Technische Universität Chemnitz

Sonderforschungsbereich 393

Numerische Simulation auf massiv parallelen Rechnern

Jungong Xue

On the blockwise perturbation of nearly uncoupled Markov chains

Preprint SFB393/00-18

Preprint-Reihe des Chemnitzer SFB 393

SFB393/00-18 **April 2000**

Contents

Author:

Jungong Xue
Department of Mathematics
Fudan University
Shanghai, 200433, P.R.China

Present address:
Fakultät für Mathematik
Technische Universität Chemnitz
09107 Chemnitz, Germany

On the blockwise perturbation of nearly uncoupled Markov chains

Jungong Xue [1]

Abstract

Let P be the transition matrix of a nearly uncoupled Markov chain. The states can be grouped into aggregates such that P has the block form $P = (P_{ij})_{i,j=1}^{k}$ where P_{ii} is square and P_{ij} is small for $i \neq j$. Let π^T be the stationary distribution partitioned conformally as $\pi^T = (\pi_1^T, \cdots, \pi_k^T)$. In this paper we present the relative error bounds for π_i^T when each block P_{ij} gets a small relative perturbation. These bounds show that although the stationary distribution of a nearly uncoupled Markov chain is very sensitive to general perturbations, small blockwise relative perturbations of P only cause small relative errors in each aggregate distribution π_i^T.

Keywords: nearly uncoupled Markov chains, blockwise perturbation, nonnegative matrices.
AMS Subject Classifications: 65F35, 15A42,

1 Introduction

A nearly uncoupled Markov chain is a discrete chain whose states can be ordered such that the transition matrix assumes the form

$$P = \begin{bmatrix} P_{11} & P_{12} & \cdots & P_{1k} \\ P_{12} & P_{22} & \cdots & P_{2k} \\ \vdots & \vdots & & \vdots \\ P_{k1} & P_{k2} & \cdots & P_{kk} \end{bmatrix}, \tag{1}$$

where all the off-diagonal blocks P_{ij} are small. Here each P_{ij} is an $n_i \times n_j$ matrix. We set

$$\epsilon = \max_{1 \leq i \leq k} \sum_{j \neq i} \|P_{ij}\|, \tag{2}$$

where $\| * \|$ is the ∞-norm. Chains of this kind are used to model systems whose states can be grouped into aggregates that are loosely connected to one another. They have been

[1] Research supported by Alexander von Humboldt Foundation of Germany and Natural Sciences Foundation of China.

addressed by many authors, see e.g. [1, 2, 3, 4, 8, 9, 10, 12, 13]. One reason why nearly uncoupled Markov chains receive so much attention is that their stationary distributions are very sensitive to the perturbations in the transition matrices. Let π^T and $\widehat{\pi}^T$ be stationary distributions of transition matrices P and $\widehat{P} = P + F$, respectively; that is, π^T and $\widehat{\pi}^T$ are row vectors satisfying

$$\pi^T P = \pi^T, \qquad \widehat{\pi}^T \widehat{P} = \widehat{\pi}^T, \qquad \pi^T \mathbf{1} = \widehat{\pi}^T \mathbf{1} = 1,$$

where $\mathbf{1}$ is the vector of all ones. According to the perturbation theory for Markov chains, see e.g. [5, 7],

$$\|\pi^T - \widetilde{\pi}^T\| \leq \|A^{\#}\| \|F\|, \tag{3}$$

where $A^{\#}$ is the group inverse of the matrix $A = I - P$. Equality in (3) can be attained for some F. It is shown in [14] that

$$\|A^{\#}\| \geq O\left(\frac{1}{\epsilon}\right).$$

This means that small perturbations in the transition matrices of nearly uncoupled Markov chains can result in large errors in their stationary distributions. The smaller ϵ is, the more sensitive the stationary distributions are to the perturbations. However, if the perturbation F has some special structure, the error bound (3) is often an overestimate. One typical example is that if F is a small entrywise relative perturbation to P, then the entrywise relative error in π^T it causes must be small and independent of any condition number, see [11, 16, 17]. In [19], Zhang studied a class of perturbations for nearly uncoupled Markov chains to which their stationary distributions are insensitive. To state his result, we partition F, π^T and $\widehat{\pi}^T$ conformally with P as

$$F = \begin{bmatrix} F_{11} & F_{12} & \cdots & F_{1k} \\ F_{12} & F_{22} & \cdots & F_{2k} \\ \vdots & \vdots & & \vdots \\ F_{k1} & F_{k2} & \cdots & F_{kk} \end{bmatrix}, \qquad \pi^T = [\pi_1^T, \cdots, \pi_k^T], \qquad \widehat{\pi}^T = [\widehat{\pi}_1^T, \cdots, \widehat{\pi}_k^T].$$

If the blocks of the perturbation F satisfy

$$\|F_{ii}\| \leq \eta \qquad \text{and} \qquad \|F_{ij}\| \leq \epsilon\eta \quad i \neq j,$$

then under some regularity conditions, it is proved in [19] that

$$\frac{\|\pi^T - \widehat{\pi}^T\|}{\|\pi^T\|} \leq c\eta. \tag{4}$$

The quantity c in (4) is bounded from above as ϵ tends to zero. However, the upper bound for c is not discussed in [19]. Besides,the error bound (4) is for the whole stationary distribution π^T and gives no information about the relative error in each aggregate distribution π_i^T. In fact, even under the regularity conditions in [19], some aggregate distributions can be very small compared to others. A small relative error in π^T does not guarantee small relative error in π_i^T when $\|\pi_i^T\|$ is tiny.

