is reachable from pX and hence also from p_0X_0. Hence, if $rY \notin \mathcal{H}$, then there must be some $\ell \geq 1$ such that $\mathbf{P}[\ell](pX) = 0$. However, then by Lemma 3.5, $\mathbf{P}[|Q||\Gamma| + 1](pX) = 0$, which implies that for all $\ell \geq |Q||\Gamma| + 1$ holds $\mathbf{P}[\ell](pX) = 0$.

□

B.5 Proof of Proposition 3.8

Here is a restatement of Propositions 3.8.

PROPOSITION 3.8. *The following equations hold for all* $n \geq |Q|^2|\Gamma| + 3$*:*

$$\mathbf{t}[n] = \mathbf{c} + \tilde{L}\mathbf{t}[n] + \tilde{Q}(\mathbf{t}[n-1], \mathbf{t}[n]) \quad \textit{and} \quad \mathbf{p}[n] = L\mathbf{p}[n] + L'\mathbf{p}[n-1] + Q(\mathbf{t}[n-2], \mathbf{p}[n])$$

Proof. The equations are equal to the ones of Propositions 3.1 and 3.2 up to some omitted terms, which are zero according to the stronger version of Lemma 3.6 that was proved above. □

B.6 Proof of Proposition 3.10

Here is a restatement of Proposition 3.10.

PROPOSITION 3.10. *Suppose* $\mathcal{P}(M_{p_0X_0} < \infty) = 1$. *Let*

$$A_n := (L + Q(\mathbf{t}[n-2], \cdot))^* L' \mathbf{p}[n-1] .$$

Then for every $n \geq |Q|^2|\Gamma| + 3$ *the matrix* A_n *exists and* $\mathbf{p}[n] = A_n \mathbf{p}[n-1]$.

We first prove the following lemma.

Lemma B.1. *Suppose* $\mathcal{P}(M_{p_0X_0} < \infty) = 1$. *The matrix* $A = (L + Q(\mathbf{t}, \cdot))^* L'$ *exists.*

Proof. Let us consider a finite state Markov chain $M = (S, \hookrightarrow, \mathit{Prob}')$ where $S = \mathcal{H} \cup \{\mathrm{term}\}$ and $pX \hookrightarrow qY$ iff

$$x = \sum_{pX \overset{y}{\hookrightarrow} qY} y + \sum_{\substack{pX \overset{y}{\hookrightarrow} rZY \\ rZq \in \mathcal{T}}} y\mathbf{t}(rZq) > 0$$

in which case $\mathit{Prob}'(pX \hookrightarrow qY) = x$. Finally, we define $pX \hookrightarrow \mathrm{term}$ iff $1 - \sum_{pX \hookrightarrow qY} \mathit{Prob}'(pX \hookrightarrow qY) > 0$ in which case we put

$$\mathit{Prob}'(pX \hookrightarrow \mathrm{term}) = 1 - \sum_{pX \hookrightarrow qY} \mathit{Prob}'(pX \hookrightarrow qY)$$

For all $pX, qY \in \mathcal{H}$ holds that $\mathit{Prob}'(pX \hookrightarrow qY)$ is the entry of $L + Q(\mathbf{t}, \cdot)$ corresponding to the heads pX and qY. We show that all $pX \in \mathcal{H}$ are transient states of M (see [27], Definition 2.4.1), and then apply [27], Corollary 3.1.2 to obtain the desired result.

Assume that $pX \in \mathcal{H}$ is not a transient state of M, i.e. that the probability of reaching pX from pX in at least one step is 1 (in the chain M). We show that

$$\mathcal{P}(pX \rightarrow^+ pX) = 1$$

Given $qY, rZ \in \mathcal{H}$, denote by $B_{qY,rZ}$ the set of all paths $p_0\alpha_0, \ldots, p_n\alpha_n$ where $n > 0$, $p_0\alpha_0 = qY$, $p_n\alpha_n = rZ$, and for all $0 < i < n$ holds $|\alpha_i| \geq 2$. It is easy to show that

$$\mathit{Prob}'(pX \hookrightarrow qY) = \mathcal{P}\Big(\bigcup_{v \in B_{pX,qY}} \mathit{Run}(v)\Big)$$

However, then $\mathcal{P}(pX \rightarrow^+ pX)$ is equal to the probability of reaching pX from pX in at least one step in M, which is 1.

Since $pX \in \mathcal{H}$, for every $n \geq 0$ there is a path v_n from pX to pX in M_Δ such that the maximal stack height in v_n is at least n. It is straightforward to show that almost all runs initiated in pX follow all paths v_n infinitely many times. Hence, for almost all runs $w \in \mathit{Run}(pX)$ holds $M_{pX}(w) = \infty$. However, then $\mathcal{P}(M_{p_0X_0} = \infty) > 0$ (because $pX \in \mathcal{H}$)

which contradicts our assumption that $\mathcal{P}(M_{p_0X_0} < \infty) = 1$. Hence, all elements of $\mathcal{H}$ are transient states of M.

Let us denote by $P^n(pX, qY)$ the probability of reaching qY from pX in M in precisely n steps. Observe that $\sum_{n=0}^{\infty} P^n(pX, qY)$ is precisely the entry of the matrix $(L + Q(\mathbf{t}, \cdot))^* = \sum_{n=0}^{\infty}(L + Q(\mathbf{t}, \cdot))^n$ corresponding to the heads pX and qY. By [27], Corollary 3.1.2, there are numbers $b > 0$ and $0 < c < 1$ such that $P^n(pX, qY) \leq b \cdot c^n$, and thus $\sum_{n=0}^{\infty} P^n(pX, qY) \leq \frac{b}{1-c} < \infty$. It follows that $(L + Q(\mathbf{t}, \cdot))^*$ exists. □

Now we can prove Proposition 3.10.

Proof. We have $\mathbf{t}[n] \leq \mathbf{t}$ for all $n \geq 1$. Since A exists by Lemma B.1, the matrix A_n also exists, using the monotonicity of the matrix star. The equality $\mathbf{p}[n] = A_n \mathbf{p}[n-1]$ follows from Proposition 3.8 using the fact that $(L + Q(\mathbf{t}[n], \cdot))^* = \big(I - (L + Q(\mathbf{t}[n], \cdot))\big)^{-1}$. □

B.7 Proof of Proposition 3.11

Here is a restatement of Proposition 3.11.

PROPOSITION 3.11. *Suppose that $EM_{p_0X_0}$ is finite. The matrix A exists and its spectral radius ρ satisfies $\rho < 1$.*

The fact that A exists follows from Lemma B.1. The statement that $\rho < 1$ follows from the following Lemma.

Lemma B.2. *Suppose $\mathcal{P}(M_{p_0X_0} < \infty) = 1$. Then $EM_{p_0X_0}$ is infinite iff $\rho \geq 1$.*

So it suffices to show Lemma B.2. It can be seen as the technical core of the paper, and the proof is quite long. We start with a proof sketch.
Proof sketch.

The direction "⇒" is quite straightforward: Let $\rho < 1$. By Proposition 3.10

$$\mathbf{p}[n] = (L + Q(\mathbf{t}[n-2], \cdot))^* L' \mathbf{p}[n-1] \leq A\mathbf{p}[n-1] \quad \text{for } n \geq n_\perp. \tag{1}$$

By an easy induction we obtain $\mathbf{p}[n_\perp + n] \leq A^n \mathbf{p}[n_\perp]$ and so $\sum_{n=0}^{\infty} \mathbf{p}[n_\perp + n] \leq A^* \mathbf{p}[n_\perp]$. By standard matrix theory, $\rho < 1$ implies that the matrix sum A^* converges, so $\sum_{n=1}^{\infty} \mathbf{p}[n]$ is finite, in particular $EM_{p_0X_0}$ is finite.

For the other (harder) direction "⇐", let $\rho \geq 1$. We have to show that $EM_{p_0X_0}$ is infinite. It suffices to show that EM_{pX} is infinite for some $pX \in \mathcal{H}$ because the configuration

p_0X_0 can reach all $pX \in \mathcal{H}$. In fact, we even show that for some $pXq \in \mathcal{T}$ the conditional expectation $E(M_{pX} \mid \mathit{Run}(pXq))$ is infinite, i.e., the expectation of M_{pX} under the condition that pX terminates at q. As $[pXq] > 0$ for $pXq \in \mathcal{T}$, it is equivalent to show that $[pXq] \cdot E(M_{pX} \mid \mathit{Run}(pXq))$ is infinite. This value equals $\sum_{n=1}^{\infty} \mathbf{e}[n](pXq)$ where we define

$$\mathbf{e}[n](pXq) := \mathcal{P}(\{w \in \mathit{Run}(pXq) \mid M_{pX}(w) \geq n\}) ,$$

i.e., $\mathbf{e}[n](pXq)$ is the probability for the configuration pX to reach a stack height of at least n and then to terminate to q. Summarizing, it suffices to show that the sum $\sum_{n=1}^{\infty} \mathbf{e}[n]$ diverges.

Notice that we have $\mathbf{e}[n] := \mathbf{t} - \mathbf{t}[n-1]$. Therefore, showing that $\sum_{n=1}^{\infty} \mathbf{e}[n]$ diverges amounts to showing that $(\mathbf{t}[n])$ converges 'slowly' to $\mathbf{t}$. Two steps remain:

1. Show that it is *impossible* that $(\mathbf{t}[n])$ converges 'fast', i.e., at least linearly.

2. Show that it follows from 1. that $\sum_{n=1}^{\infty} \mathbf{e}[n]$ diverges.

Both steps crucially depend on Perron-Frobenius theory. Let us give the intuition.

For step 1., we return to $(\mathbf{p}[n])$ and define a small variant $(\mathbf{p}'[n])_{n \geq n_\perp}$ by $\mathbf{p}'[n_\perp] = \mathbf{p}[n_\perp]$ and $\mathbf{p}'[n] = A\mathbf{p}'[n-1]$. By Perron-Frobenius theory, $\rho \geq 1$ implies that $(\mathbf{p}'[n])$ does not converge to $\mathbf{0}$. On the other hand, $(\mathbf{p}[n])$ converges to $\mathbf{0}$, because $\mathcal{P}(M_{pX} = \infty) = 0$ for all $pX \in \mathcal{H}$. But comparing the definition of $\mathbf{p}'[n]$ with Equation (1) for $\mathbf{p}[n]$ shows that $\mathbf{p}[n]$ and $\mathbf{p}'[n]$ differ only in that the matrix $(L + Q(\mathbf{t}[n-2], \cdot))^* L'$ has changed to $(L + Q(\mathbf{t}, \cdot))^* L'$. So it is natural to expect that the coefficients $\mathbf{t}[n]$ converge 'slowly' to $\mathbf{t}$, because otherwise $(\mathbf{p}[n])$ and $(\mathbf{p}'[n])$ would have the same limit. We can show that this is in fact true. More precisely, it is impossible that the sequence $(\mathbf{t}[n])$ converges (at least) *linearly* to $\mathbf{t}$.

