AN "ALTERNIERENDE VERFAHREN" FOR GENERAL POSITIVE OPERATORS[1]

BY GIAN-CARLO ROTA[2]

Communicated by Felix Browder, October 25, 1961

1. **Introduction.** Two types of limiting processes involving linear operators are of frequent occurrence. One type is associated with ergodic limits of Cesaro averages $n^{-1}(1+T+ \cdots +T^{n-1})f$ of a linear operator T, for f in suitable function spaces, the so-called "ergodic theorems." The strongest such results establish almost everywhere convergence for f in a suitable Lebesgue space, under suitable hypotheses on T. This type of theorem is, nowadays, fairly well understood.[3]

A second type of limiting theorem of frequent occurrence concerns the limit of *products* $T_1T_2 \cdots T_nf$ of a sequence of operators T_n. The two noteworthy special cases are (a) Limiting theorems of the type $\lim_{t\to\infty} P^tf$, where P^t is a semigroup of selfadjoint operators. Results of this type express the "tendency to equilibrium" in certain physical processes (typically, in diffusion theory). (b) One considers two or more noncommuting projections (most commonly *conditional expectations* in a probability space) say F_1 and F_2, and asks for $\lim_{n\to\infty} (F_1F_2)^nf$. This is an abstract rendering of the so-called *"alternierende Verfahren,"* (see [**8**]) which originated in function theory and has recently merged with various probabilistic and other considerations.

While results relating to the *mean convergence* of such iterations are easily obtained (for (a) and (b) they were obtained independently by von Neumann and Wiener [**10**; **15**] in the 1930's), the corresponding statements relating to *almost everywhere* convergence have been missing. A result in this direction has recently been obtained by Burkholder and Chow [**1**], but their results, though more general in some respects, are limited to L_2.[4]

[1] I wish to thank Professors J. Feldman, P. R. Halmos and C. C. Moore for listening to expositions of the contents of this paper, and offering much helpful advice.

[2] Work supported in part by Grant No. NSF G-19684 with the National Science Foundation and in part by the Office of Naval Research. The contents of this paper were delivered in an invited address at the Symposium in Ergodic Theory held at Tulane University in November 1961.

[3] For the most refined results see [**2**; **7**].

[4] Another important set of results has been recently obtained by E. M. Stein (personal communication, whose methods are of an altogether different character from the present ones; to him are due the statement and a first proof Theorem 2b below.

It is our present purpose to establish a single general procedure which yields *almost everywhere* convergence in L_p of certain products of operators. The main result stated below (Theorem 1) is general enough to include the case of $\lim_{n\to\infty} P^n f$, as well as the limit of iteration of several noncommuting conditional expectations.

2. **Doubly stochastic operators.** Given a *positive measure space* (S, Σ, μ), we consider a linear operator P whose domain and range are certain subspaces of the space $TM(S, \Sigma, \mu)$ of all real-valued totally measurable functions on (S, Σ, μ), as specified below. We make the following assumption on the operator P.

(1) P is defined on all of $L_1(S, \Sigma, \mu)$ and on all of $L_\infty(S, \Sigma, \mu)$, maps each of those spaces into itself. It is a *positive*[5] operator (that is, $Pf \geqq 0$ if $f \geqq 0$), and it is a contraction (that is, is of norm at most one) in each of these spaces. In other words,

$$\sup_{s\in S} |Pg(s)| \leqq \sup_{s\in S} |g(s)|,$$

and

$$\int_S |Pf(s)| \mu(ds) \leqq \int_S |f(s)| \mu(ds)$$

for bounded and integrable g and f, respectively. (It follows by the Riesz convexity theorem that P will be a contraction in $L_p(S, \Sigma, \mu)$, $1 \leqq p \leqq \infty$.)

(2) The adjoint operator P^*, bounded by the identity $\int_S fPg d\mu = \int_S P^* f g d\mu$, is also defined in $L_1(S, \Sigma, \mu)$ and $L_\infty(S, \Sigma, \mu)$.

