Sensitivity of the Stationary Distribution Vector for an Ergodic Markov Chain*

R. E. Funderlic
Mathematics and Statistics Research Section
Engineering Physics and Mathematics Division
Oak Ridge National Laboratory
Oak Ridge, Tennessee 37831

and

C. D. Meyer, Jr.
Mathematics Department
North Carolina State University
Raleigh, North Carolina 27695-8205

Submitted by Robert J. Plemmons

ABSTRACT

Stationary distribution vectors p^{∞} for Markov chains with associated transition matrices T are important in the analysis of many models in the mathematical sciences, such as queuing networks, input-output economic models, and compartmental tracer analysis models. The purpose of this paper is to provide insight into the sensitivity of p^{∞} to perturbations in the transition probabilities of T and to understand some of the difficulties in computing an accurate p^{∞}. The group inverse $A^{\#}$ of $I-T$ is shown to be of fundamental importance in understanding sensitivity or conditioning of p^{∞}. The main result shows that if there is a state that is accessible from every other state and the corresponding column of T has no small off-diagonal elements, then p^{∞} cannot be sensitive to small perturbations in T. Ecological examples are given. A new algorithm for calculating $A^{\#}$ is described.

1. INTRODUCTION

For an n-state finite, homogeneous, ergodic Markov chain with transition matrix $T=[p_{ij}]$, the stationary distribution is the unique row vector p^{∞}

*Work supported by the National Science Foundation Grant MCS-82-19500 under subcontract with Oak Ridge National Laboratory, operated by Martin Marietta Energy Systems, Inc. for the U.S. Department of Energy under Contract No. DE-AC05-840R21400.

0024-3795/86/$0.00

satisfying

$$p^{\infty}T = p^{\infty}, \qquad \sum p_i^{\infty} = 1.$$

Letting $A_{n\times n}$ and $e_{n\times 1}$ denote the matrices $A = I - T$ and $e = [1,1,\dots,1]^T$, the stationary distribution p^{∞} can be characterized as the unique solution to the linear system of equations defined by

$$p^{\infty}A = 0 \quad \text{and} \quad p^{\infty}e = 1.$$

(See [14] for an elementary exposition of finite ergodic chains.)

The theory of finite Markov chains has long been a fundamental tool in the analysis of social and biological phenomena. More recently the ideas embodied in Markov-chain models along with the analysis of a stationary distribution have proven to be useful in applications which do not fall directly into the traditional Markov-chain setting. Some of these applications include the analysis of queuing networks [13], the analysis of compartmental ecological models [6], and least-squares adjustment of geodetic networks [1]. Recently, the behavior of the numerical solution of systems of nonlinear reaction-diffusion equations has been analyzed by making use of the stationary distribution of a finite Markov chain in conjunction with the concept of the group inverse [7].

An ergodic chain manifests itself in the transition matrix T, which must be row stochastic and irreducible. Of central importance is the sensitivity of the stationary distribution p^{∞} to perturbation in the transition probabilities of T. This area has been addressed by several authors, substantial contributions being made by Schweitzer [17], Meyer [16], Golub and Meyer [8], Harrod [10], and Conlisk [3]. This work has provided a solid base for the theoretical analysis of the sensitivity question, but little has been written concerning the practical problem of how one can easily decide *a priori* whether or not a given chain might be sensitive to small perturbations. The tests for sensitivity which can be implemented numerically include the computation or estimation of the "chain condition number" $|\max_{i,j} a_{ij}^{\#}|$ proposed in Section 2 (see Section 2 for the definition and summary of properties concerning $A^{\#}$), computation of the eigenvalue of A of second smallest magnitude (proposed by Funderlic and Heath [5]), and computation of the second smallest singular value of A (proposed by Harrod and Plemmons [11]). Each of these may be a formidable numerical task which can be quite expensive relative to the computation of the stationary distribution p^{∞}.

The purpose of this paper is to demonstrate that in many common situations it is only necessary to inspect the magnitudes of the transition probabilities in order to be certain that p^{∞} is not sensitive to small perturbations in T. In particular, we will show that if there exists at least one state S_j which is directly accessible from every other state, and if p_{ij} is not close to zero for $i \neq j$, then p^{∞} cannot be sensitive to small perturbations in T.

