Ann. Inst. Statist. Math.
37 (1985), Part B, 173-187

THE CONJUGATE GRADIENT METHOD FOR COMPUTING ALL THE EXTREMAL STATIONARY PROBABILITY VECTORS OF A STOCHASTIC MATRIX

KUNIO TANABE

(Received Apr. 8, 1981; revised Aug. 7, 1984)

Summary

The conjugate gradient method is developed for computing stationary probability vectors of a large sparse stochastic matrix P, which often arises in the analysis of queueing system. When unit vectors are chosen as the initial vectors, the iterative method generates all the extremal probability vectors of the convex set formed by all the stationary probability vectors of P, which are expressed in terms of the Moore-Penrose inverse of the matrix $(P-I)$. A numerical method is given also for classifying the states of the Markov chain defined by P. One particular advantage of this method is to handle a very large scale problem without resorting to any special form of P.

1. Introduction

We consider the problem of computing the stationary probability vector $\alpha=(a_1, a_2, \cdots, a_m)$ of an $m\times m$ stochastic matrix $P=(p_{ij})$ such that

$$(1) \qquad \alpha P=\alpha ,$$

where $a_i\geqq 0$, $\sum_{i=1}^{m} a_i=1$, $p_{ij}\geqq 0$ and $\sum_{j=1}^{m} p_{ij}=1$ for $i=1, 2, \cdots, m$. The set $C(P)$ of all the stationary probability vectors of P forms a convex set in the linear space of m-dimensional row vectors. It is well-known that if P is a regular or a cyclic ergodic transition matrix then there exists an unique positive probability vector α which satisfies Eq. (1), $a_i>0$ and $\sum_{i=1}^{m} a_i=1$ [2, 6].

For any initial probability vector α_0 we have

Key words and phrases: Conjugate gradient method, extremal stationary probability vector, finite Markov chain, fixed point, Moore-Penrose inverse, stochastic matrix.

(2) $$\lim_{n\to\infty} \alpha_0 P^n = \alpha$$

and

(3) $$\lim_{n\to\infty} \alpha_0 (hI+(1-h)P)^n = \alpha \qquad \text{for } 0<h<1\,,$$

respectively when P is regular and when P is periodic. These equalities suggest iterative methods for computing α. When P has the second eigenvalue which is close to one in absolute value the convergence of these methods is tediously slow. For the cyclic ergodic case, Odell and Decell [5] proposed a method which uses the Moore-Penrose inverse $(I-P)^{\dagger}$ of submatrices of $P-I$, based on the following result [1].

THEOREM 1 (Decell and Odell). *If P is an ergodic (i.e. irreducible) transition matrix and*

(1) $$\alpha = \frac{\mathbf{1}^t(I-(P-I)(P-I)^{\dagger})}{\mathbf{1}^t(I-(P-I)(P-I)^{\dagger})\mathbf{1}}\,,$$

then α is the unique stationary probability vector of P, where $\mathbf{1}=(1, 1, \cdots, 1)^t$.

When P is of large order and has not the cyclic partitioned form, a fairly large amount of computation seems to be required in their method. In this paper we develop the conjugate gradient algorithm for computing all the extremal stationary probability vectors which span the convex set $C(P)$ of a large sparse stochastic matrix P. In § 2 an iterative algorithm is introduced, its behavior is analysed and the extremal stationary probability vectors are characterized in terms of the Moore-Penrose inverse of $(P-I)$. In § 3 a numerical method is given for classifying the states of the Markov chain defined by P. Our method can handle a large scale problem without resorting to any special form of P such as required in the method [5]. This is a great advantage when P is a very large sparse stochastic matrix.

In the following a vector β is said to be positive if every entry is positive, semi-positive if every entry is non-negative and $\beta\neq 0$ and semi-negative if every entry is non-positive and $\beta\neq 0$, which will be expressed by writing $\beta>0$, $\beta\geqq 0$ and $\beta\leqq 0$ respectively.

We shall use the following notations.

Im A : The linear subspace spannes by the row vectors of a matrix A, i.e., $\{\gamma A; \gamma\}$.

Ker A: The linear subspace spanned by the row vectors annihilated by A, i.e., $\{\gamma; \gamma A=0\}$.

A^t : The transpose of A.

$A^\dagger$: The Moore-Penrose inverse of A defined by $AA^\dagger A=A$, $A^\dagger AA=A^\dagger$, $(AA^\dagger)^t=AA^\dagger$ and $(A^\dagger A)^t=A^\dagger A$.

$\langle\alpha, \beta\rangle$: The inner product of two vectors $\alpha=(a_1, a_2, \cdots, a_m)$ and $\beta=(b_1, b_2, \cdots, b_m)$, defined by $\langle\alpha, \beta\rangle=\sum_{i=1}^{m} a_i b_i$.

$\|\alpha\|_2$: The norm of a vector α, defined by $\|\alpha\|_2^2=\langle\alpha, \alpha\rangle$.

