COMPOSITIO MATHEMATICA

PETER FLOR

On groups of non-negative matrices

Compositio Mathematica, tome 21, nº 4 (1969), p. 376-382
<http://www.numdam.org/item?id=CM_1969__21_4_376_0>

COMPOSITIO MATHEMATICA, Vol. 21, Fasc. 4, 1969, pag. 376-382
Wolters-Noordhoff Publishing
Printed in the Netherlands

On groups of non-negative matrices

by

Peter Flor

Index of notation

$\mathscr{G}$: some group of non-negative matrices
I: the neutral element of $\mathscr{G}$
$N = \ker I$
$S = \operatorname{im} I$
P: the cone of non-negative vectors in R^n
$P_0 = IP$
$A_1, \cdots, A_k$: the edges of P_0
$a_j \in A_j$ $(j = 1, \cdots, k)$, $a_j \neq 0$
$E = \{e_1, \cdots, e_n\}$, the set of unit vectors in R^n
$M = N \cap P$
$M_j = \{x \in P | Ix \in A_j\}$
$E_0 = N \cap E$
$E_j = (M_j \cap E) \setminus E_0$ $(j = 1, \cdots, k)$
K_j: the linear hull of E_j
$a_j = b_j + c_j$, $b_j \in K_j$, $c_j \in M$.

There has been some interest, in recent years, in groups of non-negative matrices. Brown [1] discovered the striking fact that any such group is finite if it is compact, and Schwarz ([6], § 4) gave a detailed description of groups of stochastic matrices. In these papers, groups of non-negative matrices were studied in the context of semigroup theory. Both authors restricted their attention to compact groups.

In the present paper a fairly complete description of *all* groups of non-negative matrices will be given, without the restriction of compactness. This includes a proof of Brown's theorem as well as an extension of some results of Schwarz. Our method is quite elementary, contrasting with Brown's proof which uses Lie theory. Indeed it was the desire to prove Brown's theorem in an elementary manner which led to the results presented here.

Let $\boldsymbol{P} \subset \boldsymbol{R}^n$ be the cone of vectors with non-negative components, and denote by $\mathscr{G}$ any group of non-negative $n \times n$ matrices. Non-negativity means $\mathscr{G}\boldsymbol{P} \subset \boldsymbol{P}$. Denote by I the unit element of $\mathscr{G}$. I is an idempotent matrix. So $\boldsymbol{R}^n = \boldsymbol{N} \oplus \boldsymbol{S}$ where $\boldsymbol{N} = \ker I = \{x \in \boldsymbol{R}^n | Ix = 0\}$ and $\boldsymbol{S} = \operatorname{im} I = \{Ix | x \in \boldsymbol{R}^n\}$. The relations $G = GI = IG$, $I = GG^{-1} = G^{-1}G$ imply $\boldsymbol{N} = \ker G$, $\boldsymbol{S} = \operatorname{im} G$ for every $G \in \mathscr{G}$. Thus dim $\boldsymbol{S}$ is the rank of every element of $\mathscr{G}$ (the *rank of* $\mathscr{G}$, as we shall say for shortness). On $\boldsymbol{S}$, $\mathscr{G}$ acts as a transformation group in the usual sense; in particular, $Ix = x$ for every $x \in \boldsymbol{S}$.

Let $\boldsymbol{P}_0 = \boldsymbol{P} \cap \boldsymbol{S}$, $G \in \mathscr{G}$. We have $G\boldsymbol{P} \subset \boldsymbol{P}$ and also $G\boldsymbol{P} \subset \boldsymbol{S}$, that is, $G\boldsymbol{P} \subset \boldsymbol{P}_0$. Furthermore, $\boldsymbol{P}_0 = I\boldsymbol{P}_0 = GG^{-1}\boldsymbol{P}_0 \subset G\boldsymbol{P}_0 \subset G\boldsymbol{P}$. So $G\boldsymbol{P} = G\boldsymbol{P}_0 = \boldsymbol{P}_0$, for every $G \in \mathscr{G}$.

