Numerical Analysis Project
Manuscript NA-85-32

March 1985

Simultaneous Computation of Stationary Probabilities With Estimates of Their Sensitivity

by

Gene H. Golub

Carl D. Meyer, Jr.

Numerical Analysis Project
Computer Science Department
Stanford University
Stanford, California 94305

SIMULTANEOUS COMPUTATION OF STATIONARY PROBABILITIES WITH

ESTIMATES OF THEIR SENSITIVITY

Gene Golub +

Computer Science Department
Stanford University
Stanford, California 94305

Carl D. Meyer, Jr.*

Department of Mathematics
North Carolina State University
P. 0. 80x 8205
Raleigh, North Carolina 27695-8205

+ This work was supported in part by the National Science Foundation under Grant No. MCS 78-11985.

* This work was supported in part by the National Science Foundation under Grant No. MCS-82-19500.

1. INTRODUCTION

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

$$\pi P = \pi, \quad \sum \pi_i = 1.$$

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

$$\pi A = 0 \text{ and } \pi e = 1.$$

(See Kemeny and Snell [11] 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 (Kaufman [7]), the analysis of compartmental ecological models (Funderlic and Mankin [5]), and least squares adjustment of geodedic networks (Brandt [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 group matrix inversion (Galeone [6]).

An ergodic chain manifests itself in the transition matrix P which must be *row* stochastic and irreducible. Of central importance is the sensitivity of the stationary distribution π to perturbations in the transition probabilities in P.

The sensitivity of π is most easily gauged by considering the transition probabilities in P to be differentiable functions. One approach, adopted by Conlisk [3], Schweitzer [11], and Funderlic and Heath [4] is to examine partial derivatives $\partial\pi/\partial p_{ij}$. Our strategy is to consider the transition probabilities $p_{ij}(t)$ as differentiable functions of a single parameter t and study the stationary distribution $\pi(t)$ as a function of t. We present a new and very simple formulation for the derivative, $d\pi(t)/dt$, of the stationary distribution directly in terms of the derivatives $dp_{ij}(t)/dt$ and entries from $\pi(t)$ and a matrix $A^{\#}(t)$, called the group inverse of $A(t) = I - P(t)$. After the derivative $d\pi(t)/dt$ has been obtained, we demonstrate its applicability by using it to deduce the relative sensitivity of a discrete Markov chain. This is followed by a first order perturbation analysis. Finally, it is demonstrated how a QR factorization can be used to simultaneously compute π along with estimates which gauge the sensitivity of π to perturbations in P.

2. BACKGROUND MATERIAL

In this paper, we take advantage of results which are phrased in terms of the group inverse $A^{\#}$ of $A = I - P$. Below is a short summary concerning the matrix $A^{\#}$. Proofs and additional background material on $A^{\#}$ may be found in Campbell and Meyer [2] and Meyer [9], [10].

BACKGROUND MATERIAL CONCERNING $A^{\#}$

(2.1) Each finite Markov chain has the property that $A = I - P$ belongs to some multiplicative matrix group. (P 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

$$P^{\infty} = \lim_{k\to\infty} \frac{I + P + P^2 + \dots + P^{k-1}}{k} = I - E = I - AA^{\#}$$

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

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

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

$$P^{\infty} = I - E = I - AA^{\#} = e\pi$$

where e is a column 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 $AA^{\#} = A^{\#}A$.

3. DIFFERENTIATION OF THE STATIONARY DISTRIBUTION

Throughout this section, we assumethat $A(t) = I - P(t)$ where $P(t)$ is a matrix which is row stochastic and irreducible for each t in some interval (a,b). Furthermore, we will assume that each entry $p_{ij}(t)$ of $P(t)$ is differentiable at each t in (a,b). It is important to note at the outset that, in general, the null vectors of a differentiable matrix need not be differentiable. However, for our special situation, normalizing a null vector of $A(t)$ so as to produce the stationary distribution vector $n(t)$ always results in a differentiable vector.

> THEOREM 3.1: If $A(t) = I - P(t)$ where $P(t)$ is row stochastic, irreducible, and differentiable on (a,b), then each component of the unique stationary distribution $\pi(t)$ of $P(t)$ satisfying
>
> $\pi(t) = \pi(t)P(t)$, $\sum \pi_i(t) = 1$, $\pi_i(t) > 0$, is differentiable on (a,b).