$P_{\mathcal{S}}$: The orthogonal projector onto the subspace $\mathcal{S}$, defined by $P_{\mathcal{S}}^2=P_{\mathcal{S}}$, $P_{\mathcal{S}}^t=P_{\mathcal{S}}$, $\operatorname{Im} P_{\mathcal{S}}=\mathcal{S}$ and $\operatorname{Ker} P_{\mathcal{S}}=\mathcal{S}^{\perp}$.

$|\alpha|$: The vector defined by $|\alpha|=(|\alpha_1|, |\alpha_2|, \cdots, |\alpha_m|)$.

$\|\alpha\|_1$: The norm of a vector, defined by $\|\alpha\|_1=\sum_{i=1}^{m}|\alpha_i|$.

$\mathcal{E}\cup\mathcal{T}$: The join of two sets $\mathcal{E}$ and $\mathcal{T}$.

$\mathcal{E}-\mathcal{T}$: The intersection of $\mathcal{E}$ and the complement of $\mathcal{T}$.

2. Conjugate gradient algorithms for computing stationary probability vectors

We shall introduce an algorithm for computing the stationary probability vectors, which is based on the following theorem established by Kammerer and Nashed [4].

THEOREM 2. *The conjugate gradient process*

$$\begin{aligned} & r_0=b-Ax_0 \\ & p_0=A^t r_0 \\ & u_i=\|A^t r_i\|_2^2/\|Ap_i\|_2^2 \\ & x_{i+1}=x_i+u_i p_i \qquad (i=0, 1, \cdots) \\ & r_{i+1}=b-Ax_{i+1}=r_i-u_i A p_i \\ & v_i=\|A^t r_{i+1}\|_2^2/\|A^t r_i\|_2^2 \\ & p_{i+1}=A^t r_{i+1}+v_i p_i \end{aligned} \tag{5}$$

gives the least squares solution

$$\hat{x}=A^\dagger b+(I_n-A^\dagger A)x_0 \tag{6}$$

of the linear least squares problem

$$\min_x \|b-Ax\|_2^2, \tag{7}$$

where A is a matrix, b and r are column vectors, and x_0 is the initial column vector of the process (5).

Noting that the stationary probability vectors α are solutions of the system of linear equations

$$(P^t - I)\alpha^t = 0 , \tag{8}$$

we apply the process (5) to this system to obtain the following algorithm for computing the fixed point α.

CG ALGORITHM. Starting with β_0, generate a sequence $\{\beta_i\}$ of m-dimensional row vectors β_i by the recurrence formula

$$\begin{aligned}
&\rho_0 = \beta_0(I - P)\\
&\pi_0 = \rho_0(P^t - I)\\
&u_i = \|\rho_i(P^t - I)\|_2^2 / \|\pi_i(P - I)\|_2^2\\
&\beta_{i+1} = \beta_i + u_i\pi_i\\
&\rho_{i+1} = \beta_{i+1}(I - P) = \rho_i - u_i\pi_i(P - I)\\
&v_i = \|\rho_{i+1}(P^t - I)\|_2^2 / \|\rho_i(P^t - I)\|_2^2\\
&\pi_{i+1} = \rho_{i+1}(P^t - I) + v_i\pi_i
\end{aligned} \tag{9}$$

stopping if ρ_i vanishes, where β_i, ρ_i and π_i are m-dimensional row vectors.

THEOREM 3. *The sequence $\{\beta_i\}$ generated by the algorithm (9) converges to the vector*

$$\tilde{\beta} = \beta_0(I - (P - I)(P - I)^\dagger) = \beta_0 P_{\mathrm{Ker}(P-I)} , \tag{10}$$

which satisfies

$$\tilde{\beta}P = \tilde{\beta} \tag{11}$$

or equivalently

$$\tilde{\beta} \in \mathrm{Ker}\,(P - I) . \tag{12}$$

PROOF. Noting that A, x, and b in (7) correspond to $P^t - I$, α^t and 0 in (8) respectively and that $(P^t - I)^\dagger = ((P - I)^\dagger)^t$, we can easily deduce the result from Theorem 2.

PROPOSITION 4. *If P is an irreducible transition matrix in the previous theorem then we have*

$$\tilde{\beta} = \beta_0((\alpha^t\alpha)/\|\alpha\|_2^2) = (\langle\beta_0, \alpha\rangle/\|\alpha\|_2^2)\alpha . \tag{13}$$

Further we have $\tilde{\beta} = (\tilde{b}_1, \tilde{b}_2, \cdots, \tilde{b}_m) \neq 0$, iff,

$$\beta_0 \notin \mathrm{Im}\,(P^t - I) , \tag{14}$$

in which case the vector,

$$\alpha \equiv \left(\sum_{i=1}^{m} \tilde{b}_i\right)^{-1} \tilde{\beta} = \|\tilde{\beta}\|_1^{-1} |\tilde{\beta}| \tag{15}$$

is the unique stationary probability vector of P.

