PACIFIC JOURNAL OF MATHEMATICS
Vol. 16, No. 3, 1966

INFINITE PRODUCTS OF SUBSTOCHASTIC MATRICES

N. J. PULLMAN

This paper is about two types of infinite products of substochastic matrices $\{A_j\}$ namely: the left product defined by the sequence of left partial products $A_1, A_2A_1, A_3A_2A_1, \cdots$; and the right product defined by the sequence of right partial products $A_1, A_1A_2, A_1A_2A_3, \cdots$.

The basic theorem is that if the A_n are each ∞ by ∞ then:

a. There is a nonempty set E of substochastic sequences each of which (except possibly the zero sequence, 0) is the componentwise limit of a sequence of rows, one from each left partial product;

b. Any sequence $\{\rho_n\}$ of rows, one from each left partial product, can be approximated by a sequence of convex combinations $\{c_n\}$ of points of E (that is, $\{\rho_n - c_n\}$ converges componentwise to the zero sequence), and c. $E = \{0\}$ if and only if every sequence of rows, one from each left partial product, converges to 0.

Similar conclusions follow immediately for the right product of ∞ by ∞ doubly substochastic matrices.

The asymptotic behaviour of the right product of a special class of $\{A_n\}$ is also considered.

The finite case (that is, when all the A_n are r by r) for stochastic A_n is treated independently for convenience, even though the result in this case (Theorem 1) is actually a direct consequence of the basic Theorem 1'. Its conclusion is that there is an m by r stochastic matrix A with $1 \leqq m \leqq r$ and permutation matrices Q_n such that

a. if $m < r$ then for some stochastic $r - m$ by m matrices C_n:

$$\lim_{n\to\infty}\left\{A_nA_{n-1}\cdots A_1 - Q_n\begin{pmatrix} A \\ C_nA\end{pmatrix}\right\} = 0$$

and b. if $m = r$ then

$$\lim_{n\to\infty}\{A_nA_{n-1}\cdots A_1 - Q_nA\} = 0 \; .$$

Some results on fixed points are obtained in the finite case which carry over, in restricted form, to the infinite case.

A real matrix is said to be stochastic if none of its entries is negative and each of its row sums is 1. Two types of infinite products which arise naturally from a given sequence $\{A_n\}$ of stochastic matrices are those whose nth partial products are $R_n = A_1A_2 \cdots A_n$ and $L_n = A_nA_{n-1} \cdots A_1$ respectively. We'll call the sequence $\{R_n\}$ the right

Received October 9, 1964.

product and the sequence $\{L_n\}$ the left product of the A_n.

The right product is of interest in the theory of Markov chains with possibly nonstationary transition probabilities because if A_n is the matrix of probabilities $a_{ij}^{(n)}$ of transition from state i at time $n-1$ to state j at time n then the ijth entry $r_{ij}^{(n)}$ of R_n is the probability of transition to state j at time n from state i at time 0.

The left product has a similar interpretation: $l_{ij}^{(n)}$ is the probability of transition from state i at time $-n$ to state j at time 0.

We shall obtain theorems on the asymptotic behaviour of these partial products and on their fixed points. For example if the A_n are ∞ by ∞ stochastic matrices we can show that there is a sequence of rows, one from each L_n, which converges componentwise.

The finite and infinite cases are treated separately for clarity.

DEFINITION. A *permutation matrix* is a matrix of zeroes and ones which exactly one 1 in each row and each column.

THEOREM 1. *If $L_n = A_nA_{n-1}\cdots A_1$ and each A_n is an r by r stochastic matrix then there exists an m by r stochastic matrix A with $1 \leqq m \leqq r$, r by r permutation matrices Q_n and, if $m < r$, stochastic $r-m$ by m matrices C_n such that*:

$$\lim_{n\to\infty}\left\| L_n - Q_n\begin{pmatrix} A \\ C_nA\end{pmatrix}\right\| = 0 \quad if \quad m < r \quad and$$
$$\lim_{n\to\infty}\| L_n - Q_nA\| = 0 \quad if \quad m = r\,.$$

