Pacific Journal of Mathematics

IN THIS ISSUE—

Vol. 9, No. 3 **July, 1959**

PACIFIC JOURNAL OF MATHEMATICS

Mathematical papers intended for publication in the *Pacific Journal of Mathematics* should be typewritten (double spaced), and the author should keep a complete copy. Manuscripts may be sent to any one of the four editors. All other communications to the editors should be addressed to the managing editor, L. J. Paige at the University of California, Los Angeles 24, California.

50 reprints per author of each article are furnished free of charge; additional copies may be obtained at cost in multiples of 50.

The *Pacific Journal of Mathematics* is published quarterly, in March, June, September, and December. The price per volume (4 numbers) is $12.00; single issues, $3.50. Back numbers are available. Special price to individual faculty members of supporting institutions and to individual members of the American Mathematical Society: $4.00 per volume; single issues, $1.25.

Subscriptions, orders for back numbers, and changes of address should be sent to Pacific Journal of Mathematics, 2120 Oxford Street, Berkeley 4, California.

Printed at Kokusai Bunken Insatsusha (International Academic Printing Co., Ltd.), No. 6, 2-chome, Fujimi-cho, Chiyoda-ku, Tokyo, Japan.

PUBLISHED BY PACIFIC JOURNAL OF MATHEMATICS, A NON-PROFIT CORPORATION

The Supporting Institutions listed above contribute to the cost of publication of this Journal, but they are not owners or publishers and have no responsibility for its content or policies.

A MINIMAL BOUNDARY FOR FUNCTION ALGEBRAS

ERRETT BISHOP

1. Introduction. An algebra $\mathfrak{A}$ of continuous functions on a compact Hausdorff space C will be understood to be a set of complex-valued functions on C which is closed under the operations of addition, multiplication, and multiplication by complex numbers. The algebra $\mathfrak{A}$ is called separating if to any two distinct points of C there exists a function in $\mathfrak{A}$ which takes distinct values at the given points. The norm $\|f\|$ of a continuous function f on a compact space is defined to be the maximum absolute value of the function. The algebra $\mathfrak{A}$ is thus a normed algebra. $\mathfrak{A}$ is called a Banach algebra if it is complete with respect to its norm, i.e., if the limit of every uniformly convergent sequence of elements of $\mathfrak{A}$ is in $\mathfrak{A}$.

An important theorem of Šilov (see [5], p. 80) asserts that if $\mathfrak{A}$ is a separating algebra of continuous functions on a compact Hausdorff space C then there is a smallest closed subset S of C having the property that every function of $\mathfrak{A}$ attains its maximum absolute value at some point of S. This set is called the Šilov boundary of $\mathfrak{A}$. A simple example is obtained by taking C to be a compact subset of the complex plane and $\mathfrak{A}$ to be the set of all continuous functions on C which are analytic at interior points; in this case the Šilov boundary of $\mathfrak{A}$ coincides with the topological boundary of C.

Given a separating normed algebra $\mathfrak{A}$ of continuous functions on a compact space C, it seems natural to ask, in view of Šilov's theorem, whether there exists a smallest subset M (not necessarily closed) of C having the property that every function in $\mathfrak{A}$ attains its maximum absolute value at some point of M. The answer in general is no. However, it will be shown (Theorem 1 below) that such a set M, called the minimal boundary of $\mathfrak{A}$, always exists if in addition it is assumed that $\mathfrak{A}$ is a Banach algebra and that there is a countable basis for the open sets of C, i.e., that C is metrizable. An example will be given to show that the metrizability of C is necessary.

If the minimal boundary M exists, it is clear that the closure of M is the Šilov boundary. An example will be given to show that M need not be closed, so that M in general is smaller than the Šilov boundary. This raises the question of the topological structure of M, which is answered (Theorem 2) by showing that M is a G_δ, i.e., a countable intersection of open sets.

Received December 1, 1958, in revised form February 23, 1959. The author holds a Sloan Foundation Fellowship.

The next portion of the paper concerns the representation of bounded linear functionals on $\mathfrak{A}$ by measures. It is an easy consequence of the classical Hahn-Banach theorem and the Riesz representation theorem that any bounded linear functional φ on $\mathfrak{A}$ of norm 1 can be represented by a (complex-valued, Borel) measure μ of norm 1 on the Šilov boundary S of $\mathfrak{A}$, in the sense that $\varphi(f) = \int_S f d\mu$ for all f in $\mathfrak{A}$. It is natural to conjecture that μ can actually be taken to be a measure on the minimal boundary M of A. The author will devote a subsequent paper to a proof of this result and a consideration of related questions. Karl de Leeuw also has a proof of this result, based on work of Choquet [3]. In the present paper we prove a special case, which is needed to prove the general result and which will be sufficient for the applications considered here. This special case, Theorem 3 below, states that for any point x in $C - M$ there exists a non-negative valued measure μ of norm 1 on $C - \{x\}$ such that $f(x) = \int f d\mu$ for all f in $\mathfrak{A}$.

The final section is concerned with problems of approximation in one complex variable. Necessary and sufficient conditions are obtained on a compact set C without interior of the complex plane that every continuous function on C be uniformly approximable by rational functions whose poles lie in the complement of C. Mergelyan [6] has obtained sufficient conditions, of a different type, that the approximation be possible.

A summary of the results of this paper was given in [1].

2. The minimal boundary.

DEFINITION 1. Let f be a continuous function on a compact space C. *Then* $S(f)$, *the maximal set of* f, *consists of all points* x *in* C *such that* $|f(x)| = ||f||$.

DEFINITION 2. Let $\mathfrak{A}$ be a separating algebra of continuous function on a compact space C. *A subset* N *of* C *is said to bound* $\mathfrak{A}$ *if* $N \cap S(f)$ *is non-void for all* f *in* $\mathfrak{A}$. *If the class of subsets of* C *which bound* $\mathfrak{A}$ *contains a smallest set* M, *the set* M *will be called the minimal boundary of* $\mathfrak{A}$.

THEOREM 1. *Let* $\mathfrak{A}$ *be a separating Banach algebra of continuous functions on a compact metrizable Hausdorff space* C. *Then* $\mathfrak{A}$ *has a minimal boundary* M *and* M *equals the subset* M_0 *of* C *consisting of all* x *in* C *such that there exists* f *in* $\mathfrak{A}$ *with* $S(f) = \{x\}$

Proof. Let N be an arbitrary subset of C which bounds $\mathfrak{A}$. For each x in M_0, there exists f in $\mathfrak{A}$ with $S(f) = \{x\}$. Thus $\{x\} \cap N = S(f) \cap N$ is non-void. Hence $x \in N$. Therefore $M_0 \subset N$.

To show that M_0 is indeed the minimal boundary of $\mathfrak{A}$, it remains to prove that M_0 bounds $\mathfrak{A}$. It must therefore be shown that $M_0 \cap S(f)$ is non-void for each f in $\mathfrak{A}$. Let f be given. Let Γ be the class of all subsets γ of C such that there exists f_γ in $\mathfrak{A}$ with $S(f_\gamma) = \gamma$. By Zorn's lemma, there is a subclass Γ_0 of Γ which contains $S(f)$, which has the finite intersection property, and which has the property that no larger subclass of Γ has the finite intersection property. Since C is compact and since each γ in Γ_0 is closed, the set $D = \bigcap_{\Gamma_0} \gamma$ is non-void and closed. Since there is a countable basis for the open sets of C, and since the family $\{C - \gamma : \gamma \in \Gamma_0\}$ of open sets covers $C - D$, there exists a sequence $\{\gamma_n\}$ from Γ_0 such that $\{C - \gamma_n\}$ covers $C - D$, i.e., such that $D = \bigcap \gamma_n$. Fix a point x_0 of D. Define

$$f_n = [f_{\gamma_n}(x_0)]^{-1} f_{\gamma_n} .$$

Clearly $S(f_n) = \gamma_n$ and $\|f_n\| = f_n(x_0) = 1$. Thus the series $\sum_{n=1}^{\infty} 2^{-n} f_n$ converges uniformly on C to a function g in $\mathfrak{A}$ with $\|g\| = g(x_0) = 1$. If $x \in C - \gamma_k$, then $|g(x)| \leqq \sum 2^{-n} |f_n(x)| < 1$ since $|f_n(x)| \leqq 1$ for all n and $|f_k(x)| < 1$. Therefore $S(g) \subset \gamma_k$. Thus $S(g) \subset \bigcap \gamma_k = D$. Assume that $S(g)$ contains more than one point. Since $\mathfrak{A}$ separates points, there exists h_0 in $\mathfrak{A}$ which is not constant on $S(g)$. We may assume that the maximum of $|h_0|$ on $S(g)$ is 1 and that h_0 takes the value 1 at some point of $S(g)$. If we set $h = h_0 + h_0^2$, it follows that the maximum of $|h|$ on $S(g)$ is 2 and that this maximum is attained only where h_0 takes the value 1. Thus $|h|$ is not constant on $S(g)$. Therefore the set

$$E = \{x : x \in S(g) \text{ and } |h(x)| \geqq |h(y)| \text{ for all } y \text{ in } S(g)\}$$

is a proper closed subset of $S(g)$.

