Journal of Computational Mathematics, Vol.18, No.3, 2000, 305–312.

BLOCKWISE PERTURBATION THEORY FOR 2×2 BLOCK MARKOV CHAINS[*1)]

Jun-gong Xue Wei-guo Gao
(*Department of Mathematics, Fudan University, Shanghai, 200433, China*)

Abstract

Let P be a transition matrix of a Markov chain and be of the form

$$P = \begin{pmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{pmatrix}.$$

The stationary distribution π^T is partitioned conformally in the form (π_1^T, π_2^T). This paper establish the relative error bound in π_i^T $(i = 1, 2)$ when each block P_{ij} get a small relative perturbation.

Key words: Blockwise perturbation, Markov chains, stationary distribution, error bound

1. Introduction

The sensitivity of the stationary distribution to general perturbations in a transition matrix have been addressed by many authors [1], [2], [4], [6]. Let P amd $\tilde{P} = P + F$ be irreducible transition matrices with respective stationary distributions π^T and $\tilde{\pi}^T$ satisfying

$$P\mathbf{1} = \tilde{P}\mathbf{1} = \mathbf{1},\ \pi^T P = \pi^T,\ \tilde{\pi}^T P = \tilde{\pi}^T,\ \pi^T \cdot \mathbf{1} = \tilde{\pi}^T \cdot \mathbf{1} = 1.$$

Here, we denote by $\mathbf{1}$ (a bold one) the vector of all ones. In later discussion, we give its size with a subscript (e.g. $\mathbf{1})_n$ for the vector with n entries) explicitly when necessary. It is well known that

$$\|\pi - \tilde{\pi}\| \leq \|F\| \cdot \|A^{\#}\|, \tag{1}$$

where $A^{\#}$ is the group inverse of $A = I - P$, and $\| * \|$ denotes the infinity norm.

For some Markov chains, such as nearly uncoupled Markov chains, $\|A^{\#}\|$ is very large, which means that small perturbations in P can cause severe perturbations in π. However, the stationary distribution π can be insensitive to some special perturbation F. See [5], [9].

* Received February 20, 1997.
1) This author is supported in part by National Science Foundation.

The purpose of this paper is to analyze the effects of small blockwise relative perturbation to a 2×2 block transition matrix. More precisely, let P has the form

$$P = \begin{pmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{pmatrix}.$$

where P_{11} and P_{22} are square matrices of order n_1 and n_2, and let F, π^T and $\tilde{\pi}^T$ be partitioned conformally as

$$F = \begin{pmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \end{pmatrix} \quad \pi^T = (\pi_1^T, \pi_2^T) \quad \text{and} \quad \tilde{\pi}^T = (\tilde{\pi}_1^T, \tilde{\pi}_2^T).$$

Under the condition that P_{ii} $(i = 1, 2)$ is irreducible and

$$\|F_{ij}\| \le \eta \cdot \|P_{ij}\|, \quad i, j = 1, 2,$$

we are to bound $\|\pi_i^T - \tilde{\pi}_i^T\| / \|\pi_i^T\|$. Under certain condition, we will show that this relative error can be small even when $\|A^{\#}\|$ is large, or when $\|\pi_1\|$ is far more large (or less) than $\|\pi_2\|$. In [8], G. W. Stewart provided the relative error bound for π_2 when P is the transition matrix of a nearly transient Markov chain, i.e., $\|P_{12}\|$ is very small and $\|P_{21}\|$ is of magnitude one. In his analysis, he assumed that $F_{12} = 0$ and $\|\pi_1 - \tilde{\pi}_1\| / \|\pi_1\|$ is small. In this paper, these restrictions are deleted. The only assumption is that P_{11} and P_{22} are irreducible.

2. Some Basic Lemmas

In this section, we present some basic lemmas for Markov chains. These lemmas are important tools in deriving the main result of this paper.

Lemma 1. *Let x and y are n-vectors satisfy $y^T x = 1$. Then there are matrices J and K such that*

$$(x, J)^{-1} = \begin{pmatrix} y^T \\ K^T \end{pmatrix}.$$

Moreover, $\|J\|_2 = 1$ and $\|K\|_2 = \|x\|_2 \cdot \|y\|_2$.

Proof. See [7].

