GRAPH-THEORETIC AND ALGEBRAIC CHARACTERIZATIONS OF SOME MARKOV PROCESSES*

BY

A. PAZ

ABSTRACT

An algebraic decidable condition for a stationary Markov chain to consist of a single ergodic set, and a graph-theoretic decidable condition for a stationary Markov chain to consist of a single ergodic noncyclic set are formulated.

In the third part of the paper a graph-theoretic condition for a nonstationary Markov chain to have the weakly-ergodic property is given.

1. **Introduction.** In this paper we are concerned with algebraic and graph-theoretic characterizations of stationary and nonstationary Markov chains (with discrete time parameter and finite number of states).

In the first part we formulate a graph-theoretic condition for a given chain to consist of a single ergodic noncyclic set. This condition is shown to be equivalent, as to characterization, tc a condition given in Doob [3, p. 173] but more economical as to computation.

In the second part we formulate a simple algebraic sufficient and necessary condition for a stationary Markov chain to consist of a single ergodic set. This condition can also be used to characterize finite graphs, since with every finite graph one can associate an infinite number of Markov chains.

In the third part we derive a theorem generalizing a result mentioned in Doob [2] for stationary chains, to a certain type of nonstationary chains. The theorem states, roughly, that such a nonstationary chain, after a sufficiently long lapse of time, "forgets" its initial state.

Some formulas characteristic of nonstationary chains are given, permitting approximation of long products of stochastic matrices satisfying certain conditions.

A *finite graph* [1] is an ordered pair $\langle S,\Gamma\rangle$ where S is a finite nonempty set and Γ a multi-valued mapping of S into S. The elements of S are called *vertices*, and the ordered pairs $\langle a,b\rangle$, such that $a \in S$ and $b \in a\Gamma$** are called *edges*. A sequence of vertices $(a_0 a_1 \cdots a_\nu)$ such that $\langle a_i a_{i+1}\rangle$ $(i = 0,\cdots,\nu-1)$ is an edge, is a path of length ν and a_ν is a *consequent* of order ν of a_0. (Notation: $a_\nu \in a_0\Gamma^\nu$).

Received May 14, 1963.

* The paper is based on part of the author's work towards the D. Sc. degree.

** $a\,\Gamma$ is the set of all Γ images of a.

A graph is strongly connected if for every pair of vertices i,j $(i \neq j)$, i is a consequent of j.

A pair of vertices i and j has a *common consequent* k (of order ν) if there exists a ν such that $k \in i\Gamma^{\nu} \cap j\Gamma^{\nu}$.

If all the vertices in the graph have a common consequent k of order ν, then k is a *universal consequent* of order ν for the given graph.

A vertex in a given graph is *transient* if it has a consequent of which it is not itself a consequent.

A vertex which is not transient is *nontransient.* Note that, if i is a consequent of j and j is nontransient, then j is a consequent of i and i is nontransient. (For let k be any consequent of i; then k is a consequent of j and j is nontransient, whence j is a consequent of k, this implying that i is also a consequent of k).

Note also that the set of nontransient vertices cannot be empty, since a maximal sequence of vertices connected by a path must terminate in a nontransient vertex.

The class of nontransient vertices is subdivided into *ergodic* subclasses (where an ergodic class is the set of vertices of a maximal strongly-connected subgraph) with two vertices belonging to the same ergodic class iff they are consequents of each other.

Finally, it can be shown [4, p. 6] that with any ergodic class E there is associated a unique positive integer d with the following properties:

1) If $i \in E$ and $i \in i\Gamma^{\nu}$, then d divides ν.

2) If $i,j \in E$ and $j \in i\Gamma^{\mu}$ as well as $j \in i\Gamma^{\nu}$, then $\mu \equiv \nu (\text{mod}\, d)$.

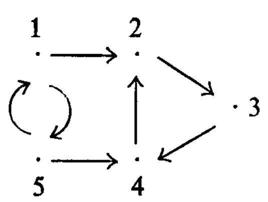
3) d is the smallest integer having properties 1) and 2).

Thus any ergodic class E may be subdivided into d cyclic classes $C_1 \cdots C_d$, as follows:

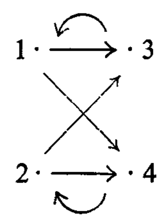
Two vertices i,j belong to the same cyclic class, iff $j \in i\Gamma^{\nu}$ and $\nu \equiv 0 (\text{mod}\, d)$.

Note that if $i \in C_1$, $j \in C_t$ $(1 \leqq t \leqq d)$ and $j \in i\Gamma^{\nu}$ then $\nu \equiv t-1 (\text{mod}\, d)$

EXAMPLES.