Let x_1 be any point in E. Define the functions

$$g_0 = [g(x_1)]^{-1} g$$

and

$$h_0 = [h(x_1)]^{-1} h .$$

Thus $\|g_0\| = g_0(x_1) = 1$ and $S(g_0) = S(g)$. Also $h_0(x_1) = 1$, $|h_0(x)| \leqq 1$ if $x \in S(g)$, and $|h_0(x)| < 1$ if $x \in S(g) - E$. Let $K = \|h_0\|$. For each positive integer n, let

$$V_n = \{x : 1 + 2^{-n}(K - 1) \leqq |h_0(x)| \leqq 1 + 2^{-n+1}(K - 1)\} .$$

Clearly $\bigcup V_n = \{x : |h_0(x)| > 1\}$. Thus $V_n \cap S(g) = V_n \cap S(g_0)$ is void for each n. Therefore $|g_0(x)| < 1$ for each x in V_n. Since V_n is compact, it follows that there exists a positive integer p_n such that $|g_0(x)|^{p_n} \leqq 1/2$

for all x in V_n. Since $||g_0|| \leqq 1$, the series

$$h_0 + 4(K-1)\sum_{n=1}^{\infty} 2^{-n} g_0^{p_n}$$

converges uniformly on C to a function k in $\mathfrak{A}$. We have

$$k(x_1) = 1 + 4(K-1)\sum_{n=1}^{\infty} 2^{-n} = 1 + 4(K-1) .$$

If $x \in S(g) - E$, then $|h_0(x)| < 1$ and $|g_0(x)| = 1$, so that $|k(x)| < 1 + 4(K-1)$. If $x \in C - \bigcup V_n$, then $|h_0(x)| \leqq 1$ and $|g_0(x)| \leqq 1$, so $|k(x)| \leqq 1 + 4(K-1)$. If $x \in V_j$, then $|h_0(x)| \leqq 1 + 2^{-j+1}(K-1)$, $|g_0(x)|^{p_n} \leqq 1$ for all n, and $|g_0(x)|^{p_j} \leqq 1/2$, so that $|k(x)| \leqq 1 + 4(K-1)$. Therefore $k(x_1) = 1 + 4(K-1) = ||k||$. Thus $x_1 \in S(k)$ and $S(k)$ is disjoint from $S(g) - E$. Since $x_1 \in S(g) \subset D = \bigcap_{\Gamma_0}\gamma$, and since $S(k) \in \Gamma$, it follows from the maximality of Γ_0 with respect to the finite intersection property that $S(k) \in \Gamma_0$. Therefore $S(g) \subset \bigcap_{\Gamma_0}\gamma \subset S(k)$. Since $S(g) - E$ is non-void, this contradicts the fact that $S(g) - E$ is disjoint from $S(k)$. Therefore the assumption that $S(g)$ contains more than one point is false. Thus $S(g)$ consists of a single point x_0. It follows that $x_0 \in M_0$. Since $S(g) \subset D = \bigcap_{\Gamma_0}\gamma \subset S(f)$, it follows that $x_0 \in S(f) \cap M_0$. Thus $S(f) \cap M_0$ is non-void, as was to be proved.

We now give an example to show that Theorem 1 fails if C is not metrizable. Let I denote the unit interval $[0, 1]$ with the usual topology. Let Γ be an uncountable set. Let C consist of all families $x = \{x_\alpha\}_{\alpha \in \Gamma}$ with $x_\alpha \in I$ for each α. Thus C is the Cartesian product of an uncountable number of intervals, and is therefore compact. Let $\mathfrak{A}$ consist of all continuous functions f on C which have the property that there exists a countable subset Δ of Γ such that $f(x) = f(y)$ whenever x and y are points in C such that $x_\alpha = y_\alpha$ for all α in Δ. It is easy to see that $\mathfrak{A}$ is a separating Banach algebra of continuous functions on C. By the Stone-Weierstrass theorem it follows that $\mathfrak{A}$ consists of all continuous functions on C. Let $N_1 = \{x\colon x_\alpha = 0$ except for a countable set of $\alpha\}$ and $N_2 = \{x\colon x_\alpha = 1$ except for a countable set of $\alpha\}$. It is easy to see that N_1 and N_2 bound $\mathfrak{A}$. Since $N_1 \cap N_2$ is void, it follows that $\mathfrak{A}$ does not have a minimal boundary.

For an example of a function algebra whose minimal boundary is distinct from its Šilov boundary, let C be the subset $\{z\colon |z| = 1\}$ of the complex plane and let $\mathfrak{A}$ consist of all continuous functions f on C which have the property that there exists a continuous function $\hat{f}$ on $\{z\colon |z| \leqq 1\}$ such that $\hat{f}(z) = f(z)$ for z in C, such that $\hat{f}$ is analytic on $\{z\colon |z| < 1\}$, and such that $\hat{f}(1) = \hat{f}(0)$. It is easy to see that $\mathfrak{A}$ is a separating Banach algebra of continuous functions on C. It is also not difficult to show that the Šilov boundary of $\mathfrak{A}$ is C, whereas the minimal boundary of $\mathfrak{A}$ is the set $\{z: |z| = 1, z \neq 1\} = C - \{1\}$.

THEOREM 2. *Let $\mathfrak{A}$ be a separating Banach algebra of continuous functions on a compact metrizable Hausdorff space C. For each positive integer n, let U_n consist of all points x in C such that there exists f in $\mathfrak{A}$ with $||f|| \leqq 1$, $|f(x)| > 3/4$, and $|f(y)| < 1/4$ for all y in $D_n(x)$, where $D_n(x) = \{y: \rho(x, y) \geqq n^{-1}\}$ and ρ is a metric on C. Then U_n is open and $\bigcap U_n = M$, where M is the minimal boundary of $\mathfrak{A}$.*

Proof. If f is any function in $\mathfrak{A}$, it is clear that the set $\sigma_n(f) = \{x: x \in C, |f(x)| > 3/4, |f(y)| < 1/4$ whenever $y \in D_n(x)\}$ is open for each n. Since U_n is the union of the sets belonging to the class

$$\{\sigma_n(f): f \in \mathfrak{A},\ ||f|| \leqq 1\} ,$$

it follows that U_n is open.

If $x \in M$, by Theorem 1 there exists f in $\mathfrak{A}$ with $S(f) = \{x\}$. It is clearly no restriction to assume that $||f|| = 1$. Hence $|f(x)| = 1$. Since $|f(y)| < 1$ when y is in the compact set $D_n(x)$, it follows that there exists a positive integer p_n such that $|f(y)|^{p_n} < 1/4$ when $y \in D_n(x)$. Thus $x \in \sigma_n(f^{p_n})$. Therefore $x \in U_n$. Since this is true for each n, it follows that $x \in \bigcap U_n$. Therefore $M \subset \bigcap U_n$.

Now consider a fixed x in $\bigcap U_n$. We must prove that $x \in M$. To this end, we construct by induction a sequence $\{g_n\}$ of functions in $\mathfrak{A}$ having the following properties:

(i) $||g_{n+1} - g_n|| \leqq 2^{-n+1}$
(ii) $||g_n|| \leqq 3(1 - 2^{-n-1})$
(iii) $g_n(x) = 3(1 - 2^{-n})$
(iv) $|g_{n+1}(y) - g_n(y)| < 2^{-n-1}$ if $y \in D_n(x)$.

