J. Appl. Prob. **40,** 970–979 (2003)
Printed in Israel

STABILITY AND EXPONENTIAL CONVERGENCE OF CONTINUOUS-TIME MARKOV CHAINS

A. YU. MITROPHANOV,* *Saratov State University*

Abstract

For finite, homogeneous, continuous-time Markov chains having a unique stationary distribution, we derive perturbation bounds which demonstrate the connection between the sensitivity to perturbations and the rate of exponential convergence to stationarity. Our perturbation bounds substantially improve upon the known results. We also discuss convergence bounds for chains with diagonalizable generators and investigate the relationship between the rate of convergence and the sensitivity of the eigenvalues of the generator; special attention is given to reversible chains.

Keywords: Markov chain; exponential convergence; perturbation bounds; sensitivity analysis; ergodicity coefficient; eigenvalue; spectral gap

AMS 2000 Subject Classification: Primary 60J27
Secondary 60J35; 60E15; 34D10; 15A42

1. Introduction

When using mathematical models to study real-world phenomena, it is often necessary to predict the uncertainty in the model output given the uncertainty in the parameter values. The purpose of this paper is to develop a new approach to sensitivity analysis for finite, homogeneous, continuous-time Markov chains. Such chains are widely used as a modeling tool; to give some examples of modern applications, we mention biosensors [2], ion channels [3], cell adhesion receptors [16]. In chemical kinetics, reversible chains are of special importance since reversibility reflects the physical property of detailed balance [17].

Inequality-based perturbation theory for continuous-time Markov chains has been the subject of several publications; see [1], [9], [18], [20], [21] (a perturbation theory for discrete-time Markov chains with general state space can be found in [12], [13]). One of the most important results was the discovery of a connection between the 'mixing rate' of a chain and its sensitivity to perturbations. We further explore this connection by deriving estimates of stability which are much sharper than the ones obtained by other authors (see Section 2). Though we restrict our attention to time-homogeneous Markov chains, our method of stability analysis can be extended to nonhomogeneous processes. In Section 3 we consider convergence bounds for chains with diagonalizable generators and show that the rate of convergence is closely related to the sensitivity of the eigenvalues of the generator. We demonstrate that, for reversible chains, the closeness of the generator to being symmetric controls the stability of its eigenvalues and the rate of convergence of the chain to stationarity. Finally, in Section 4 we compare the convergence bounds discussed in Sections 2 and 3.

Received 23 January 2003; revision received 1 July 2003.
* Postal address: Faculty of Computer Science and Information Technology, Saratov State University, 83 Astrakhanskaya str., Saratov 410012, Russia. Email address: mitrophanovay@info.sgu.ru

2. Perturbation bounds and exponential convergence

Consider two continuous-time Markov chains, $X(t)$ and $\tilde{X}(t)$, $t \geq 0$, with finite state space $\mathcal{S} = \{0, 1, \dots, N\}$, $N \geq 1$, and respective generators $\boldsymbol{Q} = (q_{ij})$ and $\tilde{\boldsymbol{Q}} = (\tilde{q}_{ij})$ (with $\boldsymbol{Q} \neq \tilde{\boldsymbol{Q}}$). Let $\boldsymbol{p}(t) = (p_i(t))$ and $\tilde{\boldsymbol{p}}(t) = (\tilde{p}_i(t))$ be the state probability vectors of $X(t)$ and $\tilde{X}(t)$. In this section our main goal is to obtain estimates for the distance between $\boldsymbol{p}(t)$ and $\tilde{\boldsymbol{p}}(t)$ which relate the stability of $\boldsymbol{p}(t)$ to the rate of its convergence as $t \to \infty$. Throughout the paper $X(t)$ has a unique stationary distribution $\boldsymbol{\pi} = (\pi_i)$.

We regard vectors as row vectors, and $\|\cdot\|$ denotes the l_1-norm (absolute entry sum) for vectors and the corresponding subordinate norm (maximum absolute row sum) for matrices. Note that for all probability vectors $\boldsymbol{p}$ and $\tilde{\boldsymbol{p}}$, the quantity $\|\boldsymbol{p} - \tilde{\boldsymbol{p}}\|$ is twice the variation distance between the respective distributions $p(\mathcal{A})$ and $\tilde{p}(\mathcal{A})$ (for $\mathcal{A} \subseteq \mathcal{S}$): $\|\boldsymbol{p} - \tilde{\boldsymbol{p}}\| = 2 \max_{\mathcal{A} \subseteq \mathcal{S}} |p(\mathcal{A}) - \tilde{p}(\mathcal{A})|$.

For probability vectors $\boldsymbol{p}_1$ and $\boldsymbol{p}_2$, define $\boldsymbol{p}_1(t) = \boldsymbol{p}_1 \boldsymbol{P}(t)$, $\boldsymbol{p}_2(t) = \boldsymbol{p}_2 \boldsymbol{P}(t)$, where $\boldsymbol{P}(t)$ is the transition probability matrix of $X(t)$, $\boldsymbol{P}(t) = \mathrm{e}^{\boldsymbol{Q}t}$.