From the relation between ∞-norm and 2-norm,

$$\frac{1}{\sqrt{n}} \|B\|_\infty \le \|B\|_2 \le \sqrt{m} \|B\|_\infty, \quad B \in \mathbb{C}^{m \times n},$$

we have

$$\|J\|_\infty \le \sqrt{n-1} \|J\|_2 = \sqrt{n-1} \tag{2}$$

and

$$\|K\|_\infty \le \sqrt{n-1} \|K\|_2 = \sqrt{n-1} \cdot \|x\|_2 \cdot \|y\|_\infty. \tag{3}$$

The following two lemmas can be found in [8]

Lemma 2. *Let P be a transition matrix of the form*

$$P = \begin{pmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{pmatrix}$$

where $P_{11},\ P_{22}$ are square matrices. The stationary distribution π^T be partitioned conformal of the form (π_1^T, π_2^T). Then

$$S_{11} = P_{11} + P_{12}(I - P_{22})^{-1}P_{21} \quad \text{and} \quad S_{22} = P_{22} + P_{21}(I - P_{11})^{-1}P_{12} \tag{4}$$

are transition matrices with stationary distributions $\omega_1^T \frac{\pi_1^T}{\|\pi_1^T\|}$ and $\omega_2^T \frac{\pi_2^T}{\|\pi_2^T\|}$ respectively.

Lemma 3. *Let $P,\ \pi^T,\ \pi_1^T,\ \pi_2^T,\ \omega_1^T,\ \omega_2^T$ be as in Lemma 2. Then the matrix*

$$C = \begin{pmatrix} \omega_1^T P_{11}\mathbf{1} & \omega_1^T P_{12}\mathbf{1} \\ \omega_2^T P_{21}\mathbf{1} & \omega_2^T P_{22}\mathbf{1} \end{pmatrix}$$

is a transition matrix with stationary distribution $(\|\pi_1^T\|, \|\pi_2^T\|)$.

Let

$$\tilde{P} = \begin{pmatrix} P_{11} + F_{11} & P_{12} + F_{12} \\ P_{21} + F_{21} & P_{22} + F_{22} \end{pmatrix}$$

be a perturbed transition matrix with

$$\|F_{ij}\| \le \eta \cdot \|P_{ij}\|, \quad i, j = 1, 2.$$

In our analysis, we need to bound $\|F_{11}(I - P_{11})^{-1}\|$ and $\|F_{22}(I - P_{22})^{-1}\|$. When P is a transition matrix of a nearly coupled or nearly transient Markov chain, $\|P_{12}\|$ or $\|P_{21}\|$ is very small. A technical difficult presents itself immediately. $P_{11},\ P_{22}$ is near a stochastic matrix and has an eigenvalue near one. Hence $I - A_{11}$ is very nearly singular, and this near singularity prevents us from applying standard perturbation theory directly. We will circumvent the problem by transforming the matrix $I - P_{11}$ or $I - P_{22}$ into a form in which the offending eigenvalue is isolated.

Let γ_1 be the Perron eigenven of P_{11} and let v_1^T be the corresponding positive left eigenvector with $\|v^T\| = 1$. From Lemma 1, we can construct V_1, U_1 such that

$$\begin{pmatrix} v_1^T \\ V_1^T \end{pmatrix}^{-1} = (\mathbf{1}\ U_1).$$

and

$$\begin{pmatrix} v_1^T \\ V_1^T \end{pmatrix} P_{11}(\mathbf{1}\ U_1) = \begin{pmatrix} \gamma_1 & \\ V_1^T P_{11}\mathbf{1} & B_1 \end{pmatrix} \tag{5}$$

where

$$B_1 = V_1^T P_{11} U, \quad \|V_1\| \le \sqrt{n_1 - 1}, \quad \|U_1\| \le n_1 - 1.$$

The eigenvalues of B_1 are the eigenvalues of P_{11} other than γ_1. Noticing $V_1^T\mathbf{1} = 0$ and $P_{11}\mathbf{1} + P_{12}\mathbf{1} = \mathbf{1}$, we have

$$\begin{pmatrix} v_1^T \\ V_1^T \end{pmatrix} (I - P_{11})(\mathbf{1}\ U_1) = \begin{pmatrix} 1 - gamma_1 & \\ V_1^T P_{12}\mathbf{1} & I - B_1 \end{pmatrix},$$

which yields

$$(I - P_{11})^{-1} = (\mathbf{1}\ U_1) \begin{pmatrix} (1-\gamma_1)^{-1} & \\ & (I - B_1)^{-1} \end{pmatrix} \begin{pmatrix} 1 & \\ \frac{1}{1-\gamma_1} V_1^T P_{12}\mathbf{1} & \end{pmatrix} \begin{pmatrix} v_1^T \\ V_1^T \end{pmatrix}. \quad (6)$$

In the remader of this paper we introduce symbols c and $c_i,\ i = 1, 2, \ldots$ to represent which are bounded above polynomials of n_1 and n_2. And these constants grows slowly as n_1 and n_2 grows.