Like $\boldsymbol{P}$, $\boldsymbol{P}_0$ is a convex pyramid. We shall have to study certain extreme subsets of $\boldsymbol{P}_0$. A subset $\boldsymbol{Q}$ of a convex cone $\boldsymbol{C}$ is called *extreme* if $\boldsymbol{Q}$ is a convex cone and $a \in \boldsymbol{C}$, $b \in \boldsymbol{C}$, $a+b \in \boldsymbol{Q}$ imply $a \in \boldsymbol{Q}$, $b \in \boldsymbol{Q}$. A convex cone $\boldsymbol{C}$ is *spanned* by a set $\boldsymbol{D} \subset \boldsymbol{C}$ if $\boldsymbol{C}$ is the set of all linear combinations of elements of $\boldsymbol{D}$ with non-negative coefficients. Alternatively, $\boldsymbol{D}$ spans $\boldsymbol{C}$ if $\boldsymbol{C}$ is the smallest convex cone containing $\boldsymbol{D}$. The following lemma is obvious.

LEMMA 1. *If $\boldsymbol{D}$ spans $\boldsymbol{C}$ and $\boldsymbol{Q}$ is extreme in $\boldsymbol{C}$, $\boldsymbol{D} \cap \boldsymbol{Q}$ spans $\boldsymbol{Q}$.*

Define $\boldsymbol{M} = \boldsymbol{N} \cap \boldsymbol{P}$, $\boldsymbol{E}_0 = \boldsymbol{N} \cap \boldsymbol{E}$ where $\boldsymbol{E} = \{e_1, e_2, \cdots, e_n\}$ is the set of unit vectors in $\boldsymbol{R}^n$. $\boldsymbol{M}$ is extreme in $\boldsymbol{P}$, and so is spanned by $\boldsymbol{E}_0$, by Lemma 1. Further let $\boldsymbol{A}_1, \cdots, \boldsymbol{A}_k$ be the edges of $\boldsymbol{P}_0$, define $\boldsymbol{M}_j = \{x \in \boldsymbol{P} | Ix \in \boldsymbol{A}_j\}$, and let $\boldsymbol{E}_j$ be the complement of $\boldsymbol{E}_0$ in $\boldsymbol{M}_j \cap \boldsymbol{E}$, for $j = 1, 2, \cdots, k$. Every $\boldsymbol{M}_j$ is extreme in $\boldsymbol{P}$ since the $\boldsymbol{A}_j$ are extreme in $\boldsymbol{P}_0$. So by Lemma 1 again, $\boldsymbol{M}_j$ is spanned by $\boldsymbol{E}_0 \cup \boldsymbol{E}_j$. For $j \neq l$, we have $\boldsymbol{M}_j \cap \boldsymbol{M}_l = \boldsymbol{M}$, $\boldsymbol{E}_j \cap \boldsymbol{E}_l = \emptyset$.

Now select a non-zero vector a_j on every $\boldsymbol{A}_j$. For every j, there is a unique decomposition $a_j = b_j + c_j$ where b_j is in the linear hull of $\boldsymbol{E}_j$, and $c_j \in \boldsymbol{M}$. $Ia_j = a_j \neq 0$; so $a_j \notin \boldsymbol{M}$, $b_j \neq 0$. Since the $\boldsymbol{E}_j$ are mutually disjoint, the b_j are linearly independent. Furthermore, $c_j \in \boldsymbol{M}$ which is spanned by $\boldsymbol{E}_0$; this is disjoint from every $\boldsymbol{E}_j$ $(j = 1, \cdots, k)$. So any linear relation involving the a_j would entail the same relation for the b_j. We have proved

LEMMA 2: *$a_1, \cdots, a_k$ are linearly independent.*

$\boldsymbol{P}_0 = \boldsymbol{P} \cap \boldsymbol{S}$ is a pointed closed convex pyramid, that is, the solution set of a system of weak homogeneous linear inequalities, of maximal rank. It is well known that a closed pointed convex pyramid is spanned by its edges. (See [4], pp. 40—43, or [3], Corollary 1B.) So P, is spanned by $a_1, \cdots, a_k$. On the other hand,

$\boldsymbol{P}_0 = I\boldsymbol{P}$ contains $Ie_1, \cdots, Ie_n$; so its linear hull contains $\boldsymbol{S}$, and therefore is equal to $\boldsymbol{S}$. This shows that $\boldsymbol{S}$ is the linear hull of $\{a_1, \cdots, a_k\}$; by Lemma 2, $\{a_1, \cdots, a_k\}$ is a basis of $\boldsymbol{S}$, and $\dim \boldsymbol{S} = k$.

