Mathematical
Inequalities
& Applications
Volume 8, Number 1 (2005), 159–168

ERGODICITY COEFFICIENT AND PERTURBATION BOUNDS FOR CONTINUOUS–TIME MARKOV CHAINS

A. YU. MITROPHANOV

(*communicated by I. Pinelis*)

Abstract. For the distribution of a finite, homogeneous, continuous-time Markov chain, we derive perturbation bounds in terms of the ergodicity coefficient of the transition probability matrix. Our perturbation bounds improve upon the known results. We give sensitivity bounds for the coefficient of ergodicity, providing a sufficient condition for the uniqueness of the stationary distribution of the perturbed Markov chain. These results are used to obtain estimates of the speed of convergence for singularly perturbed Markov chains.

1. Introduction

In many areas of application of Markov chains, such as physics and chemistry, the numerical values of some parameters of a model chain must be taken from experiment. The experimental data surely are not absolutely accurate, and we are interested in knowing how the uncertainties in the parameter values affect the distribution of the chain under study. We may also want to solve the inverse problem: how accurate should our experimental data be to guarantee a given accuracy of determination of the distribution vector? This can be accomplished if we have computable estimates of sensitivity of the distribution vector to changes in the parameters.

We shall investigate the question of sensitivity to perturbations in the following setting. Consider two homogeneous, continuous-time Markov chains, $X = \{X(t), t \geqslant 0\}$ and $\tilde{X} = \{\tilde{X}(t), t \geqslant 0\}$, with finite state space $\mathscr{S} = \{1, \ldots, m\}$ ($m \geqslant 2$) and generators $\mathbf{Q} = (q_{ij})$ and $\tilde{\mathbf{Q}} = (\tilde{q}_{ij})$, respectively ($\mathbf{Q} \neq \tilde{\mathbf{Q}}$). Let $\mathbf{p}(t) = (p_i(t))$ and $\tilde{\mathbf{p}}(t) = (\tilde{p}_i(t))$ be the respective state probability vectors of X and $\tilde{X}$ (we regard vectors as row vectors). Our goal is to estimate the change in the distribution, $\mathbf{z}(t) := \tilde{\mathbf{p}}(t) - \mathbf{p}(t)$, at a time $t > 0$, given the change in the generator, $\mathbf{E} := \tilde{\mathbf{Q}} - \mathbf{Q}$, and the change in the initial distribution.

For arbitrary X and $\tilde{X}$, the following inequality holds:

$$\|\tilde{\mathbf{p}}(t) - \mathbf{p}(t)\| \leqslant \|\tilde{\mathbf{p}}(0) - \mathbf{p}(0)\| + T\|\mathbf{E}\|, \qquad 0 \leqslant t \leqslant T < \infty, \tag{1.1}$$

where $\|\cdot\|$ denotes the l_1-norm (absolute entry sum) for vectors and the corresponding induced norm (maximum absolute row sum) for matrices; see [8, 10, 15]. It was shown

Mathematics subject classification (2000): 60J27, 60J35, 34D10, 34D15.

Key words and phrases: Markov chain, ergodicity coefficient, variation distance, perturbation bounds, sensitivity analysis, singular perturbations.

© ELEMENT, Zagreb
Paper No. 08-15/8.00$

in [9] that if X has a unique stationary distribution, then the inequality in (1.1) is strict. The uniqueness of the stationary distribution also makes it possible to obtain bounds that are uniform over infinite time intervals. We study such bounds in this paper, and from now on we assume that X does have a unique stationary distribution, π.

The two main approaches to bounding sensitivity to perturbations on infinite time intervals are:

(a) to obtain sensitivity bounds using exponential bounds on the speed of convergence to stationarity [1, 5, 9, 15, 16];

(b) to obtain sensitivity bounds in terms of the l_1 ergodicity coefficient of the transition probability matrix, $\mathbf{P}(t)$ [2, 3, 4].

When used as a modeling tool, a Markov chain is often solved numerically by applying standard methods. In such cases, it may be desirable to analyze the sensitivity to perturbations using the knowledge of $\mathbf{P}(t)$ for some $t > 0$ (exponential convergence bounds often are not easy to obtain). This justifies the development of the approach (b), which is the main purpose of our paper. In Section 2 we obtain new perturbation bounds improving upon the results of Anisimov [3]. In Section 3 we study perturbations of the ergodicity coefficient of $\mathbf{P}(t)$, which allows us to analyze the sensitivity of the perturbed chain. Section 4 is devoted to applications of our results to singularly perturbed Markov chains; for a background on such chains, see [6, 7, 13, 14].

2. The ergodicity coefficient and sensitivity with respect to perturbations

In this section we obtain upper bounds on the l_1-distance between the distributions of X and $\tilde{X}$. Before proceeding, we note that l_1-distance has a clear probabilistic interpretation: for all random variables V and $\tilde{V}$ taking values in $\mathscr{S}$ with the respective distribution vectors $\mathbf{p}$ and $\tilde{\mathbf{p}}$, the quantity $\|\mathbf{p} - \tilde{\mathbf{p}}\|$ is twice the variation distance between the distributions of V and $\tilde{V}$:

$$\|\mathbf{p} - \tilde{\mathbf{p}}\| = 2 \max_{\mathscr{A} \subseteq \mathscr{S}} |P\{V \in \mathscr{A}\} - P\{\tilde{V} \in \mathscr{A}\}| \tag{2.1}$$

(some authors define the variation distance as being equal to the right-hand side of (2.1)).