Lemma 4. *Let P, B_1 be as above. Define*

$$\theta_1 = \frac{\|(I - P_{11})^{-1} P_{11}\|}{\|(I - P_{11})^{-1}\| \cdot \|P_{12}\|} = \frac{1}{\|(I - P_{11})^{-1}\| \cdot \|P_{12}\|}$$

Then

$$\|F_{11}(I - P_{11})^{-1}\| \le c_1 \cdot \|(I - B_1)^{-1}\| \cdot \frac{1}{\theta_1} \cdot \eta$$

Proof. From (6) and noting $F_{11}\mathbf{1} = -F_{12}\mathbf{1}$, we have

$$F_{11}(I - P_{11})^{-1} = (\frac{1}{1-\gamma_1} F_{12}\mathbf{1}\ F_{11}U_1)(I - B_1)^{-1} \begin{pmatrix} 1 & \\ \frac{1}{1-\gamma_1} V_1^T P_{12}\mathbf{1} & \end{pmatrix} \begin{pmatrix} v_1^T \\ V_1^T \end{pmatrix}. \quad (7)$$

Since

$$|\frac{1}{1-\gamma_1}| \le \|(I - P_{11})^{-1}\| \le \frac{1}{\theta \cdot \|P_{12}\|},$$

then

$$\|\frac{1}{1-\gamma_1} F_{12}\mathbf{1}\| \le \frac{\eta}{\theta_1}$$

and

$$\|\frac{1}{1-\gamma_1} (I - B_1)^{-1} V_1^T P_{12}\mathbf{1}\| \le c_2 \cdot \frac{1}{\theta_1} \cdot \|(I - B_1)^{-1}\|.$$

On taking norm of each term in (7), we have

$$\|F_{11}(I - P_{11})^{-1}\| \le c_1 \cdot \frac{1}{\theta_1} \|(I - B_1)^{-1}\| \cdot \eta.$$

Remark: If the second large eignevalue of P_{11} is not near to 1, we can expect $\|(I - B_1)^{-1}\|$ is not large. The main term in $\|(I - P_{11})^{-1}$ is $\frac{1}{1-\gamma_1}$. Since $1 - \gamma_1 = v_1^T P_{12}\mathbf{1}$ thus θ_1 can be estimated by $|v_1^T P_{12}\mathbf{1}|/\|P_{12}\|$.

Analoguously, there exist v_2, V_2, U_2 such that

$$\begin{pmatrix} v_2^T \\ V_2^T \end{pmatrix}^{-1} = (\mathbf{1}\ U_2), \quad \begin{pmatrix} v_2^T \\ V_2^T \end{pmatrix} P_{22}(\mathbf{1}\ U_2) = \begin{pmatrix} \gamma_2 & \\ V_2^T P_{22}\mathbf{1} & B_2 \end{pmatrix}$$

where

$$B_2 = V_2^T P_{22} U, \quad \|V_2\| \le \sqrt{n_2 - 1}, \quad \|U_2\| \le n_2 - 1,$$

where $B_2 = V_2^T P_{22} U_2$. Defining

$$\theta_2 = \frac{1}{\|(I - P_{22})^{-1}\| \cdot \|P_{21}\|},$$

we can also have

$$\|F_{22}(I - P_{11})^{-1}\| \le c_4 \cdot \frac{1}{\theta_2} \|(I - B_2)^{-1}\|\eta.$$

3. Main Result

Theorem 5. *Let*

$$P = \begin{pmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{pmatrix} \quad \textit{and} \quad \tilde{P} = \begin{pmatrix} P_{11} + F_{11} & P_{12} + F_{12} \\ P_{21} + F_{21} & P_{22} + F_{22} \end{pmatrix}$$

be transition matrices with stationary distributions $\pi^T = (\pi_1^T, \pi_2^T)$ *and* $\tilde{\pi}^T = (\tilde{\pi}_1^T, \tilde{\pi}_2^T)$ *respectively. Denote*

$$\theta = \min\{\theta_1, \theta_2\} \quad \xi = \min\{\frac{\pi_1^T P_{12}\mathbf{1}}{\|\pi_1^T\| \cdot \|P_{12}\|}, \frac{\pi_2^T P_{21}\mathbf{1}}{\|\pi_2^T\| \cdot \|P_{21}\|}\}$$