The goal of this paper is to analyze the sensitivity of each aggregate distribution π_i^T to small relative blockwise perturbations in the transition matrix P. More precisely, we aim to bound the relative error in the aggregate distributions

$$\frac{\|\pi_i^T - \widetilde{\pi_i}^T\|}{\|\pi_i^T\|}, \qquad i = 1, \ldots, k \tag{5}$$

under the assumption that

$$\|F_{ij}\| \leq \eta \|P_{ij}\|, \qquad i, j = 1, \ldots, k. \tag{6}$$

The error (5) will be bounded in terms of η and some quantities, which we discuss in detail. The error bounds demonstrate that however small $\|\pi_i^T\|$ is, π_i^T is very insensitive to the small relative blockwise perturbation F. This result reveals the fact that if we can obtain information with high relative accuracy both in the aggregates and between the aggregates, we can trust the solution from the measured system. Our result extends entrywise perturbation theory for general Markov chains to the blockwise case for nearly uncoupled Markov chains. Also it strengthens the result in [19] for more restrictive perturbations. Blockwise perturbation theory has been studied in [15] and [18] for a class of Markov chains with transition matrices assuming block p-cyclic form. In this paper, we generalize those results to general nearly uncoupled Markov chains.

Throughout this paper we always assume that P is a primitive matrix of order n and for each diagonal block P_{ii}, the second largest eigenvalue is bounded away from 1.

This paper is organized as follows. In Section 2 we present some notation and lemmas, especially we introduce a special decomposition of nonnegative matrices. In Section 3 we use this decomposition to define some quantities in terms of which we bound the error (5). There we also analyze these quantities through the spectral analysis of P_{ii}. In Section 4 we investigate the structure of each block of the inverse of the matrix $I - P_i$, where P_i is the principal submatrix of P with the ith row and column of blocks removed. This structure will be exploited in Section 5 to bound the error (5).

2 Notation and lemmas

Throughout this article $\| * \|$ denotes the ∞-norm for matrices and column vectors and the 1-norm for row vectors. Let B be the matrix with entries b_{ij} and C be the matrix with entries c_{ij}. We denote by $|B|$ the matrix with entries $|b_{ij}|$ and let $B \leq C$ mean $b_{ij} \leq c_{ij}$ for all i and j. For vectors, $|y|$ and $y \leq x$ are defined in an analogous way. We denote by $\mathbf{1}$ the column vector of all ones regardless of its dimension. For transition matrices P as in (1), we denote by P_{i*} the ith block row of P with P_{ii} deleted, P_{*i} the ith block column of P with P_{ii} deleted, and P_i the principal matrix of P obtained by deleting the ith block row and block column. We let S_{ii} denote the stochastic complement of P_{ii} in P, that is,

$$S_{ii} = P_{ii} + P_{i*}(I - P_i)^{-1} P_{*i}. \tag{7}$$

It was shown in [10] that S_{ii} is stochastic and $\pi_i^T / \|\pi_i^T\|$ is its stationary distribution.
Each nonnegative matrix A can be decomposed in the form

$$A = \mathbf{1} r^T + R, \tag{8}$$

where r^T is a nonnegative row vector and R is a nonnegative matrix with at least one zero in each column. In other words, the i-th entry of r is the minimun of the entries in the i-th column of A. The decomposition (8) is called the *column parallel decomposition* for nonnegative matrices. Based on (8), we define the *column parallel rate* of a nonnegative matrix A as

$$s(A) = \begin{cases} \frac{\|R\|}{r^T \mathbf{1}} & r^T \mathbf{1} \neq 0 \\ 0 & r^T \mathbf{1} = 0. \end{cases}$$

Now we present two basic properties of the *column parallel rate.*

Lemma 2.1 *Let A_1 and A_2 be nonnegative matrices, then*

$$s(A_1 + A_2) \leq \max\{s(A_1), s(A_2)\}.$$

Proof Let A_1 and A_2 have the *column parallel decompositions*

$$A_1 = \mathbf{1} r_1^T + R_1, \qquad A_2 = \mathbf{1} r_2^T + R_2,$$

respectively. Let u^T be a nonnegative row vector whose i-th entry is the smallest entry of the i-th column of $R_1 + R_2$. Then $A_1 + A_2$ has the *column parallel decomposition*

$$A_1 + A_2 = \mathbf{1}(r_1 + r_2 + u)^T + R_1 + R_2 - \mathbf{1} u^T,$$

from which it is straightforward to get that

$$s(A_1 + A_2) \leq \max(s(A_1), s(A_2)).$$

$\square$

Lemma 2.2 *Let A_1, A_2 and S be nonnegative matrices of orders $m_1 \times p_1$, $m_2 \times p_2$ and $p_1 \times m_2$, respectively. Let A_1 and A_2 have the column parallel decompositions*

$$A_1 = \mathbf{1}r_1^T + R_1, \qquad A_2 = \mathbf{1}r_2^T + R_2.$$

Set

$$\nu = \frac{r_1^T S \mathbf{1}}{r_1^T \mathbf{1} \|S\|},$$

then

$$s(A_1 S A_2) \leq \frac{s(A_1)(1 + s(A_2))}{\nu}.$$

Proof We have

$$A_1 S A_2 = (r_1^T S \mathbf{1})\mathbf{1}r_2^T + \mathbf{1}r_1^T S R_2 + R_1 S \mathbf{1} r_2^T + R_1 S R_2.$$

Let u^T be the nonnegative row vector whose i-th entry is the minimun of the entries in the i-th column of matrix $R_1 S \mathbf{1} r_2^T + R_1 S R_2$. Then $A_1 S A_2$ has the *column parallel decomposition*

$$A_1 S A_2 = \mathbf{1} r_3^T + R_3,$$

where

$$r_3^T = (r_1^T S \mathbf{1}) r_2^T + r_1^T S R_2 + u^T$$

and

$$R_3 = R_1 S \mathbf{1} r_2^T + R_1 S R_2 - \mathbf{1} u^T.$$

Using the nonnegativity of matrices and norm inequalities we get

$$r_3^T \mathbf{1} \leq (r_1^T S \mathbf{1}) r_2^T \mathbf{1} = \nu (r_1^T \mathbf{1})(r_2^T \mathbf{1}) \|S\| \tag{9}$$

and

$$\|R_3\| \leq \|R_1 S \mathbf{1} r_2^T\| + \|R_1 S R_2\| \leq \|S\| \|R_1\| (r_2^T \mathbf{1} + \|R_2\|). \tag{10}$$

Combining (9) and (10) completes the proof. $\square$

These two lemmas will be used in Section 4 to investigate the decomposition of each block of $(I - P_i)^{-1}$. In the next section, we will bound $s((I - P_{ii})^{-1})$ through the spectral analysis of P_{ii}. To do this, we need the following lemma.