Theorem 2.1. *If $b > 0$ and $c > 2$ are constants such that, for all $\boldsymbol{p}_1$, $\boldsymbol{p}_2$,*

$$\|\boldsymbol{p}_1(t) - \boldsymbol{p}_2(t)\| \leq c\mathrm{e}^{-bt}, \qquad t \geq 0, \tag{2.1}$$

then, for $\boldsymbol{z}(t) = \tilde{\boldsymbol{p}}(t) - \boldsymbol{p}(t)$ and $\boldsymbol{E} = \tilde{\boldsymbol{Q}} - \boldsymbol{Q}$,

$$\|\boldsymbol{z}(t)\| < \begin{cases} \|\boldsymbol{z}(0)\| + t\|\boldsymbol{E}\|, & 0 < t \leq b^{-1}\log\left(\dfrac{c}{2}\right), \\ \dfrac{c\mathrm{e}^{-bt}\|\boldsymbol{z}(0)\|}{2} + b^{-1}\left(\log\left(\dfrac{c}{2}\right) + 1 - \dfrac{c\mathrm{e}^{-bt}}{2}\right)\|\boldsymbol{E}\|, & t \geq b^{-1}\log\left(\dfrac{c}{2}\right). \end{cases} \tag{2.2}$$

If $\tilde{\boldsymbol{\pi}}$ is a stationary distribution of $\tilde{X}(t)$, then

$$\|\tilde{\boldsymbol{\pi}} - \boldsymbol{\pi}\| < b^{-1}\left(\log\left(\frac{c}{2}\right) + 1\right)\|\boldsymbol{E}\|. \tag{2.3}$$

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

$$\frac{\mathrm{d}\tilde{\boldsymbol{p}}(t)}{\mathrm{d}t} = \tilde{\boldsymbol{p}}(t)\tilde{\boldsymbol{Q}}, \qquad \frac{\mathrm{d}\boldsymbol{p}(t)}{\mathrm{d}t} = \boldsymbol{p}(t)\boldsymbol{Q}, \qquad t \geq 0;$$

therefore, the vector $\boldsymbol{z}(t)$ is the solution to the initial-value problem

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

which implies that

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

$$\|\boldsymbol{z}(t)\| \leq \|\boldsymbol{z}(0)\boldsymbol{P}(t)\| + \int_0^t \|\tilde{\boldsymbol{p}}(t-u)\boldsymbol{E}\boldsymbol{P}(u)\|\,\mathrm{d}u. \tag{2.4}$$

Define

$$\beta(t) = \sup_{\substack{\|\boldsymbol{v}\|=1 \\ \boldsymbol{v}\boldsymbol{e}^\top = 0}} \|\boldsymbol{v}\boldsymbol{P}(t)\| = \tfrac{1}{2}\max_{i,j \in \mathcal{S}} \|(\boldsymbol{e}_i - \boldsymbol{e}_j)\boldsymbol{P}(t)\|,$$

where $\boldsymbol{e}$ is the row vector all of whose entries are 1, $^\top$ denotes transpose and $\boldsymbol{e}_i$ is the vector whose ith entry equals 1 and all the other entries equal 0. The quantity $\beta(t)$ is the l_1 ergodicity coefficient of $\boldsymbol{P}(t)$; see [15]. Since $\tilde{\boldsymbol{p}}(t)\boldsymbol{E}\boldsymbol{e}^\top \equiv 0$ (implied by $\boldsymbol{E}\boldsymbol{e}^\top = \boldsymbol{0}^\top$) and

$$\|\tilde{\boldsymbol{p}}(t-u)\boldsymbol{E}\boldsymbol{P}(u)\| \le \|\boldsymbol{E}\| \left\| \frac{\tilde{\boldsymbol{p}}(t-u)\boldsymbol{E}}{\|\tilde{\boldsymbol{p}}(t-u)\boldsymbol{E}\|} \boldsymbol{P}(u) \right\|$$

(if $\tilde{\boldsymbol{p}}(t-u)\boldsymbol{E} \neq \boldsymbol{0}$), we have $\|\tilde{\boldsymbol{p}}(t-u)\boldsymbol{E}\boldsymbol{P}(u)\| \le \|\boldsymbol{E}\|\beta(u)$. Similarly, $\|z(0)\boldsymbol{P}(t)\| \le \|z(0)\|\beta(t)$. This, together with (2.4), gives

$$\|z(t)\| \le \beta(t)\|z(0)\| + \|\boldsymbol{E}\| \int_0^t \beta(u)\,\mathrm{d}u. \tag{2.5}$$

The inequality $\beta(t) < 1$ holds if and only if, for every two states $i, j \in \mathcal{S}$, there exists a state k such that $p_{ik}(t) > 0$ and $p_{jk}(t) > 0$. This condition is satisfied for all $t > 0$ because every state in the unique closed irreducible class of $X(t)$ can be reached from every state in $\mathcal{S}$ in any time $t > 0$. This implies that

$$\int_0^t \beta(u)\,\mathrm{d}u < t, \qquad t > 0. \tag{2.6}$$

It follows from (2.1) that

$$\beta(t) \le \frac{c\mathrm{e}^{-bt}}{2}, \qquad t \ge 0. \tag{2.7}$$

Using (2.6) and (2.7), we obtain that

$$\begin{aligned}\int_0^t \beta(u)\,\mathrm{d}u &< b^{-1}\log\Big(\frac{c}{2}\Big) + \int_{b^{-1}\log(c/2)}^t \frac{c}{2}\mathrm{e}^{-bu}\,\mathrm{d}u \\ &= b^{-1}\Big(\log\Big(\frac{c}{2}\Big) + 1 - \frac{c\mathrm{e}^{-bt}}{2}\Big), \qquad b^{-1}\log\Big(\frac{c}{2}\Big) \le t \le \infty.\end{aligned} \tag{2.8}$$

This, together with (2.5)–(2.7) and the fact that $\boldsymbol{Q} \neq \tilde{\boldsymbol{Q}}$, gives (2.2).

Setting $\tilde{\boldsymbol{p}}(0) = \tilde{\boldsymbol{\pi}}$ and passing to the limit as $t \to \infty$ in (2.5), we get

$$\|\tilde{\boldsymbol{\pi}} - \boldsymbol{\pi}\| \le \|\boldsymbol{E}\| \int_0^\infty \beta(u)\,\mathrm{d}u.$$

This inequality and (2.8) prove (2.3).