We first construct g_1. Since $x \in U_1$, there exists a function f in $\mathfrak{A}$ such that $||f|| \leqq 1$ and $x \in \sigma_1(f)$. Let

$$g_1 = \frac{3}{2}[f(x)]^{-1}f .$$

Since $|f(x)| > 3/4$, we have $||g_1|| \leqq 3/2 \cdot 4/3 = 2 < 3(1 - 2^{-2})$, so that g_1 satisfies (ii). Clearly $g_1(x) = 3(1 - 2^{-1})$, so that g_1 satisfies (iii). Hence g_1 satisfies all of the relevant conditions. Assume now that $g_1, \cdots, g_k$ have been chosen to satisfy all of the relevant conditions. Since $g_k(x) = 3(1-2^{-k})$, there exists an integer $j>k$ such that $|g_k(y)|<3(1-2^{-k})+2^{-k-2}$ for $\rho(x, y) < j^{-1}$, i.e., for y in $C - D_j(x)$. Since $x \in U_j$, there exists a function f in $\mathfrak{A}$ such that $||f|| \leqq 1$ and $x \in \sigma_j(f)$. Define $h = 3 \cdot 2^{-k-1}[f(x)]^{-1}f$. Thus $h(x) = 3 \cdot 2^{-k-1}$. Since $||f|| \leqq 1$ and $|f(x)| > 3/4$, we see that $||h|| \leqq 2^{-k+1}$. Since also $|f(y)| < 1/4$ for y in $D_j(x)$, we see that $|h(y)| < 2^{-k-1}$ for y in $D_j(x)$. Let $g_{k+1} = g_k + h$. It follows immediately that

$$\text{(i)} \quad \|g_{k+1} - g_k\| \leqq 2^{-k+1} ,$$

that

$$\text{(iv)} \quad |g_{k+1}(y) - g_k(y)| < 2^{-k-1} \text{ if } y \in D_k(x) ,$$

and that

$$\begin{aligned}\text{(iii)} \quad g_{k+1}(x) = g_k(x) + h(x) &= 3(1 - 2^{-k}) + 3 \cdot 2^{-k-1} \\ &= 3(1 - 2^{-k-1}) .\end{aligned}$$

If $y \in D_j(x)$, then

$$\begin{aligned}|g_{k+1}(y)| &\leqq |g_k(y)| + |h(y)| \\ &< \|g_k\| + 2^{-k-1} \leqq 3(1 - 2^{-k-1}) + 2^{-k-1} \\ &= 3 - 2^{-k} < 3(1 - 2^{-k-2}) .\end{aligned}$$

If $y \in C - D_j(x)$, then

$$\begin{aligned}|g_{k+1}(y)| &\leqq |g_k(y)| + |h(y)| \leqq 3(1 - 2^{-k}) + 2^{-k-2} + \|h\| \\ &\leqq 3(1 - 2^{-k}) + 2^{-k-2} + 2^{-k+1} = 3(1 - 2^{-k-2}) .\end{aligned}$$

It follows that

$$\text{(ii)} \quad \|g_{k+1}\| \leqq 3(1 - 2^{-k-2}) .$$

Thus g_{k+1} has the relevant properties. We have thus constructed the sequence $\{g_n\}$. By condition (i), the sequence $\{g_n\}$ converges uniformly on C to a function g in $\mathfrak{A}$. By (ii), $\|g\| \leqq 3$. By (iii), $g(x) = 3$. If $y \in D_n(x)$, then

$$|g(y)| \leqq \|g_n\| + \sum_{k=n}^{\infty} |g_{k+1}(y) - g_k(y)| < 3(1 - 2^{-n-1}) + \sum_{k=n}^{\infty} 2^{-k-1} < 3 .$$

Thus $S(g) = \{x\}$. Therefore $x \in M$, as was to be proved.

COROLLARY. *If $\mathfrak{A}$ is a separating Banach algebra of continuous functions on a compact metrizable Hausdorff space C, then the minimal boundary M of $\mathfrak{A}$ is a countable intersection of open sets.*

3. Representation by measures.

We now prove the fundamental result of this paper.

THEOREM 3. *Let $\mathfrak{A}$ be a separating Banach algebra of continuous functions on a compact metrizable Hausdorff space C. Let $\mathfrak{A}$ contain the function* 1. *Let x be a point of $C - M$, where M is the minimal* *boundary of* $\mathfrak{A}$. *Then there exists a non-negative Borel measure* γ *of norm* 1 *on* $C-\{x\}$ *such that* $f(x)=\int f d\gamma$ *for all* f *in* $\mathfrak{A}$.

Proof. We assume that a metric ρ on C is given. For each positive integer n, let

$$D_n=\{y\colon y\in C, \rho(x, y)\geqq n^{-1}\}\ .$$

Let b and c be real numbers such that $0<b<1/4<3/4<c<1$. For each positive integer n and each positive integer m, let h_{nm} be a continuous function on C such that $h_{nm}(y)=b^{1/m}$ for $y\in D_n$, $h_{nm}(y)=c^{1/m}$ for $y\in C-D_{2n}$, and $b^{1/m}\leqq h_{nm}(y)\leqq c^{1/m}$ for all y. Such a function exists because the closures of the sets D_n and $C-D_{2n}$ are disjoint. There are two cases to consider. Either there exists (Case 1) for each positive integer n a positive integer m and a function f in $\mathfrak{A}$ such that $|f(x)|>(3/4)^{1/m}$ and $|f(y)|\leqq h_{nm}(y)$ for all y in C, or (Case 2) there exists a positive integer n such that for all positive integers m and for all f in $\mathfrak{A}$ either $|f(x)|\leqq(3/4)^{1/m}$ or $|f(y)|>h_{nm}(y)$ for some y in C. We shall show that Case 1 is impossible and that Case 2 implies the theorem to be proved.

Assume now that Case 1 obtains. Let the positive integer n be given, and choose f in $\mathfrak{A}$ and a positive integer m such that $|f(x)|>(3/4)^{1/m}$ and $|f(y)|\leqq h_{nm}(y)$ for all y.

Write $g=f^m$. Since $|f(y)|\leqq h_{nm}(y)\leqq c^{1/m}$ for all y, we have $|g(y)|\leqq c$ for all y. Thus $\|g\|\leqq c<1$. Since $|f(x)|>(3/4)^{1/m}$ we have $|g(x)|>3/4$. Since $|f(y)|\leqq h_{nm}(y)=b^{1/m}$ for y in D_n, we have $|g(y)|\leqq b<1/4$ for y in D_n. It follows that $x\in U_n$, where U_n is the set defined in Theorem 2. Since this is true for each n, we have $x\in\bigcap U_n=M$, by Theorem 2. This contradicts the hypothesis of Theorem 3. Therefore Case 1 is impossible.

We are therefore justified in assuming that Case 2 obtains. Thus there exists a positive integer n, henceforth fixed, such that for all positive integers m and all f in $\mathfrak{A}$ either $|f(x)|\leqq(3/4)^{1/m}$ or $|f(y)|>h_{nm}(y)$ for some y in C. Consider now a positive integer m. For each f in $\mathfrak{A}$ either $|f(x)|\leqq(3/4)^{1/m}$ or $\|fh^{-1}\|>1$, where $h=h_{nm}$. Thus $|f(x)|\leqq(3/4)^{1/m}$ whenever $f\in\mathfrak{A}$ and $\|fh^{-1}\|\leqq 1$. Let $\mathfrak{B}$ be the Banach space of all continuous functions on C, under the uniform norm, and let $\mathfrak{B}_0$ be the subspace $\{fh^{-1}\colon f\in\mathfrak{A}\}$ of $\mathfrak{B}$. Define the linear functional φ on $\mathfrak{B}_0$ by defining $\varphi(fh^{-1})=f(x)$ for each f in $\mathfrak{A}$. Since $|f(x)|\leqq(3/4)^{1/m}$ if $f\in\mathfrak{A}$ and $\|fh^{-1}\|\leqq 1$, it follows that $\|\varphi\|\leqq(3/4)^{1/m}$. By the Hahn-Banach theorem, there exists an extension φ_0 of φ which is a linear functional on $\mathfrak{B}$ with $\|\varphi_0\|\leqq(3/4)^{1/m}$. By the Riesz representation theorem, there exists a measure ν_m on C such that $\|\nu_m\|\leqq(3/4)^{1/m}$ and $\varphi_0(f)=\int f d\nu_m$ for all continuous functions f on C. Thus