2. BACKGROUND MATERIAL

The sensitivity of p^{∞} is most easily gauged by considering the transition probabilities in T to be differentiable functions of a parameter t and by examining the derivatives of the stationary probabilities with respect to t. This approach has been adopted by Golub and Meyer [8] and Conlisk [3]. An alternative approach taken by Schweitzer [17] and Funderlic and Heath [5] is to examine partial derivatives $\partial p^{\infty}/\partial p_{ij}$. In this paper, we will take advantage of Golub and Meyer's results, which are phrased in terms of the group inverse $A^{\#}$ of $A = I - T$. Below is a short summary concerning the matrix $A^{\#}$. Proofs and additional background material on $A^{\#}$ may be found in [2, 15, 16].

(2.1) Each finite Markov chain has the property that $A = I - T$ belongs to some multiplicative matrix group. (T is the transition matrix.) Let G denote the maximal subgroup containing A. The inverse of A with respect to G is denoted by $A^{\#}$, and the identity element in G is denoted by E.

(2.2) For all finite Markov chains, the limiting matrix is the difference of the two identities I and E in the sense that

$$T^{\infty} = \lim_{k \to \infty} \frac{I + T + T^2 + \cdots + T^{k-1}}{k} = I - E = I - AA^{\#}.$$

Of course, if the chain has a limiting matrix in the strong sense, then

$$T^{\infty} = \lim_{k \to \infty} T^k = I - E.$$

(2.3) If the chain is ergodic (i.e., T is irreducible), then

$$T^{\infty} = I - E = I - AA^{\#} = ep^{\infty},$$

where e is a column vector of 1's.

(2.4) The group inverse $A^{\#}$ of A can be characterized as the unique matrix satisfying the three equations $AA^{\#}A = A$, $A^{\#}AA^{\#} = A^{\#}$, and $AA^{\#} = A^{\#}A$.

The following results of Golub and Meyer [8] make it transparent that the matrix $A^{\#}$ is a fundamental quantity governing the sensitivity of the stationary distribution of an ergodic chain.

THEOREM 2.1. *Let $T(t)$ be a matrix that on some interval (a,b) has entries $p_{ij}(t)$ which are differentiable functions of a parameter t. Furthermore, suppose that $T(t)$ is row stochastic and irreducible on (a,b), so that there is a uniquely defined stationary distribution $p^{\infty}(t)$ for each t in (a,b). The derivatives of the stationary probabilities are given by*

$$\frac{dp^{\infty}(t)}{dt} = p^{\infty}(t)\frac{dT(t)}{dt}A^{\#}(t),$$

where $A^{\#}(t) = [I - T(t)]^{\#}$. In particular, the derivative of the ith stationary probability is

$$\frac{dp_i^{\infty}(t)}{dt} = p^{\infty}(t)\frac{dT(t)}{dt}A_i^{\#}(t),$$

where $A_i^{\#}(t)$ is the ith column of $A^{\#}(t)$.

In loose terms, Theorem 2.1 says that in gauging the sensitivity of p^{∞}, small perturbations in the transition probabilities are "magnified" by the entries in $A^{\#}$. The sensitivity of the ith stationary probability depends on the entries of the ith column of $A^{\#}$.

The "condition number" for an ergodic chain arises from the following result of Meyer [16], which is a discretized version of the previous theorem.

THEOREM 2.2. *Let C and $\tilde{C}$ be ergodic chains with transition matrices T and $\tilde{T} = T - F$ and stationary distributions p^{∞} and $\tilde{p}^{\infty}$, respectively. If $A = I - T$, then $I + FA^{\#}$ is nonsingular for all $F = T - \tilde{T}$ and*

$$\tilde{p}^{\infty} = p^{\infty}(I + FA^{\#})^{-1}. \tag{2.5}$$

Use (2.5) to write $p^{\infty} - \tilde{p}^{\infty} = \tilde{p}^{\infty}FA^{\#}$ and multiply on the right by e_i, the ith unit column vector, to obtain

$$p_i^{\infty} - \tilde{p}_i^{\infty} = \tilde{p}^{\infty}FA_i^{\#},$$

where $A_i^{\#}$ denotes the ith column of $A^{\#}$. By the Hölder inequality

$$|p_i^\infty - \tilde{p}_i^\infty| \leqslant \|\tilde{p}^\infty\|_1 \|FA_i^{\#}\|_\infty \leqslant \|FA_i^{\#}\|_\infty. \tag{2.6}$$

This gives the following theorem, which again shows the importance of the group inverse in the sensitivity of the ith stationary probability.

THEOREM 2.3. *If C and $\tilde{C}$ are ergodic chains with transition matrices T and $\tilde{T} = T - F$ and stationary distributions p^∞ and $\tilde{p}^\infty$, respectively, then*

$$|p_i^\infty - \tilde{p}_i^\infty| \leqslant \|F\|_\infty \max_j |a_{ij}^{\#}|, \tag{2.7}$$

where $a_{ij}^{\#}$ denotes the (i, j) entry of $A^{\#}$.

