THE SEMIGROUP OF DOUBLY-STOCHASTIC MATRICES

by H. K. FARAHAT

(Received 21 July, 1965)

1. Introduction. The set D_n of all $n \times n$ doubly-stochastic matrices is a semigroup with respect to ordinary matrix multiplication. This note is concerned with the determination of the maximal subgroups of D_n. It is shown that the number of subgroups is finite, that each subgroup is finite and is in fact isomorphic to a direct product of symmetric groups. These results are applied in § 3 to yield information about the least number of permutation matrices whose convex hull contains a given doubly-stochastic matrix.

2. Groups of doubly-stochastic matrices. A square matrix with non-negative real elements is called *doubly-stochastic* if every row sum and every column sum is equal to unity. The set D_n of all $n \times n$ doubly-stochastic matrices is easily seen to be a semigroup with respect to ordinary matrix multiplication. The set P_n of all $n \times n$ permutation matrices (i.e. matrices obtained by permuting the columns of the identity matrix 1) is a subgroup of D_n which is obviously isomorphic to the symmetric group on n letters. We prove

(2.1) *If a matrix and its inverse belong to D_n, they belong to P_n.*

Proof. It is well known that the roots of a doubly-stochastic matrix lie in the closed unit disc. If $x \in D_n$, $x^{-1} \in D_n$, then, for every root λ of x we have $|\lambda| \leqq 1$ and $|\lambda^{-1}| \leqq 1$, and so $|\lambda| = 1$. This implies that $x \in P_n$ (see Lemma 1 and Theorem 5 of [7]).

For an arbitrary idempotent e of D_n we let G_e denote the maximal subgroup of D_n which contains e (cf. [3, Theorem 1]). When $e = 1$, the identity matrix, we have the group G_1 of all invertible elements of the semigroup D_n, i.e. of all invertible matrices of D_n whose inverses also belong to D_n. It follows from (2.1) that $G_1 = P_n$. We shall determine all subgroups G_e.

A mapping of the form $x \to u^{-1}xu$ defined by a permutation matrix u will be called a *cogredience*. Such a mapping obviously takes each maximal subgroup G_e to an isomorphic group $u^{-1}G_eu$, and the maximal subgroups thus fall into various cogrediency classes. In order to determine the structure of the subgroups G_e it is sufficient to consider one subgroup from each class. We begin by the determination of the possible idempotents. A matrix which is cogredient to one of the form

$$\begin{pmatrix} a & 0 \\ b & c \end{pmatrix}$$

will be called *reducible*; otherwise it is called *irreducible*. It is easily seen that every cogredience maps D_n into itself and that a reducible member of D_n is *decomposable* in the sense that b is necessarily also zero. Thus for doubly-stochastic matrices the notions of reducibility and decomposability coincide.

(2.2) *If e is an idempotent indecomposable doubly-stochastic $n \times n$ matrix, then every element of e is equal to $1/n$.*

https://doi.org/10.1017/S2040618500035401 Published online by Cambridge University Press

Proof.† The roots of an idempotent matrix e are 1 or 0, and the number of roots equal to 1 is the rank of e. If e is indecomposable and doubly-stochastic, it follows from the Perron-Frobenius theorem on non-negative matrices (cf. [8]) that 1 is a *simple* root of e and hence that e has rank one. The result now follows easily.