Now it is easy to describe the way in which $\mathscr{G}$ operates on $\boldsymbol{S}$. Since $G\boldsymbol{P}_0 = \boldsymbol{P}_0$ for every $G \in \mathscr{G}$, G permutes the edges of $\boldsymbol{P}_0$:

$$(1) \qquad Ga_j = \lambda_j(G) a_{\pi_G(j)} \qquad (j = 1, 2, \cdots, k)$$

where $\lambda_j(G) > 0$, and π_G is some permutation of the set $\{1, 2, \cdots, k\}$. Clearly, $\lambda_j(I) = 1$ for all j, and π_I is the identity.

Once I is given, and $\boldsymbol{N}$, $a_1, \cdots, a_k$ have been defined as above, equation (1) completely describes all possible G. Indeed let $\lambda_1, \cdots, \lambda_k$ be any positive numbers, and π any permutation of $\{1, \cdots, k\}$. Define $G(\lambda_1, \cdots, \lambda_k; \pi)$ to be that matrix G which satisfies

$$(2) \qquad G\boldsymbol{N} = \{0\}, \quad Ga_j = \lambda_j a_{\pi(j)} \qquad (j = 1, 2, \cdots, k).$$

Since the a_j are a basis of $\boldsymbol{S}$, and $\boldsymbol{R}^n = \boldsymbol{N} \oplus \boldsymbol{S}$, a matrix $G(\lambda_1, \cdots, \lambda_k; \pi)$ is uniquely defined by (2). I claim that G *is non-negative*, i.e. that $G\boldsymbol{P} \subset \boldsymbol{P}$. Now $G\boldsymbol{P}_0 = \boldsymbol{P}_0$ is obvious from (2) and the fact that the λ_j are positive. Furthermore, we have $I\boldsymbol{N} = \{0\}$, $Ia_j = a_j$ $(j = 1, \cdots, k)$; so $G = GI$, $G\boldsymbol{P} = GI\boldsymbol{P} = G\boldsymbol{P}_0 = \boldsymbol{P}_0$.

A simple calculation shows that the different $G(\lambda_1, \cdots, \lambda_k; \pi)$ multiply according to the rule

$$(3) \quad G(\lambda_1, \cdots, \lambda_k; \pi) G(\mu_1, \cdots, \mu_k; \rho) = G(\lambda_{\rho(1)}\mu_1, \cdots, \lambda_{\rho(k)}\mu_k; \pi\rho).$$

In an abstract setting, where the λ_j are elements of an arbitrary group $\mathscr{B}$, the group defined by (1) (or (3)) was introduced by Ore [5] who called it *the complete monomial group of degree k over* $\mathscr{B}$. So we have proved

THEOREM 1. *Every maximal group of non-negative matrices of rank k is isomorphic to the complete monomial group of degree k over the reals.*

(Of course, by an elementary theorem of semigroup theory, every subgroup of a semigroup is contained in a unique maximal subgroup.)

In fact, our proof shows somewhat more than just the isomorphism stated in Theorem 1: for any two maximal groups of non-negative $n \times n$ matrices whose ranks are identical, one is the image of the other under some inner automorphism of the semigroup of *all* $n \times n$ matrices. This is not so, however, within the semigroup $\mathscr{H}_n$ of all *non-negative* $n \times n$ matrices. Indeed consider the idempotents

$$I_1 = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{pmatrix} \text{ and } I_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

both of rank two. The equation $I_i X = 0$ has non-trivial solutions in $\mathscr{H}_3$ for $i = 2$ but not for $i = 1$. So there is no automorphism of $\mathscr{H}_3$, inner or not, which maps (the maximal subgroup containing) I_1 onto (the maximal subgroup containing) I_2.

The considerations preceding Theorem 1 give a complete description of all maximal groups of non-negative matrices provided we know all possible systems $(N, a_1, \cdots, a_k)$. Some information about these systems is implicit in our next theorem which describes the non-negative idempotent matrices, generalizing Doob's description of stochastic idempotents ([2], Theorem 2).