PROOF: If $D_i(t)$ denotes the i-th principal minor of $A(t)$ obtained by deleting the i-th row and i-th column of $A(t)$, then for each t in (a,b), $D_i(t) > 0$ and $\pi(t)$ is given by

$$\pi(t) = \frac{1}{\sum D_i(t)} [D_1(t), D_2(t), \ldots, D_m(t)]. \tag{3.1}$$

This formula for $\pi(t)$ is a simple consequence of the fact that $(\operatorname{adj} A)A = A(\operatorname{adj} A) = 0$ and $\{e\}$ is a basis for $N(A)$. Because the entries of $A(t)$ are differentiable, each $D_i(t)$ must be differential and hence each component of $\pi(t)$ must be differentiable at each t in (a,b). ■

In the sequel, we will omit writing the argument t (e.g., instead of writing $\pi(t)$, simply write π) and we sometimes will also use the notation $(\dot{})$ to indicate the differentiation with respect to t (e.g., write $\dot{\pi}$ instead of $d\pi(t)/dt$).

THEOREM 3.2: If $P = P(t)$ is row stochastic, irreducible, and differentiable for t in (a,b), then the derivative of the stationary distribution associated with P is given by

$$\dot{\pi} = \pi \dot{P} A^{\#} \tag{3.2}$$

where $A^{\#}$ denotes the group inverse of $A = I - P$ as described in the previous section.

PROOF: Identify $R(A^T)$ with the row space of A. If $A:\mathbb{R}^{1\times m} \to R(A^T)$ is the mapping defined by $A(x) = xA$ and if $A_r:R(A^T) \to R(A^T)$ denotes the restriction of A to the space $R(A^T)$, then A_r is a bijection on $R(A^T)$ and it is not difficult to show that $A_r^{-1}:R(A^T) \to R(A^T)$ is given by $A_r^{-1}(x) = xA^{\#}$. (See pp. 121-122 in Campbell and Meyer [2].) Thus for $b \in R(A^T)$, the unique vector x in $R(A^T)$ satisfying $xA = b$ is given by $x = bA^{\#}$. Apply the elementary product rule for differentiation to the equation $\pi = \pi P$ to obtain

$$\dot{\pi} = \dot{\pi} P + \pi \dot{P}$$

or

$$\dot{\pi} A = \pi \dot{P} \tag{3.3}$$

and deduce that $\pi\dot{P} \in R(A^T)$. To see that $\dot{\pi} \in R(A^T)$, differentiate the equation $\pi e = 1$ to obtain

$$\dot{\pi}e = 0.$$

Hence $\dot{\pi}$ is orthogonal to e. Since $\{e\}$ is a basis for N(A) it follows that

$$\dot{\pi} \in N(A)^{\perp} = R(A^T). \tag{3.4}$$

Because $A^{\#}$ is the inverse of A on $R(A^T)$, (3.3) and (3.4) imply

$$\dot{\pi} = \pi\dot{P}A^{\#},$$

which is the desired conclusion. ■

By multiplying (3.2) on the right by the i-th unit column e_i, we may extract the expression for the derivative of the i-th stationary probability π_i.

COROLLARY: The derivative of the i-th stationary probability is given by,

$$\dot{\pi}_i = \pi\dot{P}A^{\#}_i \tag{3.5}$$

where A_i is the i-th column of $A^{\#}$.

One of the most pleasing aspects of Theorem 3.2 and its Corollary is the sheer simplicity. The simple structure of (3.2) and (3.5) make it absolutely clear how the stationary distribution changes as the transition probabilities change. It shows that $A^{\#}$ acts as a "magnification factor". If, at a particular point t_0, the derivatives of the transition probabilities are all relatively small and the i-th column of $A^{\#}$ contains only relatively small entries, then the i-th stationary probability π_i must have a relatively small derivative. Because the entries of $A^{\#} = A^{\#}(t)$ are continuous functions of t (Corollary 3.1 in Meyer [10]), it follows that at t_0, π_i cannot be extremely sensitive to small perturbations in the transition probabilities whenever the i-th column of $A^{\#}(t_0)$ has no entries of relatively large magnitude. On the other hand, if the i-th column of $A^{\#}(t_0)$ contains some entries of large magnitude, then small perturbations in $P(t_0)$ can be greatly magnified so as to make π_i very sensitive near t_0.