For step 2., we take advantage of the fact that $\mathbf{e}[n]$ satisfies the following recurrence:

$$\mathbf{e}[n] = \left(\tilde{L} + \tilde{Q}(\mathbf{t} - \mathbf{e}[n-1], \cdot)\right)^* \tilde{Q}(\cdot, \mathbf{t})\mathbf{e}[n-1] \tag{2}$$

Let $B = \left(\tilde{L} + \tilde{Q}(\mathbf{t}, \cdot)\right)^* \tilde{Q}(\cdot, \mathbf{t})$. Assume first $\rho(B) < 1$. We have $\mathbf{e}[n] \leq B\mathbf{e}[n-1]$ by (2), so by standard matrix theory, the sequence $(\mathbf{e}[n])$ converges linearly to $\mathbf{0}$, a situation that we have excluded in step 1. So we have in fact $\rho(B) \geq 1$.

Like the sequence $(\mathbf{p}'[n])$ above, define a sequence $(\mathbf{e}'[n])_{n \geq n_\perp}$ by $\mathbf{e}'[n_\perp] = \mathbf{e}[n_\perp]$ and $\mathbf{e}'[n] = B\mathbf{e}'[n-1]$. By Perron-Frobenius theory, $\rho(B) \geq 1$ implies that $(\mathbf{e}'[n])$ does not converge to $\mathbf{0}$, which in turn implies that $\sum_n \mathbf{e}'[n]$ diverges. Comparing the definition of $\mathbf{e}'[n]$ with Equation (2) for $\mathbf{e}[n]$ shows that $\mathbf{e}[n]$ and $\mathbf{e}'[n]$ differ only in that the matrix $\left(\tilde{L} + \tilde{Q}(\mathbf{t} - \mathbf{e}[n-1], \cdot)\right)^* \tilde{Q}(\cdot, \mathbf{t})$ has changed to $\left(\tilde{L} + \tilde{Q}(\mathbf{t}, \cdot)\right)^* \tilde{Q}(\cdot, \mathbf{t})$. So, intuitively, if $(\mathbf{e}[n])$ decays quickly then $\mathbf{e}[n]$ 'behaves like' $\mathbf{e}'[n]$, i.e., $\sum_n \mathbf{e}[n]$ diverges. On the other hand, if $(\mathbf{e}[n])$ decays slowly then it is immediately conceivable that $\sum_n \mathbf{e}[n]$ diverges as well. Therefore, we prove a carefully chosen inductive invariant that shows that the divergence of $\sum_n \mathbf{e}[n]$ cannot be avoided, again using Perron-Frobenius theory for nonnegative matrices. □

We now give a full proof of Lemma B.2 for which we have given a sketch above. Unlike the proof sketch, the following proof is organized in a bottom-up fashion in order to avoid forward references. It is divided into a sequence of lemmata.

The following lemma is similar to Lemma 3.5 in that it is also based on a "pumping" argument.

Lemma B.3. *For every* $pXq \in \mathcal{T}$ *we have that either* $\mathbf{t}[|Q|^2|\Gamma|](pXq) = [pXq]$*, or for all* $n \geq 1$ *holds* $\mathbf{t}[n](pXq) < [pXq]$.

Proof. Let $\xi = |Q|^2|\Gamma|$. Assume that $\mathbf{t}[\xi](pXq) < [pXq]$. Then there is a path ν from pX to $q\varepsilon$ with maximal stack height at least $\xi + 1$. There are numbers $i_1 < \ldots < i_{\xi+1}$ such that for every $1 \leq k \leq \xi + 1$ holds $|tail(\nu(i_k))| = k$ and for $i_k \leq l \leq i_{\xi+1}$ holds $|tail(\nu(l))| \geq k$. For every k let j_k be the least number greater than i_k such that $|tail(\nu(j_k))| = k - 1$. By the pigeonhole principle there exist $k < \ell$ such that $head(\nu(i_k)) = head(\nu(i_\ell))$ and the configurations $\nu(j_k)$ and $\nu(j_\ell)$ have the same control state.

Let $t_1, \ldots, t_m$ be the sequence of transitions in δ that induce ν. Given $n \geq 1$, we define a path ν_n induced by the following transitions of δ (for the meaning of *induce* see the beginning of Appendix):

$$t_1, \ldots, t_{i_k}, s_{up}^n, t_{i_\ell+1}, \ldots, t_{j_\ell}, s_{down}^n, t_{j_k+1}, \ldots, t_m$$

where

$$s_{up} := t_{i_k+1}, \ldots, t_{i_\ell}$$

$$s_{down} := t_{j_\ell+1}, \ldots, t_{j_k}$$

and s_{up}^n and s_{down}^n are concatenations of n copies of s_{up} and s_{down}, respectively.

It is easy to verify that maximum stack height of ν_n is at least $\xi + n$. Hence, $\mathbf{t}[n](pXq) < [pXq]$ for all $n \geq 1$. □

We define, for $n \geq n_\perp$, the vector $\mathbf{e}[n] \in \mathbb{R}^{\mathcal{T}}$ as

$$\mathbf{e}[n] := \mathbf{t} - \mathbf{t}[n-1] .$$

Notice that for all $pXq \in \mathcal{T}$ we have

$$\mathbf{e}[n](pXq) = \mathcal{P}(\{w \in \mathit{Run}(pXq) \mid M_{pX}(w) \geq n\}) \ ,$$

i.e., $\mathbf{e}[n](pXq)$ is the probability of all runs of $\mathit{Run}(pX)$ that terminate at q and reach a height of at least n. For $\mathbf{e}[n]$ we give a recurrence similar to those given in Proposition 3.8 for $\mathbf{t}[n]$ and $\mathbf{p}[n]$:

Lemma B.4. *For all $n \geq n_\perp$:*

$$\mathbf{e}[n] \ = \ \big(\tilde{L} + \tilde{Q}(\mathbf{t} - \mathbf{e}[n-1], \cdot)\big)\mathbf{e}[n] + \tilde{Q}(\cdot, \mathbf{t})\mathbf{e}[n-1]$$

Proof. It is easy to see (see e.g. [16]) that the vector $\mathbf{t}$ of termination probabilities satisfies

$$\mathbf{t} = \mathbf{c} + \tilde{L}\mathbf{t} + \tilde{Q}(\mathbf{t}, \mathbf{t}) \ . \tag{3}$$

Then, using $\mathbf{t} = \mathbf{t}[n-1] + \mathbf{e}[n]$, we have:

$$\begin{aligned}
\mathbf{e}[n] &= \mathbf{t} - \mathbf{t}[n-1] \\
&= \mathbf{t} - \mathbf{c} - \tilde{L}\mathbf{t}[n-1] - \tilde{Q}(\mathbf{t}[n-2], \mathbf{t}[n-1]) \\
&\qquad \text{(by Proposition 3.8)} \\
&= \mathbf{t} - \mathbf{c} - \tilde{L}(\mathbf{t} - \mathbf{e}[n]) - \tilde{Q}(\mathbf{t} - \mathbf{e}[n-1], \mathbf{t} - \mathbf{e}[n]) \\
&= \tilde{L}\mathbf{e}[n] + \tilde{Q}(\mathbf{t} - \mathbf{e}[n-1], \mathbf{e}[n]) + \tilde{Q}(\mathbf{e}[n-1], \mathbf{t}) \\
&\qquad \text{(by (3))}
\end{aligned}$$

□

The following lemma (and its proof) is analogous to Lemma B.1.

Lemma B.5. *The matrix $\big(\tilde{L} + \tilde{Q}(\mathbf{t}, \cdot)\big)^*$ exists.*

Proof. Let us consider a finite state Markov chain $M = (S, \hookrightarrow, \mathit{Prob}')$ where $S = \mathcal{T} \cup \{\mathrm{term}\}$ and $pXs \hookrightarrow qYs$ iff

$$x = \sum_{pX \stackrel{y}{\hookrightarrow} qY} y + \sum_{\substack{pX \stackrel{y}{\hookrightarrow} rZY \\ rZq \in \mathcal{T}}} y\mathbf{t}(rZq) > 0$$

in which case $\mathit{Prob}'(pXs \hookrightarrow qYs) = x$. Finally, we define $pXs \hookrightarrow \mathrm{term}$ iff $1 - \sum_{pXs \hookrightarrow qYs} \mathit{Prob}'(pXs \hookrightarrow qYs) > 0$ in which case we put

$$\mathit{Prob}'(pXs \hookrightarrow \mathrm{term}) = 1 - \sum_{pXs \hookrightarrow qYs} \mathit{Prob}'(pXs \hookrightarrow qYs)$$

For all $pXs, qYr \in \mathcal{T}$ holds that $\mathit{Prob}'(pXs \hookrightarrow qYr)$ is the entry of $\tilde{L} + \tilde{Q}(\mathbf{t}, \cdot)$ corresponding to the triples pXs and qYr. We show that all $pXs \in \mathcal{T}$ are transient states of M (see [27], Definition 2.4.1), and then apply [27], Corollary 3.1.2 to obtain the desired result.

Assume that $pXs \in \mathcal{T}$ is not a transient state of M, i.e., the probability of reaching pXs from pXs in at least one step is 1 (in the chain M). We show that

$$\mathcal{P}(pX \rightarrow^{+} pX) = 1 .$$

Denote by $B_{qY,rZ}$ the set of all paths $p_0\alpha_0, \ldots, p_n\alpha_n$ where $n > 0$, $p_0\alpha_0 = qY$, $p_n\alpha_n = rZ$, and for all $0 < i < n$ holds $|\alpha_i| \geq 2$.

It is easy to show that

$$\mathit{Prob}'(pXs \hookrightarrow qYs) = \mathcal{P}(\bigcup_{v \in B_{pX,qY}} \mathit{Run}(v))$$

Then $\mathcal{P}(pX \rightarrow^{+} pX)$ is equal to the probability of reaching pXs from pXs in at least one step in M, which is 1.

But $\mathcal{P}(pX \rightarrow^{+} pX) = 1$ implies $[pXs] = 0$ contradicting that $pXs \in \mathcal{T}$. Hence, all triples $t \in \mathcal{T}$ are transient states of M.

