

SENSITIVITY AND CONVERGENCE OF UNIFORMLY ERGODIC MARKOV CHAINS

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

Abstract

For uniformly ergodic Markov chains, we obtain new perturbation bounds which relate the sensitivity of the chain under perturbation to its rate of convergence to stationarity. In particular, we derive sensitivity bounds in terms of the ergodicity coefficient of the iterated transition kernel, which improve upon the bounds obtained by other authors. We discuss convergence bounds that hold in the case of finite state space, and consider numerical examples to compare the accuracy of different perturbation bounds.

Keywords: Markov chain; uniform ergodicity; rate of convergence; ergodicity coefficient; perturbation bound; sensitivity analysis

2000 Mathematics Subject Classification: Primary 60J05

Secondary 60J10; 60J35

1. Introduction

There exists an extensive literature on perturbation bounds for Markov chains. One group of results concerns the sensitivity of the stationary distribution of a finite, homogeneous Markov chain, and the bounds are derived using methods of matrix analysis; see the review [2] and recent papers [12]–[14], [21] (also see [29] for a graph-theoretic approach). Another group includes perturbation bounds for finite-time and invariant distributions of Markov chains with general state space; see [1], [9], [10] (the results of [9] and [10] are also described in [11]). In [1] and [9], the bounds for general Markov chains are expressed in terms of ergodicity coefficients of the iterated transition kernel, which are difficult to compute for infinite state spaces. These results were obtained using operator-theoretic and probabilistic methods. Here, we improve upon such bounds and also prove new bounds which explicitly relate the sensitivity under perturbation to geometric convergence estimates. Bounds of the latter kind have not previously appeared in the discrete-time Markov chain literature. Much has been done to investigate the relationship between the convergence rate and the sensitivity under perturbation for Markov chains in continuous time; see [20] and references therein. In this paper, we extend the approach to sensitivity analysis for Markov chains developed in [20] to the case of discrete-time chains.

The need for perturbation bounds for uniformly ergodic general Markov chains comes from the fact that many practically important chains are uniformly ergodic. In the area of Markov chain Monte Carlo (MCMC) methods, it has been shown that some special cases of the Gibbs sampler [23], [24], the independence Hastings algorithm [16] and the Metropolis algorithm [30] are uniformly ergodic general Markov chains. It is indeed important to study perturbations for MCMC algorithms because running a Markov chain on a computer introduces an error associated with finite-precision computation, so that, in fact, a perturbed chain is simulated [25]. In the modeling context, transition probabilities of Markov chains are often not known

Received 13 April 2004; revision received 24 April 2005.

* Current address: School of Biology, Georgia Institute of Technology, 310 Ferst Drive, Atlanta, GA 30332, USA.
Email address: amit@amber.biology.gatech.edu

exactly due to measurement errors, which necessitates the use of perturbation bounds to assess the uncertainties in the model output. As a recent example, we mention the application of perturbation bounds for Markov chains in bioinformatics [27].

The paper is organized as follows. Section 2 contains the necessary definitions and notation. In Section 3, we give new perturbation bounds for Markov chains with general state space. In Section 4, we discuss two approaches to bounding the speed of convergence in the case of finite state space. Finally, in Section 5, we numerically compare different perturbation bounds. In the appendix, we compare some perturbation bounds of Section 3 with the results obtained in [1] and [9].

2. Preliminaries

Let $\mathbb{N} = \{1, 2, \dots\}$ and $\mathbb{Z}_+ = \{0, 1, 2, \dots\}$. We consider a Markov chain $X = \{X_n, n \in \mathbb{Z}_+\}$ on a measurable space $(\mathcal{S}, \mathfrak{B})$, with transition kernel $P(x, \mathcal{G})$. For a background on general Markov chains and their convergence properties, see [19]. In this paper we deal with operators defined by a kernel which is either a (Markov) transition kernel or a difference of two such kernels. A kernel $A(x, \mathcal{G})$ on $(\mathcal{S}, \mathfrak{B})$ defines an operator, A , acting on a space, \mathcal{M} , of finite signed measures on \mathfrak{B} as follows:

$$(qA)(\mathcal{G}) = \int_{\mathcal{S}} q(dx)A(x, \mathcal{G}), \quad q \in \mathcal{M}, \mathcal{G} \in \mathfrak{B}.$$

If $A(x, \mathcal{G})$ is a transition kernel, then we call A a *transition operator*. For a measure $q \in \mathcal{M}$, define the total variation norm by

$$\|q\| = |q|(\mathcal{S}),$$

that is, $\|q\|$ is the total variation of q on \mathcal{S} . This functional-analytic definition of total variation is twice the quantity $\sup_{\mathcal{G} \in \mathfrak{B}} |q(\mathcal{G})|$, which is often used by probabilists. The corresponding operator norm for A is

$$\|A\| = \sup_{\|q\|=1} \|qA\| = \sup_{x \in \mathcal{S}} \|A_x\|, \tag{2.1}$$

where $A_x(\mathcal{G}) = A(x, \mathcal{G})$, $\mathcal{G} \in \mathfrak{B}$. Thus, $\|q\| = 1$ for a probability measure q , and $\|A\| = 1$ for a transition operator A .

