ELSEVIER
Linear Algebra and its Applications 326 (2001) 173–191
LINEAR ALGEBRA AND ITS APPLICATIONS
www.elsevier.com/locate/laa

Blockwise perturbation theory for nearly uncoupled Markov chains and its application

Jungong Xue*,[1]

Department of Mathematics, Fudan University, Shanghai 200433, People's Republic of China

Received 4 April 2000; accepted 16 October 2000

Submitted by J.L. Barlow

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 $\pi^{\mathrm{T}} = (\pi_1^{\mathrm{T}}, \ldots, \pi_k^{\mathrm{T}})$. In this paper we bound the relative error in each aggregate distribution π_i^{T} caused by small relative perturbations in P_{ij}. The error bounds demonstrate that nearly uncoupled Markov chains usually lead to well-conditioned problems in the sense of blockwise relative error. As an application, we show that with appropriate stopping criteria, iterative aggregation/disaggregation algorithms will achieve such structured backward errors and compute each aggregate distribution with high relative accuracy. © 2001 Elsevier Science Inc. All rights reserved.

AMS classification: 65F35; 15A42

Keywords: Nearly uncoupled Markov chains; Blockwise perturbation; Nonnegative matrices

* Present address: Fakultät für Mathematik, Technishe Universität Chemnitz, Reichenhainer Str. 41, D-09126 Chemnitz, Germany. Tel.: +49-371-531-2708; fax: +49-371-531-2657.

E-mail address: jungong.xue@mathematik.tu-chemnitz.de (J. Xue).

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

PII: S 0 0 2 4 - 3 7 9 5 (0 0) 0 0 2 8 0 - 9

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

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

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 addressed by many authors, see e.g. [1–4,9–11,13,14]. 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 $\widehat{\pi}^{\mathrm{T}}$ be stationary distributions of transition matrices P and $\widehat{P} = P + F$, respectively; that is, π^{T} and $\widehat{\pi}^{\mathrm{T}}$ are row vectors satisfying

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

where $\mathbf{1}$ is the vector of all ones. According to the standard perturbation theory for Markov chains, see e.g. [5,8],

$$\|\pi^{\mathrm{T}} - \widehat{\pi}^{\mathrm{T}}\| \leqslant \|A^{\#}\|\|F\|, \tag{3}$$

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 [15] that

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

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 [12,17,18].

In [20], 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 $\widehat{\pi}^{\mathrm{T}}$ conformally with P as

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

$$\pi^{\mathrm{T}} = [\pi_1^{\mathrm{T}}, \ldots, \pi_k^{\mathrm{T}}], \quad \widehat{\pi}^{\mathrm{T}} = [\widehat{\pi}_1^{\mathrm{T}}, \ldots, \widehat{\pi}_k^{\mathrm{T}}].$$

If the blocks of the perturbation F satisfy

$$\|F_{ii}\| \leqslant \eta \quad \text{and} \quad \|F_{ij}\| \leqslant \epsilon\eta, \ \ i \neq j, \tag{4}$$

then under some regularity conditions, it is proved in [20] that

$$\frac{\|\pi^{\mathrm{T}} - \widehat{\pi}^{\mathrm{T}}\|}{\|\pi^{\mathrm{T}}\|} \leqslant c\eta. \tag{5}$$

The quantity c in (5) is bounded from above as ϵ tends to 0. However, the upper bound for c is not discussed in [20]. Under the same assumption (4), Barlow [1] bounded the error in another way and obtained

$$\frac{\|\pi^{\mathrm{T}} - \widehat{\pi}^{\mathrm{T}}\|}{\|\pi^{\mathrm{T}}\|} \leqslant c_1\eta + c_2\epsilon, \tag{6}$$

where c_1 and c_2 are well defined. Both error bounds (5) and (6) demonstrate that structured perturbations (4) cause small relative errors in the entire stationary distribution.

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. Under the assumption that

$$\|F_{ij}\| \leqslant \eta\|P_{ij}\|, \quad i, j = 1, \ldots, k, \tag{7}$$

we will prove that

$$\frac{\|\pi_i^{\mathrm{T}} - \widehat{\pi}_i^{\mathrm{T}}\|}{\|\pi_i^{\mathrm{T}}\|} \leqslant 2k\bar{f}(\epsilon, \eta)\eta + \mathrm{O}(\eta^2), \quad i = 1, \ldots, k. \tag{8}$$

Here $\bar{f}(\epsilon, \eta)$ is usually of moderate size. The error bound (8) shows that small relative blockwise perturbations in P induce small relative errors in each aggregate distribution π_i^{T}.