Let us denote by $P^n(pXs, qYs)$ the probability of reaching qYs from pXs in M in precisely n steps. Observe that $\sum_{n=0}^{\infty} P^n(pXs, qYs)$ is precisely the entry of the matrix $(\tilde{L} + \tilde{Q}(\mathbf{t}, \cdot))^* = \sum_{n=0}^{\infty}(\tilde{L} + \tilde{Q}(\mathbf{t}, \cdot))^n$ corresponding to the triples pXs and qYs. By [27], Corollary 3.1.2, there are numbers $b > 0$ and $0 < c < 1$ such that $P^n(pXs, qYs) \leq b \cdot c^n$, and thus $\sum_{n=0}^{\infty} P^n(pXs, qYs) \leq \frac{b}{1-c} < \infty$. It follows that $(\tilde{L} + \tilde{Q}(\mathbf{t}, \cdot))^*$ exists. □

This immediately implies the following:

Corollary B.6. *For all* $n \geq n_\perp$:

$$\mathbf{e}[n] \quad = \quad \left(\tilde{L} + \tilde{Q}(\mathbf{t} - \mathbf{e}[n-1], \cdot)\right)^* \tilde{Q}(\cdot, \mathbf{t})\mathbf{e}[n-1]$$

We will need that the ratio between certain components of $\mathbf{e}[n]$ cannot be not arbitrarily large:

Lemma B.7. *There is a constant* $c > 0$ *such that* $\mathbf{e}[n](t_1) \geq c\mathbf{e}[n](t_2)$ *holds for all* $n \in \mathbb{N}$ *and all* $t_1, t_2 \in \mathcal{T}$ *such that* t_2 *is reachable from* t_1 *in* $\tilde{L} + \tilde{Q}(\mathbf{t}, \cdot) + \tilde{Q}(\cdot, \mathbf{t})$. *(Here, by "reachable" we mean that* $\left((\tilde{L} + \tilde{Q}(\mathbf{t}, \cdot) + \tilde{Q}(\cdot, \mathbf{t}))^i\right)_{t_1, t_2} \neq 0$ *for some* $i \geq 0$.*)*

Proof. Let $t_1 = pXq$ and $t_2 = rYs$. We prove that there is $\alpha \in \Gamma^*$ such that $\mathcal{P}(pX \to^* rY\alpha) > 0$ and $\mathcal{P}(s\alpha \to^* q\varepsilon) > 0$. From this we obtain that $\mathbf{e}[n](t_1) \geq \mathcal{P}(pX \to^* rY\alpha)\mathcal{P}(s\alpha \to^* q\varepsilon)\mathbf{e}[n](t_2)$ and it suffices to put $c = \mathcal{P}(pX \to^* rY\alpha)\mathcal{P}(s\alpha \to^* q\varepsilon)$.

We prove the statement by induction on the length of a shortest path from t_1 to t_2 in $\tilde{L} + \tilde{Q}(\mathbf{t}, \cdot) + \tilde{Q}(\cdot, \mathbf{t})$. If the length is 0, then $t_1 = t_2$ and the claim follows immediately. Assume that $s_1, \ldots, s_n$ ($n \geq 1$) is a shortest path in $\tilde{L} + \tilde{Q}(\mathbf{t}, \cdot) + \tilde{Q}(\cdot, \mathbf{t})$ from $t_1 = s_1$ to $t_2 = s_n$. Assume that $s_2 = tUv$. By induction, there is $\alpha' \in \Gamma^*$ such that $\mathcal{P}(tU \to^* rY\alpha') > 0$ and $\mathcal{P}(s\alpha' \to^* v\varepsilon) > 0$. There are two cases:

- If $(\tilde{L} + \tilde{Q}(\cdot, \mathbf{t}))_{pXq,tUv} > 0$ then $pX \to tU\gamma$ where $\mathcal{P}(v\gamma \to^* q\varepsilon) > 0$.
- If $\tilde{Q}(\mathbf{t}, \cdot)_{pXq,tUv} > 0$ then $pX \to uVU$ where $[uVt] > 0$ and $q = v$.

In the former case we have $\mathcal{P}(pX \to^* rY\alpha'\gamma) > 0$ and $\mathcal{P}(s\alpha'\gamma \to^* q\varepsilon) > 0$. In the latter case we have $\mathcal{P}(pX \to^* tU) > 0$, which implies that $\mathcal{P}(pX \to^* rY\alpha') > 0$ and $\mathcal{P}(s\alpha' \to^* v\varepsilon) > 0$. □

Let $\mathcal{T}_\uparrow \subseteq \mathcal{T}$ denote the set of those triples $pXq \in \mathcal{T}$ for which $\mathbf{e}[n](pXq) > 0$ for all $n \in \mathbb{N}$. In the following we denote, for any vector $\mathbf{v} \in \mathbb{R}^{\mathcal{T}}$, by $\mathbf{v}_\uparrow$ the projection of $\mathbf{v}$ on $\mathcal{T}_\uparrow$, i.e., $\mathbf{v}_\uparrow := \mathbf{v}_{|\mathcal{T}_\uparrow}$. We denote, for any $\mathbf{u} \in \mathbb{R}^{\mathcal{T}}$, by $\bar{L}, \bar{P}(\mathbf{u}), \bar{Q} \in \mathbb{R}^{\mathcal{T}_\uparrow \times \mathcal{T}_\uparrow}$ the principal submatrices of $\tilde{L}, \tilde{Q}(\mathbf{u}, \cdot)$ and $\tilde{Q}(\cdot, \mathbf{t})$, respectively, obtained by deleting all rows and columns not indexed by $\mathcal{T}_\uparrow$. Then Corollary B.6 can be restricted to $\mathbf{e}_\uparrow[n]$ as follows:

Lemma B.8. *For all* $n \geq n_\perp$:

$$\mathbf{e}_\uparrow[n] \quad = \quad \left(\bar{L} + \bar{P}(\mathbf{t} - \mathbf{e}[n-1])\right)^* \bar{Q}\mathbf{e}_\uparrow[n-1]$$

Proof. By Corollary B.6 we have

$$\mathbf{e}_\uparrow[n] = \left(\left(\tilde{L} + \tilde{Q}(\mathbf{t} - \mathbf{e}[n-1], \cdot)\right)^* \tilde{Q}(\cdot, \mathbf{t})\mathbf{e}[n-1]\right)_\uparrow .$$

So it suffices to show that

$$\left(\left(\tilde{L} + \tilde{Q}(\mathbf{t} - \mathbf{e}[n-1], \cdot)\right)^j \tilde{Q}(\cdot, \mathbf{t})\mathbf{e}[n-1]\right)(t) = 0$$

for all $j \geq 0$ and all $t \notin \mathcal{T}_\uparrow$. This is true because otherwise we would have, again by Corollary B.6, a $t \notin \mathcal{T}_\uparrow$ with $\mathbf{e}[n_\perp](t) > 0$, and hence, by Lemma B.3, $\mathbf{e}[n](t) > 0$ for all $n \in \mathbb{N}$, which contradicts the definition of $\mathcal{T}_\uparrow$. □

The following lemma is the key to the proof of Lemma B.2. It provides a dichotomy result on $\mathbf{e}[n]$.

Lemma B.9. *If* $\mathcal{T}_\uparrow = \emptyset$ *then* $\mathbf{e}[n] = \mathbf{0}$ *for all* $n \geq n_\perp$. *Otherwise, let* $r = \rho\big(\bar{L} + \bar{P}(\mathbf{t}) + \bar{Q}\big)$. *If* $r < 1$ *then the sequence* $(\mathbf{e}[n])_{n\in\mathbb{N}}$ *converges linearly to* $\mathbf{0}$. *If* $r \geq 1$ *then* $\sum_{n\in\mathbb{N}} \mathbf{e}[n]$ *diverges.*

Remark on the lemma: It follows from results by Etessami and Yannakakis [24] that $r > 1$ (strictly greater) is impossible, but we do not need that here.

Proof. By Lemma B.3 we have $\mathbf{e}[n](t) > 0$ for $n \geq n_\perp$ iff $\mathbf{e}[n_\perp](t) > 0$ iff $t \in \mathcal{T}_\uparrow$. So, for the rest of the proof we can assume $\mathcal{T}_\uparrow \neq \emptyset$ and it suffices to consider the sequence $(\mathbf{e}_\uparrow[n])_{n\geq n_\perp}$. Notice that $\mathbf{e}_\uparrow[n] \gg \mathbf{0}$. (For this notation see the beginning of the appendix.)

Let $r < 1$. With Lemma A.2 it follows

$$\rho\left(\left(\bar{L} + \bar{P}(\mathbf{t})\right)^* \bar{Q}\right) < 1.$$

It follows using standard matrix theory (see e.g. [26]) that there is a norm $\|\cdot\|$ such that

$$r' := \left\|\left(\bar{L} + \bar{P}(\mathbf{t})\right)^* \bar{Q}\right\| < 1.$$

In this norm we have by Lemma B.8

$$\begin{aligned}
\|\mathbf{e}_\uparrow[n]\| &= \left\|\left(\bar{L} + \bar{P}(\mathbf{t} - \mathbf{e}[n-1])\right)^* \bar{Q}\mathbf{e}_\uparrow[n-1]\right\| \\
&\leq \left\|\left(\bar{L} + \bar{P}(\mathbf{t})\right)^* \bar{Q}\mathbf{e}_\uparrow[n-1]\right\| \\
&\leq r' \left\|\mathbf{e}_\uparrow[n-1]\right\| .
\end{aligned}$$

Using the equivalence of norms (see e.g. [26]) linear convergence is now immediate.

Now let $r \geq 1$. Let $C = \left(\bar{L} + \bar{P}(\mathbf{t})\right)^* \bar{Q}$. Then by Lemma A.2 we have $\rho(C) \geq 1$. By Lemma A.1 there is a set $\mathcal{T}_{\uparrow\uparrow} \subseteq \mathcal{T}_\uparrow$ such that the principal submatrix D obtained from C by deleting all rows and columns not indexed by $\mathcal{T}_{\uparrow\uparrow}$ is strongly connected and satisfies $\rho(D) \geq 1$.

In the following we denote, for any vector $\mathbf{v} \in \mathbb{R}^{\mathcal{T}}$, by $\mathbf{v}_\uparrow$ and $\mathbf{v}_{\uparrow\uparrow}$ the projection of $\mathbf{v}$ on $\mathcal{T}_\uparrow$ and $\mathcal{T}_{\uparrow\uparrow}$, respectively. We will show that $\sum_{n\in\mathbb{N}} \mathbf{e}_{\uparrow\uparrow}[n]$ diverges. We can restrict our attention to those $t \in \mathcal{T}$ that are reachable from $\mathcal{T}_{\uparrow\uparrow}$ via $\tilde{L} + \tilde{Q}(\mathbf{t}, \cdot) + \tilde{Q}(\cdot, \mathbf{t})$. To avoid notational clutter we simply assume that all $t \in \mathcal{T}$ are reachable from $\mathcal{T}_{\uparrow\uparrow}$. (Otherwise the corresponding coefficients and the corresponding matrix rows and columns can be removed in the straightforward way, without affecting the validity of what has been said. In particular, the matrix D stays exactly the same and Lemma B.8 stays valid for the remaining components and coefficients.)

