SIAM J. MATRIX ANAL. APPL.
Vol. 15, No. 4, pp. 1061–1074, October 1994

002

UNIFORM STABILITY OF MARKOV CHAINS*

ILSE C. F. IPSEN† AND CARL D. MEYER‡

Abstract. By deriving a new set of tight perturbation bounds, it is shown that all stationary probabilities of a finite irreducible Markov chain react essentially in the same way to perturbations in the transition probabilities. In particular, if at least one stationary probability is insensitive in a relative sense, then all stationary probabilities must be insensitive in an absolute sense. New measures of sensitivity are related to more traditional ones, and it is shown that all relevant condition numbers for the Markov chain problem are small multiples of each other. Finally, the implications of these findings to the computation of stationary probabilities by direct methods are discussed, and the results are applied to stability issues in nearly transient chains.

Key words. Markov chains, stationary distribution, stochastic matrix, sensitivity analysis, perturbation theory, stability of a Markov chain, condition numbers

AMS subject classifications. 65U05, 65F35, 60J10, 60J20, 15A51, 15A12, 15A18

1. Introduction.

The purpose of this paper is to analyse the sensitivity of individual stationary probabilities to perturbations in the transition probabilities of finite irreducible Markov chains. In addition to providing perturbation bounds that are much sharper than the traditional bounds, our analysis demonstrates that all stationary probabilities in an irreducible chain react in a somewhat uniform manner to perturbations in the transition probabilities. This property of uniform sensitivity markedly distinguishes Markov problems from general linear systems. Examples are presented in §3 to illustrate why a Markov problem should not be treated as just another linear system.

Previous perturbation theory for irreducible chains focused on the derivation of norm-based bounds of the following kind. Let P and $\tilde{P} = P + E$ be transition probability matrices with respective stationary probability vectors π^T and $\tilde{\pi}^T$ satisfying

$$\pi^T P = \pi^T, \quad \tilde{\pi}^T \tilde{P} = \tilde{\pi}^T, \quad \sum_i \pi_i = 1 = \sum_i \tilde{\pi}_i.$$

For suitable vector and matrix norms, it is known that

$$\|\pi^T - \tilde{\pi}^T\| \leq \kappa \|E\|$$

where values for the condition number κ can be derived in various ways. Schweitzer (1968) derives a value for κ from the fundamental matrix of Kemeny and Snell (1960) whereas the group inverse $A^{\#}$ of $A = I - P$ is used by Meyer (1980), Golub and Meyer (1986), Funderlic and Meyer (1986), Meyer (1994), Meyer and Stewart (1988), Barlow (1993), and Stewart (1991). Seneta (1991) suggests using a coefficient of ergodicity for κ.

*Received by the editors September 24, 1992; accepted for publication (in revised form) July 26, 1993.

†Computer Science, Yale University, New Haven, Connecticut 06520 (ipsen@cs.yale.edu). The work of this author was supported in part by National Science Foundation grant CCR-9102853.

‡Mathematics Department, North Carolina State University, Raleigh, North Carolina 27695–8205 (meyer@math.ncsu.edu). The work of this author was supported in part by National Science Foundation grants DMS-9020915 and DDM-8906248.

These norm-based bounds are not satisfying for two reasons. First, there exist irreducible chains for which the bounds are not tight, so the condition number κ may seriously overestimate the sensitivity to perturbations. Secondly, the bounds generally provide little information about the relative error $|\pi_j - \tilde{\pi}_j|/\pi_j$ in individual stationary probabilities. We remedy this situation in §4 by deriving tight perturbation bounds for individual stationary probabilities. On the basis of these bounds, we prove a uniform sensitivity theorem saying that if at least one stationary probability has low relative sensitivity, or if at least one large stationary probability has low absolute sensitivity, then all probabilities have low absolute sensitivity.

In §5 we relate our measure of sensitivity to the traditional condition numbers for the Markov problem, and we prove that all relevant condition numbers for the problem $\pi^T A = 0$ are small multiples of each other. After discussing the ramifications of the perturbation results on direct methods for computing the stationary probabilities, we consider the case of nearly transient chains in §§6 and 7. We show that under special perturbations even small stationary probabilities may have low relative sensitivity. In addition, we give conditions under which a nearly transient chain is absolutely stable under general perturbations.

2. Norms and notation. Throughout the article the infinity-norm is exclusively used for matrices and column vectors, and the one-norm is used for row vectors. Since it will always be clear from the context whether a quantity is a matrix, column, or row, the subscripts on $\|\star\|_\infty$ and $\|\star\|_1$ are suppressed. Row vectors will always be transposed (e.g., π^T), and column vectors will be untransposed. The jth column of the identity matrix I is denoted by e_j and the column of all ones is denoted by e. The matrix P denotes the transition probability matrix of an n-state irreducible Markov chain with stationary distribution π^T whose entries satisfy $\pi_i > 0$ and $\sum_i \pi_i = 1$. We define $A = I - P$ and $A^{\#}$ denotes the group inverse of A, properties of which can be found in Campbell and Meyer (1991), Meyer (1975), and Meyer (1982). The matrix $\tilde{P} = P + E$ is a perturbation of P that represents the transition matrix of another irreducible chain with stationary distribution $\tilde{\pi}^T$. The perturbation matrix E is not necessarily constrained to be small. We use $E^{(j)}$ to denote the matrix obtained by deleting the jth column of E, and A_j denotes the principal submatrix obtained by deleting the jth row and column from $A = I - P$. Finally, we let N denote the matrix obtained by replacing the last column of A by a column of ones.

3. Absolutely stable chains. The solution of a general ill-conditioned linear system $Ax = b$ need not be uniformly sensitive to small perturbations. Some components of x can be sensitive while others are not. Furthermore, as shown in Chandrasekaran and Ipsen (1992), the sensitivity of the x_i's need not be a result of their size. Our purpose is to demonstrate that this cannot happen for Markov chains, but first it is important to distinguish between absolute sensitivity and relative sensitivity in the Markov chain setting.