I



II

In graph I vertices 1 and 5 are transient while 2, 3, 4 are nontransient, forming an ergodic (strongly-connected) subgraph. Vertices 5 and 3 have vertex 4 as common consequent of order 1.

In graph II all vertices are nontransient. The graph is strongly connected and subdivisible into the cyclic subclasses $\{1,2\}$ and $\{3,4\}$.

2. **Conditions H_1 and H_2.** We shall consider the following conditions for a given graph:

Condition H_1., Every pair of vertices in the graph has a common consequent.

Condition H_2. The graph has a universal consequent.

THEOREM 1. *A graph which satisfies condition H_1 contains a single ergodic class which is not divisible into cyclic subclasses.*

Proof. Suppose there are several ergodic classes in the graph, $G_1, G_2 \cdots G_r$. Let $i_1 \in G_1$, $i_t \in G_t$ be a pair of vertices of different classes. By our assumption, i_1 and i_t have a common consequent k, where k is nontransient being a consequent of nontransient vertices. Hence i_1 and i_t are consequents of k, and this implies that $k \in G_1$ and $k \in G_t$. Thus G_1 and G_t are identical classes. To prove the second part of the theorem, assume that the ergodic class in the graph is divisible into several cyclic subclasses $C_1, C_2 \cdots C_d$. Let $c_1 \in C_1$ and $c_t \in C_t$ be a pair of vertices of different classes. They have a common consequent k which is nontransient and hence belongs to a cyclic class C_k. This implies that k is a consequent of order $k-1 (\mathrm{mod}\, d)$ of c_1 and a consequent of order $k-t (\mathrm{mod}\, d)$ of c_t. Now k is a common consequent of c_1 and c_t, which implies that $k-1 \equiv k-t (\mathrm{mod}\, d)$ or $1 \equiv t (\mathrm{mod}\, d)$; thus c_1 and c_t are identical cyclic classes.

THEOREM 2. Let *$\langle S, \Gamma \rangle$ be a graph with n vertices.* If a pair of vertices i *and j, $i, j \in S$, has a common consequent, then it has a common consequent of order ν where $\nu \leqq n(n-1)/2$.*

Proof. If states i and j have a common consequent, then there exists a sequence of (unordered) pairs of vertices (with $i = i_0, j = j_0$):

$$(i_0 j_0), (i_1 j_1) \cdots (i_\mu j_\mu)$$

such that

(1) $i_k \neq j_k$ $\qquad k = 0, 1, 2, \cdots, \mu - 1$

(2) $i_k \in i\Gamma^k$, $j_k \in j\Gamma^k$

(3) $i_\mu = j_\mu$

If the sequence contains two equal pairs, then omit the part of the sequence between these pairs, including the second of the equal pairs. Repeat this procedure until a reduced sequence is obtained:

$$(i_0 j_0), (i'_1 j'_1) \cdots (i'_k j'_k) \cdots (i'_\nu j'_\nu)$$

such that

(1′) $i'_k \neq j'_k$ $\qquad k = 0, 1 \cdots \nu - 1$

(2′) $i'_k \in i\Gamma^k$ $\quad j'_k \in j\Gamma^k$

(3′) $(i'_l j'_l) \neq (i'_t j'_t)$ $\qquad l \neq t,\ k, t = 1, 2 \cdots \nu$

(4′) $i'_\nu = j'_\nu$

Now by (2′) and (4′) $i'_\nu = j'_\nu$ is a common consequent of order ν of the vertices i and j, while by (1′) and (3′) ν is at most $n(n-1)/2$. Q.e.d.

Note that if vertices i and j have a common consequent of order v, then they also have a common consequent of order $v + k, k = 1,2 \cdots$.

THEOREM 3. *Condition H_1 is equivalent to H_2.*

Proof. That H_2 implies H_1 is trivial. Now assume H_1 to hold and let i and j be a pair of vertices in the graph. These vertices have a common consequent k of order v_1. Let t be a vertex different from i and j and l any consequent of t of order v_1. By H_1, vertices l and k have a common consequent of order v_2, which is a common consequent of order $v_1 + v_2$ of vertices i, j and t. This argument can be repeated a finite number of times to give a universal consequent. Q.e.d.

