Probability in the Engineering and Informational Sciences, **21**, 2007, 381–400. Printed in the U.S.A.
DOI: 10.1017/S0269964807000034

SERIES EXPANSIONS FOR FINITE-STATE MARKOV CHAINS

BERND HEIDERGOTT
Department of Econometrics
Vrije Universiteit Amsterdam
1081 HV Amsterdam, The Netherlands
E-mail: bheidergott@feweb.vu.nl

ARIE HORDIJK
Mathematical Institute
Leiden University
2300 RA Leiden, The Netherlands
E-mail: hordijk@math.leidenuniv.nl

MIRANDA VAN UITERT
Department of Econometrics
Vrije Universiteit Amsterdam
1081 HV Amsterdam, The Netherlands
E-mail: muitert@feweb.vu.nl

This article provides series expansions of the stationary distribution of a finite Markov chain. This leads to an efficient numerical algorithm for computing the stationary distribution of a finite Markov chain. Numerical examples are given to illustrate the performance of the algorithm.

1. INTRODUCTION

Let P denote the transition kernel of a Markov chain defined on a finite state space $\mathcal{S}$ having unique stationary distribution π_P. For example, think of P as the embedded jump chain of an M/M/1/c queue, where c is a finite number denoting the buffer capacity of the queue. What would be the effect on the stationary behavior of the queue if, for example, we increase the buffer capacity or adjusted the service rate of the queue? Let Q denote the Markov transition kernel of the Markov chain modeling the alternative system and assume that Q has unique stationary distribution

© 2007 Cambridge University Press 0269-9648/07 $25.00

https://doi.org/10.1017/S0269964807000034 Published online by Cambridge University Press

π_Q. The question about the effect of switching from P to Q on the stationary behavior is expressed by $\pi_P - \pi_Q$, the difference between the stationary distributions. Let $\|\cdot\|_{\text{tv}}$ denote the total variation norm; then the above problem can be phrased as follows: Can $\|\pi_P - \pi_Q\|_{\text{tv}}$ be approximated or bounded in terms of $\|P - Q\|_{\text{tv}}$? This is known as *perturbation analysis of Markov chains* (PAMC) in the literature.

In this article we will show that $\pi_P - \pi_Q$ can be arbitrarily closely approximated by a polynomial in $(Q - P)D_P$, where D_P denotes the deviation matrix associated with P; a precise definition will be given later. The starting point is the representation

$$\pi_Q = \sum_{n=0}^{k} \pi_P((Q - P)D_P)^n + \pi_Q((Q - P)D_P)^{k+1} \tag{1}$$

for any $k \geq 0$, a proof of which will be given in Section 3.1. This series expansion of π_Q provides the means of approximating π_Q by Q and entities given via the P Markov chain only. In order to obtain a bound for the remainder term, we propose to work with the weighted supremum norm, denoted by $\|\cdot\|_v$, where v is some vector with positive nonzero elements, and for any $\omega \in \mathbb{R}^{\mathcal{S}}$,

$$\|w\|_v \stackrel{\text{def}}{=} \sup_{i \in \mathcal{S}} \frac{|w(i)|}{v(i)}; \tag{2}$$

see, for example, [8]. We will show that

$$\pi_Q(s) - \left(\sum_{n=0}^{k} \pi_P((Q - P)D_P)^n \right)(s) \leq d\|(Q - P)D_P\|_v^{k+1}$$

for any $k \in \mathbb{N}$ and any $s \in \mathcal{S}$, where v can be any vector satisfying $v(s) \geq 1$ for $s \in \mathcal{S}$ and d is some finite computable constant. In particular, the above error bound can be computed without knowledge of π_Q. The key idea of our approach is to solve, for all k, the optimization problem

$$\begin{aligned} &\min \|((Q - P)D_P)^k\|_v \\ &\text{subject to} \\ &\qquad v(s) \geq 1 \quad \text{for } s \in \mathcal{S}. \end{aligned} \tag{3}$$

Surprisingly enough, it will turn out that the solution v^* of the above optimization problem is independent of k. The vector v^* thus yields the optimal measure of the rate of convergence of the series in (1). Moreover, the series in (1) tends to converge extremely fast due to the fact that in many examples, v^* can be found such that $\|(Q - P)D_P\|_{v^*} \ll 1$. To the best of our knowledge, the limit of the series in (1) first appeared in [2]; however, neither upper bounds for the remainder term nor numerical examples were given there.

The use of series expansions for computational purposes is not new. It has been used in various fields for different purposes—for instance, in the field of linear algebra [3]. In [1] the authors derived the power series for the stationary distribution in a slightly

https://doi.org/10.1017/S0269964807000034 Published online by Cambridge University Press

different way than [6] and the approach we take in this article, but they do not use it for real problems or real computations. The novelty of our work is the fact that we combine the ideas obtained and extend them such that problems can be really solved.

The series expansion for the finite-state Markov chains in this article is derived in a very elegant manner using the Poisson equation. We remark that this series expansion holds in a very general format under proper conditions. The derivation of this has been done in [6], which is a generalization of [2].

The work presented in this article is part of a major research project on numerical algorithms based on series expansions of Markov chains. The present article establishes the main theoretical results. In a follow-up article, we will study our methodology for large-scale problems.

The article is organized as follows. Section 2 presents basic facts on finite Markov chains. The series expansion (1) is discussed in Section 3. In particular, numerical examples are provided. Section 4 presents the extension to convex combinations of Markov chains.

2. PRELIMINARIES ON FINITE MARKOV CHAINS

Let $\mathcal{S}$ denote a finite set with $0 < S < \infty$ elements. For notational convenience we will identify $\mathcal{S}$ with the set $\{1, \ldots, S\}$. We consider Markov kernels on state space $\mathcal{S}$. Such a Markov kernel, say P, can be written as a square matrix $P \in [0,1]^{S\times S}$. Element (i, j) of P is denoted by $P(i, j)$ and represents the probability of jump from state i to state j, which implies $\sum_{j\in\mathcal{S}} P(i, j) = 1$, for all $i \in \mathcal{S}$. The probability of going from state i to state j in n steps is denoted by $P^n(i, j)$, where the Markov kernel P^n is simply obtained from taking the nth power of P. Provided it exists, we denote the unique stationary distribution of P by π_P and its ergodic projector by Π_P; that is, for any distribution μ, it holds that $\mu\Pi_P = \pi_P$. In order to simplify the notation, we will—with a slight abuse of notation—identify π_P with Π_P.

Throughout the article we assume that P is aperiodic and unichain, which means that there is one closed irreducible set of states and a (possibly empty) set of transient states. We write $|A|(i, j)$ to denote the (i, j)th element of the matrix of absolute values of $A \in \mathbb{R}^{S\times S}$; additionally, we use the notation $|A|$ for the matrix of absolute values of A.