The relation

$$\max_i |p_i^\infty - \tilde{p}_i^\infty| \leqslant \|F\|_\infty \max_{ij} |a_{ij}^{\#}|, \tag{2.8}$$

motivates the following definition.

DEFINITION. The number

$$\kappa(C) = \max_{i,j} |a_{ij}^{\#}|$$

is defined to be the *condition number for the chain C.*

In passing, it is worth noting that it is always the case that

$$\|F\|_\infty = \|T - \tilde{T}\|_\infty \leqslant 2.$$

Hence (2.7) and (2.8) yield the following bounds which hold for all perturbations:

$$|p_i^\infty - \tilde{p}_i^\infty| \leqslant 2 \max_j |a_{ij}^{\#}| \tag{2.9}$$

and

$$\max_i |p_i^\infty - \tilde{p}_i^\infty| \leqslant 2\kappa(C). \tag{2.10}$$

3. THE INSENSITIVITY OF A CHAIN WITH A STRONGLY ACCESSIBLE STATE

Throughout this section it is assumed that $A = I - T$, where $T_{n\times n} = [p_{ij}]$ is the transition matrix of an n-state ergodic chain C; i.e., T is an irreducible row stochastic matrix.

DEFINITION. The order of accessibility of state S_j is defined to be that number δ_j such that $\min_{i\neq j} p_{ij} = \delta_j$. The state S_j is considered to be a *strongly accessible* state if $\delta_j \neq 0$ and δ_j is large relative to 1.

In other words, a strongly accessible state S_j is one which is *directly* accessible from every state and such that the probability of moving into S_j from each other state is large. In terms of $A = I - T$, the jth column of A has all its off-diagonal elements relatively close to -1.

Our major purpose is to demonstrate rigorously that if a chain possesses at least one strongly accessible state, then the chain must be well conditioned in the sense that the stationary distribution p^{∞} must be relatively insensitive to small perturbations in the transition probabilities. This will be accomplished by showing that chains which possess a strongly accessible state must necessarily have a small condition number $\kappa(C)$, as described in the previous section. To best facilitate the development, we will always use the l_{∞} matrix norm, and we will obtain the major result as a sequence of two theorems.

It is well known that if $A_{n\times n}$ is irreducible, then $\operatorname{rank}(A) = n-1$ and hence every subset of $n-1$ rows (or columns) of T is linearly independent. Hence every $(n-1)\times(n-1)$ submatrix of A is nonsingular. Our first major result relates the magnitudes of the inverses of these submatrices to the magnitude of $A^{\#}$.

THEOREM 3.1. *For every $(n-1)\times(n-1)$ submatrix B of $A = I - T$,*

$$\|A^{\#}\|_{\infty} \leqslant 4\|B^{-1}\|_{\infty}.$$

Proof. Let B be any $(n-1)\times(n-1)$ submatrix of A. There exist permutation matrices P and Q such that

$$PAQ = \begin{bmatrix} B & c \\ d^T & \alpha \end{bmatrix}.$$

It is well known that the condition

$$\operatorname{rank}\begin{bmatrix} B & c \\ d^T & \alpha \end{bmatrix} = \operatorname{rank}(B)$$

guarantees

$$\alpha = d^T B^{-1} c$$

[2, Corollary 6.3.8, p. 103]. Write A as

$$A = P^T \begin{bmatrix} B & c \\ d^T & \alpha \end{bmatrix} Q^T,$$

and define A^- to be the matrix

$$A^- = Q \begin{bmatrix} B^{-1} & 0 \\ 0 & 0 \end{bmatrix} P. \tag{3.1}$$

Because $\alpha = d^T B^{-1} c$, it is easy to verify that $AA^-A = A$. Recall from (2.4) that $A^\# = A^\# A A^\#$, and use the fact that $A = AA^-A$, where A^- is the matrix of (3.1), to obtain

$$A^\# = A^\#(AA^-A)A^\# = EA^-E, \tag{3.2}$$

where E is the identity element of the maximal subgroup which contains A. From (2.2), $E = I - T^\infty$ and hence

$$\|E\|_\infty = \|I - T^\infty\|_\infty \leqslant \|I\|_\infty + \|T^\infty\|_\infty \leqslant 2.$$

Since $\|\cdot\|_\infty$ is invariant under all row and column permutations, taking norms in (3.2) produces

$$\|A^\#\|_\infty \leqslant \|E\|_\infty \|A^-\|_\infty \|E\|_\infty = 4\|A^-\|_\infty = 4\|B^{-1}\|_\infty. \qquad \blacksquare$$

Since $\kappa(C) = \max_{ij}|a^\#_{ij}| \leqslant \|A^\#\|_\infty$, the preceding theorem yields the following corollary.