Example 3.1. For the three-state chain whose transition matrix is

$$P(\epsilon) = \begin{pmatrix} 0 & 1-\epsilon & \epsilon \\ 1-\epsilon & 0 & \epsilon \\ 1 & 0 & 0 \end{pmatrix},$$

the associated stationary distribution is

$$\pi^T(\epsilon) = \left(\frac{1}{(2-\epsilon)(1+\epsilon)}, \ \frac{1-\epsilon}{(2-\epsilon)(1+\epsilon)}, \ \frac{\epsilon}{1+\epsilon} \right).$$

If $P = P(10^{-8})$ is perturbed to become $\tilde{P} = P(10^{-4})$, then the magnitude of the perturbation $E = \tilde{P} - P$ is

$$\|E\| = 2(10^{-4} - 10^{-8}).$$

Consider the change in the respective stationary distributions

$$\pi^T = \pi^T(10^{-8}) \quad \text{and} \quad \tilde{\pi}^T = \pi^T(10^{-4}).$$

The *absolute change* (the change relative to 1) in π_3 is

$$|\pi_3 - \tilde{\pi}_3| = \left| \frac{10^{-8}}{1+10^{-8}} - \frac{10^{-4}}{1+10^{-4}} \right| = \frac{10^{-4} - 10^{-8}}{(1+10^{-4})(1+10^{-8})} \approx 10^{-4} - 10^{-8} = \frac{\|E\|}{2},$$

but the *relative change* (the change relative to the original value) is

$$\left| \frac{\pi_3 - \tilde{\pi}_3}{\pi_3} \right| = \left| 1 - \frac{10^{-4}(1+10^{-8})}{10^{-8}(1+10^{-4})} \right| \approx 10^4.$$

If the change in probabilities is assessed in an absolute sense by comparing it to 1, then π_3 is not at all sensitive to the perturbation because the change of magnitude $\|E\|$ in the transition probabilities produces a change in π_3 of only $\|E\|/2$. We say that π_3 is *absolutely insensitive.* But if the change in probabilities is assessed in a relative sense then the change in π_3 is large, so π_3 is *relatively sensitive.* As for the sensitivity of the other two probabilities π_1 and π_2, if $a^{\#}_{ij}$ is element (i, j) in the group inverse $A^{\#}$ of $A = I - P$, then, as shown by Funderlic and Meyer (1986), the absolute error in the jth stationary probability is bounded by

$$|\pi_j - \tilde{\pi}_j| \le \kappa_j \|E\|, \qquad \kappa_j = \max_i |a^{\#}_{ij}|.$$

In this example, $\max_{i,j} |a^{\#}_{ij}| < 1$, so all three stationary probabilities are insensitive in the absolute sense. Because π_1 and π_2 are both very close to .5, they are insensitive in the relative sense as well. This example motivates the following definition.

DEFINITION 3.1. *An irreducible chain is said to be* absolutely stable *whenever each* π_j *is insensitive to perturbations in* P *in the absolute sense; i.e., whenever there is a small constant* κ *such that for all perturbations* E,

$$|\pi_j - \tilde{\pi}_j| \le \kappa \|E\| \quad \textit{for each } 1 \le j \le n,$$

where the term "small" is to be interpreted in the context of the underlying application.

Sufficient conditions for absolute stability are well-known. The results in Barlow (1993), Funderlic and Meyer (1986), Golub and Meyer (1986), Meyer (1980), Meyer (1994), Meyer and Stewart (1988), Stewart (1991), for instance, use the fact that a chain is absolutely stable if the group inverse $A^{\#}$ of $A = I - P$ has no large entries (relative to 1). The results of §5 will establish that the converse of this statement is also true.

4. Componentwise analysis.

In this section we derive tight upper bounds on the relative change in individual stationary probabilities, and we prove that all stationary probabilities show essentially the same sensitivity to perturbations in the transition probabilities.

We make use of the following properties of M–matrices, details of which can be found in the text by Berman and Plemmons (1979). If P is an irreducible stochastic matrix of order n, then $A = I - P$ is a singular M–matrix of rank $n-1$. Moreover, if A_j is the principal submatrix of A obtained by deleting the jth row and column from A, then A_j is a nonsingular M–matrix. Hence $A_j^{-1} > 0$, and if e is the column vector of all ones, then $\|A_j^{-1}e\| = \|A_j^{-1}\|$. The following theorem demonstrates that the entries in A_j^{-1} determine the relative sensitivity of the jth stationary probability to perturbations in the transition probabilities.

THEOREM 4.1. *If $E^{(j)}$ denotes the matrix obtained by deleting the jth column of E, then*

$$\frac{\pi_j - \tilde{\pi}_j}{\pi_j} = \tilde{\pi}^T E^{(j)} A_j^{-1} e.$$

Furthermore,

$$\left|\frac{\pi_j - \tilde{\pi}_j}{\pi_j}\right| \leq \|E^{(j)}\|\,\|A_j^{-1}\|,$$

and there always exists a perturbation E (dependent on j) for which equality is attained.

Proof. By applying a symmetric permutation to P, the states may be reordered so that a particular stationary probability occurs in the last position of π^T. Thus it suffices to prove the theorem for $j = n$. With the partitioning

$$\pi^T = (\,\bar{\pi}^T \quad \pi_n\,), \qquad A = \begin{pmatrix} A_n & b \\ c^T & \delta \end{pmatrix},$$

$\pi^T A = 0^T$ implies $\bar{\pi}^T = -\pi_n c^T A_n^{-1}$. Replacing the last column of A by the vector of all ones produces a nonsingular matrix

$$N = \begin{pmatrix} A_n & e \\ c^T & 1 \end{pmatrix} \quad \text{with inverse} \quad N^{-1} = \begin{pmatrix} A_n^{-1}(I - e\bar{\pi}^T) & -\pi_n A_n^{-1} e \\ \bar{\pi}^T & \pi_n \end{pmatrix}.$$

