Applied Probability Trust (5 December 2003)

PERTURBATION ANALYSIS FOR DENUMERABLE MARKOV CHAINS WITH APPLICATION TO QUEUEING MODELS

EITAN ALTMAN,* *INRIA*

KONSTANTIN E. AVRACHENKOV,** *INRIA*

RUDESINDO NÚÑEZ-QUEIJA,*** *CWI*

Abstract

We study the parametric perturbation of Markov chains with denumerable state spaces. We consider both regular and singular perturbations. By the latter we mean that transition probabilities of a Markov chain, with several ergodic classes, is perturbed such that (rare) transitions among the different ergodic classes of the unperturbed chain are allowed. Singularly perturbed Markov chains have been studied in literature under more restrictive assumptions such as strong recurrence ergodicity or Doeblin conditions. We relax these conditions so that our results can be applied to queueing models (where the conditions mentioned above typically fail to hold). Assuming ν-geometric ergodicity, we are able to explicitly express the steady state distribution of the perturbed Markov chain as a Taylor series in the perturbation parameter. We apply our results to quasi birth and death processes and queueing models.

Keywords: Denumerable Markov Chains; Perturbation Analysis; Geometric Ergodicity; Quasi Birth and Death Processes; Queueing Models

AMS 2000 Subject Classification: Primary 60J10

Secondary 60J22;60J27;60K27

* Postal address: INRIA Sophia Antipolis, 2004 Route des Lucioles, B.P. 93, Sophia Antipolis Cedex, 06902, France, e-mails: altman@sophia.inria.fr, k.avrachenkov@sophia.inria.fr

*** Postal address: CWI, P.O. Box 94079, Amsterdam, 1090 GB, The Netherlands, e-mail: sindo@cwi.nl. Also affiliated with Eindhoven University of Technology.

1. Introduction

Let us consider a Markov chain on the denumerable state space $E = \{1, 2, ...\}$, whose generator depends on a small parameter ε, that is

$$G(\varepsilon) = G^{(0)} + \varepsilon G^{(1)}, \quad 0 \le \varepsilon \le \bar{\varepsilon}. \tag{1}$$

Throughout this paper we restrict ourselves to a discrete-time Markov chain, in which case the transition matrix of the chain is given by $P(\varepsilon) = \mathcal{I} + G(\varepsilon)$. However, our analysis equally applies to a continuous-time Markov process with transition rates uniformly bounded by some constant $r > 0$; in that case we shall focus on the uniformized discrete-time chain

$$P(\varepsilon) = \mathcal{I} + \Delta G(\varepsilon) \tag{2}$$

with $\Delta < r^{-1}$ [27]. We call $G(\varepsilon)$ and $G^{(0)}$ the perturbed generator and the unperturbed generator, respectively. From now on, we assume that there exists a non-zero $\bar{\varepsilon}$ such that $G(\varepsilon)$ is the generator of an *irreducible* Markov chain for $\varepsilon \in (0, \bar{\varepsilon}]$ (i.e. all states communicate; transient states are excluded from our consideration). We consider two types of perturbations: *regular* perturbation, if the unperturbed generator $G^{(0)}$ is also irreducible, and *singular* perturbation, if the unperturbed Markov chain has several ergodic classes. Let us note that the irreducibility of the perturbed Markov chain implies that, if the invariant probability measure denoted by $\pi(\varepsilon)$ exists, it is unique and satisfies

$$\sum_{i \in E} \pi_i(\varepsilon) G_{ij}(\varepsilon) = 0, \quad j \in E, \tag{3}$$

$$\sum_{i \in E} \pi_i(\varepsilon) = 1, \tag{4}$$

for $0 < \varepsilon \le \bar{\varepsilon}$. We give conditions for its existence in terms of the characteristics of the unperturbed Markov chain (e.g. probability measure(s), deviation matrix) and the perturbation term $G^{(1)}$. Furthermore, we show that, under rather weak conditions, the invariant probability measure of the perturbed chain $\pi(\varepsilon)$ is analytic in ε in the punctured neighborhood of zero. Specifically,

$$\pi(\varepsilon) = \pi^{(0)} + \varepsilon\pi^{(1)} + ..., \quad 0 < \varepsilon \leq \bar{\varepsilon}. \tag{5}$$

Note that $\pi(0)$, the invariant probability measure of the unperturbed chain, is not well defined if the perturbation is singular. We shall show that the coefficients of power series (5) form a geometric sequence and, hence, there exists a computationally stable updating formula for $\pi(\varepsilon)$.

Before proceeding, let us discuss the existing results on perturbation analysis of Markov chains and Markov Processes. There is a significant body of literature on perturbation analysis of *finite state* Markov chains and Markov processes, for the extensive bibliography see the books [8, 25, 29] and the survey paper [4]. However, there are only few references available on perturbation analysis of Markov chains with an *infinite state space*. Singularly perturbed Markov chains on general measurable state spaces have been analyzed in the book of Korolyuk and Turbin [18] and in the paper of Bielecki and Stettner [5]. We would like to note that in [5, 18] the authors impose Doeblin type conditions for the unperturbed Markov chain. These conditions are quite restrictive. For instance, the simple M/M/1 queueing model does not satisfy them. Furthermore, the case of an infinite number of ergodic classes in the unperturbed Markov chain cannot be considered under the Doeblin conditions. Cao and Chen [6] have analyzed regularly perturbed Markov chains on countable state spaces under a strong ergodicity assumption. This assumption also excludes such simple models as an M/M/1 queue. In [30] Yin and Zhang analyze singularly perturbed continuous-time Markov processes on denumerable state space under conditions equivalent to the Doeblin conditions. The authors mistakenly stated that the M/M/1 model satisfies the Doeblin conditions (see for instance Section 7.1 of [16] for explanations why the M/M/1 model does not satisfy the Doeblin conditions). In the present work, we use the concepts of Lyapunov functions and ν-geometric ergodicity [9, 16, 21, 27, 28], that allow us to treat the cases that satisfy neither Doeblin conditions nor strong ergodicity condition. We would also like to mention a number of works on the application of singularly perturbed Markov chains to Quasi Birth and Death models, Queueing models and Reliability Theory [1, 2, 7, 11, 20, 24]. The singular perturbation techniques allow to solve models with significantly larger state space than in the case of direct application of standard tools. At the end of the paper we consider some of these models as examples of our framework. In particular, with the help of these examples we show how the general Lyapunov function based conditions presented in the paper can be applied to queueing models.

