Hindawi
Journal of Applied Mathematics
Volume 2020, Article ID 2678374, 23 pages
https://doi.org/10.1155/2020/2678374

Research Article

Finite-State-Space Truncations for Infinite Quasi-Birth-Death Processes

Hendrik Baumann

Clausthal University of Technology, Institute of Mathematics, Erzstr. 1, 38678 Clausthal-Zellerfeld, Germany

Correspondence should be addressed to Hendrik Baumann; hendrik.baumann@tu-clausthal.de

Received 3 March 2020; Revised 14 June 2020; Accepted 23 June 2020; Published 7 August 2020

Academic Editor: Elisa Francomano

For dealing numerically with the infinite-state-space Markov chains, a truncation of the state space is inevitable, that is, an approximation by a finite-state-space Markov chain has to be performed. In this paper, we consider level-dependent quasi-birth-death processes, and we focus on the computation of stationary expectations. In previous literature, efficient methods for computing approximations to these characteristics have been suggested and established. These methods rely on truncating the process at some level N, and for $N \longrightarrow \infty$, convergence of the approximation to the desired characteristic is guaranteed. This paper's main goal is to quantify the speed of convergence. Under the assumption of an f-modulated drift condition, we derive terms for a lower bound and an upper bound on stationary expectations which converge quickly to the same value and which can be efficiently computed.

1. Introduction

In many applications, discrete-time or continuous-time Markov chains are used to model real-life problems. For these models, characteristics of interest can be determined, enabling us to understand relationships between parameters and characteristics, solve optimization problems, and so on. When Markov models are used, the characterization of the dynamics by its probability transition matrix P or generator matrix Q, respectively, is quite simple due to all theoretical problems being solved. However, in applications, the determination of interesting characteristics of the process is very important. In many models, there is no chance of finding explicit analytical representations of these characteristics, and hence, we have to use numerical calculations or simulation methods. In most practical situations, interesting characteristics refer to stationary probabilities or stationary expectations which can be derived from the invariant distribution. In queueing theory, elementary examples are given by moments of the stationary number of customers in a queueing system or queueing network, or by the stationary probability that the number of customers exceeds some threshold. Similarly, such characteristics are interesting in Markovian population models, epidemic models, etc. Formally, stationary expectations are given by $\pi f = \sum_{x \in E} \pi_x f(x)$, where E is the state space of the Markov chain, π is the invariant distribution, and f is some cost or reward function.

The invariant distribution for an irreducible and positive recurrent Markov chain is characterized by the unique probability vector which solves a certain homogeneous system of linear equations. The number of states coincides with the number of equations and the number of variables. Unfortunately, in realistic models, we have a very large number of states or even infinitely many states. In these situations, numerically computing stationary probabilities or stationary expectations requires a truncation of the system of equations (which means a truncation of the state space), and this truncation results in inevitable errors.

Hence, numerical calculations only compute some approximation A to the desired characteristic πf, and from both a mathematical and a practical point of views, bounds on the truncation error $\pi f - A$ are interesting. In this paper, we will derive lower and upper bounds on πf which is equivalent to computing an approximation to πf and bounds on $\pi f - A$. Since all values are computed numerically, our results and methods provide a posteriori error estimates.

Competing methods for finding a posteriori error estimates were given in [1–3], and an approach for finding a priori estimates can be found in [4, 5]. In Section 8, we will briefly compare these methods to the method suggested in this paper.

Although our theoretical results will be quite general, we will only discuss the efficient implementation of the computation of the bounds for level-dependent quasi-birth-death processes (LDQBDs). These processes are characterized by the block-tridiagonal structure of the transition probability matrix (discrete-time) or generator matrix (continuous time). This structure is very typical for a large class of queueing models, population and epidemic models, etc.

The rest of the paper is organized as follows: After introducing some notation in Section 2, we will present our theoretical results on lower and upper bounds on stationary expectations πf in Section 3. In Section 4, we will derive a method for computing these bounds efficiently for LDQBDs. Afterwards, we will apply our method to the elementary $M/M/1$ queue (Section 5), a variant of the $M/\mathrm{PH}/1$ queue (Section 6), and a popular retrial queueing model (Section 7). Whereas we have an explicit analytical representation of the invariant distribution and of stationary characteristics in the first example, numerical computations are the only chance to find stationary expectations in the latter two examples. Finally, we give a comparison of the suggested method to competing ones (Section 8).

2. Preliminaries

2.1. Notations and an Auxiliary Result. In order to introduce our goals formally, we review some well-known facts on the limit behaviour of Markov chains. In this context, and in the whole paper, we will use the following notations:

(i) I is a finite or infinite identity matrix. The dimension will become clear from the context

(ii) $\mathbf{1}$ denotes a finite or infinite (column) vector with all entries being 1. Alternatively, $\mathbf{1}$ refers to a function with constant value 1. The meaning will be clear from the context

(iii) $\mathbf{1}_A$ denotes the indicator function of set A, that is, $\mathbf{1}_A(x) = 1$ if $x \in A$, and $\mathbf{1}_A(x) = 0$ otherwise. With slight abuse of notation, we use $\mathbf{1}_B$ for Boolean expressions B. If B is true, $\mathbf{1}_B$ is 1, and otherwise, $\mathbf{1}_B$ is 0

(iv) id is the identity function

(v) We use the notation $\mathbb{N} = \{1, 2, 3, \cdots\}$ and $\mathbb{N}_0 = \{0, 1, 2, 3, \cdots\}$.

Furthermore, we will use a probabilistic proof of invertibility frequently throughout the paper: If P is finite and transient, $I - P$ is invertible. We give some details: A finite substochastic matrix $P = (p_{ij})_{i,j=1}^{m}$ is referred to as *transient* if $\lim_{n\to\infty} P^n \longrightarrow 0$ (component-wise). This is true if and only if P contains no stochastic submatrix $(p_{ij})_{i,j\in F}$ for some $F \subset \{1, \cdots, m\}$. Hence, P is transient if and only if, for all $i \in \{1, \cdots, m\}$, there are $i_0, i_1, \cdots, i_k \in \{1, \cdots, m\}$ with $i_0 = i, p_{i_{r-1}, i_r} > 0$ for $r = 1, \cdots, k$, and $\sum_{j=1}^{m} p_{i_k, j} < 1$.

Due to P being finite, $P^n \longrightarrow 0$ entails $|\lambda| < 1$ for all eigenvalues of P, and hence, $\sum_{n=0}^{\infty} P^n$ converges, and by a standard argument, we obtain

$$(I - P)\sum_{n=0}^{\infty} P^n = \lim_{N\to\infty}(I - P)\sum_{n=0}^{N} P^n \lim_{N\to\infty}\left(I - P^{N+1}\right) = I, \quad (1)$$

which means that $I - P$ is invertible.

2.2. Results on the Limit Behaviour of the Markov Chains. We briefly summarize some results on the limit behaviour of Markov chains since these results are the reason why it is important to be able to compute (approximations to) stationary expectations πf.

We consider Markov chains $(X_n)_{n\in\mathbb{N}_0}$ in discrete time or Markov chains $(Y_t)_{t\geq 0}$ in continuous time with countable state space E. Throughout the paper, we assume that $(X_n)_{n\in\mathbb{N}_0}$ is irreducible and recurrent with transition probability matrix $P = (p_{xy})_{x,y\in E}$ or that $(Y_t)_{t\geq 0}$ is nonexplosive, irreducible and recurrent with generator matrix $Q = (q_{xy})_{x,y\in E}$. Then, there is an invariant measure ψ, that is, a positive vector ψ with $\psi P = \psi$ or $\psi Q = 0$, respectively. In both cases, ψ is unique up to constant multiples. For $f : E \longrightarrow \mathbb{R}$, we introduce the notation $\psi f := \sum_{x\in E} \psi_x f(x)$. If ψf and ψg converge absolutely with $\psi g \neq 0$ for functions $f, g : E \longrightarrow \mathbb{R}$, a fundamental result on the limit behaviour of Markov chains is

$$\lim_{n\to\infty} \frac{\sum_{k=0}^{n-1} f(X_k)}{\sum_{k=0}^{n-1} g(X_k)} = \frac{\psi f}{\psi g} \quad \text{or} \quad \lim_{n\to\infty} \frac{\int_0^t f(Y_s)ds}{\int_0^t g(Y_s)ds} = \frac{\psi f}{\psi g}. \quad (2)$$

in the sense of almost sure convergence (see [6], pp. 85-86, 203-209). If we have the stronger assumption of positive recurrence, the sum of the entries of ψ is finite. Hence, by multiplying with an appropriate constant, we obtain the uniquely determined invariant distribution π which still is an invariant measure, but additionally satisfies $\sum_{x\in E} \pi_x = 1$. Then, we can set $\psi = \pi$ and $g = \mathbf{1}$ constant, and obtain

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} f(X_k) = \pi f \quad \text{or} \quad \lim_{t\to\infty} \frac{1}{t}\int_0^t f(Y_s)ds = \pi f, \quad (3)$$

almost surely if πf converges absolutely. If f measures costs or rewards that are associated with the time spent in state x, πf is the long-run average of costs or rewards per time unit. In particular, for $A \subset E$, we can specify

$$f(x) = \mathbf{1}_A(x) = \begin{cases} 1, & x \in A, \\ 0, & x \notin A, \end{cases} \quad (4)$$

and $\pi \mathbf{1}_A$ is the long-run average of the proportion of time spent in set A. For $A = \{x\}$, $\pi \mathbf{1}_{\{x\}} = \pi_x$ is the long-run average

of the proportion of time spent in state x. Note that there is another interpretation of the entries of the invariant distribution π: For irreducible and positive recurrent discrete-time Markov chains which are additionally aperiodic, X_n converges to some random variable X_∞ in distribution, and for nonexplosive, irreducible, and positive recurrent continuous-time Markov chains, Y_t converges to some random variable Y_∞ in distribution. Then, π is the distribution of X_∞ or Y_∞ (see any textbook on Markov chains, e.g., [6, 7]), that is, $\mathbb{P}(X_\infty = x) = \pi_x$ and $\mathbb{P}(Y_\infty = x) = \pi_x$. Hence, if πf converges absolutely for some function $f : E \longrightarrow \mathbb{R}$, we have $\pi f = \mathbb{E}[f(X_\infty)]$ or $\pi f = \mathbb{E}[f(Y_\infty)]$, respectively.

Due to these results on the limit behaviour, in many applications, all characteristics of interest can be deduced from $\psi f^{(\ell)}$ for functions $f^{(0)}, \cdots, f^{(L)}$, where we will set $f^{(0)} = \mathbf{1}$ in case of positive recurrence. Furthermore, by splitting functions into positive and negative part, we can assume $f^{(\ell)} \geq 0$ for all $\ell = 0, \cdots, L$ without restriction. We remark that ψf is vector valued, $\psi f = \psi f^{(0)}, \cdots, \psi f^{(L)}$.

2.3. Notations for Block-Structured Markov Chains. Throughout this paper, let the state space $E = E_0 \cup E_1 \cup E_2 \cup \cdots$ be partitioned into finite and pairwise disjoint sets E_i which will be referred to as *levels*. We introduce some notation:

(i) We fix $K \in \mathbb{N}_0$. K will be chosen such that the f-modulated drift condition (see below) holds on $E_{N+1} \cup E_{N+2} \cup \cdots$. Furthermore, throughout this paper, $x_0 \in E_K$ is some arbitrary, but fixed state within level E_K

(ii) $\psi = (\psi_x)_{x \in E}$ denotes the invariant measure with $\psi_{x_0} = 1$. In case of positive recurrence, $\pi = (\pi_x)_{x \in E}$ is the invariant distribution

(iii) For $i, j \in \mathbb{N}_0$, P_{ij} or Q_{ij}, respectively, shall denote the transition probabilities or transition rates, respectively, for transitions from states $\in E_i$ to states $\in E_j$.

(iv) For $i \in \mathbb{N}_0$, we write $\psi_i = (\psi_x)_{x \in E_i}, \pi_i = (\pi_x)_{x \in E_i}, f_i = (f^{(\ell)}(x))_{\substack{x \in E_i \\ 0 \leq \ell \leq L}}$ and $g_i = (g^{(\ell)}(x))_{\substack{x \in E_i \\ 0 \leq \ell \leq L}}$ for functions $f, g : E \longrightarrow \mathbb{R}^{L+1}$.

Later on, multidimensional functions f will be useful for computing approximations to several stationary expectations at the same time. With this notation, we have $\psi f = \sum_{i=0}^{\infty} \psi_i f_i$, and we are interested in computing approximations to ψ_i and use them for approximating ψf, or we want to find an approximation to ψf directly by other means. As in other approaches for finding error bounds on the approximation of stationary characteristics, we assume that an *f-modulated drift condition* is satisfied, that is,

$$\sum_{j=K+1}^{\infty} P_{ij} g_j - g_i \leq -f_i, \; i > K, \tag{5}$$

$$\sum_{j=K+1}^{\infty} Q_{ij} g_j \leq -f_i, \quad i > K. \tag{6}$$

In scalar notation, (6) reads as $\sum_{y \in E_{K+1} \cup E_{K+2} \cup \cdots} q_{xy} g(y) \leq -f(x)$ for $x \in E_{K+1} \cup E_{K+2} \cup \cdots$. f-modulated drift conditions are a popular tool to prove positive recurrence (for $f(x) = 1$) and convergence of ψf or πf. In our examples in Sections 5–7, we will use this tool, too. For details, we refer to Theorem 1 and Theorem 2 in [8]. Note that the idea of making use of f-modulated drift conditions for this purpose is much older (see, e.g., [9–13], but some of these classical references formally require $f(x) = 1$, whereas the results in [8] only require $f(x) \geq 0$.

Finally, throughout the paper, we will require that

$$P_{ij} = 0 \quad \text{for } i > K > j \quad \text{or} \quad Q_{ij} = 0 \quad \text{for} \quad i > K > j. \tag{7}$$

We will comment this structural requirement in Section 8.

3. Upper and Lower Bounds on ψf: Theoretical Results

Our goal is to find a lower bound and an upper bound on ψf such that

(i) Both bounds converge to ψf as the truncation level N tends to ∞

(ii) Both bounds can be computed numerically in an efficient manner.

In this section, we focus on the "mathematical" criteria, that is, we derive exact terms for the lower and upper bounds in which both converge to ψf. The latter goal of computational efficiency will be considered in Section 4 for the special case of QBDs.

First, we focus on the discrete-time setting. We use the notation for block-structured Markov chains as introduced above, and we begin with a representation of invariant measures and their approximations. For this purpose, we introduce the hitting times $\sigma(N) = \inf\{n \in \mathbb{N} : X_n \in E_K \cup E_{N+1} \cup E_{N+2} \cup \cdots\}$, $\tau = \inf\{n \in \mathbb{N} : X_n \in E_K\}$, and $\tau_0 = \inf\{n \in \mathbb{N} : X_n = x_0\}$.

Lemma 1. *Let $(X_n)_{n \in \mathbb{N}_0}$ be an irreducible recurrent discrete-time Markov chain with block-structured transition probability matrix P as in the preliminaries, let $K, N \in \mathbb{N}_0$ with $K \leq N$ and $x_0 \in E_K$. Let*

$$\begin{aligned} S_K(N) &= I, \\ S_j(N) &= \sum_{i=0}^{N} S_i(N) P_{ij}, \quad j \in \mathbb{N}_0 \setminus \{K\}, \end{aligned} \tag{8}$$

and $T(N) = \sum_{i=0}^{N} S_i(N) P_{i,K}$.

(a) $S_i(N)$ is well defined, and we have

$$S_i(N) = \left(\mathbb{E}\left[\sum_{n=1}^{\sigma(N)} 1_{X_n=y} | X_0 = x\right]\right)_{\substack{x\in E_K \\ y\in E_i}}, \quad i \neq K,$$

$$T(N) = \left(\mathbb{E}\left[\sum_{n=1}^{\sigma(N)} 1_{X_n=y} | X_0 = x\right]\right)_{x,y\in E_K}. \tag{9}$$

$T(N)$ and all $S_i(N)$ increase monotonically in N.