Let us denote the $m \times m$ idempotent matrix all of whose elements are equal to $1/m$ by $e(m)$. More generally, if $\lambda = (\lambda_1, \ldots, \lambda_k)$ is any partition of n, i.e. if $n = \lambda_1 + \ldots + \lambda_k$ with $\lambda_1 \geqq \lambda_2 \geqq \ldots \geqq \lambda_k > 0$, let us denote by $e(\lambda)$ the idempotent $n \times n$ matrix which is the direct sum of $e(\lambda_1), \ldots, e(\lambda_k)$:

$$e(\lambda) = e(\lambda_1) \oplus \ldots \oplus e(\lambda_k).$$

Clearly $e(\lambda)$ is an idempotent member of D_n, and, according to (2.2), every idempotent is cogredient to some $e(\lambda)$. It is clear that the cogrediency class of $e(\lambda)$ corresponds uniquely to the partition λ (with decreasing parts $\lambda_1 \geqq \lambda_2 \geqq \ldots \geqq \lambda_k > 0$), so that distinct partitions λ yield non-cogredient idempotents $e(\lambda)$. Since P_n is finite and the number of partitions of n is also finite, it follows that D_n has only a finite number of idempotents. Hence

(2.3) *The number of maximal subgroups of D_n is finite.*

In fact, we can easily determine this number, by computing the number of idempotents cogredient to $e(\lambda)$. There are altogether $n!$ idempotents $u^{-1}e(\lambda)u$ with $u \in P_n$, but each is repeated a number of times equal to the number of permutation matrices u which commute with $e(\lambda)$. If the partition λ has ρ_α parts equal to α $(1 \leqq \alpha \leqq n)$, then it is easily seen that this number is equal to $\prod_\alpha (\alpha!)^{\rho_\alpha}\rho_\alpha!$. It follows that the number of distinct idempotents in D_n is equal to

$$\sum \frac{n!}{(1!)^{\rho_1}\rho_1!(2!)^{\rho_2}\rho_2! \ldots (n!)^{\rho_n}\rho_n!},$$

where the sum extends over all partitions $n = \sum \rho_\alpha \alpha$ of n.

We now proceed to determine the structure of the maximal group $G(\lambda) = G_{e(\lambda)}$ containing the idempotent $e(\lambda)$, where λ is a fixed partition of n. To this end, let $e(p;\ q)$ denote the $p \times q$ matrix all of whose elements are equal to $1/q$. Thus it is obvious that $e(q;\ q) = e(q)$, and it is easily verified that $e(p;\ q)e(q;\ r) = e(p;\ r)$. If $\lambda = (\lambda_1, \ldots, \lambda_k)$ is any partition of n, put‡

$$e(*\lambda) = e(1;\ \lambda_1) \oplus \ldots \oplus e(1;\ \lambda_k),$$

$$e(\lambda*) = e(\lambda_1;\ 1) \oplus \ldots \oplus e(\lambda_k;\ 1),$$

so that $e(*\lambda)$ has k rows and n columns while $e(\lambda*)$ has n rows and k columns. The preceding remarks concerning $e(p;\ q)$ imply at once that

$$e(\lambda*)e(*\lambda) = e(\lambda), \quad e(*\lambda)e(\lambda*) = 1,$$

where of course 1 denotes the identity $k \times k$ matrix.

(2.4) *If x is an $r \times s$ matrix satisfying $e(r)x = x = xe(s)$, then x is a scalar multiple of $e(r;\ s)$.*

† I owe this simple proof to Miss Hazel Perfect.

‡ The direct sum $a \oplus b$ of rectangular matrices a, b is the matrix given in blocks (of obvious sizes) as follows:

$$a \oplus b = \left[\begin{array}{c|c} a & 0 \\ \hline 0 & b \end{array}\right].$$

Proof. The rows (and also the columns) of $e(r)$ are all equal. From $e(r)x = x$ it follows that the rows of x are all equal. Similarly, from $x = xe(s)$ we deduce that all the columns of x are equal and hence that all the elements of x are equal. The result follows.