2. Preliminaries

Let us recall the notions of ν-geometric ergodicity, Lyapunov functions and the relation between these two concepts. Fix a denumerable vector ν with strictly positive entries. Let x be a vector defined on a subset $\mathcal{X}$ of the nonnegative integers. Then its ν-norm is defined as follows:

$$||x||_\nu = \sup_{i\in\mathcal{X}} \frac{|x_i|}{\nu_i}.$$

The corresponding induced ν-norm for any operator A on $\mathcal{X}\times\mathcal{X}$ is given by

$$||A||_\nu = \sup_{i\in\mathcal{X}} \nu_i^{-1} \sum_{j\in\mathcal{X}} |A_{ij}|\nu_j.$$

A Markov chain on a denumerable (or finite) state space with transition operator P is said to be ν-geometrically ergodic if

$$||P^k - \Pi||_\nu \le c\beta^k, \quad k = 0, 1, 2, ... \tag{6}$$

where $\Pi = \underline{1}\pi$ is the ergodic projection of the unperturbed Markov chain, i.e. $\Pi = \lim_{k\to\infty} \frac{1}{k}\sum_{j=1}^{k} P^j$; $\underline{1}$ is a column vector of 1's, π is a row vector and c, $\beta < 1$ are (scalar) constants. The notion of ν-geometric ergodicity is a powerful theoretical tool. However, in practice the condition (6) is often difficult to verify. The main difficulty lies in the computation of the powers of the transition operator. Often even for little n the expression for P^n becomes very cumbersome. Thus, instead of actually checking all of the moments, it would be simpler if we could obtain the second largest (in absolute value) eigenvalue of P or a bound on it. As we shall see (Theorem 1), this can be done using a Lyapunov function approach.

Let us consider the following stability condition based on Lyapunov function: There exist a strongly aperiodic state $\alpha \in E$ (i.e. a state satisfying $P_{\alpha\alpha} > p_0$), constants $\delta < 1$, $b < \infty$ and a Lyapunov function $\mathcal{V}$, $\mathcal{V}_i \geq 1, \forall i \in E$, such that

$$P\mathcal{V} \leq \delta\mathcal{V} + b\mathbf{I}_\alpha. \tag{7}$$

The above vector inequality should be considered component-wise, that is, for two vectors v and w, the inequality $v \leq w$ means $v_i \leq w_i$. The following theorem [22] shows that the Lyapunov function based stability implies the ν-geometric ergodicity.

Theorem 1. *Let the Lyapunov function based stability condition* (7) *hold for a Markov chain. Then the chain is $\mathcal{V}$-geometrically ergodic; it satisfies* (6) *for any $\beta > \theta$ and $c = \beta/(\beta - \theta)$, where $\theta = 1 - M_\alpha^{-1}$,*

$$M_\alpha = \frac{1}{(1-\delta)^2}[1 - \delta + b + b^2 + \zeta_\alpha(b(1-\delta) + b^2)],$$

and where ζ_α is some positive constant satisfying

$$\zeta_\alpha \leq \frac{32 - 8p_0^2}{p_0^3}\left(\frac{b}{1-\delta}\right)^2.$$

The above theorem demonstrates that the Lyapunov function $\mathcal{V}$ can be used as a bounding function for the ν-norm.

3. Regular Perturbation

We first make several rather weak assumptions. The first assumption guarantees that the perturbation is regular.

Assumption (R1): The unperturbed Markov chain is irreducible.

Assumption (R2): The unperturbed Markov chain satisfies the stability condition (7) for some strongly aperiodic state $\alpha \in E$ ($P_{\alpha\alpha} > p_0$), some constants $\delta < 1$, $b < \infty$ and a Lyapunov function $\mathcal{V}$, $\mathcal{V}_i \geq 1, \forall i \in E$.

Assumption (R3): The perturbation matrix $G^{(1)}$ is $\mathcal{V}$-bounded, that is, $||G^{(1)}||_{\mathcal{V}} \leq g_1$, where $g_1 > 0$ is some constant.

Assumptions (R1) and (R2) imply that the unperturbed Markov chain has a unique invariant probability measure π, which is a solution of the following system $\sum_{i \in E} \pi_i G^{(0)}_{ij} = 0, j \in E, \quad \sum_{i \in E} \pi_i = 1$, see [21, 27]. Furthermore, from Assumption (R2) and Theorem 1 we conclude that there exist constants c and β ($c > 0$, $0 < \beta < 1$) such that (6) holds for $P(0)$, where $\Pi = \underline{1}\pi$ is the ergodic projection of the unperturbed Markov chain. Hence, by Lemma 4.1 from [27] there exists a $\mathcal{V}$-bounded deviation matrix $H = \sum_{j=0}^{\infty}(P(0) - \Pi)^j$ which is the unique solution of the following equations

$$HG^{(0)} = G^{(0)}H = \Pi - \mathcal{I}, \tag{8}$$

$$H\Pi = \Pi H = 0. \tag{9}$$

We have the following bound for the $\mathcal{V}$-norm of the deviation matrix

$$||H||_{\mathcal{V}} \leq \frac{c}{1-\beta}. \tag{10}$$

Now we are able to formulate and prove the main result of this section.

Theorem 2. *Let Assumptions (R1), (R2) and (R3) be satisfied and let g_1, β and c be as in (6) and Assumption (R3). Then, the perturbed Markov chain has a unique invariant probability measure $\pi(\varepsilon)$, which is an analytic function of ε*

$$\pi(\varepsilon) = \pi^{(0)} + \varepsilon\pi^{(1)} + \varepsilon^2\pi^{(2)} + ..., \quad 0 \le \varepsilon \le \min\{\bar{\varepsilon}, \frac{1-\beta}{g_1 c}\},$$

with $\pi^{(k)} = \pi[(G^{(1)}H)^k]$, where π is the invariant probability measure of the unperturbed Markov chain. Moreover, the invariant probability measure of the perturbed chain can be calculated by the updating formula

$$\pi(\varepsilon) = \pi[\mathcal{I} - \varepsilon G^{(1)}H]^{-1}, \quad 0 \le \varepsilon \le \min\{\bar{\varepsilon}, \frac{1-\beta}{g_1 c}\}. \tag{11}$$