For the proof it will be crucial that the Perron-Frobenius theorem (see e.g. [26]) implies that there is a nonnegative vector $\mathbf{u} \in \mathbb{R}^{\mathcal{T}_\uparrow}$ with $\mathbf{u}_{\uparrow\uparrow} \gg \mathbf{0}$ and $D\mathbf{u}_{\uparrow\uparrow} = \rho(D)\mathbf{u}_{\uparrow\uparrow} \geq \mathbf{u}_{\uparrow\uparrow}$ and $\mathbf{u}(t) = 0$ for $t \notin \mathcal{T}_{\uparrow\uparrow}$. As D is a principal submatrix of $\left(\bar{L} + \bar{P}(\mathbf{t})\right)^* \bar{Q}$ we also have

$$\left(\bar{L} + \bar{P}(\mathbf{t})\right)^* \bar{Q}\mathbf{u} \geq \mathbf{u}. \tag{4}$$

Let $t_{\min} = \min_{t\in\mathcal{T}}\{\mathbf{t}(t)\}$. Define a sequence $(\varepsilon_n)_{n\geq n_\perp}$ by $\varepsilon_n := t_{\min}^{-1} \cdot \max_{t\in\mathcal{T}}\{\mathbf{e}[n](t)\}$. As $\mathbf{e}_\uparrow[n] \gg \mathbf{0}$ and by definition of $\mathbf{e}[n]$, we have $\varepsilon_n \geq \varepsilon_{n+1} > 0$ for all n. We can assume w.l.o.g. that (ε_n) converges to 0, because otherwise the conclusion of the lemma ($\sum_{n\in\mathbb{N}} \mathbf{e}[n]$ diverges) is already satisfied. With the constant c from Lemma B.7 we have $\mathbf{e}_\uparrow[n](t) \geq c \cdot t_{\min} \cdot \varepsilon_n$ for all $t \in \mathcal{T}_{\uparrow\uparrow}$. By scaling down the vector $\mathbf{u}$ by multiplying it with a small positive number, we can accomplish $c \cdot t_{\min} \geq \mathbf{u}(t)$ for all t without changing the stated properties of $\mathbf{u}$. Since $\mathbf{u}(t) = 0$ for $t \notin \mathcal{T}_{\uparrow\uparrow}$, we can summarize:

$$\varepsilon_n \cdot \mathbf{t} \geq \mathbf{e}[n] \text{ and } \mathbf{e}_\uparrow[n] \geq \varepsilon_n \cdot \mathbf{u}\ . \tag{5}$$

Next, we show that there is an $n_1 \geq n_\perp$ and a $d > 0$ such that for all $n \geq n_1$ we have $\varepsilon_n d < 1$ and for all $i \in \mathbb{N}$

$$\mathbf{e}_\uparrow[n+i] \geq (1 - \varepsilon_n d)^i \varepsilon_n \mathbf{u}\ . \tag{6}$$

As $\mathbf{u}(t) = 0$ for $t \notin \mathcal{T}_{\uparrow\uparrow}$, it suffices to show $\mathbf{e}_\uparrow[n+i] \geq_{\uparrow\uparrow} (1 - \varepsilon_n d)^i \varepsilon_n \mathbf{u}$ where by the notation $\mathbf{v} \geq_{\uparrow\uparrow} \mathbf{w}$ we mean $\mathbf{v}_{\uparrow\uparrow} \geq \mathbf{w}_{\uparrow\uparrow}$.

We will determine the constants on the fly and proceed by induction on i. The base case ($i = 0$) is immediate from (5). Let $i \geq 0$. Then

$$
\begin{aligned}
& \mathbf{e}_{\uparrow}[n+i+1] \\
& \quad = \quad \big(\bar{L} + \bar{P}(\mathbf{t} - \mathbf{e}[n+i])\big)^{*}\bar{Q}\mathbf{e}_{\uparrow}[n+i] \\
& \qquad\qquad \text{(by Lemma B.8)} \\
& \quad \geq \quad \big((1-\varepsilon_{n+i})\big(\bar{L} + \bar{P}(\mathbf{t})\big)\big)^{*}\,\bar{Q}\mathbf{e}_{\uparrow}[n+i] \\
& \qquad\qquad \text{(by (5))} \\
& \quad \geq \quad \big((1-\varepsilon_{n})\big(\bar{L} + \bar{P}(\mathbf{t})\big)\big)^{*}\,\bar{Q}(1-\varepsilon_n d)^{i}\varepsilon_n\mathbf{u} \\
& \qquad\qquad \text{(induction hypothesis)} \\
& \quad \geq \quad (I - \varepsilon_n K)\big(\bar{L} + \bar{P}(\mathbf{t})\big)^{*}\bar{Q}(1-\varepsilon_n d)^{i}\varepsilon_n\mathbf{u} \\
& \qquad\qquad \text{(for a large } n_1 \text{ and some } K \text{ by Lemma A.3)} \\
& \quad \geq \quad (1-\varepsilon_n d)^{i}\varepsilon_n\big(\mathbf{u} - \varepsilon_n K\big(\bar{L} + \bar{P}(\mathbf{t})\big)^{*}\bar{Q}\mathbf{u}\big) \\
& \qquad\qquad \text{(by (4))} \\
& \quad \geq_{\uparrow\uparrow} \quad (1-\varepsilon_n d)^{i}\varepsilon_n(\mathbf{u} - \varepsilon_n d\mathbf{u}) \\
& \qquad\qquad \text{(for a large } d \text{ such that} \\
& \qquad\qquad\quad K\big(\bar{L} + \bar{P}(\mathbf{t})\big)^{*}\bar{Q}\mathbf{u} \leq_{\uparrow\uparrow} d\mathbf{u}) \\
& \quad = \quad (1-\varepsilon_n d)^{i+1}\varepsilon_n\mathbf{u}\ .
\end{aligned}
$$

We have

$$
\begin{aligned}
& \sum_{i=0}^{k} \mathbf{e}_{\uparrow}[n+i] \\
& \qquad \geq \sum_{i=0}^{k} (1-\varepsilon_n d)^{i}\varepsilon_n\mathbf{u} && \text{(by (6))} \\
& \qquad \geq \frac{1-(1-\varepsilon_n d)^{k+1}}{1-(1-\varepsilon_n d)} \cdot \varepsilon_n \cdot \mathbf{u} \\
& \qquad = \frac{1-(1-\varepsilon_n d)^{k+1}}{d} \cdot \mathbf{u}\ ,
\end{aligned}
$$

so, for every $n \geq n_{\perp}$, there exists some $k(n)$ such that $\sum_{i=n}^{k(n)} \mathbf{e}_{\uparrow}[i] \geq \frac{1}{2d} \cdot \mathbf{u}$. Hence the sum

$$
\begin{aligned}
\sum_{i=n_{\perp}}^{\infty} \mathbf{e}_{\uparrow}[i] &= \sum_{i=n_{\perp}}^{k(n_{\perp})} \mathbf{e}_{\uparrow}[i] + \sum_{i=k(n_{\perp})+1}^{k(k(n_{\perp})+1)} \mathbf{e}_{\uparrow}[i] + \cdots \\
&\geq \frac{1}{2d} \cdot \mathbf{u} + \frac{1}{2d} \cdot \mathbf{u} + \cdots
\end{aligned}
$$

diverges since $\mathbf{u}_{\uparrow\uparrow} \gg \mathbf{0}$. □

The following lemma provides a lower bound on $\mathbf{p}[n](p_0X_0)$ in terms of the spectral radius ρ of A. Its proof is, in large parts, quite similar to the previous proof.

Lemma B.10. *Suppose that $EM_{p_0X_0}$ is finite (sic!). Then there is a number $b > 0$ with*

$$\mathbf{p}[n](p_0X_0) \geq b\rho^n \quad \textit{for all } n \geq 1.$$

Proof. We first show that $\sum_n \mathbf{e}[n](pXq)$ converges. Assume for a contradiction that $\sum_n \mathbf{e}[n](pXq)$ diverges for some $pXq \in \mathcal{T}$. Then $\sum_n \mathbf{p}[n](pX)$ diverges as well because $\mathbf{e}[n](pXq) \leq \mathbf{p}[n](pX)$. Notice that in fact $pX \in \mathcal{H}$ because $\mathbf{e}[n](pXq) > 0$ for all $n \geq 1$. Hence, EM_{pX} is infinite which implies that $EM_{p_0X_0}$ is infinite because p_0X_0 can reach pX. This contradicts the precondition of the lemma requiring $EM_{p_0X_0}$ to be finite.

Hence, Lemma B.9 guarantees that $(\mathbf{e}[n])_n$ converges (at least) linearly, i.e., there are constant numbers $r, s > 0$ with $s < 1$ such that $\mathbf{e}[n] \leq rs^n\mathbf{t}$, because $\mathbf{t} \gg \mathbf{0}$. Setting $\varepsilon_n := rs^n$ we have

$$\mathbf{e}[n] \leq \varepsilon_n \mathbf{t} \; . \tag{7}$$

By Lemma A.1 there is a set $\mathcal{H}_{\uparrow\uparrow} \subseteq \mathcal{H}$ such that the principal submatrix D obtained from A by deleting all rows and columns not indexed by $\mathcal{H}_{\uparrow\uparrow}$ is strongly connected and satisfies $\rho(D) \geq 1$.