We assume that X is uniformly ergodic with unique invariant measure π . By definition, uniform ergodicity means that there exist positive constants $\rho < 1$ and $C < \infty$ such that, for all $x \in \mathcal{S}$,

$$\|e_x P^n - \pi\| \leq C\rho^n, \quad n \in \mathbb{Z}_+, \tag{2.2}$$

where e_x is the probability measure concentrated at x . A classical bound of the form (2.2) is obtained through minorization conditions [19, Chapter 16] (bounds of this type were used in [16] and [30]). A new method of deriving uniform geometric bounds (using a drift condition) was developed in [18].

For a probability measure p_0 , define $p_n = p_0 P^n$ (the distribution of X after n steps). We approximate p_n by the distribution of a Markov chain $\tilde{X} = \{\tilde{X}_n, n \in \mathbb{Z}_+\}$ on $(\mathcal{S}, \mathfrak{B})$; \tilde{P} denotes the transition operator of \tilde{X} and \tilde{p}_n denotes the distribution of \tilde{X} after n steps. The chain X is regarded as the unperturbed Markov chain, whereas \tilde{X} is a perturbed Markov chain. Perturbation bounds studied in this paper are upper bounds on the variation distance between \tilde{p}_n and p_n .

To obtain our perturbation bounds, we shall use the notion of the ergodicity coefficient of a transition operator. For a transition operator A , define the ergodicity coefficient by

$$\tau(A) := \sup_{q \in \mathcal{M}_0} \|qA\| = \frac{1}{2} \sup_{x,y \in \mathcal{S}} \|(e_x - e_y)A\|, \tag{2.3}$$

where $\mathcal{M}_0 = \{q \in \mathcal{M} : \|q\| = 1, q(\mathcal{S}) = 0\}$ [4], [5]. We have $\tau(A) \leq 1$ and, for transition operators A_1 and A_2 , the following submultiplicativity property holds:

$$\tau(A_1A_2) \leq \tau(A_1)\tau(A_2). \tag{2.4}$$

If \mathcal{S} is finite, then we can define a measure q on \mathcal{S} by defining a vector $\mathbf{q} = (q_i)$, where $q_i = q(\{i\})$, $i \in \mathcal{S}$ (we regard vectors as row vectors). An operator A can be defined by the matrix $\mathbf{A} = (a_{ij})$, where $a_{ij} = (e_iA)(\{j\})$, $i, j \in \mathcal{S}$. To apply our sensitivity bounds to finite chains, we must use the l_1 -norm for vectors (absolute entry sum) and the corresponding induced norm for matrices (maximum absolute row sum); these norms are also denoted by $\|\cdot\|$, which will cause no confusion. It is easily seen that $\|q\| = \|\mathbf{q}\|$ and $\|A\| = \|\mathbf{A}\|$. If A is a transition operator then, using the notation $\tau(\cdot)$ for matrices as well, we have

$$\tau(A) = \tau(\mathbf{A}) := \sup_{\substack{\|\mathbf{v}\|=1 \\ \mathbf{v}\mathbf{e}^T=0}} \|\mathbf{v}\mathbf{A}\| = \frac{1}{2} \max_{i,j \in \mathcal{S}} \|(e_i - e_j)\mathbf{A}\|, \tag{2.5}$$

where the supremum is taken over real vectors \mathbf{v} ; here \mathbf{e} is the row vector of 1s, e_i is the vector whose i th entry is equal to 1 with all other entries equal to 0, and T denotes transpose. The quantity $\tau(\mathbf{A})$ is the l_1 -norm ergodicity coefficient of the matrix \mathbf{A} [26]. It should be noted that (2.5) can be straightforwardly generalized to the case of countably infinite state space (in this connection, see [4] and [5]).

In what follows, $\lfloor x \rfloor$ is the largest integer less than or equal to x and $\lceil x \rceil$ is the smallest integer greater than or equal to x .

3. Perturbation bounds and uniform ergodicity

In this section, we prove perturbation bounds which relate the sensitivity of X to the speed of its convergence in the case of general state space. Our primary goal is to provide bounds in terms of C and ρ under the condition that (2.2) holds, but we also give several bounds in terms of $\tau_m := \tau(P^m)$, where $m \in \mathbb{N}$ is such that $\tau_m < 1$. It should be noted that the quantity τ_m gauges the speed of convergence to stationarity through the inequalities

$$\|p_n - \pi\| \leq \tau_m^{\lfloor n/m \rfloor} \|p_0 - \pi\| \leq 2\tau_m^{\lfloor n/m \rfloor}, \quad n \in \mathbb{N},$$

which follow from the definition of $\tau(\cdot)$, (2.4), and the fact that $\tau(P^i) \leq 1$ for all $i \in \mathbb{Z}_+$.

Theorem 3.1. *If (2.2) holds then $C \geq 1$ and, for $z_n := \tilde{p}_n - p_n$ and $E := \tilde{P} - P$, we have*

$$\|z_n\| \leq \begin{cases} \|z_0\| + n\|E\|, & n \leq \hat{n}, \\ C\rho^n \|z_0\| + \left(\hat{n} + C\frac{\rho^{\hat{n}} - \rho^n}{1 - \rho}\right)\|E\|, & n \geq \hat{n} + 1, \end{cases} \tag{3.1}$$

where $\hat{n} = \lceil \log_\rho C^{-1} \rceil$.