Proof: Recall that $G(\varepsilon)$ is an irreducible generator. Hence, if a solution of (3) and (4) exists, it is unique. Next, we show constructively that $\pi(\varepsilon)$ can be represented by a power series (5) with non-zero radius of convergence. Toward this end, let us substitute (1) and (5) into (3) and equate coefficients with the same powers of ε. This leads to the following system of equations

$$\pi^{(0)}G^{(0)} = 0, \tag{12}$$

$$\pi^{(k)}G^{(0)} + \pi^{(k-1)}G^{(1)} = 0, \quad k = 1, 2, ... \tag{13}$$

The normalization condition (4) implies

$$\pi^{(0)}\underline{1} = 1, \tag{14}$$

$$\pi^{(k)}\underline{1} = 0, \quad k = 1, 2, ... \tag{15}$$

Since the unperturbed Markov chain has a unique invariant probability measure, from (12) and (14) we conclude that the first term in the power series of (5) is equal to the invariant probability measure of the unperturbed Markov chain, that is $\pi^{(0)} = \pi$. Next let us consider equation (13). As H is the group generalized inverse of $-G^{(0)}$ (see equations (8) and (9)) [27], we can write the general solution of (13) in the form

$$\pi^{(k)} = c^{(k)}\pi + \pi^{(k-1)}G^{(1)}H,$$

where $c^{(k)}$ is some constant. Now we use condition (15) and the fact that $\underline{1}\pi = 1$.

$$\pi^{(k)}\underline{1} = c^{(k)} + \pi^{(k-1)}G^{(1)}H\underline{1} = 0$$

Note that it follows from (9) that $H\underline{1} = 0$, hence $c^{(k)} = 0, k = 1, 2, ...$ and

$$\pi^{(k)} = \pi^{(k-1)}G^{(1)}H, \quad k = 1, 2, ...$$

Since the matrices $G^{(1)}$ and H are $\mathcal{V}$-bounded, the power series (5) is absolutely convergent with non-zero radius of convergence (the radius of convergence is greater than or equal to $||G^{(1)}||_{\mathcal{V}}^{-1}||H||_{\mathcal{V}}^{-1}$). This justifies the substitution of (5) into (3) and (4). Finally, the updating formula (11) is an immediate consequence of the fact that the coefficients of power series (5) form a geometric sequence.

□

Remark 3.1. The updating formula (11) can alternatively be expressed as

$$\pi(\varepsilon) = \pi + \varepsilon\pi(\varepsilon)G^{(1)}H.$$

Thus, new approximations of $\pi(\varepsilon)$ can be computed recursively (hence, the term "updating" formula).

4. Singular Perturbation

Next, we study singularly perturbed Markov chains. In the case of singular perturbation, several ergodic classes are connected in a single Markov chain by "small" transition probabilities.

Assumption (S1): The unperturbed Markov chain consists of several ergodic classes and there are no transient states. Denote the ergodic classes by $E_I, I \in \bar{E}$, where $\bar{E}$ is either a finite or denumerable set. Each E_I itself is either finite or denumerable. We denote the transition operator of each ergodic class by $A_I, I \in \bar{E}$. Thus, the transition operator of the unperturbed Markov chain can be written in the form

$$P^{(0)} = \begin{bmatrix} A_1 & 0 & \cdots \\ 0 & A_2 & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}.$$

Assumption (S2): The Markov chains corresponding to ergodic classes of the unperturbed Markov chain are *uniformly* Lyapunov stable. That is, for each ergodic class there exist a strongly aperiodic state $\alpha(I) \in E_I$, $P_{\alpha(I)\alpha(I)} > p_0$ (without loss of generality we can take $\alpha = (I, 1)$), constants $\delta < 1$, $b < \infty$ and a Lyapunov function $\mathcal{V}$, $\mathcal{V}_i \geq 1, \forall i \in E_I$, such that

$$A_I \mathcal{V} \leq \delta \mathcal{V} + b \mathbf{1}_{\alpha(I)}, \quad I \in \bar{E}. \tag{16}$$

We would like to emphasize that the Lyapunov function $\mathcal{V}$ as well as the constants δ and b are the same for all ergodic classes.

Assumption (S2) together with Theorem 1 implies that the Markov chains corresponding to the ergodic classes of the unperturbed Markov chain are *uniform* $\mathcal{V}$-geometrically ergodic. Namely, there exist constants c and β ($c > 0$, $0 < \beta < 1$) such that

$$||A_I^n - \Pi_I||_{\mathcal{V}} \leq c\beta^n, \quad k = 0, 1, 2, \ldots, \tag{17}$$

where Π_I is the ergodic projection for the I-th ergodic class.

Remark 4.1. We allow that some – or, all – the ergodic classes have a finite number of states. A Markov chain on a finite state space is always geometrically ergodic. Still, we need to ensure that the condition (17) holds uniformly for all ergodic classes. In particular, if an ergodic class of the unperturbed Markov chain is finite, we use the restriction of the Lyapunov function $\mathcal{V}$ to that class and require the condition (17) to be satisfied with the same constants c and $\beta < 1$.

We now introduce the so-called "aggregated" Markov chain [3, 8, 10, 18, 25]. Define $V \in R^{\bar{E} \times E}$ to be a matrix whose I-th row corresponds to the invariant probability measure of the unperturbed Markov chain given that the process starts in the I-th ergodic class. Also we introduce a matrix $W \in R^{E \times \bar{E}}$, whose J-th column has ones in the components corresponding to the J-th ergodic class and zeros in the other components. Note that the columns of matrix W form a basis for the null space of $G^{(0)}$ and the rows of matrix V form a basis for the null space of the adjoint operator $G^{(0)*}$. Now define the generator of the aggregated Markov chain by $\Gamma = VG^{(1)}W \in R^{\bar{E} \times \bar{E}}$, or in the component form

$$\Gamma_{IJ} = \pi_I G^{(1)}_{IJ} \underline{1}_J, \quad I, J \in \bar{E},$$

where π_I is the invariant probability measure of the I-th ergodic class before perturbation, $G^{(1)}_{IJ}$ is the block I, J of the perturbation matrix and $\underline{1}_J$ is a vector of ones whose length is equal to $|E_J|$.

Assumption (S3): The aggregated Markov chain is irreducible and Lyapunov stable, i.e., there exist a strongly aperiodic state $\bar{\alpha} \in \bar{E}$, $P_{\bar{\alpha}\bar{\alpha}} > \bar{p}_0$, constants $\bar{\delta} < 1$, $\bar{b} < \infty$ and a Lyapunov function $\bar{\mathcal{V}}$, $\bar{\mathcal{V}}_I \geq 1, \forall I \in \bar{E}$, such that

$$(\mathcal{I} + \Gamma)\bar{\mathcal{V}} \leq \bar{\delta}\bar{\mathcal{V}} + \bar{b}\mathbf{I}_{\bar{\alpha}}.$$

Again invoking Theorem 1 we conclude from Assumption (S3) that the aggregated Markov chain is $\bar{\mathcal{V}}$-geometrically ergodic, that is,

$$||(\mathcal{I} + \Gamma)^k - \bar{\Pi}||_{\bar{\mathcal{V}}} \leq \bar{c}\bar{\beta}^k, \quad k = 0, 1, 2, ...$$

where $\bar{\Pi} = \underline{1}\bar{\pi}$ is the ergodic projection and $\bar{\pi}$ is the invariant probability measure of the aggregated Markov chain. Using again Lemma 4.1 from [27], we conclude that there exists a $\bar{\mathcal{V}}$-bounded deviation matrix of the aggregated Markov chain. Let us denote it by Φ.