In the following we denote, for any vector $\mathbf{v} \in \mathbb{R}^{\mathcal{H}}$, by $\mathbf{v}_{\uparrow\uparrow}$ the projection of $\mathbf{v}$ on $\mathcal{H}_{\uparrow\uparrow}$. The Perron-Frobenius theorem (see e.g. [26]) implies that there is a nonnegative vector $\mathbf{u} \in \mathbb{R}^{\mathcal{H}}$ with $\mathbf{u}_{\uparrow\uparrow} \gg \mathbf{0}$ and $D\mathbf{u}_{\uparrow\uparrow} = \rho\mathbf{u}_{\uparrow\uparrow}$ and $\mathbf{u}(h) = 0$ for $h \notin \mathcal{H}_{\uparrow\uparrow}$. As D is a principal submatrix of A we also have

$$A\mathbf{u} \geq \rho\mathbf{u} \; . \tag{8}$$

Now we show that there is an $n_1 \geq n_\perp$ and a $d > 0$ such that $\varepsilon_{n_1-1}d < 1$ and for all $n \geq 0$

$$\mathbf{p}[n_1 + n] \geq \left(\prod_{i=1}^{n} (1 - \varepsilon_{n_1+i-1}d) \right) \rho^n \mathbf{u} \; . \tag{9}$$

As $\mathbf{u}(h) = 0$ for $h \notin \mathcal{H}_{\uparrow\uparrow}$, it suffices to show $\mathbf{p}[n_1 + n] \geq_{\uparrow\uparrow} \prod_{i=1}^{n}(1 - \varepsilon_{n_1+i-1}d)\rho^n\mathbf{u}$ where by the notation $\mathbf{v} \geq_{\uparrow\uparrow} \mathbf{w}$ we mean $\mathbf{v}_{\uparrow\uparrow} \geq \mathbf{w}_{\uparrow\uparrow}$.

We proceed by induction on n. For the induction base ($n = 0$) observe that, due to $\mathbf{p}[n_1] \gg \mathbf{0}$, we can enforce $\mathbf{p}[n_1] \geq \mathbf{u}$ by scaling down $\mathbf{u}$ by multiplying it with a small scalar. This does not change the stated properties of $\mathbf{u}$.

Let $n \geq 0$. Now we have

$$\begin{aligned}
&\mathbf{p}[n_1 + n + 1] \\
&\quad = (L + Q(\mathbf{t}[n_1 + n - 1], \cdot))^* L' \mathbf{p}[n_1 + n] \\
&\quad = (L + Q(\mathbf{t} - \mathbf{e}[n_1 + n], \cdot))^* L' \mathbf{p}[n_1 + n] \\
&\quad \geq (L + Q((1 - \varepsilon_{n_1+n})\mathbf{t}, \cdot))^* L' \mathbf{p}[n_1 + n] \\
&\qquad\quad \text{(by (7))} \\
&\quad \geq \big((1 - \varepsilon_{n_1+n})(L + Q(\mathbf{t}, \cdot))\big)^* L' \mathbf{p}[n_1 + n] \\
&\quad \geq \big((1 - \varepsilon_{n_1+n})(L + Q(\mathbf{t}, \cdot))\big)^* L' \prod_{i=1}^{n} (1 - \varepsilon_{n_1+i-1} d) \rho^n \mathbf{u} \\
&\qquad\quad \text{(induction hypothesis)} \\
&\quad \geq (I - \varepsilon_{n_1+n} K) A \prod_{i=1}^{n} (1 - \varepsilon_{n_1+i-1} d) \rho^n \mathbf{u} \\
&\qquad\quad \text{(for a large } n_1 \text{ and some } K \text{ by Lemma A.3)} \\
&\quad \geq \prod_{i=1}^{n} (1 - \varepsilon_{n_1+i-1} d)(\rho^{n+1} \mathbf{u} - \varepsilon_{n_1+n} K A \rho^n \mathbf{u}) \\
&\qquad\quad \text{(by (8))} \\
&\quad \geq_{\uparrow\uparrow} \prod_{i=1}^{n} (1 - \varepsilon_{n_1+i-1} d)(\rho^{n+1} \mathbf{u} - \varepsilon_{n_1+n} d \rho^{n+1} \mathbf{u}) \\
&\qquad\quad \text{(for a large } d \text{ such that } KA\mathbf{u} \leq_{\uparrow\uparrow} d\rho\mathbf{u}\text{)} \\
&\quad = \prod_{i=1}^{n+1} (1 - \varepsilon_{n_1+i-1} d) \rho^{n+1} \mathbf{u}
\end{aligned}$$

This proves (9). We have

$$b' := \prod_{i=n_1}^{\infty} (1 - \varepsilon_i d) = \prod_{i=n_1}^{\infty} (1 - r s^i d) > 0\,,$$

as $1 - rs^i d \geq 1 - 1/i^2$ is true for almost all natural numbers i and $\prod_{i=2}^{\infty}(1 - 1/i^2) = \frac{1}{2} > 0$.

Let $p_1 X_1 \in \mathcal{H}_{\uparrow\uparrow}$. Recall that $\mathbf{u}(p_1 X_1) > 0$. We have for all $n \geq 1$:

$$\begin{aligned}
\mathbf{p}[n](p_0 X_0) &\geq b'' \mathbf{p}[n](p_1 X_1) && \text{(for a } b'' > 0\text{, as } p_1 X_1 \Gamma^* \text{ is reachable from } p_0 X_0\text{)} \\
&\geq b'' \mathbf{p}[n_1 + n](p_1 X_1) && \text{(as } (\mathbf{p}[n])_n \text{ is monotonically decreasing)} \\
&\geq b'' b' \mathbf{u}(p_1 X_1) \rho^n && \text{(by Equation (9))}\,,
\end{aligned}$$

so the lemma holds for $b := b'' b' \mathbf{u}(p_1 X_1)$. □

Now we complete the proof of Lemma B.2.

Proof. Let $\rho < 1$. By Proposition 3.10 we have

$$\mathbf{p}[n] = (L + Q(\mathbf{t}[n-2], \cdot))^* L' \mathbf{p}[n-1] \leq A\mathbf{p}[n-1] \ .$$

By an easy induction we obtain $\mathbf{p}[n_\perp + n] \leq A^n \mathbf{p}[n_\perp]$ and so $\sum_{n=0}^{\infty} \mathbf{p}[n_\perp + n] \leq A^* \mathbf{p}[n_\perp]$. By Lemma A.1, $\rho < 1$ implies that the matrix sum A^* converges, so $\sum_{n=1}^{\infty} \mathbf{p}[n]$ is finite, in particular $EM_{p_0 X_0}$ is finite.

For the other direction, let $EM_{p_0 X_0}$ be finite. We have $EM_{p_0 X_0} = \sum_{n \geq 1} \mathbf{p}[n](p_0 X_0)$, so $\sum_{n \geq 1} \mathbf{p}[n](p_0 X_0)$ converges. By Lemma B.10, $\sum_{n \geq 1} b\rho^n$ converges as well. Hence, $\rho < 1$. □

B.8 Proof of Theorem 3.13

Here is a restatement of Theorem 3.13.

THEOREM 3.13. *Let* $n_\perp := |Q|^2|\Gamma| + 3$ *and* $\widehat{\mathbf{p}}[n] := \mathbf{p}[n]$ *for* $n < n_\perp$ *and* $\widehat{\mathbf{p}}[n_\perp + n] := A^n \mathbf{p}[n_\perp]$ *for* $n \geq 0$*. Then we have* $\mathbf{p}[n] \leq \widehat{\mathbf{p}}[n]$ *for all* $n \geq 1$*. Moreover, there exists* d *with* $0 < d \leq 1$ *and*

$$d \cdot \widehat{\mathbf{p}}[n](p_0 X_0) \leq \mathbf{p}[n](p_0 X_0) \leq \widehat{\mathbf{p}}[n](p_0 X_0) \ .$$

Proof. We first prove the upper bound by induction on n. The induction base ($n \leq n_\perp$) is trivial. For $n \geq n_\perp$ we have:

$$\begin{aligned}
\mathbf{p}[n+1] &= A_{n+1}\mathbf{p}[n] && \text{(Proposition 3.10)} \\
&\leq A\mathbf{p}[n] && (A_n \leq A) \\
&\leq A\widehat{\mathbf{p}}[n] && \text{(induction hypothesis)} \\
&= \widehat{\mathbf{p}}[n+1] && \text{(definition of } \widehat{\mathbf{p}}[n+1])
\end{aligned}$$

It remains to show the lower bound. Letting ρ denote the spectral radius of A, there is, by standard matrix theory [26], a number $c > 0$ with $\widehat{\mathbf{p}}[n](p_0 X_0) \leq c\rho^n$ for all $n \geq 1$. By Lemma B.10 we have $b\rho^n \leq \mathbf{p}[n](p_0 X_0)$. So the theorem holds with $d := b/c$. □

B.9 Proof of Theorem 3.16

Here is a restatement of Theorem 3.16.

THEOREM 3.16. *Suppose that $\mathrm{EM}_{p_0X_0}$ is finite. Let $\mathrm{UM}_{p_0X_0}(k) := \sum_{n=1}^{k} \mathbf{p}[n](p_0X_0)$. For all $k \geq |Q|^2|\Gamma| + 3$*

$$\mathrm{EM}_{p_0X_0} - \mathrm{UM}_{p_0X_0}(k) \leq \|A^*\|_1 \|\mathbf{p}[k]\|_1 \leq ab^k$$

where $a > 0$ and $0 < b < 1$ are computable rational numbers. Hence, the sequence $(\mathrm{UM}_{p_0X_0}(k))_k$ converges linearly to $\mathrm{EM}_{p_0X_0}$.

Proof. Let $\mathrm{UM}_{pX}(k) := \sum_{n=1}^{k} \mathbf{p}[n](pX)$ for all $pX \in \mathcal{H}$. Recall that $\mathrm{EM}_{pX} = \lim_{k\to\infty} \mathrm{UM}_{pX}(k)$. Hence, the "error vector" $\boldsymbol{\delta}[k] \in \mathbb{R}^{\mathcal{H}}$ with

$$\delta[k](pX) = \mathrm{EM}_{pX} - \mathrm{UM}_{pX}(k-1) = \sum_{n=k}^{\infty} \mathbf{p}[n](pX)$$

converges to $\mathbf{0}$. We prove the theorem by showing $\|\boldsymbol{\delta}[k](pX)\|_1 \leq \|A^*\|_1 \|\mathbf{p}[k]\|_1 \leq ab^k$.

By Proposition 3.10, for $n \geq n_\perp = |Q|^2|\Gamma| + 3$

$$\mathbf{p}[n] = (L + Q(\mathbf{t}[n-2], \cdot))^* L' \mathbf{p}[n-1] \leq A\mathbf{p}[n-1]$$

where $\rho < 1$ by Lemma B.2. By a simple induction, we get $\mathbf{p}[k+n] \leq A^n \mathbf{p}[k]$ and so $\boldsymbol{\delta}[k] \leq A^* \mathbf{p}[k]$ which converges. So we have

$$\|\boldsymbol{\delta}[k]\|_1 \leq \|A^*\|_1 \|\mathbf{p}[k]\|_1 \text{ for all } k \geq n_\perp$$

which shows the first inequality.