Under the stronger and yet reasonable assumption (7), our result improves that of Barlow in two aspects. First, instead of that in the entire stationary distribution π^{T}, we bound the relative error in each π_i^{T}, and show that it is small however small $\|\pi_i^{\mathrm{T}}\|$ is. This cannot be concluded from Barlow's result (6) when some $\|\pi_i^{\mathrm{T}}\|$ is tiny compared to others. We should mention that even under the regularity conditions in [20], some aggregate distributions can be very small compared to others. To illustrate this, consider the following example. Let P be a 10×10 block transition matrix of the form

$$P = \begin{bmatrix} A_1 & 4E & & & \\ E & A & \ddots & & \\ & \ddots & \ddots & 4E & \\ & & E & A & 4E \\ & & & E & A_2 \end{bmatrix},$$

where

$$A = \begin{bmatrix} 0.5 & 0.5-\epsilon \\ 0.5-\epsilon & 0.5 \end{bmatrix}, \quad E = \begin{bmatrix} 0.1\epsilon & 0.1\epsilon \\ 0.1\epsilon & 0.1\epsilon \end{bmatrix}$$

and

$$A_1 = \begin{bmatrix} 0.5 & 0.5-0.8\epsilon \\ 0.5-0.8\epsilon & 0.5 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 0.5 & 0.5-0.2\epsilon \\ 0.5-0.2\epsilon & 0.5 \end{bmatrix}.$$

For any ϵ, we have

$$\pi_i^{\mathrm{T}} = \beta[4^{i-1},\ 4^{i-1}], \quad \beta = \frac{3}{2(4^{10}-1)}.$$

Therefore, $\|\pi_1^{\mathrm{T}}\|$ is tiny compared to $\|\pi_{10}^{\mathrm{T}}\|$. Small relative error in π^{T} does not mean small relative error in π_1^{T}. In fact, assumption (4) is not enough to guarantee small relative error in π_1^{T}. For example, let perturbations $10^{-4}E$ and $-10^{-4}E$ be introduced to the $(10, 1)$ and $(10, 9)$ blocks of P, respectively. We have

$$\frac{\|\pi_1^{\mathrm{T}} - \tilde{\pi}_1^{\mathrm{T}}\|}{\|\pi_1^{\mathrm{T}}\|} \approx 8.7,$$

which shows π_1^{T} has no accuracy at all.

The second improvement is that we drop the term $c_2\epsilon$ in error bound (6). This makes our result consistent with the fact that the relative error should tend to 0 as η tends to 0.

As an application of our perturbation theory, we show that with appropriate stopping criteria, iterative aggregation/disaggregation algorithms will achieve small blockwise backward error and thus compute each aggregate distribution with high relative accuracy.

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 the quantities involved in $\bar{f}(\epsilon, \eta)$ in (8). 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 get the error bound (8). Finally we discuss the application in iterative aggregation/disaggregation methods.

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 (in real part) is bounded away from 1.

2. Notation and lemmas

Throughout this paper $\| * \|$ 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 $B \leqslant C$ mean $b_{ij} \leqslant c_{ij}$ for all i and j. For vectors, $|y|$ and $y \leqslant x$ 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,

$$S_{ii} = P_{ii} + P_{i*}(I - P_i)^{-1} P_{*i}. \tag{9}$$

It was shown in [11] that S_{ii} is stochastic and $\pi_i^{\mathrm{T}} / \|\pi_i^{\mathrm{T}}\|$ is its stationary distribution.

Each nonnegative matrix A can be decomposed in the form

$$A = \mathbf{1} r^{\mathrm{T}} + R, \tag{10}$$

where r^{T} is a nonnegative row vector and R is a nonnegative matrix with at least one 0 in each column. In other words, the ith entry of r is the minimum of the entries in the ith column of A. Decomposition (10) is called the *column parallel decomposition* for nonnegative matrices. Based on (10), we define the *column parallel rate* of a nonnegative matrix A as

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

We now present two basic properties of the *column parallel rate*.

Lemma 2.1. *Let A_1 and A_2 be nonnegative matrices. Then*

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

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

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

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

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

from which it is straightforward to get that

$$s(A_1 + A_2) \leqslant \max(s(A_1), s(A_2)). \qquad \square$$

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*

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

Set

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

Then

$$s(A_1SA_2) \leqslant \frac{s(A_1)(1 + s(A_2))}{\nu}.$$

Proof. We have

$$A_1SA_2 = (r_1^{\mathrm{T}} S\mathbf{1})\mathbf{1}r_2^{\mathrm{T}} + \mathbf{1}r_1^{\mathrm{T}} SR_2 + R_1S\mathbf{1}r_2^{\mathrm{T}} + R_1SR_2.$$