Corollary 2.1. *In the setting of Theorem 2.1,*

$$\sup_{t\ge 0} \|z(t)\| < \begin{cases} b^{-1}\Big(\log\Big(\dfrac{c}{2}\Big) + 1\Big)\|\boldsymbol{E}\| & \text{if } \|\boldsymbol{E}\| \ge b\|z(0)\|, \\ \|z(0)\| + b^{-1}\log\Big(\dfrac{c}{2}\Big)\|\boldsymbol{E}\| & \text{otherwise.} \end{cases} \tag{2.9}$$

Proof. Denote by $f(t)$ the right-hand side of (2.2). Suppose that $\|\boldsymbol{E}\| \ge b\|z(0)\|$. If $\sup_{t\ge 0}\|z(t)\| = \|z(t_0)\|$ for some $t_0 < \infty$, then (2.9) follows from (2.2) and the fact that $f(t)$ is an increasing function. If $\sup_{t\ge 0}\|z(t)\| = \lim_{t\to\infty}\|z(t)\|$, then (2.9) follows from (2.3). Suppose now that $\|\boldsymbol{E}\| < b\|z(0)\|$. In this case, $f(t)$ is a strictly decreasing function for $t > b^{-1}\log(c/2)$. This fact, together with (2.2), gives (2.9).

Remark 2.1. If (2.1) holds with some $b, c > 0$, then $c \geq 2$. In the case when $c = 2$ we can obtain results similar to Theorem 2.1 and Corollary 2.1; the analogues of (2.2) for $t \geq b^{-1}\log(c/2)$, and of (2.3) and (2.9) have the same form, but the inequalities are not strict.

Remark 2.2. If a, d are positive constants such that, for all $\boldsymbol{p}(0)$,

$$\|\boldsymbol{p}(t) - \boldsymbol{\pi}\| \leq d\mathrm{e}^{-at}, \qquad t \geq 0, \tag{2.10}$$

then we have (2.1) with $c = 2d$ and $b = a$. The convergence bound (2.10) holds for arbitrary $\boldsymbol{p}(0)$ if it holds for $\boldsymbol{p}(0) = \boldsymbol{e}_i$ for each $i \in \mathcal{S}$; similarly, (2.1) holds for arbitrary $\boldsymbol{p}_1$ and $\boldsymbol{p}_2$ if it holds for $\boldsymbol{p}_1 = \boldsymbol{e}_i$ and $\boldsymbol{p}_2 = \boldsymbol{e}_j$ for each $i, j \in \mathcal{S}$.

We now compare our perturbation bounds with the ones derived in [9] and [21]. These papers deal with the case $\boldsymbol{z}(0) = \boldsymbol{0}$. Let C and α be positive constants such that

$$\|\boldsymbol{p}_2(t) - \boldsymbol{p}_1(t)\| \leq C\mathrm{e}^{-\alpha t}\|\boldsymbol{p}_2 - \boldsymbol{p}_1\|, \qquad t \geq 0, \tag{2.11}$$

for all $\boldsymbol{p}_1, \boldsymbol{p}_2$. Following the proof of Theorem 2.1 of [21], we arrive at

$$\sup_{t\geq 0}\|\boldsymbol{z}(t)\| \leq 12C\alpha^{-1}\|\boldsymbol{E}\|. \tag{2.12}$$

Our result for this situation is much sharper: if (2.11) holds, then Corollary 2.1, Remark 2.1 and the inequality $\|\boldsymbol{p}_2 - \boldsymbol{p}_1\| \leq 2$ give

$$\sup_{t\geq 0}\|\boldsymbol{z}(t)\| \leq \alpha^{-1}(\log(C) + 1)\|\boldsymbol{E}\|. \tag{2.13}$$

Suppose that $X(t)$ is a birth–death process with birth rates $\{b_i,\ i \in \mathcal{S}\}$ and death rates $\{d_i,\ i \in \mathcal{S}\}$, all positive except $b_N = d_0 = 0$. Let $g_0, g_1, \dots, g_{N-1}$ be positive numbers; set $g_{-1} = g_0 = 1$, $g_N = 0$. Consider

$$\boldsymbol{G} = \begin{pmatrix} g_0 & g_0 & g_0 & \cdots & g_0 \\ 0 & g_1 & g_1 & \cdots & g_1 \\ 0 & 0 & g_2 & \cdots & g_2 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & g_{N-1} \end{pmatrix},$$

$$\boldsymbol{G}^{-1} = \begin{pmatrix} g_0^{-1} & -g_1^{-1} & 0 & \cdots & 0 \\ 0 & g_1^{-1} & -g_2^{-1} & \ddots & \vdots \\ 0 & 0 & g_2^{-1} & \ddots & 0 \\ \vdots & \vdots & \ddots & \ddots & -g_{N-1}^{-1} \\ 0 & 0 & \cdots & 0 & g_{N-1}^{-1} \end{pmatrix},$$

$$\alpha_k = b_k + d_{k+1} - g_{k+1}g_k^{-1}b_{k+1} - g_{k-1}g_k^{-1}d_k, \qquad k = 0, 1, \dots, N-1, \tag{2.14}$$

and assume that the g_k are such that $\underline{\alpha} := \min_{0\leq k\leq N-1}\alpha_k > 0$. Following the proofs of Theorems 6 and 8 of [9], we obtain that

$$\|\boldsymbol{p}_2(t) - \boldsymbol{p}_1(t)\| \leq 2\kappa_\infty(\boldsymbol{G})\mathrm{e}^{-\underline{\alpha}t}\|\boldsymbol{p}_2 - \boldsymbol{p}_1\|, \qquad t \geq 0, \tag{2.15}$$

$$\sup_{t\geq 0}\|\boldsymbol{z}(t)\| \leq 6\underline{\alpha}^{-1}\kappa_\infty(\boldsymbol{G})\|\boldsymbol{E}\|,$$

where $\kappa_\infty(\boldsymbol{G}) = \|\boldsymbol{G}\|_\infty \|\boldsymbol{G}^{-1}\|_\infty$ (the ∞-norm condition number of $\boldsymbol{G}$) and $\|\cdot\|_\infty$ denotes the maximum absolute column sum of a matrix. Note that the above perturbation bound is tighter than the combination of (2.12) and (2.15). The inequalities (2.13) and (2.15) imply that

$$\sup_{t\geq 0} \|z(t)\| \leq \underline{\alpha}^{-1}(\log(2\kappa_\infty(\boldsymbol{G})) + 1)\|\boldsymbol{E}\|.$$

Our approach can also be compared with that of [1] and [20] in a similar manner.