THEOREM 2. *Let I be a non-negative idempotent matrix of rank k. There exists a permutation matrix P such that*

$$P^{-1}IP = \begin{pmatrix} J & JB & 0 & 0 \\ 0 & 0 & 0 & 0 \\ AJ & AJB & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \tag{4}$$

$$J = \begin{pmatrix} J_1 & & & 0 \\ & J_2 & & \\ & & \ddots & \\ 0 & & & J_k \end{pmatrix} \tag{5}$$

where the J_j are non-negative idempotent matrices of rank one, that is, matrices $(\alpha_s \beta_t)$, $\alpha_s \geqq 0$, $\beta_s \geqq 0$, $\sum \alpha_s \beta_s = 1$. Conversely, every matrix (4) *where J is given by* (5) *while A and B are arbitrary non-negative matrices of appropriate sizes, is idempotent and of rank k.*

Remember that the change from I to $P^{-1}IP$ means that a certain permutation is simultaneously performed on the rows and columns of I. This operation is an isomorphism for matrix multiplication and thus does not interfere with idempotence.

We begin by proving a particular case of Theorem 2. Assume that *no column of I is zero*. This means that $\boldsymbol{E}_0$ is empty. Assume further that *no row of I is zero either*. It follows that every e_i occurs in the canonical decomposition of some $x \in \boldsymbol{S}$, and so of some a_j; in other words, $\cup \boldsymbol{E}_j = \boldsymbol{E}$. Now let $\boldsymbol{K}_j$ be the linear hull of $\boldsymbol{E}_j$. Then $\boldsymbol{R}^n = \boldsymbol{K}_1 \oplus \cdots \oplus \boldsymbol{K}_k$. Furthermore, $I\boldsymbol{K}_j$ is the line through a_j, for every j. So I is reduced by the $\boldsymbol{K}_j$, and each restriction of

I to some $\boldsymbol{K}_j$ has rank one. By rearranging the coordinates in such a way that the unit vectors belonging to each $\boldsymbol{K}_j$ are grouped together, we find a matrix

$$P^{-1}IP = J = \begin{pmatrix} J_1 & & & \\ & J_2 & & 0 \\ 0 & & \ddots & \\ & & & J_k \end{pmatrix}$$

where the J_j are idempotents of rank one. So we have proved Theorem 2 for the particular case indicated above.

In the general case, we begin by grouping the indices $i = 1, 2, \cdots, n$ in four sets according to whether the i-th row and column of I are both zero, or the i-th row is zero but the i-th column is not, and so on. By simultaneously rearranging rows and columns, we find a matrix

$$P_1^{-1}IP_1 = I_1 = \begin{pmatrix} K & L & 0 & 0 \\ 0 & 0 & 0 & 0 \\ M & N & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

where P_1 is a permutation matrix, and K, L, M are such that K and L have no zero row in common, and K and M have no zero column in common. I_1 is idempotent: $K^2 = K, KL = L, MK = M$, $ML = N$. Since K and $L = KL$ have no zero row in common, K cannot have a zero row. Similarly, no column of K is zero. By the first part of the proof, a further permutation performed simultaneously on the rows and columns of K produces a matrix of the form (5). This completes the proof of Theorem 2 in one direction. The converse is trivial.

Equation (3) shows that the mapping $\pi : G \to \pi_G$ is a homomorphism of $\mathscr{G}$ into the symmetric group $\mathscr{S}_k$. In the case in which $\mathscr{G}$ is bounded, more can be said.

LEMMA 3. *If $\mathscr{G}$ is bounded, π is an isomorphism.*
This is an immediate consequence of

LEMMA 4. *If $\mathscr{G}$ is bounded, and if for some index j and for some $G \in \mathscr{G}$, $\pi_G(j) = j$, then $\lambda_j(G) = 1$.*

Under the assumptions made in Lemma 4, $Ga_j = \lambda_j(G)a_j$ and $G^{-1}a_j = (\lambda_j(G))^{-1}a_j$; so by iteration,

$$G^t a_j = (\lambda_j(G))^t a_j \qquad (t = 0, \pm 1, \pm 2, \cdots).$$

Since $\mathscr{G}$ is bounded, so is the set $\{G^t a_j | t = 0, \pm 1, \pm 2, \cdots\}$. Since $\lambda_j(G) > 0$, this implies $\lambda_j(G) = 1$.

To derive Lemma 3, assume that $G \in \ker \pi$, i.e. π_G is the identity. So by Lemma 4, $Ga_j = a_j$ for every j. Since S is the linear hull of the a_j, it follows that $Gx = x$ for every $x \in S$, i.e. $G = I$.