$$f(x) = \varphi(fh^{-1}) = \varphi_0(fh^{-1}) = \int fh^{-1}d\nu_m$$

for all f in $\mathfrak{A}$. If we define the measure μ_m by

$$\mu_m(S) = \int_S h^{-1}d\nu_m$$

for all Borel subsets S of C, it follows that $f(x) = \int fd\mu_m$ for all f in $\mathfrak{A}$. In particular, $1 = \int d\mu_m$. Thus $||\mu_m|| \geqq 1$. Let the measure ν_m^0, the restriction of ν_m to the set D_{2n}, be defined by $\nu_m^0(S) = \nu_m(S \cap D_{2n})$ for each Borel set S. Let ν_m^1, the restriction of ν_m to $C - D_{2n}$, be defined similarly. Thus

$$||\nu_m^0|| + ||\nu_m^1|| = ||\nu_m|| \leqq \left(\frac{3}{4}\right)^{1/m} .$$

Similarly, let μ_m^0 be the restriction of μ_m to D_{2n}, and let μ_m^1 be the restriction of μ_m to $C - D_{2n}$. Thus

$$||\mu_m^0|| + ||\mu_m^1|| = ||\mu_m|| \geqq 1 .$$

Since $[h(y)]^{-1} = c^{-1/m}$ for all y in $C - D_{2n}$ and since $\mu_m^1(S) = \int_{S-D_{2n}} h^{-1}d\nu_m$, for all Borel sets S, we see that $\mu_m^1 = c^{-1/m}\nu_m^1$, so that $c^{1/m}||\mu_m^1|| = ||\nu_m^1||$. Since $|h(y)|^{-1} \leqq b^{-1/m}$ for all y, and therefore for all y in D_{2n}, we see similarly that $b^{1/m}||\mu_m^0|| \leqq ||\nu_m^0||$. Thus

$$b^{1/m}||\mu_m^0|| + c^{1/m}||\mu_m^1|| \leqq ||\nu_m^0|| + ||\nu_m^1|| \leqq \left(\frac{3}{4}\right)^{1/m} .$$

Combined with the inequality

$$b^{1/m}||\mu_m^0|| + b^{1/m}||\mu_m^1|| \geqq b^{1/m} ,$$

this gives

$$[c^{1/m} - b^{1/m}]||\mu_m^1|| \leqq \left(\frac{3}{4}\right)^{1/m} - b^{1/m}$$

Thus

$$||\mu_m^1|| \leqq \left[\left(\frac{3}{4}\right)^{1/m} - b^{1/m}\right][c^{1/m} - b^{1/m}]^{-1} .$$

Since $||\mu_m|| = ||\mu_m^0|| + ||\mu_m^1|| \leqq b^{-1/m}||\nu_m^0|| + b^{-1/m}||\nu_m^1||$
$\leqq b^{-1/m}||\nu_m|| \leqq b^{-1/m}$,

there exists a subsequence $\{\mu_{m_i}\}$ of $\{\mu_m\}$ which converges in the weak star topology for measures on C to a measure μ on C with $||\mu|| \leqq 1$. Also,

$$\int f d\mu = \lim \int f d\mu_{m_i} = f(x)$$

for each f in $\mathfrak{A}$.

Since $C - D_{2n}$ is open and since

$$||\mu_m^1|| \leqq \left[\left(\frac{3}{4}\right)^{1/m} - b^{1/m}\right][c^{1/m} - b^{1/m}]^{-1}$$

for each m, we have

$$||\mu^1|| \leqq \lim \left[\left(\frac{3}{4}\right)^{1/m} - b^{1/m}\right][c^{1/m} - b^{1/m}]^{-1} ,$$

where μ^1 is the restriction of μ to $C - D_{2n}$. Now

$$\lim_{h\to 0}\left[\left(\frac{3}{4}\right)^h - b^h\right][c^h - b^h]^{-1}$$

$$= \lim_{h\to 0}\left[\left(ln\frac{3}{4}\right)\left(\frac{3}{4}\right)^h - (lnb)b^h\right][(lnc)c^h - (lnb)b^h]^{-1}$$

$$= \left[ln\frac{3}{4} - lnb\right][lnc - lnb]^{-1} = ln\left(\frac{3}{4}b^{-1}\right)[ln(cb^{-1})]^{-1} < 1 .$$

Thus, if μ^2 denotes the restriction of μ to the set $\{x\}$, we have $||\mu^2|| \leqq ||\mu^1|| < 1$. Thus there exists a constant a with $|a| < 1$ such that $\int f d\mu^2 = af(x)$ for all continuous functions f on C. Let μ^3 be the restriction of μ to $C - \{x\}$, so that $\mu^2 + \mu^3 = \mu$ and $||\mu^2|| + ||\mu^3|| = ||\mu|| \leqq 1$. Thus

$$\int f d\mu = \int f d\mu_2 + \int f d\mu_3 = af(x) + \int f d\mu_3$$

for all f in $\mathfrak{A}$. Therefore $(1 - a)f(x) = \int f d\mu_3$ for all f in $\mathfrak{A}$. Since $||\mu^3|| \leqq 1 - ||\mu_2|| = 1 - |a|$, and since $1 \in \mathfrak{A}$, we have $1 - |a| \leqq |1 - a| = \left|\int 1 d\mu_3\right| \leqq ||\mu_3|| \leqq 1 - |a|$. Thus a is positive. Define $\gamma = (1 - a)^{-1}\mu_3$. Thus

$$f(x) = (1 - a)^{-1}\int f d\mu_3 = \int f d\gamma$$

for all f in $\mathfrak{A}$. Also, $||\gamma|| \leqq (1 - a)^{-1}||\mu_3|| \leqq 1$. Since $1 \in \mathfrak{A}$ we have $1 = \int d\gamma$. Therefore γ is a non-negative valued measure. This completes the proof of the theorem.

COROLLARY. *Let $\mathfrak{A}$ be a separating Banach algebra containing the* *unit function of continuous functions on the compact metrizable Hausdorff space C. Let* $\mathfrak{R}_0$ *consist of all functions on C which are real parts of functions in* $\mathfrak{A}$, *and let* $\mathfrak{R}$ *be the uniform closure of* $\mathfrak{R}_0$. *Let* M_0 *consist of all points x in C such that there exists f in* $\mathfrak{R}$ *with* $|f(x)| > |f(y)|$ *for all* $y \neq x$ *in C. Then* M_0 *equals the minimal boundary M of* $\mathfrak{A}$.

Proof. If $x \in M$, there exists g in C such that $|g(x)| > |g(y)|$ for all $y \neq x$ in C. It is no loss of generality to assume that $g(x) = 1$. If we let f be the real part of g, then $f \in \mathfrak{R}_0 \subset \mathfrak{R}$ and $|f(x)| = |g(x)| > |g(y)| \geqq |f(y)|$ for all $y \neq x$. Hence $x \in M_0$.

If x is not in M, then there exists a real-valued measure γ on $M - \{x\}$ of norm 1 such that $g(x) = \int g d\gamma$ for all g in $\mathfrak{A}$, by Theorem 3. Since γ is real-valued, it follows that $f(x) = \int f d\gamma$ for all f in $\mathfrak{R}_0$. Thus $f(x) = \int f d\gamma$ for all f in $\mathfrak{R}$. If x where in M_0, there would exist f in $\mathfrak{R}$ with $1 = f(x) > |f(y)|$ for all $y \neq x$. Thus $1 = f(1) = \int f d\gamma \leq \left|\int f d\gamma\right| < 1$, since $|f(y)| < 1$ for $y \neq x$ and since $||\gamma|| = 1$. This contradiction shows that x is not M_0. Hence $M = M_0$, as was to be proved.

DeLeeuw has found a proof of Theorem 3 which is somewhat simpler than the one given here.

4. Applications.

We now apply the results of the previous sections to certain problems of approximation in one complex variable.