The main tool for our analysis is the weighted supremum norm, also called the v-norm, as defined in (2). For a matrix $A \in \mathbb{R}^{S\times S}$, the v-norm is given by

$$\|A\|_v \stackrel{\text{def}}{=} \sup_{i,\|w\|_v\le 1} \frac{\sum_{j=1}^{S} |A(i,j)w(j)|}{v(i)}.$$

Observe that

$$\|A\|_v = \sup_i \frac{\sum_{j=1}^{S} |A|(i,j)v(j)}{v(i)},$$

https://doi.org/10.1017/S0269964807000034 Published online by Cambridge University Press

which is due to $\|v\|_v = 1$ and the fact that $\|w\|_v \leq 1$ implies $|w(i)| \leq v(i)$ for $i \in \mathcal{S}$ Obviously, this implies

$$v(i)\|A\|_v \geq \sum_{j=1}^{S} |A|(i,j)v(j), \qquad i \in \mathcal{S},$$

and hence, using $v(i) \geq 1$, $i \in \mathcal{S}$, we obtain

$$\max_{j \in \mathcal{S}} |A|(i,j) \leq \|A\|_v v(i), \qquad i \in \mathcal{S}. \tag{4}$$

From (4) it readily follows that v-norm convergence to zero implies elementwise convergence to zero. More precisely, let $\{A(n) \in \mathbb{R}^{S \times S}, n = 1, 2, \ldots\}$ be given such that $\lim_{n\to\infty} \|A(n)\|_v = 0$ exists; then $\lim_{n\to\infty} |A(n)|(i,j) = 0$ exists $\forall i, j \in \mathcal{S}$. Next, we introduce v-geometric ergodicity (also called v-normed ergodicity) of P; see [12] for details.

Definition 2.1: *A Markov chain P is v-geometric ergodic if $c < \infty$, $\beta < 1$ and $N < \infty$ exist such that*

$$\|P^n - \Pi_P\|_v \leq c\beta^n$$

for all $n \geq N$.

The following lemma shows that any finite-state aperiodic Markov chain is v-geometric ergodic.

Lemma 2.2: *For finite-state and aperiodic P, a finite number N exists such that*

$$\|P^n - \Pi_P\|_v \leq c\beta^n$$

for all $n \geq N$, where $c < \infty$ and $\beta < 1$.

Proof: Because of the finite state space and aperiodicity,

$$\lim_{n\to\infty} P^n(i,j) = \Pi_P(i,j), \quad i,j \in \mathcal{S}.$$

Moreover, it is possible to take N sufficiently large such that

$$\forall i \in \mathcal{S}: \quad \frac{\sum_{j=1}^{S} |P^N - \Pi_P|(i,j)v(j)}{v(i)} < \epsilon,$$

with $\epsilon < 1$. Because $P^n - \Pi_P = (P - \Pi_P)^n$ for $n \geq 1$, this means that

$$\|(P - \Pi_P)^N\|_v < \epsilon. \tag{5}$$

For $n > N$, taking appropriate integers n, k, and l such that $n = kN + l$, we find

$$\|P^n - \Pi_P\|_v \leq \epsilon^k \|P^l - \Pi_P\|_v, \tag{6}$$

https://doi.org/10.1017/S0269964807000034 Published online by Cambridge University Press

using norm inequalities. Define

$$c \stackrel{\text{def}}{=} \sup_{l=0,\ldots,N-1} \|P^l - \Pi_P\|_v \tag{7}$$

and $\beta \stackrel{\text{def}}{=} \epsilon^{1/2N}$. Since $\epsilon < 1$, it follows that $\beta < 1$, which implies

$$\epsilon^k = \beta^{2kN} \leq \beta^{kN+N} \leq \beta^{kN+l} = \beta^n, \tag{8}$$

where we use the fact that $0 \leq l < N$ and that $n > N$ implies $k \geq 1$. Inserting (7) and (8) in (6), we obtain the lemma. ■

We write D_P for the *deviation matrix* associated with P; in symbols,

$$D_P = \sum_{m=0}^{\infty} (P^m - \Pi_P). \tag{9}$$

Note that D_P is finite for any aperiodic finite-state Markov chain; see Lemma 2.2. Moreover, the deviation matrix can be rewritten as

$$D_P = \sum_{m=0}^{\infty} (P - \Pi_P)^m - \Pi_P,$$

where $\sum_{m=0}^{\infty} (P - \Pi_P)^m$ is often referred to as the group inverse; see, for instance, [2,4]. A general definition that is valid for any possibly periodic Markov chain can be found in, [14], for example.

3. SERIES EXPANSIONS

We are interested in the performance of a system when some of its parameters or characteristics are changed. The system as given is modeled as a Markov chain with kernel P, the changed system with kernel Q. We assume that both Markov chains have a common finite state space $\mathcal{S}$ as defined earlier (i.e., $P, Q \in [0, 1]^{S \times S}$). Note that sometimes phantom states have to be added to Q or P in order to achieve this; see Section 3.5 for an example. We also assume, as indicated earlier, that both Markov kernels are aperiodic and unichain. The goal of Section 3.1 is to obtain the stationary distribution of Q, denoted by π_Q, via a series expansion in P. In Section 3.2 we comment on the speed of convergence of this series. When applying these concepts to actually compute the stationary distribution, we have to solve the optimization problem stated in (3). We explain in Section 3.3 how to find an optimal solution. We summarize our results in an algorithm, presented in Section 3.4. Finally, we illustrate our approach with numerical examples in Section 3.5.

https://doi.org/10.1017/S0269964807000034 Published online by Cambridge University Press

3.1. Series Representation for π_Q

Let P be unichain (not necessarily finite). Elementary calculation shows (use the definition of D_P in (9))

$$(I - P)D_P = I - \Pi_P.$$

This is the Poisson equation in matrix format. Multiply this equation by Π_Q, where Π_Q denotes the projective operator of Q. Since P is unichain, it holds that $\Pi_Q\Pi_P = \Pi_P$, and we obtain

$$\Pi_Q(I - P)D_P = \Pi_Q - \Pi_P.$$

Using $\Pi_Q = \Pi_Q Q$, we obtain

$$\Pi_Q = \Pi_P + \Pi_Q(Q - P)D_P. \tag{10}$$

Inserting (10) for Π_Q in the right-hand side of (10), we obtain

$$\Pi_Q = \Pi_P + \Pi_P(Q - P)D_P + \Pi_Q((Q - P)D_P)^2.$$

Repeating this step k times yields

$$\Pi_Q = \Pi_P \sum_{n=0}^{k} ((Q - P)D_P)^n + \Pi_Q((Q - P)D_P)^{k+1} \tag{11}$$

for $k \geq 0$. Based on the above equation, we introduce the following notation. Let $k \geq 0$, then $H(k)$, with

$$H(k) \stackrel{\text{def}}{=} \Pi_P \sum_{n=0}^{k} ((Q - P)D_P)^n,$$

is called a *series approximation of degree k* for Π_Q, $T(k)$, with

$$T(k) \stackrel{\text{def}}{=} \Pi_P((Q - P)D_P)^k, \tag{12}$$

denotes the kth element of $H(k)$, and $R(k)$, with

$$R(k) \stackrel{\text{def}}{=} \Pi_Q((Q - P)D_P)^{k+1},$$

is called the *remainder term*. Notice that the remainder term $R(k)$ almost equals the $(k + 1)$st term of $H(k + 1)$ (i.e., $T(k + 1)$), except for the prefactor, which is Π_Q in $R(k)$ and Π_P in $T(k + 1)$. The quality of the approximation provided by $H(k)$ is given through the remainder term $R(k)$. This issue is discussed in the next subsection.