Lemma 2.3 *Let A be an $m \times m$ nonnegative matrix of the form $A = \mathbf{1}v^T + Q$, where v^T is a nonnegative row vector and $\|Q\|$ is small compared to $\|v^T\|$. Note that we do not assume that Q is nonnegative. Let*

$$\delta = \frac{\|Q\|}{\|v^T\|}$$

and let A have the column parallel decomposition

$$A = \mathbf{1}r^T + R.$$

If $m\delta < 1$, then

$$s(A) \leq \frac{(m+1)\delta}{1 - m\delta} \qquad \text{and} \qquad \frac{\|r^T - v^T\|}{\|v^T\|} \leq m\delta.$$

Proof Let u^T be the row vector whose i-th entry is the minimun of the entries in the i-th column of Q. Obviously, $|u^T| \leq \|Q\|\mathbf{1}^T$ and thus

$$\|u^T\| \leq m\|Q\| = m\delta\|v^T\|.$$

We have the *column parallel decomposition* of A with

$$r^T = v^T + u^T \quad \text{and} \quad R = Q - \mathbf{1}u^T.$$

Thus

$$\frac{\|r^T - v^T\|}{\|v^T\|} = \frac{\|u^T\|}{\|v^T\|} \leq m\delta$$

and

$$s(A) = \frac{\|R\|}{\|r^T\|} \leq \frac{\|Q\| + \|u^T\|}{\|v^T - u^T\|} \leq \frac{(m+1)\delta}{1 - m\delta}.$$

□

3 Spectral analysis

In this section we will define some quantities in terms of which we bound the relative error (5). These quantities are somewhat complicated at first sight. However, we will give insight into them through spectral analysis of the diagonal blocks P_{ii}.

Let $(I-P_{ii})^{-1}$ have the *column parallel decomposition* $(I-P_{ii})^{-1} = \mathbf{1}r_i^T + R_i$. We define

$$\tau_i = s((I-P_{ii})^{-1}) = \begin{cases} \frac{\|R_i\|}{\|r_i^T\mathbf{1}\|} & r_i^T\mathbf{1} \neq 0 \\ 0 & r_i^T\mathbf{1} = 0 \end{cases} \tag{11}$$

and for $j \neq i$

$$\phi_{ij} = \begin{cases} \frac{r_i^T P_{ij}\mathbf{1}}{(r_i^T\mathbf{1})\|P_{ij}\|.} & \|P_{ij}\| \neq 0 \\ 1 & \|P_{ij}\| = 0. \end{cases} \tag{12}$$

In the following we derive bounds for τ_i and ϕ_{ij}.

Let γ_i be the Perron root of P_{ii} and let v_i^T be the corresponding left eigenvector normalized so that $v_i^T\mathbf{1} = 1$. Let the columns of U_i form an orthonormal basis for the space orthogonal to v_i and the columns of J_i form an orthonormal basis for the space orthogonal to $\mathbf{1}$. In other words,

$$U_i^T v_i = 0, \quad U_i^T U_i = I, \quad J_i^T\mathbf{1} = 0, \quad J_i^T J_i = I.$$

Let

$$V_i = J_i(J_i^T U_i)^{-T},$$

then it is proved in [9] that

$$\begin{bmatrix} v_i^T \\ V_i^T \end{bmatrix}^{-1} = [\mathbf{1}\ U_i]$$

and

$$\|U_i\|_2 = 1, \quad \|V_i\|_2 = \|(J_i^T U_i)^{-1}\|_2 \leq \sqrt{n_i},$$

where $\| * \|_2$ is Euclidean norm. The following theorem bounds τ_i and ϕ_{ij}.

Theorem 3.1 *Let P_{ii} of order n_i be the i-th diagonal block of P in (1). Let $B_i = V_i^T(I - P_{ii})U_i$, $\delta_i = \|U_i B_i^{-1} V_i^T\|$ and let ϵ be as in (2). For $i \neq j$, set*

$$q_{ij} = \begin{cases} \frac{v_i^T P_{ij}\mathbf{1}}{\|P_{ij}\|} & \|P_{ij}\| \neq 0 \\ 1 & \|P_{ij}\| = 0 \end{cases}.$$

If $2n_i\delta_i\epsilon < 1$, then τ_i in (11) is bounded as

$$\tau_i \leq \frac{2(n_i+1)\delta_i\epsilon}{1-2n_i\delta_i\epsilon}. \tag{13}$$

Moreover, if $2n_i\delta_i\epsilon \leq q_{ij}$, then ϕ_{ij} is bounded as

$$\phi_{ij} \geq \frac{q_{ij}-2n_i\delta_i\epsilon}{1+2n_i\delta_i\epsilon}. \tag{14}$$

Proof We have

$$\begin{bmatrix} v_i^T \\ V_i^T \end{bmatrix} (I - P_{ii})[\mathbf{1}\ U_i] = \begin{bmatrix} 1-\gamma_i & \\ V_i^T(I-P_{ii})\mathbf{1} & B_i \end{bmatrix}.$$