For a square real matrix $\mathbf{A} = (a_{ij})$ of order m, define the l_1 coefficient of ergodicity by

$$\tau(\mathbf{A}) = \sup_{\substack{\|\mathbf{v}\|=1 \\ \mathbf{v}\mathbf{e}'=0}} \|\mathbf{v}\mathbf{A}\| = \tfrac{1}{2} \max_{i,j \in \mathscr{S}} \sum_{k \in \mathscr{S}} |a_{ik} - a_{jk}|,$$

where $\mathbf{e} = (1, 1, \dots, 1)$ and $'$ denotes transpose. For a background on ergodicity coefficients, see [11, 12]. Set $\beta_t = \tau(\mathbf{P}(t))$, $t \geqslant 0$. The uniqueness of the stationary distribution of X implies that $\beta_s < 1$ for all $s > 0$; this property will be used in the sequel. The magnitude of β_s also allows us to judge whether or not X has a unique stationary distribution, as is shown by Proposition 2.1.

PROPOSITION 2.1. *The following three statements are equivalent:*

(a) $\beta_s < 1$ *for all* $s > 0$*;*

(b) *there exists* $s > 0$ *such that* $\beta_s < 1$*;*
(c) *the chain* X *has a unique stationary distribution.*

The next theorem gives perturbation bounds for the distribution of X in terms of β_s.

THEOREM 2.1. *If* $0 < s < t$*, then*

$$\|\mathbf{z}(t)\| < \beta_s^{\lfloor t/s \rfloor} \|\mathbf{z}(0)\| + \left(\frac{s\left(1 - \beta_s^{\lfloor t/s \rfloor}\right)}{1 - \beta_s} + \beta_s^{\lfloor t/s \rfloor} (t - s\lfloor t/s \rfloor) \right) \|\mathbf{E}\|. \tag{2.2}$$

For all $s > 0$*,*

$$\sup_{t \geqslant 0} \|\mathbf{z}(t)\| < \|\mathbf{z}(0)\| + \frac{s\|\mathbf{E}\|}{1 - \beta_s}; \tag{2.3}$$

if $\tilde{\pi}$ *is a stationary distribution of* $\tilde{X}$*, then*

$$\| \tilde{\pi} - \pi \| < \frac{s\|\mathbf{E}\|}{1 - \beta_s}. \tag{2.4}$$

Proof. The vectors $\tilde{\mathbf{p}}(t)$ and $\mathbf{p}(t)$ satisfy the equations

$$\mathrm{d}\tilde{\mathbf{p}}(t)/\mathrm{d}t = \tilde{\mathbf{p}}(t)\tilde{\mathbf{Q}}, \qquad \mathrm{d}\mathbf{p}(t)/\mathrm{d}t = \mathbf{p}(t)\mathbf{Q}, \qquad t \geqslant 0,$$

which implies that

$$d\mathbf{z}(t)/\mathrm{d}t = \mathbf{z}(t)\mathbf{Q} + \tilde{\mathbf{p}}(t)\mathbf{E}, \qquad \mathbf{z}(0) = \tilde{\mathbf{p}}(0) - \mathbf{p}(0).$$

The solution to this initial-value problem has the form

$$\mathbf{z}(t) = \mathbf{z}(0)\mathbf{P}(t) + \int_0^t \tilde{\mathbf{p}}(u)\mathbf{E}\mathbf{P}(t-u)\mathrm{d}u,$$

which yields

$$\|\mathbf{z}(t)\| \leqslant \|\mathbf{z}(0)\mathbf{P}(t)\| + \int_0^t \|\tilde{\mathbf{p}}(t-u)\mathbf{E}\mathbf{P}(u)\|\mathrm{d}u. \tag{2.5}$$

Since the matrices $\mathbf{Q}$ and $\tilde{\mathbf{Q}}$ have zero row sums, the same holds for $\mathbf{E}$, which implies that $\tilde{\mathbf{p}}(t)\mathbf{E}\mathbf{e}' \equiv 0$. We have

$$\|\tilde{\mathbf{p}}(t-u)\mathbf{E}\mathbf{P}(u)\| \leqslant \left\| \frac{\tilde{\mathbf{p}}(t-u)\mathbf{E}}{\|\tilde{\mathbf{p}}(t-u)\mathbf{E}\|}\mathbf{P}(u) \right\| \|\mathbf{E}\| \quad \text{for} \quad \tilde{\mathbf{p}}(t-u)\mathbf{E} \neq 0;$$

from the above inequality and the definition of $\tau(\cdot)$ we obtain that $\|\tilde{\mathbf{p}}(t-u)\mathbf{E}\mathbf{P}(u)\| \leqslant \beta_u\|\mathbf{E}\|$. Similarly, $\|\mathbf{z}(0)\mathbf{P}(t)\| \leqslant \beta_t\|\mathbf{z}(0)\|$. This, together with (2.5), gives

$$\|\mathbf{z}(t)\| \leqslant \beta_t\|\mathbf{z}(0)\| + \|\mathbf{E}\| \int_0^t \beta_u \mathrm{d}u, \qquad t \geqslant 0 \tag{2.6}$$

(this inequality was first proved in [9]).

