

## PRODUCTS OF INDECOMPOSABLE, APERIODIC, STOCHASTIC MATRICES<sup>1</sup>

J. WOLFOWITZ

1. **Introduction.** A finite square matrix  $P = \{p_{ij}\}$  is called stochastic if  $p_{ij} \geq 0$  for all  $i, j$ , and  $\sum_j p_{ij} = 1$  for all  $i$ . A stochastic matrix  $P$  is called indecomposable and aperiodic (SIA) if

$$Q = \lim_{n \rightarrow \infty} P^n$$

exists and all the rows of  $Q$  are the same. SIA matrices are defined differently in books on probability theory; see, for example, [1] or [2]. The latter definition is more intuitive, takes longer to state, is easier to verify, and explains why the probabilist is interested in SIA matrices. A theorem in probability theory or matrix theory then says that the customary definition is equivalent to the one we have given. The latter is brief and emphasizes the property which will interest us in this note.

We define  $\delta(P)$  by

$$\delta(P) = \max_j \max_{i_1, i_2} |p_{i_1 j} - p_{i_2 j}|.$$

Thus  $\delta(P)$  measures, in a certain sense, how different the rows of  $P$  are. If the rows of  $P$  are identical,  $\delta(P) = 0$  and conversely.

Let  $A_1, \dots, A_k$  be any square matrices of the same order. By a word (in the  $A$ 's) of length  $t$  we mean the product of  $t$   $A$ 's (repetitions permitted).

The object of this note is to prove the following:

**THEOREM.** *Let  $A_1, \dots, A_k$  be square stochastic matrices of the same order such that any word in the  $A$ 's is SIA. For any  $\epsilon > 0$  there exists an integer  $\nu(\epsilon)$  such that any word  $B$  (in the  $A$ 's) of length  $n \geq \nu(\epsilon)$  satisfies*

$$\delta(B) < \epsilon.$$

In words, the result is that any sufficiently long word in the  $A$ 's has all its rows approximately the same.

It is sufficient to require that any word in the  $A$ 's be indecomposable. (This means, in the terminology of [1], that it contains at most one closed set of states other than the set of all states, or, in the

---

Presented to the Society, June 6, 1962; received by the editors June 7, 1962.

<sup>1</sup> Research under contract with the Office of Scientific Research of the U. S. Air Force.

terminology of [2], that it contains only one ergodic class.) For, as pointed out by Thomasian [6], if  $C$  is a word which is indecomposable and has period  $r > 1$ ,  $C^r$  is decomposable.

The theorem has applications to coding (information) theory (see [3]) and obvious applications to the study of nonhomogeneous Markov chains.

For  $2 \times 2$  matrices  $A$  the theorem is trivial. For then the determinant of each matrix is less than one in absolute value. The determinant of a long enough word has a value close to zero, so that the rows of the word are almost identical.

It would be desirable to have the hypothesis of the theorem require only that the  $A$ 's be SIA, but it is easy to prove by a counter-example that this would be insufficient. Indeed (e.g., [4, Equation 4(b)]), one can construct two SIA matrices  $A_1$  and  $A_2$  such that  $A_1 A_2$  is decomposable. Hence for no  $n$  does the word  $(A_1 A_2)^n$  have all its rows approximately the same. Thus the condition of the theorem is not only sufficient but also necessary.

Thomasian [6] has proved a result which will be described at the end of §2 and which implies an algorithm such that, in a bounded number of arithmetical operations, starting with  $A_1, \dots, A_k$ , one can determine whether *every* word in the  $A$ 's (of whatever length) is SIA. The number of arithmetical operations required by his method may still be very large. At the end of §2 we make a suggestion which will usually substantially reduce the amount of calculation required.

A result related to ours has been proved by Sarymsakov [5]. Lemma 1 of his paper (which he considers its most important result) is Lemma 4 of the present paper, with our  $t$  replaced by two less than the order of the  $A$ 's. Sarymsakov's conditions are only on the individual  $A$ 's (and not on all the words in the  $A$ 's) which is a pleasant feature. However, his conditions are rather strong and not obviously meaningful in a probabilistic sense. Essentially they are that each  $A$  remain SIA under permutation of rows and columns, the permutation of rows being *independent* of the permutation of the columns. As an example of the implications of this consider the matrix  $D$ , where

$$D = \begin{Bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ x & x & x \end{Bmatrix}.$$

Here the  $x$ 's stand for positive elements whose actual value is immaterial.  $D$  is obviously SIA. It does not fulfill the above described conditions of [5], as may be seen by putting the first column in third position.

2. **Proof of the theorem.** Henceforth, all matrices under discussion will be square and stochastic without any further mention of this fact. Capital Roman letters will always denote matrices. We will say that  $P_1$  and  $P_2$  are of the same type,  $P_1 \sim P_2$ , if they have zero elements and positive elements in the same places. Whether or not a matrix is SIA depends solely on its type.

Define

$$\lambda(P) = 1 - \min_{i_1, i_2} \sum_j \min(p_{i_1 j}, p_{i_2 j}).$$

If  $\lambda(P) < 1$  we will call  $P$  a scrambling matrix.  $\lambda(P) = 0$  if and only if  $\delta(P) = 0$  and conversely.  $\lambda(P) < 1$  implies that, for every pair of rows  $i_1$  and  $i_2$ , there exists a column  $b$  (which may depend on  $i_1$  and  $i_2$ ) such that  $p_{i_1 b} > 0$  and  $p_{i_2 b} > 0$ , and conversely. The introduction of  $\lambda$  is due to Hajnal [4]. Russian writers (e.g., [5]) use a related coefficient of "ergodicity."