With every finite graph $\langle S,\Gamma\rangle$ where $S = \{v_1 v_2 \cdots v_n\}$ one can associate an $n \times n$ Boolean matrix (transition matrix) $M_{\langle S,\Gamma\rangle} = \| m_{ij} \|$, such that

$$m_{ij} = \begin{cases} 1 - \text{if } \langle v_i v_j \rangle \text{ is an edge of the graph,} \\ 0 - \text{otherwise.} \end{cases}$$

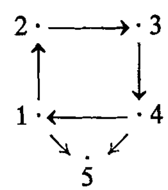
It is easily proved that the powers of $M_{\langle S,\Gamma\rangle}$, namely $M^k_{\langle S,\Gamma\rangle} = \| m^{(k)}_{ij} \|$, $k = 1,2,\cdots$, are such that:

$$m^{(k)}_{ij} = \begin{cases} 1 - \text{if there is a path of length } k \text{ from } i \text{ to } j, \\ 0 - \text{otherwise.} \end{cases}$$

This result, combined with Theorem 2, provides a computational method for deciding whether a given graph satisfies H_1. The question is decided by raising the transition matrix of the graph to the $n \cdot (n-1)/2$-th power at most. By Theorem 3, the same method will decide whether a given graph satisfies H_2.

Note that H_1, although equivalent to H_2, is more economical in the sense that more steps would be required to decide directly whether a given graph satisfies H_2.

EXAMPLE. Consider the following graph



In order to verify whether vertices 2 and 5 have a common consequent, we consider the sequence of consequent sets:

$$\begin{matrix}\{2\}\\ \{5\}\end{matrix} \xrightarrow{1} \begin{matrix}\{3\}\\ \{1\}\end{matrix} \xrightarrow{2} \begin{matrix}\{4\}\\ \{2\}\end{matrix} \xrightarrow{3} \begin{matrix}\{5,1\}\\ \{3\}\end{matrix} \xrightarrow{4} \begin{matrix}\{1,2\}\\ \{4\}\end{matrix} \xrightarrow{5} \begin{matrix}\{2,3\}\\ \{5,1\}\end{matrix} \xrightarrow{6} \begin{matrix}\{3,4\}\\ \{1,2\}\end{matrix} \xrightarrow{7} \begin{matrix}\{4,5,1\}\\ \{2,3\}\end{matrix} \xrightarrow{8}$$

$$\xrightarrow{8} \begin{matrix}\{5,1,2\}\\ \{3,4\}\end{matrix} \xrightarrow{9} \begin{matrix}\{1,2,3\}\\ \{4,5,1\}\end{matrix}$$

One sees that vertex 1 is a common consequent at vertices 2 and 4 of order 9 (the bound given by Theorem 2 in this case being 10). Inspection also shows that any other pair of vertices in the graph have a common consequent at smaller order than veitices 2 and 5.

Further consideration of this example shows that 13 steps would be required to decide directly whether this graph satisfies H_2.

REMARK. The above graph can be generalized:

Let $S = \{1,2,\cdots,n\}$ and define:

$$i\Gamma = i+1$$
$$(n-2)\Gamma = \begin{cases} n-1 \\ 1 \end{cases}$$
$$(n-1)\Gamma = 1$$

It can be shown that this graph satisfies the condition H_1 but the smallest order for which the condition is satisfied is $(n^2 - 2n + 2)/2$ if n is even and $(n^2 - 2n + 3)/2$ if n is odd. These numbers are close enough to the bound given in Theorem 2.

3. **Stochastic matrices. Stationary case.** In what follows familiarity with finite stochastic matrices and their properties is assumed. The reader is referred to [2, 3, 4] for a detailed account on these topics.

A *finite stochastic* matrix is a square matrix $A = \| a_{ij} \|$ such that $a_{ij} \geqq 0$, $i,j = 1,2,\cdots,n$, and $\sum_{j=1}^{n} a_{ij} = 1$, $i = 1,2,\cdots,n$.

With every stochastic finite matrix A there can be associated a finite graph $\langle S, \Gamma \rangle$ such that $S = \langle v_1 v_2 \cdots v_n \}$, where n is the order of A and the ordered pair of vertices $\langle v_1, v_j \rangle$ is an edge of the graph iff $a_{ij} > 0$. More generally, the elements of the stochastic matrix are interpreted as the probability of transition from state to state in a system having n internal states.

Using the previous classification of vertices in a given graph (internal states in a given system) we can now state:

THEOREM 4*. *If A is a stochastic matrix, there is a stochastic matrix $Q = \| q_{ij} \|$ such that*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} A^m = Q.$$

THEOREM 5**. *The limit q_{ij} in Theorem 4 is independent of i iff there is a single ergodic class in the graph related to A.*

THEOREM 6***. *The limit q_{ij} of Theorem 4 can be taken as ordinary limit*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} A^m = \lim_{n\to\infty} A^m = Q$$

* For a proof, see [2, p. 175].