For all stochastic matrices $\mathbf{P}_1$ and $\mathbf{P}_2$, $\tau(\mathbf{P}_1\mathbf{P}_2) \leqslant \tau(\mathbf{P}_1)\tau(\mathbf{P}_2)$ and $\tau(\mathbf{P}_1) \leqslant 1$. This gives $\beta_t \leqslant \beta_s^{\lfloor t/s \rfloor}$ for all $s > 0$, $t \geqslant 0$. If $t > s$ and t/s is not an integer, then $\beta_s < 1$ implies that $\beta_t < \beta_s^{\lfloor t/s \rfloor}$. Thus, for $t > s$, we have

$$\int_0^t \beta_u \mathrm{d}u < \int_0^t \beta_s^{\lfloor u/s \rfloor} \mathrm{d}u = s\left(1 + \beta_s + \beta_s^2 + \cdots + \beta_s^{\lfloor t/s \rfloor - 1}\right) + \beta_s^{\lfloor t/s \rfloor}(t - s\lfloor t/s \rfloor)$$
$$= \frac{s\left(1 - \beta_s^{\lfloor t/s \rfloor}\right)}{1 - \beta_s} + \beta_s^{\lfloor t/s \rfloor}(t - s\lfloor t/s \rfloor).$$

This, together with (2.6), gives (2.2). We also have

$$\sup_{t \geqslant 0} \|\mathbf{z}(t)\| \leqslant \|\mathbf{z}(0)\| + \|\mathbf{E}\| \int_0^\infty \beta_u \mathrm{d}u,$$
$$\int_0^\infty \beta_u \mathrm{d}u < \int_0^\infty \beta_s^{\lfloor u/s \rfloor} \mathrm{d}u = s\left(1 + \beta_s + \beta_s^2 + \cdots\right) = \frac{s}{1 - \beta_s}, \qquad (2.7)$$

hence (2.3) follows.

Setting $\tilde{\mathbf{p}}(0) = \tilde{\pi}$ and passing to the limit as $t \to \infty$ in (2.6), we obtain that

$$\|\tilde{\pi} - \pi\| \leqslant \|\mathbf{E}\| \int_0^\infty \beta_u \mathrm{d}u.$$

This inequality and (2.7) prove (2.4). $\square$

Now we compare our results with those of Anisimov [3]. It follows from Theorem A.1, Lemma A.1 and (2.1) that if there exists such $s > 0$ that

$$\beta_s < 1 \qquad \text{and} \qquad s \max_{i \in \mathscr{S}} \sum_{j \in \mathscr{S} \setminus \{i\}} |q_{ij} - \tilde{q}_{ij}| \leqslant \varepsilon,$$

then

$$\tfrac{1}{2}\|\tilde{\mathbf{P}}(t) - \mathbf{P}(t)\| \leqslant \frac{\varepsilon}{1 - \beta_s}\left(1 - \beta_s^{\lfloor t/s \rfloor + 1}\right),$$

where $\tilde{\mathbf{P}}(t) = \exp(t\tilde{\mathbf{Q}})$. Using this inequality, together with Proposition 2.1 and the triangle inequality, we obtain the following bound: for arbitrary $s > 0$,

$$\|\mathbf{z}(t)\| \leqslant \|\mathbf{z}(0)\mathbf{P}(t)\| + \|\tilde{\mathbf{P}}(t) - \mathbf{P}(t)\|$$
$$\leqslant \|\mathbf{z}(0)\|\beta_s^{\lfloor t/s \rfloor} + \frac{2s\|\mathbf{E}_0\|}{1 - \beta_s}\left(1 - \beta_s^{\lfloor t/s \rfloor + 1}\right), \quad s < t, \qquad (2.8)$$

where $\mathbf{E}_0$ is the matrix obtained from $\mathbf{E}$ by replacing its diagonal entries with zeros. It is easy to see that

$$\frac{s\left(1 - \beta_s^{\lfloor t/s \rfloor}\right)}{1 - \beta_s} + \beta_s^{\lfloor t/s \rfloor}(t - s\lfloor t/s \rfloor) < \frac{s\left(1 - \beta_s^{\lfloor t/s \rfloor + 1}\right)}{1 - \beta_s}, \qquad 0 < s < t.$$

This inequality and Proposition 2.2 below show that our bound (2.2) is sharper than (2.8).

PROPOSITION 2.2. *The following inequality holds:*

$$\|\mathbf{E}\| \leqslant 2\|\mathbf{E}_0\|;$$

in this inequality an equality is attained if and only if in every row of $\mathbf{E}$ *all off-diagonal non-zero entries are of the same sign. In this case,* $\|\mathbf{E}\| = 2\max_{i\in\mathscr{S}} |e_{ii}|$, *where* e_{ij} *are the entries of* $\mathbf{E}$.