For the second inequality it suffices to compute rational numbers $r, b > 0$ with $b < 1$ such that $\|\mathbf{p}[n_\perp + k]\|_1 \leq rb^k$. By the argument above we have $\mathbf{p}[n_\perp + k] \leq A^k \mathbf{p}[n_\perp]$ with $\rho < 1$. First we compute an upper bound $b < 1$ on ρ as follows. Note that one can express $\mathbf{t}$ in $\mathbf{ExTh}(\mathbb{R})$. Furthermore one can decide for any rational number s whether $\rho \geq s$ because this is equivalent to the existence of a nonnegative, nonzero vector $\mathbf{x}$ with $A\mathbf{x} \geq s\mathbf{x}$, see [4] Thm. 2.1.11 and cf. [24]. So, a bound b can be computed using a bisection method.

Now we can compute a rational number r with $\|\mathbf{p}[n_\perp + k]\|_1 \leq rb^k$ as follows. By standard matrix theory there exists a vector norm $\|\cdot\|_*$ that induces a matrix norm with $\|A\|_* \leq b$. Following the proof of Lemma 5.6.10 in [26] the norm $\|\cdot\|_*$ is given by $\|\mathbf{x}\|_* := \|DQ\mathbf{x}\|_1$ where Q is the unitary similarity matrix of a Schur factorization of A and D is a nonsingular diagonal matrix with large enough entries. (Strictly speaking, [26] defines a matrix norm directly, but it is easy to verify that the vector norm $\|\cdot\|_*$ defined above induces this matrix norm.)

Recall that a Schur factorization of A is given by two matrices Q, U such that $A = Q^{-1}UQ$, where Q is a unitary matrix and U is an upper diagonal matrix. (Those properties imply that the entries on the main diagonal of U are the eigenvalues of A.) The Schur factorization of A can be expressed in **ExTh**($\mathbb{R}$) by handling complex numbers using pairs of real numbers. It is easy to see that the matrix D given in [26] can also be expressed in **ExTh**($\mathbb{R}$).

Now we have

$$\begin{aligned}
\|\mathbf{p}[n_\perp + k]\|_1 &\le \left\|A^k \mathbf{p}[n_\perp]\right\|_1 \\
&= \left\|Q^{-1}D^{-1}DQA^k \mathbf{p}[n_\perp]\right\|_1 \\
&\le \left\|Q^{-1}D^{-1}\right\|_1 \left\|DQA^k \mathbf{p}[n_\perp]\right\|_1 \\
&= \left\|Q^{-1}D^{-1}\right\|_1 \left\|A^k \mathbf{p}[n_\perp]\right\|_* \\
&\le \left\|Q^{-1}D^{-1}\right\|_1 \|A\|_*^k \|\mathbf{p}[n_\perp]\|_* \\
&= \left\|Q^{-1}D^{-1}\right\|_1 \|A\|_*^k \|DQ\mathbf{p}[n_\perp]\|_1 \\
&\le \left\|Q^{-1}D^{-1}\right\|_1 b^k \|DQ\|_1 \|\mathbf{p}[n_\perp]\|_1 \\
&\le \left\|Q^{-1}D^{-1}\right\|_1 b^k \|DQ\|_1 |\mathcal{H}|,
\end{aligned}$$

so choose $r \ge \left\|Q^{-1}D^{-1}\right\|_1 \|DQ\|_1 |\mathcal{H}|$ and it follows $\|\mathbf{p}[n_\perp + k]\|_1 \le rb^k$. □

C Proofs of Section 4

C.1 Proof of Theorem 4.1 and Proposition 4.2

Here are restatements of Theorem 4.1 and Proposition 4.2.

THEOREM 4.1. *The problem whether $EM_{p_0X_0}$ is finite is decidable in polynomial space.*

PROPOSITION 4.2. *Suppose $\mathcal{P}(M_{p_0X_0} < \infty) = 1$. Then the matrix A exists. Moreover, its spectral radius ρ satisfies $\rho < 1$ if and only if $EM_{p_0X_0}$ is finite.*

Proposition 4.2 is immediate from Lemma B.1 and Lemma B.2. So it suffices to show Theorem 4.1.

Proof. In [17] it was shown that the problem whether $\mathcal{P}(M_{p_0X_0} = \infty) > 0$ is decidable in polynomial space. If $\mathcal{P}(M_{p_0X_0} = \infty) > 0$, then clearly $\mathrm{E}M_{p_0X_0}$ is infinite. Otherwise, i.e. if $\mathcal{P}(M_{p_0X_0} < \infty) = 1$, we use the criterion of Lemma B.2, i.e., we check whether or not the spectral radius ρ of $A = (L + Q(\mathbf{t}, \cdot))^* L'$ satisfies $\rho < 1$. The vector $\mathbf{t}$ can be expressed in **ExTh**($\mathbb{R}$), see e.g. [17]. Hence, the same applies to A.

Standard theory on nonnegative matrices implies (see e.g. [4] Thm. 2.1.11 and cf. [24]) that $\rho \geq s$ iff there is a nonnegative nonzero vector $\mathbf{x}$ with $A\mathbf{x} \geq s\mathbf{x}$, so $\rho \geq 1$ is expressible in **ExTh**($\mathbb{R}$) which is decidable in polynomial space [12, 33]. □

C.2 Discussion of the Lower Bound and Proof of Theorem 4.3

The following exposition of those problems is essentially taken from Etessami and Yannakakis [24]. Also the reduction to our finite-stack decision problem is based on a reduction of [24].

SQRT-SUM is the following problem: given natural numbers $d_1, \ldots, d_n \in \mathbb{N}$ and another number $k \in \mathbb{N}$, decide whether $\sqrt{d_1} + \cdots + \sqrt{d_n} \leq k$. The PosSLP (Positive Straight-Line Program) decision problem asks whether a given straight-line program or, equivalently, arithmetic circuit with operations $+, -, \cdot$, and inputs 0 and 1, and a designated output gate, outputs a positive integer or not [2]. SQRT-SUM, a long-standing open problem in the complexity of numerical computation, reduces to PosSLP via a P-time Turing reduction, see [2]. Both the SQRT-SUM and the PosSLP problem can be solved in PSPACE. Their complexity was recently lowered slightly to the 4th level of the Counting Hierarchy [2]. PosSLP is a fundamental problem of numerical computation; it is complete for the class of decision problems that can be solved in polynomial time on models with unit-cost exact rational arithmetic, see [2, 24] for more details.
Now we prove Theorem 4.3 which we restate here:

THEOREM 4.3. *The* PosSLP *problem is P-time many-one reducible to the decision problem whether the expected maximal height of a pPDA is finite.*

Proof. As a gadget for their reduction to the decision problem whether a pPDA terminates with probability one, Etessami and Yannakakis [24] compute, given a PosSLP instance, a pPDA Δ with the following properties. (Strictly speaking, they construct an equivalent Recursive Markov Chain.) The starting configuration is pX, and after having left the initial configuration, Δ reaches the control state p again with probability 1. When

Δ reaches p, the configuration is $p\varepsilon$ with some probability a, and pXX with probability $1 - a$. Moreover, the time needed to reach either of those configurations is essentially bounded by the size of the given PosSLP instance. In particular, this time is finite, and so is the expected maximal height of the stack. Furthermore, the given PosSLP instance is a "yes instance" iff $a > \frac{1}{2}$.

Now it is easy to see that EM_{pX} is finite in Δ iff it is finite in the pPDA Δ' that consists only of the transitions $pX \overset{a}{\to} p\varepsilon$ and $pX \overset{1-a}{\to} pXX$. So it suffices to show that EM_{pX} is finite in Δ' iff $a > \frac{1}{2}$.

If $a < \frac{1}{2}$, then it is easy to show with the method from [17] that $\mathcal{P}(M_{pX} = \infty) > 0$ and so $EM_{pX} = \infty$.

If $a \geq \frac{1}{2}$, then by [16, 24] we clearly have $[pXp] = 1$. Consequently, we can apply Lemma B.2 that states that EM_{pX} is finite iff

$$(1-a)^* \cdot (1-a) = \frac{1-a}{1-(1-a)} = \frac{1-a}{a} < 1$$

which holds iff $a > \frac{1}{2}$. This completes the proof.

We remark that Δ contains more control states than just p. So, the reduction does not work for pBPA. $\square$

C.3 Proof of Proposition 4.5

Here is a restatement of Proposition 4.5.

PROPOSITION 4.5. *If* $X_0 \in$ *Term, the problem whether* $EM_{X_0} < \infty$ *is decidable in polynomial time.*

Proof. As $X_0 \in$ *Term*, all symbols reachable from X_0 are in *Term* as well. It follows $\mathbf{t} = \mathbf{1}$ and so all entries of the matrix $A = (L + Q(\mathbf{t}, \cdot))^* L'$ are rational. Since the sets $\mathcal{H}$ and $\mathcal{T}$ are computable in polynomial time using Lemma 3.6, it follows that A can be computed in polynomial time.

As $X_0 \in$ *Term*, we clearly have $\mathcal{P}(M_{X_0} < \infty) = 1$, so the criterion of Lemma B.2 is applicable, which states that $EM_{X_0} = \infty$ holds iff the spectral radius ρ of A satisfies $\rho \geq 1$. Standard theory on nonnegative matrices implies (see e.g. [4] Thm. 2.1.11 and cf. [24]) that $\rho \geq 1$ iff there is a nonnegative nonzero vector $\mathbf{x}$ with $A\mathbf{x} \geq \mathbf{x}$. As A has rational entries, that condition can be decided in polynomial time using linear programming. $\square$

C.4 Proof of Proposition 4.6

Here is a restatement of Proposition 4.6.

PROPOSITION 4.6. *If $X_0 \in NTerm$, then the problem whether $\mathcal{P}(M_{X_0} < \infty) = 1$ is decidable in polynomial time.*

Proof. The result follows immediately from [17] Theorem 6.2 and the fact, that the transition structure of the finite Markov chain **X**, as defined in [17], can be computed in polynomial time for pBPA using methods of [24] (see also [7]). □

C.5 Proof Sketch and Full Proof of Theorem 4.8

Here is a restatement of Theorem 4.8.

THEOREM 4.8. *The algorithm on page 15 returns 'yes' iff EM_{X_0} is finite. It runs in polynomial time.*

Proof sketch.

Clearly, if the algorithm returns 'no', then $EM_{X_0} = \infty$. Assume that the algorithm returns 'yes'.

Let RH denote the set of all $Y \in NTerm$ such that almost all runs of $Run(Y)$ reach $Y\Gamma^* \cup \{\varepsilon\}$. We start by proving that for every Y in RH we have that $\mathcal{P}(Y \to^+ Y) = 1$ and that all runs $w \in Run(Y)$ satisfy $M_Y(w) \leq |\Gamma|$.