We note the following for future reference:

LEMMA 1. *If one or more matrices in a product of matrices is scrambling, so is the product.*

This is Lemma 2 of [4].

LEMMA 2. *For any  $k$*

$$\delta(P_1 P_2 \cdots P_k) \leq \prod_{i=1}^k \lambda(P_i).$$

This is Theorem 2 of [4].

We now prove

LEMMA 3. *If  $P_2$  is an SIA matrix and*

$$(2.1) \quad P_1 P_2 \sim P_1,$$

*then  $P_1$  is a scrambling matrix.*

PROOF. From (2.1) we obtain that, for every positive integral  $n$ ,

$$(2.2) \quad P_1 P_2^n \sim P_1,$$

because

$$P_1 P_2^2 \sim P_1 P_2 \sim P_1.$$

From the definition of an SIA matrix it follows that, for  $n$  suffi-

ciently large,  $P_2^n$  is a scrambling matrix. The desired result now follows from (2.2) and Lemma 1.

The conclusion of this lemma could easily be strengthened, but the present lemma will suffice for our purposes.

Let  $t$  be the number of different types of all SIA matrices of the same order as the  $A$ 's.

**LEMMA 4.** *All words in the  $A$ 's of length  $\geq t+1$  are scrambling matrices.*

**PROOF.** Consider any word in the  $A$ 's of length  $t+1$ , say

$$B_1 \cdot B_2 \cdot \cdots \cdot B_{t+1}.$$

From the definition of  $t$  it follows that there exist integers  $a$  and  $b$ ,  $0 < a < b \leq t+1$  such that

$$(2.3) \quad B_1 \cdot B_2 \cdot \cdots \cdot B_a \sim B_1 \cdot B_2 \cdot \cdots \cdot B_b.$$

It follows from Lemma 3 that  $B_1 \cdot B_2 \cdot \cdots \cdot B_a$  is a scrambling matrix. The lemma now follows from Lemma 1.

**LEMMA 5.** *There exists a constant  $d$ ,  $0 \leq d < 1$ , with the following property: Let  $C$  be any word of length  $(t+1)$  in the  $A$ 's. Then  $\lambda(C) \leq d$ .*

**PROOF.** There are a finite number of words of length  $(t+1)$  in the  $A$ 's. By Lemma 4 each of these words has a  $\lambda$  which is less than one. Let  $d$  be the largest of these values of  $\lambda$ .

**PROOF OF THE THEOREM.** Let  $h$  be such that

$$d^h < \epsilon.$$

The theorem now follows from Lemma 2, with

$$v(\epsilon) = h(t+1).$$

The result of Thomasian's referred to earlier is as follows: Let  $\Gamma(k)$  be the set of types of all words in the  $A$ 's of length  $\leq k$ . If, for some  $k$ ,  $\Gamma(k) = \Gamma(k+1)$  and all the types in  $\Gamma(k)$  are those of SIA matrices, then all words in the  $A$ 's are SIA; the converse also holds. Let  $s$  be the order of the  $A$ 's. If all words in the  $A$ 's are SIA then the above property must hold for  $k \leq 2^s$ . (Actually it must hold for  $k \leq t$ .)

Since any word which has a scrambling matrix as a factor is SIA, we can at once disregard all  $A$ 's which are scrambling and consider the classes  $\Gamma$  only for those  $A$ 's which are not scrambling. Thus if all the  $A$ 's are scrambling no computations are necessary. In computing the  $\Gamma$ 's for the  $A$ 's which are not scrambling we need not consider further any word, a factor of which is scrambling. This procedure

will generally considerably reduce the amount of computation required. It is easy to verify whether or not a given matrix is scrambling.

The only place where we make use of the fact that there are only finitely many  $A$ 's is in Lemma 5. If there are infinitely many  $A$ 's but the conclusion of Lemma 5 holds then the theorem also holds. Some of the results on Markov chains in the literature are immediate consequences of this remark. In particular also, Lemma 4 always holds, even if there are infinitely many  $A$ 's.

#### REFERENCES

1. W. Feller, *An introduction to probability theory and its applications*, 2nd ed., Vol. 1, Wiley, New York, 1957.
2. J. L. Doob, *Stochastic processes*, Wiley, New York, 1953.
3. J. Wolfowitz, *Strong converse of the coding theorem for indecomposable channels*, *Sankhya* (1963). See also the abstract by the same title in *Ann. Math. Statist.* **33** (1962), 1212.
4. J. Hajnal, *Weak ergodicity in non-homogeneous Markov chains*, *Proc. Cambridge Philos. Soc.* **54** (1958), 233–246.
5. T. A. Sarymsakov, *Inhomogeneous Markov chains*, *Teor. Veroyatnost. i Primenen.* **6** (1961), 194–201.
6. A. J. Thomasian, *A finite criterion for indecomposable channels*, *Ann. Math. Statist.* **34** (1963), 337–338.

CORNELL UNIVERSITY