More precisely, translate the discussion to the origin and write

$$\pi_i(t) - \pi_i(0) = \dot{\pi}_i(0)t + O(t^2)$$

and

$$t\dot{P}(0) = P(t) - P(0) + O(t^2).$$

Theorem 3.2 now produces the following perturbation formula.

$$\pi_i(t) - \pi_i(0) = \pi_i(0)[P(t) - P(0)]A^{\#}(0) + O(t^2). \tag{3.6}$$

It is transparent from (3.6) that the entries of $A^{\#}(0)$ are the fundamental quantities which govern the sensitivity of the stationary probabilities. Assuming that t is small enough so that higher order terms may be neglected, apply Hölder's inequality to (3.6) and obtain the statement

$$(3.7) \qquad |\pi_i(t) - \pi_i(0)| \leq \|\pi(0)\|_p \| [P(t) - P(0)]A_i^{\#}(0)\|_q$$

where $(1/p) + (1/q) = 1$. For all of the Hölder norms it is the case that $\|\pi(t)\|_p \leq 1$. Thus for every Hölder vector norm $\|\cdot\|_q$ and compatible matrix norm $\|\cdot\|_m$, it follows from (3.7) that

$$(3.8) \qquad |\pi_i(t) - \pi_i(0)| \leq \|P(t) - P(0)\|_m \|A_i^{\#}(0)\|_q$$

The observations made throughout this section motivate the following definition.

<u>DEFINITION:</u> For an ergodic Markov chain with transition matrix P and stationary distribution π, the <u>condition number for the i-th stationary probability</u> π_i is defined to be the number

$$\mathrm{Cond}_q(\pi_i) = \|A_i^{\#}\|_q$$

where $\|\cdot\|_q$ is any Hölder vector norm and $A_i^{\#}$ is the i-th column of the group inverse of $A = I - P$. For a matrix norm $\|\cdot\|_m$, the number

$$\mathrm{Cond}_m(\pi) = \|A^{\#}\|_m$$

is defined to be the <u>condition number for</u> π. This number will also be referred to as the <u>condition of the underlying Markov chain</u>.

4. LINEAR PERTURBATIONS

A special case of the preceding analysis which is of particular interest is that in which the perturbations are linear functions. That is, for a fixed row stochastic irreducible matrix P_0, let F be a constant matrix such that

$$P(t) = P_0 + tF$$

is row stochastic and irreducible on (a,b). As before, let $A(t) = I - P(t)$ and $A_0 = A(0) = I - P_0$. By making use of our earlier results, we can obtain a *very* simple and explicit formula for $\dot{\pi}$, the derivative of the stationary distribution associated with P(t).

THEOREM 4.1: If $P(t) = P_0 + tF$ is row stochastic and irreducible on $[0,\beta)$, then the derivative of the stationary distribution $\pi(t)$ associated with $P(t)$ is given by

$$\dot{\pi}(t) = \pi(t)FA_0^{\#}[I - tFA_0^{\#}]^{-1} \quad \text{for } t \text{ in } [0,\beta) \tag{4.1}$$

where $A^{\#}$ is the group inverse of $A_0 = I - P_0$.

PROOF: Using Theorem 3.2, we obtain

$$\dot{\pi}(t) = \pi(t)F(I - P_0 - tF)^{\#} = \pi(t)F(A_0 - tF)^{\#}. \tag{4.2}$$

Let $\pi_0 = \pi(0)$. From Theorem 3.1 of Meyer [10], the matrix $(I - tFA_0^{\#})$ is always nonsingular and the term $(A_0 - tF)^{\#}$ can be expanded as follows.

(4.3) $$(A_0 - tF)^{\#} = A_0^{\#} + tA_0^{\#}FA_0^{\#}(I - tFA_0^{\#})^{-1} - P_0^{\infty}(I - tFA_0^{\#})^{-1}A_0^{\#}(I - tFA_0^{\#})^{-1}$$

where $P_0^{\infty} = e\pi$ is as described in (2.2). Since

$$e = P(t)e = P_0e + tFe = e + tFe$$

must hold for all t in $[0,\beta)$, it follows that $Fe = 0$ and hence

$$FP_0^{\infty} = Fe\pi_0 = 0.$$

Use this when substituting (4.3) into (4.2) to obtain

(4.4) $$\dot{\pi}(t) = \pi(t)FA_0^{\#}[I + tFA_0^{\#}(I - tFA_0^{\#})^{-1}].$$

By making use of the identity

$$I = (I - tFA_0^{\#})^{-1} - tFA_0^{\#}(I - tFA_0^{\#})^{-1},$$

(4.4) reduces to

$$\dot{\pi}(t) = \pi(t)FA_0^{\#}(I - tFA_0^{\#})^{-1},$$

which is the desired conclusion. $\blacksquare$

There are at least two interesting features to this theorem. The first is to notice that at $t = 0$, the behavior of $\dot{\pi}(t)$ is governed strictly by F, $A_0^{\#}$, and $\pi(0)$.