Proof. Let S be the convex hull of the basis vectors $v_1 = (1, 0, 0, \cdots, 0)$, $v_2 = (0, 1, 0, \cdots, 0), \cdots$, $v_r = (0, 0, 0, \cdots, 1)$. Each $(S)L_n$ is a convex polytope (that is, the convex hull of p points), these polytopes are nested (that is, $(S)L_{n+1} \subseteqq (S)L_n$ for all n) and none of them has more than r vertices (a point x of a polytope is a *vertex* if it is on no open line segment contained in the polytope). It can be shown that the intersection of such a family of convex polytopes is a convex polytope of r or fewer vertices. Let $K = \bigcap_{n\geq 1}(S)L_n$ and denote its vertices by $k_1, \cdots, k_m$. Let A be the m by r matrix whose ith row is k_i. Let $v_i^{(n)}$ denote $(v_i)L_n$. For each n and each $t \leqq$ m there is a $v_{i_t}^{(n)}$ such that $k_t = \lim_{n\to\infty} v_{i_t}^{(n)}$. We can assume that for each n there are only m such $v_{i_t}^{(n)}$ so chosen. If $m < r$ extend the definition of i_t so that $\{v_{i_t}^{(n)} : m < t \leqq r\}$ is the set of $v_i^{(n)}$ not already chosen. Q_n is the matrix $(q_{ij}^{(n)})$ for which $q_{it}^{(n)}$ is 1 if $i = i_t$ and is 0 otherwise. If $m < r$ and $t > m$ let $k_t^{(n)}$ be the point of K closest to $v_{i_t}^{(n)}$. Since K is convex, $k_t^{(n)}$ is a convex combination, $\sum_{j=1}^m c_{tj}^{(n)}k_j$, of the vertices K. Therefore $C_n = (c_{ij}^{(n)})$ is an $r-m$ by m stochastic matrix and $k_t^{(n)} = (v_t)\begin{pmatrix} A \\ C_nA\end{pmatrix}$ for each $m < t \leqq r$. Consequently $(v_{i_t})Q_n\begin{pmatrix} A \\ C_nA\end{pmatrix} = (v_t)\begin{pmatrix} A \\ C_nA\end{pmatrix}$ if

$m < r$ and $(v_{i_t})Q_n A = (v_t)A$ if $m = r$. Theorem 1 then follows from the fact that $\lim_{n\to\infty} v_{i_t}^{(n)} = k_t$ if $1 \leqq t \leqq m$ and $\lim_{n\to\infty} v_{i_t}^{(n)} - k_t^{(n)} = 0$ if $t > m$.

Notice that $\lim_{n\to\infty} L_n = \begin{pmatrix} k_1 \\ k_1 \\ \vdots \\ k_1 \end{pmatrix}$ if m = 1 because K then consists of the one vertex k_1.

DEFINITION. A sequence $\{P_n\}$ of r by r matrices is *descending* if and only if $(S)P_{n+1} \subseteqq (S)P_n$ for all n sufficiently large. (S is as in the proof of Theorem 1). As a first corollary to Theorem 1 we have: m, Q_n, A and C_n (if $m < r$) such that $\lim_{n\to\infty} \left\| P_n - Q_n \begin{pmatrix} A \\ C_n A \end{pmatrix} \right\| = 0$ if $m < r$ and $\lim_{n\to\infty} \| P_n - Q_n A \| = 0$ if $m = r$, for all descending sequences because each such sequence (with the first N terms omitted) is the left product of some sequence of stochastic matrices. (All left products of stochastic matrices are, of course, descending sequences.) Another immediate corollary concerns doubly stochastic matrices (that is, stochastic matrices whose transposes are also stochastic). We shall state the corollary emphasizing the matrix entries for variety's sake.