COROLLARY. *If a chain C is ill conditioned in the sense that $\kappa(C)$ is large relative to 1, then the inverse of every $(n-1)\times(n-1)$ submatrix B of A must have some entries of large magnitude.*

(The converse of this corollary has neither been proven nor disproven.)

The next theorem is the fundamental result of this paper. It says that if an ergodic (irreducible) chain possesses at least one strongly accessible state, then the chain must be well conditioned in the sense that the stationary distribution p^{∞} is relatively insensitive to small perturbations in the transition probabilities contained in T.

THEOREM 3.2. *Let δ_j be the order of state S_j of a chain C. Then*

$$\kappa(C) \leqslant 4/\delta_j.$$

Proof. For each j, $\delta_j = \min_{i \neq j} p_{ij}$. The proof will hinge on the fact that

$$\|A^{\#}\|_{\infty} \leqslant \frac{4}{\delta_j} \tag{3.3}$$

for each $j = 1, 2, \ldots, n$. This will guarantee that if there is a strongly accessible state, then $\|A^{\#}\|_{\infty}$ must be relatively small in magnitude. Since it is always the case that $\kappa(C) \leqslant \|A^{\#}\|_{\infty}$, it will follow that for each j

$$\kappa(C) \leqslant \frac{4}{\delta_j}. \tag{3.4}$$

Hence $\kappa(C)$ will be relatively small whenever there exists a strongly accessible state. To prove (3.3), proceed as follows. Suppose the jth state is strongly accessible, so that in particular, $p_{ij} \neq 0$ for $i \neq j$. Reorder the states to make state j the last state of the chain. That is, execute a symmetric permutation on the entries of T to give

$$Q^T T Q = \begin{bmatrix} S & c \\ d^T & \alpha \end{bmatrix},$$

where Q is the permutation matrix obtained by interchanging columns j and n in I. In a symmetric permutation, diagonals go to diagonals, and thus c has no zero entries, i.e., $c > 0$. This implies that *every* row sum of S is strictly less than 1 so that $\|S\|_{\infty} < 1$. Write

$$Q^T A Q = Q^T(I - T)Q = \begin{bmatrix} I - S & -c \\ -d^T & 1-\alpha \end{bmatrix} = \begin{bmatrix} U & -c \\ -d^T & 1-\alpha \end{bmatrix}.$$

Since $\|S\|_\infty < 1$, it follows that $U = I - S$ is nonsingular and

$$U^{-1} = (I - S)^{-1} = I + S + S^2 + S^3 + \cdots .$$

Hence

$$\|U^{-1}\|_\infty \leqslant 1 + \|S\|_\infty + \|S\|_\infty^2 + \|S\|_\infty^3 + \cdots = \frac{1}{1 - \|S\|_\infty}.$$

It is easily verified that $(Q^TAQ)^{\#} = Q^TA^{\#}Q$, so that

$$\|A^{\#}\|_\infty = \|Q^TA^{\#}Q\|_\infty = \left\|(Q^TAQ)^{\#}\right\|_\infty.$$

From the previous theorem, we know that

$$\left\|(Q^TAQ)^{\#}\right\|_\infty \leqslant 4\|U^{-1}\|_\infty.$$

Therefore,

$$\|A^{\#}\|_\infty \leqslant 4\|U^{-1}\|_\infty \leqslant \frac{4}{1 - \|S\|_\infty},$$

If e denotes the column of 1's, then

$$c = (I - S)e = e - Se.$$

This implies that the smallest component of c lies in the position corresponding to the row of S which has maximal row sum. That is,

$$1 - \|S\|_\infty = \min_i c_i = \delta_j.$$

Hence

$$\|A^{\#}\|_\infty \leqslant \frac{4}{1 - \|S\|_\infty} = \frac{4}{\delta_j},$$

and (3.3) is proven. ■

COROLLARY. *From the preceding proof, we have that*

$$\kappa(C) \leqslant \|A^{\#}\|_{\infty} \leqslant \frac{4}{\max_{j} \min_{i \neq j} p_{ij}}.$$

It is worth noting that the converse to Theorem 3.2 is not true. That is, a chain with no strongly accessible states need not be badly conditioned. For example, the chain with transition matrix

$$T = \begin{bmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

has no strongly accessible states, but

$$A = \frac{1}{2}\begin{bmatrix} 1 & 0 & -1 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix}$$

and

$$A^{\#} = \frac{1}{3}\begin{bmatrix} 2 & -2 & 0 \\ 0 & 2 & -2 \\ -2 & 0 & 2 \end{bmatrix},$$

so that $\kappa(C) = \frac{2}{3}$ is relatively small and hence the chain cannot be sensitive to small perturbations.