Let u^{T} be the nonnegative row vector whose ith entry is the minimum of the entries in the ith column of matrix $R_1S\mathbf{1}r_2^{\mathrm{T}} + R_1SR_2$. Then A_1SA_2 has the *column parallel decomposition*

$$A_1SA_2 = \mathbf{1}r_3^{\mathrm{T}} + R_3,$$

where

$$r_3^{\mathrm{T}} = (r_1^{\mathrm{T}} S\mathbf{1})r_2^{\mathrm{T}} + r_1^{\mathrm{T}} SR_2 + u^{\mathrm{T}}$$

and

$$R_3 = R_1S\mathbf{1}r_2^{\mathrm{T}} + R_1SR_2 - \mathbf{1}u^{\mathrm{T}}.$$

Using the nonnegativity of matrices and norm inequalities we get

$$r_3^{\mathrm{T}}\mathbf{1} \geqslant (r_1^{\mathrm{T}} S\mathbf{1})r_2^{\mathrm{T}}\mathbf{1} = \nu(r_1^{\mathrm{T}}\mathbf{1})(r_2^{\mathrm{T}}\mathbf{1})\|S\| \tag{11}$$

and

$$\|R_3\| \leqslant \|R_1S\mathbf{1}r_2^{\mathrm{T}}\| + \|R_1SR_2\| \leqslant \|S\|\|R_1\|(r_2^{\mathrm{T}}\mathbf{1} + \|R_2\|). \tag{12}$$

Combining (11) and (12) completes the proof. $\square$

These two lemmas will be used in Section 4 to investigate the *column parallel 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 $A = \mathbf{1}v^{\mathrm{T}} + Q$, where v^{T} is a nonnegative row vector and $\|Q\|$ is small compared to $\|v^{\mathrm{T}}\|$. Note that we do not assume that Q is nonnegative. Let*

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

and let A have the column parallel decomposition

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

If $m\delta < 1$, *then*

$$s(A) \leqslant \frac{(m+1)\delta}{1 - m\delta} \quad \textit{and} \quad \frac{\|r^{\mathrm{T}} - v^{\mathrm{T}}\|}{\|v^{\mathrm{T}}\|} \leqslant m\delta.$$

Proof. Let u^{T} be the row vector whose ith entry is the minimum of the entries in the ith column of Q. Obviously, $|u^{\mathrm{T}}| \leqslant \|Q\|\mathbf{1}^{\mathrm{T}}$ and thus

$$\|u^{\mathrm{T}}\| \leqslant m\|Q\| = m\delta\|v^{\mathrm{T}}\|.$$

We have the *column parallel decomposition* of A with

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

Thus

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

and

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

3. Spectral analysis of diagonal blocks

In this section we will define some quantities in terms of which we bound the relative error (8). 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* $(I - P_{ii})^{-1} = \mathbf{1}r_i^{\mathrm{T}} + R_i$. We define

$$\tau_i = s((I - P_{ii})^{-1}) = \begin{cases} \frac{\|R_i\|}{\|r_i^{\mathrm{T}}\mathbf{1}\|} & r_i^{\mathrm{T}}\mathbf{1} \neq 0, \\ \infty & r_i^{\mathrm{T}}\mathbf{1} = 0, \end{cases} \tag{13}$$

and for $j \neq i$

$$\phi_{ij} = \begin{cases} \frac{r_i^{\mathrm{T}} P_{ij}\mathbf{1}}{(r_i^{\mathrm{T}}\mathbf{1})\|P_{ij}\|} & \|P_{ij}\| \neq 0, \\ 1 & \|P_{ij}\| = 0. \end{cases} \tag{14}$$

We now analyze τ_i and ϕ_{ij} via the eigenpairs of $I - P_{ii}$.

Let γ_i be the Perron root of P_{ii} and let v_i^{T} be the corresponding left eigenvector normalized so that $v_i^{\mathrm{T}}\mathbf{1} = 1$. 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,

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

Let

$$V_i = J_i (J_i^{\mathrm{T}} U_i)^{-\mathrm{T}}.$$

Then it is proved in [10] that

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

and

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

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

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

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

If $2n_i\delta_i\epsilon < 1$, then τ_i in (13) *is bounded as*

$$\tau_i \leqslant \frac{2(n_i + 1)\delta_i\epsilon}{1 - 2n_i\delta_i\epsilon}. \tag{15}$$

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

$$\phi_{ij} \geqslant \frac{q_{ij} - 2n_i\delta_i\epsilon}{1 + 2n_i\delta_i\epsilon}. \tag{16}$$

Proof. We have

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

Then

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

where

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

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

$$1-\gamma_i \leqslant \epsilon \quad \text{and} \quad \|C_i\| \leqslant 2\epsilon.$$

Therefore

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

Applying Lemma 2.3 gives (15). Let $(I-P_{ii})^{-1}$ have the *column parallel decomposition* $(I-P_{ii})^{-1} = \mathbf{1}r_i^{\mathrm{T}} + R_i$. Applying Lemma 2.3 once more we have

$$\|v_i^{\mathrm{T}} - (1-\gamma_i)r_i^{\mathrm{T}}\| \leqslant 2n_i\delta_i\epsilon.$$

It follows that

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

The eigenvalues of B_i are those of $I-P_{ii}$ other than $1-\gamma_i$. Throughout this paper we always assume that the second largest eigenvalue (in real part) 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 a nonsingular matrix T such that $T^{-1}B_iT$ is a diagonal matrix, then $\|B_i^{-1}\| \leqslant \|T\|\|T^{-1}\|/|\lambda|$, 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 $\|V_i\|_2 \leqslant \sqrt{n_i}$, 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 $q_{ij} \leqslant 1/(n_i\rho_i)$. Therefore ϕ_{ij} can be bounded away from 0 as long as ρ_i is not very large.

