University of Maryland College Park

Institute for Advanced Computer Studies TR–92–14
Department of Computer Science TR–2835

On the Perturbation of Markov Chains with Nearly Transient States*

G. W. Stewart†

February, 1992
Revised October, 1992

ABSTRACT

Let A be an irreducible stochastic matrix of the form

$$A = \begin{pmatrix} A_{11} & E_{12} \\ A_{21} & A_{22} \end{pmatrix}.$$

If E_{12} were zero, the states corresponding to A_{22} would be transient in the sense that if the steady state vector y^{T} is partitioned conformally in the form $(y_1^{\mathrm{T}} \ y_2^{\mathrm{T}})$ then $y_2^{\mathrm{T}} = 0$. If E_{12} is small, then y_2^{T} will be small, and the states are said to be nearly transient. It this paper it is shown that small relative perturbations in A_{11}, A_{21}, and A_{22}, though potentially larger than y_2^{T}, induce only small relative perturbations in y_2^{T}.

*This report is available by anonymous ftp from `thales.cs.umd.edu` in the directory `pub/reports`.

†Department of Computer Science and Institute for Advanced Computer Studies, University of Maryland, College Park, MD 20742. This work was supported in part by the Air Force Office of Scientific Research under Contract AFOSR-87-0188 and the National Science Foundataion under grant CCR 9115568 and was done while the author was a visiting faculty member at the Institute for Mathmatics and Its Applications, The University of Minnesota, Minneapolis, MN 55455.

On the Perturbation of Markov Chains with Nearly Transient States

G. W. Stewart

ABSTRACT

Let A be an irreducible stochastic matrix of the form

$$A = \begin{pmatrix} A_{11} & E_{12} \\ A_{21} & A_{22} \end{pmatrix}.$$

If E_{12} were zero, the states corresponding to A_{22} would be transient in the sense that if the steady state vector y^{T} is partitioned conformally in the form $(y_1^{\mathrm{T}}\ y_2^{\mathrm{T}})$ then $y_2^{\mathrm{T}} = 0$. If E_{12} is small, then y_2^{T} will be small, and the states are said to be nearly transient. It this paper it is shown that small relative perturbations in A_{11}, A_{21}, and A_{22}, though potentially larger than y_2^{T}, induce only small relative perturbations in y_2^{T}.

1. Introduction

The concerns of the paper are best illustrated by a 2×2 example. Consider the nonnegative matrix

$$A_2 = \begin{pmatrix} 1-\epsilon & \epsilon \\ \alpha & 1-\alpha \end{pmatrix},$$

where ϵ is small and α is of order of magnitude one (e.g., $\alpha = \frac{1}{2}$). When $\epsilon = 0$, the second state of the Markov chain corresponding to A_2 is transient: it eventually goes away, never to return. When ϵ is small but positive, we shall say that the state is *nearly transient*. The near transience of the state is reflected by the second component of the steady state vector

$$y^{\mathrm{T}} = \frac{(1\ \epsilon/\alpha)}{1+\epsilon/\alpha},$$

which tends to zero with ϵ.

Now consider the perturbed matrix

$$\tilde{A}_2 = \begin{pmatrix} 1-\epsilon-\eta & \epsilon \\ \alpha & 1-\alpha \end{pmatrix},$$

where η is small compared to one, but not necessarily small compared to ϵ. If we seek the eigenvalues of $I - \tilde{A}_2$ in the form $\lambda_1 = \eta + \delta$ and $\lambda_2 = \alpha + \epsilon - \delta$, then from the equation $\lambda_1 \lambda_2 = \alpha\eta$, we obtain the approximation

$$\delta \cong -\frac{\eta\epsilon}{\alpha}.$$

If we seek the eigenvector corresponding to the smallest eigenvalue $\eta + \delta$ in the form $\tilde{y}^{\mathrm{T}} = (1\ \xi)$, then from the first component of the equation $\tilde{y}^{\mathrm{T}}(I - \tilde{A}_2) = (\eta + \delta)\tilde{y}^{\mathrm{T}}$ it follows that

$$\epsilon + \eta - \xi\alpha = \eta + \delta,$$

or

$$\xi = \frac{\epsilon - \delta}{\alpha} \cong \frac{\epsilon}{\alpha}\Big(1 + \frac{\eta}{\alpha}\Big). \tag{1.1}$$

In other words, a change of order η in the leading element of A_2 makes a *relative* change of only $O(\eta)$ in the components of the steady-state vector. This implies that the probability of being in the nearly transient state, however small, is insensitive to potentially much larger perturbations in the (1,1)-element of the matrix.