It follows that P^* is also a contraction in each of these spaces.

Evidently the various conditions specified in (1) and (2) are far from independent from one another, and could in fact be reduced in many ways to smaller sets of conditions. The redundance of these specifications is more than made up for by the ease with which they can be verified.

(3) Let 1 denote the function defined on S and with real values, which takes identically the value 1. We assume that $P1 = 1$ and that $P^*1 = 1$.

It follows, by an easy argument, which we omit, that in a finite measure space $P1 = 1$ implies $P^*1 = 1$ for an operator satisfying (1) and (2).

We shall call an operator satisfying conditions (1), (2), (3) a *doubly stochastic operator.*

---

[5] *Positive* operators should not be confused with *positive-definite* operators in Hilbert space, which are also sometimes called positive operators.

Consider now a sequence P_n ($n=1, 2, 3, \cdots$) of doubly stochastic operators in a probability space (S, Σ, μ). The notion of the *path space* associated with this sequence[6] is an easy generalization of the classical probabilistic construction of the sample space of a Markov process (which in turn is a special case of the present notion), as found for example in Doob [4] or Loève [9]. It is defined as follows.

Consider the *measurable space* (S', Σ'') consisting of the one-sided infinite product of replicas of the measurable space (S, Σ); that is, $S' = \prod_{n=0}^{\infty} S_n$, and $S_n = S$ for all n. The σ-field Σ'' is the ordinary product σ-field.

A particular class of measurable real-valued functions on (S', Σ'') consists of functions $F(s_0, s_1, s_2 \cdots) = f_0(s_0)f_1(s_1)f_2(s_2) \cdots$, where each of the functions f_n is a measurable real-valued function defined on S, and where $f_n = 1$ for almost all n (that is, all integers n except for a finite number). We shall denote by $\boldsymbol{A}$ the algebra generated by these functions, when all the f_n are assumed to be essentially bounded functions.

The algebra $\boldsymbol{A}$ inherits a natural (and obvious) *lattice structure* from $L_\infty(S, \Sigma, \mu)$, which makes it into a lattice-ordered algebra.

With the aid of the given sequence P_n of doubly stochastic operators, we can define an L-space structure[7] on $\boldsymbol{A}$, considered as a vector space. A positive linear functional L on functions of the form $F(s_0, s_1, s_2, \cdots) = f_0(s_0)f_1(s_1)f_2(s_2) \cdots$ belonging to the algebra $\boldsymbol{A}$ is defined by the formula

$$(*) \qquad L(F) = \int_S f_0 P_1[f_1 P_2[f_2 P_3[f_3 \cdots]]] \cdots]d\mu,$$

and then extended to all of $\boldsymbol{A}$ by linearity. The linear functional L is *well-defined* on $\boldsymbol{A}$. First, note that the integral in (*) involves only a *finite* number of terms, since $f_n = 1$ for almost all n and $P_n 1 = 1$ for all n; hence the integral makes sense. Secondly, the extension to $\boldsymbol{A}$ is well-defined, because $\boldsymbol{A}$ is essentially a tensor product of algebras isometric to $L_\infty(S, \Sigma, \mu)$, and L in (*) is compatible with the distributive laws defining the tensor product. Thirdly, from the definition of the lattice structure of $\boldsymbol{A}$ it is clear that $|F| = |f_0| \cdot |f_1| \cdot |f_2| \cdot \cdots$,

[6] The notion of a path space, and much of the theory to follow, can be studied for an arbitrary measure space, which need not even be σ-finite. In the present account we have refrained from such generalizations, for the sake of clarity of exposition. In many cases the reader himself may easily supply the necessary steps to achieve the needed generalization.