3.2. Convergence of the Series

In this section we investigate the limiting behavior of $H(k)$ as k tends to ∞. We first establish sufficient conditions for the existence of the series.

https://doi.org/10.1017/S0269964807000034 Published online by Cambridge University Press

LEMMA 3.1: *The following assertions are equivalent:*

(i) The series $\sum_{k=0}^{\infty}((Q-P)D_P)^k$ *is convergent.*
(ii) There are N *and* $\delta_N \in (0,1)$ *such that* $\|((Q-P)D_P)^N\|_v < \delta_N$.
(iii) There are N, κ, *and* $\delta < 1$ *such that* $\|((Q-P)D_P)^k\|_v \le \kappa\delta^k$ *for* $k \ge N$.
(iv) There are N *and* $\delta \in (0,1)$ *such that* $\|((Q-P)D_P)^k\|_v < \delta^k$ *for any* $k \ge N$.

PROOF: We first show that (i) implies (ii). Provided that $\sum_{k=0}^{\infty}((Q-P)D_P)^k$ converges, it holds that

$$\lim_{k\to\infty}[((Q-P)D_P)^k](i,j) = 0$$

for $i, j \in \mathcal{S}$. Since $\mathcal{S}$ is finite, it is possible to take k sufficiently large such that

$$\frac{\sum_{j\in\mathcal{S}}|((Q-P)D_P)^k|(i,j)v(j)}{v(i)} < 1,$$

which implies (ii).

Now suppose that (ii) holds true, and set $\varepsilon \stackrel{\text{def}}{=} \|((Q-P)D_P)^N\|_v$. Let $k \ge N$ and let m and l be integers such that $k = mN + l$. Then

$$\begin{aligned}\|((Q-P)D_P)^k\|_v &= \|((Q-P)D_P)^{mN+l}\|_v \\ &\le \|((Q-P)D_P)^{mN}\|_v \,\|((Q-P)D_P)^l\|_v \\ &\le \varepsilon^m\|((Q-P)D_P)^l\|_v. \end{aligned} \tag{13}$$

Set

$$\kappa \stackrel{\text{def}}{=} \sup_{i=0,\ldots,N-1}\|((Q-P)D_P)^i\|_v$$

and $\delta = \varepsilon^{1/2N}$. By (ii), $\varepsilon < 1$ and, thus, $\delta < 1$. Hence, the following holds:

$$\delta^k = \delta^{mN+l} \ge \delta^{2mN} = \varepsilon^m,$$

where we use the fact that $0 \le l < N$ and that for $k \ge N$, it holds that $m \ge 1$. Inserting the above in (13) yields (iii).

Suppose (iii) holds. In the case that $\kappa \le 1$, (iv) is immediate. In the case that $\kappa > 1$, let, N and $\beta < 1$ be such that

$$\kappa^{1/n}\delta \le \beta, \quad n \ge N.$$

Then

$$\kappa\delta^n \le \beta^n, \quad n \ge N,$$

and (iv); is satisfied.

https://doi.org/10.1017/S0269964807000034 Published online by Cambridge University Press

Suppose (iv); then

$$\sum_{k=0}^{\infty} \|((Q-P)D_P)^k\|_v \leq \frac{\delta^N}{1-\delta} + \sum_{k=0}^{N-1} \|((Q-P)D_P)^k\|_v.$$

Hence, the series is convergent with respect to the v-norm. Since v-norm convergence implies elementwise convergence, (i) follows. ■

Remark 3.2: Note that the fact that the maximal eigenvalue of $|(Q-P)D_P|$ is smaller than one is not necessary for $\sum_{k=0}^{\infty}((Q-P)D_P)^k$ to converge, which stems from the fact that $((Q-P)D_P)^k$ has positive and negative entries.

Existence of the limit of $H(k)$ [see (i) in Lemma 3.1], is equivalent to an exponential decay in the v-norm of the elements of the series [see (iv) in Lemma 3.1]. For practical purposes, one needs to identify the decay rate δ and the threshold value N after which the exponential decay occurs. We found in our numerical experiments that it is most convenient to work with condition (ii) in Lemma 3.1. More specifically, we work with the following condition.

(C) There exists a finite number N such that we can find $\delta_N \in (0, 1)$ that satisfies

$$\|((Q-P)D_P)^N\|_v < \delta_N,$$

and we set

$$c^v_{\delta_N} \stackrel{\text{def}}{=} \frac{1}{1-\delta_N} \left\| \sum_{k=0}^{N-1} ((Q-P)D_p)^k \right\|_v.$$

As shown in the following lemma, the factor $c^v_{\delta N}$ in condition **(C)** allows one to establish an upper bound for the remainder term that is independent of Π_Q.

LEMMA 3.3: *Under* **(C)**, *it holds that*

(i) $\|R(k-1)\|_v \leq c^v_{\delta_N} \|T(k)\|_v$ for all k
(ii) $\lim_{k\to\infty} H(k) = \Pi_P \sum_{n=0}^{\infty}((Q-P)D_p)^n - \Pi_Q$

PROOF: By definition, we have

$$\|R(k-1)\|_v = \left\| \pi_Q - \sum_{l=0}^{k-1} T(l) \right\|_v = \left\| \sum_{l=k}^{\infty} T(l) \right\|_v.$$

https://doi.org/10.1017/S0269964807000034 Published online by Cambridge University Press

This is obviously equal to

$$\left\|\Pi_P \sum_{l=k}^{\infty} ((Q-P)D_P)^l\right\|_v = \left\|\Pi_P((Q-P)D_P)^k \sum_{l=0}^{\infty} ((Q-P)D_P)^l\right\|_v$$
$$\leq \|T(k)\|_v \left\|\sum_{l=0}^{\infty} ((Q-P)D_P)^l\right\|_v$$
$$\leq \|T(k)\|_v \left\|\sum_{l=0}^{\infty} ((Q-P)D_P)^{lN}\right\|_v \left\|\sum_{k=0}^{N-1} ((Q-P)D_P)^k\right\|_v$$
$$\leq \frac{1}{1-\delta_N} \|T(k)\|_v \left\|\sum_{k=0}^{N-1} ((Q-P)D_P)^k\right\|_v,$$

which proves the first part of the lemma.

With (4), it is sufficient to show that we have v-norm convergence for $\Pi_Q - H(k)$. Using claim (i), we obtain

$$0 \leq \|R(k)\|_v \leq c^v_{\delta_N} \|\Pi_P\|_v \|((Q-P)D_P)^{k+1}\|_v.$$

By Lemma 3.1(iv), the right-hand side in the above inequality converges to zero. This proves the claim. ■

An example in which the series $H(k)$ fails to converge is illustrated in the following.