PROOF. Under the condition, the dimension of the subspace $\operatorname{Ker}(P-I)$ is one and the subspace is spanned by the stationary probability vector α. Hence $P_{\operatorname{Ker}(P-I)} = (\alpha^t \alpha)/\|\alpha\|_2^2$ and Eq. (13) follows directly from Eq. (10). Since

$$\operatorname{Ker}(P-I)^{\perp} = \operatorname{Im}(P^t - I),$$

the condition (14) implies $\tilde{\beta} \neq 0$, in which case it follows from the form (13) of $\tilde{\beta}$ that either $\tilde{\beta} > 0$ or $\tilde{\beta} < 0$ is satisfied. Hence $\sum_{i=1}^{m} \tilde{b}_i \neq 0$ and α is given by the formula (15).

COROLLARY 1. *Under the same condition, if we put either $\beta_0 \geqq 0$ or $\beta_0 \leqq 0$ then the condition* (14) *is satisfied, hence we can obtain the stationary probability vector α by the formula* (15).

PROOF. Let $\beta_0 = (b_1^\circ, b_2^\circ, \cdots, b_m^\circ)$. When $\beta_0 \geqq 0$ we have

$$\langle \beta_0, \alpha \rangle = \sum_{i=1}^{m} a_i b_i^\circ > 0,$$

since $a_i > 0$ for any i and there exists an integer j such that $b_j > 0$. Hence it follows from (13) that $\tilde{\beta} \neq 0$. The case where $\beta_0 \leqq 0$ is similarly proved. Now we obtain a generalization of Theorem 1.

COROLLARY 2. *If P is an irreducible transition matrix, the unique stationary probability vector α is given by*

$$\alpha = \frac{\beta_0(I-(P-I)(P-I)^{\dagger})}{\beta_0(I-(P-I)(P-I)^{\dagger})1} \tag{16}$$

for any vector β_0 which satisfies either

$$\beta_0 \geqq 0 \quad \text{or} \quad \beta_0 \leqq 0 .$$

Before proceeding to the discussion of general case we recall the following result [6, 10].

THEOREM 5 (Romanovsky). *Every stochastic matrix P has the following standard form*

$$P = D^t \begin{bmatrix} T_0 & T_1 & T_2 \cdots & T_r \\ & E_1 & & \\ & & E_2 & \\ & & & \ddots \\ & & & & E_r \end{bmatrix} D \tag{17}$$

where D is a permutation matrix, T_0 is a square matrix of order n_0 and E_i's are irreducible transiton matrices of order n_i respectively.

Let $\hat{\alpha}_i$ be the unique positive stationary probability vector of E_i, $i=1, 2, \cdots, r$. Then the m-dimensional semi-positive vectors

$$\begin{aligned} & \quad\;\; n_0 \;\; n_1 \;\; n_2 \;\; \cdots \;\; n_r \\ \alpha_1 &= (0\,;\hat{\alpha}_1\,;0\,;\cdots;0)D\,, \\ \alpha_2 &= (0\,;0\,;\hat{\alpha}_2\,;\cdots;0)D\,, \\ & \cdots\cdots\cdots\cdots\cdots \\ & \cdots\cdots\cdots\cdots\cdots \\ \alpha_r &= (0\,;0\,;0\,;\cdots;\hat{\alpha}_r)D \end{aligned} \tag{18}$$

are mutually orthogonal stationary probability vectors of P. Every α_i has n_i positive entries which correspond to the states in the same ergodic class $\mathcal{E}_i$. Every stationary probability vector of P is a convex combination of $\alpha_1, \alpha_2, \cdots, \alpha_r$.

For a vector $\beta=(b_1, b_2, \cdots, b_m)$ the set of states

$$\mathrm{Supp}\,(\beta) \equiv \{j\,;\, b_j \neq 0\}$$

is said to be the support of β. It should be noted that the supports of α_i, $i=1, 2, \cdots, r$, are mutually disjoint ergodic classes and $\{1, 2, \cdots, m\} - \bigcup_{i=1}^{r} \mathrm{Supp}\,(\alpha_i)$ is the set of transient states of P.

The following theorem is a generalization of Proposition 4.