Proof. For each $n \in \mathbb{N}$,

$$z_n = z_0 P^n + \sum_{i=0}^{n-1} \tilde{p}_i E P^{n-i-1}, \tag{3.2}$$

which easily follows by induction. Since $E(x, \mathfrak{J}) = 0$ for all $x \in \mathfrak{J}$, we have $(\tilde{p}_i E)(\mathfrak{J}) = 0$ for all $i \in \mathbb{Z}_+$. This, together with the inequality

$$\|\tilde{p}_i E P^{n-i-1}\| \leq \|E\| \left\| \frac{\tilde{p}_i E}{\|\tilde{p}_i E\|} P^{n-i-1} \right\| \quad \text{for } \|\tilde{p}_i E\| \neq 0,$$

gives $\|\tilde{p}_i E P^{n-i-1}\| \leq \|E\| \tau(P^{n-i-1})$. Similarly, $\|z_0 P^n\| \leq \|z_0\| \tau(P^n)$. From (3.2), using the above relations we obtain

$$\|z_n\| \leq \|z_0\| \tau(P^n) + \|E\| \sum_{i=0}^{n-1} \tau(P^i), \quad n \in \mathbb{N}. \tag{3.3}$$

Since $\tau(P^n) \leq 1$, we obtain

$$\sum_{i=0}^{n-1} \tau(P^i) \leq n, \quad n \in \mathbb{N}. \tag{3.4}$$

From (2.2), (2.3), and the triangle inequality, we deduce that

$$\tau(P^n) \leq \sup_{x \in \mathfrak{J}} \|e_x P^n - \pi\| \leq C \rho^n, \quad n \in \mathbb{Z}_+. \tag{3.5}$$

The relations (3.5) with $n = 0$ and the equality $\tau(P^0) = 1$ imply that $C \geq 1$. Noting the fact that $C \rho^n \leq 1$ for $n \geq \hat{n}$, and using (2.2), (3.4), and (3.5), we obtain

$$\begin{aligned} \sum_{i=0}^{n-1} \tau(P^i) &\leq \hat{n} + C \sum_{i=\hat{n}}^{n-1} \rho^i \\ &= \hat{n} + C \rho^{\hat{n}} \sum_{i=0}^{n-\hat{n}-1} \rho^i \\ &= \hat{n} + C \rho^{\hat{n}} \frac{1 - \rho^{n-\hat{n}}}{1 - \rho}, \quad n \geq \hat{n} + 1. \end{aligned} \tag{3.6}$$

This, together with (2.2) and (3.3)–(3.5), proves (3.1).

Corollary 3.1. *In the setting of Theorem 3.1,*

$$\sup_{n \in \mathbb{N}} \|z_n\| \leq \|z_0\| + \left(\hat{n} + \frac{C \rho^{\hat{n}}}{1 - \rho} \right) \|E\|. \tag{3.7}$$

If, in addition, the chain \tilde{X} has an invariant measure, $\tilde{\pi}$, then

$$\|\tilde{\pi} - \pi\| \leq \left(\hat{n} + \frac{C \rho^{\hat{n}}}{1 - \rho} \right) \|E\|. \tag{3.8}$$

Proof. The bound (3.7) follows directly from (3.1). Setting $p_0 = \pi$ and $\tilde{p}_0 = \tilde{\pi}$, and taking the limit as $n \rightarrow \infty$ in (3.3), we obtain

$$\|\tilde{\pi} - \pi\| \leq \|E\| \sum_{i=0}^{\infty} \tau(P^i).$$

This relation and (3.6) prove (3.8).

The inequality (3.3) also allows us to obtain perturbation bounds in terms of the ergodicity coefficient of the iterated transition kernel. Such bounds are provided by the following theorem.

Theorem 3.2. *If (2.2) holds then there exists a positive integer m such that $\tau_m < 1$ and*

$$\sup_{k \in \mathbb{N}} \|z_{km}\| \leq \tau_m \|z_0\| + \frac{\|\tilde{P}^m - P^m\|}{1 - \tau_m}, \tag{3.9}$$

$$\|z_n\| \leq \begin{cases} \|z_0\| + \max_{0 < i < m} \|\tilde{P}^i - P^i\|, & n < m, \\ \tau_m \left(\|z_0\| + \max_{0 < i < m} \|\tilde{P}^i - P^i\| \right) + \frac{\|\tilde{P}^m - P^m\|}{1 - \tau_m}, & n \geq m. \end{cases} \tag{3.10}$$

If, in addition, the chain \tilde{X} has an invariant measure, $\tilde{\pi}$, then

$$\|\tilde{\pi} - \pi\| \leq \frac{\|\tilde{P}^m - P^m\|}{1 - \tau_m}. \tag{3.11}$$