Example 3.4: Suppose the state space $\mathcal{S} = \{0,1,\ldots,\mathcal{S}\}$. Let $P(i,j) = Q(i,j)$ for $i \neq 0$ and all j, $P(0,0) = Q(0,S) = 1$, and $P(i,i-1) = 1$ for $i \neq 0$. Then

$$\Pi_P(i,0) = 1,$$
$$D_P(i,j) = \sum_{k=0}^{\infty} (P^k - \Pi_P)(i,j) = 1$$

for $i \geq j \geq 1$ and is equal to zero for $i \geq 1$, $i < j$ and

$$D_P(i,0) = -(i-1),$$
$$D_P(0,j) = 0$$

for all $j \geq 0$. Hence, $((Q-P)D)(i,j) = 0$ for $i \neq 0$,

$$((Q-P)D)(0,j) = -(S-j)$$

for $j \geq 1$, and

$$((Q-P)D)(0,0) = -(S-1),$$

https://doi.org/10.1017/S0269964807000034 Published online by Cambridge University Press

and we have that

$$((Q-P)D)^k(0,0) = (-(S-1))^k.$$

Therefore, the series is divergent.

Remark 3.5: The series expansion for Π_Q put forward in Lemma 3.3(ii) is well known; see [2] and [9] for the case of finite Markov chains and [6] for the general case. It is, however, worth noting that in the aforementioned articles, the series was obtained via a differentiation approach, whereas the representation is derived in this article is from the elementary equation (11).

Remark 3.6: Provided that $\det(I-(Q-P)D_P) \neq 0$, (10) determines π_Q uniquely and one can obtain π_Q from

$$\pi_Q = \Pi_P(I-(Q-P)D_P)^{-1}. \tag{14}$$

Moreover, provided that the limit

$$\lim_{k\to\infty} H(k) = \lim_{k\to\infty} \pi_P \sum_{n=0}^{\infty} ((Q-P)D_P)^n$$

exists (see Lemma 3.1 for sufficient conditions), it yields π_Q as $\pi_P \sum_{n=0}^{\infty} ((Q-P)D_P)^n$; see Lemma 3.3.

Equation (14) is the starting point for derivation of the series expansion put forward in Lemma 3.3(ii) in [2].

Remark 3.7: Note that a sufficient (but not necessary) condition for (**C**) is

$$\|(Q-P)D_P\|_v < \delta, \qquad \delta < 1. \tag{15}$$

We will present examples for which it turns out that $\|(Q-P)D_P\|_v > 1$, although (**C**) is satisfied for $N \geq 2$. In [1,3] it is even assumed that

$$\|(Q-P)\|_v < g_1 \tag{16}$$

with $g_1 > 0$, a finite constant, and

$$\|D_P\|_v < \frac{c}{1-\beta} \tag{17}$$

with $c > 0$ and $0 < \beta < 1$ finite constants. If

$$\frac{g_1 c}{1-\beta} < 1, \tag{18}$$

then (15) and hence (**C**) is clearly fulfilled. However, we will illustrate with some examples presented in Section 3.5 and Section 4.1 that we cannot find v such that (16)–(18) hold. Hence, for numerical purposes, these conditions are too strong.

https://doi.org/10.1017/S0269964807000034 Published online by Cambridge University Press

3.3. Bounding the Remainder Term

Until now we have not specified v. The quality of approximation by $H(k-1)$ is given by the remainder term $R(k-1)$, and in applications, v should be chosen such that it minimizes $c^v_{\delta_N}\|T(k)\|_v$, thus minimizing our upper bound for the remainder term. Since $c^v_{\delta_N}$ is independent of k, we focus on $T(k)$ for finding an optimal upper bound. Specifically, we have to find a bounding vector v that minimizes $\|T(k)\|_v$ uniformly w.r.t. k. As the following theorem shows, the unit vector, denoted by $\mathbf{1}$, with all components equal to one, yields the minimal value for $\|T(k)\|_v$ for any k. The proof of the theorem is given in the Appendix.

THEOREM 3.8: *The unit vector* $\mathbf{1}$ *minimizes* $\|T(k)\|_v$ *uniformly over* k; *that is*,

$$\forall k \geq 1: \quad \inf_v \|T(k)\|_v = \|T(k)\|_{\mathbf{1}}. \tag{19}$$

Remark 3.9: As for the results in [1,3], following the line of argument put forward in the Appendix, it can be shown that the smallest g_1 in (16) is the maximal eigenvalue of $|(Q-P)|$, and the smallest $c/(1-\beta)$ is precisely the maximal eigenvalue of $|D_P|$. Again, we note that often the product of these maximal eigenvalues is not smaller than one. In Sections 3.5 and 4.1 we will present examples for which $cg_1/(1-\beta) > 1$. If this is the case, then according to [1, 3] we cannot decide whether the series $H(k)$ converges to Π_Q. Hence, their condition is too restrictive for numerical purposes.

3.4. Algorithm

In this subsection we describe a numerical approach to computing our upper bound for the remainder term $R(k)$. We search for N such that $1 > \delta_N \stackrel{\text{def}}{=} \|((Q-P)D_P)^N\|_{\mathbf{1}}$, which implies that **(C)** holds for N and δ_N. Then the upper bound for $R(k)$ is obtained from $c^{\mathbf{1}}_{\delta_N}\|((Q-P)D_P)^{k+1}\|_{\mathbf{1}}$. Based on the above, we can now describe an algorithm that yields an approximation for π_Q with ε precision. The algorithm has two main parts. First, $c^{\mathbf{1}}_{\delta N}$ is computed. Then, the series can be computed in an iterative way until a predefined level of precision is reached.

Algorithm 1

Chose precision $\varepsilon > 0$. Set $k = 1$, $T(1) = \Pi_P(Q-P)D_P$ and $H(0) = \Pi_P$.

Step 1: Find N such that $\|((Q-P)D_P)^N\|_{\mathbf{1}} < 1$. Set $\delta_N = \|((Q-P)D_P)^N\|_{\mathbf{1}}$ and compute

$$c^{\mathbf{1}}_{\delta_N} = \frac{1}{1-\delta_N}\left\|\sum_{k=0}^{N-1}((Q-P)D_P)^k\right\|_{\mathbf{1}}.$$

Step 2: If

$$c^{\mathbf{1}}_{\delta_N}\|T(k)\|_{\mathbf{1}} < \epsilon,$$

https://doi.org/10.1017/S0269964807000034 Published online by Cambridge University Press

the algorithm terminates and $H(k-1)$ yields the desired approximation. Otherwise, go to Step 3.

Step 3: Set $H(k) = H(k-1) + T(k)$. Set $k := k+1$ and $T(k) = T(k-1)(Q - P)D_P$. Go to Step 2.