THEOREM 6. *Let α_i's be defined in* (18) *then the sequence $\{\beta_i\}$ generated by the algorithm* (9) *converges to*

$$\begin{aligned} \tilde{\beta} &= \beta_0 \Big(\sum_{i=1}^{r} (\alpha_i^t \alpha_i)/\|\alpha_i\|_2^2 \Big) \\ &= \sum_{i=1}^{r} (\langle \beta_0, \alpha_i \rangle / \|\alpha_i\|_2^2)\alpha_i\,, \end{aligned} \tag{19}$$

where β_0 is the initial vector of the process (9). *If*

$$(20)\qquad \beta_0 \notin \operatorname{Im}(P^t - I),$$

then $\tilde{\beta}=(\tilde{b}_1, \tilde{b}_2, \cdots, \tilde{b}_m)\neq 0$ *and the semi-positive vector*

$$(21)\qquad \tilde{\alpha}\equiv\left(\sum_{i=1}^{m}|\tilde{b}_i|\right)^{-1}|\tilde{\beta}|=\|\tilde{\beta}\|_1^{-1}|\tilde{\beta}|$$

is a stationary probability vector of P where

$$|\tilde{\beta}|=(|\tilde{b}_1|, |\tilde{b}_2|, \cdots, |\tilde{b}_m|).$$

PROOF. Since E_i's are irreducible stochastic matrices and T_0-I is non-singular, $\operatorname{Ker}(P-I)$ is spanned by the mutually orthogonal vectors $\alpha_1, \alpha_2, \cdots, \alpha_r$. Hence we have

$$(22)\qquad \begin{aligned} I-(P-I)(P-I)^\dagger &= P_{\operatorname{Ker}(P-I)} \\ &= \sum_{i=1}^{r}(\alpha_i^t\alpha_i)/\|\alpha_i\|_2^2 \end{aligned}$$

which, together with Eq. (10) of Theorem 3, implies Eq. (19). Since $\operatorname{Im}(P^t-I)$ is the orthogonal complement of $\operatorname{Ker}(P-I)$, the condition (20) implies $\tilde{\beta}\neq 0$, in which case the vector

$$|\tilde{\beta}|=\sum_{i=1}^{r}(|\langle\beta_0, \alpha_i\rangle|/\|\alpha_i\|_2^2)\alpha_i$$

is also a semi-positive vector such that $|\tilde{\beta}|P=|\tilde{\beta}|$. Hence normalizing it by the factor $\|\tilde{\beta}\|_1=\sum_{i=1}^{r}|\langle\beta_0, \alpha_i\rangle|/\|\alpha_i\|_2^2$ we have a stationary probability vector $\tilde{\alpha}$.

The unit vectors will be denoted by

$$\varepsilon_1=(1, 0, 0, \cdots, 0),$$

$$\varepsilon_2=(0, 1, 0, \cdots, 0),$$

$$\cdots\cdots\cdots\cdots$$

and $$\varepsilon_m=(0, 0, 0, \cdots, 1).$$

Now we have the main result.

COROLLARY 1. *Under the condition of the previous theorem if we put* $\beta_0=c\varepsilon_j$ $(c\neq 0)$ *then*

$$(23)\qquad \tilde{\beta}\begin{cases} =0, & \text{if } j\in\{1, 2, \cdots, m\}-\bigcup_{i=1}^{r}\operatorname{Supp}(\alpha_i), \\ \neq 0, & \text{otherwise}. \end{cases}$$

If $\tilde{\beta}\neq 0$, *there exists the unique integer k such that*

$$j \in \operatorname{Supp}(\alpha_k)$$

and

$$\alpha_k = \left(\sum_{i=1}^{m} \tilde{b}_i\right)^{-1} \tilde{\beta} = \|\tilde{\beta}\|_1^{-1} |\tilde{\beta}| \,,$$

where $\tilde{\beta} = (\tilde{b}_1, \tilde{b}_2, \cdots, \tilde{b}_m)$ *and* α_i*'s are defined in* (18).

PROOF. If the state j is transient, $\langle \varepsilon_j, \alpha_i \rangle = 0$ for all i, otherwise there exists the unique extremal stationary probability vector α_k such that $\langle \varepsilon_j, \alpha_k \rangle \neq 0$ since $\operatorname{Supp}(\alpha_i)$'s are mutually disjoint. This implies the desired result, since in this case

$$\tilde{\beta} = c \sum_{i=1}^{r} (\langle \varepsilon_j, \alpha_i \rangle / \|\alpha_i\|_2^2) \alpha_i \,.$$

The following corollary is a restatement of the previous corollary.

COROLLARY 2. *For every* ε_j, $j = 1, 2, \cdots, m$, *either of the following propositions holds.*
1. *The vector*

$$\gamma_j \equiv \frac{\varepsilon_j(I - (P-I)(P-I)^\dagger)}{\varepsilon_j(I - (P-I)(P-I)^\dagger)1} \tag{24}$$

coincides with one of the extremal stationary probability vectors α_i*'s of* P *defined in* (18), *in which case the state* j *belongs to the ergodic class* $\operatorname{Supp}(\gamma_j)$.
2. *The vector*

$$\tilde{\beta}_j \equiv \varepsilon_j (I - (P-I)(P-I)^\dagger) \tag{25}$$

vanishes, in which case the state j *is transient. The Moore-Penrose inverse* $(P-I)^\dagger$ *in the formula* (24) *can be replaced by an arbitrary least square type generalized inverse* $(P-I)_l^-$ *of* $(P-I)$, *that is defined by*

$$\begin{aligned} &(P-I)(P-I)_l^-(P-I) = (P-I) \quad \textit{and} \\ &((P-I)(P-I)_l^-)^t = (P-I)(P-I)_l^- \,. \end{aligned} \tag{26}$$