The stationary distribution of the original chain is the solution of the system

$$\pi^T N = e_n^T \quad \text{where} \quad e_n^T = (\,0 \quad \cdots \quad 0 \quad 1\,),$$

and the stationary distribution for the perturbed chain is the solution of

$$\tilde{\pi}^T (N - F) = e_n^T \quad \text{where} \quad F = (\,E^{(n)} \quad 0\,).$$

Consequently,

$$\pi^T - \tilde{\pi}^T = -\tilde{\pi}^T F N^{-1}, \tag{4.1}$$

so

$$\pi_n - \tilde{\pi}_n = -\tilde{\pi}^T \Big(E^{(n)} \ \ 0\Big) \begin{pmatrix} -\pi_n A_n^{-1} e \\ \pi_n \end{pmatrix} = \pi_n\Big(\tilde{\pi}^T E^{(n)} A_n^{-1} e\Big),$$

and therefore

$$\frac{\pi_n - \tilde{\pi}_n}{\pi_n} = \tilde{\pi}^T E^{(n)} A_n^{-1} e.$$

Applying Hölder's inequality and $\|A_n^{-1} e\| = \|A_n^{-1}\|$ yields

$$\left|\frac{\pi_n - \tilde{\pi}_n}{\pi_n}\right| \leq \|\tilde{\pi}\| \, \|E^{(n)} A_n^{-1} e\| \leq \|E^{(n)}\| \, \|A_n^{-1}\|.$$

To see that equality is always attainable, let k be the position where the largest component of $A_n^{-1} e$ occurs so that

$$e_k^T A_n^{-1} e = \|A_n^{-1} e\| = \|A_n^{-1}\|,$$

and let $E = \epsilon e (e_k - e_n)^T$. Then $\tilde{\pi}^T E^{(n)} = \epsilon e_k^T$ and $\|E^{(n)}\| = \epsilon$, so that

$$\frac{\pi_n - \tilde{\pi}_n}{\pi_n} = \tilde{\pi}^T E^{(n)} A_n^{-1} e = \epsilon e_k^T A_n^{-1} e = \epsilon \|A_n^{-1}\| = \|E^{(n)}\| \, \|A_n^{-1}\|. \qquad \square$$

COROLLARY 4.1. *An irreducible chain is absolutely stable if and only if* $\pi_j \|A_j^{-1}\|$ *is small for every* $1 \leq j \leq n$.

The results of Theorem 4.1 and its corollary suggest the following definitions.

DEFINITION 4.1. *Let* A_j *be the principal submatrix obtained by deleting the* j*th row and column from* A, *and let* π_j *denote the* j*th stationary probability. The* relative condition number *for* π_j *is defined to be*

$$\rho_j = \|A_j^{-1}\| \quad \textit{and we set} \quad \rho = \min_j \{\rho_j\}.$$

The absolute condition number *for* π_j *is defined to be*

$$\alpha_j = \pi_j \|A_j^{-1}\| \quad \textit{and we set} \quad \alpha = \max_j \{\alpha_j\}.$$

In terms of this notation, Theorem 4.1 states

$$\left|\frac{\pi_j - \tilde{\pi}_j}{\pi_j}\right| \leq \rho_j \, \|E^{(j)}\|, \quad |\pi_j - \tilde{\pi}_j| \leq \alpha_j \, \|E\|, \quad \text{and} \quad \|\pi - \tilde{\pi}\| \leq \alpha \|E\|,$$

so α is the *absolute condition number for the entire chain.*

Notice that if a stationary probability is relatively well-conditioned, then it is absolutely well-conditioned but not conversely, cf., Example 3.1. It may be of interest to note that the existence of a small ρ_j means that the $(n-1)$st singular value of A is large (Barlow (1993)).

We now arrive at one of our principal conclusions which states that the sensitivity of the stationary distribution is uniform in the sense that all π_j 's are absolutely well-conditioned if and only if at least one π_j is relatively well-conditioned.

THEOREM 4.2. *For every* $1 \leq j \leq n$,

$$|\pi_j - \tilde{\pi}_j| \leq \rho \, \|E\|.$$

Consequently, an irreducible chain is absolutely stable if and only if at least one π_j *is relatively well-conditioned.*

Proof. As in the proof of Theorem 4.1, assume that the states have been permuted so the best relatively conditioned stationary probability is in the last position; i.e., $\rho_n = \rho$. If

$$N = \begin{pmatrix} A_n & e \\ c^T & 1 \end{pmatrix}$$

is the matrix obtained by replacing the last column of A by ones then, as in (4.1),

$$\pi^T - \tilde{\pi}^T = -\tilde{\pi}^T F N^{-1}, \quad \text{where} \quad F = (\, E^{(n)} \quad 0\,).$$

From

$$N^{-1} = \begin{pmatrix} A_n^{-1}(I - e\overline{\pi}^T) & -\pi_n A_n^{-1} e \\ \overline{\pi}^T & \pi_n \end{pmatrix} = \begin{pmatrix} A_n^{-1} & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} I - e\overline{\pi}^T & -\pi_n e \\ \overline{\pi}^T & \pi_n \end{pmatrix}, \tag{4.2}$$

it follows that

$$\pi_j - \tilde{\pi}_j = -\tilde{\pi}^T F N^{-1} e_j = \begin{cases} -\tilde{\pi}^T E^{(n)} A_n^{-1}(e_j - \pi_j e) & \text{if} \quad j < n. \\ -\tilde{\pi}^T E^{(n)} A_n^{-1}(-\pi_j e) & \text{if} \quad j = n. \end{cases}$$

Since $\|e_j - \pi_j e\| = \max\{\pi_j,\, 1 - \pi_j\} < 1$ and $\|A_n^{-1} e\| = \|A_n^{-1}\| = \rho_n$, we have that

$$|\pi_j - \tilde{\pi}_j| \le \rho_n \|E^{(n)}\| \le \rho_n \|E\|, \qquad 1 \le j \le n.$$

Therefore, if at least one stationary probability is relatively well-conditioned, then all stationary probabilities are absolutely well-conditioned. The converse follows from Corollary 4.1 because at least one π_j must be greater than or equal to $1/n$. □

The following two statements are direct consequences of Theorem 4.2, but they are important to state because they drive home the extent to which there exists uniform stability in Markov chains.