COROLLARY: For $P(t) = P_0 + tF$, the derivative $\dot{\pi}(t)$ of the stationary distribution evaluated at $t = 0$ is given by

(4.5) $$\dot{\pi}(0) = \pi(0)FA_0^{\#}.$$

The derivative $\dot{\pi}_i(t)$ of the i-th stationary probability at $t = 0$ is

(4.6) $$\dot{\pi}_i(0) = \pi(0)F[A_0^{\#}]_i$$

where $[A_0^{\#}]_i$ is the i-th column of $A_0^{\#}$. For any Hölder vector norm and compatible matrix norm,

(4.7) $$|\dot{\pi}_i(0)| \leq \|F[A_0^{\#}]_i\| \leq \|F\| \, \|[A_0^{\#}]_i\|.$$

Another important point to be made concerning Theorem 4.1 and its corollary is the fact that neither t nor F is required to be "small." The formulas for $\dot{\pi}$ in (4.1) as well as those in (4.5) - (4.7) are global in the sense that they hold for all t and F for which $P(t) = P_0 + tF$ represents an irreducible transition matrix. However, if either t or F is small enough in magnitude to insure that $\|tFA_0^{\#}\| < 1$ for compatible vector and matrix norms such that $\|I\| < 1$, then

$$(I - tFA_0^{\#})^{-1} = \sum_{k=0}^{\infty} (tFA_0^{\#})^k$$

so that taking norms in (4.1) produces the following corollary.

COROLLARY: If $\|tFA_0^{\#}\| < 1$, then

(4.8) $$\|\dot{\pi}\| \leq \|\pi\| \frac{\|FA_0^{\#}\|}{1 - t\|FA_0^{\#}\|}$$

Furthermore, if $t\|F\|\|A_0^{\#}\| \leq 1$, then

(4.9) $$\frac{\|\dot{\pi}\|}{\|\pi\|} \leq \frac{\frac{\|F\|}{\|A_0\|}\|A_0\| \|A_0^{\#}\|}{1 - t\frac{\|F\|}{\|A_0\|}\|A_0\| \|A_0^{\#}\|} = \frac{\frac{\|F\|}{\|A_0\|}\kappa(A_0)}{1 - t\frac{\|F\|}{\|A_0\|}\kappa(A_0)}$$

where $\kappa(A_0) = \|A_0\|\|A_0^{\#}\|$.

The expression (4.9) is a continuous counterpart of the discrete formula given by Meyer in [10].

The results of sections 3 and 4 make it absolutely clear that for a finite homogeneous ergodic Markov chain, the sensitivity of the stationary probabilities are directly governed by the entries of the $A'^{\#}$ matrix. There appears to be ample evidence to support the use of $A'^{\#}$ as the fundamental quantity in gauging the "condition of a finite Markov chain" and it seems apparent that any perturbation or sensitivity analysis of a finite Markov chain should revolve around the matrix $A'^{\#}$.

In dealing with almost any aspect of a finite chain, the entries of $A^{\#}$ seem to be relevant, regardless of whether these quantities are used explicitly or whether they appear only implicitly being incorporated in different terms or notations.

In one way or another the $A^{\#}$ matrix always seems to be present.

5. UTILIZING A QR FACTORIZATION

The utility of orthogonal triangularitation is well documented in the vast literature on matrix computations. The purpose of this section is to demonstrate how one can use a QR factorization of $A = I - P$ to not only compute the stationary distribution π, but to also gain some insight into the relative sensitivity which π is expected to exhibit.

For every nxn irreducible row stochastic matrix P, it is well known that $A = I - P$ must have Rank(A) $= n - 1$. Moreover, any subset of $n - 1$ columns from A is linearly independent. There is "essentially" a unique QR factorization A. The R-factor is uniquely determined by A and the Q-factor is unique up to the algebraic sign of the last column.