$\mathscr{G}$ thus is isomorphic to a group of permutations on k elements. Actually, $\mathscr{G}$ itself acts as a group of permutations on a set of suitably chosen multiples $\mu_1 a_1, \cdots, \mu_k a_k$. For suppose the indices i and j are in the same class of transitivity with respect to $\pi(\mathscr{G})$. If G and H are elements of $\mathscr{G}$ satisfying $\pi_G(i) = \pi_H(i) = j$, Lemma 4 shows that $Ga_i = Ha_i$ (= some multiple of a_j). So it is possible to choose $\mu_1, \cdots, \mu_k$ in such a way that

$$G(\mu_j a_j) = \mu_{\pi_G(j)} a_{\pi_G(j)},$$

for all j. (This choice is unique, up to a common factor, if and only if $\pi(\mathscr{G})$ is transitive.) We also see that $\mathscr{G}$ is contained in (at least) one maximal bounded group of non-negative matrices: once multiples $d_j = \mu_j a_j$ have been fixed, consider the set of all linear transformations vanishing on N and permuting the d_j. We have proved

THEOREM 3. *Every bounded group of non-negative matrices is contained in at least one maximal bounded group of non-negative matrices. Every maximal bounded group of non-negative matrices is isomorphic to the symmetric group $\mathscr{S}_k$ where k is the rank of the matrices in the group.*

In particular, every bounded group of non-negative matrices is finite. This is Brown's theorem.

In the special case of stochastic matrices, our Theorem 3 is related to Theorem 8 of [6] which says that every maximal group of stochastic matrices of rank k is isomorphic to $\mathscr{S}_k$. To deduce this fact from our Theorem 3, we have to show that if $\mathscr{G}$ consists of stochastic matrices, then among the maximal bounded groups of non-negative matrices containing $\mathscr{G}$ there is one, at least, which consists entirely of stochastic matrices. (Since the stochastic $n \times n$ matrices form a semigroup, the theorem on semigroups mentioned at an earlier occasion implies that there is at most one.)

Now a non-negative matrix G is stochastic if and only if $Ge = e$ where $e = e_1 + e_2 + \cdots + e_n$. Let $e = x + \sum_{j=1}^{k} \mu_j a_j$, $x \in N$. Since I is stochastic, $e = Ie = 0 + \sum \mu_j a_j$. So $x = 0$. Thus

$$e = \sum e_i = \sum \mu_j a_j = \sum \mu_j (b_j + c_j).$$

If $e_i \in E_l$, say, $\mu_l b_l$ is the only term on the right whose i-th coordinate may be different from zero. Since the i-th coordinate of e is one, we have $\mu_j \neq 0$, for all j. Now for every $G \in \mathscr{G}$,

$$e = Ge = \sum \mu_j \lambda_j (G) a_{\pi_G(j)} = \sum \mu_j a_j.$$

So

$$\mu_j \lambda_j (G) = \mu_{\pi_G(j)}, \quad G(\mu_j a_j) = \mu_{\pi_G(j)} a_{\pi_G(j)}.$$

This shows that $\mathscr{G}$ acts as a group of permutations on the set $\{\mu_j a_j | j = 1, 2, \cdots, k\}$, and may be extended to the group of all linear transformations vanishing on N and permuting the $\mu_j a_j$. In the proof of Theorem 1 it was shown that transformations of this kind are necessarily non-negative. Since they leave invariant $e = \sum \mu_j a_j$, they are stochastic. This proves Schwarz's theorem.

In concluding, I am glad to acknowledge a suggestion from the referee which resulted in simplifying the proofs of Lemma 2, Theorem 2, and Schwarz's theorem.

REFERENCES

D. R. Brown

[1] On clans of non-negative matrices. Proc. Amer. Math. Soc. 15 (1964) 671—674.

J. L. Doob

[2] Topics in the theory of Markoff chains. Transactions Amer. Math. Soc. 52 (1942) 37—64.

A. J. Goldman and A. W. Tucker

[3] Polyhedral convex cones. In: Kuhn-Tucker (ed.), Linear inequalities and related systems. Annals of Mathematics Studies 38, Princeton University Press, Princeton, N.J. (1956) pp. 19—40.

H. Minkowski

[4] Geometrie der Zahlen. Teubner, Leipzig und Berlin 1910.

O. Ore

[5] Theory of monomial groups. Transactions Amer. Math. Soc. 51 (1942) 15—64.

Št. Schwarz

[6] On the structure of the semigroup of stochastic matrices. Publ. Math. Inst. Hungar. Acad. Sci., Ser. A, 9 (1964) 297—311.

(Oblatum 16–10–68)