COROLLARY 3. *The probability vector*

$$\tilde{\alpha} \equiv \frac{1^t(I - (P-I)(P-I)^\dagger)}{1^t(I - (P-I)(P-I)^\dagger)1} \tag{27}$$

is a convex combination of all the extremal stationary probability vectors α_i*'s such that*

$$\tilde{\alpha} = \sum_{i=1}^{r} c_i \alpha_i \tag{28}$$

where $c_i=\|\alpha_i\|_2^{-2}\Big/\Big(\sum_{i=1}^{r}\|\alpha_i\|_2^{-2}\Big)$. *We also have the inequalities*

$$m\geqq\|\mathbf{1}^t(I-(P-I)(P-I)^\dagger\|_1=\sum_{i=1}^{r}\|\alpha_i\|_2^{-2}\geqq r\,. \tag{29}$$

In this case Supp$(\tilde{\alpha})$ *is the set of all the ergodic states of* P.

PROOF. The inequalities in (29) are easily deduced respectively from the Schwartz inequality, $\|\alpha_i\|_2^2 n_i\geqq\|\alpha_i\|_1^2=1$, $(i=1,\cdots,r)$, and from the inequality, $\|\alpha_i\|_2^2\leqq\|\alpha_i\|_1=1$, $(i=1,\cdots,r)$.

It should be noted here that given a state j, using the algorithm (9) we can determine whether the state is recurrent or transient and if it is recurrent we can determine the ergodic class to which it belongs. Hence theoretically we can completely classify the states of an arbitrary stochastic matrix P in the following way.

CONCEPTUAL ALGORITHM.

1) Put $\mathcal{U}=\{1,2,\cdots,m\}$ and $l=0$. $\mathcal{U}$ is a set of unlabeled states.
2) Draw a state, say j, from $\mathcal{U}$.
3) Compute $\tilde{\beta}_j$, putting $\beta_0=\varepsilon_j$ in the algorithm (9).
4) If $\tilde{\beta}_j=0$ then the state j is transient, so label it 'T' and eliminate it from $\mathcal{U}$, otherwise Supp$(\tilde{\beta}_j)$ is the ergodic class of P to which the state j belongs, so set $l:=l+1$, label the states in Supp(γ_j) '$\mathcal{E}_l$' and eliminate them from $\mathcal{U}$.
5) If $\mathcal{U}$ is empty the classification is completed otherwise return to 2).

Since only a numerical approximation $\bar{\beta}_j$ to $\tilde{\beta}_j$ is obtained practically one might suspect that we can neither check whether $\tilde{\beta}_j$ vanishes or not, nor determine the set Supp$(\tilde{\beta}_j)$ accurately. In the next section we will give a practical algorithm to overcome the difficulty.

3. Practical algorithm

Before introducing an algorithm for computing extremal stationary probability vectors and classifying the states of a stochastic matrix P, we give some results which will be used in the algorithm.

Let P^B be the Boolian matrix of zeros and ones defined by

$$P^B=(p_{i,j}^B)\,,\qquad \text{where } p_{i,j}^B=\begin{cases}1 & \text{if } p_{i,j}>0\\ 0 & \text{if } p_{i,j}=0\,.\end{cases} \tag{30}$$

Given a vector $\xi=(x_1,x_2,\cdots,x_m)$, let ξ^B be the Boolian vector defined by

(31) $$\xi^B=(x_1^B, x_2^B, \cdots, x_m^B), \quad \text{where } x_i^B=\begin{cases} 1 & \text{if } x_i>0 \\ 0 & \text{if } x_i\leqq 0 . \end{cases}$$

Let $\oplus$ and $\otimes$ be denote the Boolian addition and multiplication, where the product $\xi^B\otimes P^B\equiv(z_1, z_2, \cdots, z_m)$ of ξ^B and P^B is defined by

(32) $$z_i=x_1^B\otimes p_{1,i}^B\oplus x_2^B\otimes p_{2,i}^B\oplus\cdots\oplus x_m^B\otimes p_{m,i}^B .$$

Given a set $\mathcal{R}$ of states, let $\psi_{\mathcal{R}}\equiv(f_1, f_2, \cdots, f_m)$ be the Boolian vector defined by

(33) $$f_i=\begin{cases} 1 & \text{if } i\in\mathcal{R} \\ 0 & \text{if } i\notin\mathcal{R} . \end{cases}$$

We will identify $\psi_{\mathcal{R}}$ with the vector $\sum_{i\in\mathcal{R}} \varepsilon_i$. Given a set $\mathcal{R}$ of states and a vector $\beta=(b_1, b_2, \cdots, b_m)$, let $\beta|_{\mathcal{R}}=(c_1, c_2, \cdots, c_m)$ be the vector defined by

(34) $$c_i=\begin{cases} b_i & \text{if } i\in\mathcal{R} \\ 0 & \text{if } i\notin\mathcal{R} . \end{cases}$$

Given a vector $\beta=(b_1, b_2, \cdots, b_m)$, let $|\beta|^+=(c_1, c_2, \cdots, c_m)$ be the vector defined by

(35) $$c_i=\begin{cases} b_i & \text{if } b_i>0 \\ 0 & \text{if } b_i\leqq 0 . \end{cases}$$

The following two statements are easily seen.