Note that in the above we defined ν-norms for the ergodic classes of the unperturbed Markov chain and the aggregated Markov chain using Lyapunov functions. We can also define a ν-norm for the whole state space $E = \bigcup E_I$. Namely, for a pair of $I \in \bar{E}, i \in E_I$ let us define $\nu_{Ii} = \bar{\mathcal{V}}_I \mathcal{V}_i$. Since

$$\begin{aligned}
||P^{(0)k} - \Pi||_\nu &= \sup_{I \in \bar{E}, i \in E_I} \frac{1}{\nu_{Ii}} \sum_{j \in E_I} |(A_I^k - \Pi_I)_{ij}| \bar{\mathcal{V}}_I \mathcal{V}_j \\
&= \sup_{I \in \bar{E}, i \in E_I} \frac{1}{\mathcal{V}_i} \sum_{j \in E_I} |(A_I^k - \Pi_I)_{ij}| \mathcal{V}_j \\
&= \sup_{I \in \bar{E}} ||A_I^k - \Pi_I||_{\mathcal{V}} \\
&\leq c\beta^k, \quad k = 0, 1, 2, \dots .
\end{aligned}$$

there exists a ν-bounded deviation matrix H of the unperturbed Markov chain. Furthermore, we have the following bound for the norm of H

$$||H||_\nu \leq \frac{c}{1-\beta}.$$

It follows from Assumption (S1) that H has a block-diagonal structure

$$H = \begin{bmatrix} H_1 & 0 & \cdots \\ 0 & H_2 & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix},$$

where $H_I, I \in \bar{E}$ is the deviation matrix of each ergodic class of the unperturbed Markov chain. Finally, as in the case of regular perturbation, we make an assumption on the perturbation matrix.

Assumption (S4): The perturbation matrix $G^{(1)}$ is ν-bounded (for $\nu_{Ii} = \bar{\mathcal{V}}_I \mathcal{V}_i$, $I \in \bar{E}, i \in E_I$). Namely, $||G^{(1)}||_\nu \leq g_1$, where $g_1 > 0$ is some constant.

Remark 4.2. Fix some $\varepsilon_1, \varepsilon_2 \in (0, \bar{\varepsilon})$ and $\varepsilon_1 \neq \varepsilon_2$. Since $G^{(1)} = (P(\varepsilon_1) - P(\varepsilon_2))/(\varepsilon_1 - \varepsilon_2)$, a sufficient condition for $(S4)$ to hold is that $||P(\varepsilon_1)||_\nu$ and $||P(\varepsilon_2)||_\nu$ are finite.

Now we are able to formulate and prove the main result of this section.

Theorem 3. *Let Assumptions $(S1) - (S4)$ hold. Then, the perturbed Markov chain has a unique invariant probability measure $\pi(\varepsilon)$, which is an analytic function of ε*

$$\pi(\varepsilon) = \pi^{(0)} + \varepsilon\pi^{(1)} + \varepsilon^2\pi^{(2)} + \ldots,$$

for $0 < \varepsilon \leq \min\{\bar{\varepsilon},\ \frac{1-\beta}{g_1 c}\left(1 + \frac{g_1\bar{c}(c+1)}{1-\bar{\beta}}\right)^{-1}\}$, where

$$\pi^{(k)} = \pi^{(0)}U^k, \qquad \pi^{(0)} = \bar{\pi}V, \tag{18}$$

and the ν-bounded matrix U is given by

$$U = G^{(1)}H(\mathcal{I} + G^{(1)}W\Phi V). \tag{19}$$

Moreover, the invariant probability measure of the perturbed Markov chain can be calculated by the updating formula

$$\pi(\varepsilon) = \pi^{(0)}[\mathcal{I} - \varepsilon U]^{-1}, \quad 0 < \varepsilon \leq \min\{\bar{\varepsilon},\ \frac{1-\beta}{g_1 c}\left(1 + \frac{g_1\bar{c}(c+1)}{1-\bar{\beta}}\right)^{-1}\}. \tag{20}$$

Proof: From the construction of the aggregated Markov chain, one can see that the irreducibility of the perturbed Markov chain is equivalent to the irreducibility of the aggregated Markov chain. Hence, from Assumption (S3) we conclude that if there exists an invariant probability measure of the perturbed Markov chain ($0 < \varepsilon \leq \bar{\varepsilon}$), it is unique. As in the proof of Theorem 2, let us formally construct a power series for $\pi(\varepsilon)$, which satisfies (3), (4), and then show that it is absolutely convergent in some non-empty region. As in the case of regular perturbation, we have to solve the infinite system of matrix equations (12)-(15). The difference with the regular case is that the equations (12),(14) do not have a unique solution. From (12),(14) we can only conclude that $\pi^{(0)}$ is a linear combination of the stationary distributions corresponding to the ergodic classes of the unperturbed Markov chain. Namely,

$$\pi^{(0)} = c^{(0)}V, \tag{21}$$

for some vector $c^{(0)} \in R^{1\times\bar{E}}$. In order for the equation (13) for $k = 1$ to be feasible, $c^{(0)}$ should be chosen to satisfy the following condition

$$(-\pi^{(0)}G^{(1)})W = 0.$$

This condition is known as the Fredholm alternative in operator theory (see e.g., [15]). Substituting expression (21) into the above condition, we get

$$c^{(0)}VG^{(1)}W = 0$$

or, equivalently,

$$c^{(0)}\Gamma = 0. \tag{22}$$

We also substitute (21) into the normalization condition (14).

$$c^{(0)}V\underline{1} = 1$$

Since each row of V is a probability measure, we have

$$c^{(0)}\underline{1} = 1. \tag{23}$$

Since, according to Assumption (S3), the aggregated Markov process has a unique invariant probability measure $\bar{\pi}$, the equations (22) and (23) imply that $c^{(0)} = \bar{\pi}$. Thus, we obtain the second formula in (18).