COROLLARY 4.2. *If any stationary probability is relatively well-conditioned, then all large stationary probabilities are relatively well-conditioned.*

COROLLARY 4.3. *If any large stationary probability is absolutely well-conditioned, then the chain is absolutely stable.*

A natural question arises at this point. We know that the existence of one relatively well-conditioned π_j implies the chain is absolutely stable, but does the existence of one absolutely well-conditioned π_j insure absolute stability? Unfortunately, the answer is "no," and this can be seen by considering

$$P = \begin{pmatrix} 1-\epsilon & \epsilon/2 & \epsilon/2 \\ \epsilon/2 & 1-\epsilon & \epsilon/2 \\ 1/2 & 1/2 & 0 \end{pmatrix}, \qquad \pi^T = \frac{1}{2+\epsilon}(\,1 \quad 1 \quad \epsilon\,)$$

for small $0 < \epsilon < 1$. The absolute and relative condition numbers are

$$\alpha_1 = \alpha_2 = \frac{1}{2+\epsilon}\left(\frac{2}{3} + \frac{4}{3\epsilon}\right), \quad \alpha_3 = \frac{2}{2+\epsilon}, \quad \rho_1 = \rho_2 = \frac{2}{3} + \frac{4}{3\epsilon}, \quad \text{and} \quad \rho_3 = \frac{2}{\epsilon},$$

so, for small ϵ, π_3 is absolutely well-conditioned, but π_1 and π_2 are not. The chain is not absolutely stable, and no π_j is relatively well-conditioned.

Small stationary probabilities are the ones that appear least likely to be relatively well-conditioned. Therefore it makes sense to try to determine features that may be responsible for the small size. The following theorem shows that those π_j whose associated submatrix A_j is well-conditioned cannot be small. It also shows that a nearly reducible matrix A that is far from being uncoupled produces small π_j.

THEOREM 4.3. *If Q is a permutation matrix such that*

$$Q^T A Q = \begin{pmatrix} A_j & b_j \\ c_j^T & \delta_j \end{pmatrix},$$

then

$$\frac{1}{1+\rho_j} \leq \pi_j \leq \frac{\|b_j\|}{\|c_j^T\| + \|b_j\|}.$$

Proof. Let $\pi^T Q = \psi^T = \begin{pmatrix} \overline{\psi}^T & \pi_j \end{pmatrix}$. Since $\psi^T A = 0$ implies $\overline{\psi}^T = -\pi_j c_j^T A_j^{-1}$, Hölder's inequality gives the lower bound

$$1 - \pi_j = \overline{\psi}^T e = \pi_j |c_j^T A_j^{-1} e| \leq \pi_j \rho_j.$$

To obtain the upper bound, use $\|c_j^T\| = -c_j^T e = \delta_j$ and $\delta_j \pi_j = -\overline{\psi}^T b_j$, and again apply Hölder's inequality,

$$\pi_j \|c_j^T\| = \pi_j \delta_j = -\overline{\psi}^T b_j \leq \|\overline{\psi}^T\| \, \|b_j\| = (1 - \pi_j)\|b_j\|. \qquad \square$$

5. Condition numbers and linear systems. It was demonstrated in the previous section that the sensitivity of the stationary distribution is governed by ρ. We now compare this measure of sensitivity to other condition numbers, and we relate these results to numerical techniques for computing stationary probabilities by solving certain linear systems.

The nonsingular matrix

$$N = \begin{pmatrix} A_n & e \\ c^T & 1 \end{pmatrix} \quad \text{and the associated system} \quad \pi^T N = e_n^T$$

are focal points of the development. The expression (4.2) together with the fact that $\rho_n = \|A_n^{-1}\| \geq 1$ (because $e = -A_n^{-1} b$) produces

$$1 \leq \|N^{-1}\| \leq 2\rho_n. \tag{5.1}$$

This means that if π_n is relatively well-conditioned, then $\pi^T N = e_n^T$ is a well-conditioned nonsingular system and therefore any stable algorithm can accurately solve it. But it is not clear that the solution of $\pi^T N = e_n^T$ should be attempted when ρ_n is large, even if the chain is absolutely stable. Theorem 4.1 insures that some ρ_j must be small, but, as Example 3.1 demonstrates, it need not be ρ_n. Of course, safety can be guaranteed if one is willing to determine a value of k such that $\rho_k = \rho$ because the same logic that produced (5.1) insures that the system $\pi^T \hat{N} = e_k^T$ is well-conditioned where $\hat{N}$ is the nonsingular matrix obtained by replacing the kth column of A by e. But determining ρ (or its position) is prohibitively expensive, and this may be why this approach is dismissed as "naive" by Paige, Styan, and Wachter (1975) and not included in their comparisons.

Surprisingly, it does not matter which column of A is replaced by e. This is a consequence of the next theorem that relates N^{-1} to the group inverse $A^{\#}$.

