

ON THE REPRESENTATION OF DOUBLY STOCHASTIC OPERATORS

JOHN V. RYFF

1. Introduction. A real n -square matrix is called *doubly stochastic* if it has nonnegative elements and each row and each column sums to 1. For any two n -vectors $x = (x_1, \dots, x_n)$, $y = (y_1, \dots, y_n)$, let $x^* = (x_1^*, \dots, x_n^*)$ and $y^* = (y_1^*, \dots, y_n^*)$ be the vectors obtained from x and y by rearranging their respective components in nonincreasing order. Then for T doubly stochastic and $y = Tx$ it is not hard to show that

$$\sum_{i=1}^k y_i^* \leq \sum_{i=1}^k x_i^* \quad 1 \leq k < n, \quad (1)$$

$$\sum_{i=1}^n y_i = \sum_{i=1}^n x_i.$$

The converse is also true. If the inequalities (1) are valid for every pair of vectors related by the equation $y = Tx$, then T must be doubly stochastic. Employing a continuous version of (1), one can define on a space of integrable functions a class of linear operators which inherits many of the familiar properties of doubly stochastic matrices. A concrete representation of these operators will be obtained and a comparison made with the doubly stochastic operators defined by Rota [6].

In the following all functions and variables will be real. Measure will be understood to mean Lebesgue measure and will be denoted by μ . Sets will always be measurable and the term "almost everywhere" will usually be suppressed. $L^1 = L^1(0, 1)$ and $L^\infty = L^\infty(0, 1)$ are to have the conventional meaning as the spaces of integrable and essentially bounded "functions" on $[0, 1]$, with $[L^1]$ and $[L^\infty]$ to represent the operator spaces of L^1 and L^∞ . Convergence is to be taken in the pointwise sense unless the contrary is indicated.

2. Rearrangements. If f is a measurable function on $[0, 1]$ consider

$$m(y) = \mu\{x : f(x) > y\}.$$

This function is nonincreasing, right-continuous and defined for all values of y . As such, the function m admits an inverse which will

Received November 26, 1962. A portion of this work was supported by the Army Research Office (Durham). Presented to the American Mathematical Society on January 24, 1963.

be denoted by f^* . To fix ideas and remove certain ambiguities of definition it is convenient to define

$$f^*(x) = \sup_{m(y) > x} y$$

for $0 < x < 1$. It is not difficult to show from this definition that f^* is also nonincreasing and right-continuous. The function f^* is called (somewhat inaccurately) the *decreasing rearrangement* of f . The definition given here represents a departure from the customary definition of f^* as the decreasing rearrangement of $|f|$ (see [1], Chapter X). If $\{f_n\}$ is a convergent sequence of measurable functions on $[0, 1]$ with limit f , then $f^* = \lim_n f_n^*$. Thus, it is possible to give a more suggestive definition of f^* by first choosing a sequence $\{\sigma_n\}$ of simple functions converging to f . Then the corresponding sequence of rearrangements $\{\sigma_n^*\}$ will form a sequence of decreasing step functions converging to f^* . Moreover, if the simple functions converge monotonically so too will the step functions (for $g \leq f$ implies $g^* \leq f^*$). The functions f and f^* are equally integrable (or non-integrable) and their integrals are related by

$$\int_0^s f \leq \int_0^s f^* \quad 0 \leq s < 1,$$

$$\int_0^1 f = \int_0^1 f^*.$$

Furthermore, if f and g are integrable, the inequality

$$\int_0^s (f + g)^* \leq \int_0^s f^* + \int_0^s g^* \quad 0 \leq s \leq 1$$

is easy to verify.

3. \mathcal{E} -operators. A partial order $<$ can be introduced in L^1 by defining $g < f$ whenever

$$\int_0^s g^* \leq \int_0^s f^* \quad 0 \leq s < 1,$$

and

$$\int_0^1 g = \int_0^1 f$$

(It is instructive to compare this with (1)). The class of linear transformations $T: L^1 \rightarrow L^1$ which satisfy $Tf < f$ for each integrable function f will be denoted by \mathcal{E} . Such transformations will be referred to as \mathcal{E} -operators. As an illustration, the transformation

which carries each $f \in L^1$ onto the constant $\int_0^1 f$ is an \mathfrak{S} -operator. Other examples arise from measure preserving transformation $s \rightarrow \pi(s)$ of $[0, 1]$ by defining $(Tf)(s) = f(\pi(s))$. Their behavior is plainly reminiscent of that of permutation matrices. On the other hand, the mapping $f \rightarrow f^*$ fails to belong to \mathfrak{S} since it is not linear.