Then

$$(I - P_{ii})^{-1} = \frac{1}{1-\gamma_i}\mathbf{1}v_i^T + Q_i,$$

where

$$Q_i = \frac{1}{1-\gamma_i} U_i B_i^{-1} V_i^T C_i \quad \text{and} \quad C_i = -(I - P_{ii})\mathbf{1}v_i^T + (1-\gamma_i)I.$$

Since $(I - P_{ii})\mathbf{1} = \sum_{j\neq i} P_{ij}\mathbf{1} \leq \epsilon\mathbf{1}$, we have

$$1 - \gamma_i \leq \epsilon \qquad \text{and} \qquad \|C_i\| \leq 2\epsilon.$$

Therefore

$$(1-\gamma_i)\|Q_i\| \leq 2\delta_i\epsilon.$$

Applying Lemma 2.3 gives (13) and

$$|v_i^T - (1-\gamma_i)r_i^T|\mathbf{1} \leq \|v_i^T - (1-\gamma_i)r_i^T\| \leq 2n_i\delta\epsilon.$$

It follows that

$$\begin{aligned} \phi_{ij} &= \frac{v_i^T P_{ij}\mathbf{1} + ((1-\gamma_i)r_i^T - v_i^T)P_{ij}\mathbf{1}}{(1-\gamma_i)r_i^T\mathbf{1}\|P_{ij}\|} \\ &\geq \frac{q_{ij} - 2n_i\delta_i\epsilon}{1 + 2n_i\delta_i\epsilon}. \end{aligned}$$

□

Remark 3.1. The eigenvalues of B_i are those eigenvalues of $I - P_{ii}$ other than $1 - \gamma_i$. Throughout this article we always assume that the second largest eigenvalue of P_{ii} is bounded away from 1. Thus the eigenvalues of B_i are bounded away from 0. If B_i is diagonalizable, that is, there exists nonsingular matrix T such that $T^{-1}B_iT$ is a diagonal matrix, then $\|B_i^{-1}\| \leq \|T\|\|T^{-1}\|/|\lambda|$, where λ is the smallest eigenvalue (in modulus) of B_i. Even though $I - P_{ii}$ is nearly singular and $\|(I - P_{ii})^{-1}\|$ must be very large, we can expect that $\|B_i\|$ is of moderate size. Noting that $\|U_i\|_2 = 1$ and $\|V_i\|_2 \leq \sqrt{n_i}$, we can also expect that δ_i is of moderate size and so τ_i is very small. The quantities q_{ij} may be large if v_i is not nearly orthogonal to $P_{ij}\mathbf{1}$. In fact, let ρ_i be the ratio between the largest and smallest entries of v_i^T, then we have $q_{ij} \leq 1/(n_i \rho_i)$. Therefore ϕ_{ij} can be bounded away from zero as long as ρ_i is not very large.

For τ_i as in (11) and ϕ_{ij} as in (12), we define

$$\tau = \max_{1\leq i\leq k} \tau_i \qquad \text{and} \qquad \phi = \max_{1\leq i\leq k} (\max_{j\neq i} \phi_{ij}). \tag{15}$$

We still need two other quantities to bound the error (5). To do this, we define the set of stochastic matrices

$$\Phi_i = \{T \mid T \geq 0,\ T\mathbf{1} = \mathbf{1},\ \|T - P_{ii}\| \leq 2\eta + \epsilon\}. \tag{16}$$

and for each set Φ_i, we define

$$\sigma_i = \sup\{\|(I-T)^{\#}\| \mid T \in \Phi_i\}$$

and

$$\psi_{ij} = \inf\left\{ \frac{v^T P_{ij}\mathbf{1}}{\|P_{ij}\|} \;\Big|\; T \in \Phi_i,\ v^T = v^T T,\ v^T\mathbf{1} = 1 \right\}.$$

Here ϵ is as in (2) and η is as in (6). The quantities σ_i and ψ_{ij} can also be bounded through the spectral analysis of the diagonal blocks P_{ii}. Let v^T be the stationary distribution of $T \in \Phi_i$, i.e., $v^T T = v^T$ and $v^T\mathbf{1} = 1$. According to the perturbation theory for the Perron vector v_i^T of P_{ii}, see [6], if $2\eta + \epsilon$ is sufficiently small, then $\|v^T - v_i^T\| \leq s_i(2\eta+\epsilon)$. Here s_i is the condition number for v_i^T in infinity norm. It is shown in [6] that the separation of the Perron root γ_i and other eigenvalues of P_{ii} has a bearing upon s_i. Since the eigenvalues other than γ_i are bounded away from 1, this separation is not small. We can expect that s_i is of moderate size. If $s_i(2\eta+\epsilon) < q_{ij}$, it is straightforward to get that

$$\psi_{ij} \geq q_{ij} - s_i(2\eta + \epsilon).$$

The following theorem bounds $\|(I-T)^{\#}\|$ for $T \in \Phi_i$.

Theorem 3.2 *Let Φ_i be as in (16), let $T \in \Phi_i$ and $(I-T)^{\#}$ be the group inverse of $I-T$, and let, furthermore,*

$$g(\epsilon, \eta) = \|U_i\|\|V_i^T\|(1 + 2s_i + s_i(2\eta+\epsilon))(2\eta+\epsilon).$$

If $\|B_i^{-1}\| g(\epsilon,\eta) < 1$, then

$$\|(I-T)^{\#}\| \leq \frac{(1 + s_i(2\eta+\epsilon))\|U_i\|\|V_i^T\|\|B_i^{-1}\|}{1 - \|B_i^{-1}\| g(\epsilon,\eta)}$$