PROPOSITION 7. *If $\mathcal{R}$ is a nonempty subset of some ergodic class of the states of P, i.e., $\mathcal{R}\subset\mathcal{E}_i\equiv\mathrm{Supp}\,(\alpha_i)$ for some i, then*

(36) $$\psi_{\mathcal{E}_i}=\psi_{\mathcal{R}}\oplus\psi_{\mathcal{R}}\otimes P^B\oplus\psi_{\mathcal{R}}\otimes P^B\otimes P^B\oplus\cdots\oplus\psi_{\mathcal{R}}\otimes\underbrace{P^B\otimes\cdots\otimes P^B}_{k} ,$$

for some k. If $\mathcal{R}$ coincides with one of the ergodic class of the states of P then

(37) $$\psi_{\mathcal{R}}\otimes P^B=\psi_{\mathcal{R}} .$$

PROPOSITION 8. *Let the maximum and the minimum values of the elements g_i's of a vector $\gamma=(g_1, g_2, \cdots, g_m)$ be denoted by $\max\gamma$ and $\min\gamma$ respectively. Then for any m-dimensional positive probability vector γ we have*

(38) $$\max\gamma\geqq 1/m\geqq\min\gamma \quad \textit{and}$$

(39) $$m \geqq \|\gamma\|_2^{-2} \geqq 1 .$$

PROPOSITION 9. *Let* $\tilde{\beta}_j$ *be defined by* (25), *then if the state* j *is transient,* $\|\tilde{\beta}_j\|_1=0$, *otherwise*

(40) $$\|\tilde{\beta}_j\|_1 \geqq \min \alpha_k , \qquad \textit{for some } k .$$

For any ergodic class $\mathcal{E}_k$ *there exists a state* l *such that*

(41) $$\|\tilde{\beta}_l\|_1 \geqq \max \alpha_k \geqq 1/m .$$

PROOF. Since we have by (19)

$$\tilde{\beta}_j = \sum_{i=1}^{r} (\langle \varepsilon_j, \alpha_i \rangle / \|\alpha_i\|_2^2) \alpha_i ,$$

if j is transient then $\tilde{\beta}_j=0$, otherwise there exists the unique α_k such that

$$\tilde{\beta}_j = (\langle \varepsilon_j, \alpha_k \rangle / \|\alpha_k\|_2^2) \alpha_k ,$$

hence we have

$$\begin{aligned} \|\tilde{\beta}_j\|_1 &= (\langle \varepsilon_j, \alpha_k \rangle / \|\alpha_k\|_2^2) \|\alpha_k\|_1 \\ &= \langle \varepsilon_j, \alpha_k \rangle / \|\alpha_k\|_2^2 \\ &\geqq \min \alpha_k / \|\alpha_k\|_2^2 \\ &\geqq \min \alpha_k . \end{aligned}$$

The inequality (41) is similarly proved. It should be noted that the vector

$$\tilde{\beta} = \alpha_i (I - (P-I)(P-I)^{\dagger})$$

which is generated by the algorithm (9) starting with $\beta_0=\alpha_i$, satisfies

(42) $$\|\tilde{\beta}\|_1 = \|\alpha_i\|_1 = 1$$

and that if $\mathcal{E}_i = \mathrm{Supp}\,(\alpha_i)$ then the vector

$$\tilde{\beta} = \psi_{\mathcal{E}_i} (I - (P-I)(P-I)^{\dagger}) ,$$

which is generated by the algorithm (9) starting with $\beta_0=\psi_{\mathcal{E}_i}$, satisfies

(43) $$\|\tilde{\beta}\|_1 = \|\alpha_i\|_2^{-2} \geqq 1 .$$

PROPOSITION 10. *If the state* j *is ergodic and let the vector* $\gamma_j = (g_1^{(j)}, g_2^{(j)}, \cdots, g_m^{(j)})$ *be defined by* (24), *then we have*

(44) $$\mathcal{R} \equiv \{i : g_i^{(j)} \geqq 1/m\} \subset \mathcal{E}_k \equiv \mathrm{Supp}\,(\alpha_k)$$

for some k.

PROOF. The result is easily deduced from Corollary 2 of Theorem 6 and Proposition 8.