For τ_i as in (13) and ϕ_{ij} as in (14), we define

$$\tau = \max_{1\leqslant i\leqslant k} \tau_i \quad \text{and} \quad \phi = \max_{1\leqslant i\leqslant k}(\max_{j\neq i}\phi_{ij}). \tag{17}$$

We still need two other quantities to bound the error (8). To get them, we first define a set of stochastic matrices for each diagonal block P_{ii}

$$\Phi_i = \{T \mid T \geqslant 0,\ T\mathbf{1} = \mathbf{1},\ \|T - P_{ii}\| \leqslant 2\eta + \epsilon\}. \tag{18}$$

Here ϵ is as in (2) and η is as in (7). On each set Φ_i, we define

$$\sigma_i = \sup\{\|(I-T)^{\#}\| \mid T \in \Phi_i\} \tag{19}$$

and

$$\psi_{ij} = \inf\left\{\frac{v^{\mathrm{T}}P_{ij}\mathbf{1}}{\|P_{ij}\|} \mid T \in \Phi_i,\ v^{\mathrm{T}} = v^{\mathrm{T}}T,\ v^{\mathrm{T}}\mathbf{1} = 1\right\}. \tag{20}$$

We can also shed light on σ_i and ψ_{ij} through the spectral analysis of diagonal blocks P_{ii}. Let v^{T} be the stationary distribution of $T \in \Phi_i$, i.e., $v^{\mathrm{T}}T = v^{\mathrm{T}}$ and $v^{\mathrm{T}}\mathbf{1} =$

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 $\|v^{\mathrm{T}}-v_i^{\mathrm{T}}\| \leqslant s_i(2\eta+\epsilon)$. 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 $s_i(2\eta+\epsilon) < q_{ij}$, then it is straightforward to get that

$$\psi_{ij} \geqslant q_{ij} - s_i(2\eta+\epsilon),$$

which implies that ψ_{ij} can be bounded away from 0 if q_{ij} is not small.

The following theorem bounds σ_i.

Theorem 3.2. *Let σ_i be as in* (19) *and let*

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

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

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

Proof. Let $T \in \Phi_i$ and $(I-T)^{\#}$ be the group inverse of $I-T$, 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 $v_i^{\mathrm{T}}\mathbf{1} = 1$. Set $u^{\mathrm{T}} = v_i^{\mathrm{T}} - v^{\mathrm{T}}$. Choosing

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

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

$$v^{\mathrm{T}}(U_i+F_i) = 0, \quad \|F_i\| \leqslant s_i\|U_i\|(2\eta+\epsilon).$$

It follows that

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

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

$$(I-T)^{\#} = (U_i+F_i)\widehat{B}_i^{-1}V_i^{\mathrm{T}}. \tag{21}$$

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

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

Taking norms we obtain

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

which implies that

$$\|\widehat{B}_i^{-1}\| \leqslant \frac{\|B_i^{-1}\|}{1-\|B_i^{-1}\|g(\epsilon,\eta)}. \tag{22}$$

Using (22) and taking norms in (21) we have

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

By the definition of σ_i, we complete the proof. $\square$

From Theorem 3.2, we can also expect that σ_i is of moderate size. We then define

$$\sigma=\max_i \sigma_i \quad \text{and} \quad \psi=\min_i(\min_{j\neq i}\psi_{ij}). \tag{23}$$

In Section 5, we will bound the relative error (8) in terms of τ, ϕ, σ and ψ.

4. Column parallel decomposition of blocks of the inverse

Since 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 **1**. This property will be exploited to bound the error (8) in next 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* (17). *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*

$$s(G_{lm}) \leqslant \frac{\tau}{\phi-\tau}. \tag{24}$$

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

$$D=\begin{bmatrix} I-P_{11} & & & \\ & I-P_{22} & & \\ & & \ddots & \\ & & & I-P_{k-1,k-1}\end{bmatrix}$$

$$E=\begin{bmatrix} 0 & P_{12} & \cdots & P_{1,k-1} \\ P_{21} & 0 & \cdots & P_{2,k-1} \\ \vdots & \vdots & & \vdots \\ P_{k-1,1} & P_{k-1,2} & \cdots & 0\end{bmatrix},$$

we have

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

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

$$G_{lm} = \sum_{j=0}^{\infty} G_{lm}^{(j)} \tag{25}$$

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

$$s(G_{lm}^{(j)}) \leqslant \frac{\tau}{\phi - \tau}, \tag{26}$$

we proceed by induction on j. Obviously, (26) 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

$$s(H_{lpm}^{(j)}) \leqslant \frac{\tau\,(1 + (\tau/(\phi - \tau))}{\phi} = \frac{\tau}{\phi - \tau}.$$

From Lemma 2.1, it follows that

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

Using (25) and Lemma 2.1 completes the proof. □

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* $G_{ij} = \mathbf{1}r^{\mathrm{T}} + R$. We have