** For proof, see [2, p. 181].

*** See [2, p. 182].

iff there are no cyclic subclasses in any ergodic class of the graph related to A.

COROLLARY 1. *If A is a stochastic matrix and the graph related to this matrix satisfies H_1 or (H_2), then there is a matrix $Q = \| q_{ij} \|$ such that*

$$\lim_{n \to \infty} \frac{1}{n} \sum_{m=1}^{n} A^m = \lim_{n \to \infty} A^m = Q$$

and q_{ij} is independent of i.

Proof. By Theorems 1, 3, 5 and 6.

The limiting matrix Q.

In the sequel, some additional definitions are needed:

Let A be an $n \times n$ matrix, η the n-component row vector and ξ the n-component column vector having all components equal to 1.

$A_{(r)}$ denotes the r-th row and $A^{(r)}$ the r-th column of A. Clearly $(A \cdot B)_{(r)} = A_{(r)}B$; $\bar{A}_{(r)}$ is obtained from $A_{(r)}$ by omitting the r-th term.

Clearly $\xi \cdot A_{(r)} = \begin{Vmatrix} A_{(r)} \\ \vdots \\ A_{(r)} \end{Vmatrix}$, i. e. a matrix all whose rows are equal to $A_{(r)}$.

If A is stochastic, then $A_{(r)}\xi = 1$ and $A\xi = \xi$.

If $A\xi = \xi$ and also $\eta A = \eta$ then A called *doubly stochastic.*

For any stochastic $n \times n$ matrix A and any $r \leqq n$ we define the r-th kernel of A, denoted by $\overline{k_r(A)}$, as $\overline{k_r(A)} = A - \xi A_{(r)}$.

Finally, $k_r(A)$ is obtained from $\overline{k_r(A)}$ by omitting the r-th column and the r-th row. ($k_r(A)$ is a $(n-1) \times (n-1)$ matrix).

The usual way to compute $\lim_{n \to \infty} A^n$ or $\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} A_k$ (when the first limit does not exist) is based on the fact that the limiting matrix Q satisfies (in both cases) the equation

$$QA = AQ - Q \quad \text{or} \quad Q[I - A] = 0$$

This equation is equivalent to a set of n linear homogeneous equations. Non-trivial solutions exist provided $|I - A| = 0$. If the system associated with the matrix consists of a single ergodic class, then it is known that a unique solution exists for the set of equations

$$(1) \qquad \begin{cases} (x_1 \cdots x_n)[I - A] = 0 \\ \sum_{i=1}^{n} x_i = 1 \end{cases}$$

This implies that $|I - A| = 0$ and the dimension of $[I - A]$ is $n-1$ (for the single ergodic class system).

From (1) we obtain, by adding to both sides $(x_1 \cdots x_n) \cdot \xi \cdot A_{(r)}$,

$$(2) \qquad (x_1 \cdots x_n)[I - (A - \xi A_{(r)})] = (x_1 \cdots x_n)\xi A_{(r)}$$

Now $\Sigma x_i = 1$, hence $(x_1 \cdots x_n)\xi = 1$ and $(x_1 \cdots x_n)\xi A_{(r)} = A_{(r)}$; thus $(x_1 \cdots x_n)[I - (A - \xi A_{(r)})] = A_{(r)}$, and by the definitions introduced above,

$$(3) \qquad (x_1 \cdots x_n)[I - \bar{K}_r(A)\} = A_{(r)}$$

Any vector $(x_1 \cdots x_n)$ which is a solution of (3) is such that $(x_1 \cdots x_n)\xi = \Sigma_{i=1}^n x_i = 1$. Indeed, $\Sigma_{i=1}^n x_i = (x_1 \cdots x_n)\xi = (x_1 \cdots x_n)[\xi - \xi + \xi] = (x_1 \cdots x_n)[I - A + \xi A_{(r)}]\xi =$ $= (x_1 \cdots x_n)[I - \bar{K}_r(A)]\xi = A_{(r)}\xi = 1$.

Any such vector is therefore a solution of (2) and (1). Thus (1) is seen to be equivalent to (3).