COROLLARY 2. *If $\{A_n\}$ is a sequence of doubly stochastic r by r matrices and $R_n = A_1 A_2 \cdots A_n$ then there exists an m by r stochastic matrix A with $1 \leqq m \leqq r$ and permutations q_n of the r indices such that for each $1 \leqq j \leqq r$:*

(a) *if $1 \leqq q_n(i) \leqq m$, $\lim_{n\to\infty} (r_{ji}^{(n)} - a_{q_n(i)j}) = 0$ and if $m < r$ there exist $r - m$ by m stochastic matrices C_n such that*

(b) *if $m < q_n(i) \leqq r$ then:*

$$\lim_{n\to\infty} (r_{ji}^{(n)} - \sum_{k=1}^{m} c^{(n)} q_n(i) k^a k_j) = 0 .$$

Some examples of $\{A_n\}$ with descending right products are provided by all those sequences of stochastic matrices $\{A_n\}$ which commute pairwise within a row permutation (i.e. $A_n A_{n'} = Q_{nn'} A_{n'} A_n$ for some permutation matrix $Q_{nn'}$). Because of their connection with Markov chains we shall investigate descending right products further. We'll impose further conditions on the A_n which are not too stringent but which give additional information about the C_n of Theorem 1. While doing so we acquire some information on the fixed points of A_n and R_n.

DEFINITION. B occurs frequently among the A_n if and only if $B = A_n$ for infinitely many n.

LEMMA. *If $\{A_n\}$ is a sequence of r by r stochastic matrices whose* *right partial products $R_n = A_1A_2 \cdots A_n$ are a descending sequence and B occurs frequently among the A_n then, in the notation of Theorem 1 there is an m by m permutation matrix D such that $AB = DA$.*

Proof. For some N, $\{R_{N+n}\}$ is the left product of some sequence of stochastic matrices A'_n. Let K be as in the proof of Theorem 1 applied to the A'_n. Then $K = \bigcap_{n>N}(S)R_n$. $(K)B \subseteqq K$ because $K = \bigcap\{(S)R_{n-1} : B = A_n \text{ and } n > N\}$. Suppose $x \in K$. Then, for infinitely many n, there are $x_n \in (S)R_{n-1}$ for which $x = (x_n)B$. A subsequence $\{x_{n_m}\}$ converges to some point $y \in S$. Therefore $(x_{n_m})B$ converges to $(y)B$ and hence $x = (y)B$. But $y \in K$ and hence $K \subseteqq (K)B$. Thus $K = (K)B$ and hence B permutes the vertices of K (rows of A). Let D be the m by m permutation matrix representing this row permutation then $AB = DA$.

B permutes all the vertices of K and fixes the barycentre, $1/m' \sum_{t=1}^{m'} k_{i_t}$, of each subset $\{k_{i_1}, k_{i_2}, \cdots, k_{i_{m'}}\}$ of m' vertices of K (rows of A) which it permutes. Therefore $(x)B = x$ for all x in the convex hull of these barycentres. There may be (left) fixed points of B outside the convex hull of the barycentres.

Let us enumerate all the matrices occurring frequently among the A_n so that A_{n_1} is the first such matrix and A_{n_p} is the pth such matrix distinct from $A_{n_{p-1}}$. Let D_{n_p} be the m by m permutation matrix corresponding to A_{n_p} (as in the lemma) and let $D_n = D_{n_p}$ if $A_n = A_{n_p}$. Applying the lemma to the first corollary to Theorem 1 we obtain:

THEOREM 2. *If $\{A_n\}$ is a sequence of r by r stochastic matrices each of which (except for finitely many n) occur frequently among the A_n and the n-th partial products $R_n = A_1A_2 \cdots A_n$ are descending then there exists an m by r stochastic matrix A (with $1 \leqq m \leqq r$), permutation matrices Q_n and, if $m < r$, $r - m$ by m stochastic matrices C_n such that given $\varepsilon > 0$ there is an N for which:*

$$\text{(a)} \quad \left\| R_n - Q_N \begin{pmatrix} D'_n A \\ C_N D'_n A \end{pmatrix} \right\| < \varepsilon \quad (if\ m < r)\,,$$

$$\text{(b)} \quad \| R_n - Q_N D'_n A \| < \varepsilon \quad (if\ m = r)\,,$$

for all $n > N$. D'_n is the permutation matrix which is the product $D_{N+1}D_{N+2} \cdots D_n$ of D_s defined in the previous paragraph. Moreover the barycentres of those sets of rows of A which are permuted by all the D_{n_p} is a (left) fixed point for all A_n (except perhaps the finitely many n for which A_n does not occur frequently). In particular the barycentre $b = 1/m \sum_{i=1}^{m} (a_{i1}, \cdots, a_{ir})$ of the rows of A is such a (left) fixed vector.