THEOREM 5.1. *For the numbers* α *and* ρ *given in Definition* 4.1,

$$\frac{\|A^{\#}\|}{2} \leq \|N^{-1}\| \leq 2\,\|A^{\#}\| + 1$$

and

$$\frac{\alpha}{2} \leq \|A^{\#}\| \leq 4\rho.$$

Proof. We derive the upper bounds first. It is easily verified (Meyer (1975)) that

$$\begin{aligned} A^{\#} &= (I - e\pi^T)\begin{pmatrix} A_n^{-1} & 0 \\ 0 & 0 \end{pmatrix}(I - e\pi^T) \\ &= \begin{pmatrix} (I - e\overline{\pi}^T)A_n^{-1}(I - e\overline{\pi}^T) & -\pi_n(I - e\overline{\pi}^T)A_n^{-1}e \\ -\overline{\pi}^T A_n^{-1}(I - e\overline{\pi}^T) & \pi_n\overline{\pi}^T A_n^{-1}e \end{pmatrix}. \end{aligned} \tag{5.2}$$

A symmetric permutation can bring any principal submatrix A_j of A to the upper left-hand corner of Q^TAQ. Then $(Q^TAQ)^{\#} = Q^TA^{\#}Q$, and

$$\psi^T = \pi^T Q = (\,\overline{\psi}^T \quad \pi_j\,)$$

imply

$$\begin{aligned} Q^TA^{\#}Q &= (I - e\psi^T)\begin{pmatrix} A_j^{-1} & 0 \\ 0 & 0 \end{pmatrix}(I - e\psi^T) \\ &= \begin{pmatrix} (I - e\overline{\psi}^T)A_j^{-1}(I - e\overline{\psi}^T) & -\pi_j(I - e\overline{\psi}^T)A_j^{-1}e \\ -\overline{\psi}^T A_j^{-1}(I - e\overline{\psi}^T) & \pi_j\overline{\psi}^T A_j^{-1}e \end{pmatrix}. \end{aligned}$$

The second upper bound is now immediate because

$$\|A^{\#}\| = \|Q^TA^{\#}Q\| \leq 4\rho_j \quad \text{for all } j.$$

The first upper bound follows from

$$N^{-1} = \begin{pmatrix} I & -e \\ -c^T & -\delta \end{pmatrix} A^{\#} + e_n e_n^T,$$

which can be verified by using (5.2), so that $\|N^{-1}\| \leq 2\|A^{\#}\| + 1$. To establish the lower bounds, use the expressions for $A^{\#}$ and $Q^TA^{\#}Q$ to write

$$A^{\#} = \begin{pmatrix} I - e\overline{\pi}^T & 0 \\ -\overline{\pi}^T & 0 \end{pmatrix} N^{-1} \quad \text{and} \quad \pi_j A_j^{-1} = (\,I \quad -e\,)\; Q^TA^{\#}Q \begin{pmatrix} \pi_j I \\ -\overline{\psi}^T \end{pmatrix}.$$

Hence $\|A^{\#}\| \leq 2\|N^{-1}\|$ and, for every j,

$$\alpha_j = \pi_j\|A_j^{-1}\| \leq 2\|A^{\#}\|. \qquad \square$$

The group inverse is relevant because

$$\pi - \tilde{\pi} = \tilde{\pi}EA^{\#} \quad \text{and} \quad |\pi_j - \tilde{\pi}_j| \leq \|E\|\,\|A^{\#}\|, \tag{5.3}$$

(Meyer (1980)), so if $\|A^{\#}\|$ is small, then the chain is absolutely stable. While conjectured, the converse of this statement has never been proven. However, on the basis of Theorems 4.2 and 5.1, the converse is now evident.

The logic used in proving Theorem 5.1 dictates that replacing *any* column of A by e results in a well-conditioned matrix when the chain is absolutely stable, and when the chain is not absolutely stable, all such matrices are ill-conditioned. Consequently, it does not matter which column of A is replaced by 1's, so the problem addressed by Harrod and Plemmons (1984) and Barlow (1986, 1993) of having to locate a well-conditioned principal submatrix A_j in order to build a well-conditioned system is obviated. Furthermore, since N is nonsingular, standard numerical techniques[1] can be applied to solve $\pi^T N = e_n^T$.

So far we have viewed the stationary distribution π^T as a solution to two different linear systems; the singular system $\pi^T A = 0$ and the nonsingular system $\pi^T N = e_n^T$. There is a yet a third linear system of which π^T is a solution, namely

$$\pi^T M = e_{n+1}^T \quad \text{where} \quad M = (A \quad e). \tag{5.4}$$

The augmented matrix M is of order $n \times (n+1)$ and has full row rank. The perturbed system is $\tilde{\pi}^T(M+E) = e_{n+1}^T$, so

$$\pi^T - \tilde{\pi}^T = \tilde{\pi}^T E M^{\dagger},$$

where $M^{\dagger}$ is the Moore-Penrose pseudo-inverse (Campbell and Meyer (1991)), and

$$|\pi_j - \tilde{\pi}_j| \le \|\tilde{\pi}\| \, \|E\| \, \|M^{\dagger}\| \quad \text{for each } j.$$

Hence $\|M^{\dagger}\|$ is a condition number for measuring absolute stability. Another such number is $\|Z\|$ where $Z = (A + e\pi^T)^{-1}$ is the Kemeny and Snell (1960) fundamental matrix because $\pi^T - \tilde{\pi}^T = \tilde{\pi}^T E Z$ and $|\pi_j - \tilde{\pi}_j| \le \|\tilde{\pi}\| \, \|E\| \, \|Z\|$ (Schweitzer (1968)). The following lemma shows that $\|M^{\dagger}\|$ and $\|Z\|$ are small multiples of each other and that they are not significantly different from $\|A^{\#}\|$.

LEMMA 5.1. *For the matrices M and Z defined above,*

$$\frac{\|Z\|}{3} \le \|M^{\dagger}\| \le 2\,\|Z\| \quad \text{and} \quad \|A^{\#}\| - 1 \le \|Z\| \le \|A^{\#}\| + 1.$$

Proof. The first set of inequalities follows from the identities

$$Z = \Big((I - e\pi^T) \;\; e \Big) M^{\dagger} \quad \text{and} \quad M^{\dagger} = \begin{pmatrix} I - ee^T/n \\ \pi^T \end{pmatrix} Z,$$

each of which is straightforward to verify. The second set of inequalities is a consequence of the fact that $Z = A^{\#} + e\pi^T$ (Meyer (1975)). □

Combining the results of Lemma 5.1 with those of Theorems 4.2 and 5.1 produces the following complete statement concerning the stability of irreducible Markov chains.

[1] Gaussian elimination with exact arithmetic generates positive pivots, but floating-point arithmetic may produce a zero or negative pivot (Funderlic and Mankin (1981)). This can be avoided with diagonal adjustment schemes as discussed by Grassmann, Taksar, and Heyman (1985), Stewart and Zhang (1991), and Barlow (1993).