Proof Let v^T be the stationary distribution of T and let v_i^T be the left Perron vector of P_{ii} normalized so that $v_i^T \mathbf{1} = 1$. Set $u^T = v_i^T - v^T$. Choosing

$$F_i = \mathbf{1} u^T U_i$$

and noting that $\|u^T\| \le s_i(2\eta + \epsilon)$ and $v_i^T U_i = 0$, we have

$$v^T(U_i + F_i) = 0, \qquad \|F_i\| \le s_i \|U_i\|(2\eta + \epsilon).$$

It follows that

$$\begin{bmatrix} v^T \\ V_i^T \end{bmatrix} (I - T)[\mathbf{1} \quad U_i + F_i] = \begin{bmatrix} 0 & \\ & \widehat{B}_i \end{bmatrix},$$

where $\widehat{B}_i = V_i^T(I - T)(U_i + F_i)$. The group inverse $(I - T)^{\#}$ can be expressed as

$$(I - T)^{\#} = (U_i + F_i)\widehat{B}_i^{-1} V_i^T. \tag{17}$$

The difference between B_i and $\widehat{B}_i$ is

$$\widehat{B}_i - B_i = V_i(P_{ii} - T)(U_i + F_i) + V_i^T(I - P_{ii})F_i.$$

Taking norms we obtain

$$\|\widehat{B}_i - B_i\| \le \|U_i\| \|V_i^T\| (2\eta + \epsilon)(1 + 2s_i + s_i(2\eta + 3\epsilon)) = g(\epsilon, \eta),$$

which implies that

$$\|\widehat{B}_i^{-1}\| \le \frac{\|B_i^{-1}\|}{1 - \|B_i^{-1}\| g(\epsilon, \eta)}. \tag{18}$$

Using (18) and taking norms in (17) completes the proof. □

By the definition of σ_i, it is easy to get that

$$\sigma_i \le \frac{(1 + s_i(2\eta + \epsilon))\|U_i\| \|V_i^T\| \|B_i^{-1}\|}{1 - \|B_i^{-1}\| g(\epsilon, \eta)}.$$

Just as pointed out Remark 3.1, we can expect that σ_i is small and ψ_{ij} is bounded away from zero.

We then define

$$\sigma = \max_i \sigma_i \qquad \text{and} \qquad \psi = \min_i(\min_{j \ne i} \psi_{ij}). \tag{19}$$

4 Decomposition of blocks of the inverse

Because the transition matrix P is irreducible, the matrix $I - P_i$ is a nonsingular M-matrix. In this section we will show that $(I - P_i)^{-1}$ has a special structure. To be precise, we partition $(I - P_i)^{-1}$ conformally with P_i. We will show that the columns of each block are nearly parallel to $\mathbf{1}$. This property will be used to bound the error (5) in the following section.

Theorem 4.1 *Let P_i be the principal submatrix of P in (1) obtained by deleting the i-th block row and block column. Let τ and ϕ be as in (15). Let $(I - P_i)^{-1}$ be partitioned conformally with P_i in the block form $(I - P_i)^{-1} = [G_{lm}]$. If $\tau < \phi$, then for all l and m, the column parallel rate of G_{lm} is bounded as*

$$s(G_{lm}) \leq \frac{\tau}{\phi - \tau}. \tag{20}$$

Proof We only prove this theorem for $i = 1$. For $i \neq 1$, it can be proved in a similar way. Writing $I - P_i$ in the form $I - P_i = D - E$, where

$$D = \begin{bmatrix} I - P_{11} & & & \\ & I - P_{22} & & \\ & & \ddots & \\ & & & I - P_{k-1,k-1} \end{bmatrix} \qquad E = \begin{bmatrix} 0 & P_{12} & \cdots & P_{1,k-1} \\ P_{21} & 0 & \cdots & P_{2,k-1} \\ \vdots & \vdots & & \vdots \\ P_{k-1,1} & P_{k-1,2} & \cdots & 0 \end{bmatrix},$$

we have

$$(I - P_i)^{-1} = (I - D^{-1}E)^{-1}D^{-1} = \sum_{j=0}^{\infty}(D^{-1}E)^j D^{-1}.$$

Let $(D^{-1}E)^j D^{-1}$ be partitioned conformally with P_i in the block form

$$(D^{-1}E)^j D^{-1} = [G_{lm}^{(j)}].$$

Obviously

$$G_{lm} = \sum_{j=0}^{\infty} G_{lm}^{(j)} \tag{21}$$

and the relation between $G_{lm}^{(j)}$ and $G_{lm}^{(j+1)}$ can be described via

$$G_{lm}^{(j+1)} = \sum_{p \neq l}(I - P_{ll})^{-1}P_{lp}G_{pm}^{(j)}.$$

To prove that for all l, m and j, we have

$$s(G_{lm}^{(j)}) \leq \frac{\tau}{\phi - \tau}, \tag{22}$$

we proceed by induction on j. Obviously, (22) holds for $j = 0$, since $G_{ll}^{(0)} = (I - P_{ll})^{-1}$ and $G_{lm}^{(0)} = 0$ for $l \neq m$. Suppose it holds for j. Setting

$$H_{lpm}^{(j)} = (I - P_{ll})^{-1} P_{lp} G_{pm}^{(j)}$$

and applying Lemma 2.2, we have

$$s(H_{lpm}^{(j)}) \leq \frac{\tau\left(1 + \frac{\tau}{\phi - \tau}\right)}{\phi} = \frac{\tau}{\phi - \tau}.$$

¿From Lemma 2.1, it follows that

$$s(G_{lm}^{(j+1)}) \leq \frac{\tau}{\phi - \tau}.$$

Using (21) and Lemma 2.1 completes the proof. $\square$

One interesting consequence of this structure of G_{ij} is that for a nonnegative matrix B, $\|BG_{ij}\|$ is near to $\|B\|\|G_{ij}\|$. To prove this, we let G_{ij} have the *column parallel decomposition* $G_{ij} = \mathbf{1}r^T + R$. We have

$$\|r^T\| \geq \frac{\phi - \tau}{\phi} \|G_{ij}\|$$

and

$$\|BG_{ij}\| = \|B\mathbf{1}r^T + BR\| \geq \|B\|\|r^T\| \geq \frac{\phi - \tau}{\phi} \|B\|\|G_{ij}\|. \tag{23}$$

5 Main result

In this section we will bound the relative errors (5). First we bound them in the case that only one row of blocks of P is perturbed.