(b) $S_i(\infty) = \lim_{n\to\infty} S_i(N)$ exists for all $i \in \mathbb{N}_0$ with $S_K(\infty) = I$ and

$$S_i(\infty) = \left(\mathbb{E}\left[\sum_{n=1}^{\tau} 1_{X_n=y} | X_0 = x\right]\right)_{\substack{x\in E_K \\ y\in E_i}}, \quad i \neq K. \tag{10}$$

$T(\infty) = \lim_{n\to\infty} T(N)$ exists with

$$T(\infty) = \left(\mathbb{E}\left[\sum_{n=1}^{\tau} 1_{X_n=y} | X_0 = x\right]\right)_{x,y\in E_K}. \tag{11}$$

For the limits, we have $S_i(\infty) = \sum_{i=0}^{\infty} S_i(\infty) P_{ij}$ for all $j \neq K$ and $T(\infty) = \sum_{i=0}^{\infty} S_i(\infty) P_{iK}$.

(c) $T(\infty)$ is stochastic and irreducible. Let $\psi_K = (\psi_x)_{x\in E_K}$ be the invariant measure for $T(\infty)$ subject to $\psi_{x_0} = 1$, and let $\psi_i = \psi_K S_i(\infty)$ for $i \neq K$. Then $\psi = (\psi_0, \psi_1, \cdots)$ is the unique invariant measure for P subject to $\psi_{x_0} = 1$.

(d) We have $\psi f = \psi_K F$ where $F = \sum_{i=0}^{\infty} S_i f_i$.

Proof. We refer to the representations of $S_i(N)$, $T(N)$, S_i, and T as "probabilistic interpretations."

(a) The definition of $S_i(N)$ implies

$$(S_i(N))_{\substack{i=0 \\ i\neq K}}^{N} \left(I = \left(P_{ij}\right)_{\substack{i,j=0 \\ i\neq K}}^{N}\right) = \left(P_{K_j}\right)_{\substack{j=0 \\ j\neq K}}^{N}. \tag{12}$$

Since $(P_{ij})_{\substack{i,j=0 \\ i\neq K}}^{N}$ is finite and transient, $(I = (P_{ij})_{\substack{i,j=0 \\ i\neq K}}^{N})$ is invertible, and hence, $S_i(N)$ is well defined for $i = 0, \cdots, K-1, K+1, \cdots, N$. Afterwards, the definition of $S_i(N)$ for $i > N$ is explicit. We omit the proof of the probabilistic interpretation since such considerations are standard in the context of finding probabilistic interpretations for invariant measures (see, e.g., [7, 14]), and since these considerations are similar to those in the proof of Lemma 3.2b) or Lemma 4 below. Note that the probabilistic interpretation directly implies the monotonicity since $\sigma(N)$ increases monotonically.

(b) More precisely, $\sigma(N)$ converges to τ almost surely. By monotone convergence, the probabilistic interpretation of $S_i(\infty)$ and $T(\infty)$ follows. Note that these expectations are finite due to recurrence. $S_i(\infty) = \sum_{i=0}^{\infty} S_i(\infty) P_{ij}$ for $j \neq K$ and $T(\infty) = \sum_{i=0}^{\infty} S_i(\infty) P_{iK}$ can be proved by means of the probabilistic interpretations. Again, we omit all further details.

(c) The probabilistic interpretation of $T(\infty)$ can be rewritten as

$$T(\infty) = (\mathbb{P}(\tau < \infty, X_\tau = y | X_0 = x))_{x,y\in E_K}. \tag{13}$$

Due to recurrence, we have $\tau < \infty$ almost surely, and stochasticity and irreducibility follow easily. Since $T(\infty)$ is finite, it admits an invariant measure ψ_K which becomes unique by requiring $\psi_{x_0} = 1$. Multiplying $S_i(\infty) = \sum_{i=0}^{\infty} S_i(\infty) P_{ij}$ by ψ_K immediately leads to $\psi_j = \sum_{i=0}^{\infty} \psi_i P_{ij}$ for $i \neq K$, and analogously, we find $\psi_K = \psi_K T(\infty) = \sum_{i=0}^{\infty} \psi_i P_{iK}$. Hence, $\psi = (\psi_0, \psi_1, \cdots)$ is invariant for P. Note that we can find the probabilistic interpretation

$$\psi_y = \mathbb{E}\left[\sum_{n=1}^{\tau_0} 1_{X_n=y} | X_0 = x_0\right], \tag{14}$$

by standard considerations (again, we omit details). This is the standard representation of invariant measures with $\psi_{x_0} = 1$ (see, e.g., [7, 14]).

(d) The statement is a direct consequence of $\psi f = \sum_{i=0}^{\infty} \psi_i f_i$ and the representation of ψ_i found in (c).

Both ψ_K and F depend on P_{ij} for all $i, j \in \mathbb{N}_0$. Our goal is to find bounds both on ψ_K and F which can be computed from $S_0(N), \cdots, S_N(N)$, since these matrices only depend on the finite matrix $P(N) = (P_{ij})_{i,j=0}^{N}$. We begin with bounds on ψ_K.

Lemma 2. *Let all requirements of Lemma 1 hold and choose $A = A(N) = (a_{xy})_{x,y\in E_K}$ such that $A(I - T(N))$ is an invertible diagonal matrix.*

(a) Such a matrix A exists, and for two such matrices $A, \tilde{A}$, we have $\tilde{A} = DA$ with an invertible diagonal matrix D.

(b) For any such choice of A, we have

$$\underline{\varphi}_x := \frac{a_{x_0,x}}{a_{x_0,x_0}} \leq \psi_x \leq \frac{a_{xx}}{a_{x_0,x_0}} =: \bar{\varphi}_x, \quad x \in E_K. \tag{15}$$

(c) *Let* $\underline{\varphi}_x(N) = \underline{\varphi}_x$ *and* $\bar{\varphi}_x(N) = \bar{\varphi}_x$ *be defined as in (b). Then* $\underline{\varphi}_x(N)$ *increases monotonically in* N, $\bar{\varphi}_x(N)$ *decreases monotonically in* N, *and both converge to* ψ_x *for all* $x \in E_K$.

Proof. We adapt ideas for proving classical bounds on quotients of entries of the invariant measure (see, e.g., [15]).

(a) The probabilistic interpretation of $T(N)$ implies that $T(N)$ is transient. Hence, $I - T(N)$ is invertible and we could choose $A = (I - T(N))^{-1}$. If both $A(I - T(N))$ and $\tilde{A}(I - T(N))$ are invertible diagonal matrices, so is $\tilde{A}(I - T(N))(A(I - T(N)))^{-1} = \tilde{A} \cdot A^{-1}$.

(b) First, we note that $\underline{\varphi}_x$ and $\bar{\varphi}_x$ do not depend on the choice of A since a left-hand multiplication by some diagonal matrix has no impact on the quotients under consideration. Now, we will directly construct a matrix A with $a_{xx} = 1$ for all $x \in E_K$. With $T(N) = (t_{xy})_{x,y \in E_K}$ and $T(\infty) = (t_{xy}(\infty))_{x,y \in E_K}$, this works as follows: We set

$$\begin{aligned} \lambda_{xy}^{(1)} &= t_{xy}, \qquad \lambda_{xy}^{(n)} = \sum_{z \neq x} \lambda_{xz}^{(n-1)} t_{zy}, \quad n \geq 2, \\ \mu_{xy}^{(1)} &= t_{xy}(\infty), \quad \mu_{xy}^{(n)} = \sum_{z \neq x} \lambda_{xz}^{(n-1)} t_{zy}(\infty), \quad n \geq 2, \\ a_{xy} &= \sum_{n=1}^{\infty} \lambda_{xy}^{(n)}, \qquad g_{xy} = \sum_{n=1}^{\infty} \mu_{xy}^{(n)}, \quad y \neq x. \end{aligned} \tag{16}$$

Additionally, we set $a_{xx} = g_{xx} = 1$. Then for all, $x \in E_K$, $(a_{xy})y \in E_K$ is the standard construction (see, e.g., [7, 14]) of the *minimal subinvariant measure* for the substochastic matrix T subject to $a_{xx} = 1$. Indeed,

$$\begin{aligned} \sum_{z \in E_K} a_{xz} t_{zy} &= t_{xy} + \sum_{z \neq x} \sum_{n=1}^{\infty} \lambda_{xz}^{(n)} t_{zy} = \lambda_{xy}^{(1)} + \sum_{n=2}^{\infty} \sum_{z \neq x} \lambda_{xz}^{(n-1)} t_{zy} \\ &= \sum_{n=1}^{\infty} \lambda_{xy}^{(n)} = a_{xy}, \end{aligned} \tag{17}$$

for $y \neq x$. Hence, $A(I - T(N))$ is a diagonal matrix. Analogously, the rows of $G = (g_{xy})_{x,y \in E_K}$ are subinvariant measures for $T(\infty)$. This latter matrix is finite and stochastic, and hence, every subinvariant measure is invariant. In particular, every row of G is a constant multiple of ψ_K, that is,

$$\frac{\psi_y}{\psi_x} = xy, \tag{18}$$

for all $x, y \in E_K$. Due to $t_{xy} \leq t_{xy}(\infty)$, we have $\lambda_{xy}^{(n)} \leq \mu_{xy}^{(n)}$ by a trivial induction, and thus, $a_{xy} \leq g_{xy}$. From $a_{xx} = 1$, we obtain

$$\frac{a_{xy}}{a_{xx}} \leq \frac{\psi_y}{\psi_x}, \tag{19}$$

for all $x, y \in E_K$. By setting $x = x_0$, we obtain $\underline{\varphi}_y \leq \psi_y$ (due to $\psi_{x_0} = 1$), and by setting $y = x_0$, we obtain $1/\bar{\varphi}_x \leq 1/\psi_x$.

(c) We have $\lim_{n \to \infty} T(N) = T(\infty)$, implying $\lim_{n \to \infty} \lambda_{xy}^{(n)} = \mu_{xy}^{(n)}$, and by monotone convergence, we obtain

$$\lim_{n \to \infty} \frac{a_{xy}}{a_{xx}} g_{xy} = \frac{\psi_y}{\psi_x}. \tag{20}$$

By setting $y = x_0$ or $x = x_0$, respectively, the statement follows.

Lemma 2 gives a lower and an upper bound on all entries of ψ_K. It remains to find bounds on F. A lower bound is a direct consequence of Lemma 1.

Lemma 3. *Let all requirements of Lemma 1 hold, let* $f \geq 0$, *and let*

$$\underline{F}(N) = \sum_{i=0}^{N} S_i(N) f_i. \tag{21}$$

(a) *We have* $\underline{F}(N) \leq F$.

(b) $\underline{F}(N)$ *increases monotonically in* N *with limit* F.

As pointed out above, for finding an upper bound, we have to use some information on p_{xy} with $x \in E_{N+1} \cup E_{N+2} \cup \cdots$ or $y \in E_{N+1} \cup E_{N+2} \cup \cdots$. This is done by using the f-modulated drift condition (5). Note that the function f is given, and the function g solving the drift condition can often be found by "educationally guessing" even if there is no chance to find explicit analytical terms for ψ or ψf.

Lemma 4. *Let all requirements of Lemma 1 hold, let the* f*-modulated drift condition (5) hold, and let the structural requirement (7) be fulfilled. Furthermore, define* $\bar{f}_i = f_i + \sum_{j=N+1}^{\infty} p_{ij} g_j$ *for* $i \leq N$ *and* $\bar{F}(N) = \sum_{i=0}^{N} S_i \bar{f}_i$.

(a) *For* $N \geq K$, *we have* $\bar{F}(N) \geq F$.

(b) *Additionally, let* $\psi g < \infty$. *Then* $\bar{F}(N)$ *converges to* F.

Proof. (a) First, remember

$$F = \sum_{i=0}^{\infty} S_i(\infty) f_i = f_K + \sum_{\substack{i=0 \\ i \neq K}}^{\infty} S_i(\infty) f_i = \left(f^{(\ell)}(x) + \sum_{y \in E \setminus E_K} \Psi_{xy} f^{(\ell)}(y) \right)_{\substack{x \in E_K \\ 0 < \ell \leq L}}, \tag{22}$$

where for $y \notin <E_K$

$$\Psi_{xy} = \mathbb{E}\left[\sum_{n=1}^{\tau} 1_{X_n = y} | X_0 = x \right] = \mathbb{E}\left[\sum_{n=0}^{\tau - 1} 1_{X_n = y} | X_0 = x \right], \tag{23}$$

implying

$$\Psi_{xy} f^{(\ell)}(y) = \mathbb{E}\left[\sum_{n=0}^{\tau - 1} 1_{X_n = y} f^{(\ell)}(y) | X_0 = x \right]. \tag{24}$$

For $y \in E_K$, we have

$$\mathbb{E}\left[\sum_{n=0}^{\tau - 1} 1_{X_n = y} f^{(\ell)}(y) | X_0 = x \right] = \delta_{xy} f^{(\ell)}(x), \tag{25}$$

and together, we obtain

$$F = \left(\sum_{y \in E} \mathbb{E}\left[\sum_{n=0}^{\tau - 1} 1_{X_n = y} f^{(\ell)}(y) | X_0 = x \right] \right)_{\substack{x \in E_K \\ 0 \leq \ell < L}} = \left(\mathbb{E}\left[\sum_{n=0}^{\tau - 1} f^{(\ell)}(X_n) | X_0 = x \right] \right)_{\substack{x \in E_K \\ 0 \leq \ell < L}}. \tag{26}$$

Analogously, we find

$$\bar{F}(N) = \left(\mathbb{E}\left[\sum_{n=0}^{\sigma(N)} \bar{f}^{(\ell)}(X_n) | X_0 = x \right] \right)_{\substack{x \in E_K \\ 0 \leq \ell < L}}. \tag{27}$$

Hence, it remains to show $\mathbb{E}[\sum_{n=0}^{\tau-1} f^{(\ell)}(X_n) | X_0 = x] \leq \mathbb{E}[\sum_{n=0}^{\sigma(N)-1} \bar{f}^{(\ell)}(X_n) | X_0 = x]$ for all $x \in E_K$. For this purpose, we first write

$$\begin{aligned} &\mathbb{E}\left[\sum_{n=0}^{\tau-1} f^{(\ell)}(X_n) | X_0 = x \right] \\ &= \mathbb{E}\left[\sum_{n=0}^{\sigma(N)-1} f^{(\ell)}(X_n) | X_0 = x \right] + \mathbb{E}\left[1_{\sigma(N) < \tau} \sum_{n=\sigma(N)}^{\tau-1} f^{(\ell)}(X_n) | X_0 = x \right] \\ &= \mathbb{E}\left[\sum_{n=0}^{\sigma(N)-1} f^{(\ell)}(X_n) | X_0 = x \right] + \mathbb{E}\left[\sum_{k=0}^{\sigma(N)-1} 1_{X_{k+1} \in E_{N+1} \cup E_{N+1} \cup \cdots} \sum_{n=k+1}^{\tau-1} f^{(\ell)}(X_n) \mid X_0 = x \right]. \end{aligned} \tag{28}$$

Let $g_{\infty}^{(\ell)}(x) = \mathbb{E}[\sum_{n=0}^{\tau-1} f^{(\ell)}(X_n) \mid X_0 = x]$. By means of the tower property of conditional expectations, we find

$$\begin{aligned} &\mathbb{E}\left[\sum_{k=0}^{\sigma(N)-1} 1_{X_{k+1} \in E_{N+1} \cup E_{N+2} \cup \cdots} \sum_{n=k+1}^{\tau-1} f^{(\ell)}(X_n) \mid X_0 = x \right] \\ &= \mathbb{E}\left[\sum_{k=0}^{\sigma(N)-1} 1_{X_{k+1} \in E_{N+1} \cup E_{N+2} \cup \cdots} \mathbb{E}\left[\sum_{n=k+1}^{\tau-1} f^{(\ell)}(X_n) \mid X_{k+1} \right] \mid X_0 = x \right] \\ &= \mathbb{E}\left[\sum_{k=0}^{\sigma(N)-1} 1_{X_{k+1} \in E_{N+1} \cup E_{N+2} \cup \cdots} g_{\infty}^{(\ell)}(X_{k+1}) \mid X_0 = x \right] \\ &= \mathbb{E}\left[\sum_{k=0}^{\sigma(N)-1} \mathbb{E}\left[1_{X_{k+1} \in E_{N+1} \cup E_{N+2} \cup \cdots} g_{\infty}^{(\ell)}(X_{k+1}) \mid X_k \right] \mid X_0 = x \right] \\ &= \mathbb{E}\left[\sum_{k=0}^{\sigma(N)-1} \sum_{y \in E_{N+1} \cup E_{N+2} \cup \cdots} p_{X_k, y} g_{\infty}^{(\ell)}(y) \mid X_0 = x \right]. \end{aligned} \tag{29}$$

Together, we obtain

$$\mathbb{E}\left[\sum_{n=0}^{\tau-1} f^{(\ell)}(X_n) \mid X_0 = x \right] = \mathbb{E}\left[\sum_{n=0}^{\sigma(N)-1} \left(f^{(\ell)}(X_n) + \sum_{y \in E_{N+1} \cup E_{N+2} \cup \cdots} p_{X_k, y} g_{\infty}^{(\ell)}(y) \right) | X_0 = x \right], \tag{30}$$

for the entries of F. Below, we will show $g_{\infty}^{(\ell)}(x) \leq g^{(\ell)}(x)$ for $x \in E_{K+1} \cup E_{K+2} \cup \cdots$, and due to $\bar{f}(x) = f(x) + \sum_{y \in E_{N+1} \cup E_{N+2} \cup \cdots} p_{xy} g(y)$ and $N \geq K$, the desired statement follows.