THEOREM 5.1: If $A_{n\times n}$ is as described in the previous sections and if $A = QR$ is a QR factorization of A, then R must have the form

$$(5.1)\qquad R = \left[\begin{array}{c|c} U & -Ue \\ \hline 0 & 0 \end{array}\right]$$

where U is a nonsingular upper triangular $(n-1)\times(n-1)$ matrix and e is the column of 1's. The stationary distribution π can be recovered from the last column, q, of Q as

$$(5.2)\qquad \pi = \frac{q^T}{\sum_{j=1}^{n} q_j}$$

PROOF: To prove that R has the form (5.1), we need to show $r_{nn} = 0$. Let e be the column of all 1's and use the fact that $0 = Ae = QRe$ to obtain

$$Re = 0.$$

This together with the fact that R is upper triangular guarantees that $r_{nn} = 0$. The fact that U is nonsingular now follows by noting that

$$\text{Rank}(U) = \text{Rank}(R) = \text{Rank}(QR) = \text{Rank}(A) = n-1.$$

To see that the stationary probabilities can be obtained from the last column of Q, recall that $A = I - P$ where P is a nonnegative irreducible matrix with spectral radius 1. One consequence of the Perron-Frobenius theorem is that if $x^TA = 0$, then $x^T > 0$ or $x^T < 0$. Moreover, the system

$$(5.3) \qquad x^TA = 0, \quad x^T > 0, \quad \|x^T\|_1 = 1$$

possesses a unique solution for x^T. Since the last row of $R = Q^TA$ is zero, it is clear that

$$0 = q^TA$$

where q is the last column of Q and hence $q^T > 0$ or $q^T < 0$. Thus

$$q^T / \sum_{j=1}^{n} q_j$$

satisfies (5.3). Since the stationary distribution also satisfies (5.3), it must be the case that

$$\pi = \frac{q^T}{\sum_{j=1}^{n} q_j} \quad . \qquad \blacksquare$$

If the last column of a Q-factor produces the stationary distribution π, of what relevance, if any, is the remaining information in the Q-factor and R-factor? It is demonstrated below how to use the QR-factors to gauge the inherent sensitivity which π can exhibit to small perturbations in P. To see how this is accomplished, recall from the previous sections that the sensitivity of the i-th stationary probability π_i is directly governed by the magnitude of the entries in the i-th column of $A^\#$. The goal therefore is to write $A^\#$ in terms of the QR-factors. Since A is orthogonally similar to RQ via

$$A = Q(RQ)Q^T,$$

it easily follows that

$$A^\# = Q(RQ)^\# Q^T.$$

(See Campbell and Meyer [2]). If we adopt the Z-norm, then

$$\|A^\#\|_2 = \|(RQ)^\#\|_2 \tag{5.4}$$

and the objective shifts to studying the structure of RQ and measuring the magnitude of its group inverse.

Unfortunately, group inversion is somewhat different than other familiar inversion processes in that $(RQ)^\# \neq Q^T R^\#$. Moreover, there is no known way to directly unravel the term $(RQ)^\#$ so as to directly gauge its magnitude. However, the special structure of our matrix $A = I - P$ does lend itself to an analysis by block decomposition.

By nature of the QR factors for A, the matrix RQ must be of the form

$$(5.5)\qquad RQ = \left[\begin{array}{c|c} U & -Ue \\ \hline 0 & 0 \end{array}\right]\left[\begin{array}{c|c} Q_1 & c \\ \hline d^T & \alpha \end{array}\right] = \left[\begin{array}{c|c} U(Q_1 - ed^T) & U(c - \alpha e) \\ \hline 0 & 0 \end{array}\right] = \left[\begin{array}{c|c} V & b \\ \hline 0 & 0 \end{array}\right]$$

Since $A = QR$ has rank $n - 1$, and since QR is similar to RQ, it follows that $\mathrm{Rank}(RQ) = n - 1$. We know from Theorem 5.1 that U is nonsingular so that it must be the case in (5.5) that $V = U(Q_1 - ed^T)$ as well as $(Q_1 - ed^T)$ is nonsingular. It is well known (see Campbell and Meyer [2]) that for a nonsingular V,

$$\left[\begin{array}{c|c} V & b \\ \hline 0 & 0 \end{array}\right]^{\#} = \left[\begin{array}{c|c} V^{-1} & V^{-2}b \\ \hline 0 & 0 \end{array}\right]$$

and hence $(RQ)^{\#}$ is derived from (5.5) as

$$(5.6)\qquad (RQ)^{\#} = \left[\begin{array}{c|c} (Q_1 - ed^T)^{-1}U^{-1} & (Q_1 - ed^T)^{-1}U^{-1}(Q_1 - ed^T)^{-1}(c - \alpha e) \\ \hline 0 & 0 \end{array}\right].$$

We could, of course, stop here and simply say that $\|A^{\#}\|_2$ (and hence the condition of the associated Markov chain) is essentially that of $\|V^{-1}\|_2$ and advocate estimating $\mathrm{Cond}_2(V)$ in order to gauge the sensitivity of the underlying chain. However, it is apparent from (5.6) that the magnitude of $(RQ)^{\#}$ (and hence of $A^{\#}$) is determined by <u>two</u> inverses; namely

(5.7) $\qquad (Q_1 - ed^T)^{-1};$ $\qquad$ which is dependent solely on Q

and

(5.8) $\qquad U^{-1},$ $\qquad$ which is dependent solely on R.