[7] In the sense of Kakutani, cf. Day [3, p. 107 ff.]. Day calls these (AL)-spaces.

where $|f|$ denotes the function equal to $|f(s)|$ at the point s. Furthermore, the formula

$$\|F\| = P(|F|) = \int_S |f_0| P_1[|f_1| P_2[|f_2| P_3[|f_3| \cdots]] \cdots] d\mu$$

defines a seminorm, and this seminorm is of "L-type" in the sense of Kakutani, (in other words, $\|F+G\| = \|F\| + \|G\|$ whenever both F and G are positive elements of $\boldsymbol{A}$). Factoring off the subspace of elements F such that $\|F\| = 0$, and completing, we obtain an abstract L-space X. By Kakutani's theorem,[8] the space X is isometric to (and will thereafter be identified with) $X = L_1(S'', \Sigma', \nu)$ for some (structurally unique) measure space (S'', Σ', ν).

The measure space (S'', Σ', ν) is the *path space* of the sequence of operators P_n. The definition of the path space does not use the full strength of the assumption that the P_n are doubly stochastic operators. All that is needed, for the definition to make sense, is that $P_n 1 = 1$ for all n, and that P_n is a positive operator.

It is clear that we can identify the set S'' with S', and the σ-field Σ with a σ-subfield of Σ''. We shall regard this identification (whose purpose is purely heuristic) as made, and "picture" the space (S', Σ', ν) as a space of sequences of points of S, with a certain measure. From $L(1) = 1$ it follows that (S', Σ', ν) is a probability space.

We next note that there is a natural imbedding of (S, Σ, ν) obtained by identifying $f \in L_1(S, \Sigma, \mu)$ with the function $F(s_0, s_1, s_2, \cdots) = f(s_0)$ of $\boldsymbol{A}$; the definition of L shows that this identification is faithful. Accordingly, we shall consider this imbedding as accomplished, and set $S = S'$, and $\Sigma \subset \Sigma'$, writing sometimes the path space as (S, Σ', ν). We shall also set $d\mu = d\nu$, when convenient. Thus, the path space is obtained essentially by *refining* the given σ-field Σ to a large enough σ-field.

3. **Main theorem.** Our main result is the following:

THEOREM 1. *Let P_n $(n = 1, 2, 3, \cdots)$ be a sequence of doubly stochastic operators in a probability space (S, Σ, μ). Then for every $p > 1$ and for every $f \in L_p(S, \Sigma, \mu)$, the sequence $P_1 P_2 \cdots P_n P_n^* P_{n-1}^* \cdots P_1^* f$ converges almost everywhere as $n \to \infty$.*

PROOF. We consider the subspace L_n of $L_1(S', \Sigma', \nu)$ generated by all functions $F(s_0, s_1, s_2, \cdots, s_n, \cdots) = f_0(s_0) f_1(s_1) \cdots f_n(s_n) \cdots$, where all f_n are essentially bounded, and where $f_0 = f_1 = \cdots = f_{n-1}$

[8] Cf. Day, loc. cit.

$=1$. Evidently the L_1-closure of L_n is also an L-space; it is therefore representable in the form $L_1(S', \Sigma_n, \nu)$, where Σ_n is a σ-subfield of Σ', uniquely determined by L_n. If $n > m$, then $\Sigma_n \subset \Sigma_m$. Let E_n be the *conditional expectation*[9] operator of $L_1(S', \Sigma', \nu)$ onto $L_1(S', \Sigma_n, \nu)$, and let E be the conditional expectation projecting onto $L_1(S, \Sigma, \mu)$. If f is any function of class $L_p(S', \Sigma', \nu)$, then the sequence of functions $f_n = E_n f$ forms a martingale reversed sequence, (cf. Loève loc. cit. p. 306 ff.) which is uniformly integrable. By the martingale convergence theorem, we infer that $f_\infty = \lim E_n f$ exists *almost everywhere* for $p > 1$. Furthermore, by the maximal ergodic theorem for martingales there is a function f^* in $L_p(S', \Sigma', \nu)$ such that $\sup_{n \geq 0} |f_n(s)| \leq f^*(s)$. In other words, the sequence f_n converges boundedly in L_p to f_∞. Since conditional expectation preserves bounded almost everywhere convergence, we conclude that the sequence $EE_n f = Ef_n$ converges almost everywhere to Ef_∞.