For proving $g_{\infty}^{(\ell)}(x) \leq g^{(\ell)}(x)$, we define $g_M^{(\ell)}(x) = \mathbb{E}[\sum_{n=0}^{\max\{\tau-1, M\}} f^{(\ell)}(X_n) | X_0 = x]$ for $x \in E_{K+1} \cup E_{K+2} \cup \cdots$. Then $g_0^{(\ell)}(x) = f(x) \leq g^{(\ell)}(x)$ for $x \in E_{K+1} \cup E_{K+2} \cup \cdots$, and by

induction, we obtain

$$\begin{aligned} g_M^{(\ell)}(x) &= f(x) + \mathbb{E}\left[1_{X_1 \in E_{K+1} \cup E_{K+2} \cup \cdots} \sum_{n=1}^{\max\{\tau-1,M\}} f^{(\ell)}(X_n) | X_0 = x\right] \\ &\quad + \mathbb{E}\left[1_{X_1 \in E_0 \cup \cdots \cup E_{K+1}} \sum_{n=1}^{\max\{\tau-1,M\}} f^{(\ell)}(X_n) | X_0 = x\right] \\ &= f^{(\ell)}(x) + \mathbb{E}\left[1_{X_1 \in E_{K+1} \cup E_{K+2} \cup \cdots} \left[\sum_{n=1}^{\max\{\tau-1,M\}} f^{(\ell)}(X_n) | X_1\right] | X_0 = x\right] + 0 \\ &= f^{(\ell)}(x) + \mathbb{E}\left[1_{X_1 \in E_{K+1} \cup E_{K+2} \cup \cdots} g_{M-1}^{(\ell)}(x_1) | X_0 = x\right] \\ &\le f^{(\ell)}(x) = f^{(\ell)}(x) + \sum_{y \in E_{K+1} \cup E_{K+2} \cup \cdots} p_{xy} g_{M-1}^{(\ell)}(y) \le g^{(\ell)}(x), \end{aligned} \tag{31}$$

for all $x \in E_{K+1} \cup E_{K+2} \cup \cdots$ and all $M \in \mathbb{N}0$, since transitions from $E_{K+1} \cup E_{K+2} \cup \cdots$ to $E_0 \cup \cdots \cup E_{K-1}$ have probability 0. Due to recurrence, τ is finite almost surely, implying $\max\{\tau - 1, M\} \longrightarrow \tau$ almost surely, and by monotone convergence, $g_M^{(\ell)}(x) \longrightarrow g_\infty^{(\ell)}(x)$, implying $g_\infty^{(\ell)}(x) \le g^{(\ell)}(x)$ for all $x \in E_{K+1} \cup E_{K+2} \cup \cdots$ as desired.

(b) Since $\underline{F}(N)$ converges to F, it suffices to prove that $\bar{F}(N) - \underline{F}(N)$ converges to 0. For this purpose, we write

$$\begin{aligned} \bar{F}(N) - \underline{F}(N) &= \sum_{i=0}^{N} S_i(N)\left(\overline{f_i} - f_i\right) = \sum_{i=0}^{N} S_i(N) \sum_{j=N+1}^{\infty} P_{ij} g_j \\ &= \sum_{j=N+1}^{\infty} \sum_{i=0}^{N} S_i(N) P_{ij} g_j = \sum_{j=N+1}^{\infty} S_i(N) g_j \\ &\le \sum_{j=N+1}^{\infty} S_j(\infty) g_j. \end{aligned} \tag{32}$$

Again, let $S_j(\infty) = (\Psi_{xy})_{y \in E_j}^{x \in E_K}$. For $x = x_0$, we have $\Psi_{xz} \le \psi_z$ (due to the respective probabilistic interpretation, see proof of Lemma 1 for the interpretation of ψ_z), and since we assume that ψ_g converges, we have convergence to 0 for $N \longrightarrow \infty$. Since the choice of $x_0 \in E_K$ is arbitrary, this is true for any $x \in E_K$.

By summarizing all previous results, we obtain our main result for the discrete-time setting where we will omit the dependency of $S_i \underline{F}, \bar{F}, \cdots$ on N. Furthermore, we will phrase the result directly for both discrete-time and continuous-time setting. All previous results may be transferred into the context of continuous-time Markov chains easily by means of the embedded jump chain. Since similar considerations have been used in the literature frequently (see, e.g., [4, 16, 17]), we omit further details.

Theorem 5. *Let $X = (X_n)_{n \in N_0}$ be an irreducible and recurrent discrete-time Markov chain, or let $Y = (Y_t)_{t \ge 0}$ be a nonexplosive, irreducible, and recurrent continuous-time Markov chain, respectively. Let the transition probability matrix $P = (P_{ij})_{i,j=0}^{\infty}$ or $Q = (Q_{ij})_{i,j=0}^{\infty}$, respectively, be block-structured as introduced in the preliminaries. Let the drift condition (5) or (6), respectively, and the structural requirement (7) hold. Furthermore, let*

(i) $S_i \in \mathbb{R}^{d_K \times d_i}$ be defined by $S_K = I$ and $S_i = \sum_{j=0}^{N} S_j P_{ji}$ or $0 = \sum_{j=0}^{N} S_j Q_{ji}$ resp., $i = 0, \cdots, K-1, K+1, \cdots N$

$$T = \sum_{j=0}^{N} S_j P_{jK} \text{ or } T = -\sum_{j=0}^{N} S_j Q_{jK}, \text{ resp.} \tag{33}$$

(ii) $A = (a_{xy})_{x,y \in E_K}$ be chosen such that $A(I - T)$ or AT, respectively, is a diagonal matrix, let $\underline{\varphi}_x = (a_{x_0,x})/a_{x_0,x_0}$ and $\bar{\varphi}_x = a_{xx}/a_{x,x_0}$ for $x \in E_K$, and let $\underline{\varphi} = (\underline{\varphi}_x)_{x \in E_K}$ and $\bar{\varphi} = (\bar{\varphi}_x)_{x \in E_K}$ as row vectors

(iii) $\sum_{i=K+1}^{\infty} P_{ij} g_j + f_i \le g_i$ or $\sum_{i=K+1}^{\infty} Q_{ij} g_j + f_i \le 0$, resp., for $i \ge K+1$

(iv) $\bar{f}_i := f_i + \sum_{j=N+1}^{\infty} P_{ij} g_j$ or $\bar{f}_i := f_i + \sum_{j=N+1}^{\infty} Q_{ij} g_j$, resp., be component-wise finite for all $i = 0, \cdots N$

(v) and

$$\underline{F} = \sum_{i=0}^{N} S_i f_i \text{ and } \bar{F} = \sum_{i=0}^{N} S_i \bar{f}_i. \tag{34}$$

(a) S_i is uniquely defined for $i = 0, \cdots N$, and $I - T$ or T, resp., is invertible, in particular, there is a matrix A with the desired properties

(b) We have $\underline{\varphi}\underline{F} \le \psi f \le \bar{\varphi}\bar{F}$

(c) If $\psi g < \infty$, we have

$$\lim_{N \to \infty} \underline{\varphi}\underline{F} = \psi F = \lim_{N \to \infty} \bar{\varphi}\bar{F}. \tag{35}$$

Remark 6. Theorem 5 is phrased for Markov chains with an infinite state space, and the goal is to find approximations to ψf which rely on the transition probabilities for transitions within a finite subset of the state space. Since even finite state spaces can be so large that an exact computation of ψf cannot be performed, we might want to apply Theorem 5 to Markov chains with a finite state space in order to reduce the state space to a smaller one.

Indeed, for $E = E_0 \cup \cdots \cup E_M$, $K \le N < M$, $\sum_{i=K+1}^{M} P_{ij} g_j + f_i \le g_i$ for $i = K+1, \cdots, M$, and $\bar{f}_i = f_i + \sum_{j=K+1}^{M} P_{ij} g_j$, (a) and (b) in Theorem 5 remain true. Note that formally, we cannot allow $N = M$ since $S_K = (H_{xy})_{x,y \in G_a}$ would become stochastic, and $I - T$ would not be invertible anymore. However, $N = M$ means that we have no reduction of the state space, and hence, this setting makes no sense. Part (c) becomes obsolete.

4. Efficiently Computing the Bounds for Quasi-Birth-Death Process (Discrete-Time Setting)

We turn now to developing methods for computing both bounds on ψf efficiently (and simultaneously). As pointed out in the introduction, we focus on quasi-birth-death processes (QBDs). QBDs are characterized by $P_{ij} = 0$ or $Q_{ij} = 0$, resp., for $i, j \in \mathbb{N}_0$ with $|i - j| \geq 2$, that is, each jump changes the level at most by 1.

Level-independent QBDs were analysed by Neuts [18] by means of matrix-geometric methods. First algorithmic approaches for level-dependent QBDs (LDQBDs) are due to Bright and Taylor [17] and Hanschke [19]. The approach in [17] generalizes Neuts' probabilistic interpretations of the matrices which arise in the matrix-geometric method, whereas the approach in [19] is motivated by the relationship between second-order vector-matrix difference equations and matrix-valued continued fractions. Remarkably, both methods are equivalent (up to suggestions regarding some initializations). More details and comparisons can be found in [20, 21]. All these methods intend to find (approximations to) an invariant measure or the invariant distribution. In [22], an algorithm was suggested which allows to compute stationary expectations directly. In the continuous-time setting, this method reads as follows:

(i) Choose large N, set $R = 0$ and $F = f_N$

(ii) For $i = N-1, N-2, \cdots, 0$, replace R by $Q_{i,i+1}(-Q_{i+1,i+1} - RQ_{i+2,i+1})^{-1}$ and F by $f_i + RF$

(iii) Determine ψ_0 as an (approximative) solution to $x(-Q_{00} - RQ_{10}) = 0$

(iv) Return $\psi_0 F$.

Since the memory requirement does not depend directly on N (with $d_i = |E_i|$, it depends on $\max\{d_i : 0 \leq i \leq N\}$), the truncation level N can be chosen very large. Despite the possibility to choose large N, results on the truncation error still remained desirable from both a mathematical and a practical points of view.

It is not difficult to prove that (with the notation of the present paper) the method from [22] computes the lower bound on ψf for $K = 0$ (K has no impact on the lower bound). The goal of this section is to generalize this method in such a way that any value $K \in \{0, \cdots, N\}$ is allowed, and that the upper bound will be computed, too. For means of conciseness, we focus on the continous-time setting.

Note that the QBD property directly implies $Q_{ij} = 0$ for $i > K > j$, that is, (7) holds. The drift condition (6) is met if

$$Q_{i,i-1}g_{i-1} + Q_{ii}g_i + Q_{i,i+1}g_{i+1} + f_i \leq 0, \quad i \geq K+1, \tag{36}$$

and the definition of $\bar{f}_i$ simplifies to

$$\bar{f}_i = \begin{cases} f_i & i < N, \\ f_N + Q_{N,N+1}g_{N+1,} & i = N. \end{cases} \tag{37}$$

Most importantly, the computation of the matrices S_i according to the system

$$0 = \sum_{i=0}^{N} S_i Q_{ij}, \quad j \neq K, \tag{38}$$

becomes much easier (the matrix-analytic methods in, e.g., [17, 19–21] use this fact). Many of the following identities are consequences of the probabilistic interpretation of the involved matrices, and we omit these proofs since the considerations are similar to those in [17] anyway.

(i) Let $R_K, \cdots, R_{N-1}$ be defined by $R_{N-1} = Q_{N-1,N}(-Q_{NN})^{-1}$ and

$$R_{i-1} = Q_{i-1,i}(-Q_{ii} - R_i Q_{i+1,i})^{-1} \quad K < i < N. \tag{39}$$

Then, the inverses in (39) exist, and we have

$$S_i = R_K R_{K+1} \cdots R_{i-1}, \quad i = K+1, \cdots N \tag{40}$$

(ii) Similarly, for $B_1, \cdots, B_K$ defined by $B_1 = Q_{10}(-Q_{00})^{-1}$ and

$$B_{i+1} = Q_{i+1,i}(I - Q_{ii} - B_i Q_{i-1,i})^{-1}, \quad 0 < i < K, \tag{41}$$

we have

$$S_i = B_K B_{K-1} \cdots B_{i+1,} \quad i = 0, \cdots, K-1 \tag{42}$$

(iii) The structure of Q allows to write $T = -Q_{KK} - B_K Q_{K-1,K} - R_K Q_{K+1,K}$.

If the level sizes E_i are relatively small, the computational effort induced by (39) and (41) is acceptable. Then, it seems natural to compute all R_i and B_i, then all S_i and finally $\underline{F}$ and $\bar{F}$. However, in [22], it was pointed out that for $K = 0$, computing $\underline{F}$ can be performed in a much more efficient way (in particular with respect to memory requirement) by using a Horner-type scheme. Here, we generalize this procedure slightly with respect to two issues: We consider an arbitrary K, and we consider the simultaneous computation of $\underline{F}$ and $\bar{F}$. Introduce

$$\begin{aligned} Z_n &= f_n + \sum_{m=n+1}^{N} \prod_{r=n}^{m-1} R_r f_m, \quad n = K, K+1, \cdots N, \\ \bar{Z}_n &= \bar{f}_n + \sum_{m=n+1}^{N} \prod_{r=n}^{m-1} R_r \bar{f}_m, \qquad n = 1, \cdots K. \\ W_n &= \sum_{m=0}^{n-1} \prod_{r=0}^{n-1-m} B_{n-r} f_m, \qquad n = 1, \cdots, K. \end{aligned} \tag{43}$$

With this notation, we have

$$\begin{aligned}
\underline{F} &= W_K + Z_K,\\
\bar{F} &= W_K + \bar{Z}_K,\\
Z_N &= f_N,\\
Z_n &= f_n + R_n Z_{n+1}, \quad n = K, K+1, \cdots, N-1,\\
\bar{Z}_N &= \bar{f}_N,\\
\bar{Z}_n &= \bar{f}_n + R_n \bar{Z}_n, \quad n = K, K+1, \cdots, N-1,\\
W_1 &= B_1 f_0,\\
W_n &= B_n(W_{n-1} + f_{n-1}), \quad n = 2, \cdots, K-1.
\end{aligned} \tag{44}$$

The recurrence schemes for R_n, Z_n, and $\bar{Z}_n$ all start at $n = N$, and are used in the sequel for computing the other values for $n = N-1, N-2, \cdots, K$. Similarly, the computation of B_n and W_n starts at $n = 1$, and then the values for $n = 2, 3, \cdots, K$ can be computed. For finding $\underline{\varphi}, \bar{\varphi}, \underline{F}, \bar{F}$, we only need the values for $n = K$. Hence, for $n > K$, R_n is only used for finding Z_n, $\bar{Z}_n$, and R_{n-1}. Similarly, the other matrices are only used in a single step. Hence, there is no need to store these matrices. Instead, we suggest to use the recurrence scheme for all these matrices as an "update" procedure. In total, we suggest the following method.

(i) Choose$N$, set $R = 0$, $Z = f_N$, and $\bar{Z} = f_N + Q_{N,N+1} g_{N+1}$.