4. EXAMPLES

Transition matrices of the form

$$T = \begin{bmatrix} q_1 & 0 & \cdots & 0 & 1-q_1 \\ 0 & q_2 & \cdots & 0 & 1-q_2 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & q_{n-1} & 1-q_{n-1} \\ p_1 & p_2 & \cdots & p_{n-1} & 1-\Sigma p_i \end{bmatrix}, \qquad q_n = 1 - \sum p_i, \quad (4.1)$$

arise from a compartmental analysis of systems called *mammillary systems* by Sheppard and Householder [18]. Clearly, these are systems in which there is one central state (or main processing unit) as depicted in Figure 1 for $n = 5$.

If the q_i's are zero, it follows from the results of the previous section that such mammillary systems are very well conditioned because the order of accessibility of the "central state" is 1 (the maximum possible) and hence $\kappa(C) \leqslant 4$ regardless of how many states are in the chain. The stationary probabilities of these chains must therefore be very insensitive to perturbations in the transition probabilities.

For mammillary systems with the $q_i = 0$, the simple structure allows us to extract explicitly the general form of $A^{\#}$ and p^{∞} so that we may independently confirm our conclusions. We use Theorems 8.5.2 and 8.5.3 in [2] to obtain

$$A^{\#} = \frac{1}{4}\begin{bmatrix} 4I - 3eq^T & -e \\ -q^T & 1 \end{bmatrix},$$

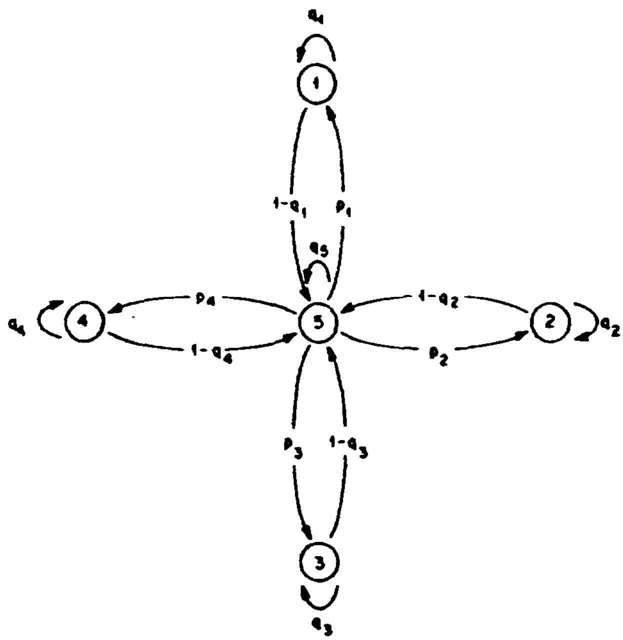

FIG. 1. A mammillary system.

where $q^T = [p_1, p_2, \ldots, p_{n-1}]$ and

$$p^\infty = \tfrac{1}{2}[p_1, p_2, \ldots, p_{n-1}, 1].$$

The actual value of $\kappa(C)$ is

$$\kappa(C) = \max\{1 - \tfrac{3}{4}p_i, \tfrac{3}{4}p_i, \tfrac{1}{4}\} < 1.$$

The upper bound of 4 provided by the corollary to Theorem 3.2 is not overly pessimistic. Moreover, the characteristic equation of the matrix $T_{n\times n}$ in (4.1) is $\lambda^{n-2}(\lambda^2 - 1) = 0$, so that the eigenvalues of T and A are $\lambda_T = \{1, 0, -1\}$ and $\lambda_A = \{0, 1, 2\}$. This observation also tends to corroborate the conclusion that the stationary distribution is insensitive to perturbations in T because of the fact that there is no other eigenvalue of A close to 0. By continuity, it follows that if the q_i's are each close to 0, then the associated stationary distribution cannot be sensitive.