The purpose of this paper is to generalize this result to stochastic matrices of the form

$$A = \begin{pmatrix} A_{11} & E_{12} \\ A_{21} & A_{22} \end{pmatrix}; \tag{1.2}$$

that is, to a chain with a group of nearly transient states. Before proceeding, however, it will be worth while to examine the above example more closely for things to generalize.

It is easy to see that the steady-state vector is also insensitive to small perturbations in the $(2,1)$- and $(2,2)$-elements of A_2. We will show that this generalizes: under suitable restrictions on A, the small components of the steady-state vector corresponding to the nearly transient states are insensitive to perturbations in A_{11}, A_{12}, and A_{22}. On the other hand, in the example the small component of the steady-state vector is very sensitive to changes in ϵ itself, and we may expect a similar sensitivity in the general case to perturbations in E_{12}.

The condition that $\alpha = O(1)$ is necessary. For as α becomes small, the approximations that lead to (1.1) become increasingly inaccurate and break down entirely when $\alpha = O(\epsilon)$. (This break-down agrees with what we know about the perturbation theory for nearly completely decomposable chains, where the steady state vector is sensitive to such perturbations [2].) In generalizing the result, however, it is not be enough to require that $A_{21} = O(1)$, and we will formulate an alternate condition in terms of A_{22}.

Finally, we note that the perturbation in the example does not leave the matrix stochastic. At first blush, this generality may seem superfluous, since in applications to Markov chains we should expect both the matrix and its perturbation to be stochastic. However, it turns out that certain numerical algorithms, among them Gaussian elimination, introduce perturbations that render the matrix in question nonstochastic.

The paper is organized as follows. In the next section, we will introduce some preliminary transformations of the problem. In Section 3 we will establish a general perturbation bound, and in Section 4 we will discuss its consequences.

Throughout this paper $\|\cdot\|$ stands for the Euclidean vector norm and the subordinate matrix norm defined by

$$\|A\| = \sup_{\|x\|=1} \|Ax\|.$$

2. The Transformed Problem

To state our problem more precisely, let the matrix A of (1.2) and its submatrices A_{11} and A_{22} be irreducible, and let

$$y = (y_1^{\mathrm{T}}\ y_2^{\mathrm{T}})$$

be its Perron vector partitioned conformally. Let

$$G = \begin{pmatrix} G_{11} & 0 \\ G_{21} & G_{22} \end{pmatrix}$$

be a matrix that is small compared to one (but not necessarily compared to E_{12}), and assume that

$$\tilde{A} \equiv A + G = \begin{pmatrix} A_{11} + G_{11} & E_{12} \\ A_{21} + G_{21} & A_{22} + G_{22} \end{pmatrix}$$

is also irreducible. Let

$$\tilde{y} = (\tilde{y}_1^{\mathrm{T}}\ \tilde{y}_2^{\mathrm{T}})$$

be the Perron vector of $\tilde{A}$. Then our problem is to establish perturbation bounds for $\tilde{y}_2^{\mathrm{T}}$.

A technical difficulty presents itself immediately. If E_{12} is small, the matrix A_{11} 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 A into a form in which the offending eigenvalue is isolated.

Let β_{11} be the Perron eigenvalue of A_{11} and let the corresponding positive left eigenvector be u_1^{T}, normalized so that $\|u_1\| = 1$. Let

$$U = (u_1 \; U_2)$$

be orthogonal. Then it is easily verified that $U^{\mathrm{T}} A_{11} U$ has the form

$$U^{\mathrm{T}} A_{11} U = \begin{pmatrix} \beta_{11} & 0 \\ b_{21} & B_{22} \end{pmatrix}.$$

The eigenvalues of B_{22} are the eigenvalues of A_{11} other than β_{11}. Since A_{11} is substochastic, $I - B_{22}$ is nonsingular.