LEMMA 3.10: *Provided that* $\sum_{k-0}^{\infty}((Q-P)D_P)^k$ *is finite, Algorithm 1 terminates in a finite number of steps.*

PROOF: By Lemma 3.1, finiteness of $\sum_{k-0}^{\infty}((Q-P)D_P)^k$ implies **(C)**. Thus, we know from Lemma 3.3 that $\|R(k)\|_{\mathbf{1}} \le c_{\delta N}^{\mathbf{1}}\|T(k+1)\|_{\mathbf{1}}$ for any k. Since

$$\|T(k+1)\|_{\mathbf{1}} \le \|\Pi_P\|_{\mathbf{1}}\|((Q-P)D_P)^k\|_{\mathbf{1}},$$

the proof of the claim follows from the fact that $\lim_{k\to\infty}\|((Q-P)D_P)^k\|_{\mathbf{1}} = 0$; see Lemma 3.1. ■

3.5. Numerical Examples

Consider a model in which customers have to be served first at station 1 and, after having successfully completed service, move on to station 2 to receive service there. We assume that customers receive a successful service at station 1 with probability $1-p$; that is, a fraction p of the customers has to go through station 1 again. Customers arrive at the system according to a Poisson process with rate λ and are served at station i, $i = 1, 2$, with an exponential service time with rate μ_i. We assume the waiting capacity at both stations to be finite [i.e., b_i customers can be at station i (including the customer in service)] and that customers that find upon arrival or service completion no free buffer places at the next station are lost. The current system with parameters λ, μ_1, μ_2, b_1, and b_2 is modeled as a uniformized embedded Markov chain with kernel P, and its stationary distribution π_P is known. In all examples we take $p = 0.25$, $\mu_1 = 2$, $\mu_2 = 2$, and $b_1 = 10 = b_2$. We are interested in the probability that customers find station 2 full; we call this probability the *overflow probability* and denote it by p_λ.

We want to investigate what happens when the waiting capacity of station 2 is expanded to $b_2 + 1$. Hence, kernel Q represents the uniformized embedded Markov chain of the system with an additional buffer place at station 2. Note that we created a phantom state $b_2 + 1$ in the P kernel, from which a jump to b_2 occurs with probability 1. We denote the prediction for p_λ based on $H(k)$ by $p_\lambda(k)$. Figure 1 shows the absolute relative error (in formula, $|p_\lambda - p_\lambda(k)|/p_\lambda$) for different values of k and different values for λ (i.e., different traffic loads). As can be seen, the overflow probability can be predicted within an error of 1% through $H(16)$, and in the light-traffic case, $H(11)$ is sufficient.

We now turn to the numerical behavior of our upper bound for the remainder term. For given λ, denote by $r_\lambda(k)$ the upper bound for $R(k)$ given in Lemma 3.3. Figure 2

https://doi.org/10.1017/S0269964807000034 Published online by Cambridge University Press

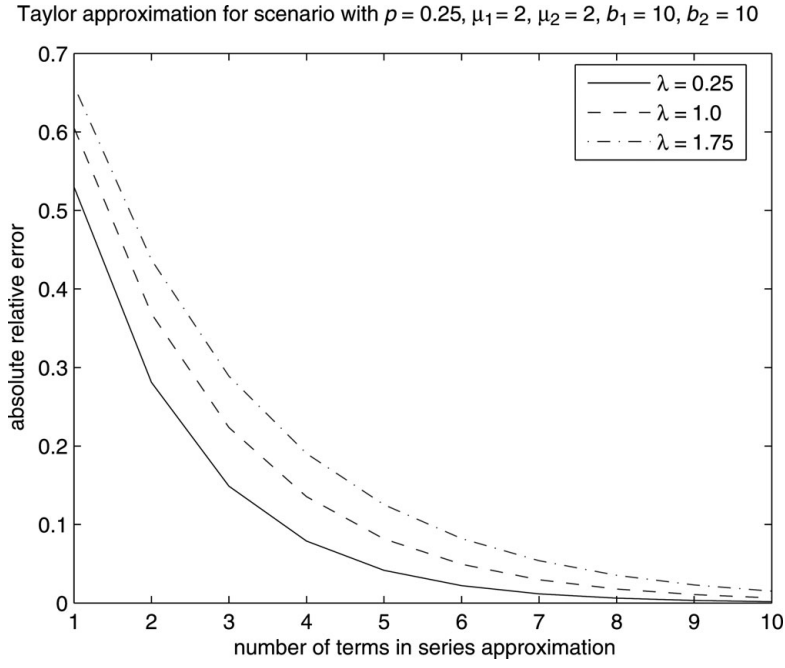


FIGURE 1. Absolute relative error.

illustrates the relative error of our upper bound (in formula, $(r_\lambda(k) - (p_\lambda - p_\lambda(k)))/p_\lambda$) or various values of λ and different values of k. For $\lambda = 0.25$, we obtained (**C**) for $N = 3$ and $c^{\mathbf{1}}_{\delta_N} = 11.17$; for $\lambda = 1.0$, we obtained (**C**) for $N = 4$ and $c^{\mathbf{1}}_{\delta_N} = 4.65$; and for $\lambda = 1.75$, we obtained (**C**) for $N = 5$ and $c^{\mathbf{1}}_{\delta_N} = 15.397$.

To illustrate the decay of the relative error induced by the upper bound for the remainder term, we show in Figure 3 the relative error of our upper bound for values of k between 11 and 20.

Remark 3.11: Let us turn to the condition of [1, 3]. For $p = 0.5$, for instance, the smallest g_1 is 0.6718 and the smallest $c/(1-\beta)$ is 6.5557, and, hence, $cg_1/(1-\beta) > 1$, although (**C**) holds; see Remark 3.7.

4. A POWER SERIES APPROACH

Let P and Q be given as in Section 2. In this section, we study the convex combination

$$P_\theta = (1 - \theta)P + \theta Q, \qquad \theta \in [0, 1], \tag{20}$$

https://doi.org/10.1017/S0269964807000034 Published online by Cambridge University Press

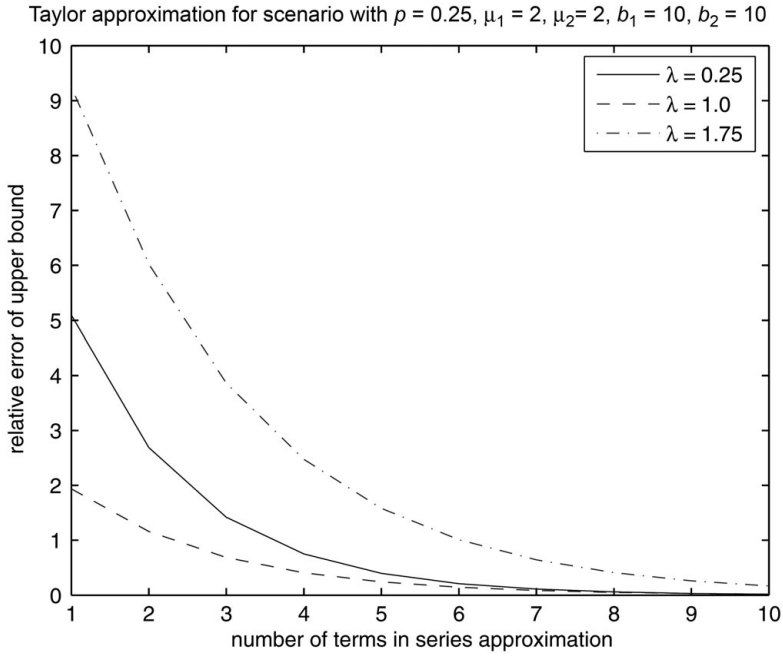


FIGURE 2. Relative error of the bound of the remainder term.