THEOREM 5.2. *For an n-state irreducible Markov chain, the following statements are equivalent.*

- *At least one stationary probability is relatively well-conditioned.*
- *The chain is absolutely stable.*
- *All entries of the group inverse $A^{\#}$ are small.*
- *The matrix N and the system $\pi^T N = e_n^T$ are well-conditioned.*
- *The matrix M and the system $\pi^T M = e_{n+1}^T$ are well-conditioned.*
- *All entries in the Kemeny and Snell fundamental matrix Z are small.*

6. Sensitivity of nearly transient chains. In this section we examine the sensitivity of stationary probabilities of irreducible chains with nearly transient states; i.e., irreducible chains in which the states can be ordered so that the transition matrix is almost block triangular in the sense that

$$P = \begin{pmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{pmatrix} \quad \text{with} \quad \|P_{21}\| << 1. \tag{6.1}$$

We prove two results, one for structured perturbations and one for more general perturbations.

The first theorem establishes a result similar to the one by Stewart (1992b). It says that *small* stationary probabilities of an absolutely stable chain are *relatively* well-conditioned if only the states corresponding to these probabilities are perturbed and all other states remain unaffected.

THEOREM 6.1. *If E can be symmetrically permuted so that*

$$E = \begin{pmatrix} 0 \\ E_2 \end{pmatrix} \quad \textit{with} \quad \|E\| = \epsilon,$$

and if $\pi^T = \begin{pmatrix} \pi_1^T & \pi_2^T \end{pmatrix}$ is partitioned conformably, then

$$\frac{|\pi_j - \tilde{\pi}_j|}{\|\tilde{\pi}_2\|} \leq 4\epsilon\, \rho, \qquad 1 \leq j \leq n.$$

Proof. Combine (5.3) with the fact $\|A^{\#}\| \leq 4\rho$ from Theorem 5.1. □

The second theorem concerns nearly transient chains, but no restriction is placed on the structure of the perturbation matrix.

THEOREM 6.2. *Suppose P_{11} in (6.1) is $s \times s$, and let*

$$A = \begin{pmatrix} A_n & b \\ c^T & \delta \end{pmatrix}, \quad A_n = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}, \quad \textit{and} \quad b = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix},$$

where B_{11} is $s \times s$, b_1 is $s \times 1$, and $b_i \neq 0$ for each i. The relative condition of π_n is bounded by

$$\rho_n < \frac{2 \max\left\{\|B_{11}^{-1}\|,\, \|B_{22}^{-1}\|\right\}}{1 - \|B_{22}^{-1} B_{21}\|},$$

so the chain is absolutely stable whenever B_{11} and B_{22} have small inverses.

Proof.

$$A_n = \begin{pmatrix} B_{11} & B_{12} \\ 0 & B_{22} \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ B_{21} & 0 \end{pmatrix} = T + K = T(I + T^{-1}K).$$

If $\|T^{-1}K\| < 1$, then, from results in §2.3.4 in Golub and Van Loan (1989),

$$\rho_n = \|A_n^{-1}\| \le \|T^{-1}\| \, \|(I + T^{-1}K)^{-1}\| \le \frac{\|T^{-1}\|}{1 - \|T^{-1}K\|}.$$

Since A is an M–matrix, $B_{ii}^{-1} > 0$, $B_{ij} \le 0$, and $b \le 0$. Consequently, $Ae = 0$ implies

$$0 \le -B_{11}^{-1}B_{12}e = e + B_{11}^{-1}b_1 \le e.$$

By assumption, $b_1 \ne 0$, so $B_{11}^{-1}b_1 < 0$, and thus

$$\|B_{11}^{-1}B_{12}\| = \|B_{11}^{-1}B_{12}e\| < 1.$$

A similar argument shows that $\|B_{22}^{-1}B_{21}\| < 1$. Since

$$T^{-1} = \begin{pmatrix} B_{11}^{-1} & -B_{11}^{-1}B_{12}B_{22}^{-1} \\ 0 & B_{22}^{-1} \end{pmatrix} = \begin{pmatrix} I & -B_{11}^{-1}B_{12} \\ 0 & I \end{pmatrix} \begin{pmatrix} B_{11}^{-1} & 0 \\ 0 & B_{22}^{-1} \end{pmatrix},$$

we have

$$\|T^{-1}\| \le (1 + \|B_{11}^{-1}B_{12}\|) \max\{\|B_{11}^{-1}\|, \|B_{22}^{-1}\|\} < 2\max\{\|B_{11}^{-1}\|, \|B_{22}^{-1}\|\}.$$

Similarly,

$$T^{-1}K = \begin{pmatrix} -B_{11}^{-1}B_{12}B_{22}^{-1}B_{21} & 0 \\ B_{22}^{-1}B_{21} & 0 \end{pmatrix}$$

implies

$$\|T^{-1}K\| \le \max\{\|B_{11}^{-1}B_{12}B_{22}^{-1}B_{21}\|, \|B_{22}^{-1}B_{21}\|\} < \|B_{22}^{-1}B_{21}\|,$$

so

$$\rho_n \le \frac{\|T^{-1}\|}{1 - \|T^{-1}K\|} < \frac{2\max\{\|B_{11}^{-1}\|, \|B_{22}^{-1}\|\}}{1 - \|B_{22}^{-1}B_{21}\|}. \qquad \square$$

7. Small probabilities in nearly transient chains. Let $\pi^T = (\bar{\pi}_1^T \;\; \bar{\pi}_2^T)$ be the stationary distribution of the nearly transient matrix P in (6.1), and set

$$A = I - P = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}, \quad \text{where} \quad \|A_{21}\| = \|P_{21}\| = \epsilon.$$