Intuitively, one feels that the influence of Q in $\|(RQ)^{\#}\|_2$ should somehow be insignificant. The next section explores to what extent this is true.

6. CAN THE ENTRIES IN Q BE NEGLECTED IN ESTIMATING $\|A^{\#}\|_2$?

The contribution from Q in $\|A^{\#}\|_2$ is manifested in the matrix of (5.7). Although the upper left hand $(n-1)\times(n-1)$ block Q_1 in a general orthogonal matrix

$$Q = \left[\begin{array}{c|c} Q_1 & c \\ \hline d^T & \alpha \end{array}\right]$$

is not necessarily invertible, the special structure of $A = I - P$ forces Q_1 to be nonsingular when Q is the Q-factor in a QR-factorization for A.

LEMMA 6.1: If A = QR is a QR-factorization for A = I - P, then the $(n-1)\times(n-1)$ matrix Q_1 in

$$\left[\begin{array}{c|c} Q_1 & c \\ \hline d^T & \alpha \end{array}\right]$$

is nonsingular.

PROOF: Recall from Theorem 5.1 that R has the form

$$\left[\begin{array}{c|c} U & -Ue \\ \hline 0 & 0 \end{array}\right]$$

where U is $(n-1)\times(n-1)$ and is nonsingular. Thus

$$A = \left[\begin{array}{c|c} A_1 & A_2 \\ \hline A_3 & A_4 \end{array}\right] = QR = \left[\begin{array}{c|c} Q_1 U & -Q_1 Ue \\ \hline d^T & -d^T Ue \end{array}\right]$$

implies $Q_1 = A_1 U^{-1}$. Recall from the ciiscussion at the beginning of Section 5 that A_1 must be nonsingular and conclude that Q_1 is nonsingular. ∎

Using the familiar formula for the inverse of a rank 1 update, we obtain that

$$S = (Q_1 - ed^T)^{-1} = Q_1^{-1} + \frac{Q_1^{-1} ed^T Q_1^{-1}}{1 - d^T Q_1^{-1} e} . \tag{6.1}$$

If it could be guaranteed that Q_1 is such that $\|Q_1^{-1}\|_2$ is not unduly large and $|1 - d^T Q_1^{-1} e| >> 0$, then we could disregard the action of Q. Unfortunately, one can exhibit ergodic Markov chains in which Q_1^{-1} can possess entries of arbitrarily large magnitude. For example, consider the chain whose transition matrix is

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

In this case,

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

and the QR factors are of the form

$$Q = \left[\begin{array}{cc|c} \frac{1}{\sqrt{1+\varepsilon^2}} & \frac{\varepsilon^2}{\sqrt{1+3\varepsilon^2+2\varepsilon^4}} & \frac{\varepsilon}{\sqrt{1+2\varepsilon^2}} \\ \frac{-\varepsilon}{\sqrt{1+\varepsilon^2}} & \frac{\varepsilon}{\sqrt{1+3\varepsilon^2+2\varepsilon^4}} & \frac{1}{\sqrt{1+2\varepsilon^2}} \\ \hline 0 & \frac{-1-\varepsilon^2}{\sqrt{1+3\varepsilon^2+2\varepsilon^4}} & \frac{\varepsilon}{\sqrt{1+2\varepsilon^2}} \end{array}\right]$$

$$R = \left[\begin{array}{cc|c} \sqrt{1+\varepsilon^2} & \frac{-1-2\varepsilon^2}{\sqrt{1+\varepsilon^2}} & \frac{\varepsilon^2}{\sqrt{1+\varepsilon^2}} \\ 0 & \frac{\sqrt{1+3\varepsilon^2+2\varepsilon^4}}{1+\varepsilon^2} & \frac{-\sqrt{1+3\varepsilon^2+2\varepsilon^4}}{1+\varepsilon^2} \\ \hline 0 & 0 & 0 \end{array}\right]$$

The leading principal submatrix Q_1 approaches singularity as $\varepsilon \to 0$ and hence $\|Q_1^{-1}\|_2 \to \infty$ as $\varepsilon \to 0$.