Now we show that $\pi^{(k)} = \pi^{(k-1)}U$ for $k = 1, 2, \ldots$. One can write the general solution of equation (13) in the following form

$$\pi^{(k)} = c^{(k)}V + \pi^{(k-1)}G^{(1)}H, \tag{24}$$

where $c^{(k)} \in R^{1\times\bar{E}}$ is an arbitrary vector. The vector $c^{(k)}$ is determined from the feasibility condition of the next equation in (13). Namely,

$$(-\pi^{(k)}G^{(1)})W = 0,$$

$$c^{(k)}VG^{(1)}W + \pi^{(k-1)}G^{(1)}HG^{(1)}W = 0.$$

From the normalization condition (15) and property (9) of the deviation matrix H, we obtain

$$c^{(k)}V\underline{1} = c^{(k)}\underline{1} = 0.$$

Thus, we get a system of equations for $c^{(k)}$

$$c^{(k)}\Gamma = -\pi^{(k-1)}G^{(1)}HG^{(1)}W,$$

$$c^{(k)}\underline{1} = 0.$$

Since the aggregated Markov chain is irreducible and ν-geometrically ergodic, the above system has a unique solution which is given by the following explicit expression

$$c^{(k)} = \pi^{(k-1)}G^{(1)}HG^{(1)}W\Phi.$$

where Φ is the deviation matrix of the aggregated Markov chain. Combining the above expression with (24), we obtain the recursion $\pi^{(k+1)} = \pi^{(k)}U$ for $k = 1, 2, ...$ with

$$U = G^{(1)}H(\mathcal{I} + G^{(1)}W\Phi V).$$

Next let us show that the matrix U is ν-bounded, and consequently, the power series for $\pi(\varepsilon)$ has a non-zero radius of convergence. First we note that

$$||\Pi_J||_{\mathcal{V}} = ||\Pi_J - \mathcal{I} + \mathcal{I}||_{\mathcal{V}} \le ||\Pi_J - \mathcal{I}||_{\mathcal{V}} + ||\mathcal{I}||_{\mathcal{V}} \le c + 1.$$

Next we give a bound for the ν-norm of $W\Phi V$. The matrix $W\Phi V$ has the following structure

$$W\Phi V = \begin{bmatrix} \varphi_{11}\Pi_1 & \varphi_{12}\underline{1}\pi_2 & \cdots \\ \varphi_{21}\underline{1}\pi_1 & \varphi_{22}\Pi_2 & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix},$$

where $\underline{1}$ is a vector of ones with appropriate dimension. Hence, we can write

$$\begin{aligned} ||W\Phi V||_\nu &= \sup_{I\in\bar{E},i\in E_I} \frac{1}{\nu_{Ii}} \sum_{J\in\bar{E}} |\varphi_{IJ}|(\sum_{j\in E_J} \alpha_{Jj}\bar{\mathcal{V}}_J\mathcal{V}_j) \\ &\le \sup_{I\in\bar{E},i\in E_I} \frac{1}{\bar{\mathcal{V}}_I} \sum_{J\in\bar{E}} |\varphi_{IJ}|\bar{\mathcal{V}}_J||\Pi_J||_{\mathcal{V}} \\ &\le \sup_{I\in\bar{E},i\in E_I} \frac{1}{\bar{\mathcal{V}}_I} \sum_{J\in\bar{E}} |\varphi_{IJ}|\bar{\mathcal{V}}_J(c+1) \\ &\le ||\Phi||_{\bar{\mathcal{V}}}(c+1) \\ &\le \frac{\bar{c}(c+1)}{1-\bar{\beta}}. \end{aligned}$$

Thus, the radius of convergence of the power series (5) is greater than or equal to

$$\frac{1-\beta}{g_1 c}\left(1+\frac{g_1\bar{c}(c+1)}{1-\bar{\beta}}\right)^{-1}.$$

Finally, the updating formula (20) immediately follows from the fact that the coefficients $\pi^{(k)}, k=0,1,2,...$ form a geometric sequence.

□

In addition, we would like to note that for the computational purposes it is more convenient to write the matrix U in terms of blocks that correspond to the ergodic classes of the unperturbed Markov chain. Specifically, we have $U = \{U_{IJ}\}_{I,J\in\bar{E}}$, where U_{IJ} is given by

$$U_{IJ} = G^{(1)}_{IJ}H_J + \sum_{L\in\bar{E}} G^{(1)}_{IL}H_L \sum_{K\in\bar{E}} G^{(1)}_{LK}\varphi_{KJ}\underline{1}\pi_J, \qquad I,J\in\bar{E}. \qquad (25)$$

5. Applications to QBDs and Queueing Models

To illustrate our results on singular perturbation we now apply them for Quasi Birth and Death (QBD) processes; we refer to [19, 23] for general discussions on QBDs. After introducing some notation we develop our results in Section 5.1 without assuming a further particular structure for so-called phase transitions. Then, in Section 5.2 we elaborate on a particular queueing example.

An (inhomogeneous) QBD process is a Markov chain whose generator has the following structure

$$\mathcal{Q} = \begin{bmatrix} A_{01} - A_{02}^{(d)} & A_{02} & 0 & 0 & \cdots \\ A_{10} & A_{11} - A_1^{(d)} & A_{12} & 0 & \cdots \\ 0 & A_{20} & A_{21} - A_2^{(d)} & A_{22} & \cdots \\ 0 & 0 & A_{30} & A_{31} - A_3^{(d)} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix},$$

where $A_k := A_{k0}^{(d)} + A_{k2}^{(d)}$ and, for any $m \times n$ matrix M, we use $M^{(d)}$ to denote the m-dimensional square diagonal matrix with ith diagonal element equal to the i-th row sum of M. (Although, by definition, $A_k = A_k^{(d)}$, we shall use $A_k^{(d)}$ to emphasize that it is a diagonal matrix.) The QBD is called *homogeneous* if, for $k \geq 1$, $A_{k1} \equiv A_{11}$ and $A_{k2} \equiv A_{12}$, and for $k \geq 2$ $A_{k0} \equiv A_{10}$. In that case the A_{20}, A_{11} and A_{12} are square matrices of the same dimension (not necessarily finite). The square matrix A_{01} may have a different dimension, while A_{02} and A_{10} need not be square.

The structure corresponds to a partition of states into so-called *levels*. The i-th *block row* of $\mathcal{Q}$ corresponds to transitions originating from states in the i-th level. The states within a given level are commonly called *phases*. The rates of transitions that do not involve a change of level are contained by the matrices

The results on regular perturbation can also be used for Infinitesimal Perturbation Analysis [6, 12, 13, 14, 26].

A_{k1} (transitions take place within level k). Transitions among levels are only possible from level k to either level $k-1$ or level $k+1$, the rates of which are gathered in A_{k0} and A_{k2}, respectively. We emphasize that we allow an infinite number of phases within a level.