Proof. The existence of a positive integer m , such that $\tau_m < 1$, follows from (3.5). In the case when $\tau_1 < 1$, from (3.3) we obtain

$$\sup_{n \in \mathbb{N}} \|z_n\| \leq \tau_1 \|z_0\| + \frac{\|E\|}{1 - \tau_1};$$

if $\tilde{\pi}$ is an invariant measure of \tilde{X} , then

$$\|\tilde{\pi} - \pi\| \leq \frac{\|E\|}{1 - \tau_1}.$$

These inequalities, applied to the m -skeletons of X and \tilde{X} , give (3.9) and (3.11). For every $k \in \mathbb{N}$ and every integer i such that $1 \leq i \leq m - 1$, we obtain

$$z_{mk+i} = \tilde{p}_i(\tilde{P}^{mk} - P^{mk}) + z_i P^{mk};$$

therefore,

$$\|z_{mk+i}\| \leq \|\tilde{P}^{mk} - P^{mk}\| + \|z_i\| \tau_m^k. \tag{3.12}$$

By (2.1), we have

$$\|\tilde{P}^{mk} - P^{mk}\| = \sup_{x \in \mathcal{S}} \|e_x(\tilde{P}^{mk} - P^{mk})\| = \sup_{x \in \mathcal{S}} \|z_{mk}^{(x)}\|, \tag{3.13}$$

where $z_{mk}^{(x)}$ is the value of z_{mk} corresponding to the case $\tilde{p}_0 = p_0 = e_x$. Since $z_n = z_0 P^n + \tilde{p}_0(\tilde{P}^n - P^n)$,

$$\|z_n\| \leq \|z_0\| + \max_{0 < i < m} \|\tilde{P}^i - P^i\|, \quad n < m. \tag{3.14}$$

Combining (3.9) and (3.12)–(3.14), we obtain (3.10).

The following theorem gives an alternative method of obtaining perturbation bounds in terms of τ_m .

Theorem 3.3. *For an $m \in \mathbb{N}$ such that $\tau_m < 1$, the following bound holds:*

$$\|z_n\| \leq \tau_m^{\lfloor n/m \rfloor} \left(\|z_0\| + \max_{0 < i < m} \|\tilde{P}^i - P^i\| \right) + \frac{1 - \tau_m^{\lfloor n/m \rfloor}}{1 - \tau_m} \|\tilde{P}^m - P^m\|, \quad n \in \mathbb{N}. \quad (3.15)$$

Proof. If $n < m$, then (3.15) reduces to (3.14). Suppose now that $n \geq m$. We obtain

$$z_n = \tilde{p}_{n-m} \tilde{P}^m - p_{n-m} P^m = (\tilde{p}_{n-m} - p_{n-m}) P^m + p_{n-m} (P^m - \tilde{P}^m);$$

therefore,

$$\|z_n\| \leq \|z_{n-m}\| \tau_m + \|\tilde{P}^m - P^m\|.$$

Applying this relation successively to $\|z_n\|, \|z_{n-m}\|, \dots, \|z_{n-m(\lfloor n/m \rfloor - 1)}\|$, and using (3.14) to bound $\|z_{n-m\lfloor n/m \rfloor}\|$, we obtain

$$\|z_n\| \leq \tau_m^{\lfloor n/m \rfloor} \left(\|z_0\| + \max_{0 < i < m} \|\tilde{P}^i - P^i\| \right) + (\tau_m^{\lfloor n/m \rfloor - 1} + \tau_m^{\lfloor n/m \rfloor - 2} + \dots + 1) \|\tilde{P}^m - P^m\|,$$

which completes the proof.

Corollary 3.2. *For an $m \in \mathbb{N}$ such that $\tau_m < 1$, the following bound holds:*

$$\sup_{n \in \mathbb{N}} \|z_n\| \leq \sup_{n \in \mathbb{N}} \tau_m^{\lfloor n/m \rfloor} \|z_0\| + \frac{m \|E\|}{1 - \tau_m}. \quad (3.16)$$

Proof. We have

$$\max_{0 < i \leq m} \|\tilde{P}^m - P^m\| \leq m \|E\|, \quad (3.17)$$

which follows by induction from the relation

$$\tilde{P}^i - P^i = \tilde{P} \tilde{P}^{i-1} - P P^{i-1} = \tilde{P} (\tilde{P}^{i-1} - P^{i-1}) + E P^{i-1}, \quad i \geq 1.$$

From (3.15) and (3.17), we obtain

$$\|z_n\| \leq \tau_m^{\lfloor n/m \rfloor} \|z_0\| + m \|E\| \frac{1 - \tau_m^{\lfloor n/m \rfloor + 1}}{1 - \tau_m}, \quad n \in \mathbb{N},$$

which implies (3.16).

Remark 3.1. Theorem 3.3 allows us to obtain the bounds (3.10) and (3.11) as corollaries.

4. The case of finite state space

In this section, we assume that $\mathcal{S} = \{1, \dots, N\}$, $2 \leq N < \infty$. Thus, X is a finite Markov chain with transition probability matrix P . Denote the distribution vector of X_n by p_n ; let π be the stationary distribution of X . Note that, for finite Markov chains, uniform ergodicity is equivalent to ergodicity (convergence of all distributions to a unique limiting distribution as $n \rightarrow \infty$).

Since P is the transition matrix of a Markov chain, one of its eigenvalues is equal to 1, and the absolute values of all nonunit eigenvalues do not exceed 1. Uniqueness of the invariant

distribution implies that the multiplicity of the unit eigenvalue of \mathbf{P} is equal to 1. Denote by $\beta_j, j = 1, \dots, M$, the distinct eigenvalues of \mathbf{P} . Choose $\beta_1 = 1$ and let $\beta = \max_{2 \leq j \leq M} |\beta_j|$. Our assumption that X is ergodic is equivalent to having $\beta < 1$.