From Theorem 5 it follows that the system related to the matrix A consists of a single ergodic set iff (1) and (3) have a unique solution such that;

$$|I - \bar{K}_r(A)| \neq 0$$

(4) and

$$(x_1 \cdots x_n) = A_{(r)}[I - \bar{K}_r(A)]^{-1}.$$

We have thus proved the following

THEOREM 7. *Let A be a stochastic matrix, then for any $r = 1,2,\cdots,n$, $|I - \bar{K}_r(A)| \neq 0$ if and only if the system related to A consists of a single ergodic class.*

REMARKS. 1) It is easily seen that the theorems in this section could have been proved with $K_r(A)$ replacing $\bar{K}_r(A)$ (see definition on p. 178). This use of $K_r(A)$ has the advantage that $K_r(A)$ is an $(n-1)$ by $(n-1)$ matrix while $\bar{K}_r(A)$ is n by n. Similarly we can write, instead of (3) and (4),

$$(3') \qquad (x_1 \cdots x_{r-1} x_{r+1} \cdots x_n)\ [I - K_r(A)] = \bar{A}_{(r)}$$

$$(4') \qquad (x_1 \cdots x_{r-1} x_{r+1} \cdots x_n) = \bar{A}_{(r)}[I - K_r(A)]^{-1}$$

The information contained in these formulas is the same as that in (1) and (2), since $(x_1 \cdots x_n)$ is a stochastic vector, whence $x_r = 1 - \Sigma_{i \neq r} x_i$.

2) Corollary 1 in the preceding section and Theorem 7 provide a method for analysing the structure of a system representable by a stochastic matrix. With the aid of Theorem 7, we first ascertain whether the system has a single ergodic class or not. If it has several ergodic classes, we can investigate each class separately. If it has a single ergodic class, we have to ascertain whether it satisfies H_1 (or H_2). If it does, the ergodic class is not divisible into cyclic subclasses. The decision procedure given here is quite simple from a computational point of view.

3) The decision procedure described in Remark 2 can be applied to finite graphs as well. With a given finite graph one can associate a stochastic matrix

$A = \| a_{ij} \|$ such that $a_{ij} > 0$ iff $\langle v_i v_j \rangle$ is an edge of the graph, and $\sum_{k=1}^{n} a_{ik} = 1$, $i = 1, 2, \cdots, n$. The matrix A is clearly not unique and we can choose the simplest possible matrix (as to computation) satisfying the above condition; then the procedure described in Remark 2 can be applied to the chosen matrix.

EXAMPLES. Consider the matrices

$$A = \begin{bmatrix} \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{3} & 0 & \frac{2}{3} & 0 \\ \frac{1}{3} & 0 & \frac{2}{3} & 0 & 0 \\ 0 & \frac{2}{3} & 0 & \frac{1}{3} & 0 \\ \frac{1}{2} & \frac{1}{4} & 0 & 0 & \frac{1}{2} \end{bmatrix} \qquad B = \begin{bmatrix} 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{3} & \frac{2}{3} & 0 \\ \frac{1}{3} & \frac{2}{3} & 0 & 0 & 0 \\ \frac{2}{3} & \frac{1}{3} & 0 & 0 & 0 \\ \frac{1}{4} & 0 & \frac{1}{2} & 0 & \frac{1}{4} \end{bmatrix} \qquad C = \begin{bmatrix} 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & \frac{1}{3} & \frac{2}{3} & 0 \\ 0 & \frac{1}{4} & \frac{1}{4} & 0 & \frac{1}{2} \\ \frac{1}{3} & 0 & \frac{1}{3} & 0 & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{2} & 0 & \frac{1}{6} & 0 \end{bmatrix}$$

For matrix A, $[I - K_5(A)] = 0$. Hence A has more than one ergodic class. We find by inspection that the sets are $\{1,3\}$ and $\{2,4\}$, while state 5 is transient. Matrix B corresponds to a chain consisting of a single ergodic set and two cyclic subclasses, and matrix C satisfies H_1.

Set
$$A = \begin{bmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{bmatrix}.$$

We shall calculate $\lim A^n$ using (4):

$$K_3(A) = \begin{bmatrix} \frac{1}{4} & 0 \\ \frac{1}{4} & -\frac{1}{4} \end{bmatrix}; I - K_3(A) = \begin{bmatrix} \frac{3}{4} & 0 \\ -\frac{1}{4} & \frac{5}{4} \end{bmatrix}; \bar{A}_{(3)} = (\tfrac{1}{4}\ \tfrac{1}{4}); [I - K_3(A)]^{-1} = \frac{16}{15}\begin{bmatrix} \frac{5}{4} & 0 \\ \frac{1}{4} & \frac{3}{4} \end{bmatrix};$$

$\lim_{n \to \infty} \bar{A}_{(3)}^{(n)} = \bar{A}_{(3)}(I - K_3(A)]^{-1} = (\tfrac{2}{5}, \tfrac{1}{5})$, hence $\lim_{n \to \infty} A_{(3)}^{n} = (\tfrac{2}{5}, \tfrac{1}{5}, \tfrac{2}{5})$ and

$$\lim_{n \to \infty} A^n = \begin{bmatrix} \frac{2}{5} & \frac{1}{5} & \frac{2}{5} \\ \frac{2}{5} & \frac{1}{5} & \frac{2}{5} \\ \frac{2}{5} & \frac{1}{5} & \frac{2}{5} \end{bmatrix}.$$

4. **Stochastic matrices — nonstationary case.** In the preceding sections we considered stationary stochastic systems representable by a constant transition matrix. Nonstationary stochastic systems (discrete time parameter and finite number of states) are systems for which the transition probabilities may change from step to step.