Proof. Since $\sum_{j\in\mathscr{S}} q_{ij} = \sum_{j\in\mathscr{S}} \tilde{q}_{ij} = 0$, $\sum_{j\in\mathscr{S}} e_{ij} = 0$ and $|e_{ii}| \leqslant \sum_{j\in\mathscr{S}\setminus\{i\}} |e_{ij}|$ for all $i \in \mathscr{S}$. This implies that $\|\mathbf{E}\| \leqslant 2\|\mathbf{E}_0\|$. The equalities $|e_{ii}| = \sum_{j\in\mathscr{S}\setminus\{i\}} |e_{ij}|$, $i \in \mathscr{S}$, hold if and only if in every row of $\mathbf{E}$ all off-diagonal non-zero entries are of the same sign. When this is the case, $\|\mathbf{E}\| = 2\|\mathbf{E}_0\| = 2\max_{i\in\mathscr{S}} |e_{ii}|$. □

3. The ergodicity coefficient of the perturbed chain

Sometimes it is clear from the structure of the perturbation that $\tilde{X}$ does have a unique stationary distribution (e.g. when $\tilde{\mathbf{Q}}$ is irreducible). In the general case, it can be shown that if X has a unique stationary distribution and $\tilde{\mathbf{Q}}$ is sufficiently close to $\mathbf{Q}$, then $\tilde{X}$ also has a unique stationary distribution. The question arises what is "sufficiently close"; we shall give one condition for this in terms of β_s. We need the following theorem.

THEOREM 3.1. *For all* $t > 0$,

$$|\tilde{\beta}_t - \beta_t| \begin{cases} < t\tau(\mathbf{E}), & \tau(\mathbf{E}) > 0, \\ = 0, & \tau(\mathbf{E}) = 0, \end{cases} \tag{3.1}$$

where $\tilde{\beta}_t = \tau(\tilde{\mathbf{P}}(t))$. *If* $\tau(\mathbf{E}) > 0$, *then, for all* $s > 0$,

$$\sup_{t\geqslant 0} |\tilde{\beta}_t - \beta_t| < \frac{s\tau(\mathbf{E})}{1-\beta_s}.$$

Proof. For all square real matrices $\mathbf{A}$ and $\mathbf{B}$, $\tau(\mathbf{A}+\mathbf{B}) \leqslant \tau(\mathbf{A}) + \tau(\mathbf{B})$. This implies that

$$|\tilde{\beta}_t - \beta_t| \leqslant \tau(\mathbf{Z}(t)),$$

where $\mathbf{Z}(t) = \tilde{\mathbf{P}}(t) - \mathbf{P}(t)$. In a similar way to the proof of Theorem 2.1, we obtain that

$$\mathbf{Z}(t) = \int_0^t \tilde{\mathbf{P}}(u)\mathbf{E}\mathbf{P}(t-u)\mathrm{d}u, \qquad t > 0,$$

which yields

$$\tau(\mathbf{Z}(t)) \leqslant \int_0^t \tau(\tilde{\mathbf{P}}(t-u)\mathbf{E}\mathbf{P}(u))\mathrm{d}u. \tag{3.2}$$

It follows from the definition of $\tau(\cdot)$ that if $\mathbf{A}$ and $\mathbf{B}$ are square real matrices, and $\mathbf{A}$ has equal row sums, then $\tau(\mathbf{AB}) \leqslant \tau(\mathbf{A})\tau(\mathbf{B})$. The matrix $\tilde{\mathbf{P}}(t)\mathbf{E}$ has zero row sums. Thus, for all $t, u > 0$, we obtain that

$$\tau(\tilde{\mathbf{P}}(t-u)\mathbf{E}\mathbf{P}(u)) \leqslant \tau(\tilde{\mathbf{P}}(t-u)\mathbf{E})\tau(\mathbf{P}(u)) \leqslant \beta_u\tau(\mathbf{E}). \tag{3.3}$$

This, together with (3.2) and the fact that $\beta_s < 1$ for all $s > 0$, proves (3.1).

If $\tau(\mathbf{E}) \neq 0$, then (2.7), (3.2) and (3.3) imply that

$$\sup_{t\geqslant 0} \tau(\mathbf{Z}(t)) \leqslant \tau(\mathbf{E}) \int_0^\infty \beta_u \mathrm{d}u < \frac{s\tau(\mathbf{E})}{1-\beta_s}, \qquad s > 0. \qquad \square$$

COROLLARY 3.1. *If $\beta_s + s\tau(\mathbf{E}) < 1$ holds for some $s > 0$, then $\tilde{X}$ has a unique stationary distribution.*

Proof. If the condition of the corollary is satisfied, then the relations (3.1) imply that $\tilde{\beta}_s < 1$. Applying Proposition 2.1 to $\tilde{X}$, we obtain the result. □

COROLLARY 3.2. *If $\tilde{\beta}_s < 1$, then*

$$\frac{s}{1-\tilde{\beta}_s} \geqslant \frac{s}{1-\beta_s + s\tau(\mathbf{E})}.$$

If $\beta_s + s\tau(\mathbf{E}) < 1$, then

$$\frac{s}{1-\tilde{\beta}_s} \leqslant \frac{s}{1-\beta_s - s\tau(\mathbf{E})}.$$

Proof. Follows directly from (3.1). □

In the bounds (2.3) and (2.4), the quantity $s/(1-\beta_s)$ is in fact a 'condition number' with respect to perturbations in the generator. If we know β_s, we can assess the sensitivity of $\tilde{X}$ to perturbations by using bounds on the respective condition number, $s/(1-\tilde{\beta}_s)$, as is shown below.