We shall be interested in the case where transitions among levels are much less frequent than transitions among the states inside the same level. Instead of $\mathcal{Q}$ we will therefore consider the generator $G(\varepsilon) = G^{(0)} + \varepsilon G^{(1)}$, where

$$G^{(0)} = \begin{bmatrix} A_{01} & 0 & 0 & \cdots \\ 0 & A_{11} & 0 & \cdots \\ 0 & 0 & A_{21} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}, G^{(1)} = \begin{bmatrix} -A_{02}^{(d)} & A_{02} & 0 & \cdots \\ A_{10} & -A_{1}^{(d)} & A_{12} & \cdots \\ 0 & A_{20} & -A_{2}^{(d)} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}. \tag{26}$$

Remark 5.1. Note that, alternatively, it would also be natural to investigate the case when transitions inside the levels are much less frequent than among levels, i.e., when $G(\varepsilon) = \varepsilon G^{(0)} + G^{(1)}$ for the same matrices $G^{(0)}$ and $G^{(1)}$. However, in that case the unperturbed chain is again a QBD process itself (with no transitions within levels), making the analysis of the unperturbed chain as involved as the (original) perturbed chain, unless a special structure within levels is assumed. A special structure arising in many applications is that where the matrices A_{k0} and A_{k2} are (square) diagonal matrices. This implies that transitions among levels do not involve a change in phase. For instance, in a two-queue system studied in [1, 24] the level index counts the number of customers in the first queue and the phase index that in the second queue. However, interchanging the role of the levels and phases (the first and the second queue), the same structure as in (26) can be obtained again. We elaborate on this example in Section 5.2

5.1. General Phase Transitions

For the unperturbed chain, the ergodic classes correspond to the levels of the QBD process. We assume that all states inside the same level are communicating. Hence, Assumption (S1) is satisfied. Let Assumption (S2) hold as well. As before, denote the stationary distribution of the I-th ergodic class (level) with the vector π_I, $I = 0, 1, \dots$. In particular, if $G^{(0)}$ is homogeneous beyond level 1, the uniform Lyapunov stability assumption is equivalent with (regular) Lyapunov stability. In that case we also have that $\pi_I \equiv \pi_1$ for all levels $I \geq 1$. In general, the structure of the matrix V is given by

$$V = \begin{bmatrix} \pi_0 & 0 & 0 & 0 & \cdots \\ 0 & \pi_1 & 0 & 0 & \cdots \\ 0 & 0 & \pi_2 & 0 & \cdots \\ 0 & 0 & 0 & \pi_3 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix},$$

which gives

$$\Gamma = \begin{bmatrix} -\pi_0 A_{02}^{(d)} \underline{1} & \pi_0 A_{02} \underline{1} & 0 & 0 & \cdots \\ \pi_1 A_{10} \underline{1} & -\pi_1 A_1^{(d)} \underline{1} & \pi_1 A_{12} \underline{1} & 0 & \cdots \\ 0 & \pi_2 A_{20} \underline{1} & -\pi_2 A_2^{(d)} \underline{1} & \pi_2 A_{22} \underline{1} & \cdots \\ 0 & 0 & \pi_3 A_{30} \underline{1} & -\pi_3 A_3^{(d)} \underline{1} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}. \tag{27}$$

We note that Γ is the generator of an ordinary (one dimensional, inhomogeneous) birth and death process on states $I \in \{0, 1, 2, \dots\}$ with birth rate $\bar{\lambda}_I := \pi_I A_{I2} \underline{1}$, for $I \geq 0$, and death rate $\bar{\mu}_I := \pi_I A_{I0} \underline{1}$ for $I \geq 1$. We have that the invariant distribution of the aggregated chain is given by

$$\bar{\pi}_I = \frac{\prod_{J=0}^{I-1} \frac{\bar{\lambda}_J}{\bar{\mu}_J}}{\sum_{K=0}^{\infty} \prod_{J=0}^{K-1} \frac{\bar{\lambda}_J}{\bar{\mu}_J}}, \qquad I \geq 0,$$

whenever the denominator is well defined (the empty product is set equal to 1). This inhomogeneous birth and death process can be shown to be Lyapunov stable if there exists an $\bar{r}$ such that $\frac{\bar{\lambda}_I}{\bar{\mu}_{I+1}} < \bar{r} < 1$, for all $I \geq 0$. It is beyond the scope in this paper to work out all details for this general case; instead we shall focus on homogeneous QBD processes. It is worthwhile to note that in Section 5.2 below we analyze an *inhomogeneous* QBD process that gives rise to an aggregated process that is a *homogeneous* birth and death process.

Homogeneous QBD

If the underlying QBD process is homogeneous then $\bar{\lambda}_I \equiv \bar{\lambda}$ for $I \geq 1$ and $\bar{\mu}_I \equiv \bar{\mu}$ for $I \geq 2$. In addition we assume that also $\bar{\lambda}_0 = \bar{\lambda}$ and $\bar{\mu}_1 = \bar{\mu}$, which is the case in the examples of the next subsection. (The analysis does not change essentially otherwise, but we do not consider this since it is not necessary for the queueing examples.) The generator of the aggregated chain coincides with that of the M/M/1 queue with arrival rate $\bar{\lambda}$ and service rate $\bar{\mu}$. We have that the aggregated chain is ergodic if and only if $\bar{\rho} := \bar{\lambda}/\bar{\mu} < 1$. In this case verification of Assumption (S3) is also straightforward. Let us choose $\bar{\alpha} = 0$. We need to find a Lyapunov function $\bar{\mathcal{V}}$ and constants $\bar{\delta} \in (0,1)$ and $\bar{b} \geq 0$ such that

$$\begin{aligned}
\bar{\mu}\bar{\mathcal{V}}_{I-1} + (1-\bar{\lambda}-\bar{\mu})\bar{\mathcal{V}}_I + \bar{\lambda}\bar{\mathcal{V}}_{I+1} &\leq \bar{\delta}\bar{\mathcal{V}}_I, \qquad I \geq 1, \\
(1-\bar{\lambda})\bar{\mathcal{V}}_0 + \bar{\lambda}\bar{\mathcal{V}}_1 &\leq \bar{\delta}\bar{\mathcal{V}}_0 + \bar{b}.
\end{aligned}$$

We aim at solving these equations with equality. Introducing the generating function $\bar{\mathcal{V}}(z) = \sum_{I=0}^{\infty} z^I \bar{\mathcal{V}}_I$ (later we verify that this can be justified) the above equations translate into

$$(\bar{\mu}z^2 + (1-\bar{\delta}-\bar{\mu}-\bar{\lambda})z + \bar{\lambda})\bar{\mathcal{V}}(z) = \bar{b}z + (\bar{\lambda}-\bar{\mu}z)\bar{\mathcal{V}}_0.$$