We have therefore proved that *for every function f in $L_p(S, \Sigma, \mu)$, $p > 1$, the sequence $EE_n f$ converges almost everywhere to some function g in $L_p(S, \Sigma, \mu)$.*

The proof of the theorem will be complete, if we can show that $EE_nE = P_1 \cdots P_{n-1}P_nP_n^*P_{n-1}^* \cdots P_1^*$. This can actually be done by a simple direct computation.

To calculate $E_n f$, for f in $L_p(S, \Sigma, \mu)$, we make use of the fact that $E_n f$ is the *unique* function such that

$$(**) \qquad \int_{S'} gE_n f \, d\nu = \int_{S'} gf \, d\nu, \qquad g \in L_q(S', \Sigma_n, \nu),$$

where q is the index conjugate to p.

To verify that (**) holds, it suffices to show that it holds when g ranges over a dense subset of $L_q(S', \Sigma_n, \nu)$. Such a dense subset will be taken to be the subspace L_n defined above. We therefore need only verify (**) when $g = F$, where F is in L_n, and F is of the "product" form

$$F(s_0, s_1, s_2, \cdots) = f_0(s_0)f_1(s_1)f_2(s_2) \cdots .$$

We shall in fact find that $E_n f = G$, where

$$G(s_0, s_1, s_2, \cdots) = P_n^*P_{n-1}^* \cdots P_1^* f(s_n) = h(s_n).$$

Now,

---

[9] The reader unfamiliar with this notion may consult Loève [9, Chapter VII], or, for an "operatorial" treatment, [11]. The particular properties that single out conditional expectations among all projections are essential in what follows.

$$\int_{S'} FGd\nu = \int_S f_0P_1[f_1P_2[\cdots P_n[(hf_{n+1})P_{n+1}[f_{n+2}\cdots]]\cdot]d\mu.$$

Since $f_0=f_1=\cdots=f_{n-1}=1$, the right side simplifies to

$$\int_S P_1P_2\cdots P_n[hf_nP_{n+1}[\cdots]]d\mu.$$

Using the assumption that $P_k^*1=1$ for all k the last expression in turn simplifies to

$$\int_{S'} FGd\nu = \int_S hf_nP_{n+1}[f_{n+2}P_{n+2}[\cdots]]d\mu$$

$$= \int_S (P_n^*P_{n-1}^*\cdots P_1^*f)(f_nP_{n+1}[f_{n+1}P_{n+2}[\cdots]])d\mu.$$

Passing from the operator P_k^* to their adjoints P_k, this last integral becomes identical with

$$\int_S f(P_1P_2\cdots P_n[f_nP_{n+1}[f_{n+1}P_{n+2}\cdots]])\,d\mu = \int_{S'} fF\,d\nu.$$

We have therefore shown that

$$\int_{S'} GF\,d\nu = \int_{S'} fF\,d\nu$$

for all F of the type indicated. But, as mentioned above, this amounts to saying that $G=E_nf$.

Next, we shall compute the conditional expectation EH, where $H\in L_p(S', \Sigma_n, \nu)$ is of the form $H(s_0, s_1, s_2, \cdots)=h(s_n)$ for some $h\in L_p(S, \Sigma, \mu)$. We shall in fact find that $EH=P_1P_2\cdots P_nh=g$. Again, this amounts to verifying that

$$\int_S gf\,d\mu = \int_{S'} fH\,d\nu$$

for all $f\in L_p(S, \Sigma, \mu)$. Now, by definition of $d\nu$, we have

$$\int_{S'} fHd\nu = \int_S fP_1P_2\cdots P_nh\,d\mu = \int_S gf\,d\mu,$$

which is exactly what we want.

Combining the results of the two computations obtained in the two above paragraphs, we obtain for $f \in L_p(S, \Sigma, \mu)$, that $EE_nf = P_1P_2 \cdots P_nP_n^*P_{n-1}^* \cdots P_2^*P_1^*f$, and the proof of convergence is therefore complete.