In $s/(1-\beta_s)$, we can put $s = p$, where $p = 1/\max_{i,j\in\mathscr{S}} |q_{ij}|$. This choice of s shows that if we multiply $\mathbf{Q}$ by a positive number greater than 1, then the sensitivity of the chain X to perturbations in the entries of $\mathbf{Q}$ will decrease. Using this approach, we can compare the sensitivity of X and $\tilde{X}$; all we need is to compare $p/(1-\beta_p)$ and $\tilde{p}/(1-\tilde{\beta}_{\tilde{p}})$, where $\tilde{p} = 1/\max_{i,j\in\mathscr{S}} |\tilde{q}_{ij}|$. This can be done using Corollary 3.2. If the quantity $\tau(\tilde{p}\tilde{\mathbf{Q}} - p\mathbf{Q})$ is sufficiently small, then Corollary 3.2, applied to the chains with generators $p\mathbf{Q}$ and $\tilde{p}\tilde{\mathbf{Q}}$, gives

$$\frac{1}{1-\tilde{\beta}_{\tilde{p}}} \leqslant \frac{1}{1-\beta_p - \tau(\tilde{p}\tilde{\mathbf{Q}} - p\mathbf{Q})}. \tag{3.4}$$

Now if $\tilde{p} < p$ so that

$$\frac{\tilde{p}}{1-\beta_p - \tau(\tilde{p}\tilde{\mathbf{Q}} - p\mathbf{Q})} < \frac{p}{1-\beta_p},$$

then (3.4) gives

$$\frac{\tilde{p}}{1-\tilde{\beta}_{\tilde{p}}} < \frac{p}{1-\beta_p},$$

which implies that, in the case being considered, the chain $\tilde{X}$ is less sensitive to perturbations in the generator than the chain X.

To conclude this section, we briefly compare our sufficient condition in Corollary 3.1 with similar conditions which follow from the results of Anisimov [2, 3]. Proposition A.1, Lemma A.1 and Proposition 2.1 imply that if $\beta_s + 2s\|\mathbf{E}_0\| < 1$ for some $s > 0$, then $\tilde{X}$ has a unique stationary distribution. If we use the inequality $\|\tilde{\mathbf{P}}(t) - \mathbf{P}(t)\| \leqslant t\|\mathbf{E}\|$ (implied by (1.1)) instead of Lemma A.1, we obtain the sufficient condition $\beta_s + s\|\mathbf{E}\| < 1$. These results are weaker than Corollary 3.1, since $\tau(\mathbf{A}) \leqslant \|\mathbf{A}\|$ for every square real matrix $\mathbf{A}$.

4. Applications to singularly perturbed Markov chains

Let ε be a small positive number. Consider two singularly perturbed continuous-time Markov chains, $X_1(\varepsilon) = \{X_1(\varepsilon, t), t \geqslant 0\}$ and $X_2(\varepsilon) = \{X_2(\varepsilon, t), t \geqslant 0\}$, with state space $\mathscr{S}$, generators $\varepsilon^{-1}\mathbf{Q}$ and $\varepsilon^{-1}\mathbf{Q} + \tilde{\mathbf{Q}}$, and distribution vectors $\mathbf{p}_1(\varepsilon, t)$ and $\mathbf{p}_2(\varepsilon, t)$, respectively. We assume that $X_1(\varepsilon)$ and $X_2(\varepsilon)$ have unique stationary distributions π_1 and $\pi_2(\varepsilon)$. Note that uniqueness of the stationary distribution in the time-homogeneous case is equivalent to weak irreducibility, an important condition in the theory of singular perturbations for Markov chains [6]. Markov chains of the same structure as $X_1(\varepsilon)$ can represent fast-changing processes in real-life problems, while chains having the same structure as $X_2(\varepsilon)$ can be used for modeling systems that have slow and fast components.

For singularly perturbed Markov chains, it is of interest to investigate their asymptotic behavior when $\varepsilon \to 0$. For chains with sufficiently smooth generators, the usual way of doing this is to obtain asymptotic expansions in terms of ε for the state probability vectors [6, 7]. Here we apply the results of Section 2. to obtain bounds on the quantity $\|\mathbf{p}_1(\varepsilon, t) - \mathbf{p}_2(\varepsilon, t)\|$, as well as estimates of the rate of convergence of $\mathbf{p}_2(\varepsilon, t)$ to π_1 as $\varepsilon \to 0$; these bounds are uniform over infinite time intervals.

Let $\hat{X}_1(\varepsilon) = \{\hat{X}_1(\varepsilon, \tau), \tau \geqslant 0\}$ and $\hat{X}_2(\varepsilon) = \{\hat{X}_2(\varepsilon, \tau), \tau \geqslant 0\}$ be continuous-time Markov chains with state space $\mathscr{S}$, generators $\mathbf{Q}$ and $\mathbf{Q} + \varepsilon\tilde{\mathbf{Q}}$, and initial distributions $\mathbf{p}_1(\varepsilon, 0)$ and $\mathbf{p}_2(\varepsilon, 0)$, respectively. For the vectors $\hat{\mathbf{p}}_1(\varepsilon, \tau) := \mathbf{p}_1(\varepsilon, \varepsilon\tau)$ and $\hat{\mathbf{p}}_2(\varepsilon, \tau) := \mathbf{p}_2(\varepsilon, \varepsilon\tau)$, the following equalities hold:

$$\mathrm{d}\hat{\mathbf{p}}_1(\varepsilon, \tau)/\mathrm{d}\tau = \hat{\mathbf{p}}_1(\varepsilon, \tau)\mathbf{Q}, \qquad \hat{\mathbf{p}}_1(\varepsilon, 0) = \mathbf{p}_1(\varepsilon, 0),$$
$$\mathrm{d}\hat{\mathbf{p}}_2(\varepsilon, \tau)/\mathrm{d}\tau = \hat{\mathbf{p}}_2(\varepsilon, \tau)(\mathbf{Q} + \varepsilon\tilde{\mathbf{Q}}), \qquad \hat{\mathbf{p}}_2(\varepsilon, 0) = \mathbf{p}_2(\varepsilon, 0),$$

therefore, $\hat{\mathbf{p}}_1(\varepsilon, \tau)$ and $\hat{\mathbf{p}}_2(\varepsilon, \tau)$ are the distributions of $\hat{X}_1(\varepsilon)$ and $\hat{X}_2(\varepsilon)$.