Concentrating on the kernel $(\bar{\mu}z^2 + (1-\bar{\delta}-\bar{\mu}-\bar{\lambda})z + \bar{\lambda}$, we see that we have two real roots for z if $\bar{\delta} \geq 1-(\bar{\mu}-\bar{\lambda})^2$; this quantity is indeed in the interval

$(0,1)$. It is convenient to take $\bar{\delta} = 1 - (\bar{\mu} - \bar{\lambda})^2$. The corresponding root (with multiplicity 2) is $z = \sqrt{\bar{\rho}}$ and if we choose $\bar{b} = \bar{\mu}(1 - \sqrt{\bar{\rho}})$ one of the two roots cancels out, leaving us with

$$\bar{\mathcal{V}}(z) = \frac{\bar{\mathcal{V}}_0}{1 - z/\sqrt{\bar{\rho}}}.$$

Finally, choosing $\bar{\mathcal{V}}_0 = 1$ we have that Assumption (S3) is satisfied with $\bar{\mathcal{V}}_I = \left(\sqrt{\frac{1}{\bar{\rho}}}\right)^I$.

Note also that, in this case, if $||A_{02}||_{\mathcal{V}}$, $||A_{10}||_{\mathcal{V}}$, $||A_{12}||_{\mathcal{V}}$ and $\max_i\{(A_{02}^{(d)})_{ii}\}$, $\max_i\{(A_1^{(d)})_{ii}\}$ are finite, then Assumption (S4) is satisfied. This is a consequence of the following bound

$$\begin{aligned} ||G^{(1)}||_\nu \quad \leq \quad \max \Bigg\{ & \max_i\{(A_{02}^{(d)})_{ii}\} + \sqrt{\frac{1}{\bar{\rho}}}||A_{02}||_{\mathcal{V}}, \\ & \sqrt{\bar{\rho}}||A_{10}||_{\mathcal{V}} + \max_i\{(A_1^{(d)})_{ii}\} + \sqrt{\frac{1}{\bar{\rho}}}||A_{12}||_{\mathcal{V}} \Bigg\}. \end{aligned}$$

The deviation matrix for the M/M/1 queue was determined in [17]:

$$\Phi_{I,J} = D(I,J,\bar{\rho},\bar{\mu}) := \frac{\bar{\rho}^{\max\{J-I,0\}} - (I+J+1)(1-\bar{\rho})\bar{\rho}^J}{\bar{\mu}(1-\bar{\rho})}. \qquad (28)$$

The above enables us to apply Theorem 3 once the invariant distributions and the deviation matrices of the levels $I = 0$ and $I \geq 1$ have been determined. This is the task of the following section where we elaborate on our results for a particular example.

5.2. Queueing Examples

We now focus on a particular queueing model and study two cases, one giving rise to a homogeneous QBD process and the other corresponding to an inhomogeneous QBD process.

Priority queue with fast dynamics Let us study a system of two M/M/1 queues with strict priorities. Customers arrive at the first queue according to a Poisson process with rate λ and are served at rate μ. The arrival rate and service rate in the second queue may both depend on the number of customers in the first queue (denoted by $X(t)$). If $X(t) = i$ then customers arrive at the second queue as a Poisson process of intensity $\varepsilon\lambda_i$ and customers depart from queue 2 (if not empty) at rate $\varepsilon\mu_i$. We denote the number of customers in the second queue with $Y(t)$. The generator of $(X(t), Y(t))$ can be written as a QBD process by letting $X(t)$ correspond to the phase of the process and $Y(t)$ be the level. Letting $\varepsilon \to 0$ corresponds to slow dynamics in the second queue, i.e., with slow transitions among levels.

The blocks of (26) are given by $A_{j0} = \text{diag}\{\mu_0, \mu_1, \dots\}$, $A_{j2} = \text{diag}\{\lambda_0, \lambda_1, \dots\}$ and A_{j1} is the generator of an ordinary M/M/1 queue with arrival rate λ and service rate μ. Since this corresponds to a homogeneous QBD, in view of the results in Section 5.1 we can apply Theorem 3 once we have determined the invariant distributions and deviation matrices of the levels (ergodic classes), which all correspond to the ordinary M/M/1 queue describing $X(t)$. The invariant distribution is well known: $\pi_{I,i} \equiv (1-\rho)\rho^i$, for all levels I and phases i, with $\rho = \lambda/\mu$. As for (28) we can use the results of [17], giving the deviation matrix $H_{i,j} = D(i, j, \rho, \mu)$. We emphasize that in this case the aggregated chain and the ergodic classes of the unperturbed chain correspond to ordinary M/M/1 queues.

Priority queue with slow dynamics Alternatively, the dynamics of the first queue could be slow. Let the arrival rate and service rate at the first queue be $\varepsilon\lambda$ and $\varepsilon\mu$, respectively, and the arrival and service rates in queue 2 be λ_i and μ_i, respectively, when $X(t) = i$. As in the above example, $X(t)$ is the number of customers in the first queue. We again have a QBD process if we let $X(t)$ correspond to the level and $Y(t)$ be the phase of the process. The block matrices in (26) are now given by $A_{k0} = \text{diag}\{\mu, \mu, \dots\}$, $A_{k2} = \text{diag}\{\lambda, \lambda, \dots\}$ and A_{k1} is the generator of an ordinary M/M/1 queue with arrival rate λ_k and service rate μ_k. Thus, the QBD process is not homogeneous.

In the unperturbed chain we have infinitely many classes, each (again) corresponding with a level of the QBD process. In the I-th level the dynamics is that of the ordinary M/M/1 queue with arrival rate λ_I and service rate μ_I. We assume that $\rho_I := \lambda_I/\mu_I < r < 1$, for some $r \in (0,1)$. This implies that Assumption (S2) is satisfied with $\alpha_I = 0$, $\mathcal{V}_j = \left(\sqrt{\frac{1}{r}}\right)^j$, $\delta = 1 - \left(\sqrt{\mu_I} - \sqrt{\lambda_I}\right)^2$ and $b \geq \sqrt{\frac{1}{r}}$. The deviation matrix of the ergodic classes is again that of an M/M/1 queue: $H^{(I)}_{i,j} = D(i,j,\rho_I,\mu_I)$.

Note that $G^{(1)}$ has a simple structure, since $A_{k0} = \mu\mathcal{I}$ and $A_{k2} = \lambda\mathcal{I}$. Assumption S4 is always satisfied, since in this example we have $||G^{(1)}||_\nu \leq \lambda+\mu+2\sqrt{\lambda\mu}$. Finally, the deviation matrix of the aggregated chain is again that of the M/M/1 queue with arrival rate λ and service rate μ: $\Phi_{I,J} = D(I,J,\rho,\mu)$, where as before $\rho = \lambda/\mu$.