A similar situation which concerns the analysis of radiophosphorus kinetics in an aquarium system is described by Whittaker [20, pp. 182–184]. Placed in a Markov-chain setting, Whittaker's analysis leads to a chain whose transition matrix is

$$T = \begin{bmatrix} .740 & .110 & 0 & 0 & 0 & 0 & 0 & .150 \\ 0 & .689 & 0 & 0 & .011 & 0 & 0 & .300 \\ 0 & 0 & 0 & .400 & 0 & 0 & 0 & .600 \\ 0 & 0 & 0 & .669 & .011 & 0 & 0 & .320 \\ 0 & 0 & 0 & 0 & .912 & 0 & 0 & .088 \\ 0 & 0 & 0 & 0 & 0 & .740 & 0 & .260 \\ 0 & 0 & 0 & 0 & 0 & 0 & .870 & .130 \\ .150 & 0 & .047 & 0 & 0 & .055 & .270 & .478 \end{bmatrix}.$$

Again, there is a "central state" which is accessible from all other states. In this case, we see that $\delta_8 = 0.088$, so that

$$\kappa(C) \leqslant 4/0.088 = 45.45.$$

Hence we expect the chain to be moderately well conditioned. Direct computation, using 16-digit precision, yields $A^\#$ as shown below (rounded to

3 significant digits): $A^{\#}=$

$$\begin{bmatrix} 3.28 & 1 & -0.0147 & -0.0585 & 0.0299 & -0.209 & -3.95 & -0.074 \\ -0.329 & 2.94 & 0.00501 & -0.0347 & 0.275 & -0.121 & -3.08 & 0.344 \\ -0.156 & -0.211 & 1.02 & 1.19 & 0.0343 & -0.0573 & -2.46 & 0.643 \\ -0.299 & -0.262 & 0.00745 & 2.99 & 0.253 & -0.11 & -2.98 & 0.396 \\ -1.39 & -0.648 & -0.0816 & -0.139 & 11.2 & -0.51 & -6.91 & -1.5 \\ -0.36 & -0.283 & 0.00245 & -0.0378 & -0.0378 & -0.128 & 3.71 & -3.2 \\ -0.888 & -0.47 & -0.0405 & -0.0898 & -0.158 & -0.326 & 2.6 & -0.625 \\ 0.167 & -0.0966 & 0.0454 & 0.0141 & -0.0984 & 0.0613 & -1.3 & 1.2 \end{bmatrix}$$

Thus the value for $\kappa(C)$ is 11.2. Of course, computation of $A^{\#}$ is generally expensive (see the next section), while there is almost no computation involved in estimating $\kappa(C) \leqslant 4/(\max_j \min_{i \neq j} p_{ij})$. This fact should mitigate the slightly conservative nature of the estimate. However, if $A^{\#}$ is computed, then much more than $\kappa(C)$ is available. In particular, the sensitivity of each stationary probability can be gauged using (2.7). In this example the entry of maximum magnitude in each column of $A^{\#}$ is shown in the following table:

Col. 1	Col. 2	Col. 3	Col. 4	Col. 5	Col. 6	Col. 7	Col. 8
3.28	2.94	1.02	2.99	11.2	3.71	6.91	1.5

Thus p_3^∞ and p_8^∞ must be very insensitive to perturbations, while p_5^∞ may be slightly more sensitive.

5. COMPUTATIONAL ASPECTS

The purpose of this section is to give a simple method for calculating or forming $A^{\#}$. This method generalizes a frequently used method for calculating the inverse for general nonsingular matrices and analgous stability comments could be made.

Even in the case of nonsingular matrices, the inverse of a matrix is seldom needed. To quote Forsythe and Moler [4, p. 79]:

> ...we recommend strongly against computing A^{-1}...almost anything you can do with A^{-1} can be done without it.

The same is true for $A^{\#}$. However, there are cases when knowledge of $A^{\#}$ may be desirable. One example concerns computation of the mean first-passage times. According to Campbell and Meyer [2, 16], the expected number of steps to move from state S_i to state S_j for the first time is the (i, j) entry of the matrix

$$M = \left(I - A^{\#} + JA^{\#}_{\mathrm{dg}}\right)D,$$

where J is the matrix of all 1's, $A^{\#}_{\text{dg}}$ is the diagonal matrix obtained by setting the off-diagonal entries of $A^{\#}$ to 0, and D is the diagonal matrix

$$D=\operatorname{diag}(1/p_1^{\infty},\, 1/p_2^{\infty},\ldots,1/p_n^{\infty}).$$

The method to be generalized (see Forsythe and Moler [4, pp. 77–79]) uses the procedure for solving linear equations. The columns of A^{-1} are simply the respective solutions of the n different linear systems:

$$Ax_1=e_1,\qquad Ax_2=e_2,\ldots,\qquad Ax_n=e_n. \tag{5.0}$$

Each solution makes use of an LU factorization of A (now permuted for stability).