Lemma 5.1 *Let P be a transition matrix of a nearly uncoupled Markov chain of the form (1). Let each block $P_{l,i}$ in the l-th block row of P be perturbed by a small perturbation F_{li} with $\|F_{li}\| \leq \eta \|P_{li}\|$ and let the blocks in other block rows be unperturbed. Let $\widehat{P}$ be the perturbed stochastic matrix with stationary distribution $\widehat{\pi}^T = (\widehat{\pi}_1^T, \cdots, \widehat{\pi}_k^T)$. Set*

$$f(\epsilon, \eta) = \frac{(1 + \sigma + \sigma\epsilon)\phi}{\psi(\phi - \tau)},$$

where τ and ϕ are defined as in (15), σ and ψ are defined as in (19). Then for sufficiently small η and for all i,

$$\frac{\|\pi_i^T - \widehat{\pi}_i^T\|}{\|\pi_i^T\|} \le 2f(\epsilon,\eta)\eta + O(\eta^2). \tag{24}$$

Proof We only prove this lemma for $l = k$. If $l \neq k$, then the proof is similar.

Set $F_{k*} = (F_{k1}, \cdots, F_{kk-1})$. The stochastic complement of $\widehat{P}_{kk}$ in $\widehat{P}$ is

$$\widehat{S}_{kk} = S_{kk} + F_{kk} + F_{k*}(I - P_k)^{-1}P_{*k}.$$

Since $(I - P_k)^{-1}P_{*k}\mathbf{1} = \mathbf{1}$,

$$\|F_{k*}(I - P_k)^{-1}P_{*k}\| \le \||F_{k*}|\mathbf{1}\| \le \sum_{1\le i\le k-1} \|F_{ki}\| \le \eta\epsilon,$$

and then $\|S_{kk} - \widehat{S}_{kk}\| \le \eta(1+\epsilon)$. Let

$$v_k^T = \frac{\pi_k^T}{\|\pi_k^T\|} \qquad \text{and} \qquad \widehat{v}_k^T = \frac{\widehat{\pi}_k^T}{\|\widehat{\pi}_k^T\|}.$$

The vectors v_k^T and $\widehat{v}_k^T$ are stationary distributions of S_{kk} and $\widehat{S}_{kk}$, respectively. With σ as in (19), we have

$$\|v_k^T - \widehat{v}_k^T\| \le \|(I - S_{kk})^{\#}\|\|S_{kk} - \widehat{S}_{kk}\| \le \sigma\eta(1+\epsilon).$$

Let

$$v^T = (v_k^T P_{k*}(I - P_k)^{-1}, v_k^T) \qquad \text{and} \qquad \widehat{v}^T = (\widehat{v}_k^T(P_{k*} + F_{k*})(I - P_k)^{-1}, \widehat{v}_k^T)$$

be partitioned conformally with P as

$$v^T = (v_1^T, \cdots, v_k^T) \qquad \text{and} \qquad \widehat{v}^T = (\widehat{v}_1^T, \cdots, \widehat{v}_k^T).$$

It was proved in [10] that

$$\pi^T = \frac{v^T}{\|v^T\|} \qquad \text{and} \qquad \widehat{\pi}^T = \frac{\widehat{v}^T}{\|\widehat{v}^T\|}.$$

Now we bound the relative errors between v_j^T and $\widehat{v}_j^T$ for $1 \le j \le k-1$. Letting $(I - P_i)^{-1}$ be partitioned conformally with P_i as $(I - P_i)^{-1} = [G_{lm}]$, then

$$v_j^T = \sum_{1\le l\le k-1} v_k^T P_{kl} G_{lj} \qquad \text{and} \qquad \widehat{v}_j^T = \sum_{1\le l\le k-1} \widehat{v}_k^T (P_{kl} + F_{k1}) G_{lj}.$$

Using (23) implies that

$$\|v_j^T\| = \sum_{1\le l\le k-1} \|v_k^T P_{kl} G_{lj}\| \ge \frac{\phi-\tau}{\phi} \sum_{1\le l\le k-1} \|v_k^T P_{kl}\|\|G_{lj}\| \ge \frac{(\phi-\tau)\psi}{\phi} \sum_{1\le l\le k-1} \|v_k^T\|\|P_{kl}\|\|G_{lj}\|.$$

Thus

$$\begin{aligned} \|v_j^T - \widehat{v}_j^T\| &\le \sum_{1\le l\le k-1} \|(v_k^T - \widehat{v}_k^T)P_{kl}G_{lj}\| + \sum_{1\le l\le k-1} \|\widehat{v}_k^T F_{kl} G_{lj}\| \\ &\le (\sigma\eta(1+\epsilon)(1+\eta)+\eta) \sum_{1\le l\le k-1} \|v_k^T\|\|P_{kl}\|\|G_{lj}\| \\ &\le (f(\epsilon,\eta)\eta + O(\eta^2))\|v_j^T\|. \end{aligned}$$

Normalizing v^T and $\widehat{v}^T$ to π^T and $\widehat{\pi}^T$, respectively, leads to (24). □