(ii) For $i = N-1, N-2, \cdots, K$, update

$$\begin{aligned}
-R &= Q_{i,i+1,} \left(-Q_{i+1,i+1} - RQ_{i+2,i+1}\right)^{-1},\\
-Z &= f_i + RZ, \text{and } \bar{Z} = f_i + r\bar{Z}.
\end{aligned} \tag{45}$$

(iii) Set $B = Q_{10}(-Q_{00})^{-1}$, and $W = Bf_0$.

(iv) For $i = 2, 3, \cdots, K$, update

$$\begin{aligned}
-B &= Q_{i,i-1}(-Q_{i-1,i-1} - BQ_{i-2,i-1})^{-1},\\
&\text{–and } W = B(f_{i-1} + W).
\end{aligned} \tag{46}$$

(v) Set $T = -Q_{KK} - BQ_K - {}_{1,K} - RQ_{K+1,K}$ and determine $A = (a_{xy})_{x,y \in E} K$ such that AT is a diagonal matrix, define $\underline{\varphi}$ by $\underline{\varphi}_x = a_{x_0,x}/a_{x_0,x_0}$, and define $\bar{\varphi}$ by $\bar{\varphi}_x = a_{x_{,},x}/a_{x,x0}$.

(vi) Return $\underline{\varphi}(W + Z)$ as a lower bound and $\bar{\varphi}(W + \bar{Z})$ as an upper bound.

Usually the matrices Q_{ij} and f_i can be generated when they are needed. Up to four of these matrices are needed at the same time, and we need memory for saving R, Z, B, W. In total, a small number of finite matrices have to be stored at the same time. In particular, if $|d_i| = E_i$ is bounded by d, the memory requirement is $\leq 10d^2 + 5d(L+1)$, and this bound does not depend on K or N. Note that the "price" for this low memory requirement is that we only calculate the values $\psi f^{(\ell)}$ for cost/reward functions $f^{(\ell)}$ which have to be specified before the computation procedure begins. Since we do not store the values ψ_x or the matrices S_i, adding any "new interesting" function requires a complete restart of the method.

A discussion of all numerical details of the algorithm is beyond the scope of this paper. For two specific aspects (avoiding ill-conditioned problems when computing ϕ and avoiding instabilities in the update step for B), we refer to the Appendix.

5. Application to the $M/M/1$ Queue

We first consider an example where we have explicit terms for $\psi_x, H_{xy}, \cdots$. Of course, we would never use numerical methods for finding bounds on ψf in such a situation, but the following considerations illustrate how the method works, and they show how sharp the upper bound can be. Numerical examples for situations in which we do not have an explicit analytical representation for ψ_x will follow in Section 6 and Section 7.

Consider the $M/M/1$ queue, that is, we have $E = \mathbb{N}_0$ and

$$Q = \begin{pmatrix}
-\lambda & \lambda & & & \\
\mu & -(\lambda+\mu) & \lambda & & \\
 & \mu & -(\lambda+\mu) & \lambda & \\
 & & \ddots & \ddots & \ddots
\end{pmatrix}. \tag{47}$$

$\lambda > 0$ is referred to as *arrival rate*, and $\mu > 0$ is called *service rate*. Q is bounded, and therefore, nonexplosive. For any $\lambda, \mu > 0$, Q is irreducible, and we will assume positive recurrence which is equivalent to $\rho := \lambda/\mu < 1$.

Here, we have $E_i = \{i\}$, and we will first consider choices of g which allow choosing $K = 0$.

The invariant measure ψ with $\psi_0 = 1$ is given by $\psi_i = \rho^i$, and the 1×1-matrices S_i shall solve $0 = \sum_{j=0}^{N} S_j Q_{ji}$ for $i > 0$ which reads as

$$\begin{aligned}
0 &= S_{i-1}\lambda - S_i(\lambda+\mu) + S_{i+1}\mu,\ i = 1, \cdots, N-1,\\
0 &= S_{N-1}\lambda - S_N(\lambda+\mu).
\end{aligned} \tag{48}$$

This system of equations can be solved explicitly (e.g., by using standard results concerning linear difference equations with constant coefficients), and we find

$$S_i = \frac{\rho^i - \rho^{N+1}}{1 - \rho^{N+1}}, \quad i = 0, \cdots, N. \tag{49}$$

Due to $|E_0| = 1$, we do not have to find the upper and lower bounds on $(\psi_x)_{x \in E0} = (1)$, and instead, we focus on bounds on F.

First, consider $f(i) = 1$. As pointed out above, we have positive recurrence, and therefore, ψf should be finite. From the explicit term, we obtain $\psi f = 1/1 - \rho < \infty$. To show the finiteness of ψf by means of the drift criteria, set $g(i) = 1/\mu - \lambda$. Then, for all $i \geq 1$, we have

$$\begin{aligned} &q_{i,i-1}g(i-1) + q_{ii}g(i) + q_{i,i+1}g(i+1) + f(i) \\ &\quad = \frac{\mu(i-1) - (\lambda+\mu)i + \lambda(i+1)}{\mu-\lambda} + 1 = \frac{\lambda-\mu}{\mu-\lambda} + 1 = 0. \end{aligned} \tag{50}$$

By Theorem 2 in [8] (or older references, see remark in the preliminaries), we obtain positive recurrence. Furthermore, the requirements of Theorem 5 are met. Therefore, we have $F \leq \psi f \leq \bar{F}$, where

$$\begin{aligned} \underline{F} &= \sum_{i=0}^{N} S_i f(i) = \sum_{i=0}^{N} \frac{\rho^i - \rho^{N+1}}{1-\rho^{N+1}} = \frac{1}{1-\rho} - \frac{(N+1)\rho^{N+1}}{1-\rho^{N+1}}, \\ \bar{F} &= \underline{F} + S_N q_{N,N+1} g(N+1) = \underline{F} + \frac{\rho^N - \rho^{N+1}}{1-\rho^{N+1}} \cdot \lambda \cdot \frac{N+1}{\mu-\lambda} \\ &= \frac{1}{1-\rho} - \frac{(N+1)\rho^{N+1}}{1-\rho^{N+1}} + \frac{\rho^N(1-\rho)}{1-\rho^{N+1}} \cdot \frac{\rho(N+1)}{1-\rho} = \frac{1}{1-\rho}. \end{aligned} \tag{51}$$

Hence, the lower bound converges to $\psi f = 1/(1-\rho)$, and the upper bound coincides with ψf which is the best possible result. It is clear that this can only happen if we have $\sum q_{ij} g(j) + f(i) = g(i)$ (instead of mere $\leq$ for all i. Consider now ψf where $f = 1_A$, that is, $f(x) = 1$ for $x \in A$ and $f(x) = 0$ otherwise. Obviously, we can still use the same function g, and with the same computation as before, we obtain

$$\begin{aligned} \underline{F} &= \sum_{i=0}^{N} \frac{\rho^i - \rho^{N+1}}{1-\rho^{N+1}} \mathbf{1}_A(i), \\ \bar{F} &= \sum_{i=0}^{N} \frac{\rho^i - \rho^{N+1}}{1-\rho^{N+1}} \mathbf{1}_A(i) + \frac{\rho^{N+1}(N+1)}{1-\rho^{N+1}}. \end{aligned} \tag{52}$$

Let $A = \{0, \cdots, M\}$, and let $N > M$. If we are interested in computing the bounds on $\psi \mathbf{1}_A / \psi \mathbf{1}$, we combine the previous bounds, and the numerical algorithm would return

$$\begin{aligned} &\frac{\sum_{i=0}^{N} \left(\left(\rho^i - \rho^{N+1}\right)/\left(1-\rho^{N+1}\right) \right) \mathbf{1}_A(i)}{(1/(1-\rho))} \leq \frac{\psi \mathbf{1}_A}{\psi \mathbf{1}} \\ &\quad \leq \frac{\sum_{i=0}^{N} \left(\left(\rho^i - \rho^{N+1}\right)/\left(1-\rho^{N+1}\right) \right) \mathbf{1}_A(i) + \left(\rho^{N+1}(N+1)\right)/\left(1-\left(\rho^{N+1}\right)\right)}{(1/(1-\rho)) - ((N+1)\rho^{N+1})/(1-\rho^{N+1})}. \end{aligned} \tag{53}$$

For any choice of A, both bounds will converge to $(1-\rho)\sum_{i \in A} \rho^i$ which is the stationary probability for set A. As an easy choice, let $A = \{j\}$ for some $j \in \mathbb{N}_0$. Then, we know that the stationary probability is given by $\pi_j = (1-\rho)\rho^j$, and for $N \geq j$, we obtain the bounds

$$\begin{aligned} &\frac{1-\rho}{1-\rho^{N+1}} \left(\rho^j - \rho^{N+1}\right) \\ &\quad \leq \pi_j \leq (1-\rho) \cdot \frac{\rho^j - \rho^{N+1} + \rho^{N+1}(N+1)}{1-\rho^{N+1} - (1-\rho)(N+1)\rho^{N+1}}. \end{aligned} \tag{54}$$

Next, let $f = id$, that is, $f(i) = i$ for all $i \in \mathbb{N}_0$. Note that $\psi \mathrm{id}/\psi 1$ is the expected stationary number of customers in the system which is one of the most important performance measures in queueing theory. We set $g(i) = (i^2/2(\mu-\lambda)) + (i(\mu+\lambda)/2(\mu-\lambda)^2)$, and then we have

$$\begin{aligned} &q_{i,i-1}g(i-1) - q_{ii}g(i) + q_{i,i+1}g(i+1) + f(i) \\ &\quad = \frac{\mu(i-1)^2 - (\lambda+\mu)i^2 + \lambda(i+1)^2}{2(\mu-\lambda)} \\ &\qquad + \frac{(\mu+\lambda)(\mu(i-1) - (\lambda+\mu)i + \lambda(i+1))}{2(\mu-\lambda)^2} + i \\ &\quad = \frac{-2(\mu-\lambda)i + \mu + \lambda}{2(\mu-\lambda)} + \frac{-(\mu+\lambda)(\mu-\lambda)}{2(\mu-\lambda)^2} + i = 0. \end{aligned} \tag{55}$$

The exact value is $\psi f = \rho/(1-\rho)^2$, the computed lower bound is given by

$$\begin{aligned} \underline{F} &= f(0) + \sum_{i=1}^{N} S_i f(i) = \sum_{i=1}^{N} \frac{\rho^i - \rho^{N+1}}{1-\rho^{N+1}} \cdot i \\ &= \frac{\rho}{(1-\rho)^2(1-\rho^{N+1})} \\ &\quad \cdot \left(N\rho^{N+1} - (N+1)\rho^N + 1\right) - \frac{\rho^{N+1}N(N+1)}{2(1-\rho^{N+1})} \\ &= \frac{\rho}{(1-\rho)^2} + \frac{\rho}{(1-\rho)^2(1-\rho^{N+1})} \\ &\quad \cdot \left(N\rho^{N+1} - (N+1)\rho^N + \rho^{N+1}\right) - \frac{\rho^{N+1}N(N+1)}{2(1-\rho^{N+1})} \\ &= \psi f - \frac{\rho^{N+1}(N+1)(1-\rho)}{(1-\rho)^2(1-\rho^{N+1})} - \frac{\rho^{N+1}N(N+1)}{2(1-\rho^{N+1})} \\ &= \psi f - \frac{\rho^{N+1}(N+1)}{2(1-\rho)(1-\rho^{N+1})}(2 + N(1-\rho)), \end{aligned} \tag{56}$$

and the upper bound is given by

$$\begin{aligned} \bar{F} &= \underline{F} + S_N \cdot \lambda \cdot g(N+1) \\ &= \underline{F} + \frac{\rho^N - \rho^{N+1}}{1-\rho^{N+1}} \left(\frac{\rho(N+1)^2}{2(1-\rho)} + \frac{\rho(N+1)(1+\rho)}{2(1-\rho)^2} \right) \\ &= \underline{F} + \frac{\rho^{N+1}(1-\rho)(N+1)}{2(1-\rho)^2(1-\rho^{N+1})} \left((N+1)(1-\rho) + 1 + \rho\right) \\ &= \psi f. \end{aligned} \tag{57}$$

Hence again, we find the best possible upper bound $\bar{F} = \psi f$. Appropriately combining the bounds on $\psi \mathbf{1}$ and ψid leads to the upper and lower bounds on the expected stationary number πf of customers in the system.

At a first glance, the choice $g(i) = (i^2/2(\mu - \lambda)) + (i(\mu + \lambda)/2(\mu - \lambda)^2)$ might be "difficult to guess," but note that we have used a function g of the form $g(i) = ci$ for dealing with $f(i) = 1$, and it is quite natural to choose a function g of the form $g(i) = c_1 i^2 + c_2 i$ for dealing with $f(i) = i$. Afterwards, it is not difficult to determine c_1, c_2 such that the f-modulated drift condition is met. Nevertheless, in other situations, we will not be able to find such an optimal function g. Therefore, we demonstrate next that the method still works if we use some function g which is far from being optimal.

Let us directly consider $f = (\mathbf{1}, \mathbf{1}_A, \mathrm{id})$, that is, $L = 2, f^{(0)}(i) = 1, f^1(i) = \mathbf{1}_A(i)$, and $f^{(2)}(i) = i$. Then, $\psi f^{(1)}/\psi f^{(0)}$ is the stationary probability for set A, and $\psi f^{(2)}/\psi f^{(0)}$ is the expected stationary number of customers in the system. Let us choose $g^{(\ell)}(i) = a^i$ for $i \in \mathbb{N}_0$ and $\ell = 0, 1, 2$. Then

$$\begin{aligned} & q_{i,i-1} g^{(\ell)}(i-1) + q_{ii} g^{(\ell)}(i) + q_{i,i} + 1 g^{(\ell)}(i+1) + f^{(\ell)}(i) \\ & \quad = \mu a^{i-1} - (\lambda + \mu) a^i + \lambda a^{i+1} + f^{(\ell)}(i) \\ & \quad = a^i (\lambda a - \mu)(a - 1) + f^{(\ell)}(i) \leq 0, \end{aligned} \tag{58}$$

for $\ell = 0, 1, 2$ and all $i > K$ for some sufficiently large K if we choose $a \in (1, \mu/\lambda) = (1, 1/\rho)$. Due to $K > 0$, the values S_i will change. Now, we have

$$\begin{aligned} & 0 = -S_0 \lambda + S_{1\mu}, \\ & 0 = S_{i-1} \lambda - S_i (\lambda + \mu) + S_{i+1} \mu, \; i = 1, \cdots, K-1, K+1, \cdots, N, \\ & 0 = S_{N-1} \lambda - S_N (\lambda + \mu). \end{aligned} \tag{59}$$

The solution to this system is given by

$$S_i = \begin{cases} \rho^{i-k} & i = 0, \cdots, K \\ \dfrac{\rho^{i-k} - \rho^{N+1-K}}{1 - \rho^{N+1-k}}, & i = K, \cdots, N. \end{cases} \tag{60}$$

We omit an explicit representation of $\underline{F}$ and $\bar{F}$ (which is strictly larger than ψf here), but we remark that

$$\bar{F} - \underline{F} = S_N q_{N,N+1} g(N+1) = \frac{\rho^{N-K}(1-\rho)}{1 - \rho^{N+1-K}} \cdot \lambda \cdot a^{N+1}, \tag{61}$$

and due to $\rho < \rho a < 1$, this difference tends to 0 as $N \longrightarrow \infty$. Note that the speed of convergence of $\bar{F} - \underline{F}$ to 0 is still exponentially fast, but slower than for the optimal choices of $g^{(\ell)}$.

6. Application to a Variant of the M/PH/1 Queue

Next, we consider an application of our method to a queueing model where we do not know the exact invariant distribution and have no chance but to use numerical methods. We consider a variant of the M/PH/1 queue where arriving customers decide whether to join the queue or not depending on the number of customers they find in the queue. Precisely, we assume that

(i) Customers arrive according to the Poisson process with intensity λ

(ii) An arriving customer joins the queue with probability α_n if there are n other customers in the system upon the arrival

(iii) The service time is phase-type distributed with parameters $\beta \in \mathbb{R}^{1 \times d}$ and $B \in \mathbb{R}^{d \times d}$, that is, the cumulative distribution function of a service time is given by $t \mapsto 1 - \beta e^{Bt} \mathbf{1}$ and its expectation by $1/\mu := \beta(-B)^{-1} 1$

(iv) There is one server.