First, if there is a path from Y to some configuration $Y\alpha$ where $|\alpha| > 0$, then in almost all non-terminating runs of $Run(Y)$ the stack height grows unboundedly.[1] It follows that $\mathcal{P}(Y \to^+ \{Y, \varepsilon\}) = 1$. Now a straightforward argument shows that $\mathcal{P}(Y \to^+ \varepsilon) = 0$, because otherwise we would have $\mathcal{P}(Y \to^+ \varepsilon) = 1$, which would contradict $Y \in RH$. It follows that $\mathcal{P}(Y \to^+ Y) = 1$.

Now assume that there is a path from Y to a configuration α satisfying $|\alpha| > |\Gamma|$. Because $\mathcal{P}(Y \to^+ Y) = 1$, there is a path from α to Y, and thus also a path from Y to Y in which the stack height exceeds $|\Gamma|$. By Lemma 3.5, for every $n \geq 0$, there is a path u_n from Y to Y such that the maximal stack height is at least n in u_n. However, almost all runs of $Run(Y)$ follow all paths u_n infinitely many times, which means that the maximal

[1] This can be confirmed using the results of [17] Section 6.1; in particular, using the terminology of [17], in this situation some BSCC of the chain **X** contains a non-limited transition $(Z, m) \to (Y, +)$ which implies the unboundedness of the stack height in nonterminating runs of $Run(Y)$.

stack height is almost surely infinite, a contradiction with the step 2. of the algorithm. Hence, $\mathcal{P}(Y \to^+ Y) = 1$ and all runs $w \in \mathit{Run}(Y)$ satisfy $M_Y(w) \leq |\Gamma|$.

Let us define

$$R = \{Y\alpha \in \Gamma^* \mid Y \in RH\} \cup \{\varepsilon\}.$$

The results of [7] Section 4.3.1 imply that almost all runs of $\mathit{Run}(X_0)$ reach R. Hence, it suffices to concentrate on the expected maximal stack height before reaching R.

Intuitively, almost all runs w of $\mathit{Run}(X_0)$ that reach R behave as follows: First, w proceeds through some configurations with the head in *NTerm* until either a configuration of R, or a configuration of the form $Y\alpha$ where $Y \in \mathit{Term}$ is entered. In the latter case, w follows a finite 'bump' from $Y\alpha$ to α (note that the expected maximal height of this bump is equal to $EM_Y + |\alpha|$ which is finite by the step 3. of the above algorithm). Then w proceeds from α through configurations with the heads in *NTerm* until either a configuration of R, or some other configuration with the head in *Term* is entered, and so on. Let $T(w)$ be the number of configurations with the head in *NTerm* plus the number of 'bumps' entered before reaching R. We prove that ET is finite, which allows us to bound EM_{X_0} in terms of the product of ET and $\max_{Y \in \mathit{Term}} EM_Y$ (more concretely, we show that $EM_{X_0} \leq ET \cdot \max_{Y \in \mathit{Term}} EM_Y + |\Gamma|$).

The intuition behind the proof of the finiteness of ET is the following. For every $Y \in \mathit{NTerm}$ there is $c_Y \geq 1$ such that, with a probability $\epsilon_Y > 0$, a run goes from Y to $RH \cdot \Gamma^*$ via c_Y configurations with the head in *NTerm*. Observe that the number of bumps in such runs cannot be greater than $2c_Y$. Let $\epsilon = \min_{Y \in \mathit{NTerm}} \epsilon_Y$ and $c = \max_{Y \in \mathit{NTerm}} c_Y$. A straightforward argument shows that $\mathcal{P}(T \geq 3cn)$ is at most $(1-\epsilon)^n$, and thus

$$ET \leq \sum_{n=0}^{\infty} 3c\mathcal{P}(T \geq 3cn) \leq \frac{3c}{\epsilon} < \infty \qquad \square$$

In the following we give a full proof of Theorem 4.8. In the course of the proof we repeat some definitions from the sketch presented above.

Assume that the algorithm deciding whether EM_{X_0} is finite returns 'yes'. Let RH denote the set of all $Y \in \mathit{NTerm}$ such that almost all runs of $\mathit{Run}(Y)$ reach $Y\Gamma^* \cup \{\varepsilon\}$. The following proposition was already proved in the sketch. Let us define

$$R = \{Y\alpha \in \Gamma^* \mid Y \in RH\} \cup \{\varepsilon\}.$$

Proposition C.1. *For every Y in RH we have that $\mathcal{P}(Y \to^+ Y) = 1$ and that all runs $w \in \mathit{Run}(Y)$ satisfy $M_Y(w) \leq |\Gamma|$. Moreover, almost all runs of $\mathit{Run}(X_0)$ reach R.*

The results of [7] Section 4.3.1 imply that almost all runs of $Run(X_0)$ reach R. (Indeed, using the terminology of [7], the set R is equal to the union of all BSCCs of $\mathbf{X}_\Delta$, and hence almost all runs in $\mathbf{X}_\Delta$ reach R. Then, by [7] Lemma 4.3.6, almost all nonterminating runs of $Run(X)$ reach $\{Y\alpha \in \Gamma^* \mid Y \in RH\}$.)

We denote by $\mathrm{CRun}(X_0)$ the set of all runs $w \in Run(X_0)$ that satisfy the following two conditions:

1. for some $i \geq 0$ holds $w(i) \in R$;
2. for every $i \geq 0$ satisfying $head(w(i)) \in Term$ there is $j \geq i$ such that $w(j) = tail(w(i))$.

Proposition C.2. $\mathcal{P}(\mathrm{CRun}(X_0)) = 1$

Proof. We argued above that almost all runs satisfy the condition 1. from the definition of $\mathrm{CRun}(X_0)$. Let $\mathcal{A} \subseteq Run(X_0)$ be the set of all runs that satisfy the condition 2. Clearly, in order to prove $\mathcal{P}(\mathrm{CRun}(X_0)) = 1$, it suffices to show $\mathcal{P}(\mathcal{A}) = 1$.

Given $Y \in Term$, we denote by D_Y the set of all paths v from X_0 to $\{Y\alpha \mid \alpha \in \Gamma^*\}$. For every $v \in \bigcup_{Y \in Term} D_Y$, we denote by $\mathcal{D}(v)$ the set of all runs $w \in Run(v)$ such that for all $i \geq |v|$ holds $|w(i)| \geq |v(|v|-1)|$ (remember that $v(|v|-1)$ is the last state of v). Clearly,

$$Run(X_0) \backslash \mathcal{A} = \bigcup_{Y \in Term} \bigcup_{v \in D_Y} \mathcal{D}(v)$$

It is easy to see that for every $v \in D_Y$ holds

$$\mathcal{P}(\mathcal{D}(v)) = \mathcal{P}(Run(v)) \cdot (1 - \mathcal{P}(Y \rightarrow^* \{\varepsilon\})) = 0$$

Hence,

$$\begin{aligned} \mathcal{P}(Run(X_0) \backslash \mathcal{A}) &= \mathcal{P}(\bigcup_{Y \in Term} \bigcup_{v \in D_Y} \mathcal{D}(v)) \\ &\leq \sum_{Y \in Term} \sum_{v \in D_Y} \mathcal{P}(\mathcal{D}(v)) \\ &= 0 \end{aligned}$$

which implies that $\mathcal{P}(\mathcal{A}) = 1$ and $\mathcal{P}(\mathrm{CRun}(X_0)) = 1$. $\square$

We define a sequence of random variables $I_1, I_2, \ldots$ as follows: Given $w \in \mathrm{CRun}(X_0)$ we put

- $I_1(w) = 0$

- for $n \geq 2$
  - if either $head(w(I_n(w))) \in NTerm$, or $w(I_n(w)) = \varepsilon$, then $I_{n+1}(w) = I_n(w) + 1$
  - if $head(w(I_n(w))) \in Term$, then $I_{n+1}(w)$ is the least number $\ell \geq I_n(w)$ such that $|w(\ell)| = |w(I_n(w))| - 1$

We define a sequence of random variables $K_1, K_2, \ldots$ as follows: Given $w \in CRun(X_0)$ and $n \geq 1$ we put

$$K_n(w) = \max\{|w(j)| - |w(I_n(w))| \mid I_n(w) \leq j < I_{n+1}(w)\}$$

Let $T : CRun(X_0) \to \mathbb{N}$ be a random variable such that $T(w)$ is the *least* number satisfying $w(I_{T(w)}(w)) \in R$.

Proposition C.3. *For all* $w \in CRun(X_0)$ *we have that*

$$M_X(w) \leq T(w) + \sum_{n=1}^{T(w)} K_n(w) + |\Gamma|$$

Proof. Let us fix $w \in CRun(X_0)$. Let us define

$$M = \sup\{|w(i)| \mid i \geq I_{T(w)}(w)\}$$

and

$$m = \max\{|w(i)| \mid 0 \leq i < I_{T(w)}(w)\}$$

Clearly, $M_X(w) \leq \max\{M, m\}$. By Proposition C.1, $M \leq T(w) + |\Gamma|$. Also,

$$\begin{aligned} m &\leq \max_{1 \leq n < T(w)} \max_{I_n(w) \leq i < I_{n+1}(w)} |w(i)| \\ &\leq \max_{1 \leq n < T(w)} K_n(w) + T(w) \\ &\leq T(w) + \sum_{n=1}^{T(w)} K_n(w) \end{aligned}$$

Hence,

$$M_X(w) \leq T(w) + \sum_{n=1}^{T(w)} K_n(w) + |\Gamma|$$

□

Proposition C.4. *The expected value* ET *of the variable* T *is finite.*

Proof. Assume that $X_0 \in \mathit{NTerm} \setminus RH$ (otherwise $ET \leq 2$). We have already argued that for every $Y \in \mathit{NTerm}$ holds $\mathcal{P}(Y \to^* R) = 1$. Thus, for every $Y \in \mathit{NTerm}$ there is a finite path ν_Y initiated in Y such that $\mathit{head}(\nu_Y(|\nu_Y| - 1)) \in RH$ (remember that $\nu_Y(|\nu_Y| - 1)$ is the last state of ν_Y) and for every $0 \leq i < |\nu_Y| - 1$ holds $\mathit{head}(\nu(i)) \notin RH$. Let $\epsilon_Y = \mathcal{P}(\mathit{Run}(\nu_Y))$. We put

$$\epsilon = \min_{Y \in \mathit{NTerm}} \epsilon_Y$$

Note that $\epsilon > 0$. Let us define $c_Y = |\nu_Y|$ and $c = \max_{Y \in \mathit{NTerm}} c_Y$.