Based on Lemma 5.1, we can bound the relative error (5) as follows. We change the block rows of P into that of $\widetilde{P}$ one row at a time. Each time with Lemma 5.1 we bound the relative errors of aggregate distributions of two subsequently changed transition matrices, since they differ only in one row of blocks. By proper permutation, we assume that the perturbation at each time is added to the last row of blocks. Except for the first time, some blocks P_{ij} in P_k and P_{*k} have been changed to $P_{ij} + F_{ij}$ when we apply Lemma 5.1. This may perturb the quantities τ, ϕ, σ and ψ. It can be easily verified that $\widehat{S}_{kk}$ is always in Φ_k, which means that the quantities σ and ψ can be used in the whole process. Now we show that the other two quantities τ and ϕ are only slightly perturbed.
¿From the *column parallel decomposition* $(I - P_{ii})^{-1} = \mathbf{1}r_i^T + R_i$, we obtain

$$\|(I-P_{ii})\mathbf{1}r_i^T\| = (r_i^T\mathbf{1})\|\sum_{j\ne i} P_{ij}\mathbf{1}\| = \|I - R_i(I-P_{ii})\| \le 2\|R_i\| + 1.$$

It follows from

$$\|F_{ii}\mathbf{1}\| = \|\sum_{j\ne i} F_{ij}\mathbf{1}\| \le \sum_{j\ne i}\|F_{ij}\| \le k\|\sum_{j\ne i} P_{ij}\mathbf{1}\|$$

that

$$\|F_{ii}(I-P_{ii})^{-1}\| = \|F_{ii}\mathbf{1}r_i^T + F_{ii}R_i\| \le (r_i^T\mathbf{1})\|F_{ii}\| + \eta\|R_i\| \le ((2k+1)\|R_i\| + k)\eta.$$

It is pointed out in [18] that we can expect that $\|R_i\|$ is of moderate size. Thus we can expect that the norm $\|F_{ii}(I-P_{ii})^{-1}\|$ is small compared to 1. Then

$$(I - P_{ii} - F_{ii})^{-1} = (I-P_{ii})^{-1}(I - F_{ii}(I-P_{ii})^{-1})^{-1} = \mathbf{1}r_i^T + R_i + C_i,$$

where
$$\frac{\|C_i\|}{\|\mathbf{1}r_i^T + R_i\|} \le \frac{\|F_{ii}(I-P_{ii})^{-1}\|}{1-\|F_{ii}(I-P_{ii})^{-1}\|} = ((2k+1)\|R\|+k)\eta + O(\eta^2).$$
Let $(I-P_{ii}-F_{ii})^{-1}$ have the decomposition $(I-P_{ii}-F_{ii})^{-1} = \mathbf{1}\tilde{r}_i^T + \tilde{R}_i$. A detailed calculation shows that
$$\tilde{\tau}_i = \frac{\|\tilde{R}_i\|}{\tilde{r}_i^T\mathbf{1}} \le (1+O(\eta))\tau_i + O(\eta\epsilon)$$
and
$$\tilde{\phi}_{ij} = \frac{\tilde{r}_i^T(P_{ij}+F_{ij})\mathbf{1}}{(\tilde{r}_i^T\mathbf{1}\|P_{ij}+F_{ij}\|)} \ge (1-O(\eta))\phi_{ij} - O(\eta\epsilon).$$
Let
$$\bar{\tau}_i = \max\{\tau_i,\ \tilde{\tau}_i\} \qquad \bar{\phi}_{ij} = \min\{\phi_{ij},\ \tilde{\phi}_{ij}\}.$$
We define

(25) $$\bar{\tau} = \max_i \tau_i \qquad \text{and} \qquad \bar{\phi} = \min_i(\min_{j\ne i}\phi_{ij}).$$

Obviously, $\bar{\tau}$ and $\bar{\phi}$ are very near to τ and ϕ, respectively.
The following theorem is the main result of this paper.

Theorem 5.2 *Let P be the transition matrix of a nearly uncoupled Markov chain of the form (1). Let $\tilde{P} = P + F$ be a perturbed transition matrix of P with $\|F_{ij}\| \le \eta\|P_{ij}\|$ for all i and j. Let*
$$\pi^T = (\pi_1^T, \cdots \pi_k^T) \quad \text{and} \quad \hat{\pi}^T = (\hat{\pi}_1^T, \cdots, \hat{\pi}_k^T)$$
be stationary distributions of P and $\hat{P}$, respectively. Set
$$\bar{f}(\epsilon,\eta) = \frac{(1+\sigma+\sigma\epsilon)\bar{\phi}}{\psi(\bar{\phi}-\bar{\tau})},$$
where σ and ψ are as in (19), $\bar{\tau}$ and $\bar{\phi}$ are as in (25). If η is sufficiently small, then for $1 \le i \le k$

(26) $$\frac{\|\pi_i^T - \hat{\pi}_i^T\|}{\|\pi_i^T\|} \le 2k\bar{f}(\epsilon,\eta)\eta + O(\eta^2).$$

Proof We change the block rows of P to those of $\tilde{P}$ in k steps, one block row at each step. From Lemma 5.1 and $\bar{\tau}$ and $\bar{\phi}$ in (25), the relative error between the aggregate distributions of two subsequently changed transition matrices is no more than $2\bar{f}(\epsilon,\eta)\eta + O(\eta^2)$. Applying Lemma 5.1 k times gives (26). □

This theorem demonstrates that the sensitivity $\bar{f}(\epsilon, \eta)$ of the aggregate distributions π_i^T to blockwise perturbation F depends on four quantities $\bar{\tau}$, $\bar{\phi}$, σ and ψ. We can expect that $\bar{\tau}$ is small, σ is of moderate size and $\bar{\phi}$ and ψ are bounded away from 0 and so $\bar{f}(\epsilon, \eta)$ is of moderate size, which implies that the aggregate distributions π_i^T are insensitive to small blockwise perturbation F.

Acknowledgement. The author thanks Prof. Mehrmann for many insightful suggestions and comments which greatly improve the presentation of the paper.

References

[1] J.L. BARLOW, *Perturbation results for nearly uncoupled Markov chains with applications to iterative methods,* Numer. Math.,65(1993), pp.51-62.