We give here the analogue of this method for matrices that arise from Markov chains. Our method is based on the observation that $A^{\#}$ is the only matrix that satisfies the equations

$$AX=I-ep^{\infty},\qquad p^{\infty}X=0. \tag{5.1}$$

Suppose that there are two solutions X_1 and X_2, so that $A(X_1-X_2)=0$. Since the null space of A is spanned by e, we have

$$X_1e_i-X_2e_i=\alpha_ie,\qquad i=1,2,\ldots,n.$$

Premultiplication by p^{∞} implies every $\alpha_i=0$. Similarly $A^{\#}$ is the only matrix that satisfies

$$XA=I-ep^{\infty},\qquad Xe=0. \tag{5.2}$$

An algorithm analogous to that suggested by (5.0) would be

(a) $A=LU$.
(b) Obtain p^{∞} from (a) by backsubstitution.
(c) Solve successively

$$LUz_i=e_i-p_i^{\infty}e.$$

(d) Normalize z_i:

$$A_i^{\#}=z_i-(p^{\infty}z_i)e.$$

(e) $A^{\#}=[A_1^{\#},A_2^{\#},\ldots,A_n^{\#}]$.

Several comments are in order. The matrix $[A_1^\#, \ldots, A_n^\#]$ satisfies (5.1) because of steps (c) and (d). As suggested by Golub and Van Loan [9, p. 32], a flop is defined to be a scalar computation of the form $\alpha\beta + \gamma$. The factorization (a) can be carried out in $n^3/3$ flops in a stable way without pivoting (e.g., see [11]), and p^∞ can be obtained by back substitution from U^T (see [6]).

Funderlic and Mankin [6] showed that once either L or U is normalized in (a), then the only nonuniqueness is that either l_{nn} or u_{nn} is zero. It is convenient to choose $l_{mn} = 1$ and $u_{nn} = 0$. Therefore for $Lq_i = e_i - p_i^\infty e$ in (b), the last component of q_i is zero and the last component of z_i may be chosen as zero in the back substitution $Uz_i = q_i$.

If (5.0) is carried out for nonsingular matrices and flops of order less than n^3 are ignored, it takes $4n^3/3$ flops to invert a matrix. However, if one takes advantage of the special structure of the e_i vectors, the flops can be reduced to n^3 (see Issacson and Keller [12, p. 36]). Since no such economy exists for the calculation of $A^\#$, $4n^3/3$ flops are used.

A dual algorithm may be based on (5.2). Here steps (c) and (d) would be

(c′) $z_i^T LU = e_i^T - p^\infty$,
(d′) $(A_i^\#)^T = z_i^T - (z_i^T e)p^\infty$,

where the row vectors in (d′) are the corresponding rows of $A^\#$.

It is well known that if a symmetric positive definite matrix is poorly conditioned, the computation of the Cholesky decomposition may break down. An example of this is given by Golub and Van Loan [9, p. 90]:

$$\begin{bmatrix} 1.00 & 0.15 & 0.01 \\ 0.15 & 0.023 & 0.01 \\ 0.01 & 0.01 & 1.00 \end{bmatrix}.$$

On a two-digit decimal computer that rounds, the second diagonal element of the factorization is calculated as zero. Wilkinson [22] proved that if a symmetric positive definite matrix is not too poorly conditioned, then the Cholesky process can be computationally carried out. Wilkinson's result and examples such as that of Golub and Van Loan suggest obvious questions for transition matrices. The following example suggested by an idea of G. W. Stewart [19] answers some of these questions. Let

$$A = I - T = \begin{bmatrix} \alpha_1 + \alpha_2 + \alpha_3 & -\alpha_1 & -\alpha_2 & -\alpha_3 & 0 \\ 0 & 1 & 0 & -1 & 0 \\ -1 & 0 & 1 & 0 & 0 \\ -s & 0 & 0 & s+\varepsilon & -\varepsilon \\ 0 & 0 & 0 & -(s+\varepsilon) & s+\varepsilon \end{bmatrix}. \tag{5.3}$$

For small ε, the leading principal submatrix of order 4 is nearly singular and ill conditioned with respect to inversion. However, the principal submatrix obtained by deleting the first row and column is well conditioned for small ε (unless s is small). Therefore, by Theorem 3.1, A is well conditioned.