Let F be the convex hull of the barycentres mentioned in Theorem

2. F is fixed (pointwise) by each of those A_n which occur frequently. If all the A_n occur frequently then $(x)R_n = x$ for all n and all $x \in F$.

The fundamental theorems on the convergence of the powers of the transition matrix and the "classification of states" of a finite Markov chain with stationary transition probabilities (see for example [4] pp. 170–184) can be obtained from Theorem 1 by examination of the position of K in S. In the interest of brevity we shall not do so here but shall instead discuss two notions from the stationary case by way of sample applications of Theorems 1 and 2.

In the notation of the proof of Theorem 1 let T be the set of all i for which v_i is not in the set of basis vectors spanning K. Following the custom (see e.g. [2]) for the stationary case we'll say that i *leads to* j (written $i \leadsto j$) if and only if $r_{ij}^{(n)} > 0$ for some n. If the right product of the A_n is descending then for each i, $\lim_{n\to\infty} r_{ij}^{(n)} = 0$ for all $j \in T$ and; each $i \in T$ leads to some $j \notin T$ by the first corollary to Theorem 1. In the stationary case (i.e. when $A_n = A_1$ for all n):

$$T = \bigcup_{j \geqq 1} \{i : i \leadsto j \text{ and } j \not\leadsto i\} .$$

This is precisely the definition of the set of *transient* (sometimes called *inessential*) *states* in the stationary case.

The notion of *regular* chain (in the terminology used in [6]) can be extended to the nonstationary case so as to obtain the same kind of basic result. Suppose the right product of the A_n is descending and that there is a product $P = A_{n_{p_1}} A_{n_{p_2}} \cdots A_{n_{p_q}}$ of frequently occuring $A_{n_{p_i}}$ (in the notation of Theorem 2) which is positive (i.e. $p_{ij} > 0$, all i, j). (The n_{p_i} are not necessarily distinct nor in increasing order). Call such $\{A_n\}$ *regular* sequences. It then follows that the right products R_n of regular sequences $\{A_n\}$ converge to a matrix all of whose rows are the vector k. No component of k is zero, $(k)R_n = k$ for all sufficiently large n (for *all* n, if $(S)R_{n+1} \subseteqq (S)R_n$ for all n) and k is the only vector in S with this property. Although this is equivalent to the corresponding result for the stationary case it is easy enough to obtain using the first corollary to Theorem 1 and the lemma preceding Theorem 2: All we need do is show that $m = 1$. To this end observe that according to the lemma, P permutes the vertices of K so that, for some n: $(x)P^n = x$ for all $x \in K$. If K had more than one vertex the line joining two of them would meet the boundary of S in a point x which is fixed by P^n. $(x)P^n$ can have no zero components because P is positive but x has zero components because it's in the boundary of S. This second application may also be found in a slightly less general form as Theorem 3 of [5].

DEFINITION. A real matrix is *substochastic* if and only if none of its entries is negative and 1 is an upper bound for its row sums.

Most of the foregoing results including Theorem 1 and its corollaries can be extended to infinite as well as finite substochastic matrices. To do so, consider the set S_0 of all substochastic sequences (i.e. the set of all real sequences of nonnegative terms whose sum is at most 1). S_0 is a compact, convex subset of the space of all real sequences under the product topology. The ∞ by ∞ substochastic matrices are associative and closed under matrix multiplication so that left and right product is defined for every sequence of such matrices.