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

and

$$\|BG_{ij}\| = \|B\mathbf{1}r^{\mathrm{T}} + BR\| \geqslant \|B\|\|r^{\mathrm{T}}\| \geqslant \frac{\phi - \tau}{\phi} \|B\|\|G_{ij}\|. \tag{27}$$

5. Main result

In this section we will bound the relative error (8). First we bound it 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 form* (1). *Let each block $P_{l,i}$ in the* lth *block row of P be perturbed by a small perturbation F_{li} with $\|F_{li}\| \leqslant \eta\|P_{li}\|$ and let the blocks in other block rows be unperturbed. Let $\widetilde{P}$ be the perturbed stochastic matrix with stationary distribution $\widetilde{\pi}^{\mathrm{T}} = [\widetilde{\pi}_1^{\mathrm{T}}, \ldots, \widetilde{\pi}_k^{\mathrm{T}}]$. Set*

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

where τ and ϕ are defined as in (17), *σ and ψ are defined as in* (23). *Then for sufficiently small η and for all i,*

$$\frac{\|\pi_i^{\mathrm{T}} - \widetilde{\pi}_i^{\mathrm{T}}\|}{\|\pi_i^{\mathrm{T}}\|} \leqslant 2f(\epsilon, \eta)\eta + \mathrm{O}(\eta^2). \tag{28}$$

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

Set $F_{k*} = [F_{k1}, \ldots, F_{kk-1}]$. The stochastic complement of $\widetilde{P}_{kk}$ in $\widetilde{P}$ is

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

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

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

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

$$v_k^{\mathrm{T}} = \frac{\pi_k^{\mathrm{T}}}{\|\pi_k^{\mathrm{T}}\|} \quad \text{and} \quad \widetilde{v}_k^{\mathrm{T}} = \frac{\widetilde{\pi}_k^{\mathrm{T}}}{\|\widetilde{\pi}_k^{\mathrm{T}}\|}.$$

The vectors v_k^{T} and $\widetilde{v}_k^{\mathrm{T}}$ are stationary distributions of S_{kk} and $\widetilde{S}_{kk}$, respectively. With σ as in (23), we have

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

Let

$$v^{\mathrm{T}} = [v_k^{\mathrm{T}}P_{k*}(I - P_k)^{-1}, v_k^{\mathrm{T}}] \quad \text{and} \quad \widetilde{v}^{\mathrm{T}} = [\widetilde{v}_k^{\mathrm{T}}(P_{k*} + F_{k*})(I - P_k)^{-1}, \widetilde{v}_k^{\mathrm{T}}]$$

be partitioned conformally with P as

$$v^{\mathrm{T}} = [v_1^{\mathrm{T}}, \ldots, v_k^{\mathrm{T}}] \quad \text{and} \quad \widetilde{v}^{\mathrm{T}} = [\widetilde{v}_1^{\mathrm{T}}, \ldots, \widetilde{v}_k^{\mathrm{T}}].$$

It was proved in [11] that

$$\pi^{\mathrm{T}} = \frac{v^{\mathrm{T}}}{\|v^{\mathrm{T}}\|} \quad \text{and} \quad \widetilde{\pi}^{\mathrm{T}} = \frac{\widetilde{v}^{\mathrm{T}}}{\|\widetilde{v}^{\mathrm{T}}\|}.$$

We now bound the relative errors between v_j^{T} and $\widetilde{v}_j^{\mathrm{T}}$ for $1 \leqslant j \leqslant k-1$. Letting $(I - P_i)^{-1}$ be partitioned conformally with P_i as $(I - P_i)^{-1} = [G_{lm}]$. Then

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

Using (27) implies that

$$\begin{aligned}\|v_j^{\mathrm{T}}\| &= \sum_{1\leqslant l\leqslant k-1} \|v_k^{\mathrm{T}} P_{kl} G_{lj}\| \\ &\geqslant \frac{\phi-\tau}{\phi} \sum_{1\leqslant l\leqslant k-1} \|v_k^{\mathrm{T}} P_{kl}\| \|G_{lj}\| \\ &\geqslant \frac{(\phi-\tau)\psi}{\phi} \sum_{1\leqslant l\leqslant k-1} \|v_k^{\mathrm{T}}\| \|P_{kl}\| \|G_{lj}\|.\end{aligned}$$

Thus

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

Normalizing v^{T} and $\widetilde{v}^{\mathrm{T}}$ to π^{T} and $\widetilde{\pi}^{\mathrm{T}}$, respectively, leads to (28). $\square$

Based on Lemma 5.1, we can bound the relative error (8) 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 between 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. We now show that the other two quantities τ and ϕ are only slightly perturbed.