Now let $x \in G(\lambda)$, and partition x into blocks corresponding to the equation

$$n = \lambda_1 + \ldots + \lambda_k.$$

Let $x(i, j)$ denote the block in the ith horizontal and jth vertical strips. From the relations $e(\lambda)x = x = xe(\lambda)$, which are valid because $e(\lambda)$ is the neutral element of $G(\lambda)$, we conclude that

$$e(\lambda_i)x(i, j) = x(i, j) = x(i, j)e(\lambda_j)$$

for all i, j. It follows from (2.4) that

$$x(i, j) = \xi_{ij}e(\lambda_i;\ \lambda_j), \quad (1 \leqq i, j \leqq k),$$

where ξ is a suitable non-negative $k \times k$ matrix. We shall prove, in fact, that ξ *is a permutation matrix*. Note firstly that

$$x(i, j) = e(\lambda_i;\ 1)\xi_{ij}e(1;\ \lambda_j),$$

whence

(2.5) $$x = e(\lambda *)\xi e(*\lambda), \quad \xi = e(*\lambda)xe(\lambda *).$$

Now each of $e(*\lambda)$, x, $e(\lambda *)$ is clearly row-stochastic (i.e. all row sums are equal to unity). It follows that ξ *itself is row-stochastic*. Observe secondly that the mapping $x \to \xi$ is a multiplicative homomorphism. Thus, if y denotes the inverse of x in $G(\lambda)$, and $\eta = e(*\lambda)ye(\lambda *)$, then we have

$$\xi\eta = e(*\lambda)xe(\lambda *)e(*\lambda)ye(\lambda *) = e(*\lambda)xe(\lambda)ye(\lambda *)$$
$$= e(*\lambda)xye(\lambda *) = e(*\lambda)e(\lambda *)e(*\lambda)e(\lambda *) = 1,$$

and similarly $\eta\xi = 1$. This means that ξ, η are both non-negative row-stochastic matrices and $\xi = \eta^{-1}$. It follows by an argument similar to that used in the proof of (2.1) that both ξ and η are permutation matrices (cf. the proof of Theorem 5 in [7], where the argument clearly applies to row-stochastic matrices.) We shall however indicate this proof briefly. The function

$$\| z \| = \max_i \sum_j | z_{ij} |$$

is a matrix norm†, and every row-stochastic matrix has unit norm: $\| \xi \| = \| \eta \| = 1$. But, for any matrix norm, we have $\| z \| \geqq | \alpha |$ for every root α of z. It now follows that ξ and $\eta = \xi^{-1}$ have all their roots on the unit circle. But Schur's inequality states that the sum of the squares of the moduli of the roots of a matrix does not exceed the sum of the squares of the moduli of the elements. Hence

$$k \leqq \sum_{i,j} | \xi_{ij} |^2 = \sum_{i,j} \xi_{ij}^2 \leqq \sum_i \Big(\sum_j \xi_{ij}\Big) = k,$$

† The axioms for a matrix norm are

(i) $\| z \| > 0$ for $z \neq 0$, (ii) $\| z' + z'' \| \leqq \| z' \| + \| z'' \|$,

(iii) $\| zw \| \leqq \| z \| \, \| w \|$, (iv) $\| \lambda z \| = | \lambda | \, \| z \|$.

They clearly imply that $\| z \| \geqq | \alpha |$ whenever $zx = \alpha x$, $x \neq 0$.

https://doi.org/10.1017/S2040618500035401 Published online by Cambridge University Press

since $\xi_{ij}^2 \leqq \xi_{ij}$, and hence equality holds throughout, so that $\xi_{ij}^2 = \xi_{ij}$ for all i, j and ξ is a permutation matrix.

We have now established that, for every $x \in G(\lambda)$, the matrix $\xi = e(*\lambda)xe(\lambda*)$ is a permutation matrix. We can say more about ξ however. As before, let λ stand for the row $(\lambda_1, \ldots; \lambda_k)$. Then clearly $\lambda e(*\lambda) = (1, 1, \ldots, 1)$, and because x is doubly-stochastic we find that