4. **Applications.** The two main applications of the above theorem are (a) the "alternierende Verfahren" for sequences of noncommuting conditional expectations, and (b) the "tendency to equilibrium" theorems for *powers* of a positive doubly stochastic operator. We begin with the second result.

THEOREM 2. *Let P be a doubly stochastic operator which is selfadjoint in $L_2(S, \Sigma, \mu)$. Then*:

(a) *There is a dilation of the sequence of operators P^{2n} into a martingale E_n.*

(b) *For f in $L_p(S, \Sigma, \mu)$, $p > 1$, $\lim_{n\to\infty} P^{2n}f$ exists almost everywhere.*

PROOF. It suffices to choose all the $P_n = P$ in Theorem 1.

The "alternierende Verfahren" is actually an application of Theorem 2 when only two conditional expectations, say F_1 and F_2 are involved. Setting $P = F_1F_2F_1$ one obtains a doubly stochastic selfadjoint operator, and hence the convergence of $(F_1F_2F_1)^nf$ follows under the same conditions as in Theorem 2.

More generally, an infinite sequence F_n ($n = 1, 2, \cdots$) of noncommuting conditional expectation operators is a special case of a sequence of operators satisfying the conditions of the Main Theorem; we therefore obtain that $\lim_{n\to\infty} F_1F_2 \cdots F_{n-1}F_nF_{n-1} \cdots F_1f$ exists almost everywhere. Actually, in this result the F_n can be arbitrary doubly stochastic selfadjoint operators.

If only a *finite* sequence $F_1, F_2, \cdots, F_m$ of noncommuting conditional expectations (or again, more generally, arbitrary doubly stochastic selfadjoint operators) are involved, we may obtain from the Main Theorem various convergent sequences of operators yielding almost everywhere convergence. For example, we may take $(F_1 \cdots F_m)^n(F_mF_{m-1} \cdots F_1)^n$, or else $(F_1F_2F_1F_3F_2 \cdots F_m)^n \cdot (F_m \cdots F_2F_3F_1F_2F_1)^n$, etc., etc.

5. **Generalizations and extensions.** The method outlined above is capable of numerous generalizations, among which we mention: a "weak type L_1" theorem, a convergence theorem with reversed indices (that is, for $\lim_{n\to\infty} P_nP_{n-1} \cdots P_1P_1^* \cdots P_n^*f$), more general semigroups, (for example the continuous P^t, t real). Such developments will be given in a forthcoming publication, which will exploit a gen-

eral theory of "dilation" in measure spaces by conditional expectation.[10]

BIBLIOGRAPHY

1. Burkholder and Chow, *Iterates of conditional expectation operators*, Proc. Amer. Math. Soc. **12** (1961), 490–495.

2. Chacon and Ornstein, *A general ergodic theorem*, Illinois J. Math. **4** (1960), 153–160.

3. Day, *Normed linear spaces*, Springer, Berlin, 1958.

4. Doob, *Stochastic processes*, Wiley, New York, 1953.

5. Dunford-Schwarz, *Linear operators*, Vol. I, Interscience, New York, 1958.

6. Halmos, *Lectures on ergodic theory*, Mathematical Society of Japan, Tokyo, 1956.

7. E. Hopf, *On the ergodic theorem for positive linear operators*, Crelle's J. **205** (1960), 101–106.

8. Hurwitz-Courant, *Funktionen Theorie*, Springer, Berlin, 1929.

9. Loève, *Probability theory*, 2nd ed. Van Nostrand, Princeton, 1960.

10. Von Neumann, *Functional operators*, 2 vols., Princeton Univ. Press, Princeton, N. J., 1950.

11. Rota, *On the representation of averaging operators*, Rend. Sem. Mat. Univ. Padova **30** (1960), 52–64.

12. Wiener, *On the factorization of matrices*, Comment. Math. Helv. **29** (1955), 97–110.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

---

[10] A proof of the *Ergodic* theorem (for almost everywhere convergence) for a doubly stochastic operators along the lines of the present paper is easily obtained by using the well-known sample space construction of a Markov process, as has been known for some time (cf. [4]).