THE GEOMETRY OF FINITE MARKOV CHAINS

N. Pullman

(received March 30, 1963)

The purpose of this paper is to present a geometric theorem which provides a proof of a fundamental theorem of finite Markov chains.

The theorem, stated in matrix theoretic terms, concerns the asymptotic behaviour of the powers of an n by n stochastic matrix, that is, a matrix of non-negative entries each of whose row sums is 1. The matrix might arise from a repeated physical process which goes from one of n possible states to another at each iteration and whose probability of going to a state depends only on the state it is in at present and not on its more distant history. The entry a_{ij} of the matrix A (called the one step transition matrix) is the probability that the process goes from state i to state j in one step. The ij-th entry in A^m, which is denoted by $a_{ij}^{(m)}$, is the probability of going from i to j in m steps. For example the process might consist of shuffling a deck of n cards by means of a machine which puts the i-th card from the top into the j-th from the top with probability a_{ij}. Then $a_{ij}^{(m)}$ is the probability of finding the i-th card in the j-th position at the m-th shuffle.

For each $m > 0$, we say that <u>i leads to j in m steps</u> iff $a_{ij}^{(m)} > 0$. We write $i \rightsquigarrow j$ iff i leads to j in m steps for some m.

Canad. Math. Bull. vol. 8, no. 3, April 1965

It is easy to see that $\rightsquigarrow$ is an equivalence relation on the set of those states on which $\rightsquigarrow$ is symmetric, i.e. on $E = \bigcap_{j=1}^{n} \{ i : i \rightsquigarrow j \text{ implies } j \rightsquigarrow i \}$. E is called the set of <u>ergodic</u> states. E is partitioned by $\rightsquigarrow$ into ν equivalence classes $E_1, E_2, \ldots, E_\nu$ called <u>ergodic</u> classes. The states not in E are called <u>transient</u>.

Theorems I and II below are respectively probabilistic and matrix theoretic statements of the fundamental theorem of finite Markov chains. We shall provide a geometric proof at the conclusion of the paper. For probabilistic proofs we refer the reader to [1], [2], [3], and [4]. An algebraic proof can be found in [5].

THEOREM I. If $A = (a_{ij})$ is the one step transition matrix of a Markov chain with n states then:

(Ia) $\lim_{m \to \infty} a_{ij}^{(m)} = 0$ whenever j is transient.

There is a partition of each $\rightsquigarrow$ equivalence class E_r into c_r non-empty subsets (called <u>cyclically moving classes</u>) $E_{r0}, E_{r1}, \ldots, E_{rc_r - 1}$ with the following properties:

(Ib) If $i \in E_{rs}$ and i leads to j in one step then $j \in E_{r, s+1}$ (the second subscript is read modulo c_r).

(Ic) To each E_{rs} corresponds an n-tuple k_{rs} of non-negative numbers whose sum is 1 for which the j-th component, $k_j^{(r,s)}$, is zero iff $j \notin E_{rs}$ and such that:

$$\lim_{m \to \infty} a_{ij}^{(mc_r + t)} = k_j^{(r,s')} \qquad \text{for all } j$$

https://doi.org/10.4153/CMB-1965-025-3 Published online by Cambridge University Press

and all $t = 0, 1, \ldots, c_r - 1$ whenever $i \in E_{rs}$ and $s' \equiv s + t \pmod{c_r}$.

THEOREM II. If $A = (a_{ij})$ is an n by n stochastic matrix then there is a permutation matrix P (i.e. a matrix of zeroes and ones which has only one non-zero entry in each row and each column) such that:

$$PAP^{-1} = \begin{vmatrix} A_{00} & A_{01} & A_{02} & \ldots & A_{0r} & \ldots & A_{0,\nu-1} & A_0 \\ 0 & A_1 & 0 & \ldots & 0 & \ldots & 0 & 0 \\ \cdot & & & & & & & \\ \cdot & & & & & & & \\ \cdot & & & & & & & \\ 0 & 0 & 0 & \ldots & A_r & \ldots & 0 & 0 \\ 0 & 0 & 0 & \ldots & 0 & \ldots & A_{\nu-1} & 0 \\ 0 & 0 & 0 & \ldots & 0 & \ldots & 0 & A_\nu \end{vmatrix} \quad (\text{if } E \neq \emptyset),$$