$$\lambda\xi = (1, \ldots, 1)e(\lambda*) = \lambda.$$

Of course, ξ is a permutation matrix, and the elements of λ are positive integers. As before, suppose that ρ_α of these elements are equal to α. The equation $\lambda\xi = \lambda$ then implies that ξ belongs to $P(\lambda) = P_{\rho_n} \oplus P_{\rho_{n-1}} \oplus \ldots \oplus P_{\rho_1}$, i.e., that ξ is the direct sum of permutation matrices of degrees $\rho_n, \rho_{n-1}, \ldots, \rho_1$ (obviously terms with $\rho_\alpha = 0$ are to be ignored). Conversely, it is plain that, if $\xi \in P(\lambda)$, then $x = e(\lambda*)\xi e(*\lambda)$ belongs to the group $e(\lambda*)P(\lambda)e(*\lambda)$, which must be $G(\lambda)$ because it contains $e(\lambda*)e(*\lambda) = e(\lambda)$. We have therefore proved the following:

(2.6) THEOREM. *For any partition λ of n, the mapping*

$$x \in G(\lambda) \to \xi \in P(\lambda),$$

where $\xi = e(*\lambda)xe(\lambda*)$, is a group isomorphism. In particular $G(\lambda)$ is a finite group of order $\rho_1! \ldots \rho_n!$.

Note that, when λ is the partition of n into n parts (each equal to 1), $e(\lambda)$ is the identity matrix 1 and $G(\lambda)$ is the group P_n of all $n \times n$ permutation matrices.

3. An application. Since the mappings $x \to \xi$, $\xi \to x$ described above are both linear, they establish an isomorphism between the convex hull $H(\lambda)$ of the elements of the group $G(\lambda)$ and the convex hull of the group $P(\lambda)$. Of course both of these are semigroups. It is well known that the convex hull of P_n is D_n (this is Birkhoff's theorem; cf. [**1**]). Thus the convex hull of $P(\lambda) = P_{\rho_n} \oplus P_{\rho_{n-1}} \oplus \ldots \oplus P_{\rho_1}$ is simply $D(\lambda) = D_{\rho_n} \oplus D_{\rho_{n-1}} \oplus \ldots \oplus D_{\rho_1}$. Thus we have

(3.1) *The semigroup $H(\lambda)$ is isomorphic with $D(\lambda)$.*

In this section we are interested in the least number $\nu(x)$ of permutation matrices whose convex hull contains a given doubly-stochastic matrix x. For a review of what is known about $\nu(x)$, see [**6**, p. 324, 325]. The main tool in giving an upper estimate for $\nu(x)$ is a theorem of Carathéodory (cf. [**2**, p. 35]), which may be stated in the following form (see also [**4**, Lemma 6]):

(3.2) (*Carathéodory*). *Let X be a finite subset of a linear variety of dimension d. Then every point of the convex hull of X lies in the convex hull of $d+1$ suitable points of X. The number $d+1$ is best possible.*

It is evident that D_n is contained in a linear variety of dimension $(n-1)^2$, and therefore the above theorem gives the estimate

$$\nu(x) \leqq (n-1)^2 + 1. \tag{3.3}$$

If no further information is given concerning x, this estimate is best possible. However, an estimate is obtained in [**5**] for indecomposable x, namely

$$\nu(x) \leqq c\left(\frac{n}{c} - 1\right)^2 + 1, \tag{3.4}$$

where c denotes the number of roots of x of unit modulus. We shall obtain a bound for $\nu(x)$, given that $x \in H(\lambda)$.

(3.5) *Let* $x_1, \ldots, x_l, y_1, \ldots, y_m$ *be elements of a real vector space* V. *Then every point in the convex hull of the points* $x_i + y_j$ $(1 \leqq i \leqq l, 1 \leqq j \leqq m)$ *lies in the convex hull of* $l+m-1$ *of them.*

Proof. The *direction* of the linear variety in V generated by the lm points $x_i + y_j$ is the vector space spanned by all differences $(x_i+y_j)-(x_\alpha+y_\beta) = (x_i - x_\alpha)+(y_j - y_\beta)$. The dimension of this linear variety (i.e. the dimension of its direction) is therefore not more than $(l-1)+(m-1) = l+m-2$. The result now follows from (3.2).