Although the Q_1 matrix may be badly conditioned, one might still hope that $Q_1 - ed^T$ is better conditioned. The following theorem establishes that this is indeed true.

THEOREM 6.1: For $A = I - P$ where P is an $(n \times n)$ irreducible stochastic matrix, let

$$Q = \left[\begin{array}{c|c} Q_1 & c \\ \hline d^T & \alpha \end{array}\right]$$

be a Q-factor from a QR decomposition of A and let e denote the column of 1's. If $S = (Q_1 - ed^T)^{-1}$ and $v = (c - \alpha e)$, then.

$$\|S\|_2 = \|(Q_1 - ed^T)^{-1}\|_2 \leq 1 + \sqrt{n-1} \tag{6.2}$$

and

$$\|Sv\|_2 = \|(Q_1 - ed^T)^{-1}(c - \alpha e)\|_2 < 1 + \sqrt{n-1} \;. \tag{6.3}$$

PROOF: We will first demonstrate that

$$S = (Q_1 - ed^T)^{-1} = Q_1^T - \frac{(Q_1^T e + d)c^T}{c^T e + \alpha} \tag{6.4}$$

and

$$Sv = S(c - \alpha e) = -\frac{Q_1^T e + d}{c^T e + \alpha} \quad . \tag{6.5}$$

It is not easy to verify that these are true by straightforward multiplication and hence an indirect approach is taken,. If β denotes the scalar

$$\beta = c^Te + \alpha$$

and if M denotes the matrix

$$M = \left[\begin{array}{c|c} Q_1 - ed^T & c - \alpha e \\ \hline 0 & 1 \end{array}\right] = \left[\begin{array}{c|c} I & -e \\ \hline c^T & \alpha \end{array}\right] \left[\begin{array}{c|c} Q_1 & c \\ \hline d^T & \alpha \end{array}\right],$$

then

$$M^{-1} = \left[\begin{array}{c|c} (Q_1 - ed^T)^{-1} & -(Q_1 - ed^T)^{-1}(c - \alpha e) \\ \hline 0 & 1 \end{array}\right] = \left[\begin{array}{c|c} Q_1^T & d \\ \hline c^T & \alpha \end{array}\right] \left[\begin{array}{c|c} I & -e \\ \hline c^T & \alpha \end{array}\right]^{-1} \tag{6.6}$$

$$= \left[\begin{array}{c|c} Q_1^T & d \\ \hline c^T & \alpha \end{array}\right] \left[\begin{array}{c|c} I - \frac{ec^T}{\beta} & \frac{e}{\beta} \\ \hline \frac{-c^T}{\beta} & \frac{1}{\beta} \end{array}\right] = \left[\begin{array}{c|c} Q_1^T - \frac{(Q_1^Te + d)c^T}{\beta} & \frac{Q_1^Te + d}{\beta} \\ \hline 0 & 1 \end{array}\right],$$

which proves (6.4) - (6.5). The desired conclusion follows by taking norms. Since Q_1^T and c^T are submatrices of an orthogonal matrix, $\|Q_1^T\|_2 \leq 1$ and $\|c^T\|_2 \leq 1$ so that

(6.7) $$\|(Q_1 - ed^T)^{-1}\|_2 = \left\|Q_1^T - \frac{(Q_1^T e + d)c^T}{\beta}\right\|_2 \le 1 + \left\|\frac{Q_1^T e + d}{\beta}\right\|_2 \quad .$$

Taking Z-norms on

$$Q^T \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} = \begin{bmatrix} Q_1^T e + d \\ \\ \beta \end{bmatrix}$$

produces

$$n = \|Q_1^T e + d\|_2^2 + \beta^2$$

so that

(6.8) $$\left\|\frac{Q_1^T e + d}{\beta}\right\|^2 = \frac{n}{\beta^2} - 1 \quad .$$

From Theorem 5.1, we have

$$\pi = \frac{1}{\beta}(c^T, \alpha)$$

so that

$$\beta = \|(c^T, \alpha)\|_1 \ge \|(c^T, \alpha)\|_2 = 1 \ .$$

Apply this to (6.8) to obtain

$$(6.9) \qquad \left\| \frac{Q_1^T e + d}{\beta} \right\|_2^2 \le n - 1.$$

Use (6.9) in (6.7) to arrive at

$$\|(Q_1 - ed^T)^{-1}\|_2 \le 1 + \sqrt{n-1} \; .$$

Similarly, formula (6.5) produces (6.3). ∎

Since $\|A^\#\|_2 = \|(RQ)^\#\|_2$, one can now look at the form for $(RQ)^\#$ in (5.6) and see that the Q matrix does not significantly contribute to the size of $\|A^\#\|_2$. The only way for $\|A^\#\|_2$ to be large is for $\|U^{-1}\|_2$ to be large. This is formalized below.