Finally, we highlight the main features of the presented method, which produces improved perturbation bounds:

(a) in [1], [9], [18], [20] and [21], the authors study the so-called reduced systems of forward Kolmogorov equations, whereas we analyze the systems of Kolmogorov equations in their unmodified form;

(b) we use the inequality $\beta(t) < 1$ to prove (2.8).

It should also be noted that, since the notion of exponential convergence can be generalized in a natural way to nonhomogeneous Markov chains [21], it is possible to extend the sensitivity bounds in Theorem 2.1 and Corollary 2.1 to chains with time-dependent generators (explicit convergence bounds for nonhomogeneous birth–death processes are given in [9] and [19]).

3. Convergence bounds for Markov chains with diagonalizable generators

In this section we assume that $\boldsymbol{Q}$ is diagonalizable. The generator $\boldsymbol{Q}$ enjoys this property if all its eigenvalues are distinct; another sufficient condition is reversibility of $X(t)$. We discuss bounds on the rate of convergence to stationarity which, in conjunction with Theorem 2.1, yield perturbation bounds for the distribution of $X(t)$.

Since $\boldsymbol{Q}$ is the generator of a Markov chain, 0 is one of its eigenvalues; real parts of all nonzero eigenvalues are negative. The uniqueness of the stationary distribution, $\boldsymbol{\pi}$, implies that the multiplicity of the zero eigenvalue of $\boldsymbol{Q}$ is 1. Denote by λ_j for $j = 0, 1, \ldots, M$, the distinct values of the eigenvalues of $\boldsymbol{Q}$; their respective multiplicities are denoted by m_j, $\sum_{j=0}^{M} m_j = N + 1$. Choose $\lambda_0 = 0$. Put $\lambda = \min_{1\leq j\leq M} |\operatorname{Re}\lambda_j|$; this quantity is the spectral gap of $\boldsymbol{Q}$.

Let $\boldsymbol{X}$ be a nonsingular matrix such that $\boldsymbol{Q} = \boldsymbol{X}^{-1}\boldsymbol{\Lambda}\boldsymbol{X}$, where $\boldsymbol{\Lambda}$ is diagonal. It is clear that the rows of $\boldsymbol{X}$, denoted by $\boldsymbol{x}_i$, are left eigenvectors of $\boldsymbol{Q}$, and $\boldsymbol{\Lambda}$ consists of the eigenvalues of $\boldsymbol{Q}$. For each λ_j, define $\mathcal{L}_j = \{k \in \mathcal{S} : \boldsymbol{x}_k\boldsymbol{Q} = \lambda_j\boldsymbol{Q}\}$. Set $\kappa(\boldsymbol{X}) = \|\boldsymbol{X}\|\|\boldsymbol{X}^{-1}\|$.

Theorem 3.1. *For all* $\boldsymbol{p}(0)$,

$$\|\boldsymbol{p}(t) - \boldsymbol{\pi}\| \leq K\mathrm{e}^{-\lambda t}, \qquad t \geq 0, \tag{3.1}$$

where

$$K = \max_{i\in\mathcal{S}} \sum_{j=1}^{M} \left\| \sum_{k\in\mathcal{L}_j} x_{ik}^{(-1)} \boldsymbol{x}_k \right\| = \max_{i\in\mathcal{S}} \sum_{j=1}^{M} \left\| \boldsymbol{e}_i(\boldsymbol{Q}/\lambda_j) \prod_{k\neq 0, j} \frac{\boldsymbol{I} - \boldsymbol{Q}/\lambda_k}{1 - \lambda_j/\lambda_k} \right\|,$$

the $x_{ik}^{(-1)}$ *are the entries of* $\boldsymbol{X}^{-1}$ *and* $\boldsymbol{I}$ *denotes the identity matrix. Also,*

$$K \leq \kappa(\boldsymbol{X}) - 1,$$

and equality is attained if all eigenvalues of $\boldsymbol{Q}$ *are distinct and* $\|\boldsymbol{x}_i\| = 1$ *for all* $i \in \mathcal{S}$.

Proof. We have $\boldsymbol{p}(t) = \boldsymbol{p}(0)\boldsymbol{P}(t)$, $\boldsymbol{P}(t) = \boldsymbol{X}^{-1}\mathrm{e}^{\boldsymbol{\Lambda}t}\boldsymbol{X}$ and

$$\boldsymbol{p}(t) = x_{i0}^{(-1)}\boldsymbol{x}_0 + \sum_{j=1}^{M}\biggl(\sum_{k\in\mathcal{L}_j} x_{ik}^{(-1)}\boldsymbol{x}_k\biggr)\mathrm{e}^{\lambda_j t}, \qquad \boldsymbol{p}(0) = \boldsymbol{e}_i. \tag{3.2}$$

If $\boldsymbol{p}(0) = \boldsymbol{e}_i$, then the following expression holds [14]:

$$\boldsymbol{p}(t) = \boldsymbol{e}_i \prod_{j=1}^{M}(\boldsymbol{I} - \boldsymbol{Q}/\lambda_j) + \sum_{j=1}^{M}\boldsymbol{e}_i(\boldsymbol{Q}/\lambda_j)\prod_{k\neq 0,j}\frac{\boldsymbol{I} - \boldsymbol{Q}/\lambda_k}{1-\lambda_j/\lambda_k}\mathrm{e}^{\lambda_j t}.$$