(3.6) *If* $x \in D_a$, $y \in D_b$, *then* $\nu(x \oplus y) \leqq \nu(x)+\nu(y)-1$.

Proof. Since $x \in D_a$, x lies in the convex hull of $\nu(x)$ permutation matrices x_i (say). Similarly, y lies in the convex hull of $\nu(y)$ permutation matrices y_j. Let $x = \sum \alpha_i x_i$, $y = \sum \beta_j y_j$, where $\alpha_i, \beta_j \geqq 0$, $\sum \alpha_i = 1$, $\sum \beta_j = 1$. Then clearly

$$x \oplus y = \sum_{i,j} (\alpha_i \beta_j)(x_i \oplus y_j),$$

so that $x \oplus y$ lies in the convex hull of the permutation matrices

$$x_i \oplus y_j \quad (1 \leqq i \leqq \nu(x), 1 \leqq j \leqq \nu(y)).$$

The result now follows from (3.5).

For $x \in H(\lambda)$, let $\nu_\lambda(x)$ denote the smallest number of elements of $G(\lambda)$ whose convex hull contains x. When λ has n parts equal to 1, $\nu_\lambda(x)$ coincides with $\nu(x)$. We prove

(3.7) *Let* $x \in H(\lambda)$ *and suppose that the non-zero* ρ_α *are* $\rho_{\alpha_1}, \ldots, \rho_{\alpha_t}$ $(\alpha_1 > \ldots > \alpha_t)$. *Then*

$$\nu_\lambda(x) \leqslant 1 + \sum_{i=1}^{t} (\rho_{\alpha_i} - 1)^2.$$

Proof. According to the remarks made at the beginning of this section, the matrix $\xi = e(*\lambda)xe(\lambda*)$ belongs to $D(\lambda)$, and has the form $\xi = \xi_1 \oplus \ldots \oplus \xi_t$, where $\xi_i \in D_{\rho_{\alpha_i}}$. Thus, by (3.3), $\nu(\xi_i) \leqq (\rho_{\alpha_i} - 1)^2 + 1$, and by repeated application of (3.6) we have

$$\nu_\lambda(x) = \nu(\xi) \leqq \sum_{i=1}^{t} \nu(\xi_i) - (t-1) \leqq \sum_{i=1}^{t} (\rho_{\alpha_i} - 1)^2 + 1,$$

as required.

Remark. If we suppose in this proof that each ξ_i is indecomposable and that ξ_i has c_i roots of unit modulus, then we can use (3.4) instead of (3.3) and obtain the inequality

$$\nu_\lambda(x) = \nu(\xi) \leqq 1 + \sum_{i=1}^{t} c_i(\rho_{\alpha_i}/c_i - 1)^2. \tag{3.8}$$

Finally in this section we prove

(3.9) *Suppose that* x *is an indecomposable member of* $H(\lambda)$. *Then the parts of* λ *are equal*: $\lambda_1 = \ldots = \lambda_k = l$ *(say). Furthermore, if* x *has* c *roots of unit modulus, then*