From the *column parallel decomposition* $(I - P_{ii})^{-1} = \mathbf{1}r_i^{\mathrm{T}} + R_i$, we obtain

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

It follows from

$$\|F_{ii}\mathbf{1}\| = \left\|\sum_{j\neq i} F_{ij}\mathbf{1}\right\| \leqslant \sum_{j\neq i}\|F_{ij}\| \leqslant k\eta \left\|\sum_{j\neq i} P_{ij}\mathbf{1}\right\|$$

that

$$\begin{aligned}\|F_{ii}(I-P_{ii})^{-1}\| &= \|F_{ii}\mathbf{1}r_i^{\mathrm{T}} + F_{ii}R_i\| \\ &\leqslant (r_i^{\mathrm{T}}\mathbf{1})\|F_{ii}\| + \eta\|R_i\|\end{aligned}$$

$$\leqslant ((2k+1)\|R_i\| + k)\eta.$$

It is pointed out in [19] 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

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

where

$$\begin{aligned}\frac{\|C_i\|}{\|\mathbf{1}r_i^{\mathrm{T}} + R_i\|} &\leqslant \frac{\|F_{ii}(I - P_{ii})^{-1}\|}{1 - \|F_{ii}(I - P_{ii})^{-1}\|} \\ &= ((2k+1)\|R\| + k)\eta + \mathrm{O}(\eta^2).\end{aligned}$$

Let $(I - P_{ii} - F_{ii})^{-1}$ have the decomposition $(I - P_{ii} - F_{ii})^{-1} = \mathbf{1}\widetilde{r}_i^{\mathrm{T}} + \widetilde{R}_i$. A detailed calculation shows that

$$\widetilde{\tau}_i = \frac{\|\widetilde{R}_i\|}{\widetilde{r}_i^{\mathrm{T}}\mathbf{1}} \leqslant (1 + \mathrm{O}(\eta))\tau_i + \mathrm{O}(\eta\epsilon)$$

and

$$\widetilde{\phi}_{ij} = \frac{\widetilde{r}_i^{\mathrm{T}}(P_{ij} + F_{ij})\mathbf{1}}{(\widetilde{r}_i^{\mathrm{T}}\mathbf{1}\|P_{ij} + F_{ij}\|)} \geqslant (1 - \mathrm{O}(\eta))\phi_{ij} - \mathrm{O}(\eta\epsilon).$$

Let

$$\bar{\tau}_i = \max\{\tau_i, \widetilde{\tau}_i\}, \quad \bar{\phi}_{ij} = \min\{\phi_{ij}, \widetilde{\phi}_{ij}\}.$$

We define

$$\bar{\tau} = \max_i \bar{\tau}_i \quad \text{and} \quad \bar{\phi} = \min_i(\min_{j \neq i} \bar{\phi}_{ij}). \tag{29}$$

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 form* (1). *Let* $\widetilde{P} = P + F$ *be a perturbed transition matrix of P with* $\|F_{ij}\| \leqslant \eta\|P_{ij}\|$ *for all i and j. Let*

$$\pi^{\mathrm{T}} = [\pi_1^{\mathrm{T}}, \ldots \pi_k^{\mathrm{T}}] \quad \textit{and} \quad \widehat{\pi}^{\mathrm{T}} = [\widehat{\pi}_1^{\mathrm{T}}, \ldots, \widehat{\pi}_k^{\mathrm{T}}]$$

be stationary distributions of P and $\widehat{P}$, *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* (23), $\bar{\tau}$ *and* $\bar{\phi}$ *are as in* (29). *If* η *is sufficiently small, then for* $1 \leqslant i \leqslant k$

$$\frac{\|\pi_i^{\mathrm{T}} - \widehat{\pi}_i^{\mathrm{T}}\|}{\|\pi_i^{\mathrm{T}}\|} \leqslant 2k\bar{f}(\epsilon, \eta)\eta + \mathrm{O}(\eta^2). \tag{30}$$

Proof. We change the block rows of P to those of $\widehat{P}$ in k steps, one block row at each step. From Lemma 5.1 and $\bar{\tau}$ and $\bar{\phi}$ in (29), the relative error between the aggregate distributions of two subsequently changed transition matrices is no more than $2\bar{f}(\epsilon, \eta)\eta + \mathrm{O}(\eta^2)$. Applying Lemma 5.1 k times gives (30). $\square$

Remark 5.1.
1. Theorem 5.2 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.
2. If each block is a scalar, i.e., $n_i = 1$ for $1 \leqslant i \leqslant k$, then $\bar{\tau} = \sigma = 0$, $\bar{\phi} = \psi = 1$. In this case, Theorem 5.2 is just the entrywise perturbation theory obtained in [12,17,18].
3. Even if $\widehat{P} = P + F$ is not nonnegative, as long as $\widehat{\pi}$, the normalized left eigenvector corresponding to eigenvalue 1 is nonnegative, the error bound (30) still holds. We will employ this fact in the following section.

6. Application in iterative aggregation/disaggregation methods

Iterative methods coupled with aggregation/disaggregation technique is an important tool to compute the stationary distribution of a large-scale nearly uncoupled Markov chain, see [2,9,10]. In this section we will show that under a proper stopping criteria, iterative aggregation/disaggregation methods can achieve small blockwise relative backward error and thus can compute the stationary distribution accurately in the sense of blockwise relative error.