Since the real parts of λ_j for $j \neq 0$ are negative, $x_{i0}^{(-1)}\boldsymbol{x}_0 = \boldsymbol{e}_i\prod_{j=1}^{M}(\boldsymbol{I} - \boldsymbol{Q}/\lambda_j) = \boldsymbol{\pi}$ for all $i \in \mathcal{S}$. The functions $\mathrm{e}^{\lambda_1 t}, \dots, \mathrm{e}^{\lambda_M t}$ are linearly independent, and, therefore,

$$\sum_{k\in\mathcal{L}_j} x_{ik}^{(-1)}\boldsymbol{x}_k = \boldsymbol{e}_i(\boldsymbol{Q}/\lambda_j)\prod_{k\neq 0,j}\frac{\boldsymbol{I} - \boldsymbol{Q}/\lambda_k}{1-\lambda_j/\lambda_k}, \qquad j = 1, \dots, M. \tag{3.3}$$

The expression (3.2) implies that

$$\max_{i\in\mathcal{S}}\|\boldsymbol{e}_i\boldsymbol{P}(t) - \boldsymbol{\pi}\| \le \mathrm{e}^{-\lambda t}\max_{i\in\mathcal{S}}\sum_{j=1}^{M}\biggl\|\sum_{k\in\mathcal{L}_j} x_{ik}^{(-1)}\boldsymbol{x}_k\biggr\|.$$

This, together with (3.3), gives (3.1).

Since $\|x_{i0}^{(-1)}\boldsymbol{x}_0\| = |x_{i0}^{(-1)}|\|\boldsymbol{x}_0\| = 1$, we obtain that

$$\begin{aligned} K &\le \max_{i\in\mathcal{S}}\sum_{j=1}^{N}|x_{ij}^{(-1)}|\|\boldsymbol{x}_j\| = \max_{i\in\mathcal{S}}\sum_{j=0}^{N}|x_{ij}^{(-1)}|\|\boldsymbol{x}_j\| - 1 \\ &\le \Bigl(\max_{j\in\mathcal{S}}\|\boldsymbol{x}_j\|\Bigr)\Bigl(\max_{i\in\mathcal{S}}\sum_{j=0}^{N}|x_{ij}^{(-1)}|\Bigr) - 1. \end{aligned}$$

If all eigenvalues of $\boldsymbol{Q}$ are distinct and $\|\boldsymbol{x}_i\| = 1$ for all $i \in \mathcal{S}$, then

$$K = \max_{i\in\mathcal{S}}\sum_{j=1}^{N}|x_{ij}^{(-1)}|\|\boldsymbol{x}_j\| = \max_{i\in\mathcal{S}}\sum_{j=0}^{N}|x_{ij}^{(-1)}| - 1 = \|\boldsymbol{X}^{-1}\| - 1.$$

Remark 3.1. It follows from Theorem 3.1 that in the case of distinct eigenvalues the scaling $\|\boldsymbol{x}_i\| = 1$ for all $i \in \mathcal{S}$ minimizes $\kappa(\boldsymbol{X})$. The same is true for the general case of diagonalizable $\boldsymbol{Q}$; this can be proved using the expression $\boldsymbol{X}^{-1} = \mathrm{adj}(\boldsymbol{X})/\det(\boldsymbol{X})$, where $\mathrm{adj}(\boldsymbol{X})$ and $\det(\boldsymbol{X})$ are, respectively, the adjugate and determinant of $\boldsymbol{X}$.

The connection between $\kappa(\boldsymbol{X})$ and the sensitivity of the eigenvalues of $\boldsymbol{Q}$ to perturbations is well known. If μ is an eigenvalue of $\tilde{\boldsymbol{Q}}$, then the following inequality holds [11]:

$$\min_{0\le j\le M}|\lambda_j - \mu| \le \kappa(\boldsymbol{X})\|\boldsymbol{E}\|. \tag{3.4}$$

The next two propositions give further results for Markov chains.

Proposition 3.1. *Suppose that*

$$\kappa(\boldsymbol{X})\|\boldsymbol{E}\| < \tfrac{1}{2} \min_{j\in\mathcal{A},k\in\mathcal{B}} |\lambda_j - \lambda_k|, \tag{3.5}$$

where $\mathcal{A} = \{j : 0 \le j \le M, |\mathrm{Re}\,\lambda_j| \ne \lambda\}$ *and* $\mathcal{B} = \{j : 1 \le j \le M, |\mathrm{Re}\,\lambda_j| = \lambda\}$. *If* $\tilde{\lambda}$ *denotes the spectral gap of* $\tilde{\boldsymbol{Q}}$, *then*

$$|\tilde{\lambda} - \lambda| \le \kappa(\boldsymbol{X})\|\boldsymbol{E}\|.$$

Proof. The inequality (3.4) implies that the eigenvalues of $\tilde{\boldsymbol{Q}}$ lie in the discs

$$\{z \in \mathbb{C} : |z - \lambda_j| \le \kappa(\boldsymbol{X})\|\boldsymbol{E}\|\}, \qquad 0 \le j \le M. \tag{3.6}$$