Since $\bar{\pi}_1^T = -\bar{\pi}_2^T A_{21}A_{11}^{-1}$ implies $\|\bar{\pi}_1^T\| \le \epsilon\|A_{11}^{-1}\|$, we see that the trailing stationary probabilities dominate the leading ones provided $\|A_{11}^{-1}\|$ is not too large. For nearly transient chains with a finer block structure, say

$$(7.1) \quad A = \begin{pmatrix} A_{11} & A_{12} & A_{13} & \cdots & A_{1k} \\ F_{21} & A_{22} & A_{23} & \cdots & A_{2k} \\ F_{31} & F_{32} & A_{33} & \cdots & A_{3k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ F_{k1} & F_{k2} & F_{k3} & \cdots & A_{kk} \end{pmatrix}, \quad \pi = \begin{pmatrix} \bar{\pi}_1 \\ \bar{\pi}_2 \\ \bar{\pi}_3 \\ \vdots \\ \bar{\pi}_k \end{pmatrix}, \quad \left\| \begin{pmatrix} F_{j+1,j} \\ F_{j+2,j} \\ F_{j+3,j} \\ \vdots \\ F_{k,j} \end{pmatrix} \right\| = \epsilon_j,$$

$1 \leq j \leq k-1$, the same should be true; i.e., the trailing stationary probabilities tend to be larger than the leading ones. We will quantify this statement by providing bounds in terms of ϵ_j on the probabilities $\bar{\pi}_j$ associated with each block.

The strategy is to proceed inductively by applying the above 2×2 case to successive diagonal blocks. This is accomplished by applying the following lemma that provides a perturbation of size ϵ that essentially uncouples A_{11} from the remaining blocks. In particular, the lemma shows that the remaining probabilities are the exact probabilities of a perturbed problem of the same form (the only difference being that the sum of the probabilities is less than one).

LEMMA 7.1. *Let*

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

with $\|A_{21}\| \leq \epsilon$. *If*

$$\Delta = \frac{A_{21} e \bar{\pi}_1^T A_{12}}{\bar{\pi}_2^T A_{21} e},$$

then $A_{22} + \Delta$ *is a singular M–matrix such that*

$$\bar{\pi}_2^T (A_{22} + \Delta) = 0, \quad (A_{22} + \Delta) e = 0, \quad \textit{and} \quad \|\Delta\| \leq \epsilon.$$

Proof. We first verify that Δ satisfies the required equations. From $\pi^T A = 0$ and $Ae = 0$ we get $r_1^T = \bar{\pi}_2^T A_{22} = -\bar{\pi}_1^T A_{12}$ and $r_2 = A_{22} e = -A_{21} e$, so one can write

$$\Delta = -\frac{r_2 r_1^T}{\bar{\pi}_2^T r_2}.$$

Since $\bar{\pi}_2^T \Delta = -r_1^T$, it follows that $\bar{\pi}_2^T (A_{22} + \Delta) = 0$, and thus Δ satisfies the first equation. To prove that Δ satisfies the second equation, observe that $\pi^T A = 0$ and $Ae = 0$ imply

$$\bar{\pi}_2^T r_2 = -\bar{\pi}_2^T A_{21} e = \bar{\pi}_1^T A_{11} e = -\bar{\pi}_2^T A_{12} e = r_1^T e.$$

Thus,

$$\Delta = -\frac{r_2 r_1^T}{r_1^T e},$$

so $\Delta e = -r_2$ and $(A_{22} + \Delta) e = 0$. As for the bound on the norm of Δ, notice that r_1 and r_2 both consist entirely of nonnegative elements since A is an M–matrix so Δ consists entirely of nonpositive elements. This means

$$\|\Delta\| = \left\| \frac{r_2 r_1^T}{r_1^T e} e \right\| = \|r_2\| \leq \epsilon.$$

Moreover, since all elements of Δ are nonpositive, the off-diagonal elements in $A_{22}+\Delta$ are more negative than those of A_{22}. This implies with $(A_{22} + \Delta) e = 0$ that the diagonal elements must be nonnegative. From $\pi > 0$ it follows that $A_{22} + \Delta$ must be irreducible, for otherwise a component of $\bar{\pi}_2$ would be zero. According to Corollary 1 in §3.5 of Varga (1962), the signs of the matrix elements and the irreducibility imply that every principal submatrix of $A_{22} + \Delta$ is an M–matrix. Therefore, $A_{22} + \Delta$ is a singular M–matrix. □

Now we can prove the following theorem that says that in a nearly transient chain, the size of the π_i in the jth block is controlled by the smallness of the preceding off-diagonal columns $1, \ldots, j-1$, and by the condition of a perturbed jth diagonal block. The size of this perturbation is again determined by the smallness of the off-diagonal columns $1, \ldots, j-1$. This implies that the trailing solution components tend to be larger than the leading ones.

THEOREM 7.1. *If A is partitioned as indicated in* (7.1), *then* $\|\bar{\pi}_1^T\| \leq \epsilon_1\kappa_1$ *with* $\kappa_1 = \|A_{11}^{-1}\|$. *Furthermore, there exist matrices* $X_{j+1,j+1}$ *such that*

$$\|A_{j+1,j+1} - X_{j+1,j+1}\| \leq \epsilon_1 + \cdots + \epsilon_j, \quad 1 \leq j \leq k-2,$$

and

$$\|\bar{\pi}_{j+1}^T\| \leq (\epsilon_1 + \cdots + \epsilon_{j+1})\kappa_{j+1}, \quad \textit{where} \quad \kappa_{j+1} = \|X_{j+1}^{-1}\|.$$

Proof. The statements for $\bar{\pi}_1$ follow from the 2×2 block partitioning. Now apply the same argument recursively to the matrix

$$\bar{A}_{22} = \begin{pmatrix} A_{22} & * & \dots & * & * \\ F_{32} & A_{33} & \dots & * & * \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \vdots & \vdots & & A_{k-1,k-1} & * \\ F_{k2} & F_{k3} & \dots & F_{k,k-1} & A_{kk} \end{pmatrix} + \Delta,$$

where Δ is given by Lemma 7.1. For instance, X_2 is the leading diagonal block of $\bar{A}_{22}$ with $\|X_2^{-1}\| = \kappa_2$. Lemma 7.1 insures $\|\Delta\| \leq \epsilon_1$ and $\|A_{22} - X_2\| \leq \Delta \leq \epsilon_1$. Since the norm of the first off-diagonal column is bounded above by $\epsilon_1 + \epsilon_2$, Lemma 7.1 gives $\|\bar{\pi}_2^T\| \leq (\epsilon_1 + \epsilon_2)\kappa_2$. □