Let P be as in (1) and $A = I - P$. Suppose that we have a computed stationary distribution

$$\widehat{\pi}^{\mathrm{T}} = [\widehat{\pi}_1^{\mathrm{T}}, \ldots, \widehat{\pi}_k^{\mathrm{T}}]$$

such that

$$\widehat{\pi}^{\mathrm{T}} A = r^{\mathrm{T}}, \quad \widehat{\pi}^{\mathrm{T}}\mathbf{1} = 1, \tag{31}$$

where r^{T} is partitioned conformally with P as

$$r^{\mathrm{T}} = [r_1^{\mathrm{T}}, r_2^{\mathrm{T}}, \ldots, r_k^{\mathrm{T}}]$$

and each r_i^{T} satisfies the stopping criteria

$$\|r_i^{\mathrm{T}}\| \leqslant \mathrm{tol}\|\widehat{\pi}_i^{\mathrm{T}}\|, \quad i = 1, 2, \ldots, k. \tag{32}$$

Each $\widehat{\pi}_i^{\mathrm{T}}$ can be decomposed as

$$\widehat{\pi}_i^{\mathrm{T}} = \|\widehat{\pi}_i^{\mathrm{T}}\|\widehat{v}_i^{\mathrm{T}}, \quad i = 1, 2, \ldots, k$$

where $\widehat{v}_i^{\mathrm{T}}$ is already available before the aggregation step while $\widehat{y}^{\mathrm{T}} = [\|\widehat{\pi}_1^{\mathrm{T}}\|, \|\widehat{\pi}_2^{\mathrm{T}}\|, \ldots, \|\widehat{\pi}_k^{\mathrm{T}}\|]$ is obtained in the aggregation step by solving the linear system

$$y^{\mathrm{T}}B = 0, \quad y^{\mathrm{T}}\mathbf{1} = 1. \tag{33}$$

Here

$$B = (b_{ij}), \quad b_{ij} = -\widehat{v}_i^{\mathrm{T}} P_{ij}\mathbf{1}, \quad i \neq j,$$

and

$$b_{ii} = -\sum_{j \neq i} b_{ij}, \quad i = 1, 2, \ldots, k.$$

Suppose we solve (33) via GTH algorithm [7], which produces an accurate solution in the sense of entrywise relative error. From the error analysis of O'Cinneide [12], we have

$$\widehat{y}_i = (1 + \epsilon_i) y_i, \quad i = 1, 2, \ldots, k,$$

where $|\epsilon_i| \leqslant 9k^2\mathsf{u}$ and u is the unit roundoff. Denote

$$\widetilde{\pi}^{\mathrm{T}} = [y_1\widehat{v}_1^{\mathrm{T}}, y_2\widehat{v}_2^{\mathrm{T}}, \ldots, y_k\widehat{v}_k^{\mathrm{T}}] \quad \text{and} \quad \widetilde{\pi}^{\mathrm{T}}A = \widetilde{r}^{\mathrm{T}} = [\widetilde{r}_1^{\mathrm{T}}, \widetilde{r}_2^{\mathrm{T}}, \ldots, \widetilde{r}_k^{\mathrm{T}}].$$

It is easy to show that

$$\widetilde{r}_i^{\mathrm{T}}\mathbf{1} = 0, \quad i = 1, 2, \ldots, k.$$

Moreover,

$$\begin{aligned}
\|\widetilde{r}_i^{\mathrm{T}}\| &= \left\| \widehat{r}_i^{\mathrm{T}} - \epsilon_i \widehat{v}_i^{\mathrm{T}}(I - P_{ii}) + \sum_{j \neq i} \epsilon_j y_i \widehat{v}_j^{\mathrm{T}} P_{ji} \right\| \\
&\leqslant \|\widehat{r}_i^{\mathrm{T}}\| + 9k^2\mathsf{u} \sum_{j=1}^{k} y_j |b_{ji}| \\
&= \|\widehat{r}_i^{\mathrm{T}}\| + 18k^2\mathsf{u} y_i b_{ii} \\
&\leqslant \left(\mathrm{tol} + \frac{18k^2\mathsf{u}}{1 - 9k^2\mathsf{u}} \right) \|\widehat{\pi}_i^{\mathrm{T}}\|.
\end{aligned}$$