$$PAP^{-1} = \begin{vmatrix} A_1 & 0 & \cdot & & \cdot & \cdot & 0 \\ 0 & A_2 & \cdot & & \cdot & \cdot & 0 \\ 0 & \cdot & \cdot & \cdot & A_r & \cdot & \cdot & 0 \\ 0 & \cdot & \cdot & \cdot & & 0 & A_\nu \end{vmatrix} \quad (\text{if } E = \emptyset),$$

https://doi.org/10.4153/CMB-1965-025-3 Published online by Cambridge University Press

$$A_r = \begin{vmatrix} 0 & A_{r1} & 0 & 0 & \cdots & & 0 & 0 \\ 0 & 0 & A_{r2} & 0 & \cdots & & 0 & 0 \\ \cdot & & & & & & & \\ \cdot & & & & & & & \\ \cdot & & & & & & & \\ 0 & 0 & 0 & 0 & \cdots & A_{rs} & \cdots & 0 \\ \cdot & & & & & & & \\ \cdot & & & & & & & \\ \cdot & & & & & & & \\ 0 & 0 & 0 & 0 & \cdots & & 0 & A_{rc_r-1} \\ A_{r0} & 0 & 0 & 0 & \cdots & & 0 & 0 \end{vmatrix},$$

where the A_r and A_{00} are square matrices and:

(IIa) $\lim_{m\to\infty} A_{00}^m = 0$,

(IIb) the entries of A_r which are in no A_{rs} are zero,

(IIc) for each $r = 1, 2, \ldots, \nu$ there are stochastic matrices $\Pi_{r0}, \Pi_{r1}, \ldots, \Pi_{rs}, \ldots, \Pi_{r,c_r-1}$ such that for each $t = 0, 1, \ldots, c_r-1$:

https://doi.org/10.4153/CMB-1965-025-3 Published online by Cambridge University Press

$$
\lim_{m\to\infty} A_r^{mc_r+t} = \left|\begin{array}{ccccccccc}
0 & \cdots & 0 & \Pi_{rt} & 0 & \cdots & & 0 \\
0 & \cdots & 0 & & \Pi_{r,t+1} & \cdots & & 0 \\
 & & & & & \ddots & & \\
0 & \cdots & & & & & 0 & \Pi_{r,c_r-1} \\
\Pi_{r0} & 0 & & & \cdots & & & 0 \\
0 & \Pi_{r1} & 0 & & \cdots & & & 0 \\
 & & \ddots & & & & & \vdots \\
0 & \cdots & 0 & \Pi_{r,t-1} & 0 & \cdots & & 0
\end{array}\right|
$$

The entries in this matrix are zero if and only if they are in no Π_{rs}, there are as many rows in Π_{rs} as there are rows in $A_{r,s-t+1}$ (second subscript modulo c_r), and all of the rows of Π_{rs} are the same vector $\pi^{(r,s)}$.

The method we shall use to prove the fundamental theorem and related theorems is briefly this: we identify the n by n stochastic matrix A with a linear operator f on the simplex, S, spanned by the basis vectors of Euclidean n-space. The intersection, K, of all the images $f^m(S)$ is a simplex whose vertices are permuted by f. The position of K in S and this permutation determine the behaviour of the $a_{ij}^{(m)}$ for large m and also locate the vertices of the simplex of its stochastic eigenvectors.

Before going further we shall state a few definitions and preliminary remarks for the reader's convenience.

A <u>convex polytope</u> P is the convex hull of finitely many points $t_1, t_2, \ldots, t_m$ in some Euclidean n-space. The point t_i is a <u>vertex</u> of P iff the convex hull of the others doesn't contain it. The convex hull of any subset of the vertices of P is called a <u>sub-polytope</u> of P. A linear function f mapping P into P is called a <u>linear operator on P</u>. A convex polytope P is a <u>simplex</u> iff none of its vertices is in the flat determined by the remaining vertices. If S is a simplex, each subpolytope is called a <u>subsimplex</u>. The subsimplices of S are themselves simplices. If A is a subset of the convex polytope P, <u>a carrier of A in P</u> is a subpolytope with fewest vertices, containing A. If P is a simplex then each non empty subset has a unique carrier in P.