$$\nu(x) \leqq lc\left(\frac{k}{c} - 1\right)^2 + l.$$

Proof. We have $x(i,j) = \xi_{ij}e(\lambda_i;\ \lambda_j)$. If ξ were decomposable, a permutation π of $1, \ldots, k$ and integers β, β' would exist such that $\xi_{\pi i,\ \pi j} = 0$ whenever $i \leqq \beta, j \geqq \beta'$. But then $x(\pi i, \pi j) = 0$ for $i \leqq \beta, j \geqq \beta'$, so that x itself would be decomposable. Thus, if x is indecomposable then ξ is also indecomposable, and in particular the number t occurring in the proof of (3.7) must be 1. Hence $\rho_\alpha = 0$ except for one value of α, say $\alpha = l$, and $\rho_l = k$. This proves the first assertion. Suppose now in addition that x has exactly c roots of unit modulus. We show that the same is true of ξ, indeed that x and ξ have the same non-zero roots with the same multiplicities. Let R_α denote the vector space of all real column matrices with α elements. Then $z \in R_n \rightarrow xz \in R_n$ and $y \in R_k \rightarrow \xi y \in R_k$ are linear transformations, and it is easy to check that the mapping $y \in R_k \rightarrow e(\lambda*)y \in_n R$ is an " operator isomorphism " because $e(\lambda*)\xi = xe(\lambda)*$. The linear transformation x restricted to the subspace $e(\lambda*)R_k$ of R_n has therefore the same roots and multiplicities as the matrix ξ. Since $e(\lambda) = e(\lambda*)e(*\lambda)$ and $e(\lambda*) = e(\lambda)e(\lambda*)$, we have $e(\lambda*)R_k = e(\lambda)R_n$; it follows that R_n is the direct sum of $e(\lambda*)R_k$ and $(1-e(\lambda))R_n$. But x vanishes on the latter. Hence x and ξ have the same non-zero roots, with the same multiplicities, as asserted. Now we have that $\xi \in D_k$ is indecomposable and has c roots of unit modulus. It follows from (3.4) that $v_\lambda(x) = v(\xi) \leqq c(k/c-1)^2+1$. Now, in this case of equal parts, we have $\lambda_1 = \ldots = \lambda_k = l$ and so $x(i,j) = \xi_{ij}e(l;\ l) = \xi_{ij}e(l)$, that is, $x = \xi \otimes e(l)$, where $\otimes$ denotes the tensor product (or Kronecker product). But obviously $v(e(l)) \leqq l$, because $e(l)$ lies in the convex hull of the matrices $1, z, z^2, \ldots, z^{l-1}$, where z is the $l \times l$ permutation matrix corresponding to the cycle of length l, i.e.

$$z = \begin{pmatrix} 0 & 1 & 0 & . & . & 0 \\ 0 & 0 & 1 & . & . & 0 \\ . & . & . & . & . & . \\ . & . & \vdots & . & . & . \\ 0 & . & . & . & 0 & 1 \\ 1 & 0 & . & . & 0 & 0 \end{pmatrix}.$$

This implies that x lies in the convex hull of the matrices $\xi \otimes z^i$ $(0 \leqq i \leqq l-1)$, and hence $v(x) \leqq l[c(k/c-1)^2+1]$, as required.

REFERENCES

1. G. Birkhoff, Tres observaciones sobre el algebra lineal, *Univ. Nac. Tacumán, Rev. Ser. A*, **5** (1946), 147–150.

2. H. G. Eggleston, *Convexity* (Cambridge, 1958).

3. H. K. Farahat and L. Mirsky, Group membership in rings of various kinds, *Math. Z.* **70** (1958), 231–244.

4. H. K. Farahat and L. Mirsky, Permutation endomorphisms and a refinement of a theorem of Birkhoff, *Proc. Cambridge Philos. Soc.* **56** (1960), 322–328.

5. M. Marcus, H. Minc and B. Moyls, Some results on non-negative matrices, *J. Res. nat. Bur Standards* **65** B (1961), 205–209.

6. L. Mirsky, Results and problems in the theory of doubly-stochastic matrices. *Z. Wahrscheinlichkeitstheorie* **1** (1963), 319–334.

7. Hazel Perfect and L. Mirsky, Spectral properties of doubly-stochastic matrices, *Monatsh. Math.* **69** (1965), 35–57.

8. H. Wielandt, Unzerlegbare nicht-negative Matrizen, *Math. Z.* **52** (1950), 642–648.

THE UNIVERSITY, SHEFFIELD