DEFINITION 3. Let C be a compact subset of the complex plane. *Then* $\Lambda_0(C)$ *will consist of all continuous functions on C which are analytic at interior points of C, and* $\Lambda_1(C)$ *will consist of all continuous functions on C which can be uniformly approximated arbitrarily closely by rational functions whose poles lie in the complement of C.*

It is clear that $\Lambda_1(C) \subset \Lambda_0(C)$, and that $\Lambda_0(C)$ and $\Lambda_1(C)$ are separating Banach algebras of continuous functions on C. Mergelyan [2] has shown that $\Lambda_0(C) = \Lambda_1(C)$ in case the complement of C consists of only a finite number of components. No necessary and sufficient condition is known that $\Lambda_1(C) = \Lambda_0(C)$. In case C has no interior, we shall obtain in Theorem 5 below a necessary and sufficient condition that every continuous function on C be uniformly approximable by rational functions with poles in $-C$.

THEOREM 4. *Let C be a compact subset of the complex plane with no interior and let M be the minimal boundary of* $\Lambda_1(C)$. *Then either*

$\Lambda_1(C) = \Lambda_0(C)$ *or* $C - M$ *has positive 2-dimensional Lebesgue measure.*

Proof. Assume that $\Lambda_1(C) \neq \Lambda_0(C)$. We must show that $C - M$ has positive 2-dimensional Lebesgue measure. Now $\Lambda_0(C)$ is the Banach space of all continuous complex-valued functions on C, and $\Lambda_1(C)$ is a proper subspace, since $\Lambda_1(C) \neq \Lambda_0(C)$. By the Hahn-Banach theorem, there exists a continuous linear functional $\varphi \neq 0$ on $\Lambda_0(C)$ which vanishes on $\Lambda_1(C)$. By the Riesz representation theorem, there exists a finite complex-valued Borel measure μ on C which represents φ. Thus $\mu \neq 0$ and $\int f d\mu = 0$ for all f in $\Lambda_1(C)$. In particular, $\int (z - \zeta)^{-1} d\mu(\zeta) = 0$ whenever z is not in C, since the function $(z - \zeta)^{-1}$ is a rational function of ζ whose pole, z, is not in C. Since the function z^{-1} of z is integrable with respect to Lebesgue measure $dxdy$ over any finite region of the plane, and since μ is a finite measure on the compact set C, the integral

$$h(z) = \int (z - \zeta)^{-1} d\mu(\zeta)$$

will exist for almost all values of z, and the function $h(z)$ so defined, called the convolution of the measure μ and the function z^{-1} and written $h = z^{-1} * \mu$, will be integrable with respect to Lebesgue measure over any finite region of the plane. Since we have seen above that $h(z) = 0$ if z is not in C, it follows that h is integrable.

Assume that the integrable function h vanishes almost everywhere, so that the integral $h(z) = \int (z - \zeta)^{-1} d\mu(\zeta)$ exists and vanishes for almost all z. To obtain a contradiction from this assumption, we use the equation $\frac{\partial}{\partial z^*}\frac{1}{z} = \pi\delta$ ([7] p. 49) from the theory of distributions. This means that for any function g on the complex plane which vanishes in a neighborhood of ∞ and which has continuous partial derivatives of all orders we have

$$-\iint (z - \zeta)^{-1} \frac{1}{2}\left(\frac{\partial}{\partial x} + i\frac{\partial}{\partial y}\right) g(z) dxdy = \pi g(\zeta)$$

for all values of ζ. If we write $g_1(z) = \frac{1}{2}\left(\frac{\partial}{\partial x} + i\frac{\partial}{\partial y}\right) g(z)$ and integrate both sides of the above equation with respect to μ we obtain

$$\begin{aligned} \pi\int g(\zeta) d\mu(\zeta) &= -\int\left\{\iint (z - \zeta)^{-1} g_1(z) dxdy\right\} d\mu(\zeta) \\ &= -\iint g_1(z)\left\{\int (z - \zeta)^{-1} d\mu(\zeta)\right\} dxdy = 0 , \end{aligned}$$

since $h(z)$ vanishes almost everywhere. The use of Fubini's theorem is justified since $z - \zeta$, and therefore $(z - \zeta)^{-1}$, is measurable with respect to the product of the measures $g_1(z)dxdy$ and $d\mu(\zeta)$ and since

$$\int\left\{\iint |g_1(z)||z - \zeta|^{-1}dxdy\right\}|d\mu(\zeta)|$$

is finite. Now every continuous function g_0 on C can be uniformly approximated by such functions g, so that $\int g_0(\zeta)d\mu(\zeta) = 0$. By the uniqueness part of the Riesz representation theorem, it follows that $\mu = 0$. This contradiction shows that there exists a set Γ of C of positive Lebesgue measure such that the integral $h(z)$ exists and does not vanish for all z in Γ. We may clearly assume that at no point of Γ does μ have point mass.

Let z_0 be any point in Γ, so that $h(z_0) = c$ exists and is not zero. Let f be any function in $\Lambda_1(C)$ such that $f(z_0) = 0$. Let $\{f_n\}$ be a sequence of rational functions with poles in the complement of C converging uniformly to f on C. Since $f_n(z_0) \to f(z_0) = 0$ as $n \to \infty$, we see that $\{g_n\}$ converges uniformly to f on C, where $g_n = f_n - f_n(z_0)$. Thus g_n is a rational function with poles in $-C$ which vanishes at z_0, so that there exists a rational function g'_n with poles in $-C$ such that $g_n(z) = g'_n(z)(z - z_0)$ for all z. Hence

$$\int g_n(z)(z - z_0)^{-1}d\mu(z) = \int g'_n(z)d\mu(z) = 0$$

for each n, since $g'_n \in \Lambda_1(C)$. Passing to the limit, we see that

$$\int f(z)(z - z_0)^{-1}d\mu(z) = 0 .$$

Since this is true for all f in $\Lambda_1(C)$ with $f(z_0) = 0$, it follows that for an arbitrary f in $\Lambda_1(C)$ we have

$$\int f(z)(z - z_0)^{-1}d\mu(z) = \int [f(z) - f(z_0)](z - z_0)^{-1}d\mu(z)$$
$$+ \int f(z_0)(z - z_0)^{-1}d\mu(z) = 0 + f(z_0)(-h(z_0)) = -cf(z_0) .$$

If we let δ_0 denote the measure of mass $+1$ at the point z_0, it follows that

$$\nu = (z - z_0)^{-1}\mu + c\delta_0$$

is a measure on C which annihilates $\Lambda_1(C)$. Now if z_0 were in M, there would exist f in $\Lambda_1(C)$ with $f(z_0) = 1$ and $|f(z)| < 1$ for all $z \neq z_0$ in C. Since $\nu(\{z_0\}) = c \neq 0$, it is clear that $\int f^n d\nu \neq 0$ if n is sufficiently large.

This is a contradiction, since $f^n \in \Lambda_1(C)$. This shows that z_0 is not in M. Since z_0 was any point in Γ we have $\Gamma \subset C - M$. Since Γ has positive measure, $C - M$ has positive measure, as was to be proved.

To restate the theorem, if for every point z on C, with the possible exception of a set of measure 0, there exists a continuous function f on C with $|f(z)| > |f(\zeta)|$ for all $\zeta \neq z$ in C which can be uniformly approximated by rational functions with poles in $-C$, then every continuous function on C can be uniformly approximated by rational functions whose poles lie in $-C$.

THEOREM 5. *Let C be a compact subset of the complex plane without interior. Let $\Lambda_0(C)$ be the algebra of all continuous complex-valued functions on C and let $\Lambda_1(C)$ be those functions in $\Lambda_0(C)$ which can be uniformly approximated by rational functions with poles in $-C$, and let M be the minimal boundary of $\Lambda_1(C)$. Let $\Delta_0(C)$ consist of all continuous real-valued functions on C, and let $\Delta_1(C)$ consist of all continuous real-valued functions on C which are uniformly approximable by real parts of functions in $\Lambda_1(C)$. Let M_0 consist of all points z in C such that there exists f in $\Delta_1(C)$ with $|f(z)| > |f(\zeta)|$ for all $\zeta \neq z$ in C. Then $M = M_0$ and the following statements are equivalent:*

(i) $\Lambda_1(C) = \Lambda_0(C)$
(ii) $C - M$ *has measure* 0
(iii) $M = C$
(iv) $\Delta_1(C) = \Delta_0(C)$.