of the two chains. We are now interested in the effect of changing θ to $\theta + \Delta$ on the stationary distribution of P_θ, denoted by π_θ, and the corresponding projective operator, denoted by Π_θ. Following the same procedure as in Section 3.1, we obtain

$$\Pi_{\theta+\Delta} = \Pi_\theta \sum_{n=0}^{k} ((P_{\theta+\Delta} - P_\theta)D_\theta)^n + \Pi_{\theta+\Delta}((P_{\theta+\Delta} - P_\theta)D_\theta)^{k+1}$$

for $k \geq 0$. From (20) it is easily seen that

$$P_{\theta+\Delta} - P_\theta = \Delta(Q - P)$$

so that the above series can be written as

$$\Pi_{\theta+\Delta} = \Pi_\theta \sum_{n=0}^{k} \Delta^n((Q - P)D_\theta)^n + \Pi_{\theta+\Delta}\Delta^{k+1}((Q - P)D_\theta)^{k+1} \tag{21}$$

for $k \geq 0$. For $k \geq 0$, we define the series approximation $H_{\theta,\Delta}(k)$ by

$$H_{\theta,\Delta}(k) \stackrel{\text{def}}{=} \Pi_\theta \sum_{n=0}^{k} \Delta^n((Q - P)D_{P_\theta})^n, \tag{22}$$

https://doi.org/10.1017/S0269964807000034 Published online by Cambridge University Press

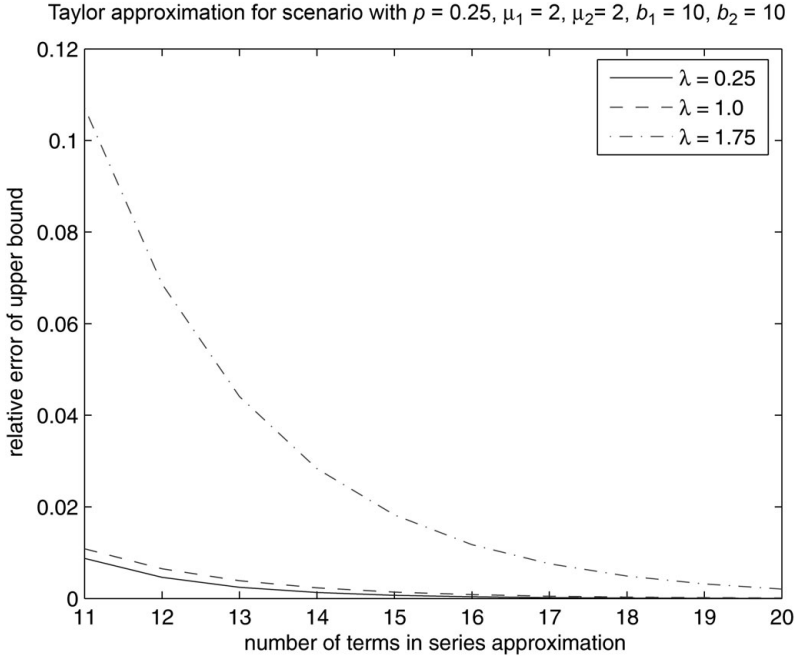


FIGURE 3. Relative error of the bound of the remainder term.

where we denote the kth element of the above series by

$$T_{\theta,\Delta}(k) \stackrel{\text{def}}{=} \Delta^k \Pi_\theta ((Q-P)D_{P_\theta})^k,$$

and the remainder term $R_\Delta(k)$ is given through

$$R_{\theta,\Delta}(k) \stackrel{\text{def}}{=} \Pi_{\theta+\Delta} \Delta^{k+1} ((Q-P)D_{P_\theta})^{k+1}.$$

For the series $H_{\theta,\Delta}(k)$ to converge, $\Delta^k T_{\theta,\Delta}(k)$ has to converge to zero as k tends to ∞. This leads to the following adaptation of (**C**).

(**C**)($\boldsymbol{\theta}$) There exist finite numbers N and $\delta_N^\theta \in (0, 1)$ such that

$$\|((Q-P)D_\theta)^N\|_v \leq \delta_N^\theta,$$

and we set

$$c_{\delta_N^\theta}^v \stackrel{\text{def}}{=} \frac{1}{1-\delta_N^\theta} \left\| \sum_{k=0}^{N-1} ((Q-P)D_\theta)^k \right\|_v.$$

Condition (**C**)($\boldsymbol{\theta}$) implies that $H_{\theta,\Delta}(k)$ converges at least for all Δ with $|\Delta| < 1/\delta_{N_\theta}$. The following lemma is a straightforward adaptation of the result for series expansion.

https://doi.org/10.1017/S0269964807000034 Published online by Cambridge University Press

LEMMA 4.1: *For* $\theta \in [0, 1]$, *let* $(\mathbf{C})(\boldsymbol{\theta})$ *be satisfied for some* δ_N^θ. *Then it holds for all* $|\Delta| < 1/\delta_N^\theta$ *that*

(i) $\|R_{\theta,\Delta}(k-1) \leq c^{\mathbf{v}}_{\delta_N^\theta} \|T_{\theta,\Delta}(k)\|_v$ *for all* k
(ii) $\lim_{k\to\infty} H_{\theta,\Delta}(k) = \Pi_{\theta+\Delta}$.

The above leads to the following algorithm, which yields an approximation for $\pi_{\theta+\Delta}$ with ε precision, where we make use of Theorem 3.8.

Algorithm 2

Chose precision $\varepsilon > 0$. Set $k = 1$, $T_{\theta,\Delta}(1) = \Delta\Pi_\theta(Q - P)D_\theta$ and $H_{\theta,\Delta}(0) = \Pi_\theta$.

Step 1: Find N such that $\|((Q - P)D_\theta)^N\|_{\mathbf{1}} < 1$. Set $\delta_N^\theta = \|((Q - P)D_\theta)^N\|_{\mathbf{1}}$ and compute

$$c^{\mathbf{1}}_{\delta_N^\theta} = \frac{1}{1 - \delta_N^\theta} \left\| \sum_{k=0}^{N-1} ((Q - P)D_\theta)^k \right\|_{\mathbf{1}}.$$

Step 2: If

$$c^{\mathbf{1}}_{\delta_N^\theta} \|T_{\theta,\Delta}(k)\|_{\mathbf{1}} < \epsilon,$$

the algorithm terminates and $H_{\theta,\Delta}(k-1)$ yields the desired approximation. Otherwise, go to Step 3.