We again note that both the aggregated chain and the ergodic classes of the unperturbed chain correspond to ordinary M/M/1 queues. Different from the case when queue 2 has slow dynamics, however, the ergodic classes of the unperturbed chain are not identical.

References

[1] Altman, E., Artiges, D., Traore, K., (1999) On the Integration of Best-Effort and Guaranteed Performance Services, *European Transactions on Telecommunications*, Special Issue on Architectures, Protocols and Quality of Service for the Internet of the Future, Issue No. 2, February-

---

Here we concentrate on the case where all classes of the unperturbed chain are ergodic. In this particular example it makes sense in some cases to allow $\rho_I > 1$ for some $\mathcal{I}$, i.e., when the phase-process requires level transitions (however infrequent) to guarantee ergodicity; see for instance [24].

March 1999. See also INRIA Research Report No. 3222 , July 1997.

[2] Anisimov, V.V., (2001). Asymptotic analysis of some type optimal maintenance policies in multicomponent systems with Markov switches, in: *Proceedings of the 10-th International Symposium on Applied Stochastic Models and Data Analysis*, Compiegne, France, pp.112-117.

[3] Avrachenkov, K.E., (1999). *Analytic Perturbation Theory and its Applications*, Ph.D. Thesis, University of South Australia. Available electronically at
http://www-sop.inria.fr/mistral/personnel/K.Avrachenkov/moi.html

[4] Avrachenkov, K.E., Filar, J.A., Haviv, M., (2002). Singular perturbations of Markov chains and decision processes, in: Feinberg, E.A., Shwartz, A., (Eds.), *Handbook of Markov Decision Processes: Methods and Applications*, Kluwer Academic Publishers, Boston, pp.113-150.

[5] Bielecki, T.R., Stettner, L., (1998). Ergodic control of singularly perturbed Markov process in discrete time with general state and compact action spaces, *Appl. Math. Optimization*, 38 (3), pp.261-281.

[6] Cao, X.-R., Chen, H.-F., (1997). Perturbation realization, potentials, and sensitivity analysis of Markov processes, *IEEE Trans. Auto. Contr.* 42 (10), pp.1382-1393.

[7] Chang, C.-S., Nelson, R., (1993). Perturbation analysis of the M/M/1 queue in a Markovian environment via the matrix-geometric method, *Commun. Statist. – Stochastic Models* 9 (2), pp.233-246.

[8] Courtois, P.J., (1977). *Decomposability: queueing and computer system applications*, Academic Press, New York.

[9] Dekker, R., Hordijk, A., Spieksma, F.M., (1994). On the relation between recurrence and ergodicity properties in denumerable Markov decision chains, *Math. Oper. Res.* 19, pp.539-559.

[10] Delebecque, F., (1983). A reduction process for perturbed Markov chain, *SIAM J. Appl. Math.* 43, pp.325-350.

[11] Gelenbe, E., Rosenberg, C., (1990). Queues with slowly varying arrival and service processes, *Management Science* 36 (8), pp.928-937.

[12] Glasserman, P., (1991). *Gradient estimation via perturbation analysis*, Kluwer Academic Publishers, Boston.

[13] Heidergott, B., Cao, X.-R., (2002). A note on the relation between weak derivatives and perturbation realization, *IEEE Trans. Auto. Contr.* 47 (7), pp.1112-1115.

[14] Heidergott, B., Hordijk, A., Weisshaupt, H., (2002). Measure-valued differentiation for stationary Markov chains, *EURANDOM Research Report 2002-27*, available electronically at http://staff.feweb.vu.nl/bheidergott/default.htm

[15] Kolmogorov, A.N., Fomin, S.V., (1999). *Elements of the theory of functions and functional analysis*, Dover Publications, New York.

[16] Kontoyiannis, I., Meyn, S.P., (2003) Spectral theory and limit theory for geometrically ergodic Markov processes, *Annals of Appl. Prob.* 13, pp.304-362.

[17] Koole, G., (1998) The deviation matrix of the $M/M/1/\infty$ and $M/M/1/N$ queue with applications to controlled queueing models, in *Proceedings of IEEE CDC'98*, Tampa, pp.56-59.

[18] Korolyuk, V.S., Turbin, A.F., (1993). *Mathematical foundations of the state lumping of large systems*, Kluwer Academic Publishers, Dordrecht, Boston.

[19] Latouche, G., Ramaswami, V., (1999). *Introduction to matrix analytic methods in stochastic modeling*, ASA-SIAM, Philadelphia.

[20] Latouche, G., Schweitzer, P.G., (1995). A Markov modulated, nearly completely decomposable M/M/1 queue, in: Stewart, W.J., (Ed.), *Computations with Markov chains*, Kluwer.

[21] Meyn, S.P., Tweedie, R.L., (1993). *Markov chains and stochastic stability*, Springer-Verlag, London.

[22] Meyn, S.P., Tweedie, R.L., (1994). Computable bounds for geometric convergence rates of Markov chains, it Annals of Appl. Prob. 4, pp.981-1012.

[23] Neuts, M.F., (1981). *Matrix-geometric solutions in stochastic models: An algorithmic approach*, The Johns Hopkins University Press, Baltimore.

[24] Núñez Queija, R., (1999). *Processor-sharing models for Integrated-Services Networks*, Ph.D. Thesis, Eindhoven University of Technology, ISBN 90-646-5667-8 (also available from http://www.cwi.nl/∼sindo).

[25] Pervozvanski, A.A., Gaitsgori, V.G., (1988). Theory of suboptimal decisions, Kluwer Academic Publishers, Dordrecht, The Netherlands, (Translation from Russian: Decomposition, aggregation and approximate optimization, Nauka, Moscow, 1979.)

[26] Pflug, G., (1992). Gradient estimates for the performance of Markov chains and discrete event processes, *Annals of Oper. Res.* 39, pp.173-194.

[27] Spieksma, F.M., (1990). *Geometrically ergodic Markov chains and the optimal control of queues*, Ph.D. thesis, Leiden University (available from the author upon request).

[28] Spieksma, F.M., Tweedie, R.L., (1994). Strengthening ergodicity to geometric ergodicity for Markov chains, *Comm. Statist. Stochastic Models* 10 (1), pp.45-74.

[29] Yin, G., Zhang, Q., (1998). *Continuous-time Markov chains and applications: A singular perturbation approach*, Series: Applications of Mathematics, 37, Springer-Verlag, New York.

[30] Yin, G., Zhang, H., (2002). Countable-state-space Markov chains with two time scales and applications to queueing systems, *Adv. Appl. Prob.* 34, pp.662-688.