Proof. The fact that $M = M_0$ is a special case of the corollary to Theorem 3. It is clear that (i) $\Rightarrow$ (iii) $\Rightarrow$ (ii). But (ii) $\Rightarrow$ (i) by Theorem 4. Thus (i), (ii), and (iii) are equivalent. It is also clear that (i) $\Rightarrow$ (iv). But (iv) implies that $M_0 = C$. Thus (iv) $\Rightarrow$ (iii). This proves Theorem 5.

Theorem 5 thus gives results concerning approximation on a nowhere dense subset of the complex plane by rational functions or by real parts of rational functions, and shows that the two problems are related. The results for approximation by the real parts of rational functions are similar in outward appearance to results of Brelot [2] and Deny [4], who consider approximation by functions harmonic in a neighborhood of C, but there does not seem to be an essential connection, due to the fact that a function harmonic in the neighborhood of C need not be the real part of an analytic function, since the conjugate harmonic function might be multiple-valued.

REFERENCES

1. E. Bishop, *Some theorems concerning function algebras*, Bull. Amer. Math. Soc. **65** (1959), 77–78.

2. M. Brelot, *Critères de régularite et de stabilité,* Acad. Roy. Belgique Bull. Cl. Sci. **25** (1939), 125-137.
3. G. Choquet, *Existence des représentations intégrales au moyen des points extrémaux dans les cônes convexes*, C. R. Acad. Sci. Paris **243** (1956), 699-702.
4. J. Deny, *Sur l'approximation des fonctions harmoniques*, Bull. Soc. Math. France **73** (1945), 71-73.
5. L. H. Loomis, *Abstract harmonic analysis*, New York, 1953.
6. S. N. Mergelyan, *On the representation of functions by series of polynomials on closed sets*, Amer. Math. Soc. Translation no. 85, 1953.
7. L. Schwartz, *Théorie des distributions* vol. 1, Actualities Scient. et Indust., Paris, 1951.

UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA

THE TOPOLOGY OF ALMOST UNIFORM CONVERGENCE

JOHN W. BRACE

The theorem of Arzelà [3, 4, 6] (see Theorem 2.2) which gives a necessary condition and a sufficient condition for a net of continuous functions to converge to a continuous function plays an important part in functional analysis. In the case of linear topological spaces it has been observed that the quasi-uniform convergence [3] (see Definition 2.1) which Arzelà presented in his theorem is related to the weak and weak* topologies [6, 9, 20]. With this fact in mind it was surmised that quasi-uniform convergence would present a useful method for topologizing function spaces. This paper presents such a topology and displays some of its properties and applications. The resulting topology will be called the topology of almost uniform convergence.

In § 1 the topology is defined by means of a base for the neighborhood system of the zero function (origin). It should be noted that there is a similarity between the development of uniform convergence topologies [17] and the topology of almost uniform convergence. Section 2 shows that convergence of a net of functions for the topology implies quasi-uniform convergence. A net of functions having the property that every subnet converges quasi-uniformly will converge for the topology. In § 3 the concept of almost uniform convergence is extended to the case where convergence occurs on each member of a family of subsets of the domain space. Section 4 examines the properties of various function spaces in regard to the topology of almost uniform convergence. In particular, Theorem 4.3 shows that convergence in this topology for a net of bounded continuous functions over a regular Hausdorff space S is equivalent to pointwise convergence of their extensions on the Stone-Čech compactification of S.

Section 5 uses the topology of almost uniform convergence to obtain the weak topology for certain locally convex linear topological spaces. It is necessary in § 5 to modify the topology of almost uniform convergence to form a finer (stronger) topology which is called the topology of convex almost uniform convergence. With this new topology, Theorem 5.6 shows that the weak topology for a function space, which was originally a locally convex linear topological space for a uniform convergence topology, is the topology of convex almost uniform convergence. Theorem 5.9 parallels a theorem in Banach's book (page 134) [5] in giving a necessary and sufficient condition for the weak conver-

Received November 3, 1958. This research was supported by the United States Air Force through the Air Force Office of Scientific Research of the Air Research and Development Command under Contract No. AF 18 (603)-78.

Presented to the Society, October 27, 1956.

gence of a net (instead of a sequence) from the Banach space of all continuous on the closed unit interval.

In the same manner Theorem 5.10 gives a necessary and sufficient condition for the convergence in the weak topology of a net in L^1. A similar theorem for sequences can be found in [12; page 89].

1. The almost uniform convergence topology. In establishing the existence of the almost uniform convergence topology for a collection of functions, the first step is to determine a class of sets that can be used to generate the neighborhood system of the zero element of the function space [7, 15, 19].

1.1 DEFINITION. Let $\mathscr{G}(S, F)$ be a linear space of functions on an abstract set S into a locally convex linear topological space F. Then a subset U of $\mathscr{G}(S, F)$ has the *property* (α) in $\mathscr{G}(S, F)$ if it satisfies the condition: for some neighborhood V of 0 in F it is true that for each finite subset $\{f_i, \cdots, f_k\}$ of $\mathscr{G}(S, F) \sim U$ there is an x in S such that $f_i(x) \notin V$ $(i = 1, \cdots, k)$.

1.2 LEMMA. *For each subset U of $\mathscr{G}(S, F)$ with property (α), there is a convex circled subset W of U with property (α). Furthermore each U is radial at the origin of $\mathscr{G}(S, F)$ if and only if $f[S]$ is bounded for each f in $\mathscr{G}(S, F)$.*

Proof. Let V_0 be a convex circled neighborhood of 0 in F whose relationship to U is as stated in property (α). Accepting the Hausdorff minimal principle, there exists a subset of U, call it W, which is minimal with respect to the property (α) for the previously mentioned neighborhood V_0 of 0 in F.

Observe that a function f is in W if and only if there exists a finite subset $\{f_i, \cdots, f_n\}$ of $\mathscr{G}(S, F) \sim W$ such that for every x in S either $f(x)$ is in V_0 or $f_j(x)$ is in V_0 for at least one f_j, $j = 1, \cdots, k$. For two functions f and g in W let the two finite sets $\{f_1, \cdots, f_m\}$ and $\{g_1, \cdots, g_n\}$ bear this relationship to f and g respectively. The union $\{f_1, \cdots, f_m, g_1, \cdots, g_n\}$ bears the same relationship to $tf + (1 - t)g$ for all t such that $0 \leq t \leq 1$. Thus it is concluded that W is a convex set.

The circularity of W is obtained in a similar manner.

Assume that for each f in $\mathscr{G}(S, F)$ there is a scalar c such that $f[S] \subset cV_0$. This implies that f is in $cW \subset cU$. In other words, U is radial at 0 in $\mathscr{G}(S, F)$.

Assuming each U to be radial and the existence of an f in $\mathscr{G}(S, F')$ such that $f[S]$ is not bounded, there is a sequence $\{x_n\} \subset S$ and a convex circled neighborhood V of the origin in F with the property that $\frac{1}{n}f(x_n)$

is not in V for $n = 1, 2, \cdots$. Let $U_0 = \mathscr{G}(S, F) \sim \left\{f, \frac{1}{2}f, \frac{1}{3}f, \cdots\right\}$. This is a contradiction because U_0 has property (α) and is not radial.

1.3 THEOREM. *Let $\mathscr{G}(S, F)$ be a linear space of functions on an abstract set S into a locally convex linear topological space F. Then all the sets of the form $U_1 \cap \cdots \cap U_n$, where U_i, has property (α) in $\mathscr{G}(S, F)$, form a local base for a locally convex topology. This topology is called the topology of almost-uniform convergence on S. Furthermore, it is a linear topology if and only if $f[S]$ is bounded for each f in $\mathscr{G}(S, F)$ and it is a Hausdorff topology if F is Hausdorff.*

Proof. The existence of the almost-uniform convergence topology (linear and non-linear) is obtained from Lemma 2.3 [7, 15, 19].