Denote by $\hat{\mathbf{P}}_1(\tau)$ and $\hat{\mathbf{P}}_2(\varepsilon, \tau)$ the respective transition probability matrices of the chains $\hat{X}_1(\varepsilon)$ and $\hat{X}_2(\varepsilon)$.

PROPOSITION 4.1. *For all $s > 0$, we have $\hat{\beta}_s := \tau(\hat{\mathbf{P}}_1(s)) < 1$ and*

$$\sup_{t\geqslant 0} \|\mathbf{p}_2(\varepsilon, t) - \mathbf{p}_1(\varepsilon, t)\| < \|\mathbf{p}_2(\varepsilon, 0) - \mathbf{p}_1(\varepsilon, 0)\| + \frac{s\varepsilon\|\tilde{\mathbf{Q}}\|}{1 - \hat{\beta}_s}.$$

Proof. Applying Theorem 2.1 to $\hat{X}_1(\varepsilon)$ and $\hat{X}_2(\varepsilon)$, for all $s > 0$ we obtain that

$$\sup_{\tau\geqslant 0} \|\hat{\mathbf{p}}_2(\varepsilon, \tau) - \hat{\mathbf{p}}_1(\varepsilon, \tau)\| < \|\mathbf{p}_2(\varepsilon, 0) - \mathbf{p}_1(\varepsilon, 0)\| + \frac{s\varepsilon\|\tilde{\mathbf{Q}}\|}{1 - \hat{\beta}_s}.$$

Since $\sup_{\tau\geqslant 0}\|\hat{\mathbf{p}}_2(\varepsilon,\tau)-\hat{\mathbf{p}}_1(\varepsilon,\tau)\| = \sup_{t\geqslant 0}\|\mathbf{p}_2(\varepsilon,t)-\mathbf{p}_1(\varepsilon,t)\|$, from the above inequality we obtain the result. □

THEOREM 4.1. *For all* $a>0$ *and* $s>0$, *the following inequalities hold:*

$$\sup_{t\geqslant a}\|\mathbf{p}_1(\varepsilon,t)-\pi_1\| \leqslant \hat{\beta}_s^{\lfloor a/(s\varepsilon)\rfloor}\|\mathbf{p}_1(\varepsilon,0)-\pi_1\|, \tag{4.1}$$

$$\sup_{t\geqslant a}\|\mathbf{p}_2(\varepsilon,t)-\pi_1\| < \hat{\gamma}_{\varepsilon,s}^{\lfloor a/(s\varepsilon)\rfloor}\|\mathbf{p}_2(\varepsilon,0)-\pi_2(\varepsilon)\| + \frac{2s\varepsilon\|\tilde{\mathbf{Q}}\|}{1-\hat{\beta}_s}, \tag{4.2}$$

where $\hat{\gamma}_{\varepsilon,s}=\tau(\hat{\mathbf{P}}_2(\varepsilon,s))$.

Proof. It follows that $\|\hat{\mathbf{p}}_1(\varepsilon,\tau)-\pi_1\|\leqslant\hat{\beta}_s^{\lfloor\tau/s\rfloor}\|\hat{\mathbf{p}}_1(\varepsilon,0)-\pi_1\|$. Setting $\tau=t/\varepsilon$, from this we obtain that

$$\|\mathbf{p}_1(\varepsilon,t)-\pi_1\|\leqslant\hat{\beta}_s^{\lfloor t/(s\varepsilon)\rfloor}\|\mathbf{p}_1(\varepsilon,0)-\pi_1\|. \tag{4.3}$$

The bound (4.1) follows from (4.3) and the fact that the variation distance between distributions of a Markov chain is a decreasing function of the time variable.

Let $W_a(\varepsilon)=\{W_a(\varepsilon,t),t\geqslant 0\}$ and $Y(\varepsilon)=\{Y(\varepsilon,t),t\geqslant 0\}$ be continuous-time Markov chains with state space $\mathscr{S}$, generators $\varepsilon^{-1}\mathbf{Q}+\tilde{\mathbf{Q}}$ and $\varepsilon^{-1}\mathbf{Q}$, and the respective initial distributions $\mathbf{p}_2(\varepsilon,a)$ and π_1. At any time $t>0$ the distribution of $W_a(\varepsilon,t)$ is $\mathbf{p}_2(\varepsilon,t+a)$, and the distribution of $Y(\varepsilon,t)$ is π_1. Applying Proposition 4.1 to $W_a(\varepsilon)$ and $Y(\varepsilon)$, for all $s>0$ we obtain that

$$\sup_{t\geqslant 0}\|\mathbf{p}_2(\varepsilon,a+t)-\pi_1\|<\|\mathbf{p}_2(\varepsilon,a)-\pi_1\|+\frac{s\varepsilon\|\tilde{\mathbf{Q}}\|}{1-\hat{\beta}_s}. \tag{4.4}$$