If (3.5) holds, then the set

$$\mathcal{M} = \bigcup_{j\in\mathcal{B}} \{z \in \mathbb{C} : |z - \lambda_j| \le \kappa(\boldsymbol{X})\|\boldsymbol{E}\|\}$$

does not intersect the discs (3.6) for $j \in \mathcal{A}$. It can be proved that in this case $\mathcal{M}$ contains $\sum_{j\in\mathcal{B}} m_j$ eigenvalues of $\tilde{\boldsymbol{Q}}$; see the proof of Theorem 6.1.1 of [11] (the Geršgorin theorem). Similarly, (3.5) implies that the disc $\{z \in \mathbb{C} : |z| \le \kappa(\boldsymbol{X})\|\boldsymbol{E}\|\}$ is isolated from the discs (3.6) for nonzero eigenvalues of $\boldsymbol{Q}$ and hence contains only one eigenvalue of $\tilde{\boldsymbol{Q}}$. Obviously, it is the zero eigenvalue. Thus, if μ is an eigenvalue of $\tilde{\boldsymbol{Q}}$ such that $|\,\mathrm{Re}\,\mu| = \tilde{\lambda}$, then either $\mu \in \mathcal{M}$ or

$$\mu \in \bigcup_{j\in\mathcal{A}} \{z \in \mathbb{C} : |z - \lambda_j| \le \kappa(\boldsymbol{X})\|\boldsymbol{E}\|\} \quad \text{and} \quad |\tilde{\lambda} - \lambda| \le \kappa(\boldsymbol{X})\|\boldsymbol{E}\|.$$

Proposition 3.2. *If* $\kappa(\boldsymbol{X})\|\boldsymbol{E}\| < \frac{1}{2}\min_{1\le j\le M} |\lambda_j|$, *then the chain* $\tilde{X}(t)$ *has a unique stationary distribution.*

Proof. It follows that the disc $\{z \in \mathbb{C} : |z| \le \kappa(\boldsymbol{X})\|\boldsymbol{E}\|\}$ is isolated from analogous discs for nonzero eigenvalues of $\boldsymbol{Q}$. Therefore, the multiplicity of the zero eigenvalue of $\tilde{\boldsymbol{Q}}$ is 1, which implies uniqueness of the stationary distribution of $\tilde{X}(t)$.

Suppose now that $X(t)$ is reversible, i.e. it is irreducible and the condition

$$\pi_i q_{ij} = \pi_j q_{ji} \tag{3.7}$$

is satisfied for all $i, j \in \mathcal{S}$. In this case, the following bound holds for all $\boldsymbol{p}(0)$ (see [6]):

$$\|\boldsymbol{p}(t) - \boldsymbol{\pi}\| \le \mathrm{e}^{-\lambda t}\sqrt{\max_{i\in\mathcal{S}} \frac{1}{\pi_i} - 1}. \tag{3.8}$$

Theorem 3.1 and (3.4) show that, for chains with diagonalizable generators, the rate of convergence to stationarity and the sensitivity of the eigenvalues of $\boldsymbol{Q}$ are both controlled by the quantity $\kappa(\boldsymbol{X})$. Below we demonstrate that, in the case of reversible chains, these two properties depend on the spread of the stationary probabilities, which in turn depends on how close $\boldsymbol{Q}$ is to being symmetric.

For a square real matrix $\boldsymbol{A}$, define the spectral norm by $\|\boldsymbol{A}\|_2 = \max_i \sqrt{\nu_i}$, where the ν_i are the eigenvalues of $\boldsymbol{A}^\top\boldsymbol{A}$.

Theorem 3.2. *If $X(t)$ is reversible and μ is an eigenvalue of $\tilde{\boldsymbol{Q}}$, then*

$$\min_{0\le j\le M} |\lambda_j - \mu| \le \sqrt{\max_{i,j\in\mathcal{S}} \frac{\pi_i}{\pi_j}}\,\|\boldsymbol{E}\|_2 \le \sqrt{(N+1)\max_{i,j\in\mathcal{S}} \frac{\pi_i}{\pi_j}}\,\|\boldsymbol{E}\|. \tag{3.9}$$

Proof. Define $\boldsymbol{D} = \operatorname{diag}(\pi_0, \dots, \pi_N)$. We then have $\boldsymbol{D}^{1/2} = \operatorname{diag}(\sqrt{\pi_0}, \dots, \sqrt{\pi_N})$ and $\boldsymbol{D}^{-1/2} = \operatorname{diag}(\sqrt{1/\pi_0}, \dots, \sqrt{1/\pi_N})$. The relations (3.7) imply that the matrix $\boldsymbol{T} = \boldsymbol{D}^{1/2}\boldsymbol{Q}\boldsymbol{D}^{-1/2}$ is symmetric, and, therefore, there exists a real orthonormal matrix $\boldsymbol{U}$ such that $\boldsymbol{U}\boldsymbol{T}\boldsymbol{U}^{-1}$ is diagonal (note that $\boldsymbol{U}^{-1}$ is also orthonormal). Thus, the rows of $\boldsymbol{U}\boldsymbol{D}^{1/2}$ are left eigenvectors of $\boldsymbol{Q}$. The spectral norm is unitarily invariant, and, therefore, $\|\boldsymbol{V}\boldsymbol{A}\hat{\boldsymbol{V}}\|_2 = \|\boldsymbol{A}\|_2$ for every square real $\boldsymbol{A}$ and for all real orthonormal $\boldsymbol{V}$, $\hat{\boldsymbol{V}}$ [11]. Using the analogue of (3.4) for the spectral norm, we obtain that

$$\min_{0\le j\le M} |\lambda_j - \mu| \le \|\boldsymbol{U}\boldsymbol{D}^{1/2}\boldsymbol{I}\|_2\, \|\boldsymbol{I}\boldsymbol{D}^{-1/2}\boldsymbol{U}^{-1}\|_2\, \|\boldsymbol{E}\|_2 = \sqrt{\max_{i,j\in\mathcal{S}} \frac{\pi_i}{\pi_j}}\,\|\boldsymbol{E}\|_2.$$

The theorem follows from this bound and the inequality $\|\cdot\|_2 \le \sqrt{n}\|\cdot\|$, which holds for all real $n \times n$ matrices.

Remark 3.2. Statements similar to Propositions 3.1 and 3.2 can be proved using (3.9).