Suppose that \mathbf{P} is diagonalizable. Let \mathbf{W} be a nonsingular matrix such that $\mathbf{P} = \mathbf{W}^{-1} \mathbf{B} \mathbf{W}$, where \mathbf{B} is diagonal. It is clear that the rows of \mathbf{W} , denoted by \mathbf{w}_i , are left eigenvectors of \mathbf{P} , and that \mathbf{B} consists of the eigenvalues of \mathbf{P} . For each β_j , define $\mathcal{B}_j = \{k \in \mathcal{S} : \mathbf{w}_k \mathbf{P} = \beta_j \mathbf{P}\}$. Set $\kappa(\mathbf{W}) = \|\mathbf{W}\| \|\mathbf{W}^{-1}\|$. The following theorem is the discrete-time counterpart of Theorem 3.1 of [20].

Theorem 4.1. *If \mathbf{P} is diagonalizable then, for all distributions \mathbf{p}_0 ,*

$$\|\mathbf{p}_n - \boldsymbol{\pi}\| \leq D\beta^n, \quad n \in \mathbb{Z}_+, \tag{4.1}$$

where

$$D = \max_{i \in \mathcal{S}} \sum_{j=2}^M \left\| \sum_{k \in \mathcal{B}_j} w_{ik}^{(-1)} \mathbf{w}_k \right\| = \max_{i \in \mathcal{S}} \sum_{j=2}^M \left\| \mathbf{e}_i \frac{\prod_{k \neq j} (\mathbf{P} - \beta_k \mathbf{I})}{\prod_{k \neq j} (\beta_j - \beta_k)} \right\|,$$

the $w_{ik}^{(-1)}$ are the entries of \mathbf{W}^{-1} , and \mathbf{I} denotes the identity matrix. Also,

$$D \leq \kappa(\mathbf{W}) - 1, \tag{4.2}$$

and equality is attained if all eigenvalues of \mathbf{P} are distinct and $\|\mathbf{w}_i\| = 1, i \in \mathcal{S}$. In (4.2), the scaling $\|\mathbf{w}_i\| = 1$ for all $i \in \mathcal{S}$ minimizes $\kappa(\mathbf{W})$.

Proof. The theorem can be proved in a similar way to Theorem 3.1 of [20], using the following expression [17, Chapter 7]:

$$\mathbf{P}^n = \sum_{j=1}^M \frac{\prod_{k \neq j} (\mathbf{P} - \beta_k \mathbf{I})}{\prod_{k \neq j} (\beta_j - \beta_k)} \beta_j^n, \quad n \in \mathbb{Z}_+.$$

Theorem 4.1 demonstrates the connection between the rate of convergence of X to stationarity and the sensitivity of the eigenvalues of \mathbf{P} under perturbation since, for every matrix $\mathbf{P} + \mathbf{F}$ with an eigenvalue ζ , the following inequality holds [7, Chapter 6]:

$$\min_{1 \leq j \leq M} |\beta_j - \zeta| \leq \kappa(\mathbf{W}) \|\mathbf{F}\|. \tag{4.3}$$

If $\mathbf{P} + \mathbf{F}$ is the transition matrix of \tilde{X} , then (4.3) can be used to study different spectrum-related properties of \tilde{X} ; see [20] for a discussion in the case of continuous time. But, since (4.3) holds for arbitrary \mathbf{F} , it can be applied in situations where perturbations of \mathbf{P} are not structured. In particular, this is the case when numerical methods are used for computing the eigenvalues of \mathbf{P} . As is well known, knowledge of the eigenvalues and eigenvectors of a diagonalizable matrix allows us to calculate the matrix powers in an effective way through the spectral decomposition, and finding \mathbf{P}^n is of course important.

The following theorem provides a convergence bound for irreducible chains. Supposing that X is irreducible (\mathbf{P} is not necessarily diagonalizable), define the matrix $\tilde{\mathbf{P}} = (\tilde{p}_{ij})$ by $\tilde{p}_{ij} = p_{ji} \pi_j / \pi_i$, where p_{ij} are the entries of \mathbf{P} and π_i are the entries of $\boldsymbol{\pi}$. Let $\theta_i, i = 1, \dots, N$, be the eigenvalues of the stochastic matrix $\tilde{\mathbf{P}} \tilde{\mathbf{P}}$. Put $\theta_1 = 1$ and $\theta = \max_{2 \leq i \leq N} \theta_i$ (the eigenvalues θ_i are all real and nonnegative [6]). The following theorem is implied by Theorem 2.7 of [6].

Theorem 4.2. For all p_0 ,

$$\|p_n - \pi\| \leq G\theta^{n/2}, \quad n \in \mathbb{Z}_+, \tag{4.4}$$

where

$$G = \sqrt{\max_{i \in \mathcal{S}} \frac{1}{\pi_i} - 1}.$$

Remark 4.1. The matrix $P\bar{P}$ is the multiplicative reversibilization of P ; it can be shown that sometimes it can be reducible, in which case $\theta = 1$. To overcome this difficulty, we may use the additive reversibilization, $(P + \bar{P})/2$, which is irreducible and aperiodic if P is; see [6].