Step 3: Set

$$T_{\theta,\Delta}(k+1) = \Delta T_{\theta,\Delta}(k)(Q - P)D_{P_\theta}$$

and $H_{\theta,\Delta}(k) = H_{\theta,\Delta}(k-1) + T_{\theta,\Delta}(k)$. Let $k := k + 1$ and go to Step 2.

The above algorithm is not guaranteed to yield the desired approximation, which stems from the fact that $|\Delta|$ might lay outside the radius of convergence of the series in (22). However, $1/\delta_N^\theta$ as computed by the above algorithm yields a lower bound for radius of convergence of $H_{\theta,\Delta}(k)$ and thus is an indication of the maximal value of $|\Delta|$.

Remark 4.2: Taking $\theta = 0$ and $\Delta = 1$, we are in the situation of Section 3.1. In symbols, $H(k) = H_{\theta,1}(k)$ for $k \geq 0$.

The series given in (21) is obviously a power series. Moreover, it is shown in [7] that under v-geometric ergodicity of P_θ, the term $\Pi_\theta((Q - P)D_{P_\theta})^n$ is equal to $d\Pi_\theta/d\theta^n$. This means that (21) is a Taylor series.

Hence, Lemma 4.1 yields a lower bound for the radius of convergence of the Taylor series $H_{\theta,\Delta}$; that is, $H_{\theta,\Delta}$ converges to the right function on $\Delta \in [0, (1/\delta_N^\theta)]$. In other words, $1/\delta_N^\theta$ is the lower bound for the domain of analyticity of Π_θ.

https://doi.org/10.1017/S0269964807000034 Published online by Cambridge University Press

We apply our theoretical results to the coupled processors model. Consider a system with two finite queues, each with their own server. Customers for queue i enter the system at a Poisson rate λ_i. As long as there are customers in both queues, the server of queue i works at service rate μ_i. However, if the server of queue i is idle, it joins the nonidle server j, making the average service rate of the nonidle queue larger than μ_j.

There are but few analytical results for this model; see [5, 11] for a derivation of the moment generating function of the joint distribution. In particular, no closed-form solution for the stationary distribution is known. In [10], a heavy-traffic approximation was found, and in [7], a large deviation analysis is given.

For P, we take the kernel of the system where customers arrive with rate λ_i to queue i and where each queue has its own server that serves at rate μ_i, independent of the state of the other queue. We denote the Markov kernel of the embedded jump chain of the coupled processor model by Q. Then the kernel P_θ (see (20) for the definition) represents a system in which the server of an idle queue joins the nonidle queue with probability θ.

Consider the following numerical example. Take $\lambda_1 = 1.9$, $\lambda_2 = 1.8$, $\mu_1 = \mu_2 = 2$, $b_1 = 5$, and $b_2 = 5$. We are interested in the probability that queue 1 is full, and we call this the *overflow probability*. We denote the overflow probability given by P_θ by $\hat{p}_\theta$. We develop the series at $\theta = 0$; that is, $D_{P\theta} = D_P$ and $\Pi_{\theta+\Delta} = \Pi_\Delta$ with

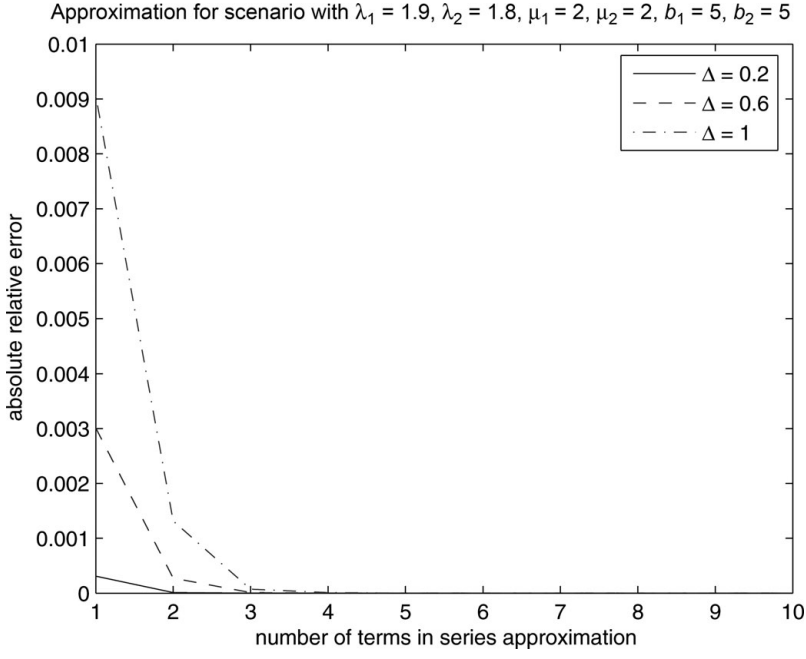


FIGURE 4. Absolute relative error.

https://doi.org/10.1017/S0269964807000034 Published online by Cambridge University Press

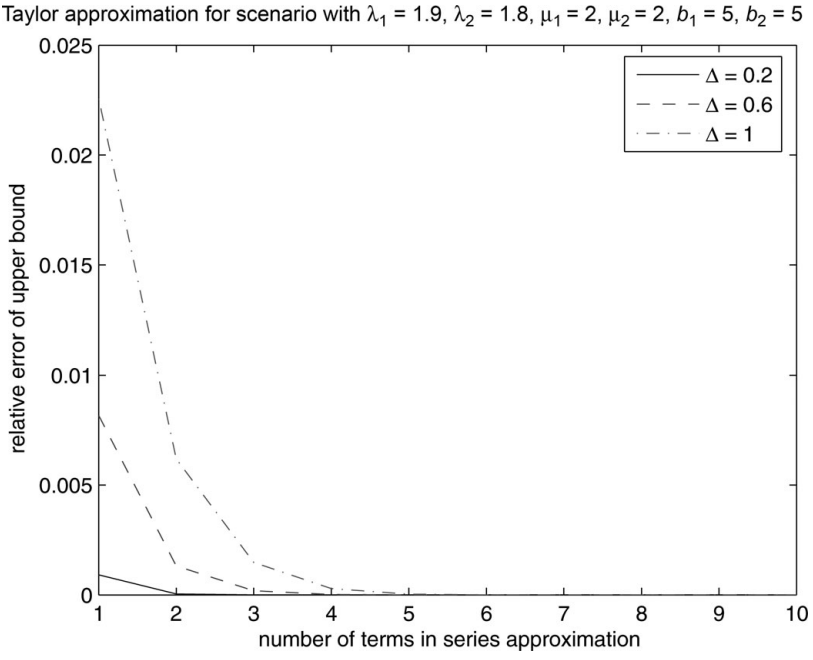


FIGURE 5. Relative error of the upper bound of the remainder term.

$\Pi_\theta = \Pi_P$. The series in (22) thus reads

$$H_{0,\Delta}(k) = \Pi_P \sum_{n=0}^{k} \Delta^n ((Q - P)D_P)^n$$

for $k \geq 0$.