PROPOSITION 11. *Let $\bar{\gamma}_j$ and $\bar{\gamma}_k$ be approximations to γ_j and γ_k respectively such that*

$$\|\bar{\gamma}_j-\gamma_j\|_1<d \quad and \quad \|\bar{\gamma}_k-\gamma_k\|_1<d\,, \tag{45}$$

for some positive number d, where each γ_j and γ_k is defined in (24) *and coincides with one of the extremal probability vectors α_i's defined in* (18). *If the two states j and k are in the same ergodic class, i.e.,*

$$\mathrm{Supp}\,(\gamma_j)=\mathrm{Supp}\,(\gamma_k) \quad and \quad \gamma_j=\gamma_k\,,$$

then

$$\|\bar{\gamma}_j-\bar{\gamma}_k\|_1\leqq 2d\,, \tag{46}$$

otherwise

$$2(1-d)\leqq\|\bar{\gamma}_j-\bar{\gamma}_k\|_1\leqq 2(1+d)\,. \tag{47}$$

PROOF. It is easily seen that

$$\begin{aligned}\|\gamma_j-\gamma_k\|_1-(\|\bar{\gamma}_j-\gamma_j\|_1+\|\bar{\gamma}_k-\gamma_k\|_1)&\\ \leqq\|\bar{\gamma}_j-\bar{\gamma}_k\|_1&\\ \leqq\|\gamma_j-\gamma_k\|_1+\|\bar{\gamma}_j-\gamma_j\|_1+\|\bar{\gamma}_k-\gamma_k\|_1\,.&\end{aligned} \tag{48}$$

Hence we have (46) and (47). This implies that if $\|\bar{\gamma}_j-\bar{\gamma}_k\|_1>2d$ then the states j and k are in different ergodic classes, and if $\|\bar{\gamma}_j-\bar{\gamma}_k\|_1<2(1-d)$ then these states are in the same ergodic class, and that the quantity $\|\bar{\gamma}_j-\bar{\gamma}_k\|_1$ does not take a value in the open interval $(2d, 2(1-d))$ when $d\leqq 1/2$.

Now we give a practical way of computing α_i's and of classifying states of P. We assume that approximations $\bar{\beta}_j$'s to $\tilde{\beta}_j$'s, which are computed by the algorithm (9), are accurate enough to make the inequalities (58)–(62) hold. We define

$$e=\max_j\|\bar{\beta}_j-\tilde{\beta}_j\|_1\,. \tag{49}$$

It should be noted that $\|\bar{\gamma}_j-\gamma_j\|_1$ is of the same order as e in this case.

A COMPUTATIONAL ALGORITHM.

1) $l:=0$,
   $\mathcal{U}:=\{1, 2, \cdots, m\}$.
2) Draw a state, say j, from the set $\mathcal{U}$ and compute an approximation $\bar{\beta}_j=(\bar{b}_1^{(j)}, \bar{b}_2^{(j)}, \cdots, \bar{b}_m^{(j)})$ to $\tilde{\beta}_j$ by the algorithm (9)

starting with ε_j.

3) If

$$\sum_{i=1}^{m} \bar{b}_i^{(j)} < t_1 , \tag{50}$$

then go to 12).

4) If

$$(\|\bar{\beta}_j\|_1 - \| |\bar{\beta}_j|^+ \|_1)/\|\bar{\beta}_j\|_1 > t_2 , \tag{51}$$

then go to 12).

5) Form an approximation $\bar{\gamma}_j = (\bar{g}_1^{(j)}, \bar{g}_2^{(j)}, \cdots, \bar{g}_m^{(j)})$ to γ_j by the formula

$$\bar{\gamma}_j = \| |\bar{\beta}_j|^+ \|_1^{-1} |\bar{\beta}_j|^+ . \tag{52}$$

If

$$\|\bar{\gamma}_j P - \bar{\gamma}_j\|_1 > t_3 , \tag{53}$$

then go to 12).

6) If

$$\|\bar{\gamma}_j - \bar{\alpha}_{i_0}\|_1 < t_4 \qquad \text{for some } i_0 \leqq l , \tag{54}$$

then go to 11).

7) If

$$\|\bar{\gamma}_j - \bar{\alpha}_{i_1}\|_1 < t_5 \qquad \text{for some } i_1 \leqq l , \tag{55}$$

then go to 12), otherwise

$$l := l + 1 ,$$

$$\bar{\mathcal{E}}_l := \{i:\ g_i^{(j)} \geqq t_6\} \cup \{j\} .$$

8) If

$$\psi_{\bar{\mathcal{E}}_l} \otimes P^B = \psi_{\bar{\mathcal{E}}_l} , \tag{56}$$

then go to 9), otherwise

$$\bar{\mathcal{E}}_l := \mathrm{Supp}\,(\psi_{\bar{\mathcal{E}}_l} \oplus \psi_{\bar{\mathcal{E}}_l} \otimes P^B) ,$$

return to 8).

9) $\bar{\alpha}_l := \|(\bar{\gamma}_j|_{\bar{\mathcal{E}}_l})\|_1^{-1} (\bar{\gamma}_j|_{\bar{\mathcal{E}}_l})$,
$\mathcal{U} := \mathcal{U} - \bar{\mathcal{E}}_l$.