Three direct consequences of these definitions which we shall refer to in the sequel are:

(1) subsimplices without vertices in common are disjoint;

(2) if the carriers of m points in a simplex S are disjoint then the convex hull of these points is a simplex;

(3) if f is a linear operator on a simplex S, then the image of the carrier of a subset X of S is contained in the carrier of f(X).

The method we use is based on a lemma which we couldn't find in the literature:

LEMMA 1. The intersection K, of a nested sequence of convex polytopes $\{P_\alpha\}$ each of which has n vertices is a convex polytope.

<u>Proof.</u> It is possible to choose a subsequence, $\{P_{\alpha_\beta}\}$ and a vertex v_β of $Q_\beta = P_{\alpha_\beta}$ such that $\{v_\beta\}$ converges, to k_1 say. Next choose a subsequence $\{Q_{\beta_\gamma}\}$ of $\{Q_\beta\}$ and a sequence of vertices w_γ of $R_\gamma = Q_{\beta_\gamma}$ with $w_\gamma \neq v_{\beta_\gamma}$ such that

https://doi.org/10.4153/CMB-1965-025-3 Published online by Cambridge University Press

w_γ converges, to k_2 say. And so on, getting $k_1, k_2, \ldots, k_n$. This process must halt in n steps because the P_α have only n vertices apiece. Let T be the convex hull of the k_i. Clearly $T \subseteq K$. Suppose $x \in K \sim T$. Let h be a hyperplane separating x from T. Let ε be the distance from h to T. For each i there are infinitely many α for which a vertex of P_α is in the sphere of radius $\varepsilon/2$ about k_i. There is therefore a member of $\{P_\alpha\}$ on the side of h opposite x hence $\bigcap P_\alpha$ and $\{x\}$ are disjoint. Thus $K \sim T = \emptyset$ and hence $K = T$.

LEMMA 2. If f is a continuous function mapping the compact set P into P and $K = \bigcap_{m \geq 1} f^m(P)$ then $f(K) = K$.

Proof. It is sufficient to show that $K \subseteq f(K)$. If $x \in K$ then $x \in f^m(P)$ for all $m > 0$ and hence $x = f(x_{m-1})$ for some $x_{m-1} \in f^{m-1}(P)$. The x_m have a convergent subsequence $\{x_{m_i}\}$ converging to a point y of P. If we can show that $y \in K$ then we are through because $x = \lim_{i \to \infty} f(x_{m_i}) = f(y)$.

If y were not in K then, for some N, y would not be in $f^N(P)$. The complement of $f^N(P)$ contains no x_{m_i} for $m_i > N$. But the complement of $f^N(P)$ is an open neighborhood of y. Therefore $y \in K$.

THEOREM 1. If f is a linear operator on a simplex S then

(i) the intersection, K, of the iterates $f^m(S)$ is a simplex, and

(ii) the vertices of K are permuted by f and hence fall into ν disjoint classes on each of which f is a cyclic permuta-

https://doi.org/10.4153/CMB-1965-025-3 Published online by Cambridge University Press

tion so that for $r = 1, 2, \ldots, \nu$ and $s = 0, 1, \ldots, c_r - 1$, (where c_r is the number of elements in the r-th class) we have

(iii) $f(k_{rs}) = k_{r, s+1}$ (the second subscript is read modulo c_r).

If C_{rs} is the carrier of k_{rs} in S then

(iv) the C_{rs} are disjoint,

(v) $f^t(C_{rs}) \subseteq C_{r, s+1}$, and

(vi) $\bigcap_{m \equiv t} f^m (C_{rs}) = k_{rs'}$ when $s' \equiv t + s \pmod{c_r}$.