Let $\hat{p}_\Delta(k)$ denote the prediction of the overflow probability based on $H_{0,\Delta}(k)$. Figure 4 shows the absolute relative error (in formula, $|\hat{p}_\Delta - \hat{p}_\Delta(k)|/\hat{p}_\Delta$) of predicting the overflow probability at queue 1 by $H_{0,\Delta}(k)$ for different values of k and different values for Δ.

Figure 5 shows the absolute relative error of the remainder term (in formula, $(\hat{r}_\Delta(k) - (\hat{p}_\Delta - \hat{p}_\Delta(k)))/\hat{p}_\Delta$, where $\hat{r}_\Delta(k)$ denotes our upper bound for $R_{0,\Delta}(k)$) for different values of k and different values of Δ. In all cases, condition $\mathbf{C}(0)$ is satisfied for $N = 2$ and $c^{\mathbf{1}}_{\delta^0_N} = 0.84$.

Remark 4.3: For this example it turns out that $cg_1/(1-\beta) = 0.3342 \times 14.1940 > 1$; see also Remark 3.7.

5. Conclusion

We presented a new algorithm for the approximative computation of characteristics of finite Markov chains. It could be shown that under quite mild conditions, the

https://doi.org/10.1017/S0269964807000034 Published online by Cambridge University Press

algorithm approximates the true solution. Moreover, upper bounds for the error of the approximations could be established. Numerical examples illustrated the fast convergence rate of our algorithm. In future work we will study our methodology for large-scale problems.

References

1. Altman, E., Avrachenkov, K.E., & Núñez-Queija, R. (2004). Perturbation analysis for denumerable Markov chains with application to queueing models. *Advances in Applied Probability* 36: 839–853.
2. Cao, X.-R. (1998). The Maclaurin series for performance functions of Markov chains. *Advances in Applied Probability* 30: 676–692.
3. Cho, G.E. & Meyer, C.D. (2001). Comparison of perturbation bounds for the stationary distribution of a Markov chain. *Linear Algebra and its Applications* 335: 137–150.
4. Coolen-Schrijner, P. & van Doorn, E.A. (2002). The deviation matrix of a continuous-time Markov chain. *Probability in the Engineering and Informational Sciences* 16: 351–366.
5. Fayolle, G. & Iasnogorodski, R. (1979). Two coupled processors: The reduction to a Riemann-Hilbert problem. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 47: 325–351.
6. Heidergott, B. & Hordijk, A. (2003). Taylor series expansions for stationary Markov chains. *Advances in Applied Probability* 35: 1046–1070.
7. Hordijk, A. & Popov, N. (2003). Large deviation analysis of a coupled processor system. *Probability in the Engineering and Informational Sciences* 17: 397–409.
8. Hordijk, A. & Spieksma, F.M. (1992). On ergodicity and recurrence properties of a Markov chain with an application to an open Jackson network. *Advances in Applied Probability* 24: 343–376.
9. Kirkland, S. (2003). Conditioning properties of the stationary distribution for a Markov chain. *Electronic Journal of Linear Algebra* 10: 1–15.
10. Knessl, C. & Morrison, J.A. (2003). Heavy traffic analysis of two coupled processors. *Queueing Systems* 43: 173–220.
11. Konheim, A.G., Meilijson, I., & Melkman, A. (1981). Processor sharing of two parallel lines. *Journal of Applied Probability* 18: 952–956.
12. Meyn, S.P. & Tweedie, R.L. (1993). *Markov chains and stochastic stability*. London: Springer-Verlag.
13. Minc, H. (1988). *Nonnegative matrices*. New York: Wiley.
14. Puterman, M.L. (1994). *Markov decision processes: Discrete stochastic dynamic programming*. New York: Wiley.

APPENDIX

We give a proof of Theorem 3.8. Let $A = |\Pi_P((Q - P)D_P)^k|$ for $k \in \mathbb{N}$. The important case is when A has nonzero elements. Indeed, if A is identical to the zero matrix, then its trivial that

$$0 = \inf_v \|A\|_v = \|A\|_{\mathbf{1}}.$$

Now suppose that A has nonzero elements. The key observation for the proof is that any two rows of A are identical. Suppose that A has m nonzero columns. Hence, after possible relabeling of states, A is of the form

$$A = \begin{pmatrix} A_1 & 0_1 \\ A_2 & 0_2 \end{pmatrix},$$

where A_1 is an $m \times m$-dimensional matrix with all elements larger than zero, A_2 is an $(S - m) \times m$ matrix with all elements larger than zero, and 0_1 and 0_2 are zero matrices of appropriate size.

https://doi.org/10.1017/S0269964807000034 Published online by Cambridge University Press

The proof has two main steps. First, we will show that

$$\lambda = \inf_v \|A_1\|_v, \tag{A.1}$$

where λ denotes the maximal eigenvalue of A_1. In a second step, we will show that (A.1) holds when A_1 is replaced by the original matrix A.

We now turn to the proof of (A.1). Observe that solving (A.1) can be interpreted as finding the minimal value of $\|A_1\|_v$. This problem can be translated into an equivalent linear programming problem as follows. We have to find the minimal value for δ such that the problem

$$\min \sum_{i \in S} v(i) \tag{A.2}$$

such that

$$(A_1 - \delta I)v \leq 0$$

and

$$v \geq 1,$$

returns a feasible solution v. The following lemma shows that the minimal δ for which (A.2) has a solution is equal to the maximal eigenvalue of A_1.

LEMMA A.1: *The linear programming problem (A.2) for any $\delta \in [\lambda, \infty)$ has a solution, where $\lambda > 0$ is the maximal eigenvalue of A_1.*

PROOF: Note that A_1 is a full matrix and thus irreducible. Furthermore, all elements of A_1 are larger than or equal to zero. Thus, Theorem 4.5 in [14, Chap. 1] applies, which gives immediately that for any δ larger than or equal to the maximal eigenvalue λ, the linear programming problem (A.2) has a solution. With the Perron Frobenius theorem, it follows that $\lambda > 0$. Finally, use Corollary 4.2 and Theorem 4.4 (both are in [14, Chap. 1]) to conclude that for any δ, the feasible vector v is the rescaled unique positive eigenvector associated with λ. ■

Matrix A_1 has identical rows; thus, any eigenvector associated with the maximal eigenvalue of A_1 has identical components. This implies $A_1\mathbf{1} = \lambda\mathbf{1}$ and, by Lemma 5.1,

$$\lambda = \|A_1\|_{\mathbf{1}} = \inf_v \|A_1\|_v,$$

which establishes (A.1).

For the second part of the proof, note that the fact that A_1 and A_2 have equal rows implies that

$$(A_1 v)(i) = (A_2 v)(j)$$

for $1 \leq i \leq m$ and $m + 1 \leq j \leq S$. Consequently,

$$\lambda = \|A\|_{\mathbf{1}} = \inf_v \|A\|_v,$$

which proves the claim.

https://doi.org/10.1017/S0269964807000034 Published online by Cambridge University Press