The use of the word *operator* for transformations in \mathfrak{S} is intended to imply that they are bounded. Indeed, any \mathfrak{S} -operator is simultaneously a contraction operator in both $[L^1]$ and $[L^\infty]$. To see this, suppose first that $f \geq 0$ and that on a set E_0 of positive measure $Tf < 0$. Then, for $1 - \mu(E_0) < s < 1$, $(Tf)^*(s) < 0$. Hence,

$$\int_{1-\mu(E_0)}^1 (Tf)^* < 0 \leq \int_{1-\mu(E_0)}^1 f^*$$

so that $Tf < f$ is not possible. Consequently, if T is in \mathfrak{S} then $T \geq 0$ (i.e., $f \geq 0$ implies $Tf \geq 0$). If f is arbitrary write $f = f^+ - f^-$, where f^+ and f^- are the positive and negative parts of f . Then from $Tf = Tf^+ - Tf^-$ and the preceding consideration,

$$\begin{aligned} (Tf)^+ &\leq Tf^+ \\ (Tf)^- &\leq Tf^- \end{aligned}$$

and so

$$\int_0^1 |Tf| \leq \int_0^1 T|f| = \int_0^1 |f|.$$

That T restricted to L^∞ is a contraction in $[L^\infty]$ follows by a similar argument.

A representation for operators in $[L^1]$ has been established [3] whereby each such operator T is given by

$$(2) \quad (Tf)(s) = \frac{d}{ds} \int_0^1 K(s, t)f(t)dt$$

where K is measurable on the square $0 \leq s, t \leq 1$ and satisfies:

- (a) $K(0, t) = 0 \quad 0 \leq t \leq 1;$
- (b) $\text{essup}_t V[K(\cdot, t)] = C < \infty;$
- (c) $\int_0^1 K(\cdot, t)f(t)dt$ is absolutely continuous for every $f \in L^1$.

By $V[K(\cdot, t)]$ is meant the total variation of K for t fixed. The kernel K may be chosen in such a way that the constant C in (b) equals $\|T\|$. It will be assumed that this has been done.

While the derivative of the integral in (2) may appear awkward, it is ordinarily not possible to replace this representation by an

integral transformation (necessary and sufficient conditions are given in [3, p. 146]). In particular, the identity transformation (an \mathfrak{S} -operator, incidentally) cannot be represented in this way. On the other hand, if one takes $K(s, \cdot)$ to be the characteristic function of the interval $[0, s]$, it is clear that (2) gives the identity. More generally, if $s \rightarrow \pi(s)$ is measure preserving, set $E_s = \pi^{-1}([0, s])$ and define $K(s, \cdot)$ to be the characteristic function of E_s . Then for $0 < s < 1$ and $h > 0$ sufficiently small,

$$\frac{1}{h} \int_0^1 [K(s+h, t) - K(s, t)] f(t) dt = \frac{1}{h} \int_s^{s+h} f(\pi(t)) dt$$

or

$$f(\pi(s)) = \frac{d}{ds} \int_0^1 K(s, t) f(t) dt.$$

The kernel $K(s, t) = s$ in (2) gives the operator which carries f onto $\int_0^1 f$ and is the only \mathfrak{S} -operator with degenerate kernel $K(s, t) = \phi(s)\psi(t)$.

The kernels in (2) corresponding to operators in \mathfrak{S} also share the following properties:

- (d) $s = \int_0^1 K(s, t) dt$;
- (e) $s_1 < s_2$ implies $K(s_1, \cdot) \leq K(s_2, \cdot)$;
- (f) $K(1, t) = 1$ $0 \leq t \leq 1$.

If 1 is used to denote the function which is identically equal to 1, it is readily seen that $T1 = 1$ whenever T is in \mathfrak{S} , which gives (d). For any measurable set E

$$\frac{d}{ds} \int_E K(s, t) dt \geq 0$$

establishing (e), while

$$\mu(E) = \int_E K(1, t) dt$$

implies (f). The importance of these three latter properties will be evident when the representation theorem for \mathfrak{S} -operators is given. However, the essential characterization of these operators is contained in the following result:

LEMMA. Let $T \in [L^\infty]$ satisfy

$$0 \leq T\chi_E \leq 1$$

$$\int_0^1 T\chi_E = \mu(E)$$

for each characteristic function χ . Then T is bounded in the L^1 -norm and admits a unique extension $\bar{T} \in [L^1]$. This extension will be an \mathfrak{S} -operator.