and

$$\lambda = \max\{(I - B_1)^{-1}, (I - B_2)^{-1}\} \quad s = \max\{\|(I - S_{11})^{\#}\|, \|(I - S_{22})^{\#}\|\}$$

where S_{11}, S_{22} *are as in* (4)*. If* P_{11} *and* P_{22} *are irreducible and*

$$\|F_{ij}\| \le \eta \cdot \|P_{ij}\|,$$

then

$$\frac{\|\pi_i^T - \tilde{\pi}_i^T\|}{\|\pi_i^T\|} \le c \cdot \frac{\lambda \cdot s}{\theta^2 \cdot \xi}\eta + O(\eta^2). \tag{8}$$

Proof. Let $\tilde{S}_{22} = \tilde{P}_{22} + \tilde{P}_{21}(I - \tilde{P}_{11})^{-1}\tilde{P}_{12}$. Then

$$\tilde{S}_{22} - S_{22} = F_{22} + F_{21}(I - \tilde{P}_{11})^{-1}\tilde{P}_{12} + P_{21}(I - \tilde{P}_{11})^{-1}[F_{11}(I - P_{11})^{-1})P_{12} + F_{12}].$$

It is easily verified

$$\begin{aligned}\|F_{22}+F_{21}(I-\tilde{P}_{11})^{-1}\tilde{P}_{12}\| &\leq \|F_{22}\|+\|F_{21}\|\cdot\|(I-\tilde{P}_{11})^{-1}\tilde{P}_{12}\| \\ &= \|F_{22}\|+\|F_{21}\| \leq \eta(\|P_{22}\|+\|P_{21}\|) \leq 2\eta.\end{aligned}$$

From Lemma 4 we have

$$\|F_{11}(I-P_{11})^{-1}P_{12}+F_{12}\| \leq c_4\frac{\lambda}{\theta}\cdot\|P_{12}\|\cdot\eta,$$

and

$$\begin{aligned}\|(I-\tilde{P}_{11})^{-1}\| &\leq \|(I-P_{11})^{-1}\|\cdot\|(I-F_{11}(I-P_{11})^{-1})^{-1}\| \\ &\leq \; = 1+c_5\frac{\lambda}{\theta}\eta+O(\eta^2),\end{aligned}$$

From which we have

$$\|(\tilde{S}_{22}-S_{22})\| \leq c_6\cdot\frac{\lambda}{\theta^2}\eta+O(\eta^2).$$

Let

$$\omega_1^T=\frac{\pi_1^T}{\|\pi_1^T\|},\quad \omega_2^T=\frac{\pi_2^T}{\|\pi_2^T\|},\quad \tilde{\omega}_1^T=\frac{\tilde{\pi}_1^T}{\|\tilde{\pi}_1^T\|},\quad \tilde{\omega}_2^T=\frac{\tilde{\pi}_2^T}{\|\tilde{\pi}_2^T\|},$$

From (1) we can get

$$\|\omega_2^T-\tilde{\omega}_2^T\| \leq c_6\cdot\frac{\lambda s}{\theta^2}\eta+O(\eta^2).$$

Similarly,

$$\|\omega_1^T-\tilde{\omega}_1\| \leq c_7\cdot\frac{\lambda s}{\theta^2}\eta+O(\eta^2).$$

Let

$$C=\begin{pmatrix}\omega_1^TP_{11}\mathbf{1} & \omega_1^TP_{12}\mathbf{1}\\ \omega_2^TP_{21}\mathbf{1} & \omega_2^TP_{22}\mathbf{1}\end{pmatrix}=(c_{ij}) \quad \text{and}\tilde{C}=\begin{pmatrix}\tilde{\omega}_1^T\tilde{P}_{11}\mathbf{1} & \tilde{\omega}_1^T\tilde{P}_{12}\mathbf{1}\\ \tilde{\omega}_2^T\tilde{P}_{21}\mathbf{1} & \tilde{\omega}_2^T\tilde{P}_{22}\mathbf{1}\end{pmatrix}=(\tilde{c}_{ij}).$$

From Lemma 3 $(\|\pi_1^T\|,\|\pi_2^T\|)$ and $(\|\tilde{\pi}_1^T\|,\|\tilde{\pi}_2^T\|)$ are their stationary distributions. Obviously,

$$\frac{|c_{ij}-\tilde{c}_{ij}}{c_{ij}} \leq c_8\cdot\frac{\lambda s}{\theta^2\xi}\eta+O(\eta^2)$$