This queueing system can be modelled as the Markov chain $Y = (Y_t)_{t \geq 0}$, where $Y_t = (N_t, U_t)$ is two-dimensional: N_t is the number of customers in the system at time t, and for $N_t > 0$, $U_t \in \{1, \cdots, d\}$ is the current service phase. By setting the service phase to 1 for $N_t = 0$, we obtain levels E_i with $E_0 = \{(0, 1)\}$ and $E_i = \{(0, 1), \cdots, (0, d)\}$ for $i \geq 1$. The generator matrix is given by

$$Q = \begin{pmatrix} -\lambda\alpha_0 & \lambda\alpha_0\beta & & & & \\ -B\mathbf{1} & B - \lambda\alpha_1 I & \lambda\alpha_1 I & & & \\ & -B\mathbf{1}\beta & B - \lambda\alpha_3 I & \lambda\alpha_2 I & & \\ & & -B\mathbf{1}\beta & B - \lambda\alpha_3 I & \lambda\alpha_3 I & \\ & & & \ddots & \ddots & \ddots \end{pmatrix}. \tag{62}$$

For $\alpha_n = 1$, we obtain a level-independent QBD (the classical $M/PH/1$ queue) for which an explicit analytical representation of the solution exists (see, e.g., [23]). For nonconstant α_n, we obtain an LDQBD, and we have to use numerical methods. In what follows, we assume that

$$\lambda . \lim_{n \to \infty} \sup \alpha_n < \mu, \tag{63}$$

which is equivalent to the existence of some $\tilde{\lambda} < \mu$ with $\lambda \alpha_n \leq \tilde{\lambda}$ for all $n \geq K_0$ with some $K_0 \in N_0$. Without restriction, we assume $K_0 \geq 2$. Condition (63) guarantees that we have positive recurrence and that the stationary number of customers in the system has finite expectation as we will prove below. Hence, under condition (63), it makes sense to compute

(i) the stationary probability $\pi f^{(1)}$ that an arriving customer joins the queue where $f_i^{(1)} = \mathbf{1} \cdot \alpha_i$ for all i

(ii) the stationary probability $\pi f^{(2)}$ that at least M customers are in the system where $f_i^{(2)} = \mathbf{1}$ for $i \geq M$ and $f_i^{(2)} = 0$ and $i < M$

(iii) the expectation $\pi f^{(3)}$ of the stationary number of customers in the system where $f_i^{(3)} = \mathbf{1} \cdot i$ for all i.

In order to use our algorithmic method, we set $f_i^{(0)} = 1$ additionally. For applying our method and for proving $\psi f^{(\ell)} < \infty$ for $\ell = 0, 1, 2, 3$ under condition (63), we check some f-modulated drift conditions. Set $h_i = Q_{i,i-1} g_{i-1} + Q_{ii} + Q_{i,i+1} g_{i+1}$ for $i \geq 1$, and then we have $h_i = -B\mathbf{1}\beta g_{i-1} + B g_i + \lambda \alpha_i (g_{i+1} - g_i) \leq -B\mathbf{1}\beta g_{i-1} + B g_i + \tilde{\lambda}(g_{i+1} - g_i)$ for $i \geq K_0$ and all choices of g for which $g_{i+1} - g_i \geq 0$ (componentwise). For $\ell = 0, 1, 2$, we have $f_i^{(\ell)} \leq 1$, and we set $g_i^{(\ell)} = (\mathbf{1} \cdot i + (-B)^{-1}\mathbf{1}\mu) \cdot 1/(\mu - \tilde{\lambda})$ for $i > 1$. Due to $\beta\mathbf{1} = 1$, we obtain

$$\begin{aligned} h_i^{(\ell)} + f_i^{(\ell)} &\leq \left(-B\mathbf{1}\beta\mathbf{1} \cdot (i-1) - B\mathbf{1}\beta(-B)^{-1}\mathbf{1} \cdot \mu \right. \\ &\quad \left. + B\mathbf{1} \cdot i - \mathbf{1} \cdot \mu + 1 \cdot \tilde{\lambda}\right) \cdot \frac{1}{\mu - \tilde{\lambda}} + f_i \\ &= \left(-B\mathbf{1} \cdot (i-1) - B\mathbf{1} + B\mathbf{1} \cdot i - \mathbf{1}\left(\mu - \tilde{\lambda}\right)\right) \\ &\quad \cdot \frac{1}{\mu - \tilde{\lambda}} + \mathbf{1} = 0, \end{aligned} \tag{64}$$

for $i \geq K_0$. From the special case $f_i^{(0)} = 1$, we obtain positive recurrence from Theorem 2 in [8], and by the same result, we have $\psi f^{(\ell)} < \infty$ for $\ell = 1, 2$. For $\ell = 3$, we have $f_i^{(3)} = \mathbf{1} \cdot i$, and we set

$$g_i^{(3)} = \left(\mathbf{1} \cdot i^2 + (-B)^{-1}\mathbf{1} \cdot 2\mu i\right) \cdot \frac{1}{\mu - \tilde{\lambda}}, \tag{65}$$

for $i \geq 1$ (note that $(-B)^{-1}$ is a nonnegative matrix, and hence, we have $g_{i+1}^{(3)} \geq g_i^{(3)}$, implying

$$\begin{aligned} &h_i^{(3)} \cdot \left(\mu - \tilde{\lambda}\right) \\ &\leq -B\mathbf{1}\beta\mathbf{1}(i+1)^2 - B\mathbf{1}\beta(-B)^{-1}\mathbf{1} \cdot 2\mu(i+1) \\ &\quad + B\mathbf{1}i^2 - \mathbf{1} \cdot 2\mu i + \mathbf{1}\tilde{\lambda}\left((i+1)^2 - i\right)^2 + (-B)^{-1}\mathbf{1} \cdot 2\tilde{\lambda}\mu \\ &= -B\mathbf{1}\left(i^2 - 2i + 1\right) - B\mathbf{1}(2i-2) + B\mathbf{1}i^2 - \mathbf{1} \cdot 2\mu i + \mathbf{1} \cdot 2\tilde{\lambda} i \\ &\quad + \mathbf{1}\tilde{\lambda} + (-B)^{-1}\mathbf{1} \cdot 2\tilde{\lambda}\mu \\ &= -\mathbf{1} \cdot 2\left(\mu - \tilde{\lambda}\right) i - B\mathbf{1} + \mathbf{1}\tilde{\lambda} + (-B)^{-1}\mathbf{1} \cdot 2\tilde{\lambda}\mu, \end{aligned} \tag{66}$$

for $i \geq K_0$. Hence,

$$h_i^{(3)} + f_i^{(3)} \leq -\left(B\mathbf{1} + \mathbf{1}\tilde{\lambda} + (-B)^{-1}\mathbf{1} \cdot 2\tilde{\lambda}\mu\right) \cdot \frac{1}{\mu - \tilde{\lambda}} - \mathbf{1} \cdot i. \tag{67}$$

Choose $K \geq K_0$ such that

$$\left(-B\mathbf{1} + \mathbf{1}\tilde{\lambda} + (-B)^{-1}\mathbf{1} \cdot 2\tilde{\lambda}\mu\right) \cdot \frac{1}{\mu - \tilde{\lambda}} \leq \mathbf{1} \cdot (K+1). \tag{68}$$

Then, we have $h_i^{(3)} + f_i^{(3)} \leq 0$ for $i > K$. The existence of such a value K proves $\sum_{i=0}^{\infty} \pi_i \mathbf{1} \cdot i < \infty$, that is, the stationary number of customers has finite expectation.

For applying our algorithmic method, we have to specify K. In our numerical examples, we set $\lambda = 5$, $\mu = 1$, $\alpha_n = 1/10 + 9/(10(n+1))$. This allows to choose $\tilde{\lambda} = 3/4$, and it is easy to derive that $\lambda\alpha_n \leq \tilde{\lambda}$ for $n \geq K_0 = 17$. Furthermore, we consider the special case where the service times are Erlang-2-distributed, that is, $\beta = (1, 0)$ and

$$B = \begin{pmatrix} -2\mu & 2\mu \\ 0 & -2\mu \end{pmatrix}, \tag{69}$$

implying

$$-B\mathbf{1} = \begin{pmatrix} 0 \\ 2\mu \end{pmatrix}, (-B)^{-1} = \begin{pmatrix} \frac{1}{2\mu} & \frac{1}{2\mu} \\ 0 & \frac{1}{2\mu} \end{pmatrix}, \tag{70}$$

and

$$(-B)^{-1}1 \cdot \mu = \begin{pmatrix} 1 \\ 1 \\ \frac{1}{2} \end{pmatrix}. \tag{71}$$

Then, (68) is guaranteed if $((2\mu + 3\tilde{\lambda})/(\mu - \tilde{\lambda})) - (K+1) \leq 0 \Leftrightarrow K \geq ((2\mu + 3\tilde{\lambda})/(\mu - \tilde{\lambda})) - 1 = (\mu + 4\tilde{\lambda})/(\mu - \tilde{\lambda})$. For $\mu = 1$ and $\tilde{\lambda} = 3/4$, this results in $K \geq 8$. Since we have to choose $K \geq K_0$, we set $K = 17$. Note that $g_i^{(\ell)}$ simplifies to

$$\begin{aligned} g_i^{(\ell)} &= \left(\mathbf{1} \cdot i + (-B)^{-1}\mathbf{1}\mu\right) \cdot \frac{1}{\mu - \tilde{\lambda}} = \begin{pmatrix} i+1 \\ i + \frac{1}{2} \end{pmatrix} 4 \quad \text{for } \ell = 0, 1, 2, \\ g_i^{(3)} &= \left(\mathbf{1} \cdot i^2 + (-B)^{-1}\mathbf{1} \cdot 2\mu i\right) \cdot \frac{1}{\mu - \tilde{\lambda}} = \begin{pmatrix} i^2 + 2 \\ i^2 + 1 \end{pmatrix} \cdot 4. \end{aligned} \tag{72}$$

In Table 1, some numerically computed results are listed. The parameters are chosen as specified above, and additionally, we set $M = 30$ for $\pi f^{(2)}$. The results clearly illustrate the convergence of both bounds to the same value. For obtaining precise results for $\pi f^{(2)}$, the truncation level N has to be chosen a bit larger than for obtaining quite precise bounds on $\pi f^{(1)}$ and $\pi f^{(3)}$. This is not surprising since due to the fact that $\pi f^{(2)}$ is the stationary probability for at least $M = 30$ customers, the truncation level should be chosen significantly larger than 30. However, the results demonstrate that even for the computation of $\pi f^{(2)}$, the truncation level 55 yields very precise results.

7. Application to a Retrial Queueing Model

As a final example, we consider the $M/M/c/d - 1$ queue with retrials which can be seen as some kind of "prototype" for LDQBDs.

Table 1: Numerical results for the considered variant of the $M/PH/1$ queue.

N	$\pi f^{(1)}$ Lower bound	Upper bound	$\pi f^{(2)} \geq$ Lower bound	Upper bound	$\pi f^{(3)}$ Lower bound	Upper bound
20	0.167971	0.329700	0	0.111918	7.80187	12.3849
25	0.196764	0.21197	0	0.010557	9.31267	9.79899
30	0.199698	0.200915	$1.05844 \times 10 - 5$	$8.76397 \times 10 - 4$	9.49181	9.53472
35	0.199927	0.200002	$6.09687 \times 10 - 5$	$1.15438 \times 10 - 4$	9.50773	9.51063
40	0.19994	0.199944	$6.8538 \times 10 - 5$	$7.13126 \times 10 - 5$	9.5088	9.50895
45	0.199941	0.199941	$6.90867 \times 10 - 5$	$6.92067 \times 10 - 5$	9.50885	9.50886
50	0.199941	0.199941	$6.91157 \times 10 - 5$	$6.91203 \times 10 - 5$	9.50886	9.50886
55	0.199941	0.199941	$6.91170 \times 10 - 5$	$6.91172 \times 10 - 5$	9.50886	9.50886

(i) Customers arrive according to the Poisson process with intensity λ

(ii) The service times are independent and identically exponentially distributed with parameter μ

(iii) There are c servers

(iv) The system capacity is $d - 1 \geq c$, that is, $d - 1 - c$ customers can wait at the same time

(v) Customers who cannot enter the system due to lack of waiting capacity enter the "orbit" of retrials

(vi) Each customer in the orbit will retry to enter the system after a time which is exponentially distributed with parameter ν and independent of all other random variables

(vii) Retrying customers who cannot enter the system due to lack of waiting capacity stay in the orbit of retrying customers.

Let $Y_t = (O_t, N_t)$, where O_t is the number of customers in the orbit of retrials, and N_t is the number of customers in the queue (including service). Then, $Y = (Y_t)_{t \geq 0}$ is a continuous-time Markov chain with state space $E \bigcup_{i=0}^{\infty} E_i$, where $E_i = \{(i, 0), \cdots, (i, d-1)\}$. Obviously, a transition with a positive rate will change the level at most by 1, and hence, we have a quasi-birth-death process with the following state transitions from state (i, u):

(i) Arrivals occur with rate λ. For $u < d - 1$, the arriving customer enters the queueing system, and the new state is $(i, u + 1)$. For $u = d - 1$, the arriving customer enters the orbit of retrials, and the new state is $(i + 1, d - 1)$

(ii) For $u > 0$, service completions occur with rate $\max\{u, c\} \cdot \mu$, and the new state is $(i, u - 1)$

(iii) For $i > 0$ and $u < d - 1$, successful retrials occur with rate $i \cdot \nu$, and the new state is $(i - 1, u + 1)$.

In matrix notation, we have

$$Q_{i,i+1} = \begin{pmatrix} 0 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 \\ 0 & \cdots & 0 & \lambda \end{pmatrix}, \tag{73}$$

for $i \geq 0$,

$$Q_{i,i-1} = \begin{pmatrix} 0 & \nu i & & & \\ & 0 & \nu i & & \\ & & \ddots & \ddots & \\ & & & 0 & \nu i \\ & & & & 0 \end{pmatrix}, \tag{74}$$

for $i \geq 1$, and

$$Q_{ii} = \begin{pmatrix} -\lambda - \nu i & \lambda & & & & & & & \\ \mu & -\lambda - \mu - \nu i & \lambda & & & & & & \\ & 2\mu & -\lambda - 2\mu - \nu i & \lambda & & & & & \\ & & \ddots & \ddots & \ddots & & & & \\ & & & c\mu & -\lambda - c\mu - \nu i & \lambda & & & \\ & & & & \ddots & \ddots & \ddots & & \\ & & & & & c\mu & -\lambda - c\mu - \nu i & \lambda & \\ & & & & & & c\mu & -\lambda - c\mu \end{pmatrix}, \tag{75}$$

for $i \geq 1$.

Note that retrial queues have been discussed intensively in the literature. We refer to the textbook [24] which gives an overview on retrial queueing models and computational methods for determining invariant distributions, stationary expectations, etc. Computational methods are very important in this context since even for $d-1=c$, there are only explicit analytical representations of the invariant distribution if $c \leq 2$ (see [24]).

Of course, there are many interesting characteristics such as the mean number of customers in the orbit and the mean number of customers in the queueing system. Our purpose is to demonstrate that our method provides reliable results for interesting characteristics, but we do not want to discuss the retrial queueing model in details. Therefore, we focus on a single characteristic, the stationary probability for arriving customers to find the queuing system full and being forced to join the orbit. If $\pi = (\pi_{(i,u)})_{(i,u)\in E}$ denotes the invariant distribution (if there is one), this probability is given by

$$p_b := \sum_{i=0}^{\infty} \pi_{(u,d-1)}. \tag{76}$$

The invariant distribution exists if and only if we have positive recurrence. Below, we will restate a proof for the fact that this is the case if $\lambda < c\mu$. Note that $\lambda < c\mu$ is also a necessary requirement for stability/positive recurrence; we refer to [24] for more details.

We demonstrate how to prove that $\lambda < c\mu$ is sufficient for positive recurrence by means of drift criteria. The function g which we use for this purpose can also be used for finding the upper bounds on $\psi\mathbf{1}$ and $\psi\mathbf{1}_A$ for any $A \subset E$. In particular, by setting $A = \{(i, d-1) : i \in \mathbb{N}_0\}$, we obtain bounds on $p_b := \psi\mathbf{1}_A/\psi\mathbf{1}$.