Now let

$$\begin{pmatrix} u_1^{\mathrm{T}} \\ U_2^{\mathrm{T}} \end{pmatrix} E_{12} = \begin{pmatrix} f_{13} \\ F_{23} \end{pmatrix}$$

and

$$A_{21}(u_1 \; U_2) = (b_{31} \; B_{32}).$$

Then

$$B \equiv \begin{pmatrix} U^{\mathrm{T}} & 0 \\ 0 & I \end{pmatrix} \begin{pmatrix} A_{11} & E_{12} \\ A_{21} & A_{22} \end{pmatrix} \begin{pmatrix} U & 0 \\ 0 & I \end{pmatrix} = \begin{pmatrix} \beta_{11} & 0 & f_{13}^{\mathrm{T}} \\ b_{21} & B_{22} & F_{23} \\ b_{31} & B_{32} & B_{33} \end{pmatrix},$$

where $B_{33} = A_{22}$.

Since both y_1^{T} and u_1^{T} are positive, $y_1^{\mathrm{T}} u_1 > 0$. It follows that we may renormalize y^{T} so that

$$y^{\mathrm{T}} U = (1 \; p_2^{\mathrm{T}} \; p_3^{\mathrm{T}}),$$

where $p_2^{\mathrm{T}} = y_1^{\mathrm{T}} U_2$ and $p_3^{\mathrm{T}} = y_2^{\mathrm{T}}$. In terms of the transformed problem, our goal is to find perturbation bounds on p_3^{T}, when the quantities B_{ij} are subject to perturbations.

It is easy to obtain a linear equation for p_3^{T}. Because $y^{\mathrm{T}}\mathrm{diag}(U, I)$ is a null vector of $I - B$ it follows that p_2^{T} and p_3^{T} satisfy

$$(p_2^{\mathrm{T}} \; p_3^{\mathrm{T}}) \begin{pmatrix} I - B_{22} & -F_{23} \\ -B_{32} & I - B_{33} \end{pmatrix} = (0 \; f_{13}^{\mathrm{T}}).$$

Eliminating p_2^{T} from this equation, we obtain

$$p_3^{\mathrm{T}}\Big(I - B_{33} - B_{32}(I - B_{22})^{-1}F_{23}\Big) = f_{13}^{\mathrm{T}}. \qquad (2.1)$$

It is equally easy to obtain an equation for the perturbed vector $\tilde{p}_3$. Let

$$H = U^{\mathrm{T}}GU = \begin{pmatrix} \eta_{11} & h_{12}^{\mathrm{T}} & 0 \\ h_{21} & H_{22} & 0 \\ h_{31} & H_{32} & H_{33} \end{pmatrix},$$

and assume that $\tilde{A} = A + G$ is stochastic. Then in the transformed system (with tildes denoting the obvious perturbations)

$$(\tilde{p}_2^{\mathrm{T}} \;\; \tilde{p}_3^{\mathrm{T}}) \begin{pmatrix} I - \tilde{B}_{22} & -F_{23} \\ -\tilde{B}_{32} & I - \tilde{B}_{33} \end{pmatrix} = (h_{12}^{\mathrm{T}} \;\; f_{13}^{\mathrm{T}}).$$

It follows that

$$\tilde{p}_3^{\mathrm{T}}\Big(I - \tilde{B}_{33} - \tilde{B}_{32}(I - \tilde{B}_{22})^{-1}F_{23}\Big) = f_{13}^{\mathrm{T}} - h_{12}^{\mathrm{T}}(I - \tilde{B}_{22})^{-1}F_{23}. \qquad (2.2)$$

3. The Perturbation Bound

In this section we will establish perturbation bounds for $\tilde{p}_3$. It will be convenient to have an abbreviated notation for the norms occurring in the bounds. Accordingly, we set

$$\begin{aligned} \beta &\equiv \|B\| = \|A\|, \\ \eta &\equiv \|H\| = \|G\|, \\ \gamma_i &\equiv \|(I - B_{ii})\|, \qquad (i = 2, 3). \end{aligned} \qquad (3.1)$$

The equalities in the above definitions follow from the fact that a transformation by the orthogonal matrix U does not change the spectral norm. The same symbols with tildes denote the norms of the perturbed quantities; e.g., $\tilde{\beta} = \|\tilde{B}\|$.

We begin by collecting some standard results from the perturbation of linear systems (see, e.g., [1, 3]).