THEOREM 6.2: Let $A = I - P$ where P is an $(n\times n)$ irreducible stochastic matrix and let $A = QR$ be a QR-factorization for A. If U is the $(n-1\times n-1)$ leading principal submatrix of R, then

$$(6.10) \qquad \|A^\#\|_2 \le O(n)\|U^{-1}\|_2.$$

PROOF: From (5.4), we know that $\|A^\#\|_2 = \|(RQ)^\#\|_2$ and from (5.6) we know that

$$(RQ)^\# = \left[\begin{array}{c|c} SU^{-1} & SU^{-1}Sv \\ \hline 0 & 0 \end{array}\right]$$

where $S = (Q_1 - ed^T)^{-1}$ and $v = C - ae$. From Theorem 6.1, it follows that

$$\|(RQ)^{\#}\|_2 \le \|S\|_2 \|U^{-1}\|_2 \sqrt{(1 + \|Sv\|_2^2)}$$

$$= O(n)\|U^{-1}\|_2. \quad \blacksquare$$

Therefore, the condition number of the $(n-1 \times n-1)$ leading principal submatrix of R may be taken as an estimate (or as a measure in itself) of the condition of the underlying chain. Since U is uppertriangular with positive diagonals, estimating $\mathrm{Cond}_2(U)$ is not overly difficult (e.g., LINPACK methods can be used.)

7. CONCLUSIONS

For an ergodic chain with transition matrix P, a QR factorization of the matrix $A = I - P$ yields complete information in the sense that both the stationary distribution π as well as measure of the sensitivity of π to perturbations in P may be deduced.

1. π is obtained by normalizing the last column of Q.
2. The sensitivity of the chain may be gauged by Cond(U) where U is the $(n-1 \times n-1)$ leading principal submatrix of R.

In general, it is well known that an upper triangular matrix may be ill-conditioned without possessing relatively small diagonal elements. However, for the special situation of an irreducible Markov chain, we have not been able to produce an example of an ergodic chain so that the factorization $A = QR$ yields an R in which $\|U^{-1}\|$ is large but U has no small diagonals. In all of our computational experience, the sensitive chains always seem to force a diagonal entry of U to be relatively small. The more sensitive the chain, the smaller some diagonal of U becomes, so it seems. There is clearly need for further study.

REFERENCES

[1] A. Brandt, "Algebra's Multigrid Theory," preprint, (1983).

[2] S. L. Campbell and C. D. Meyer, "Generalized Inverses of Linear Transformations," Surveys and Reference Works in Mathematics, Pitman, London, (1979)

[3] 3. Conslisk, "Comparative Statics for Markov Chains," preprint, (1983).

[4] R. E. Funderlic and M. T. Heath, "Linear Compartmental Analysis of Ecosystems," ORNL/IBP-71/4, Oak Ridge National Laboratory, Oak Ridge, Tennessee, (1971).

[S] R. E. Funderlic and 3. B. Mankin, "Solution of Homogeneous Systems of Equations Arising from Compartmental Models," SIAM J. Sci. Stat. Comp., 2, 375-413, (1981).

[6] 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, 461-474, (1983). ,

[7] L. Kaufman, "Matrix Methods for Queuing Problems," SIAM J. Sci. Stat. Comp., 4, 525-552, (1984).

[8] 3. G. Kemeny and 3. L. Snell, Finite Markov Chains, Van Nostrand, New York, (1960).

[9] C. D. Meyer, "The Role 'of the Group Generalized Inverse in the Theory of Finite Markov Chains," SIAM Review, 17, 443-464, (1975).

[10] C. D. Meyer, "The Condition of a Finite Markov Chain and Perturbation Bounds for the Limiting Probabilities," SIAM 3. Alg. Disc. Meth., 1, 273-283, (1980).

[11] P. 3. Schweitzer, "Perturbation Theory and Finite Markov Chains," J. Appl. Prob., 5, 401-413, (1968).