We now turn to construct the backward error for the computed solution $\widehat{\pi}^{\mathrm{T}}$. Denoting

$$D = \begin{bmatrix} \frac{1}{1+\epsilon_1} I_{n_1} & & & \\ & \frac{1}{1+\epsilon_2} I_{n_2} & & \\ & & \ddots & \\ & & & \frac{1}{1+\epsilon_k} I_{n_k} \end{bmatrix},$$

where I_{n_i} is the identity matrix of order n_i, we have

$$\widehat{\pi}^{\mathrm{T}} D A = \widetilde{r}^{\mathrm{T}}. \tag{34}$$

We can attribute each residual $\widetilde{r}_i^{\mathrm{T}}$ to the perturbation

$$\frac{1}{\|\widehat{\pi}_i^{\mathrm{T}}\|} \mathbf{1} \widetilde{r}_i^{\mathrm{T}}$$

in the ith diagonal block of DA. Thus we can rewrite (34) as

$$\widehat{\pi}^{\mathrm{T}}(P + F) = \widehat{\pi}^{\mathrm{T}},$$

where

$$F_{ij} = -\frac{\epsilon_i}{1 + \epsilon_i} P_{ij}, \quad j \neq i$$

and

$$F_{ii} = \frac{\epsilon_i}{1 + \epsilon_i}(I - P_{ii}) - \frac{1}{\|\widehat{\pi}_i^{\mathrm{T}}\|}\mathbf{1}\widetilde{r}_i^{\mathrm{T}}.$$

Obviously,

$$F\mathbf{1} = 0$$

and

$$\|F_{ij}\| \leqslant (\mathrm{tol} + 18k^2\mathsf{u} + \mathrm{O}(\mathsf{u}^2))\|P_{ij}\|, \quad i, j = 1, 2, \ldots, k,$$

With the stopping criteria (32), iterative aggregation/disaggregation methods can achieve a small blockwise relative backward error. Applying Theorem 5.2 and Remark 5.1, we know the computed stationary distribution $\widehat{\pi}^{\mathrm{T}}$ is accurate in the sense of blockwise relative error.

Acknowledgements

I am grateful to Prof. Mehrmann for many insightful suggestions and comments which greatly improve the presentation of the paper. I would like to thank the referee for his very pertinent remarks and particularly for his pointing out the application in aggregation/disaggregation methods.

References

[1] J.L. Barlow, Perturbation results for nearly uncoupled Markov chains with applications to iterative methods, Numer. Math. 65 (1993) 51–62.
[2] W.L. Cao, W.J. Stewart, Iterative aggregation/disaggregation techniques for nearly uncoupled Markov chains, J. ACM 32 (1985) 702–719.
[3] P.J. Courtois, Decomposability, Academic Press, New York, 1977.
[4] P. Deuflhard, W. Huisinga, A. Fischer, Ch. Schütte, Identification of almost aggregates in reversible nearly uncoupled Markov chains, Linear Algebra Appl. 315 (2000) 39–59.
[5] G.H. Golub, 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) 273–281.
[6] G. Golub, C. Van Loan, Matrix Computations, third ed., Johns Hopkins University Press, Baltimore, MD, 1996.
[7] W.K. Grassman, M.I. Taksar, D.P. Heyman, Regenerative analysis and steady state distribution for Markov chains, Oper. Res. 33 (1985) 1107–1116.

[8] R.E. Funderlic, C.D. Meyer, Sensitivity of the stationary distribution vector for an ergodic Markov chain, Linear Algebra Appl. 76 (1986) 1–17.
[9] M. Haviv, Aggregation/disaggregation methods for computing the stationary distribution of a Markov chain, SIAM J. Numer. Anal. 24 (1987) 952–966.
[10] D.F. McAllister, G.W. Stewart, W.J. Stewart, On a Reyleigh–Ritz refinement technique for nearly stochastic matrices, Linear Algebra Appl. 60 (1984) 1–25.
[11] C.D. Meyer, Stochastic complementation, uncoupling Markov chains, and the theory fo nearly reducible systems, SIAM Rev. 31 (1989) 240–272.
[12] C.A. O'Cinneide, Entrywise perturbation theory and error analysis for Markov chains, Numer. Math. 65 (1993) 109–120.
[13] H.A. Simon, A. Ando, Aggregation of variables in dynamic systems, Econometrica 29 (1961) 111–138.
[14] G.W. Stewart, On the structure of nearly uncoupled Markov chains, in: G. Iazeolla, P.J. Courtois, A. Hordijk (Eds.), Mathematical Computer Performance and Reliability, Elsevier, Amsterdam, 1984, pp. 287–302.
[15] G.W. Stewart, On the sensitivity of nearly uncoupled Markov chains, in: W.J. Stewart (Ed.), Numerical Methods for Markov Chains, North-Holland, Amsterdam, 1990, pp. 102–120.
[16] G.W. Stewart, On the perturbation of Markov chains with nearly transient states, Numer. Math. 65 (1993) 135–141.
[17] Y. Takahashi, On the effects of small deviations in the transition matrix of a finite Markov chain, J. Oper. Res. Soc. Jpn 16 (1973) 104–129.
[18] R.L. Tweedie, Perturbation of countable Markov chains and processes, Ann. Inst. Stat. Math. 32 (1980) 283–290.
[19] J. Xue, Blockwise perturbation theory for block p-cyclic stochastic matrices, Numer. Math. 84 (2000) 519–526.
[20] G. Zhang, On the sensitivity of the solution of nearly uncoupled Markov chains, SIAM J. Matrix Anal. Appl. 4 (1993) 1112–1123.