Theorem 3.1. *Let C be nonsingular and let $\tilde{C} = C + Q$, where*

$$\|C^{-1}\|\|Q\| < 1.$$

Then $\tilde{C}$ is nonsingular,

$$\|\tilde{C}^{-1}\| \le \frac{\|C^{-1}\|}{1-\|C^{-1}\|\|Q\|}, \tag{3.2}$$

and

$$\|\tilde{C}^{-1} - C^{-1}\| \le \frac{\|C^{-1}\|\|Q\|}{1-\|C^{-1}\|\|Q\|}, \tag{3.3}$$

Moreover, if

$$x^{\mathrm{T}}C = d^{\mathrm{T}} \qquad \text{and} \qquad \tilde{x}^{\mathrm{T}}\tilde{C} = d^{\mathrm{T}} + q^{\mathrm{T}},$$

then

$$\frac{\|\tilde{x}-x\|}{\|x\|} = \frac{\|C^{-1}\|}{1-\|C^{-1}\|\|Q\|}\left(\|Q\| + \frac{\|q\|}{\|x\|}\right). \tag{3.4}$$

Now let C denote the matrix in equation (2.1) for p_3^{T} and let d^{T} denote the right-hand side. Let $\tilde{C}$ denote the matrix in the perturbed system (2.2) and $\tilde{d}^{\mathrm{T}}$ denote the right-hand side. To apply Theorem 3.1, we must bound $\|\tilde{C}-C\|$ and $\|\tilde{d}^{\mathrm{T}} - d^{\mathrm{T}}\|$.

We have

$$\tilde{C}-C = (B_{33}-\tilde{B}_{33}) + (B_{32}-\tilde{B}_{32})(I-\tilde{B}_{22})^{-1}F_{23} + B_{32}\Big((I-B_{22})^{-1} - (I-\tilde{B}_{22})^{-1}\Big)F_{23}.$$

On taking norms we get

$$\|\tilde{C}-C\| \le \eta + \eta\tilde{\gamma}_2\epsilon + \beta\frac{\eta\epsilon\gamma_2}{1-\gamma_2\eta},$$

The third term of the bound follows from (3.3) under the assumption that $\eta\gamma_2 < 1$. Since from (3.2) we have $\tilde{\gamma}_2 \le \gamma_2/(1-\eta\gamma_2)$, if we set $\bar{\eta} = \eta/(1-\eta\gamma_2)$, we have

$$\|\tilde{C}-C\| \le \bar{\eta}(1+\gamma_2\epsilon+\beta\gamma_2\epsilon).$$

Similarly,

$$\|\tilde{d}-d\| \le \bar{\eta}\epsilon.$$

If we now use these bounds in (3.4), we get the following theorem (remember that $y_2^{\mathrm{T}} = p_3^{\mathrm{T}}$).

Theorem 3.2. *Let the irreducible stochastic matrix A have the form*

$$A = \begin{pmatrix} A_{11} & E_{12} \\ A_{21} & A_{22} \end{pmatrix},$$

where A_{11} and A_{22} are irreducible, and let

$$\tilde{A} = A + G \equiv \begin{pmatrix} A_{11} & E_{12} \\ A_{21} & A_{22} \end{pmatrix} + \begin{pmatrix} G_{11} & 0 \\ G_{21} & G_{22} \end{pmatrix}$$

also be stochastic. In the notation of (3.1), assume that $\eta\gamma_2 < 1$, and set

$$\bar{\eta} = \frac{\eta}{1 - \eta\gamma_2} \qquad \text{and} \qquad \mu = 1 + \gamma_2\epsilon + \beta\gamma_2\epsilon.$$

If

$$\bar{\eta}\mu\gamma_3 < 1,$$

then

$$\frac{\|\tilde{y}_2^{\mathrm{T}} - y_2^{\mathrm{T}}\|}{\|y_2^{\mathrm{T}}\|} \le \frac{\bar{\eta}\gamma_3}{1 - \bar{\eta}\mu\gamma_3}\left[\mu + \frac{\epsilon}{\|y_2^{\mathrm{T}}\|}\right]. \tag{3.5}$$

4. Discussion

The bound (3.5) gives the promised result. Provided $\|y_2^{\mathrm{T}}\|$ is of order ϵ (more on this point later), the *relative* perturbation of y_2^{T} is a small multiple of η. The condition that $\bar{\eta}\mu\gamma_3 < 1$ is the condition on A_{22} mentioned in the introduction. It essentially says that the eigenvalues of A_{22} are bounded away from one. In particular, it prevents the matrix A_{21} from being small — the condition used in the 2×2 example in the introduction.