Proof. Let $\sigma = \sum a_i \chi_{E_i}$ be a simple function and suppose for the present that $0 < a_1 < a_2 < \dots < a_n$. Write

$$\begin{aligned} b_1 &= a_1 & F_1 &= \bigcup_{i=1}^n E_i \\ b_2 &= a_2 - a_1 & F_2 &= \bigcup_{i=2}^n E_i \\ &\dots\dots\dots & & \\ b_n &= a_n - a_{n-1} & F_n &= E_1. \end{aligned}$$

Then $\sigma = \sum_{i=1}^n b_i \chi_{F_i}$ ($b_i > 0$) and

$$\sigma^* = (\sum b_i \chi_{F_i})^* = \sum b_i \chi_{F_i}^{*1}$$

From the hypothesis then

$$\begin{aligned} 0 \leq \int_0^s (T\sigma)^* &= \int_0^s (\sum b_i T\chi_{F_i})^* \leq \int_0^s \sum b_i (T\chi_{F_i})^* \\ &\leq \int_0^s \sum b_i (\chi_{F_i})^* = \int_0^s \sigma^*. \end{aligned}$$

A small modification of this argument covers the case where some or all of the coefficients of σ are negative. Thus, $T\sigma < \sigma$ for all simple functions σ . Furthermore, in the L^1 -norm, T acts on the simple functions as a contraction and so, extends uniquely to an operator $\bar{T} \in [L^1]$.

Suppose now that $f \in L^1$ and $f \geq 0$. Select an increasing sequence $\{\sigma_n\}$ of nonnegative simple functions converging to f . It follows that $\{T\sigma_n\}$ will also be increasing and nonnegative. Therefore, $T\sigma_n \rightarrow \bar{T}f$ both pointwise and in the L^1 -sense. Accordingly, $\bar{T}f < f$. This reasoning may now be extended to arbitrary f by splitting the function into its positive and negative parts, and completes the proof.

REMARK. The pointwise convergence of the sequence $\{T\sigma_n\}$ was required to guarantee the convergence of $\{(T\sigma_n)^*\}$ to $(Tf)^*$.

Representation theorem. A linear transformation $T: L^1 \rightarrow L^1$ is an \mathfrak{S} -operator if and only if T admits a representation (2) where the kernel K is measurable and satisfies conditions (a)-(f).

Proof. Properties (a)-(c) are necessary and sufficient for a trans-

¹ This device appears to have been first employed by F. Riesz ([5, p. 164]).

formation (2) to belong to $[L^1]$, so that it is (d), (e) and (f) which single out the class \mathfrak{S} . From (e) one infers that $T \geq 0$, while (f) gives $\int_0^1 Tf = \int_0^1 f$ for each $f \in L^1$. With $h > 0$, $0 < s < 1$ and χ an arbitrary characteristic function, (d), (e) and (f) combine to give

$$0 \leq \frac{1}{h} \int_0^1 [K(s+h, t) - K(s, t)]\chi(t)dt \leq \frac{1}{h} \int_0^1 [K(s+h, t) - K(s, t)]dt = 1.$$

That is, $0 \leq T\chi \leq 1$, hence the lemma applies.

It is natural to inquire whether the class \mathfrak{S} is at all comparable with the doubly stochastic operators introduced by Rota [6]. A linear transformation P is said to be *doubly stochastic* if

- (i) $P \geq 0$ and is simultaneously a contraction in $[L^1]$ and $[L^\infty]$;
- (ii) $P1 = 1$ (recall that 1 represents the function which is identically equal to 1);
- (iii) $P^* \in [L^1]$.

Condition (iii) is equivalent to the requirement that to each $f \in L^1$ there corresponds a function $P^*f \in L^1$ such that

$$\int_0^1 fPg = \int_0^1 gP^*f$$

whenever $g \in L^\infty$. (It should be remarked that this definition is equally applicable when $[0, 1]$ is replaced by an arbitrary (positive) measure space (S, Σ, ν) .) From the lemma and (i) and (ii) one sees that each doubly stochastic operator is also in \mathfrak{S} . On the other hand, it has already been shown that each \mathfrak{S} -operator satisfies (i) and (ii). The behavior of the adjoint T^* of an \mathfrak{S} -operator T requires some further study. It is a fact, however, that the adjoint of any \mathfrak{S} -operator (restricted to L^1) is again in \mathfrak{S} (this is not surprising, the transpose of a doubly stochastic matrix is again doubly stochastic). To see this, let χ_s be the characteristic function of the interval $[0, s]$ and take any $g \in L^\infty$. Then, using (2)

$$\begin{aligned} \int_0^1 \chi_s Tg &= \int_0^s \left(\frac{d}{d\sigma} \int_0^1 K(\sigma, t)g(t)dt \right) d\sigma \\ &= \int_0^1 K(s, t)g(t)dt = \int_0^1 g T^*\chi_s. \end{aligned}$$