The square-root factor in (3.8) is related to the spread of the stationary probabilities:

$$\max_{i,j\in\mathcal{S}} \frac{\pi_i}{\pi_j} < \max_{i\in\mathcal{S}} \frac{1}{\pi_i} \le (N+1)\max_{i,j\in\mathcal{S}} \frac{\pi_i}{\pi_j}, \tag{3.10}$$

which follows from the inequality $\max_{i\in\mathcal{S}} \pi_i \ge 1/(N+1)$. If $\boldsymbol{Q}$ is symmetric, then the conditions (3.7) imply that $\pi_i = 1/(N+1)$ for all $i \in \mathcal{S}$; in this case, equality is attained in the second inequality in (3.10). Let Γ be a nondirected graph with vertex set $\mathcal{S}$ and edge set $\mathcal{E}$ such that, for all $i, j \in \mathcal{S}$ with $i \ne j$, $(i, j) \in \mathcal{E}$ if and only if $q_{ij} > 0$. For distinct states $i, j \in \mathcal{S}$, let $i, v_1, v_2, \dots, v_d, j$ be a path from i to j in Γ. The relations (3.7) imply that

$$\max_{q_{ij}>0} \frac{q_{ij}}{q_{ji}} \le \max_{i,j\in\mathcal{S}} \frac{\pi_i}{\pi_j} = \max_{\substack{i,j\in\mathcal{S}\\ i\ne j}} \frac{q_{v_1 i} q_{v_2 v_1} \cdots q_{v_d v_{d-1}} q_{j v_d}}{q_{i v_1} q_{v_1 v_2} \cdots q_{v_{d-1} v_d} q_{v_d j}}. \tag{3.11}$$

In the inequality in (3.11), equality is attained if and only if $\max_{i,j\in\mathcal{S}} \pi_i/\pi_j = \pi_{i_0}/\pi_{j_0}$ and $q_{i_0 j_0} > 0$ for some $i_0, j_0 \in \mathcal{S}$.

If $\boldsymbol{Q}$ is almost symmetric, then the right-hand side of (3.11) is near 1 and we can expect relatively low sensitivity of the eigenvalues to perturbations, as well as a not very large square-root factor (compared to $\sqrt{N}$) in the convergence bound (3.8). If $\boldsymbol{Q}$ is far from symmetric, then the left-hand side of (3.11) and the spread of the π_i will be large, forcing the square-root factors in (3.8) and (3.9) to be large.

4. Comparing the convergence bounds

The bounds (3.1) and (3.8) belong to the family of eigenvalue convergence bounds [6], [7]. To prove the inequality (2.15), we need to consider the reduced system of forward Kolmogorov equations and make use of the notion of logarithmic norm; see [9], [19]. The derivation of (3.1) involves a similarity transformation; the same is true for (2.15) and (3.8). In the case of (2.15), it is the transformation $\boldsymbol{B} \mapsto \boldsymbol{G}\boldsymbol{B}\boldsymbol{G}^{-1}$, where $\boldsymbol{B}$ is the reduced transposed generator of the chain $X(t)$. This transformation does not diagonalize $\boldsymbol{B}$. For (3.8), the transformation is the same as in the proof of Theorem 3.2, with $\boldsymbol{U}\boldsymbol{D}^{1/2}$ being the transformation matrix. The inequalities (3.10) relate the square-root factor in (3.8) to the spectral-norm condition number of $\boldsymbol{U}\boldsymbol{D}^{1/2}$. Thus, in all of the considered convergence bounds, the rate of convergence can be estimated in terms of condition numbers of the transformation matrices.

It can be proved that a set of g_k needed for (2.15) always exists, but determining such a set may require guessing. The quantity $\underline{\alpha}$ is a lower bound on the spectral gap, λ, and there exists a set of g_k such that $\underline{\alpha} = \lambda$; this set can be found by solving a system of nonlinear algebraic equations [8], [10]. The idea behind (2.15) can be extended to general continuous-time Markov chains, but the question of finding the corresponding transformation matrix has not been addressed [1].

The use of eigenvalue bounds on the rate of convergence to stationarity is somewhat more straightforward, and they are widely applicable (note that (3.8) can be generalized to nonreversible chains [7]). Eigenvalue convergence bounds can be used in combination with estimates for the spectral gap such as the Poincaré and the Cheeger bounds; see [6] and [7] for an introduction to these techniques. There also exist related approaches to estimating the rate of convergence that make use of logarithmic Sobolev inequalities [4] and Nash inequalities [5].

A natural question that can be asked about the convergence bounds discussed above is: which bound is the most accurate? Here we show that there is no universal answer to this question.

Let $X(t)$ be a birth–death process with birth rates $b_i = k_+(N-i)$ and death rates $d_i = k_- i$ for $i \in \mathcal{S}$, where $k_+, k_- > 0$. This is a special case of the Prendiville process [22]. The Prendiville process is widely used as a stochastic model in various fields of science. As an example, consider the chemical reaction $A + B \rightleftharpoons AB$. Suppose that the reaction cell contains so many molecules of species A that the changes in the concentration of A in the course of the reaction can be neglected. If N is the number of molecules of B in the cell, then the number of complexes AB may be described by the Prendiville process (see e.g. [2]).