8. Concluding remarks.

Our goal was to better understand how individual stationary probabilities are affected by unstructured perturbations to the transition probabilities. Consequently, we measured all perturbations relative to 1 rather than relative to $A = I - P$ or relative to the structure of P. In other words, we measured the magnitude of a perturbation by $\|E\|/\|P\| = \|E\|$ instead of $\|E\|/\|A\|$ or $\max_{ij} |e_{ij}|/p_{ij}$. The latter two measures result in significantly different interpretations of sensitivity. For example, perturbations that are small relative to 1 can greatly affect the stationary probabilities of

$$P = \begin{pmatrix} 1-\epsilon & \epsilon \\ \epsilon & 1-\epsilon \end{pmatrix}, \qquad \epsilon << 1,$$

but measured relative to $\|A\| = 2\epsilon$, or measured by $\max_{ij} |e_{ij}|/p_{ij}$, small perturbations cannot have a drastic effect (Meyer (1980), O'Cinneide (1993), and Zhang (1993)).

REFERENCES

J. L. BARLOW (1986), *On the smallest positive singular value of a singular M–matrix with applications to ergodic Markov chains*, SIAM J. Algebraic Discrete Meth., 7, pp. 414–424.

——— (1993), *Error bounds for the computation of null vectors with applications to Markov chains*, SIAM J. Matrix Anal. Appl., 14, pp. 598–618.

A. BERMAN AND R. J. PLEMMONS (1979), *Nonnegative Matrices in the Mathematical Sciences*, Academic Press, New York.

S. L. CAMPBELL AND C. D. MEYER (1991), *Generalized Inverses of Linear Transformations*, Dover Publications (1979 edition by Pitman Pub. Ltd., London), New York.

S. CHANDRASEKARAN AND I. IPSEN (1992), *On the sensitivity of solution components in linear systems of equations*, Computer Science Technical Report, Yale University, New Haven, CT; SIAM J. Matrix Anal. Appl., to appear.

R. E. FUNDERLIC AND J. MANKIN (1981), *Solution of homogeneous systems of linear equations arising from compartmental models*, SIAM J. Sci. Statist. Comput., 2, pp. 375–383.

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

G. H. Golub and C. D. Meyer (1986), *Using the QR factorization and group inversion to compute, differentiate, and estimate the sensitivity of stationary probabilities for Markov chains*, SIAM J. Algebraic Discrete Meth., 7, pp. 273–281.

G. H. Golub and C. F. Van Loan (1989), *Matrix Computations, Second Ed.*, The Johns Hopkins Press, Baltimore.

W. K. Grassmann, M. I. Taksar, and D. P. Heyman (1985), *Regenerative analysis and steady state distributions for Markov chains*, Oper. Res., 33, pp. 1107–1116.

W. J. Harrod and R. J. Plemmons (1984), *Comparison of some direct methods for computing stationary distributions of Markov chains*, SIAM J. Sci. Statist. Comput., 5, pp. 453–469.

J. G. Kemeny and J. L. Snell (1960), *Finite Markov Chains*, D. Van Nostrand, New York.

C. D. Meyer (1975), *The role of the group generalized inverse in the theory of finite Markov chains*, SIAM Rev., 17, pp. 443–464.

——— (1982), *Analysis of finite Markov chains by group inversion techniques*, in Recent Applications of Generalized Inverses, S. L. Campbell, ed., Research Notes in Mathematics, Vol. 66, Pitman, Boston, pp. 50–81.

——— (1980), *The condition of a finite Markov chain and perturbation bounds for the limiting probabilities*, SIAM J. Algebraic Discrete Meth., 1, pp. 273–283.

——— (1994), *Sensitivity of Markov chains*, SIAM J. Matrix Anal. Appl., 15, pp. 715–728.

C. D. Meyer and G. W. Stewart (1988), *Derivatives and perturbations of eigenvectors*, SIAM J. Numer. Anal., 25, pp. 679–691.

C. A. O'Cinneide (1993), *Entrywise perturbation theory and error analysis for Markov chains*, Numer. Math., 65, pp. 109–120.

C. Paige, G. Styan, and P. Wachter (1975), *Computation of the stationary distribution of a Markov chain*, J. Statist. Comput. Simul., 4, pp. 173–186.

P. J. Schweitzer (1968), *Perturbation theory and finite Markov chains*, J. Appl. Probab., 5, pp. 401–413.

E. Seneta (1991), *Sensitivity analysis, ergodicity coefficients, and rank-one updates for finite Markov chains*, in Numerical Solution of Markov Chains, W. J. Stewart, ed., Probability: Pure and Applied, No. 8, Marcel Dekker, New York, pp. 121–129.

G. W. Stewart (1991), *On the sensitivity of nearly uncoupled Markov chains*, in Numerical Solution of Markov Chains, W. J. Stewart, ed., Probability: Pure and Applied, No. 8, Marcel Dekker, New York, pp. 105–119.

——— (1992b), *On the perturbation of Markov chains with nearly transient states*, Computer Science Technical Report, University of Maryland, College Park, MD.

G. W. Stewart and G. Zhang (1991), *On a direct method for the solution of nearly uncoupled Markov chains*, Numer. Math., 59, pp. 1–11.

R. Varga (1962), *Matrix Iterative Analysis*, Prentice-Hall, Englewood Cliffs, NJ.

G. Zhang (1993), *On the sensitivity of the solution of nearly uncoupled Markov chains*, SIAM J. Matrix Anal. Appl., 14, pp. 1112–1123.