The chain X is reversible if it is irreducible and the detailed balance conditions

$$\pi_i p_{ij} = \pi_j p_{ji}$$

are satisfied for all $i, j \in \mathcal{S}$. The detailed balance conditions are equivalent to $\bar{P} = P$. Thus, for reversible chains, $\sqrt{\theta} = \beta$. Many practically important chains are reversible; as an example, we mention MCMC algorithms for a finite state space [8], [28]. For a reversible chain X , there exists a connection between the rate of convergence to stationarity given by Theorem 4.2, the stability of the eigenvalues of the transition matrix, and the closeness of the latter to being symmetric. See [20] for a discussion of these issues in the case of continuous time; the case of discrete time can be treated in a similar manner.

Remark 4.2. The exact values of β or θ may not be easily available, and it is desirable to have bounds on these quantities. For different approaches, see [3], [6], [28].

It is interesting to note that there exists a class of irreducible, aperiodic Markov chain for which sensitivity bounds have been derived with ‘condition number’ depending only on the cardinality of the state space. These are the so-called Markov chains whose transition matrix has large exponent (another term for ‘exponent’ is ‘primitivity index’) [15]. But many chains that model real-world phenomena do not belong to this class; e.g. if at least one diagonal element of an irreducible transition matrix is positive, then the matrix does not have large exponent (see [7, Chapter 8]).

5. Numerical results

In this section, we provide some examples of the use of the perturbation bounds of Section 3 and the convergence bounds of Section 4. We shall work with the following two matrices which arise in the context of ion channel modeling [22, p. 54]:

$$P_1 = \begin{pmatrix} 0.9841 & 0.0119 & 0.0040 \\ 0.3178 & 0.5764 & 0.1058 \\ 0.0126 & 0.0349 & 0.9525 \end{pmatrix}, \quad P_2 = \begin{pmatrix} 0.999 & 0.001 & 0 \\ 0.025 & 0.95 & 0.025 \\ 0 & 0.03 & 0.97 \end{pmatrix}.$$

The matrix P_1 has been slightly modified in order to make the row sums equal to 1. We have $\tau(P_1) \approx 0.972$ and $\tau(P_2) \approx 0.999$. It is also important to note that P_2 is reversible.

For the chain X with transition matrix \mathbf{P} such that $\tau(\mathbf{P}) < 1$, consider the following quantities:

$$\begin{aligned}
 B_1(\mathbf{P}) &= \frac{1}{1 - \tau(\mathbf{P})}, \\
 B_2(\mathbf{P}) &= \hat{n}_2 + \frac{D\beta^{\hat{n}_2}}{1 - \beta}, & \hat{n}_2 &= \lceil \log_{\beta} D^{-1} \rceil, \\
 B_3(\mathbf{P}) &= \hat{n}_3 + \frac{G\theta^{\hat{n}_3/2}}{1 - \sqrt{\theta}}, & \hat{n}_3 &= \lceil \log_{\sqrt{\theta}} G^{-1} \rceil.
 \end{aligned}$$

Let $\mathbf{z}_n = \tilde{\mathbf{p}}_n - \mathbf{p}_n$, where $\tilde{\mathbf{p}}_n$ is the distribution of \tilde{X} after n steps. Also, let $\mathbf{E} = \tilde{\mathbf{P}} - \mathbf{P}$, where $\tilde{\mathbf{P}}$ is the transition matrix of \tilde{X} . In this section, we consider the case where $\|\mathbf{z}_0\| = 0$. In this case, for $i = 1, 2, 3$, we obtain

$$\sup_{n \in \mathbb{N}} \|\mathbf{z}_n\| \leq B_i(\mathbf{P}) \|\mathbf{E}\|.$$

The bound for $i = 1$ can be obtained either from (3.9) or from (3.16) by setting $m = 1$; the bounds for $i = 2$ and $i = 3$ are obtained by combining (3.7) with (4.1) and (3.7) with (4.4), respectively. We have

$$\begin{aligned}
 B_1(\mathbf{P}_1) &\approx 35.088, & B_2(\mathbf{P}_1) &\approx 38.277, & B_3(\mathbf{P}_1) &\approx 58.056, \\
 B_1(\mathbf{P}_2) &\approx 1000.000, & B_2(\mathbf{P}_2) &\approx 187.507, & B_3(\mathbf{P}_2) &\approx 233.652.
 \end{aligned}$$

We see that the bounds in terms of $B_2(\cdot)$ and $B_3(\cdot)$ for \mathbf{P}_1 are worse than the one in terms of $B_1(\cdot)$, and that they are better for \mathbf{P}_2 . Now we shall show that a small change in \mathbf{P} may cause a considerable change in the absolute and relative magnitudes of the quantities $B_i(\cdot)$. Consider the matrix

$$\mathbf{P}_3 = \begin{pmatrix} 0.9841 & 0.0119 & 0.0040 \\ 0.3178 & 0.5764 & 0.1058 \\ 0.0006 & 0.0349 & 0.9645 \end{pmatrix}$$

obtained from \mathbf{P}_1 by a slight modification in the third row. This modification leads to an increase in the value of the ergodicity coefficient: $\tau(\mathbf{P}_3) \approx 0.984$. We have

$$B_1(\mathbf{P}_3) \approx 60.606, \quad B_2(\mathbf{P}_3) \approx 50.120, \quad B_3(\mathbf{P}_3) \approx 79.723.$$

Thus, for \mathbf{P}_3 , the bound in terms of $B_2(\cdot)$ is better than the one in terms of $B_1(\cdot)$. Notice that, for all of the matrices $\mathbf{P}_1, \mathbf{P}_2$, and \mathbf{P}_3 , the $B_2(\cdot)$ -bound is sharper than the $B_3(\cdot)$ -bound.