In any case (both for proving recurrence if $\lambda < c\mu$ and for applying our truncation method), we want to find a function g such that we have $h_i + f_i := Q_{i,i-1}g_{i-1} + Q_{ii}g_i + Q_{i,i+1}g_{i+1} + f_i \leq 0$ for $i > K$ with some sufficiently large K, where $0 \leq f_i \leq 1$. We use the approach $g((i,u)) = \gamma i + \delta u$. For $i \geq 1$ and $u < d-1$, the entries of h_i compute as

$$\begin{aligned} h((i,u)) &= \nu i(g((i-1,u+1)) - g((i,u))) \\ &\quad + \lambda(g((i,u+1)) - g((i,u))) \\ &\quad + \max\{c,u\}\mu(g((i,u-1)) - g((i,u))) \\ &= \nu i(\delta - \gamma) + \lambda\delta - \max\{c,u\}\mu\delta \\ &\leq \nu i(\delta - \gamma) + \lambda\delta \,[= h((i,0))]. \end{aligned} \tag{77}$$

If we choose $\gamma > \delta$, we surely have $h((i,u)) + f((i,u)) \leq 0$ for $i > K$ with some sufficiently large K (due to $f((i,u)) \leq 1$). Additionally, we have to take

$$\begin{aligned} h((i,d-1)) &= \lambda(g((i+1,d-1)) - g((i,d-1))) \\ &\quad + c\mu(g((i,d-2)) - g((i,d-1))) \\ &= \lambda\gamma - c\mu\delta, \end{aligned} \tag{78}$$

into account. Hence, we have to guarantee $\lambda\gamma - c\mu\delta \leq -1$. Due to $\lambda < c\mu$, such a choice with $\gamma > \delta$ is possible. For example, we can choose $\delta = 2/c\mu - \lambda > 0$ and $\gamma = 1 + (c\mu/\lambda)/(c\mu - \lambda) > 2/c\mu - \lambda$, since then we have

$$\lambda\gamma - c\mu\delta = \frac{1}{c\mu - \lambda}(\lambda + c\mu - 2\lambda) = -1. \tag{79}$$

For this choice, we have $h((i,d-1)) + 1 \leq 0$ for all $i \geq 1$. As pointed out above, we have to guarantee that

$$\begin{aligned} h((i,u)) + 1 &\leq h((i,0)) + 1 \\ &= -\nu i(\gamma - \delta) + \lambda\delta + 1 \\ &= -\frac{\nu i}{\lambda} \cdot \frac{\lambda(1 + (c\mu/\lambda) - 2)}{c\mu - \lambda} + \frac{2\lambda + c\mu - \lambda}{c\mu - \lambda} \\ &= -\frac{\nu i}{\lambda} \cdot i + \frac{c\mu + \lambda}{c\mu - \lambda} \leq 0, \end{aligned} \tag{80}$$

for $i \geq K+1$. Hence, we choose

$$K + 1 = \left\lceil \frac{\lambda}{\nu} \cdot \frac{c\mu + \lambda}{c\mu - \lambda} \right\rceil. \tag{81}$$

In total, we set

$$\begin{aligned} f_i &= \begin{pmatrix} 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ 1 & 1 \end{pmatrix}, \\ g_i &= \begin{pmatrix} \gamma i & \gamma i \\ \gamma i + \delta & \gamma i + \delta \\ \vdots & \vdots \\ \gamma i + \delta(d-1) & \gamma i + \delta(d-1) \end{pmatrix}, \end{aligned} \tag{82}$$

where $\gamma = [1/\lambda \cdot (c\mu + \lambda)/(c\mu - \lambda)]$ and $\delta = 2/(c\mu - \lambda)$, and we choose K as above. Then, we can apply our algorithm which returns bounds on $\psi f^{(0)}$ and $\psi f^{(1)}$ where ψ is the invariant measure with $\psi_{(K,0)} = 1$. By dividing lower and upper bounds appropriately, we find bounds on

$$p_b = \frac{\psi f^{(1)}}{\psi f^{(0)}}. \tag{83}$$

In Tables 2 and 3, some numerically computed values for lower and upper bound are listed for different choices of the arrival rate λ and fixed values μ, ν, c, d. In both cases, we see

Table 2: Numerical results for $\lambda = 1.0$, $\mu = 1.0$, $\nu = 0.5$, $c = d - 1 = 5$, and $K = 4$.

Truncation level N	Lower bound on p_b	Upper bound on p_b
5	$1.85061 \times 10 - 3$	$5.62599 \times 10 - 3$
6	$2.79784 \times 10 - 3$	$3.71132 \times 10 - 3$
7	$3.12382 \times 10 - 3$	$3.32205 \times 10 - 3$
8	$3.20060 \times 10 - 3$	$3.24191 \times 10 - 3$
9	$3.21688 \times 10 - 3$	$3.22540 \times 10 - 3$
10	$3.22025 \times 10 - 3$	$3.22200 \times 10 - 3$
11	$3.22094 \times 10 - 3$	$3.22130 \times 10 - 3$
12	$3.22109 \times 10 - 3$	$3.22116 \times 10 - 3$
13	$3.22112 \times 10 - 3$	$3.22113 \times 10 - 3$
14	$3.22112 \times 10 - 3$	$3.22112 \times 10 - 3$

Table 3: Numerical results for $\lambda = 4.0$, $\mu = 1.0$, $\nu = 0.5$, $c = d - 1 = 5$, and $K = 73$.

Truncation level N	Lower bound on p_b	Upper bound on p_b
75	$2.67599 \times 10 - 5$	7218.46
80	$8.89475 \times 10 - 5$	2171.49
90	$7.39895 \times 10 - 4$	261.043
100	$5.79399 \times 10 - 3$	33.3353
110	$4.23642 \times 10 - 2$	4.55914
120	0.203545	0.948903
130	0.384933	0.501761
140	0.432069	0.447022
150	0.438578	0.440387
160	0.439374	0.439590
170	0.439469	0.439494
180	0.439480	0.439483
190	0.439481	0.439482
200	0.439482	0.439482

that lower and upper bounds converge to the same limit (which coincides with the value computed in Table 3.8 in the textbook [24] by other means). In Table 2, we have low traffic due to $\lambda/c\mu = 0.2$. Then, the blocking probability is quite low, and—more important for the evaluation of our method—we obtain precise results for low truncation levels. In Table 3, we have more traffic due to $\lambda/c\mu = 0.8$. Hence, the blocking probability is much larger, and we need higher truncation levels to obtain precise results.

The fact that we have to choose higher truncation levels N in the situation of Table 3 is anything but surprising since due to more traffic, the Markov chain will assume states within higher levels much more often. However, it is important to remark that it is not clear how large N should be chosen. The suggested method can be iteratively applied with increased truncation levels until a prescribed accuracy is achieved. Unfortunately, when increasing N, we have to restart most of the calculations (B and W do not change, but R, Z, and $\bar{Z}$ do). However, for each such calculation, the memory requirement is the same since it does not depend on N. In particular, in all situations considered in Tables 2 and 3, the memory requirement coincides.

As pointed out above, p_b is an important performance measure, but there are more interesting characteristics for this model. Note that with the same choice for g, we can find bounds on the stationary probability for any finite set of (finite or infinite) sets. In particular, if p_u denotes the stationary probabilities for u customers in the queueing system (waiting or in service), we could determine bounds on p_u for $u = 0, \cdots, d - 1$ simultaneously with little additional effort. Other performance measures (e.g., moments of the numbers of retrying customers) will require other choices for g.

We conclude this discussion of the basic retrial queueing model by remarking that there are many extensions of the retrial queueing model which improve the applicability. For example, we could consider Markovian arrival processes, phase-type distributed service times, impatient customers who leave the queue and enter the orbit, impatient customers who leave the orbit, queueing networks where declined external arrivals join the orbit of retrying customers, etc. For all such models, the suggested algorithm allows to compute important performance measures precisely and efficiently. Note that such retrial queues provide good models for many problems in telecommunications and computer networks, e.g., in the design of wireless networks. However, a deep discussion of a realistic (and hence complex) a model justifies a separate publication and is far beyond the scope of this methodological paper.

8. Evaluation of the Method

8.1. The f-Modulated Drift Condition. The major requirement for finding the bounds in Theorem 5 and for the resulting algorithmic procedure is the f-modulated drift condition (5) or (6). We give some comments on this condition.

(i) As pointed out above, f-modulated drift conditions are very popular for proving convergence/finiteness of ψf or πf. For $f \geq 0$, we remark that the existence of an f-modulated drift condition is even equivalent to the convergence of ψf. Hence, from a purely theoretical point of view, the requirement that an f-modulated drift condition is satisfied is no restriction

(ii) Most often, finding a function g which satisfies the f-modulated drift condition is the easiest way (or even the only way) to prove positive recurrence or to prove $\psi f < \infty$. If convergence of ψf is not guaranteed, it makes no sense to compute approximations to ψf (which are always finite) by numerical means. Hence, also from a practical point of view, finding a function g which satisfies the appropriate f-modulated drift condition does not require any additional effort in many situations

(iii) Every function g which can be used for proving $\psi f < \infty$ can be used in Theorem 5. Note that this is not true for all truncation bounds which rely on f-modulated drift conditions, see Section 8.6.

8.2. The Structural Requirement (7). We give some comments on condition (7). Remember that K is determined by the drift condition. If (7) is not met, we can always restructure the levels by setting $\tilde{E}_0 = E_0 \cup \cdots \cup E_K$ and $\tilde{E}_i = E_{K+i}$ for $i \geq 1$, and set $\tilde{K} = 0$. Then (7) is fulfilled. If we have the QBD structure for the levels E_i, this is preserved for the levels $\tilde{E}_i$. Consequently, we have $\tilde{d}_0 = d_0 + \cdots + d_K$. This is unfortunate since d_K (which becomes $\tilde{d}_0$ after the restructuring) should be quite small due to the operations involving matrices of dimension $d_K \times d_K$.

In total, from a purely mathematical point of view, $K = 0$ can be guaranteed without restriction, and then the structural requirement (7) can be omitted. From a practical point of view, it is advantageous to keep $d_0, d_1, \cdots$, and in particular, d_K small, and in this context, we benefit from allowing $K > 0$. Then, the price is the requirement (7).

8.3. Asymptotic Properties of the Bounds. In order to compare our bounds to competing methods, we consider the asymptotic behaviour of the difference between upper and lower bounds.

In order to keep things simple and concise, we focus on the approximation of $\pi f(1) = \psi f(1)/\psi f(0)$ (where $f^{(0)} = \mathbf{1}$) for positive recurrent continuous-time Markov chains, and we assume $d_K = 1$ (which is a clearly restrictive condition), implying $\bar{\varphi} = \underline{\varphi} = \psi_K = 1$. The bounds on $\pi f^{(1)}$ are given by $\underline{F}^{(1)}/\bar{F}^{(0)}$ and $\bar{F}^{(1)}/\underline{F}^{(0)}$, and for $N \longrightarrow \infty$, the difference behaves asymptotically like

$$\begin{aligned}\frac{\bar{F}^{(1)}}{\underline{F}^{(0)}} - \frac{\underline{F}^{(1)}}{\bar{F}^{(0)}} &= \frac{\bar{F}^{(1)} - \underline{F}^{(1)}}{\underline{F}^{(0)}} + \frac{\underline{F}^{(1)}}{\underline{F}^{(0)}} \cdot \frac{\bar{F}^{(0)} - \underline{F}^{(0)}}{\bar{F}^{(0)}} \\ &= \sum_{i=0}^{N} S_i \sum_{j=i+1}^{\infty} P_{ij} g_j^{(1)} + \frac{\bar{F}^{(1)}}{\underline{F}^{(0)}} \cdot \frac{\sum_{i=0}^{N} S_i + P_{ij} g_j^{(0)}}{\bar{F}^{(0)}} \\ &\sim \sum_{i=0}^{N} \pi_i \sum_{j=i+1}^{\infty} P_{ij} g_j^{(1)} \pi f^{(\ell)} \cdot \sum_{i=0}^{N} \pi_i \sum_{j=i+1}^{\infty} P_{ij} g_j^{0},\end{aligned} \tag{84}$$

where we have used that $S_i/\bar{F}^{(0)} = \psi_i/\psi_K \bar{F}^{(0)}$ converges to π_i. In particular, if $g_j^{(1)} = g_j^{(0)}$ (e.g., if $f^{(1)}$ is an indicator function), we have the asymptotic expansion

$$\frac{\bar{F}^{(1)}}{\underline{F}^{(0)}} - \frac{\underline{F}^{(1)}}{\bar{F}^{(0)}} \sim \left(1 + \pi f^{(1)}\right) \sum_{i=0}^{N} \pi_i \sum_{j=1+1}^{\infty} P_{ij} g_j^{(0)}. \tag{85}$$

These considerations clearly motivate to choose $g_j^{(\ell)}$ as small as possible.

8.4. Augmented Matrices versus Nonaugmented Matrices. Theorem 5 and the resulting method for computing bounds on ψf or πf rely on considering the northwest-corner truncation $P(N) = (P_{ij})_{i,j=0}^{N}$ which is substochastic. From $P(N)$, we obtain an approximation $\psi(N)$ to ψ, where $\psi(N)$ is uniquely determined by $\psi_{x_0}(N) = 1$ and by satisfying $\psi(N)P(N) = \psi(N)$ up to the scalar equation corresponding to state x_0.

There are many authors who prefer to "repair" the row sums of $P(N)$, that is, they deal with an augmented stochastic matrix $\tilde{P}(N) \geq P(N)$. Due to finiteness and stochasticity, it is possible to determine an invariant distribution $\tilde{\pi}(N)$ of $\tilde{P}(N)$ directly. Under some mild constraints, $\tilde{\pi}(N)$ converges to π (see, e.g., [3]). The fact that $\tilde{P}(N)$ admits an invariant distribution (in contrast to $P(N)$) can be interpreted as an advantage of methods relying on augmented matrices. There are more theoretical issues why considering augmented matrices is popular, and additionally, the augmented matrices often allow meaningful probabilistic interpretations. A more detailed list of advantages can be found in [3] (for the continuous-time setting).

Despite these undeniable advantages, the approach relying on nonaugmented matrices was chosen in this paper. The simple reason is that $\psi(N)$ converges (pointwise) monotonically to ψ which makes lower bounds on ψf trivial. As we will see in the next lines, this fact is more important than all the advantages of augmentation procedures.

8.5. Comparison to Other A Posteriori Bounds on Truncation Errors. The most recent results for truncation bounds on stationary expectations can be found in [1–3] where the authors consider continuous-time Markov chains. They consider northwest-corner truncations $Q = (Q_{ij})_{i,j=0}^{N}$ and their augmented versions $\tilde{Q}(N)$ (with row sums 0). With our notation and the invariant distribution $\tilde{\pi}(N)$ of $\tilde{Q}(N)$, Theorem 2.1 in [2] reads as

$$\begin{aligned}\sum_{x \in E} |\pi_x - \tilde{\pi}_x(N)| f(x) \leq \left(1 + \frac{\pi f}{\inf f(x)}\right) \cdot \tilde{\pi}(N) \left(\tilde{Q} - Q\right) \\ \cdot \left(g + \frac{b}{\beta \overline{\phi}^{(\beta)}}\right).\end{aligned} \tag{86}$$

The authors require that

(i) the Markov chain is irreducible and positive recurrent

(ii) $f \geq 1$

(iii) $\tilde{Q} = \tilde{Q}(N) = (\tilde{q}_{xy})_{x,y \in E}$ is constructed such that $(\tilde{q}_{xy})_{x,y \in E} 0_{\cup \cdots \cup E} N$ arises from truncation and augmentation and the other entries are arbitrary such that $Q \sim$ is conservative

(iv) e.g., all other entries are 0

(v) g satisfies the f-modulated drift condition $\sum_{y\in E} q_{xy} g(y) \le -f(x)$ for $x \in E_{K+1} \cup E_{K+2} \cup \cdots$

$$b = \max\left\{\sum_{y\in E} q_{xy} g(y) + f(x) : x \in E_0 \cup \cdots E_K\right\} < \infty \tag{87}$$

(vi) $\beta > 0$ is arbitrary

(vii) $\overline{\phi}^{(\beta)} = \sup_{j\in E} \min_{i\in E_0\cup\cdots\cup E_K} \phi_{ij}^{(\beta)}$, where $\Phi^{(\beta)} = (\phi_{ij}^{(\beta)})_{i,j\in E} = \int_0^\infty \beta e^{-\beta t} P^{(t)} dt$ with the transition function $t \mapsto P^{(t)}$ arising from Q.