A stable and efficient method to solve systems $A^T x = 0$ for the stationary vector is Gaussian elimination without pivoting, which for these problems is equivalent to one step of inverse iteration; see [11] and [6]. Mathematically, Gaussian elimination gives an LU factorization of these matrices such that L and U are M-matrices with L having a unit diagonal and U having positive diagonal elements except for $u_{nn} = 0$. However, under usual computational assumptions, the α's, s and ε for the matrix given by (5.3) may be chosen such that Gaussian elimination breaks down; i.e., the pivot in the (4,4) position can become zero or negative. To illustrate this, we give for simplicity an example with the α's and s chosen as integers, though these integers could be scaled so that $A = I - T$ with T a transition matrix. On a two-digit decimal computer that rounds, $\alpha_1 = 2$, $\alpha_2 = 2$, $\alpha_3 = 3$, $s = 4$, $\varepsilon = 0.10$ will cause the fourth pivot to be zero when Gaussian elimination without pivoting is carried out on A^T, and $s = 5$ will cause this pivot to be -0.10. Furthermore, if the fifth diagonal entry of A is increased, then A is nonsingular and the same breakdown occurs. As has been suggested by Wilkinson [21], replacement of a zero pivot (the fourth in our case) by a small number will allow an accurate calculation of a well-conditioned eigenvector. This example shows that Gaussian elimination with finite-precision arithmetic can break down for matrices arising from Markov chains even though the associated chain is well conditioned. Though the example of (5.3) is illustrated with two-digit arithmetic, the example is general enough to illustrate the conclusions with any finite precision.

REFERENCES

1 A. Brandt, Algebra's multigrid theory, to appear.
2 S. L. Campbell and C. D. Meyer, *Generalized Inverses of Linear Transformations*, Surveys and Reference Works in Mathematics, Pitman, London, 1979.
3 J. Conlisk, Comparative statics for Markov chains, to appear.
4 G. E. Forsythe and C. B. Moler, *Computer Solution of Linear Algebraic Systems*, Prentice-Hall, Englewood Cliffs, N.J., 1967.
5 R. E. Funderlic and M. T. Heath, Linear compartmental analysis of ecosystems, ORNL/IBP-71/4, Oak Ridge National Laboratory, Oak Ridge, Tenn., 1971.
6 R. E. Funderlic and J. B. Mankin, Solution of homogeneous systems of equations arising from compartmental models, *SIAM J. Sci. Statist. Comput.* 2:375–313 (1981).

7 L. Galeone, The use of positive matrices for the analysis of the large time behavior of the numerical solution of reaction-diffusion systems, *Math. Comp.* 41:461–472 (1983).
8 G. H. Golub and C. D. Meyer, Using the QR factorization and group inverse to compute, differentiate and estimate the sensitivity of stationary probabilities for Markov chains, *SIAM J. Algebraic Discrete Methods*, to appear.
9 G. H. Golub and C. F. Van Loan, *Matrix Computations*, Johns Hopkins, Baltimore, 1983.
10 W. J. Harrod, Rank modification method for certain singular systems of linear equations, Ph.D. Dissertation, Univ. of Tennessee, Knoxville, Tenn., 1982.
11 W. J. Harrod and R. J. Plemmons, Comparison of some direct methods for computing stationary distributions of Markov chains, *SIAM J. Sci. Statist. Comput.* 5:453–469 (1984).
12 E. Issacson and H. B. Keller, *Analysis of Numerical Methods*, Wiley, New York, 1966.
13 L. Kaufman, Matrix methods for queuing problems, *SIAM J. Sci. Statist. Comput.* 4:525–552 (1984).
14 J. G. Kemeny and J. L. Snell, *Finite Markov Chains*, Van Nostrand, New York, 1960.
15 C. D. Meyer, The role of the group generalized inverse in the theory of finite Markov chains, *SIAM Rev.* 17:443–464 (1975).
16 C. D. Meyer, The condition of a finite Markov chain and perturbation bounds for the limiting probabilities, *SIAM J. Algebraic Discrete Methods* 1:273–283 (1980).
17 P. J. Schweitzer, Perturbation theory and finite Markov chains, *J. Appl. Probab.* 5:401–413 (1968).
18 C. W. Sheppard and A. S. Householder, The mathematical basis of the interpretation of tracer experiments in closed steady-state systems, *J. Appl. Phys.* 22:510–520 (1951).
19 G. W. Stewart, Personal correspondence.
20 R. H. Whittaker, Experiments with radiophosphorus tracer in aquarium microcosms, *Ecological Monographs* 31:157–188 (1961).
21 J. H. Wilkinson, *The Algebraic Eigenvalue Problem*, Clarendon, Oxford, 1965.
22 J. H. Wilkinson, *A priori* error analysis of algebraic processes, in *Proceedings of the International Congress on Mathematics*, Izdat. Mir, Moscow, 1968, pp. 629–639.

Received 24 September 1984; revised 19 February 1985