If K_r is the subsimplex of K whose vertices are $k_{r0}, k_{r1}, \ldots, k_{r, c_r - 1}$ then

(vii) the K_r are disjoint and

(viii) the set of all f-fixed points in S is a simplex whose vertices are the barycenters of the K_r.

<u>Proof.</u>

(a) Lemmas 1 and 2 establish that K is a convex polytope and that $f(K) = K$.

(b) f permutes the vertices of K.

Let k be a vertex of K, $X_k = [f^{-1}(k)] \cap K$ and $C_K(X_k)$ denote a carrier of X_k in K. Then $f(C_K(X_k)) = \{k\}$ by remark (3) and hence $C_K(X_k) = X_k$. Therefore there are as many $C_K(X_k)$ as there are vertices of K, since the X_k are pairwise disjoint

https://doi.org/10.4153/CMB-1965-025-3 Published online by Cambridge University Press

and hence each carrier $C_K(X_k)$ has only one vertex. Thus f^{-1} and hence f permute the vertices of K.

The family of sets, $\{\bigcup_{m \geq 0} \{f^m(k)\} : k$ is a vertex of $K\}$, partition the vertices of K into ν disjoint classes on each of which f is a cyclic permutation. Denote the convex hulls of these partitioning sets by $K_1, K_2, \ldots, K_\nu$. Let k_{r0} be any vertex of K_r; let $k_{rs} = f^s(k_{r0})$ for $r = 1, 2, \ldots, \nu$ and $s = 0, 1, 2, \ldots, c_r - 1$. Let C_{rs} denote the carrier in S of k_{rs}.

(c) Each C_{rs} meets K in only one point, namely k_{rs}, and hence $C_{rs} = C_{r's'}$ iff $(r, s) = (r', s')$. If not C_{rs} would contain two distinct points k_{rs} and k' of K. The line they determine would meet C_{rs} in a line segment contained in K neither of whose endpoints is k_{rs}, contradicting the assumption that k_{rs} is a vertex.

(d) $f(C_{rs}) \subseteq C_{r,s+1}$ and hence $f^t(C_{rs}) \subseteq C_{rs'}$ if $s' \equiv t + s \pmod{c_r}$; because, by remark (3), $f(C_{rs})$ is contained in the carrier in S of $f(k_{rs})$ which is $C_{r,s+1}$ by definition.

(e) $\bigcap_{m \equiv t} f^m(C_{rs}) = \{k_{rs'}\}$ if $s' \equiv t + s \pmod{c_r}$.

To see why this is so we observe first that if $m \equiv t \pmod{c_r}$, then $k_{rs'} \in f^m(C_{rs})$ because $f^m(k_{rs}) = k_{r,s+m} = k_{rs'}$; and secondly that $(\bigcap_{m \equiv t} f^m(C_{rs})) \subseteq K$ so that $\{k_{rs'}\} \subseteq \bigcap_{m \equiv t} f^m(C_{rs})$ $= K \cap \bigcap_{m \equiv t} f^m(C_{rs}) \subseteq \bigcap_{m \equiv t} K \cap f^m(C_{rs}) \subseteq \bigcap_{m \equiv t} K \cap C_{rs'}$.

According to (c), $K \cap C_{rs'} = \{k_{rs'}\}$ and hence $\bigcap_{m \equiv t} f^m(C_{rs}) = \{k_{rs'}\}$. An immediate consequence of this is:

(f) The C_{rs} are pairwise disjoint. Applying remarks (2) and (1) we have:

(g) K is a simplex and the K_r are disjoint.

Evidently the set F of all fixed points is a convex subset of K. By linearity, the barycenter $b_r = \frac{1}{c_r} \sum_{s=0}^{c_r - 1} k_{rs}$ of K_r is fixed by f and hence F contains the convex hull of these barycenters. Conversely, if x is fixed, then, since $x \in K$,

$$x = \sum_{r=1}^{\nu} \sum_{s=0}^{c_r - 1} x_{rs} k_{rs} \quad (\text{where } x_{rs} \geq 0 \text{ and } \Sigma\Sigma\, x_{rs} = 1)$$

and hence $f(x) = \sum_{r=1}^{\nu} \sum_{s=0}^{c_r - 1} x_{rs} k_{r,s+1} = x$. Therefore $x_{rs} = x_{rs'}$ where $s' \equiv s + 1 \pmod{c_r}$ and hence, given r, either $x_{rs} = 0$ for each $0 \leq s < c_r$ or for all $0 \leq s < c_r : x_{rs} = \frac{1}{c_r}$.