Mathematical representation of such a system is provided by a sequence of stochastic matrices $A(1), A(2), \cdots, A(i), \cdots$, where $A(i)$ is the transition matrix in the i-th step.

The set $\{A(k) \mid k = 1,2,\cdots.\}$ for a given system may contain an infinite number of distinct matrices. However, we confine ourselves here to a finite set of distinct matrices $N = \{A_k \mid k = 1,2,\cdots,n\}$ and consider all products of the form $\prod_{i=1}^{m} A(i)$, $m = 1,2,\cdots$, where $A(i)$ is as above and $A(i) \in N$, $i = 1,2,\cdots,m$. Set $\prod_{i=1}^{m} A(i) = B = \| b_{ij} \|$, $M_j^{(B)} = \max_i b_{ij}$, $m_j^{(B)} = \min_i b_{ij}$.

The counterpart of H_1 for the nonstationary case is the following:

Condition H_3. A finite graph satisfies H_3 if every pair of vertices in the graph has a common consequent of order one.

In what follows we shall say that a matrix satisfies H_1 (H_2 or H_3) if the graph related to the matrix satisfies H_1 (H_2 or H_3).

We can now prove the following:

THEOREM 8. *If N is a finite set of matrices all satisfying H_3 and if $B = \prod_{i=1}^{m} A(i)$, $A(i) \in N$, then*

$$M_j^{(B)} - m_j^{(B)} \leqq (1 - \delta)^m \qquad j = 1,2,\cdots,n$$

where δ is the minimal nonzero element among all elements of the matrices in N.

Proof. (This proof generalizes the one given in [3, p. 173] for the stationary case). Let A be any matrix in $N, A = \| a_{ij} \|$. Consider the sum $\sum_{k=1}^{n} (a_{\alpha k} - a_{\beta k})$ for fixed α and β.

Divide this sum into two sums, $\sum_{k_1}^{+}$ denoting summation over values of k for which $a_{\alpha k} \geqq a_{\beta k}$, and $\sum_{k_2}^{-}$ denoting summation over values of k for which $a_{\alpha k} < a_{\beta k}$. Then:

$$\sum_{k_1}^{+} (a_{\alpha k_1} - a_{\beta k_1}) + \sum_{k_2}^{-} (a_{\alpha k_2} - a_{\beta k_2}) = \sum_{k=1}^{n} (a_{\alpha k} - a_{\beta k}) = \sum_{k=1}^{n} a_{\alpha k} - \sum_{k=1}^{n} a_{\beta k} = 1 - 1 = 0.$$

Hence

$$(*) \qquad \textstyle\sum_{k_1}^{+} = \sum_{k_2}^{-}.$$

States α and β have a common consequent of order one (condition H_3); hence there exists $a_{\bar{k}}$ such that $a_{\alpha\bar{k}} > 0$ and $a_{\beta\bar{k}} > 0$. If $a_{\alpha\bar{k}} \geqq a_{\beta\bar{k}}$, then $(a_{\alpha\bar{k}} - a_{\beta\bar{k}}) \in \sum_{k_1}^{+}$, and

$$\textstyle\sum_{k_1}^{+} (a_{\alpha k_1} - a_{\beta k_1}) = \sum_{k_1}^{+} a_{\alpha k_1} - \sum_{k_1}^{+} a_{\beta k_1} \leqq \displaystyle\sum_{k=1}^{n} a_{\alpha k} - a_{\beta\bar{k}} \leqq 1 - \delta$$

If $a_{\beta\bar{k}} < a_{\beta\bar{k}}$, then $(a_{\alpha\bar{k}} - a_{\beta\bar{k}}) \notin \sum_{k_1}^{+}$, and

$$\textstyle\sum_{k_1}^{+} (a_{\alpha k_1} - a_{\beta k_1}) \leqq \sum_{k_1}^{+} a_{\alpha k_1} \leqq 1 - \delta$$

Thus in both cases

$$(**) \qquad \textstyle\sum_{k_1}^{+} (a_{\alpha k_1} - a_{\beta k_1}) \leqq 1 - \delta.$$