We define a sequence $B_0, B_1, \ldots$ of sets of finite paths initiated in X_0 as follows (for an intuitive description see below): We define $B_0 = \{X_0\}$. Let ν be a finite path initiated in X_0. We put $\nu \in B_{n+1}$ iff

- for all $0 \leq i < |\nu|$ holds $\nu(i) \notin R$
- there is $\ell \geq 0$ such that the prefix $\nu(0) \cdots \nu(\ell)$ of ν is in B_n
- there is $k > \ell$ and a sequence $\omega = t_1, \ldots, t_{k-\ell}$ of transitions of δ such that
  - ω induces $\nu(\ell) \cdots \nu(k)$ (for the meaning of *induce* see the beginning of Appendix)
  - $t_1, \ldots, t_{k-\ell-1}$ induces a (proper) prefix of ν_Y for some $Y \in \mathit{NTerm}$ (i.e. $t_1, \ldots, t_{k-\ell-1}$ induces a path u such that $|u| < |\nu_Y|$ and for all $0 \leq i < |u|$ holds $u(i) = \nu_Y(i)$)
  - ω does *not* induce a prefix of ν_Y for any $Y \in \mathit{NTerm}$
- for $k \leq i < |\nu| - 1$ holds $\mathit{head}(\nu(i)) \in \mathit{Term}$
- $\mathit{head}(\nu(|\nu| - 1)) \in \mathit{NTerm} \setminus RH$

The intuition behind the definition of B_n, where $n \geq 1$, is following. Let us follow a finite path initiated in X_0. This path, say ν, is in B_n if it can be described as follows. We start by following a part of the path ν_{X_0}. At some point, before reaching the end of ν_{X_0}, we leave the path ν_{X_0}. Consequently, we erase all symbols of *Term* from the top of the stack and enter a configuration with the head in $\mathit{NTerm} \setminus RH$. Then we repeat a very similar behavior $n - 1$ times in a row: Assuming that the current configuration has the form $Y\beta$, we follow a part of the path ν_Y within the context β. At some point, before reaching the end of ν_Y, we leave the path ν_Y. Consequently, we erase all symbols of *Term* from the top of the stack and enter a configuration with the head in $\mathit{NTerm} \setminus RH$.

For every $n \geq 0$ we write $Run(B_n) = \bigcup_{v \in B_n} Run(v)$. Note that for all $n \geq 0$ holds $Run(B_n) \supseteq Run(B_{n+1})$ and that $\mathcal{P}(Run(B_n)) \leq (1-\epsilon)^n$.

It is easy to see that for every $i \geq 1$ and $v \in B_i$ there are at most $ci - 1$ numbers $\ell < |v| - 1$ such that $head(v(\ell)) \in NTerm$. Then for $i \geq 2$ and $w \in Run(B_{i-1}) \backslash Run(B_i)$ holds $T(w) < 3c(i-1) + 3c = 3ci$ because every symbol of *NTerm* can generate at most two symbols of *Term*. Also for $w \in Run(B_0) \backslash Run(B_1)$ holds $T(w) < 3c$. It follows that

$$Run(X_0) \backslash Run(B_n) = \bigcup_{i=1}^{n} Run(B_{i-1}) \backslash Run(B_i)$$

and hence for all $w \in Run(X_0) \backslash Run(B_n)$ holds $T(w) < 3cn$. Thus if $T(w) \geq 3cn$, then $w \in Run(B_n)$. Consequently, for every $n \geq 1$ holds

$$\mathcal{P}(T \geq 3cn) \leq \mathcal{P}(Run(B_n)) \leq (1-\epsilon)^n$$

It follows that

$$ET \leq 3c \sum_{n=0}^{\infty} \mathcal{P}(T \geq 3cn) \leq \frac{3c}{\epsilon}$$

□

Proposition C.5.

$$EM_X \leq ET \cdot \max_{Y \in Term} EM_Y + |\Gamma| < \infty$$

Proof. By Proposition C.3 and linearity of expectation,

$$EM_X \leq ET + E(\sum_{n=1}^{T} K_n) + |\Gamma|$$

We prove that

$$E(\sum_{n=1}^{T} K_n) \leq ET \cdot (\max_{Y \in Term} EM_Y - 1) \tag{10}$$

from which we immediately obtain

$$\begin{aligned} EM_X &\leq ET + E(\sum_{n=1}^{T} K_n) + |\Gamma| \\ &\leq ET + ET \cdot \max_{Y \in Term} EM_Y - ET + |\Gamma| \\ &\leq ET \cdot \max_{Y \in Term} EM_Y + |\Gamma| \end{aligned}$$

In order to prove (10) we need some additional notation. For every $Y \in \Gamma_0$ we denote

$$e_Y = \begin{cases} EM_Y & \text{if } Y \in \mathit{Term} \\ 1 & \text{if } Y \in \mathit{NTerm} \end{cases}$$

Given $m \geq 1$, $1 \leq n < m$, $Y \in \Gamma_0$, and $\alpha \in \Gamma^*$, we define $A[m, n, Y, \alpha]$ to be the set of all $w \in CRun(X_0)$ such that $T(w) = m$ and $w(I_n(w)) = Y\alpha$.

In order to prove the equation (10) we proceed in the following steps:

1. we prove that

$$E(K_n \mid A[m, n, Y, \alpha]) = e_Y - 1 \tag{11}$$

2. for $A[m, n, \alpha] = \bigcup_{Y \in \Gamma_0} A[m, n, Y, \alpha]$ we prove that

$$E(K_n \mid A[m, n, \alpha]) = \max_{Y \in \Gamma_0} EM_Y - 1$$

3. for $A[m] = \bigcup_{n=1}^{m} \bigcup_{\alpha \in \Gamma^*} A[m, n, \alpha]$ we prove that

$$E(\sum_{n=1}^{T} K_n \mid A[m]) = m \max_{Y \in \mathit{Term}} EM_Y - 1$$

4. we prove the equation (10).

ad 1.

For $Y \in \mathit{NTerm}$ the equation (11) follows immediately from definition of K_n. Assume that $Y \in \mathit{Term}$. Let us define

$$U = \{w^{I_n(w)-1} \mid w \in A[m, n, Y, \alpha]\}$$

where $w^{I_n(w)-1}$ is the path $w(0), \ldots, w(I_n(w) - 1)$. Let us define

$$W = \{w_{I_{n+1}(w)+1} \mid w \in A[m, n, Y, \alpha]\}$$

where $w_{I_{n+1}(w)+1}$ is the run

$$w(I_{n+1}(w) + 1), w(I_{n+1}(w) + 2), \ldots$$

Let V_k be the set of all paths of the form $\beta_1\alpha, \ldots, \beta_j\alpha$, where $\beta_1 = Y$, for all $1 \leq i < j$ holds $|\beta_i| > 0$, $\beta_j = \varepsilon$, and

$$\max\{|\beta_i| - 1 \mid 1 \leq i \leq j\} = k$$

Let V be the set of all paths of the form $\beta_1\alpha, \ldots, \beta_j\alpha$, where $\beta_1 = Y$, for all $1 \leq i < j$ holds $|\beta_i| > 0$, and $\beta_j = \varepsilon$.

Observe that $A[m, n, Y, \alpha] = U \cdot V \cdot W$ and that $U \cdot V_k \cdot W$ is the set of all runs $w \in A[m, n, Y, \alpha]$ such that $K_n(w) = k$. It is also easy to see that

$$\begin{aligned} \mathcal{P}(U \cdot V_k \cdot W) &= \mathcal{P}(Run(U))\mathcal{P}(Run(V_k))\mathcal{P}(W) \\ &= \mathcal{P}(Run(U))\mathcal{P}(M_Y = k+1)\mathcal{P}(W) \end{aligned}$$

$$\begin{aligned} \mathcal{P}(U \cdot V \cdot W) &= \mathcal{P}(Run(U))\mathcal{P}(Run(V)\mathcal{P}(W) \\ &= \mathcal{P}(Run(U))\mathcal{P}(W) \end{aligned}$$

Hence,

$$\begin{aligned} \mathcal{P}(K_n = k \mid A[m, n, Y, \alpha]) &= \frac{\mathcal{P}(K_n = k \wedge A[m, n, Y, \alpha])}{\mathcal{P}(A[m, n, Y, \alpha])} \\ &= \frac{\mathcal{P}(U \cdot V_k \cdot W)}{\mathcal{P}(U \cdot V \cdot W)} \\ &= \mathcal{P}(M_Y = k+1) \end{aligned}$$

It follows that for $Y \in Term$ holds

$$E(K_n \mid A[m, n, Y, \alpha]) = EM_Y - 1 = e_Y - 1$$

This proves (11) for $Y \in Term$.

ad 2.

For every $Y \in \Gamma_0$ we define

$$p_Y = \mathcal{P}(A[m, n, Y, \alpha] \mid A[m, n, \alpha])$$

Now

$$\begin{aligned} E(K_n \mid A[m, n, \alpha]) &= \sum_{Y \in \Gamma_0} E(K_n \mid A[m, n, Y, \alpha])p_Y \\ &\leq \sum_{Y \in \Gamma_0} (e_Y - 1)p_Y \\ &\leq \max_{Y \in \Gamma_0} e_Y - 1 \\ &= \max_{Y \in Term} EM_Y - 1 \end{aligned}$$

Here the last equality follows from the fact that for every $Y \in Term$ holds $EM_Y \geq 1$.

ad 3.

Define $p_{n,\alpha} = \mathcal{P}(A[m, n, \alpha] \mid A[m])$. Then

$$\begin{aligned}
E(\sum_{n=1}^{T} K_n \mid A[m]) &= \sum_{n=1}^{m} E(K_n \mid A[m]) \\
&= \sum_{n=1}^{m} \sum_{\alpha \in \Gamma^*} E(K_n \mid A[m, n, \alpha]) p_{n,\alpha} \\
&\leq \sum_{n=1}^{m} \sum_{\alpha \in \Gamma^*} p_{n,\alpha} \max_{Y \in \mathit{Term}} EM_Y - 1 \\
&= m \max_{Y \in \mathit{Term}} EM_Y - 1
\end{aligned}$$

ad 4.

We have

$$\begin{aligned}
E(\sum_{n=1}^{T} K_n) &= \sum_{m=1}^{\infty} E(\sum_{n=1}^{T} K_n \mid A[m]) \mathcal{P}(A[m]) \\
&\leq \sum_{m=1}^{\infty} m \mathcal{P}(A[m]) (\max_{Y \in \mathit{Term}} EM_Y - 1) \\
&= \sum_{m=1}^{\infty} m \mathcal{P}(T_Y = m) (\max_{Y \in \mathit{Term}} EM_Y - 1) \\
&= ET \cdot \max_{Y \in \mathit{Term}} EM_Y - 1
\end{aligned}$$

which proves the equation (10). □