Consequently x is in the convex hull of the barycenters. Therefore F is a convex polytope spanned by the ν barycenters of the K_r. The barycenters of the K_r are the vertices of F because the K_r are distinct. By applying remark (2) and (g) we obtain:

(h) The set F of all f-fixed points is a simplex whose vertices are the barycenters of the K_r.

This completes the proof of the theorem.

We shall present a proof of the probabilistic form of the fundamental theorem (theorem I) after a few preliminary remarks showing the correspondence between the pertinant geometric and probabilistic ideas.

https://doi.org/10.4153/CMB-1965-025-3 Published online by Cambridge University Press

Each state $i = 1, 2, \ldots, n$ of the Markov chain whose one step transition matrix is $A = (a_{ij})$ corresponds to the n-tuple v_i whose only non-zero component, 1, is its i-th component. Let $f(x) = xA$ (i.e. the j-th component of $f(x)$ is $\sum_{i=1}^{n} x_i a_{ij}$) for each x in the convex hull S, of the v_i. S is a simplex and f is a linear operator on S. If $X \subseteq S$ let $C(X)$ denote the carrier of X in S. We then have:

(4) $i \rightsquigarrow j$ iff $v_j \in \bigcup_{m>0} C(f^m(v_i))$ because of the definitions of $\rightsquigarrow$ and C.

We shall show that

(5) i is ergodic iff v_i is a vertex of $C(K)$

after we have established (5a) and (5b) below.

(5a) If $E(K) = \{i : v_i \in C(K)\}$ then for each j there is an $i \in E(K)$ such that $j \rightsquigarrow i$.

Proof of (5a). Let $D = C(\{f^m(v_j) : m > 0\})$ then, using (3), we have $f(D) \subseteq D$ and hence $\bigcap_{m>0} f^m(D)$ is a non-empty subset of both K and D. There is, therefore, a vertex v_i of $C(K)$ which is also a vertex of D. But the vertices of D are also those of $\bigcup_{m>0} C(f^m(v_j))$. Consequently $j \rightsquigarrow i \in E(K)$.

(5b) If $i \in E(K)$ and $i \rightsquigarrow j$ then $j \in E(K)$ and $j \rightsquigarrow i$.

Proof of (5b). $i \in E(K)$ and $i \rightsquigarrow j$ imply that v_i is a vertex of some C_{rs} and v_j is a vertex of some $C(f^t(v_i))$ by (4). Therefore $f^t(v_i) \in f^t(C_{rs})$. But $f^t(C_{rs}) \subseteq C_{rs'}$, when $s' \equiv s + t \pmod{c_r}$ according to Theorem 1 part (v) and hence

https://doi.org/10.4153/CMB-1965-025-3 Published online by Cambridge University Press

v_j is a vertex of $C_{rs'}$. Consequently $j \in E(K)$.

$C(f^m(v_j)) = C_{rs}$ for a sufficiently large $m \equiv s - s' \pmod{c_r}$ by parts (v) and (vi) of theorem 1, but $v_i \in C_{rs}$ and hence $j \rightsquigarrow i$ by (4).

Proof of (5). $E(K) \subseteq E$ by (5b) and the definition of E. If $j \notin E(K)$ then $j \rightsquigarrow i \in E(K)$ by (5a) and $i \not\rightsquigarrow j$ by (5b). Consequently $j \notin E$ and hence $E \subseteq E(K)$.