THEOREM 1′. *If $\{L_n\}$ is the left product of a sequence of ∞ by ∞ substochastic matrices then there is a nonempty set, E, of substochastic sequences with the following properties:*

(a) *For each $k \in E$ (except possibly the zero sequence) and each n there is an integer $i_{n,k}$ such that for all j:*

$$\lim_{n\to\infty} l^{(n)}_{i_{n,k}j} = k_j \ .$$

(b) *For each sequence $\{i_n\}$ there is a convex combination $x^{(i,n)}$ of elements of E such that for all j:*

$$\lim_{n\to\infty} (l^{(n)}_{i_nj} - x^{(i,n)}_j) = 0 \ .$$

(c) *The zero sequence is the only element of E if and only if for all sequences $\{i_n\}$ and all j:*

$$\lim_{n\to\infty} l^{(n)}_{i_nj} = 0 \ .$$

Proof. For each subset F of S_0 let $co(F)$ be the set of convex combinations of elements of F and $\overline{co}(F)$ be the intersection of all closed convex sets containing F. Let W_n be the set consisting of 0 and all the rows of L_n, let $\bar{L}_n = \overline{co}(W_n)$ and $K = \bigcap_{n>1} \bar{L}_n$. K is convex and compact and $0 \in K$. Let E be the set of extremals of K (that is, $k \in E$ if and only if $k \in K$ and k is an interior point of no line segment in K) then $K = \overline{co}(E)$ by the Krein-Milman theorem. Part (a) of Theorem 1′ is proven by contradiction. Suppose $k \in E$ and a neighbourhood of k excludes 0 and all rows of L_n for all n in an infinite set Ω. Then, for a finite set Λ and some $\varepsilon > 0$, W_n is in the complement of $Z \equiv \bigcap_{j\in\Lambda} \{x \in S_0 : |x_j - k_j| < \varepsilon\}$ for each $n \in \Omega$.

Let $T_j^+ = \{x \in S_0 : x_j \geqq k_j + \varepsilon\}$, $T_j^- = \{x \in S_0 : x_j \leqq k_j - \varepsilon\}$ and $T_j = T_j^+ \cup T_j^-$. Then

$$K \subseteqq \bar{L}_n = \overline{co}\left(\bigcup_{j\in\Lambda} (T_j \cap W_n)\right)$$
$$= co\left(\bigcup_{j\in\Lambda} \overline{co}(T_j \cap W_n)\right) \qquad \text{(see [3] V 2.5)}$$

$$= co\left(\bigcup_{j\in\Lambda}\overline{co}((T_j^+ \cap W_n) \cup (T_j^- \cap W_n))\right)$$

$$= co\left(\bigcup_{j\in\Lambda} co(\overline{co}(T_j^+ \cap W_n) \cup \overline{co}(T_j^- \cap W_n))\right)$$

(again by [3] V 2.5)

$$\subseteqq co\left(\bigcup_{j\in\Lambda} co((T_j^+ \cap \bar{L}_n) \cup (T_j^- \cap \bar{L}_n))\right)$$

$$\subseteqq co\left(\bigcup_{j\in\Lambda} co(T_j \cap \bar{L}_n)\right).$$

If $U_{jn} = T_j \cap \bar{L}_n$ is empty for some $j \in \Lambda$, $n \in \Omega$ then $U_{jm} = \phi$ for all sufficiently large m because the U_{jn} are nested for fixed j. Rather than change notation, we can assume that $U_{jn} \neq \phi$ for all $n \in \Omega$ and all $j \in \Lambda$. Thus k is a convex combination, $\sum_{j\in\Lambda}\lambda_{jn}u_{jn}$, of elements u_{jn} of $co(U_{jn})$. U_{jn} is the union of $U_{jn}^+ = T_j^+ \cap \bar{L}_n$ and $U_{jn}^- = T_j^- \cap \bar{L}_n$. Assuming first that U_{jn}^+ and U_{jn}^- are nonempty for all $n \in \Omega$ we have $0 \leqq \mu_{jn} \leqq 1$ such that $u_{jn} = \mu_{jn}u_{jn}^+ + (1-\mu_{jn})u_{jn}^-$ for some $u_{jn}^+ \in U_{jn}^+$ and some $u_{jn}^- \in U_{jn}^-$. By successive extraction of subsequences we obtain u_j^+, u_j^-, μ_j and λ_j such that

$$\lim_{m\to\infty} u_{jn_m}^+ = u_j^+ , \qquad \lim_{m\to\infty} u_{jn_m}^- = u_j^- , \qquad \lim_{m\to\infty} \mu_{jn_m} = \mu_j ,$$

$$\lim_{m\to\infty} \lambda_{jn_m} = \lambda_j , \qquad 1 \geqq \mu_j \geqq 0 , \qquad 1 \geqq \lambda_j \geqq 0 \quad \text{and} \quad \sum_{j\in\Lambda}\lambda_j = 1 .$$