10) If $\mathcal{U}$ is empty the classification is completed, in which case

$$\bar{\mathcal{T}} := \{1, 2, \cdots, m\} - \bigcup_{i=1}^{l} \bar{\mathcal{E}}_i ,$$

is the set of transient states of P, and $\bar{\mathcal{E}}_i$, $i=1, 2, \cdots, l$ are ergodic classes of P, otherwise go to 2).

11) $\bar{\mathcal{E}}_{i_0} := \bar{\mathcal{E}}_{i_0} \cup \{j\}$
12) $\mathcal{U} := \mathcal{U} - \{j\}$
go to 2).

The statement 3) is to check whether j is transient or not. If we choose t_1 so as to satisfy

$$e \ll t_1 < \min_i (\min \alpha_i) \tag{57}$$

then it works perfectly. However, this condition need not to hold because, by Proposition 7, since if only we can pick up at least a single state in each ergodic class then we can find out all the other states in the same ergodic class in the statement 8). Hence by Proposition 9, we may set t_1 so as to satisfy

$$e \ll t_1 \leqq 1/m - e\,. \tag{58}$$

The statements 4) and 5) are to check whether $\bar{\beta}_j$ and $\bar{\gamma}_j$ are reliable approximations to $\tilde{\beta}_j$ and γ_j respectively on the assumption that j is ergodic state. We may choose

$$t_2 = 100e(\text{say}) \quad \text{and} \quad t_3 = 100e(\text{say})\,. \tag{59}$$

The statement 6) is to check, based on Proposition 11, whether j belongs to some ergodic class which has already found out, where t_4 is set to satisfy

$$2e < t_4 \ll 1\,. \tag{60}$$

The statement 7) is to check, based on Proposition 11, whether j is ergodic or not, where t_5 is set to satisfy

$$t_4 \ll t_5 < 2(1-e)\,. \tag{61}$$

If j is ergodic it is a state in a new ergodic class. If we set t_6 so as to satisfy

$$e \ll t_6 < 1/m - e \tag{62}$$

then $\bar{\mathcal{E}}_l$ is not empty by Proposition 10. Hence we can find out, in the statement 8), all the states in the new ergodic class to which j belongs.

If we want only to discriminate ergodic states from transient states and to obtain a stationary probability vector, or if we have a prior knowledge that there is only a single ergodic class we had better start the algorithm (9) with $\beta_0=\mathbf{1}^t$, since $\|\bar{\beta}\|_1$ is expected to be much larger than r by the inequality (29). Finally it should be noted that the algorithm is not self-correcting. However this does not affect our method seriously since it is self-correcting with respect to the component in the subspace $\mathrm{Im}\,(P^t-I)$ and we can restart it without affecting too much on the error of the final approximation $\bar{\beta}_j$ to $\tilde{\beta}_j$.

THE INSTITUTE OF STATISTICAL MATHEMATICS

REFERENCES

[1] Decell, H. P. and Odell, P. L. (1967). On fixed point probability vector of regular or ergodic transition matrices, *J. Amer. Statist. Ass.*, **62**, 600-602.

[2] Feller, W. (1957). *An introduction to probability theory and its applications*, Vol. 1 (2nd ed.), Wiley, New York.

[3] Hestenes, M. R. and Stiefel, E. (1952). Method of conjugate gradients for solving linear systems, *J. Res. Nat. Bur. Standards*, **49**, 409-436.

[4] Kammerer, W. J. and Nashed, M. Z. (1972). On the convergence of the conjugate gradient method for singular operator equations, *SIAM J. Numer. Anal.*, **9**, 165-181.

[5] Odell, P. L. and Decell, H. P. (1967). On computing the fixed-point probability vector of ergodic transition matrices, *J. ACM*, **14**, 765-768.

[6] Pearl, M. (1973). *Matrix theory and finite mathematics*, McGraw-Hill, New York.

[7] Penrose, R. (1955). A generalized inverse for matrices, *Proc. Cambridge Phil. Soc.*, **51**, 406-413.

[8] Rao, C. R. and Mitra, S. K. (1971). *Generalized inverse of a matrices and its applications*, Wiley, New York.

[9] Reid, J. K. (1971). On the method of conjugate gradients for the solution of large sparse systems of linear equations, *Proc. Conference on "Large sparse set of linear equations"*, Academic Press, New York.

[10] Romanovsky, V. (1936). Recherches sur les chaines de Markoff, *Acta Math., Stockh.*, **66**, 147.

[11] Tanabe, K. (1975). The conjugate gradient method for computing all the extremal stationary probability vectors of a stochastic matrix, *The Institute of Statistical Mathematics Research Memorundum*, **72**.

[12] Tanabe, K. (1976). Generalized inverses of matrices (3) (in Japanese), *Communication of the Operations Research Society of Japan*, **21**, 324-326.