In the bound (3.16), of which the $B_1(\cdot)$ -bound is a special case, we may choose different values of m . Below we show that an increase in m may lead to a better perturbation bound. Let

$$\mathbf{E} = \begin{pmatrix} 0.0001 & -0.0001 & 0 \\ 0.0003 & 0 & -0.0003 \\ 0 & 0 & 0 \end{pmatrix},$$

and define

$$\begin{aligned}
 F_1(\mathbf{P}, \mathbf{E}, m) &= (m - 1)\|\mathbf{E}\| + \frac{\|\tilde{\mathbf{P}}^m - \mathbf{P}^m\|}{1 - \tau(\mathbf{P}^m)}, \\
 F_2(\mathbf{P}, \mathbf{E}, m) &= \frac{m\|\mathbf{E}\|}{1 - \tau(\mathbf{P}^m)}.
 \end{aligned}$$

TABLE 1: Numerical experiments.

m	$F_1(\mathbf{P}_2, \mathbf{E}, m)$	$F_2(\mathbf{P}_2, \mathbf{E}, m)$
3	0.339	0.352
5	0.239	0.257
10	0.143	0.165
20	0.089	0.113
50	0.068	0.087
70	0.072	0.089
100	0.084	0.095
150	0.110	0.113
200	0.138	0.136
300	0.196	0.187

It follows from (3.10), (3.16), and (3.17) that, for $i = 1, 2$,

$$\sup_{n \in \mathbb{N}} \|z_n\| \leq F_i(\mathbf{P}, \mathbf{E}, m).$$

The numerical values for $\mathbf{P} = \mathbf{P}_2$ and \mathbf{E} as above are given in Table 1. We also have

$$B_1(\mathbf{P}_2)\|\mathbf{E}\| \approx 0.600, \quad B_2(\mathbf{P}_2)\|\mathbf{E}\| \approx 0.113, \quad B_3(\mathbf{P}_2)\|\mathbf{E}\| \approx 0.140.$$

Thus, in our example the $F_i(\cdot, \cdot, m)$ -bounds with m of order ten are sharper than the $B_i(\cdot)$ -bounds for $i = 2, 3$, and they are less sharp for small and large m .

The bounds considered above do not require computation of quantities of the type

$$\max_{0 < i < m} \|\tilde{P}^i - P^i\|,$$

as do the bounds (3.10) and (3.15). The latter bounds are more difficult to compute, but they produce sharper estimates of $\sup_{n \in \mathbb{N}} \|z_n\|$. Again, optimal results may correspond to relatively large m .

Appendix A. Anisimov’s and Kartashov’s results

Here we briefly compare some results of Section 3 with the corresponding results of Anisimov [1] and Kartashov [9]. In [1], Anisimov obtained perturbation bounds for inhomogeneous Markov processes in discrete and continuous time, and Kartashov in [9] studied so-called strongly stable Markov chains, a wider class than the uniformly ergodic chains studied in this paper. It should also be noted that Kartashov considered a more general class of norm than the total variation norm that we use here. We shall show that our results improve upon the bounds derived in [1] and [9] for uniformly ergodic Markov chains. In fact, Anisimov and Kartashov obtained their perturbation bounds for quantities of the type $\|\tilde{P}^n - P^n\|$. Here, we treat the case of $\|z_0\| = 0$; in this case,

$$\|z_n\| \leq \|\tilde{P}^n - P^n\|.$$

Theorem 2 of [1] gave the bound

$$\|z_n\| \leq \frac{1 - \tau_m^{\lfloor n/m \rfloor + 1}}{1 - \tau_m} \max_{0 < i \leq m} \|\tilde{P}^i - P^i\|. \tag{A.1}$$

It is easy to see that our bound (3.15) (for $\|z_0\| = 0$) is sharper than (A.1). Also, our derivation of (3.15) is somewhat more straightforward than Anisimov's proof of (A.1). Lemma 1 of [1] gave a bound which in our case reduces to (3.17); the combination of Lemma 1 and the uniform bound implied by (A.1) coincides with our inequality (3.16). Results similar to (3.9)–(3.11) were not considered in [1].

The bound [9, Equation (16)], for

$$\tau_m + \|\tilde{P}^m - P^m\| < 1, \quad (\text{A.2})$$

gives

$$\sup_{k \in \mathbb{N}} \|z_{km}\| \leq \frac{\|\tilde{P}^m - P^m\|}{1 - \tau_m - \|\tilde{P}^m - P^m\|},$$

which is clearly less sharp than our bound (3.9). The bound [9, Equation (17)] differs from our inequality (3.10) in a similar way (note that Kartashov's bound [9, Equation (17)] holds for $n \geq m$ in our notation; this follows directly from the proof of [9, Equation (17)], which is based on an analogue of (3.12)). Kartashov's bound [9, Equation (14)] for the stationary distribution holds only if (A.2) is satisfied, and gives

$$\|\tilde{\pi} - \pi\| \leq \frac{\|\tilde{P}^m - P^m\|}{1 - \tau_m - \|\tilde{P}^m - P^m\|},$$

which is less sharp than our bound (3.11). Corollary 2 of [1], for

$$\tau_m + \max_{0 < i \leq m} \|\tilde{P}^i - P^i\| < 1,$$

gives the bound

$$\|\tilde{\pi} - \pi\| \leq (1 - \tau_m)^{-1} \max_{0 < i \leq m} \|\tilde{P}^i - P^i\|,$$

which is also less sharp than (3.11).