Let $C = \| c_{ij} \| \in N$. We obtain

$$M_j^{(AC)} - m_j^{(A\cdot C)} = \max_{\alpha,\beta} \sum_k (a_{\alpha k} - a_{\beta k}) c_{kj} \leqq$$

$$\leqq \max_{\alpha,\beta} \left\{ \Sigma_{k_1}^{+}(a_{\alpha k_1} - a_{\beta k_1}) M_j^{(C)} + \Sigma_{k_2}^{-}(a_{\alpha k_1} - a_{\beta k_2}) m_j^{(C)} \right\} =$$

$$= \max_{\alpha,\beta} \sum_{k_1}{}^{+}(a_{\alpha k} - a_{\beta k})(M_j^{(C)} - m_j^{(C)}) \leqq (1-\delta)(M_j^{(C)} - m_j^{(C)});$$

due to (*) and (**)

$$M_j^{(C)} - m_j^{(C)} = c_{\alpha j} - c_{\beta j} \leqq \Sigma_{k_1}^{+}(c_{\alpha k} - c_{\beta k}) \leqq 1 - \delta$$

for fixed α and β, so that $M_j^{(AC)} - m_j^{(A\cdot C)} \leqq (1-\delta)^2$. The theorem follows by induction.

THEOREM 9. *If A and B are stochastic matrices, then* $K_r(AB) = K_r(A)\cdot K_r(B)$.

Proof. By definition $\bar{K}_r(AB) = AB - \xi(AB)_{(r)} = AB - \xi A_{(r)}B =$
$= AB - \xi A_{(r)}B + \xi B_{(r)} - \xi B_{(r)} = AB - \xi A_{(r)}B + \xi A_{(r)}\xi B_{(r)} - A\xi B_{(r)}$
$= [A - \xi A_{(r)}][B - \xi B_{(r)}] = \bar{K}_r(A)\bar{K}_r(B)$. Thus $\bar{K}_r(AB) = \bar{K}_r(A)\cdot \bar{K}_r(B)$.
The r-th rows of the matrices on both sides of this equations are zero rows. The r-th column on the right is obtained by multiplying $\bar{K}_r(A)$ by the r-th column of $\bar{K}_r(B)$. Thus omitting the r-th rows and columns on both sides does not affect the equality, and therefore

$$K_r(AB) = K_r(A)\cdot K_r(B).$$

THEOREM 10. *If A and B are stochastic matrices then*

$$(\overline{AB})_{(r)} = \bar{A}_{(r)}K_{(r)}(B) + \bar{B}_{(r)}$$

Proof. $A_{(r)}\bar{K}_r(B) + B_{(r)} = A_{(r)}[B - \xi B_{(r)}] + B_{(r)} = A_{(r)}B - A_{(r)}\xi B_{(r)} + B_{(r)} =$
$= A_{(r)}B - B_{(r)} + B_{(r)} = A_{(r)}B = (AB)_{(r)}$. Hence, $(AB)_{(r)} = A_{(r)}\bar{K}_{(r)}(B) + B_{(r)}$, and by considerations similar to those in the proof of the preceding theorem we obtain

$$(\overline{AB})_{(r)} = \bar{A}_{(r)}K_{(r)}(B) + \bar{B}_{(r)}.$$

By induction, we obtain the formula

$$\overline{\left(\prod_{i=1}^{m} A(i)\right)}_{(r)} = \bar{A}_{(r)}K_r(A(2))\cdot K_r(A(3))\cdot \cdots \cdot K_r(A(m)) +$$
$$+ \bar{A}(2)_{(r)}K_r(A(3))K_r(A(4))\cdot \cdots \cdot K_r(A(m)) + \cdots + \bar{A}(m)_{(r)}.$$

THEOREM 11. *If H is a set of matrices all satisfying* H_3, *then*

$$\lim_{m\to\infty} K_r\left(\prod_{i=1}^{m} A(i)\right) = 0,$$

where $A(i) \in N$ *and* 0 *is the zero matrix.*

Proof. The terms in $K_r(\prod^m A(i))$ are differences of pairs of terms in the same column of the product matrix $(\prod^m(A(i))$. By Theorem 8, these differences are smaller than $(1-\delta)^m$ where δ is the smallest nonzero element in all matrices contained in N. But $\lim_{m\to\infty}(1-\delta)^m = 0$, and the proof is complete.

REMARKS AND CONCLUSIONS.

1) Consider the following condition for a set N of stochastic matrices:

Condition H_4. A finite set of stochastic matrices of the same order satisfies H_4 (of order k) if there is a k such that every product of k or more matrices from N is a matrix satisfying H_3.