The right-hand side in (86) has some similarities to the asymptotic expansion (85). Asymptotically, the entries of $\tilde{\pi}(N)(\tilde{Q} - Q)g$ behave like $\sum_{i=0}^{N} \pi_i \sum_{j=i+1}^{\infty} P_{ij} g_j$. Hence, (86) mainly differs by the factor $\inf f(x)$ and the summand $b/\beta\, \overline{\phi}^{(\beta)} 1$. Furthermore, note that the right-hand side of (86) provides a bound on the difference between πf and its approximation. Hence, "upper bound – lower bound" is larger by factor 2.

Finally, it is important to remark that $\Phi^{(\beta)}$ cannot be determined by means of numerical computations. Therefore, it is suggested in [2] to use $(I - 1/\beta \cdot Q(N))^{-1}$ that increases monotonically in N with limit $\Phi^{(\beta)}$. Hence, the entries of $(I - 1/\beta \cdot Q(N))^{-1}$ can be used to find computable bounds on $\overline{\phi}^{(\beta)}$, and thus, on $\sum_{x\in E} |\pi_x - \tilde{\pi}_x(N)| f(x)$.

Replacing $\Phi^{(\beta)}$ in this manner has two effects:

(i) There is no evidence that the bound in (86) is sharp. Even if it was, the bound which results from this replacement cannot be sharp anymore

(ii) Evaluating the computable bound which arises from (86) requires computing at least some rows of $(I - 1/\beta \cdot Q(N))^{-1}$ which can be costly in terms of computational effort.

Together, we see that the algorithm suggested in this paper will provide slightly better bounds which are significantly easier to compute at the same time. Finally, $f \ge 1$ in [1–3] is clearly restrictive. In particular, this requirement prevents us from using indicator functions in order to obtain stationary probabilities.

8.6. Comparison to A Priori Bounds on Truncation Errors. Theorem 5 provides a posteriori truncation bounds (similar to (86) from [2]), since both the approximation to πf and the bounds are computed numerically (simultaneously).

A priori bounds do not rely on numerical computations. Hence, before any computation scheme starts, a priori bounds would allow to determine the truncation level N in such a way that the (relative) error of the approximation is smaller than a prescribed bound.

An approach in this direction was introduced in [5] and generalized in [4] where a method was developed which determines a level N such that

$$\frac{\sum_{i=0}^{N} \psi_i f_i}{\sum_{i=0}^{\infty} \psi_i f_i} \ge 1 - \epsilon\,, \tag{88}$$

for some prescribed $\epsilon > 0$ and some scalar function $f > 0$. The method works as follows:

(i) Choosesome function g and determine (continuous-time setting) $h(x) = \sum_{y\in E} q_{xy} g(y)$ for all $x \in E$.

(ii) Set $C_0 = \{x : h(x) > 0\}$ and $c = \max_{x\in C_0} (h(x)/f(x))$.

(iii) Set $\gamma = c \cdot (1 - \epsilon)/\epsilon$ and $C = \{x \in E : h(x) > -\gamma f(x)\}$.

(iv) If C is finite, we have $\sum_{x\in C} \psi_x f(x) / \sum_{x\in E} \psi_x f(x) \ge 1 - \epsilon$. In particular, we can find N such that $E_0 \cup \cdots \cup E_N \supset C$ and obtain (88).

(v) If C is not finite, we have to change our choice of g.

At a first glance, this method provides a priori error bounds, but there are some problems. In the original papers [4, 5], it was already pointed out that it is not easy to find appropriate functions g, and that very often, the bounds are quite pessimistic. In order to illustrate these effects, we consider the $M/M/1$ queue from Section 5 and set $f = \mathbf{1}$. Whereas Theorem 5 can be applied with $g(i) = b \cdot i$ for an appropriate constant b, this choice will fail here: We obtain $h(0) = b\lambda$ and $h(i) = b(\lambda - \mu)$ for $i \ge 1$, that is, $h(i)$ is constant for $i \ge 1$. In particular, if $\epsilon > 0$ is very small (and hence, γ is large), we have $C = \{i \in E : h(i) > -\gamma f(i)\} = \mathbb{N}_0$. Note that the choice of b is unimportant since it cancels out.

The next intuitive choice is $g(i) = i^2$, resulting in $h(0) = \lambda$ and $h(i) = 2(\lambda - \mu)i + \lambda + \mu$. If $\lambda < \mu < 2\lambda$, we have $c = h(1) = 3\lambda - \mu$, $\gamma = (3\lambda - \mu) \cdot (1 - \epsilon)/\epsilon$, and

$$\begin{aligned} C &= \left\{i : 2(\lambda - \mu)i + \lambda + \mu > -(3\lambda - \mu) \cdot \frac{(1 - \epsilon)}{\epsilon}\right\} \\ &= \left\{0, 1, \cdots, \left\lceil \frac{3\lambda - \mu}{2(\mu - \lambda)} \cdot \frac{1}{\epsilon} \right\rceil\right\}. \end{aligned} \tag{89}$$

For $\lambda/\mu = 2/3$, we obtain $C = \{0, 1, \cdots, N\}$ where $N = \lceil 3/2 \cdot 1/\epsilon \rceil$. Hence, the method guarantees that $\sum_{i=0}^{N} \pi_i \ge 1 - \epsilon$ whereas the exact value is $\sum_{i=0}^{N} \pi_i = 1 - (2/3)^N$. For $\epsilon \longrightarrow 0$ and $N = N(\epsilon)$, the latter term converges to 1 much faster than $1 - \epsilon$. For example, for $\epsilon = 0.05$, we have $N = 30$, and hence, $1 - (2/3)^N \approx 1 - 5.215 \cdot 10^{-6}$.

Summarizing this brief numerical example, the most intuitive choice for g does not work, and the next choice provides very pessimistic estimates. It might be possible to find functions g which provide sharp estimates, but it does not seem likely that such functions can be "guessed."

There is an even more serious problem when interpreting the results from [4, 5] as a priori estimates for the

truncation error: The numerator in (88) refers to the exact invariant measure ψ which cannot be computed by numerical means (up to some trivial exceptions like the $M/M/1$ queue). For honest a priori bounds, we should be interested in guaranteeing

$$\frac{\sum_{i=0}^{N}\psi_i(N)f_i}{\sum_{i=0}^{\infty}\psi_i f_i} \geq 1-\epsilon, \tag{90}$$

with a numerically calculated approximation $\psi(N)$ to ψ. Unfortunately, the methods in [4, 5] cannot be modified in this direction. One might argue that this slight modification is negligible in view of the pessimistic bounds which are obtained. However, if we guessed an optimal function g by accident, we would not have a bound on $\sum_{i=0}^{N}\psi_i(N)f_i/\sum_{i=0}^{\infty}\psi_i f_i$ anymore. In total, the results from [4, 5] may be exploited to find theoretical estimates on stationary probabilities or expectations. With respect to numerical means, they can only be used for rough preliminary considerations (how large should the truncation level N be chosen) before applying the method suggested in this paper. Even that might be hard due to the increased difficulty to find appropriate functions g (in comparison to finding appropriate functions g for applying Theorem 5).

9. Conclusion and Further Research

In this paper, we have developed an exact and efficient method for numerically computing lower and upper bounds on ψf where ψ is an invariant measure for an irreducible recurrent discrete-time or continuous-time QBD and f is a multidimensional nonnegative cost/reward function. By combining the lower and upper bounds appropriately, we obtain lower and upper bounds on πf in the case of positive recurrence, where π is the unique invariant distribution and f is an arbitrary cost/reward function. The term πf can be interpreted as the stationary expectation of the Markov chain measured by f, or as the long-run average of cost/rewards measured by f. If $f = \mathbf{1}_A$ is chosen to be an indicator function, we obtain stationary probabilities. The computation of $\pi f^{(2)}$ in Section 6 illustrates that even very small probabilities can be computed with high relative precision.

The algorithm was developed as an extension of previous matrix-analytic methods which also dealt algorithmically with level-dependent QBDs, but for which no error bounds were known up to now. Due to the fact that we compute lower and upper error bounds, we obtain automatically an approximation to πf and (a posteriori) error bounds. In our examples, we have seen that the difference between lower and upper bounds converges to 0 very fast, that is, we obtain very precise results. For optimal choices of $g^{(\ell)}$, we have observed that the upper bound coincides with the true value of ψf.

Obvious directions of future publications concern detailed applications of our algorithm to specific Markov models, e.g., we could consider enhanced retrial queuing models with Markovian arrival processes, phase-type distributed service times, impatient customers, etc., but there is a huge variety of other practical processes which are modelled by infinite LDQBDs. If we focus on the analysis of a specific model, we should spend effort into finding very good functions g which may result in obtaining very precise results for small truncation levels.

In this context, it may be worth to find not only an upper truncation level but also a lower one. Then, the numerical computation would focus on a very small region of the state space, and the method would become even more efficient.

Note that the main results of this paper are not restricted to QBDs. Theorem 5 only requires that no transitions from E_i to E_j occur with positive probability/rate if $i > K > j$. Hence, these results also apply if P or Q, respectively, has upper block-Hessenberg structure (also referred to as the block-$M/G/1$ structure). We have focused on QBDs here since this specialized structure allows a very efficient computation of lower and upper bounds. Future research could deal with developing efficient algorithms for more general structures.

We remark that some of the key results also transfer to nondiscrete state spaces: Let $E = E_0 \cup E_1 \cup \cdots$, and let the levels E_i be petite in the sense of [10] (this property is satisfied by compact sets for many Markov chains). Then P_{ij} becomes a (sub-Markovian) kernel, and S_i does so, too. The multiplication S_jP_{ji} is defined straightforward by $S_jP_{ji}(x, dz) = \int_y S_j(x, dy)P_{ji}(y, dz)$. We still have $\psi f = \psi_K F$, where ψ_K solves $\psi_K \lim_{N\to\infty}(I - T) = 0$, and

$$F = \lim_{N\to\infty}\left(f_K + \sum_{\substack{i=0\\ i\neq K}}^{N} S_i f_i\right), \tag{91}$$

where $S_i f_i^{(\ell)}(x) = \int_y S_i(s, dy) f_i^{(\ell)}(y)$. Under the structural assumptions (no transitions from E_i to E_j for $i > K > j$), the results on the bounds on F remain unchanged. However, for finding $S_0, \cdots, S_N$ by means of numerical calculations (and for finding a lower and an upper bound on ψ_K), it seems that some sort of discretization is inevitable.

We conclude with a remark concerning the relationship to nonprobabilistic topics: As mentioned above, for $K = 0$, the basic algorithm for computing approximations to the subvectors ψ_i of the invariant measure ψ can be deduced from general considerations concerning (matrix-valued) continued fractions. Hence, in some sense, our computation of bounds on ψf is related to the speed-of-convergence statements for matrix-valued continued fractions. Note that the probabilistic interpretation of continued fractions (arising in nonprobabilistic contexts) and their convergents has led to new convergence criteria for continued fractions and their generalizations (see [25, 26]), and with further research, considerations similar to those in this paper might provide speed-of-convergence results.

Appendix

A. Alternative Algebraic Proof of the Upper Bound on F

Additionally, we give a less stochastic, more algebraic proof of the key result, that is, the upper bound in Lemma 4, where we concentrate on discrete setting. Let $P_{ab}, P_{ac}, \cdots$ be blocks with transition probabilities where the index

(i) a corresponds to states within E_K

(ii) b corresponds to states within $E_0 \cup \cdots \cup E_{K-1}$

(iii) c corresponds to states within $E_{K+1} \cup \cdots \cup E_N$

(iv) d corresponds to states within $E_{N+1} \cup E_{N+2} \cup \cdots$.

Then, the structural requirement (7) on the transitions reads as $P_{db} = 0$ and $P_{cb} = 0$. The f-modulated drift condition (5) can be rewritten as

$$P_{cc}g_c + P_{cd}g_d + fc \le g_c, \tag{A.1}$$

$$P_{dc}g_c + P_{dd}g_d + f_d \le g_d. \tag{A.2}$$

Let us set

$$S_a := P_{aa} + (P_{ab}, P_{ac}, P_{ad})V^{-1}\begin{pmatrix} P_{ba} \\ P_{ca} \\ P_{da} \end{pmatrix}, \tag{A.3}$$

where

$$V = \begin{pmatrix} I - P_{bb} & -P_{bc} & -P_{bd} \\ -P_{cb} & I - P_{cc} & -P_{cd} \\ -P_{db} & -P_{dc} & I - P_{dd} \end{pmatrix}. \tag{A.4}$$

S_a is stochastic, and the property of irreducibility is inherited to S_a from P. Let the blocks of an invariant measure for P be denoted by $\psi_a, \psi_b, \psi_c, \psi_d$. Then

$$(\psi_b, \psi_c, \psi_d) = \psi_a(P_{ab}, P_{ac}, P_{ad})V^{-1}, \tag{A.5}$$

and ψ_a is an invariant measure for S_a. These formulas can be proven in a purely algebraic manner although they have a stochastic interpretation: If the original Markov chain with transition probability matrix P is only observed at return times to set E_K, we obtain a new Markov chain with transition probability matrix S_a (see [27]). The above representation of ψ allows to write

$$\psi f = \psi_a\left(f_a + (P_{ab}, P_{ac}, P_{ad})V^{-1}\begin{pmatrix} f_b \\ f_c \\ f_d \end{pmatrix}\right) =: \psi_a \cdot F, \tag{A.6}$$

and due to $\psi_a = \psi_K$, F has the same meaning as before. As pointed out above, we do not want to give alternative proofs for all results here. Instead, we focus on the upper bound $\bar{F}$ on F given in Lemma 4.