<u>Proof of Theorem I.</u>

(Ia) If i is any state let $x^{(m)}$ be the point of $C(K)$ closest to $f^m(v_i)$. Because of the definition of K the sequence of distances d_m between $f^m(v_i)$ and $x^{(m)}$ converges to zero. According to (5) $x_j^{(m)} = 0$ for all transient j; therefore $\lim_{m \to \infty} a_{ij}^{(m)} = 0$ because the j-th component of $f^m(v_i)$ is $a_{ij}^{(m)}$.

Let $E_{rs} = \{ i : v_i \in C_{rs} \}$; then, evidently, $E_r = \bigcup_{s=0}^{c_r - 1} E_{rs}$. The E_{rs} are pairwise disjoint and non-empty because of parts (iii) and (iv) of Theorem 1.

(Ib) If $i \in E_{rs}$ and i leads to j in one step then $v_i \in C_{rs}$ and $v_j \in C(f(v_i))$. But according to Theorem 1 part (v): $f(v_i) \in C_{r,s+1}$ and hence $v_j \in C_{r,s+1}$.

Consequently $j \in E_{r,s+1}$.

(Ic) k_{rs} is an n-tuple of non-negative numbers summing to 1 because $k_{rs} \in S$. The j-th component, $k_j^{(r,s)}$ of k_{rs} is 0 iff $j \notin E_{rs}$ because of the definitions of C_{rs} and E_{rs}.

$$\lim_{m\to\infty} a_{ij}^{(mc_r+t)} = k_j^{(r,s')} \qquad \text{for all } j \text{ and all}$$

$t = 0, 1, \ldots, c_r - 1$ whenever $i \in E_{rs}$ and $s' \equiv s + t \pmod{c_r}$ by (vi).

This completes the proof of theorem I.

Theorem II can be proven either directly from Theorem I (their statements are equivalent) or from Theorem 1.

To obtain a proof of Theorem II the latter way let $c(r, s)$ be the number of vertices in C_{rs} and let P be the permutation matrix which performs the change of basis mapping the first c basis vectors (vertices) $v_1, v_2, \ldots, v_c$ onto each of the c vertices not in $C(K)$, mapping the next $c(1,1)$ basis vectors $v_{c+1}, v_{c+2}, \ldots, v_{c+c(1,1)}$ onto the vertices of C_{11}; and so forth until all the last $c(\nu, c_\nu - 1)$ basis vectors are mapped onto the vertices of $C_{\nu, c_\nu - 1}$.

(IIa) is proven analogously to (Ia). (IIb) is a result of (v).

To prove (IIc), define $\pi^{(r,s)}$ by letting $\pi_j^{(r,s)}$ be the j-th component of k_{rs} for each $j = 1, 2, \ldots, c(r, s)$; each $s = 0, 1, \ldots, c_r - 1$ and each $r = 1, 2, \ldots, \nu$. (IIc) then follows from (vi).

We have extended the techniques used here to study inhomogeneous chains, that is, to study the asymptotic behaviour of products $A_1 \cdot A_2 \cdot \ldots \cdot A_n \cdot \ldots$ of stochastic matrices A_n which are not necessarily the same matrix and all of which might be infinite. Some of these results are contained in a forthcoming paper on infinite products of substochastic matrices [7].

https://doi.org/10.4153/CMB-1965-025-3 Published online by Cambridge University Press

REFERENCES

1. Stochastic Processes, J. Doob, John Wiley and Sons, New York, 1953.

2. An Introduction to Probability Theory and its Applications, Volume one, W. Feller, 2nd edition, John Wiley and Sons, New York, 1950.

3. Markov Chains with Stationary Transition Probabilities, Kai Lai Chung, Springer, Berlin 1960.

4. Finite Markov Chains, J. G. Kemeny and J. Laurie Snell, D. van Nostrand Co., New Jersey 1960.

5. Applications of the Theory of Matrices, F. R. Gantmacher, Interscience, New York 1959.

6. Convexity, H. G. Eggleston, Cambridge Tracts in Mathematics and Mathematical Physics No. 47, Cambridge University Press, 1958.

7. N. Pullman, Infinite Products of Substochastic Matrices, Pacific Journal (to appear).

McGill University

https://doi.org/10.4153/CMB-1965-025-3 Published online by Cambridge University Press