In a forthcoming paper we shall prove the following:

THEOREM. *Let* $N = \{A_i \mid i = 1 \cdots n\}$ *be a finite set of stochastic matrices o the same order. Let* B_m *be any matrix of the form* $B_m = \prod_{j=1}^{m} A_j$, $A_j \in N$. *Define* $\| B_m \|$ *as:* $\| B_m \| = \max_i (M_i^{(B_m)} - m_i^{(B_m)})$ ($M_i^{(B_m)}$ and $m_i^{(B_m)}$ *are as in Theorem* 8). *If and only if* N *satisfies* H_4 *then* $\lim_{m\to\infty} \| B_m \| = 0$.

This theorem generalizes Theorem 8. Moreover, it will be shown in that paper that H_4 is a decidable condition, i.e. one can check in a finite number of steps whether a given set of matrices N satisfies H_4.

2) If N contains a single matrix, we revert to the stationary case. In this case Theorem 12 states that $\lim_{n\to\infty}[K_r(A)]^n = 0$. By Theorem 11 we find that

$$\bar{A}^n_{(r)} = \bar{A}_{(r)}[K_r(A)]^{n-1} + \bar{A}_{(r)}[K_r(A)]^{n-2} + \cdots + \bar{A}_{(r)} = \bar{A}_{(r)} \sum_{i=0}^{m-1} [K_r(A)]^i.$$

In the limit this becomes

$$\lim_{n\to\infty} \bar{A}^n_{(r)} = \bar{A}_{(r)} \lim_{n\to\infty} \sum_{i=1}^{n-1} [K_r(A)]^i.$$

Consider the following

THEOREM 12*. *If* A *is a matrix such that* $\lim A^n = 0$ *then* $|I - A| \neq 0$ *and* $\lim \sum_{i=0}^{n} A_i = [I - A]^{-1}$.

With the aid of this theorem, noting that $K_r(A)$ satisfies its condition, (if A satisfies H_3) we obtain

$$\lim_{n\to\infty} \bar{A}^n_{(r)} = \bar{A}_{(r)}[I - K_r(A)]^{-1},$$

a formula derived earlier in a different manner.

3) If N contains an infinite set of matrices, Theorems 8 and 11 may be proved by stipulating that all nonzero terms in all matrices of the (infinite) set N have a lower bound $\varepsilon > 0$ and that all matrices in N satisfy H_3.

4) If N is a finite set of doubly stochastic matrices satisfying H_3, then the products of the form $\Pi_i^m A_i, A_i \in N$, have a limit or, more precisely, the matrix Q all whose terms equal $1/n$ satisfies the equation

*) For proof, see [4, p. 22].

$$\lim_{m \to \infty} \prod_{i}^{m} A_i = Q.$$

5) With regard to long products of stochastic matrices it is of interest to inquire whether there exists a procedure for approximating such a product. Suppose, for example, that A is a stochastic matrix which can be written in the form $A = I + \varepsilon$ where terms in ε are small. The usual approximation $A^n = I + n\varepsilon$, ceases to be stochastic for sufficiently large n. However, it is seen from Theorems 8, 10 and 11 that a good stochastic approximation to a long product of stochastic matrices (all satisfying H_3 and all belonging to a finite set N of matrices) is obtained by omitting the first k matrices in the product, where k is an easily computable function of the error allowed. For, as is seen from the above theorems, the first k matrices in the product contribute to the terms in the rth row of the product of m matrices at most:

$$(1-\delta)^{m-1} + (1-\delta)^{m-2} + \cdots + (1-\delta)^{m-k} = \frac{(1-\delta)^{m-k} - (1-\delta)^m}{\delta}$$

while those in the other rows differ from their counterparts in the r-th row by at most $(1-\delta)^m$, where δ is defined as in Theorem 9 and $(1-\delta)$ is the maximal term in all kernels of the matrices in the finite set N.

Acknowledgements. The author is indebted to Dr. M. Yoeli for guidance, advice and a number of suggestions followed up in the paper; to Prof. H. Hanani, for guidance and advice; to Prof. M. O. Rabin of the Hebrew University for encouragement is undertaking the present study; and to Mr. E. Goldberg for editorial assistance.

BIBLIOGRAPHY

1. Berge, C., 1962, *The Theory of Graphs and its Applications*, Wiley, New York.
2. Doob, J. L., 1953, *Stochastic Processes*, Wiley, New York.
3. Feller, W., 1950, *Probability Theory and Applications*, Wiley, New York.
4. Kemeny, J. C. and Snell, J. L., 1960, *Finite Markov Chains*, Van Nostrand, Princeton.

TECHNION–ISRAEL INSTITUTE OF TECHNOLOGY,
HAIFA