If F is Hausdorff and f is a non-zero element of $\mathscr{G}(S, F)$ then there is a point x in S and a neighborhood V of 0 such that $f(x)$ is not in V. Thus $\mathscr{G}(S, F) \sim \{f\}$ has property (α) and the almost uniform convergence topology is Hausdorff.

2. The convergence of nets.

The study of almost uniform convergence topologies is partially motivated by quasi-uniform convergence and almost uniform convergence for nets of functions [17]. Now that the topology has been obtained it is time to consider its relationship to the almost uniform convergence of a net of functions.

2.1 DEFINITION. [3, 6, 21]. A net $\{f_\alpha, \alpha \in D\}$ in $\mathscr{G}(S, F)$ converges to f_0 quasi-uniformly on S if $\lim f_\alpha(x) = f_0(x)$ for each x in S and for every neighborhood V of the zero element of F and α_0 in D there is a finite subset $\{\alpha_1, \alpha_2, \cdots, \alpha_n\}$ of D, $\alpha_i \geq \alpha_0$, $i = 1, 2, \cdots, n$, such that for every x in A, $f_{\alpha_i}(x) - f_0(x)$ is in V for at least one α_i, $i = 1, 2, \cdots, n$.

The importance of quasi-uniform convergence stems from the following theorem.

2.2 THEOREM (Arzelà) [3, 4, 6]. *If a net of continuous functions on a topological space X converges to a continuous limit, then the convergence is quasi-uniform on every compact subset of X. Conversely, if the net converges quasi-uniformly on a subset of X, the limit is continuous on this subset.*

Since a net converges in a topological space if and only if every subnet converges, the following modification of quasi-uniform convergence is the natural thing to expect.

2.3 DEFINITION [19]. A net $\{f_\alpha, \alpha \in D\}$ in $\mathscr{G}(S, F)$ converges almost uniformly to f_0 on S if and only if every subnet converges quasi-uniformly to f_0 on A.

2.4 THEOREM. *Let $\mathscr{G}(S, F)$ be a function space with the topology of almost uniform convergence on S. A net $\{f_\alpha, \alpha \in D\}$ in $\mathscr{G}(S, F)$ converges almost uniformly to f_0 on S if and only if the net converges to f_0 in the topology.*

Proof. Considering f_0 to be the zero function, it is assumed that the net converges almost uniformly to f_0 without being eventually in a set U with property (α). This gives a subnet which would not converge quasi-uniformly and thus a contradiction.

In order to obtain the converse, assume first that the net which converges for the topology does not converge pointwise at the point x in S. It follows that there is a subnet $\{f_\beta, \beta \in D'\}$ and a convex circled neighborhood V of 0 in F such that $f_\beta(x)$ is not in V for each β in D'. This leads to a contradiction because the set $\mathscr{G}(S, F) \sim \{f_\beta, \beta \in D'\}$ has property (α) and the net is not eventually in it.

A similar contradiction is obtained when it is assumed that there is a subnet which converges pointwise and in the topology, but does not converge quasi-uniformly.

3. The topology of almost uniform convergence on a collection of subsets of S. In the above discussion the convergence has occurred over the entire set S. Without difficulty, the convergence can be restricted to a family of subsets of S.

3.1 DEFINITION. A subset U of $\mathscr{G}(S, F)$ has the property (α) over a subset A of S if it has property (α) in $\mathscr{G}(A, F)$, where $\mathscr{G}(A, F)$ is the linear space which is obtained by restricting the functions in $\mathscr{G}(S, F)$ to having A as their domain of definition.

The analogue of Theorem 1.3 is now stated.

3.2 THEOREM. *Let $\mathscr{G}(S, F)$ be a linear space of function on a set S into a locally convex linear topological space F and let $\mathscr{A}$ be a collection of subsets of S. Then all the sets in $\mathscr{G}(S, F)$ of the form $U_1 \cap \cdots \cap U_u$, where each U_i has property (α) over some A in $\mathscr{A}$, form a local base for a locally convex topology. This topology of almost uniform convergence on members of $\mathscr{A}$. Furthermore, it is a linear topology if and only if $f[A]$ is bounded for each A in $\mathscr{A}$ and each f in $\mathscr{G}(S, F)$ and it is a Hausdorff topology if F is Hausdorff and for each f in $\mathscr{G}(S, F)$ there is a point x in at least one member of $\mathscr{A}$ such that $f(x) \neq 0$.*

The expected analogues of the theorems in § 1 are also valid. In analogy with topologies of uniform convergence it is noted that $\mathscr{A}$ can be enlarged to contain all the finite unions of its members without affecting the topology.

4. Function spaces. It is interesting to note that the subset $\mathscr{B}$ of $\mathscr{F}(S, F)$ consisting of all functions which are bounded on each member of $\mathscr{A}$ is a closed subspace of $\mathscr{F}(S, F)$ for the topology of almost uniform convergence on members $\mathscr{A}$. If S is a topological space it is also clear that the functions which are continuous on each member of $\mathscr{A}$ form a closed subspace.

4.1 THEOREM. *Let $C(S, F)$ be a linear space of continuous functions on a topological space S to a locally convex linear topological space F. Then the topology of almost uniform convergence on members of $\mathscr{A}$ and the topology of almost uniform convergence on the members of a collection composed of the closures of finite unions of members of $\mathscr{A}$ are the same topology on $C(S, F)$.*

Proof. Assuming that the subset U of $C(S, F)$ has property (α) over the closure of A, where A is in $\mathscr{A}$, there is a closed convex circled neighborhood V of 0 in F such that for each finite subset $\{f_1, \cdots, f_k\}$ of $C(S, F) \sim U$ there exists an x_0 in the closure of A and a neighborhood W of x_0 with the property that $f_j(W)$ is disjoint from V for $j = 1, \cdots, k$. Since there is an x in W which is also in A, the proof is completed by concluding that U has property (α) over A.

The following theorems give some indication of the relationship between almost uniform convergence and pointwise convergence.

4.2 THEOREM. *If A is a compact subset of a topological space S then the topology of almost uniform convergence on A is equivalent to the topology of pointwise convergence on A for a function space $C(S, F)$ of continuous functions defined on S with range in the locally convex linear topological space F.*

Proof. It has already been shown that almost uniform convergence implies pointwise convergence. The converse is immediately obtained by nothing that Arzelà's Theorem 2.2 establishes the quasi-uniform convergence of every subnet of a pointwise convergent net in $C(S, F)$.

Theorems 4.1 and 4.2 combine to obtain the following result.

4.3 THEOREM. *Let S be a completely regular Hausdorff space. The topology of almost uniform convergence on S is equivalent to the topology of pointwise convergence on the Stone-Čech compactification* [17] *of S for the function space of bounded continuous functions on S with range in the complex or real numbers.*

A noticeable difference between the uniform convergence topologies and almost uniform convergence topologies occurs on questions of completeness. For example, the almost uniform convergence topology on the linear space of all bounded real valued functions defined on the closed unit interval [0, 1] is not complete. If it were a complete topology the subspace of continuous function would be required to be complete in the topology of pointwise convergence (see Theorems 2.2 and 4.3).

5. Adjoint spaces and the weak topology. [2, 8, 11, 18, 19]. Several people have observed that almost uniform convergence is in some manner related to the adjoint space of a locally convex linear topological space. The first adjoint space is the collection of all continuous scalar valued linear functions defined on the linear topological space which is under consideration. If the adjoint space is topologized it is possible to speak of the adjoint space of the adjoint space.

The adjoint space E^* of a locally convex linear topological space E defines a natural topology on E which is called the weak topology. If E is considered as a collection of linear functions defined on E^*, the weak topology on E is the topology of pointwise convergence on E^*.

By interchanging the roles of E and E^* in the above discussion it is seen that E gives rise to a natural topology on E^* which is called the weak* topology on E^*.

The next theorem gives a small degree of insight into the structure of the adjoint space.