If $N = 1$, we can analyze the stability of the Prendiville process via (2.5); in this case, $\beta(t) = \mathrm{e}^{-(k_+ + k_-)t}$. The cases when $N > 1$ require estimation of $\beta(t)$. Since $X(t)$ is a birth–death process, it is reversible and the eigenvalues of $\boldsymbol{Q}$ are distinct, which means that all of the convergence bounds of Sections 2 and 3 are applicable. We numerically compare three bounds of the form (2.1):

$$\|\boldsymbol{p}_1(t) - \boldsymbol{p}_2(t)\| \le C_n \mathrm{e}^{-r_n t}, \qquad t \ge 0,\ n = 1, 2, 3,$$

where

$$\begin{aligned} C_1 &= 4\kappa_\infty(\boldsymbol{G}), & r_1 &= \underline{\alpha}, \\ C_2 &= 2(\|\boldsymbol{X}^{-1}\| - 1)\ (\|\boldsymbol{x}_i\| = 1,\ i \in \mathcal{S}), & r_2 &= \lambda; \\ C_3 &= 2\sqrt{\max_{i \in \mathcal{S}} \frac{1}{\pi_i} - 1}, & r_3 &= \lambda; \end{aligned}$$

see (2.15), (3.1) and (3.8). In (2.14), we choose $g_k = 1$ for $k = 1, \dots, N-1$; this gives $\underline{\alpha} = k_+ + k_-$. It follows from Theorem 3 of [8] that $k_+ + k_- = \lambda$; since, for our choice of the g_k, $\|\boldsymbol{G}\|_\infty = N$ and $\|\boldsymbol{G}^{-1}\|_\infty = 2$, we put $C_1 = 8N$ and $r_1 = \lambda$. Thus, all we need to do is to compare the C_n. The parameter values for the numerical experiments and the corresponding values of the C_n are given in Table 1. Numerical experiments show that, for any permutation, σ, of the numbers 1, 2, 3, there exists a set of the parameter values such that $C_{\sigma(1)} < C_{\sigma(2)} < C_{\sigma(3)}$.

TABLE 1: Numerical experiments.

N	k_+	k_-	C_1	C_2	C_3
2	13	2	16	12.018	14.866
2	15	2	16	12.457	16.882
3	4	2	24	13.630	10.198
5	4	2	40	50.568	31.113
6	4	2	48	91.128	53.963
4	10	2	32	67.901	71.972

Acknowledgement

The author would like to thank the referee for valuable comments on earlier versions of the paper.

References

[1] ANDREEV, D. B. *et al.* (2002). On ergodicity and stability estimates for some nonhomogeneous Markov chains. *J. Math. Sci. (New York)* **112,** 4111–4118.
[2] ARAKELIAN, V. B., WILD, J. R. AND SIMONIAN, A. L. (1998). Investigation of stochastic fluctuations in the signal formation of microbiosensors. *Biosensors Bioelectron.* **13,** 55–59.
[3] BALL, F. G., MILNE, R. K. AND YEO, G. F. (2000). Stochastic models for systems of interacting ion channels. *IMA J. Math. Appl. Med. Biol.* **17,** 263–293.
[4] DIACONIS, P. AND SALOFF-COSTE, L. (1996). Logarithmic Sobolev inequalities for finite Markov chains. *Ann. Appl. Prob.* **6,** 695–750.
[5] DIACONIS, P. AND SALOFF-COSTE, L. (1996). Nash inequalities for finite Markov chains. *J. Theoret. Prob.* **9,** 459–510.
[6] DIACONIS, P. AND STROOK, D. (1991). Geometric bounds for eigenvalues of Markov chains. *Ann. Appl. Prob.* **1,** 36–61.
[7] FILL, J. A. (1991). Eigenvalue bounds on convergence to stationarity for nonreversible Markov chains, with an application to the exclusion process. *Ann. Appl. Prob.* **1,** 62–87.
[8] GRANOVSKY, B. L. AND ZEIFMAN, A. I. (1997). The decay function of nonhomogeneous birth–death processes, with application to mean-field models. *Stoch. Process. Appl.* **72,** 105–120.
[9] GRANOVSKY, B. L. AND ZEIFMAN, A. I. (2000). Nonstationary Markovian queues. *J. Math. Sci. (New York)* **99,** 1415–1438.
[10] GRANOVSKY, B. L. AND ZEIFMAN, A. I. (2000). The N-limit of spectral gap of a class of birth–death Markov chains. *Appl. Stoch. Models Business Industry* **16,** 235–248.
[11] HORN, R. A. AND JOHNSON, C. R. (1985). *Matrix Analysis*. Cambridge University Press.
[12] KARTASHOV, N. V. (1986). Inequalities in theorems of ergodicity and stability for Markov chains with common state space. I. *Theory Prob. Appl.* **30,** 247–259.
[13] KARTASHOV, N. V. (1986). Inequalities in theorems of ergodicity and stability for Markov chains with common state space. II. *Theory Prob. Appl.* **30,** 507–515.
[14] PENG, N. F. (1996). Spectral representations of the transition probability matrices for continuous time finite Markov chains. *J. Appl. Prob.* **33,** 28–33.
[15] SENETA, E. (1979). Coefficients of ergodicity: structure and applications. *Adv. Appl. Prob.* **11,** 576–590.
[16] THOUMINE, O. AND MEISTER, J.-J. (2000). A probabilistic model for ligand-cytoskeleton transmembrane adhesion: predicting the behaviour of microspheres on the surface of migrating cells. *J. Theoret. Biol.* **204,** 381–392.
[17] VAN KAMPEN, N. G. (1981). *Stochastic Processes in Physics and Chemistry*. North Holland, Amsterdam.
[18] ZEIFMAN, A. I. (1985). Stability for continuous-time nonhomogeneous Markov chains. In *Stability Problems for Stochastic Models* (Lecture Notes Math. **1155**), eds V. V. Kalashnikov and V. M. Zolotarev, Springer, Berlin, pp. 401–414.
[19] ZEIFMAN, A. I. (1995). Upper and lower bounds on the rate of convergence for nonhomogeneous birth and death processes. *Stoch. Process. Appl.* **59,** 157–173.
[20] ZEIFMAN, A. I. (1998). Stability of birth-and-death processes. *J. Math. Sci. (New York)* **91,** 3023–3031.
[21] ZEIFMAN, A. I. AND ISAACSON, D. L. (1994). On strong ergodicity for nonhomogeneous continuous-time Markov chains. *Stoch. Process. Appl.* **50,** 263–273.
[22] ZHENG, Q. (1998). Note on the non-homogeneous Prendiville process. *Math. Biosci.* **148,** 1–5.