Acknowledgement

The author is grateful to the referee for valuable comments on earlier versions of this paper.

References

- [1] ANISIMOV, V. V. (1988). Estimates for the deviations of the transition characteristics of nonhomogeneous Markov processes. *Ukrainian Math. J.* **40**, 588–592.
- [2] CHO, G. E. AND MEYER, C. D. (2001). Comparison of perturbation bounds for the stationary distribution of a Markov chain. *Linear Algebra Appl.* **335**, 137–150.
- [3] DIACONIS, P. AND STROOK, D. (1991). Geometric bounds for eigenvalues of Markov chains. *Ann. Appl. Prob.* **1**, 36–61.
- [4] DOBRUSHIN, R. (1956). Central limit theorem for non-stationary Markov chains. I. *Theory Prob. Appl.* **1**, 63–80.
- [5] DOBRUSHIN, R. (1956). Central limit theorem for non-stationary Markov chains. II. *Theory Prob. Appl.* **1**, 329–383.
- [6] 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.
- [7] HORN, R. A. AND JOHNSON, C. R. (1985). *Matrix Analysis*. Cambridge University Press.
- [8] INGRASSIA, S. (1994). On the rate of convergence of the Metropolis algorithm and Gibbs sampler by geometric bounds. *Ann. Appl. Prob.* **4**, 347–389.
- [9] 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.

- [10] 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.
- [11] KARTASHOV, N. V. (1996). *Strong Stable Markov Chains*. VSP, Utrecht.
- [12] KIRKLAND, S. (2002). On a question concerning condition numbers for Markov chains. *SIAM J. Matrix Anal. Appl.* **23**, 1109–1119.
- [13] KIRKLAND, S. (2003). Conditioning properties of the stationary distribution for a Markov chain. *Electron. J. Linear Algebra* **10**, 1–15.
- [14] KIRKLAND, S. (2004). Digraph-based conditioning for Markov chains. *Linear Algebra Appl.* **385**, 81–93.
- [15] KIRKLAND, S. J. AND NEUMANN, M. (2000). Regular Markov chains for which the transition matrix has large exponent. *Linear Algebra Appl.* **316**, 45–65.
- [16] MENGENSEN, K. L. AND TWEEDIE, R. L. (1996). Rates of convergence of the Hastings and Metropolis algorithms. *Ann. Statist.* **24**, 101–121.
- [17] MEYER, C. D. (2000). *Matrix Analysis and Applied Linear Algebra*. SIAM, Philadelphia, PA.
- [18] MEYN, S. P. AND TWEEDIE, R. L. (1994). Computable bounds for convergence rates of Markov chains. *Ann. Appl. Prob.* **4**, 981–1011.
- [19] MEYN, S. P. AND TWEEDIE, R. L. (1996). *Markov Chains and Stochastic Stability*, 2nd edn. Springer, London.
- [20] MITROPHANOV, A. YU. (2003). Stability and exponential convergence of continuous-time Markov chains. *J. Appl. Prob.* **40**, 970–979.
- [21] NEUMANN, M. AND XU, J. (2004). Improved bounds for a condition number of Markov chains. *Linear Algebra Appl.* **386**, 225–241.
- [22] PULFORD, G. W. *et al.* (1995). Evaluation and estimation of various Markov models with applications to membrane channel kinetics. *Biom. J.* **37**, 39–63.
- [23] ROBERTS, G. O. AND POLSON, N. G. (1994). On the geometric convergence of the Gibbs sampler. *J. R. Statist. Soc. B* **56**, 377–384.
- [24] ROBERTS, G. O. AND ROSENTHAL, J. S. (1998). On convergence rates of Gibbs samplers for uniform distributions. *Ann. Appl. Prob.* **8**, 1291–1302.
- [25] ROBERTS, G. O., ROSENTHAL, J. S. AND SCHWARTZ, P. O. (1998). Convergence properties of perturbed Markov chains. *J. Appl. Prob.* **35**, 1–11.
- [26] SENETA, E. (1984). Explicit forms for ergodicity coefficients and spectrum localization. *Linear Algebra Appl.* **60**, 187–197.
- [27] SHMULEVICH, I., DOUGHERTY, E. R. AND ZHANG, W. (2002). Gene perturbation and intervention in probabilistic Boolean networks. *Bioinf.* **18**, 1319–1331.
- [28] SINCLAIR, A. (1992). Improved bounds for mixing rates of Markov chains and multicommodity flow. *Comb. Prob. Comput.* **1**, 351–370.
- [29] SOLAN, E. AND VIEILLE, N. (2003). Perturbed Markov chains. *J. Appl. Prob.* **40**, 107–122.
- [30] TIERNEY, L. (1994). Markov chains for exploring posterior distributions. *Ann. Statist.* **22**, 1701–1728.