It is instructive to examine the asymptotic form of the bound as ϵ and η approach zero. In this case, μ approaches one and $\bar{\eta}$ approaches η. Consequently, (3.5) has the asymptotic form

$$\frac{\|\tilde{y}_2^{\mathrm{T}} - y_2^{\mathrm{T}}\|}{\|y_2^{\mathrm{T}}\|} \lesssim \eta\gamma_3\left[1 + \frac{\epsilon}{\|y_2^{\mathrm{T}}\|}\right]. \tag{4.1}$$

Thus if $\epsilon/\|y_2^{\mathrm{T}}\|$ is near one, the factor controlling the size of the perturbation is γ_3; i.e., the norm of $(I - A_{22})^{-1}$.

The requirement that $\epsilon/\|y_2^{\mathrm{T}}\|$ be near one may seem awkward, but it is necessary. If y_2^{T} is smaller than ϵ, perturbations due to the interaction of G and E can obliterate it [see the right-hand side of (2.2)]. More insight into this phenomena can be gained by replacing $\|y_2^{\mathrm{T}}\|$ by a lower bound. Since $p_3^{\mathrm{T}}(I - B_{33}) = f_{13}^{\mathrm{T}}$, it follows that $\|p_3\| \le \|f_{13}^{\mathrm{T}}\|/\|I - B_{33}\|$. Hence another, weaker asymptotic bound is

$$\frac{\|\tilde{y}_2^{\mathrm{T}} - y_2^{\mathrm{T}}\|}{\|y_2^{\mathrm{T}}\|} \lesssim \eta\gamma_3\left[1 + (1 + \beta)\frac{\epsilon}{\|f_{13}^{\mathrm{T}}\|}\right]. \tag{4.2}$$

Since β is of order one, we see that the bound can become large when f_{13} to be small compared with the matrix E_{12}.

Finally, we return to the case where $\tilde{A}$ is not stochastic. The problem here is that we have assumed the existence of a null vector for $I - \tilde{A}$ in deriving (2.2). We will circumvent this problem by perturbing $\tilde{A}$ so that $I - \tilde{A}$ is singular.

First note that from Theorem 3.1 and (2.1) we have the following bound:

$$\|y_2^{\mathrm{T}}\| \leq \frac{\gamma_3 \epsilon}{1 - \beta\gamma_2\epsilon};$$

i.e., the near transient states have probability of order ϵ. Since one is a simple eigenvector of A, for η sufficiently small there is a corresponding eigenvalue of $\tilde{A}$ of the form

$$\begin{aligned}\lambda &= 1 + (y_1^{\mathrm{T}}\ y_2^{\mathrm{T}}) \begin{pmatrix} G_{11} & 0 \\ G_{21} & G_{22} \end{pmatrix} \begin{pmatrix} \mathbf{e} \\ \mathbf{e} \end{pmatrix} + O(\eta^2) \\ &= 1 + y_1^{\mathrm{T}} G_{11}\mathbf{e} + O(\eta^2) + O(\epsilon)\end{aligned}$$

(see [3, Theorem IV.2.3]). Hence $\|1 - \lambda\| \leq \|y_1\|\eta + O(\eta^2) + O(\epsilon)$. Thus if $\tilde{\tilde{A}} = \tilde{A} = (1 - \lambda)I$, then $\tilde{\tilde{A}}$ comes from a perturbation of A whose norm is asymptotically bounded by $\eta(1 + \|y_1^{\mathrm{T}}\|)$. Moreover, $I - \tilde{\tilde{A}}$ is exactly singular. Consequently, the asymptotic bounds (4.1) and (4.2) continue to hold with η replaced by $\eta(1 + \|y_1^{\mathrm{T}}\|)$.

Acknowledgement

I would like the referee for many useful comments and particularly for the derivation of (1.1).

References

[1] G. H. Golub and C. F. Van Loan. *Matrix Computations.* Johns Hopkins University Press, Baltimore, Maryland, 2nd edition, 1989.

[2] G. W. Stewart. Perturbation theory for nearly uncoupled Markov chains. In W. J. Stewart, editor, *Numerical Methods for Markov Chains*, pages 105–120, North Holland, Amsterdam, 1990.

[3] G. W. Stewart and G.-J. Sun. *Matrix Perturbation Theory.* Academic Press, Boston, 1990.