We have

$$\|\mathbf{p}_2(\varepsilon,a)-\pi_1\|\leqslant\|\mathbf{p}_2(\varepsilon,a)-\pi_2(\varepsilon)\|+\|\pi_2(\varepsilon)-\pi_1\|. \tag{4.5}$$

Applying Theorem 2.1 to $\hat{X}_1(\varepsilon)$ and $\hat{X}_2(\varepsilon)$, we get

$$\|\pi_2(\varepsilon)-\pi_1\|<\frac{s\varepsilon\|\tilde{\mathbf{Q}}\|}{1-\hat{\beta}_s},\qquad s>0. \tag{4.6}$$

It is easily seen that $\|\mathbf{p}_2(\varepsilon,t)-\pi_2(\varepsilon)\|\leqslant\hat{\gamma}_{\varepsilon,s}^{\lfloor t/(s\varepsilon)\rfloor}\|\mathbf{p}_2(\varepsilon,0)-\pi_2(\varepsilon)\|$. This, together with (4.4)–(4.6), gives (4.2). □

In Theorem 4.1, the quantity $\hat{\gamma}_{\varepsilon,s}$, which gauges the speed of convergence when $\varepsilon\to 0$, itself depends on ε in a complicated way; from the definition of $\tau(\cdot)$ it follows that $\hat{\gamma}_{\varepsilon,s}\to\hat{\beta}_s$ as $\varepsilon\to 0$. The next corollary provides a bound which is somewhat easier to use.

COROLLARY 4.1. *For all* $a>0$ *and* $s>0$,

$$\sup_{t\geqslant a}\|\mathbf{p}_2(\varepsilon,t)-\pi_1\|<\hat{\theta}_{\varepsilon,s}^{\lfloor a/(s\varepsilon)\rfloor}\|\mathbf{p}_2(\varepsilon,0)-\pi_2(\varepsilon)\|+\frac{2s\varepsilon\|\tilde{\mathbf{Q}}\|}{1-\hat{\beta}_s}, \tag{4.7}$$

where $\hat{\theta}_{\varepsilon,s}=\hat{\beta}_s+s\varepsilon\tau(\tilde{\mathbf{Q}})$.

Proof. Applying (3.1) to $\hat{X}_1(\varepsilon)$ and $\hat{X}_2(\varepsilon)$, and using (4.2), we obtain the result. □

REMARK 4.1. The right-hand sides of (4.2) and (4.7) can be modified by using the inequality

$$\|\mathbf{p}_2(\varepsilon,0) - \pi_2(\varepsilon)\| \leqslant \|\mathbf{p}_2(\varepsilon,0) - \pi_1\| + \|\pi_2(\varepsilon) - \pi_1\|$$

together with (4.6). We can also simplify the right-hand sides of (4.1), (4.2), and (4.7) by using the fact that $\|\mathbf{p} - \tilde{\mathbf{p}}\| \leqslant 2$ for all probability vectors $\mathbf{p}$, $\tilde{\mathbf{p}}$.

A. Appendix

In this appendix, for the reader's convenience, we give some important results of Anisimov [2, 3] that are used in the paper.

Let $X^{(i)} = \{X^{(i)}(t), t \geqslant 0\}$, $i = 1, 2$, be two Markov processes taking values in a measurable space $(\mathscr{X}, \mathfrak{B})$ and having initial distributions $\rho^{(i)}(\mathscr{A})$ and transition probabilities $p^{(i)}(t_1, x, t_2, \mathscr{A})$, $x \in \mathscr{X}$, $\mathscr{A} \in \mathfrak{B}$, $0 \leqslant t_1 \leqslant t_2$ (the time variable, t, can be either discrete or continuous). Define

$$\varphi^{(i)}(t_1, t_2) = \sup_{x_1, x_2 \in \mathscr{X}, \mathscr{A} \in \mathfrak{B}} |p^{(i)}(t_1, x_1, t_2, \mathscr{A}) - p^{(i)}(t_1, x_2, t_2, \mathscr{A})|,$$

$$\psi(t_1, t_2) = \sup_{x \in \mathscr{X}, \mathscr{A} \in \mathfrak{B}} |p^{(1)}(t_1, x, t_2, \mathscr{A}) - p^{(2)}(t_1, x, t_2, \mathscr{A})|, \qquad t_1 \leqslant t_2.$$

PROPOSITION A.1. (Corollary 1 of [2]) *If there exist $s > 0$ and $q \in (0, 1)$ such that*

$$\sup_{t \geqslant 0} \varphi^{(1)}(t, t+s) \leqslant q \tag{A.1}$$

and

$$\sup_{t \geqslant 0} \psi(t, t+s) \leqslant \alpha,$$

where α is such that $\tilde{q} := q + 2\alpha < 1$, then $\varphi^{(2)}(t_1, t_2) \leqslant \tilde{q}^{\lfloor (t_2 - t_1)/s \rfloor}$ for all $t_1 < t_2$.