From (A.1), we find

$$g_c \ge (I - P_{cc})^{-1}(P_{cd}g_d + f_c), \tag{A.7}$$

and with (A.2), we obtain

$$\begin{aligned} g_d &\ge \left(I - P_{dd} - P_{dc}(I - P_{cc})^{-1}P_{cd}\right)^{-1}\left(P_{dc}(I - P_{cc})^{-1}f_c + f_d\right) \\ &=: T^{-1}\left(P_{dc}(I - P_{cc})^{-1}f_c + f_d\right). \end{aligned} \tag{A.8}$$

Therefore, we can write

$$\begin{aligned} \bar{F} &= \bar{f}_a + (P_{ab}, P_{ac})\tilde{V}^{-1}\begin{pmatrix} \bar{f}_b \\ \bar{f}_c \end{pmatrix} = f_a + P_{ad}g_d + (P_{ab}, P_{ac})\tilde{V}^{-1}\begin{pmatrix} f_b + P_{bd}g_d \\ f_c + P_{cd}g_d \end{pmatrix} \\ &\ge f_a + (P_{ab}, P_{ac}, P_{ad})\begin{pmatrix} \tilde{V}^{-1} + \tilde{V}^{-1}\begin{pmatrix} P_{bd} \\ P_{cd} \end{pmatrix}T^{-1}\left(0, P_{dc}(I - P_{cc})^{-1}\right) & \tilde{V}^{-1}\begin{pmatrix} P_{bd} \\ P_{cd} \end{pmatrix}T^{-1} \\ T^{-1}\left(0, P_{dc}(I - P_{cc})^{-1}\right) & T^{-1} \end{pmatrix}\begin{pmatrix} f_b \\ f_c \\ f_d \end{pmatrix}. \end{aligned} \tag{A.9}$$

In the next lines, we will prove that the middle-term matrix is V^{-1} which obviously yields the desired statement $\bar{F} \ge F$. We have

$$
\begin{pmatrix}
\tilde{V}^{-1} + \tilde{V}^{-1}\begin{pmatrix} P_{bd} \\ P_{cd} \end{pmatrix} T^{-1}\left(0, P_{dc}(I-P_{cc})^{-1}\right) & \tilde{V}^{-1}\begin{pmatrix} P_{bd} \\ P_{cd} \end{pmatrix} T^{-1} \\
T^{-1}\left(0, P_{dc}(I-P_{cc})^{-1}\right) & T^{-1}
\end{pmatrix} V
$$

$$
= \begin{pmatrix}
\tilde{V}^{-1} + \tilde{V}^{-1}\begin{pmatrix} P_{bd} \\ P_{cd} \end{pmatrix} T^{-1}\left(0, P_{dc}(I-P_{cc})^{-1}\right) & \tilde{V}^{-1}\begin{pmatrix} P_{bd} \\ P_{cd} \end{pmatrix} T^{-1} \\
T^{-1}\left(0, P_{dc}(I-P_{cc})^{-1}\right) & T^{-1}
\end{pmatrix}
\begin{pmatrix}
\tilde{V} & \begin{pmatrix} -P_{bd} \\ -P_{cd} \end{pmatrix} \\
(-P_{db}, -P_{dc}) & I - P_{dd}
\end{pmatrix}
$$

$$
= \begin{pmatrix}
I + \tilde{V}^{-1}\begin{pmatrix} P_{bd} \\ P_{cd} \end{pmatrix} T^{-1}\left(-P_{dc}(I-P_{cc})^{-1}P_{cb}, P_{dc}\right) + \tilde{V}^{-1}\begin{pmatrix} P_{bd} \\ P_{cd} \end{pmatrix} T^{-1}(-P_{db}, -P_{dc}) & -\tilde{V}^{-1}\begin{pmatrix} P_{bd} \\ P_{cd} \end{pmatrix} - T^{-1}P_{dc}(I-P_{cc})^{-1} \\
T^{-1}\left(-P_{dc}(I-P_{cc})^{-1}P_{cb}, P_{dc}\right) - T^{-1}(P_{db}, P_{dc}) & -T^{-1}P_{dc}(I-P_{cc})^{-1}P_{cd} + T
\end{pmatrix}
$$

$$
= \begin{pmatrix}
I + \tilde{V}^{-1}\begin{pmatrix} P_{bd} \\ P_{cd} \end{pmatrix} T^{-1}\left(-P_{dc}(I-P_{cc})^{-1}P_{cb} - P_{dc}, 0\right) & -\tilde{V}^{-1}\begin{pmatrix} P_{bd} \\ P_{cd} \end{pmatrix} + \tilde{V}^{-1}\begin{pmatrix} P_{bd} \\ P_{cd} \end{pmatrix} T^{-1}T \\
T^{-1}\left(-P_{dc}(I-P_{cc})^{-1}P_{cb} - P_{dc}, 0\right) & T^{-1}T
\end{pmatrix} = \begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix},
\tag{A.10}
$$

where we used the assumptions $P_{cb} = 0, P_{db} = 0$.

B. Remarks Concerning Avoidance of Numerical Problems

B.1. On Finding ψ_K. In the discrete-time setting of Theorem 5, the matrix T converges to a stochastic matrix as $N \longrightarrow \infty$. Hence, for relatively small truncation levels N, we can choose $A = (I - T)^{-1}$ and compute this matrix inverse by standard methods. For large N (when we want to obtain precise results), the inversion of $I - T$ is ill-conditioned. This problem can be avoided easily with a little additional effort: Remember that for an appropriate choice of A, the xth row of A is the minimal subinvariant measure for the transient substochastic matrix T subject to $a_{xx} = 1$. Hence, we fix $a_{xx} = 1$ and solve the inhomogeneous system of equations $a_{xy} = \sum_{z \in E_K} a_{xz} t_{zy}$ for $y \neq x$, where $T = (t_{zy})_{z,y \in E_K}$. With $\tilde{T} = (t_{zy})_{z,y \in E_K \setminus (x)}$ and the row vectors $\tilde{t} = (t_{xy})_{y \in E_K \{x\}}$, $\tilde{a} = (a_{xy})_{y \in E_K \setminus \{x\}}$, this system reads as $\tilde{a} = \tilde{t} + \tilde{a}\tilde{T}$.

As $N \longrightarrow \infty$, T converges to an irreducible stochastic matrix, and hence, $\tilde{T}$ converges to a transient substochastic matrix. Hence, $I - \tilde{T}$ is invertible, and this new system of equations is not susceptible to numerical problems. Of course, we have to find all rows of A separately, that is, we solve d_K systems of linear equations of the dimension $(d_K - 1) \times (d_K - 1)$. However, if d_K is relatively small, this effort can be neglected in comparison to the other computational steps. These considerations directly transfer to the continuous-time setting.

B.2. Exploiting the GTH Advantage in the Update Step for the Matrix B. For many systems of linear equations arising in the context of the Markov chains, the coefficient matrix satisfies some row sum conditions which is preserved during all steps of solving the linear system of equations. This fact can be used for implementing these steps in a way which is not affected by numerical instability. Such procedures were introduced by Grassmann et al. [28] and therefore referred to as *GTH advantage*. In the present paper, this advantage should be used in the update step for the matrix B. We give some more details.

First, note that the probabilistic interpretation guarantees that all entries of B_i are nonnegative for all i. Hence, computing the diagonal entries of $Q_{i-1,i-1} + B_{i-1}Q_{i-2,i-1}$ results in computing differences of positive numbers. Furthermore, when performing any elimination procedure, the updates of the diagonal entries result in computing differences of positive numbers again. All other operations only compute sums of nonnegative numbers (which cannot cause numerical problems). For avoiding the need of subtractions, we remark that the row sums of $Q_{i-1,i-1} + B_{i-1}Q_{i-2,i-1}$ and of $Q_{i-1,i}$ add up to 0 (this statement can be deduced from the probabilistic interpretation). Let h_{i-1} be the column vector of row sums of $Q_{i-1,i}$. Then the $d_{i-1} \times (d_{i-1} + 1)$-dimensional matrix $(-Q_{i-1,i-1} - B_{i-1}Q_{i-2,i-1}, -h_{i-1})$ has row sums of 0 where only the diagonal entries can be positive. An important finding on the application of (scaled) Gaussian elimination to systems with coefficient matrices of this type guarantees that these properties are preserved in each step of the elimination procedure. Hence, the pivot element in each step (which are the entries which might be affected by numerical instabilities) can be "corrected" by computing the negative sum of the

other entries in the same rows. We demonstrate this by means of a simple example.

(i) Let

$$-Q_{i-1,i-1} - B_{i-1}Q_{i-2,i-1} = \begin{pmatrix} 4 & -2 & -1 \\ -2 & 3 & 0 \\ 0 & -2 & 4 \end{pmatrix} \text{ with } h_{i-1} = \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix} \tag{B.1}$$

We want to compute $(-Q_{i-1,i-1} - B_{i-1}Q_{i-2,i-1})^{-1}$.

(ii) First, write

$$\left(\begin{array}{ccc|c|ccc} 4 & -2 & -1 & -1 & 1 & 0 & 0 \\ -2 & 3 & 0 & -1 & 0 & 1 & 0 \\ 0 & -2 & 4 & -2 & 0 & 0 & 1 \end{array}\right). \tag{B.2}$$

By performing the steps of Gaussian elimination, we convert the left-hand side to the unity matrix, and the right-hand side becomes the desired inverse matrix. The column in between acts as a control column. Before starting the Gaussian elimination, we can correct the first pivot element 4 by recomputing it as $-(-2\ -1\ -1)$.

(iv) The first row is scaled by 1/4. Afterwards a constant multiple (with factor 2) of the first row is added to the second row. This results in

$$\left(\begin{array}{ccc|c|ccc} 1 & -1/2 & -1/4 & -1/4 & 1/4 & 0 & 0 \\ 0 & 2 & -1/2 & -3/2 & 1/2 & 1 & 0 \\ 0 & -2 & 4 & -2 & 0 & 0 & 1 \end{array}\right). \tag{B.3}$$

The submatrix still has

$$\begin{pmatrix} 2 & -1/2 & -3/2 \\ -2 & 4 & -2 \end{pmatrix} \tag{B.4}$$

row sums of 0. The upper left entry 2 can be corrected by replacing it by $-(-1/2 - 3/2)$ before applying the next step.

(v) After the next step, we obtain

$$\left(\begin{array}{ccc|c|ccc} 1 & 0 & -3/8 & -5/8 & 3/8 & 1/4 & 0 \\ 0 & 1 & -1/4 & -3/4 & 1/4 & 1/2 & 0 \\ 0 & 0 & 7/2 & -7/2 & 1/2 & 1 & 1 \end{array}\right). \tag{B.5}$$

Again, $(7/2, -7/2)$ has a row sum of 0 which can be used for correcting the pivot element 7/2 which is used for the last step of the elimination procedure.

Due to the need of using the control vector h, and due to computing the inverse first instead of solving a matrix-valued system of linear equations directly, we have slightly more computational effort. The effects which are caused by instability justify this additional effort.

As an example for instability, consider the update procedure of B in the retrial queueing model in Section 7. In fact, note that for this specific model, we could even simplify our algorithmic approach since we are able to determine $M_i := -Q_{ii} - B_i Q_{i-1,i}$ explicitly. Hence, there is no need to compute $B_i = Q_{i,i-1}M_{i-1}^{-1}$ recursively. This explicit representation is given by

$$M_i = \begin{pmatrix} \lambda + \nu i & -\lambda & & & & & & & -\nu i \\ -\mu & \lambda + c\mu + \nu i & -\lambda & & & & & & -\nu i \\ & -2\mu & \lambda + 2\mu + \nu i & -\lambda & & & & & -\nu i \\ & & \ddots & \ddots & \ddots & & & & \vdots \\ & & & -c\mu & \lambda + c\mu + \nu i & -\lambda & & & -\nu i \\ & & & & \ddots & \ddots & \ddots & & \vdots \\ & & & & & -c\mu & \lambda + c\mu + \nu i & -\lambda - \nu i \\ & & & & & & -c\mu & \lambda + c\mu \end{pmatrix}, \tag{B.6}$$

which can be proven by induction. The key ideas for the induction step are the recurrence relation $M_i = -Q_{ii} - Q_{i,i-1} M_{i-1}^{-1} Q_{i-1,i}$ and the observation that the last column of M_i^{-1} is given by $1/\lambda \cdot \mathbf{1}$; we omit further details.

Instead, we use this explicit representation to evaluate numerically computed values of B_i and M_i and to demonstrate the effect of the GTH advantage. Let $\lambda = \mu = 1.0$, $c = d - 1 = 5$, and $\nu = 0.05$. If we apply some ordinary elimination procedures to compute the values of B_i (and give out the values of M_i as a test), and if we use double precision in C++, the first four decimal places of the entries of M_i are "correct" for $i \leq 16$. The first four entries in the last column of M_{17} are −0.8501 instead of −0.8500, the first four entries in the last column of M_{18} are −0.9003 instead of −0.9000, and the first four entries in the last column of M_{19} are −0.9517 instead of −0.9500. Obviously, the error of these values increases drastically. The corresponding entries in M_{23} have values between −8.9 and −9.4 (instead of −1.15), and they differ. For larger indices, some of these entries assume even positive values. Unsurprisingly, the resulting values for lower and upper bound on ψf are completely wrong. When using the GTH advantage, this effect does not occur, and all matrices M_i coincide with the values which can be obtained from the explicit representation.

Data Availability

No data were used to support this study.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The author acknowledges support by Open Access Publishing Fund of Clausthal University of Technology.

References

[1] Y. Liu and W. Li, "Error bounds for augmented truncation approximations of Markov chains via the perturbation method," *Advances in Applied Probability*, vol. 50, no. 2, pp. 645–669, 2018.

[2] Y. Liu, W. Li, and H. Masuyama, "Error bounds for augmented truncation approximations of continuous-time Markov chains," *Operations Research Letters*, vol. 46, no. 4, pp. 409–413, 2018.

[3] H. Masuyama, "Error bounds for last-column-block-augmented truncations of blockstructured Markov chains," *Journal of the Operations Research Society of Japan*, vol. 60, no. 3, pp. 271–320, 2017.

[4] H. Baumann and W. Sandmann, "Bounded truncation error for long-run averages in infinite Markov chains," *Journal of Applied Probability*, vol. 52, no. 3, pp. 609–621, 2015.

[5] T. Dayar, W. Sandmann, D. Spieler, and V. Wolf, "Infinite level–dependent QBD processes and matrix-analytic solutions for stochastic chemical kinetics," *Advances in Applied Probability*, vol. 43, no. 4, pp. 1005–1026, 2011.

[6] K. L. Chung, *Markov Chains with Stationary Transition Probabilities*, Springer-Verlag, 1960.

[7] S. Karlin and H. M. Taylor, *A First Course in Stochastic Processes (Second Edition)*, Academic Press, 1975.

[8] H. Baumann and W. Sandmann, "On finite long run costs and rewards in infinite markov chains," *Statistics & Probability Letters*, vol. 91, pp. 41–46, 2014.

[9] F. G. Foster, "On the stochastic matrices associated with certain Queuing processes," *The Annals of Mathematical Statistics*, vol. 24, no. 3, pp. 355–360, 1953.

[10] S. P. Meyn and R. L. Tweedie, *Markov Chains and Stochastic Stability*, Springer, London, 1993.

[11] A. G. Pakes, "Some conditions for ergodicity and recurrence of Markov chains," *Operations Research*, vol. 17, no. 6, pp. 1058–1061, 1969.

[12] R. L. Tweedie, "Sufficient conditions for regularity, recurrence and ergodicity of Markov processes," *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 78, no. 1, pp. 125–136, 1975.

[13] R. L. Tweedie, "The existence of moments for stationary Markov chains," *Journal of Applied Probability*, vol. 20, no. 1, pp. 191–196, 1983.

[14] E. Seneta, *Non-negative Matrices and Markov Chains*, Springer, 1981.

[15] R. L. Tweedie, "Truncation procedures for non-negative matrices," *Journal of Applied Probability*, vol. 8, no. 2, pp. 311–320, 1971.

[16] H. Baumann and T. Hanschke, "Inherent numerical instability in computing invariant measures of Markov chains," *Applied Mathematics*, vol. 8, no. 9, pp. 1367–1385, 2017.

[17] L. Bright and P. G. Taylor, "Calculating the equilibrium distribution in level dependent quasi-birth-and-death processes," *Communications in Statistics Stochastic Models*, vol. 11, no. 3, pp. 497–525, 1995.

[18] M. F. Neuts, *Matrix-Geometric Solutions in Stochastic Models*, The John Hopkins University Press, Baltimore, MD, USA, 1981.

[19] T. Hanschke, "A matrix continued fraction algorithm for the multiserver repeated order queue," *Mathematical and Computer Modelling*, vol. 30, no. 3-4, pp. 159–170, 1999.

[20] H. Baumann and W. Sandmann, "Numerical solution of level dependent quasi-birth-and-death processes," *Procedia Computer Science*, vol. 1, no. 1, pp. 1561–1569, 2010.

[21] T. Phung-Duc, H. Masuyama, S. Kasahara, and Y. Takahashi, "A simple algorithm for the rate matrices of level-dependent QBD processes," in *Proceedings of the 5th International Conference on Queueing Theory and Network Applications - QTNA '10*, pp. 46–52, New York, NY, USA, 2010.

[22] H. Baumann and W. Sandmann, "Computing stationary expectations in level-dependent qbd processes," *Journal of Applied Probability*, vol. 50, no. 1, pp. 151–165, 2013.

[23] Q. M. He, *Fundamentals of Matrix-Analytic Methods*, Springer, New York, NY, USA, 2015.

[24] J. R. Artalejo and A. Gómez-Corral, *Retrial Queueing Systems*, Springer, 2008.

[25] H. Baumann, "Two-sided continued fractions in Banach algebras- A Śleszyński-Pringsheim- type convergence criterion and applications," *Journal of Approximation Theory*, vol. 199, pp. 13–28, 2015.

[26] H. Baumann, "Generalized continued fractions: a unified definition and a Pringsheim-type convergence criterion,"

Advances in Difference Equations, vol. 2019, no. 1, Article ID 406, 2019.

[27] Y. Q. Zhao and D. Liu, "The censored Markov chain and the best augmentation," *Journal of Applied Probability*, vol. 33, no. 3, pp. 623–629, 1996.

[28] W. K. Grassmann, M. I. Taksar, and D. P. Heyman, "Regenerative analysis and steady state distributions for Markov chains," *Operations Research*, vol. 33, no. 5, pp. 1107–1116, 1985.