Therefore $k = \sum_{j\in\Lambda}\lambda_j(\mu_j u_j^+ + (1-\mu_j)u_j^-)$, and for all $j \in \Lambda$: u_j^+, $u_j^- \in K$ and u_j^+, $u_j^- \in T_j$. The extremality of k implies that $k = u_j^+$ or u_j^- for some j and hence that $k \in T_j$. Consequently $k \notin Z$, a contradiction. If, however, U_{jn}^+ or U_{jn}^- is ϕ for some (and hence all subsequent) n we can use a similar argument using the u_{jn} instead of the u_{jn}^+ and u_{jn}^-.

If $k \neq 0$ we can therefore assert that each sufficiently small neighbourhood of k excludes 0 but contains an element of W_n for all sufficiently large n. These elements must be rows of the L_n. Therefore k is the componentwise limit of a sequence of rows, one from each L_n.

To prove part (b) let d be the metric on S_0 which induces the product topology (see [1] II prop. 6, p. 97). Let $y_n \in L_n$ and z_n be a point of K closest to y_n in the metric. $d(z_n, y_n)$ is a null sequence because the $\bar{L}_n$ are nested. A sequence $\{x_n\}$ in $co(E)$ can be found for which $d(x_n, y_n)$ is a null sequence because $co(E)$ is dense in $\overline{co}(E)$ (see [3] V 2.4). Part (b) then follows if the i_nth row of L_n is used for y_n. Part (c) follows directly from parts (a) and (b). This completes the proof of Theorem 1′.

The conclusion of Theorem 1′ is valid if $\{L_n\}$ is replaced by any descending sequence $\{P_n\}$ of ∞ by ∞ substochastic matrices using the previous definition of "descending" with S replaced by S_0. Such sequences too are, except for finitely many terms, the left product of some sequence of substochastic matrices.

The statements about commutivity also carry over to the infinite case.

Corollary 2 extends to:

COROLLARY 2'. *If* $\{R_n\}$ *is the right product of* ∞ *by* ∞ *doubly stochastic matrices then there is a nonempty set, E, of substochastic sequences with the following properties:*

(a) *For each non-zero* $k \in E$ *and each* n *there is an integer* $i_{n,k}$ *such that for all* j:

$$\lim_{n\to\infty} r^{(n)}_{ji_{n,k}} = k_j \qquad \textit{and}$$

(b) *For each sequence* $\{i_n\}$ *there is a convex combination* $x^{(i,n)}$ *of elements of* E *such that for all* j:

$$\lim_{n\to\infty} r^{(n)}_{ji_n} - x^{(i,n)}_j = 0 \, ,$$

(c) *The zero sequence is the only element of* E *if and only if for all* $\{i_n\}$ *and for all* j:

$$\lim_{n\to\infty} r^{(n)}_{ji_n} = 0 \, .$$

A substochastic matrix is continuous on S_0 if and only if all of its columns are null sequences. If a continuous B occurs frequently among the A_n and their right product is descending then $(K)B = K$.

Theorem 2 and the remarks following it concerning fixed points also hold for ∞ by ∞ substochastic matrices A_n provided each A_n is continuous and K has only finitely many extremals.

REFERENCES

1. N. Bourbaki, Eléments de mathématiques, Livre V, *Espaces vectoriel topologiques*, Hermann, Paris, 1942.
2. K. L. Chung, *Markov Chains with Stationary Transition Probabilities*. Springer, Berlin, 1960.
3. N. Dunford and J. Schwartz, *Linear Operators, Part I*, Interscience, New York 1958.
4. J. L. Doob, *Stochastic Processes*, Wiley, New York, 1953.
5. J. Hajnal, The ergodic properties of non-homogeneous finite Markov chains, Proc. Camb. Phil. Soc. **52** (1956), 67-77.
6. J. Kemeny and J. Snell, *Finite Markov Chains*, Van Nostrand, New Jersey 1960.

MCGILL UNIVERSITY