THEOREM A.1. (Theorem 2 of [2]; Theorem 1.2 of [3, Chapter 3]) *Let* (A.1) *hold and* $\sup_{t \geqslant 0} \sup_{u \leqslant s} \psi(t, t+u) \leqslant \alpha$. *Then, for all $t_1 < t_2$,*

$$\psi(t_1, t_2) \leqslant \frac{\alpha}{1-q}\left(1 - q^{\lfloor (t_2 - t_1)/s \rfloor + 1}\right).$$

Suppose now that $X^{(i)}$, $i = 1, 2$, are continuous-time jump Markov processes with transition rates $q^{(i)}(x, t, \mathscr{A})$, that is, $p^{(i)}(t, x, t+\Delta, \mathscr{A}) = q^{(i)}(x, t, \mathscr{A})\Delta + o(\Delta)$, $x \in \mathscr{X} \setminus \mathscr{A}$, $\mathscr{A} \in \mathfrak{B}$. We assume that:

(a) the processes $X^{(i)}$ are uniquely defined by their transition rates;
(b) $q^{(i)}(x, t, \mathscr{A})$ are bounded functions of t on every finite interval;
(c) $q^{(i)}(x, t, \mathscr{A})$ are countably additive measure functions of $\mathscr{A}$;
(d) $q^{(i)}(x, t, \mathscr{A}) = q^{(i)}(x, t, \mathscr{A} \setminus \{x\})$.

LEMMA A.1. (Lemma 1.3 of [3, Chapter 3]) *If there exist* $t > 0$ *and* $s > 0$ *such that*

$$\int_t^{t+s} \sup_{x\in\mathscr{X},\mathscr{A}\in\mathfrak{B}} |q^{(1)}(x,u,\mathscr{A}) - q^{(2)}(x,u,\mathscr{A})|\mathrm{d}u \leqslant \alpha,$$

then $\psi(t, t+s) \leqslant 2\alpha$.

Acknowledgements. The author would like to thank the two independent referees for their comments on an earlier version of the paper.

REFERENCES

[1] D. B. ANDREEV, M. A. ELESIN, E. A. KRYLOV, A. V. KUZNETSOV AND A. I. ZEIFMAN, *On ergodicity and stability estimates for some nonhomogeneous Markov chains*, J. Math. Sci. (New York) **112** (2002), 4111–4118.

[2] V. V. ANISIMOV, *Estimates for the deviations of the transition characteristics of nonhomogeneous Markov processes*, Ukraïn. Mat. Zh. **40** (6) (1988), 699–704 (in Russian) (English translation in Ukrainian Math. J. **40** (6) (1988), 588–592).

[3] V. V. ANISIMOV, *Stochastic Processes with a Discrete Component. Limit Theorems*, Vishcha Shkola, Kiev, 1988 (in Russian).

[4] V. V. ANISIMOV AND M. F. TAIROV, *Estimates for the nearness of the transition characteristics of Markov chains*, in: Fourth International Vilnius Conference on Probability Theory and Mathematical Statistics (Abstracts of Communications), Institute of Mathematics and Cybernetics, Vilnius **1** (1985), 28–29 (in Russian).

[5] B. L. GRANOVSKY AND A. I. ZEIFMAN, *Nonstationary Markovian queues*, J. Math. Sci. (New York) **99** (2000), 1415–1438.

[6] R. Z. KHASMINSKII, G. YIN AND Q. ZHANG, *Asymptotic expansions of singularly perturbed systems involving rapidly fluctuating Markov chains*, SIAM J. Appl. Math. **56** (1996), 277–293.

[7] R. Z. KHASMINSKII, G. YIN AND Q. ZHANG, *Constructing asymptotic series for probability distributions with weak and strong interactions*, Quart. Appl. Math. **LV** (1997), 177–200.

[8] A. YU. MITROPHANOV, *Stochastic Markov models for the process of binary complex formation and dissociation*, Mat. Model. **13** (9) (2001), 101–109 (in Russian).

[9] A. YU. MITROPHANOV, *Stability and exponential convergence of continuous-time Markov chains*, J. Appl. Probab. **40** (2003), 970–979.

[10] A. V. RAMESH AND K. S. TRIVEDI, *On the sensitivity of transient solution of Markov models*, in: Proceedings of 1993 ACM SIGMETRICS, Santa Clara (1993), 122–134.

[11] E. SENETA, *Coefficients of ergodicity: structure and applications*, Adv. in Appl. Probab. **11** (1979), 576–590.

[12] E. SENETA, *Explicit forms for ergodicity coefficients and spectrum localization*, Linear Algebra Appl. **60** (1984), 187–197.

[13] G. YIN, Q. ZHANG AND G. BADOWSKI, *Singularly perturbed Markov chains: convergence and aggregation*, J. Multivariate Anal. **72** (2000), 208–229.

[14] Q. ZHANG AND G. YIN, *Markov chains with weak and strong interactions: structural properties*, Nonlinear Anal. **30** (1997), 309–316.

[15] A. I. ZEIFMAN AND D. L. ISAACSON, *On strong ergodicity for nonhomogeneous continuous-time Markov chains*, Stoch. Proc. Appl. **50** (1994), 263–273.

[16] A. I. ZEIFMAN, *Stability of birth-and-death processes*, J. Math. Sci. (New York) **91** (1998), 3023–3031.

(Received October 14, 2002)
(Revised October 15, 2003)

A. Yu. Mitrophanov
School of Biology
Georgia Institute of Technology
310 Ferst Drive
Atlanta, GA 30332
USA
e-mail: amit@amber.biology.gatech.edu