The entrywise perturbation theorem in gives

$$\frac{|\|\pi_i^T\|-\|\tilde{\pi}_i^T\||}{\|\pi_i^T\|} \leq c_9\frac{\lambda s}{\theta^2\xi}\cdot\eta+O(\eta^2),\quad i=1,2.$$

Thus

$$\begin{aligned}\frac{\|\pi_i^T-\tilde{\pi}_i^T\|}{\|\pi_i^T\|} &= \frac{\|(\|\pi_i^T\|-\|\tilde{\pi}_i)^T\|)\pi_i^T+\|\tilde{\pi}_i\|(\tilde{\pi}_i)^T-\pi_i^T}{\|\pi_i^T\|} \\ &le \quad c\cdot\frac{\lambda\cdot s}{\theta^2\cdot\xi}\eta+O(\eta^2)\end{aligned}$$

Remark: For nearly uncoupled Markov chain, though $A^{\#}$ can be very large. $\frac{\lambda s}{\theta \cdot \xi}$ can be small. For example

$$P = \begin{pmatrix} 0.5 & 0.5-2\epsilon & \epsilon & \epsilon \\ 0.5-2\epsilon & 0.5 & \epsilon & \epsilon \\ \epsilon & \epsilon & 0.5 & 0.5-2\epsilon \\ \epsilon & \epsilon & 0.5-2\epsilon & 0.5 \end{pmatrix} = \begin{pmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{pmatrix},$$

where P_{11} and P_{22} are 2×2 matrices. It is easy to see $\theta = \xi = 1,\ \lambda = \frac{1}{1-2\epsilon}$ and $s \approx 2$, while

$$\|A^{\#}\| = O(\frac{1}{\epsilon}).$$

When ϵ is small, the stationary distribution is sensitive to general perturbations, however, it is in sensitive to small relative blockwise perturbations. When the Markov chain has nearly transien states, for example

$$P = \begin{pmatrix} \frac{1}{2} & \frac{1}{2}-2\epsilon & \epsilon & \epsilon \\ \frac{1}{2}-2\epsilon & \frac{1}{2} & \epsilon & \epsilon \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{pmatrix} = \begin{pmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{pmatrix},$$

with the stationary distribution

$$\pi = \frac{1}{2+8\epsilon}(1, 1, 4\epsilon, 4\epsilon) = (\pi_1^T, \pi_2^T)$$

Since

$$\|\pi_1^T\| \gg \|\pi_2^T\|,$$

π_2 may be sensitive in a relative sense to general perturbations though π is insensitive. However, this can not happen for small blockwise perturbations because of

$$\theta = \xi = 1 \quad \lambda = \frac{2}{1-2\epsilon} \quad \text{and} \quad s \leq 2.$$

Acknowledgement. The authors are grateful to Prof. Jiang Erxiong for his careful reading of this paper and illuminating suggestions.

References

[1] R.E.Funderlic and C.D.Meyer, *Senstivity of the stationary distribution vector for an ergodic Markov chain,* Linear Algebra Appl.,76(1986), 1-17

[2] G.H.Golub and C.D.Meyer, *Using the QR factorization and group inversion to compute, differetiate, and estimate the sensitivity of stationary probabilities for Markov chains,* SIAM.J.Alg.Disc.Meth.,7(1985), 273-281

[3] McAllister, G.W.Stewart and W.J.Stewart, *On a Rayleigh-Ritz refinement technique for nearly stochastic matrices,* Linear Algebra Appl.,60(1984), 1-25.

[4] C.D.Meyer, *The condition of a finite Markov chain and perturbation bounds for the limitting probabilities,* SIAM.J.Alg.Disc.Meth.,1(1980), 273-283.

[5] C.A.O'cinneide, *Entrywise perturbation theory and error analysis for Markov chains,* Numer.Math.,65(1993), 190-120.

[6] P.J.Schweitzer, *Perturbation theory and finite Markov chains,* J.Appl.Prob., 5(1968), 135-141.

[7] G.W.Stewart, *On the perturbation of Markov chains with nearly transient states,* Numer. Math.,65(1993), 135-141

[8] W.J.Stewart, *On the effects of using the Grassmann-Taksar-Heyman method in iterative Aggregation-Disaggregation,* SIAM J.Sci.Comput.,1(1996), 287-303.

[9] G.Zhang, On the sensitivity of the solution of nearly uncoupled Markov chains, SIAM J.Matrix Anal.Appl.,4(1993), 1112-1123.