[2] W.L. CAO AND W.J. STEWART, *Iterative aggregation/disaggregation techniques for nearly uncoupled Markov chains,* Journal of the Association for computing Machinery, 32(1985), pp.702-719.

[3] P.J. COURTOIS, *Decomposability,* Academic Press, New York, 1977.

[4] P. Deuflhard, W. Huisinga, A. Fischer and Ch. Schütte, *Identification of almost aggregates in reversible nearly uncoupled Markov chains,* to appear in Linear Algebra Appl.

[5] G.H. GOLUB AND C.D. MEYER, *Using the QR factorization and group inversion to compute, differentiate and estimate the sensitivity of stationary probabilities for Markov chains,* SIAM J. Alg. Disc. Meth., 7(1985), pp. 273-281.

[6] G. GOLUB AND C. VAN LOAN, *Matrix computations,* 3rd ed., The Johns Hopkins University Press, Baltimore, MD, 1996.

[7] R.E. FUNDERLIC AND C.D. MEYER, *Sensitivity of the stationary distribution vector for an ergodic Markov chain,* Linear Algebra Appl., 76(1986), pp.1-17.

[8] M. HAVIV, *Aggregation/Disaggregation methods for computing the stationary distribution of a Markov chain,* SIAM J. Numer. Anal., 24(1987), pp.952-966.

[9] D.F. MCALLISTER, G.W. STEWART AND W.J. STEWART, *On a Reyleigh-Ritz refinement technique for nearly stochastic matrices,* Linear Algebra. Appl. 60(1984), pp.1-25.

[10] C.D. Meyer, *Stochastic complementation, uncoupling Markov chains, and the theory fo nearly reducible systems,* SIAM Rev., 31(1989), pp.240-272.

[11] C.A. O'cinneide, *Entrywise perturbation theory and error analysis for Markov chains,* Numer. Math., 65(1993), pp.109-120.

[12] H.A. Simon and A. Ando, *Aggregation of variables in dynamic systems,* Econometrica, 29(1961), pp.111-138.

[13] G.W. Stewart, *On the structure of nearly uncoupled Markov chains,* Mathematical Computer Performance and Reliability, G. Iazeolla, P.J. Courtois and A. Hordijk ed. North Holland, Elsevier, 1984, PP. 287-302.

[14] G.W. Stewart, *On the sensitivity of nearly uncoupled Markov chains,* in Numerical Methods for Markov chains, W.J. Stewart ed., North Holland, Amsterdam, 1990, pp. 102-120.

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

[16] Y. Takahashi, *On the effects of small deviations in the transition matrix of a finite Markov chain,* Journal of the Operations Research Society of Japan, 16(1973), pp. 104-129.

[17] R.L. Tweedie, *Perturbation of countable Markov chains and processes,* Ann. Inst. Statistical Math., 32(1980), pp. 283-290.

[18] J. Xue, *Blockwise perturbation theory for block p-cyclic stochastic matrices,* Numer. Math., 84(2000), pp.519-526.

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

Other titles in the SFB393 series:

00-01 G. Kunert. Anisotropic mesh construction and error estimation in the finite element method. January 2000.

00-02 V. Mehrmann, D. Watkins. Structure-preserving methods for computing eigenpairs of large sparse skew-Hamiltonian/Hamiltonian pencils. January 2000.

00-03 X. W. Guan, U. Grimm, R. A. Römer, M. Schreiber. Integrable impurities for an open fermion chain. January 2000.

00-04 R. A. Römer, M. Schreiber, T. Vojta. Disorder and two-particle interaction in low-dimensional quantum systems. January 2000.

00-05 P. Benner, R. Byers, V. Mehrmann, H. Xu. A unified deflating subspace approach for classes of polynomial and rational matrix equations. January 2000.

00-06 M. Jung, S. Nicaise, J. Tabka. Some multilevel methods on graded meshes. February 2000.

00-07 H. Harbrecht, F. Paiva, C. Perez, R. Schneider. Multiscale Preconditioning for the Coupling of FEM-BEM. February 2000.

00-08 P. Kunkel, V. Mehrmann. Analysis of over- and underdetermined nonlinear differential-algebraic systems with application to nonlinear control problems. February 2000.

00-09 U.-J. Görke, A. Bucher, R. Kreißig, D. Michael. Ein Beitrag zur Lösung von Anfangs-Randwert-Problemen einschließlich der Materialmodellierung bei finiten elastisch-plastischen Verzerrungen mit Hilfe der FEM. März 2000.

00-10 M. J. Martins, X.-W. Guan. Integrability of the D_n^2 vertex models with open boundary. March 2000.

00-11 T. Apel, S. Nicaise, J. Schöberl. A non-conforming finite element method with anisotropic mesh grading for the Stokes problem in domains with edges. March 2000.

00-12 B. Lins, P. Meade, C. Mehl, L. Rodman. Normal Matrices and Polar Decompositions in Indefinite Inner Products. March 2000.

00-13 C. Bourgeois. Two boundary element methods for the clamped plate. March 2000.

00-14 C. Bourgeois, R. Schneider. Biorthogonal wavelets for the direct integral formulation of the heat equation. March 2000.

00-15 A. Rathsfeld, R. Schneider. On a quadrature algorithm for the piecewise linear collocation applied to boundary integral equations. March 2000.

00-16 S. Meinel. Untersuchungen zu Druckiterationsverfahren für dichteveränderliche Strömungen mit niedriger Machzahl. März 2000.

00-17 M. Konstantinov, V. Mehrmann, P. Petkov. On Fractional Exponents in Perturbed Matrix Spectra of Defective Matrices. April 2000.

The complete list of current and former preprints is available via
`http://www.tu-chemnitz.de/sfb393/preprints.html`.