Evidently, $T^*\chi_s = K(s, \cdot)$. If χ_G is the characteristic function of an open subset $G = \bigcup_i (a_i, b_i)$ of $[0, 1]$ then

$$T^*\chi_G = \sum_i [K(b_i, \cdot) - K(a_i, \cdot)],$$

which converges to an essentially bounded function (property (b)). Since the constant in (b) can be taken to be 1, for each open subset

\mathcal{G} one has

$$0 \leq T^* \chi_G \leq 1$$

$$\int_0^1 T^* \chi_G = \sum_i (b_i - a_i) = \mu(G).$$

Limiting arguments show now that T^* satisfies the hypothesis of the lemma and proves the assertion. In particular then, (iii) is satisfied by every \mathcal{S} -operator. Therefore, *the doubly stochastic operators coincide with the class \mathcal{S} .*

4. Concerning the extreme points.² The convexity of \mathcal{S} leads naturally to questions of extreme points and convex closure. Consider the space $[L^\infty]$ endowed with the weak* operator topology. A subbasic neighborhood of the null transformation in this topology consists of all $T \in [L^\infty]$ for which $\left| \int_0^1 f Tg \right| < \epsilon$, where $g \in L^\infty$ and $f \in L^1$. \mathcal{S} is closed in this topology. Hence, by a direct application of the general compactness theorem of Kadison [2], \mathcal{S} is also compact.

It is not really necessary to appeal to the Krein-Milman theorem for the existence of extreme points. For instance, those operators induced by measure preserving transformations, as described earlier, are extreme points of \mathcal{S} . Indeed, any \mathcal{S} -operator which carries characteristic functions onto characteristic functions is an extreme point. These operators may also be characterized as the elements of \mathcal{S} which are *multiplicative* on L^∞ ($Tfg = TfTg$). Proofs of the preceding remarks are straightforward.

Now consider the operator T defined by $(Tf)(s) = 1/2f(s/2) + 1/2f((s+1)/2)$. This operator is the adjoint (restricted to L^1) of the \mathcal{S} -operator S induced by the measure preserving transformation $s \rightarrow \pi(s) = 2s \pmod{1}$. It follows that T is also in \mathcal{S} . Suppose that $T = 1/2T_1 + 1/2T_2$ with T_1 and T_2 in \mathcal{S} . The corresponding adjoints then satisfy $S = 1/2T_1^* + 1/2T_2^*$. Since S is an extreme point, $T_1^* = T_2^*$. Hence, $T_1 = T_2$ (as norms are preserved) so that T is also an extreme point. However, T is *not multiplicative*. A complete characterization of the extreme points of \mathcal{S} does not seem to be known.

A second problem is suggested by a corollary to Muirhead's theorem [1, p. 49]. It can be shown that if x and y are two n -vectors whose components are related by the inequalities (1), then there is a doubly stochastic matrix which carries x onto y . Now suppose that f and g are in L^1 and $g < f$. Does it follow that there exists an \mathcal{S} -operator T such that $g = Tf$?

² The author is indebted to Professor R. R. Phelps for several enlightening conversations and for access to his forthcoming paper [4].

REFERENCES

1. G. H. Hardy, J. E. Littlewood and G. E. Polya, *Inequalities*, Cambridge, 1952.
2. R. V. Kadison, *The trace in finite operator algebras*, Proc. Amer. Math. Soc., **2** (1961), 973-977.
3. L. Kantorovitch and B. Vulich, *Sur la représentation des opérations linéaires*, Compositio Math., **5** (1937), 119-165.
4. R. R. Phelps, *Extreme positive operators and homomorphisms*, Trans. Amer. Math. Soc., (to appear).
5. F. Riesz, *Sur une inégalité intégrale*, J. London Math. Soc., **5** (1930), 162-168.
6. G. C. Rota, *An "Alternierende Verfahren" for general positive operators*, Bull. Amer. Math. Soc., **68** (1962), 95-102.