5.1 THEOREM. *Consider a linear space $\mathscr{G}(S, K)$ consisting of functions with domain S and range in the scalar field K. If $\mathscr{G}(S, K)$ is given a linear topology of uniform convergence on members of a collection $\mathscr{A}$ of subsets of S and $\mathscr{G}^*$ is the adjoint space of $\mathscr{G}(S, F)$, then there exists a natural mapping ϕ from the subset $\bigcup_{A \in \mathscr{A}} A$ of S into $\mathscr{G}^*$ such that for each $x \in \bigcup_{A \in \mathscr{A}} A$, $\phi(x)f = f(x)$ for every $f \in \mathscr{G}(S, K)$ and for each f^* in $\mathscr{G}^*$ there is an A in $\mathscr{A}$ and positive scalar ε with the property that εf^* is in the weak* closed circled convex hull of $\phi[A]$.*

Proof. For each $x \in \bigcup_{A \in \mathscr{A}} A$, $\phi(x)$ is clearly an element of $\mathscr{G}^*$. Considering an arbitrary f_0^* in $\mathscr{G}^*$, let $G = \{f \in \mathscr{G}(S, K): |f_0^*(f)| \leq 1\}$. The continuity of f_0^* gives a positive number ε and an A in $\mathscr{A}$ such that G contains the neighborhood $H = \{f \in \mathscr{G}(S, K): |f(x)| \leq \varepsilon$ for all $x \in A\}$. It can be shown that the set $\{f^* \in \mathscr{G}^*: |f^*(f)| \leq \varepsilon$ for each f in $H\}$ which contains εf_0^* is the weak* closed circled convex hull of $\phi[A]$[8].

5.2 COROLLARY. *Let E be a locally convex linear topological space with an adjoint space E^*. If E^* is given the topology of uniform convergence on the bounded subsets of E then the adjoint space E^{**} of E^* is the union of the weak* closures of the images in E^{**} of the bounded subsets of E under the natural mapping of E into E^{**}.*

In the specific case of a Banach space the results presented in the corollary were proved by M. M. Day [10] and H. H. Goldstine [13].

The weak topology on the Banach space of all continuous functions defined on the closed unit interval is finer (stronger) than the topology of pointwise convergence on the closed unit interval. Since the topology of pointwise convergence is the topology of almost uniform convergence on the closed unit interval it is clear that almost uniform convergence must be modified if it is to give the weak topology. The following definition is presented with this purpose in mind.

5.3 DEFINITION. Let $\mathscr{G}(S, F)$ be a linear space of functions on an abstract set S into a locally convex linear topological space F. Then a subset U of $\mathscr{G}(S, F)$ has property β over a subset A of S if it satisfies the following condition: for some neighborhood V of 0 in F it is true that for each finite subset $\{f_1, \cdots, f_k\}$ of $\mathscr{G}(S, F) \sim U$ there is a finite subset $\{x_1, x_2, \cdots, x_n\}$ of A and a finite set of positive numbers $\{a_1, a_2, \cdots, a_n\}$, $\sum_{i=1}^n a_i = 1$, such that $\sum_{i=1}^n a_i f_j(x_i)$ is not in V for $j = 1, 2, \cdots, k$.

5.4 THEOREM. *Consider the function space $\mathscr{G}(S, F)$. Then all sets of the form $U_1 \cap \cdots \cap U_n$, where each U_i has property (β) over some A in $\mathscr{A}$, form a local base for a locally convex topology. This is called the topology of convex almost uniform convergence on members of $\mathscr{A}$. Furthermore, it is a linear topology if and only if $f[A]$ is bounded for each A in $\mathscr{A}$ and each f in $\mathscr{G}(S, F)$ and it is a Hausdorff topology if F is Hausdorff and for each f in $\mathscr{G}(S, F)$ there is a point x in at least one member of $\mathscr{A}$ such that $f(x) \neq 0$.*

The omitted proof of the above theorem is essentially the same as Theorem 2.4.

5.5 THEOREM. *Let S be a linear topological space and let $\mathscr{L}(S, F)$ be a collection of continuous linear functions defined on S with range in a locally convex linear topological space F. If $\mathscr{A}$ is a family of subsets of S such that $\mathscr{L}(S, F)$ is a linear topological space for the topology of convex almost uniform convergence on members of $\mathscr{A}$ and if $\mathscr{A}'$ is the collection of closed convex hulls of finite unions of members of $\mathscr{A}$, then the topology of almost uniform convergence on the members of $\mathscr{A}'$ is the same topology.*

Proof. In collaboration with Theorem 4.1 it is sufficient to show that a subset U of $\mathscr{L}(S, F)$ has property (β) over a subset A of S if and only if it has property (α) on the convex hulls of A. Because of the linearity of the members of U the result becomes apparent upon inspecting Definitions 5.3 and 1.2.

The theorem which follows shows that it is possible to work with the weak topology on a function space without knowing anything about the first adjoint space. In many cases this avoids the necessity of obtaining a representation of the adjoint space.

5.6 THEOREM. *Consider a linear space $\mathscr{G}(S, K)$ consisting of functions with domain an abstract set S and range in the scalar field K. If $\mathscr{G}(S, K)$ is given a linear topology of uniform convergence on the members of a collection $\mathscr{A}$ of subsets of S, then the weak topology on $\mathscr{G}(S, K)$ is the topology of convex almost uniform convergence on members of $\mathscr{A}$.*

Proof. Considering $\mathscr{G}(S, K)$ as a collection of scalar valued functions defined on $\mathscr{G}^*$, the weak topology on $\mathscr{G}(S, K)$ is the topology of pointwise convergence on the union of the weak* closed circled convex hulls of the collection $\{\phi[A]: A \in \mathscr{A}\}$ (see Theorem 5.1). The weak* closed circled convex hull of each $\phi[A]$ is weak* compact because it is the polar of a neighborhood of 0 in $\mathscr{G}(S, K)$[8]. Since each member of $\mathscr{G}(S, K)$ is a weak* continuous function on $\mathscr{G}^*$[8], the weak topology is the topology of convex almost uniform convergence on the collection $\{\phi[A]: A \in \mathscr{A}\}$ (see Theorems 5.5 and 4.2), which in turn is the topology of convex almost uniform convergence on members of $\mathscr{A}$.

5.7 COROLLARY. *If E is a locally convex linear topological space and the first adjoint space E^* is given the strong topology, then the weak topology on E^* is the topology of almost uniform convergence on the bounded subsets of E.*

Further results of this type can be obtained for a function space or an operator space whose range is contained in a locally convex linear topological space.

The above relationship of the weak topology to an almost uniform convergence topology again displays the restrictive nature of a completeness requirement on an almost uniform convergence topology [16].

In the case where S is a linear space, a topology of almost uniform convergence on the members of $\mathscr{A}$ is not always equivalent to the topology af almost uniform convergence on the convex hulls of the members of $\mathscr{A}$. To clarify this point consider the weak topology on a locally convex linear topological space E. It is the topology of almost uniform convergence on the weak* compact subsets of the first adjoint space E^* (see Theorem 4.2) If the weak topology on E was also the topology of almost uniform convergence on the convex hulls of the weak* compact subsets of E^* the topology of uniform convergence on the same subsets would be a linear topology for E with the same adjoint space. This leads to contradiction in cases where the new topology would be properly finer than the Mackey topology on E [18, 19].

In the case of a normed linear space the following theorem due to G. Sirvint [21] is of interest. The original form of the theorem concerns sequences and not nets.

5.8 THEOREM. *Let F be a normed linear space. A net $\{f_\alpha, \alpha \in D\}$ in the function space $\mathscr{C}(S, F)$ converges almost uniformly to f_0 on a subset A of S if and only if $\lim_\alpha \underline{\lim}_\beta \|f_\alpha(x_\beta) - f_0(x_\beta)\| = 0$ for every net $\{x_\beta, \beta \in D'\}$ in A.*

In Banach's book [5] on page 134 there is a necessary and sufficient condition for a sequence from the Banach space of all continuous functions on the closed unit interval to converge weakly. This theorem does not hold in the case of nets. It is now possible to state a similar theorem for nets by making use of Theorems 5.6 and 5.8. Let Ξ be the collection of all sets consisting of a finite number of ordered pairs $\{(a_1, t_1), (a_2, t_2), \cdots, (a_n, t_n)\}$ of numbers from the closed unit interval $[0, 1]$ with the property that $\sum_{i=1}^n a_i = 1$. For each $\psi \varepsilon \Xi$, $\psi = \{(a_1, t_1), (a_2, t_2), \cdots, (a_n, t_n)\}$, and each real valued function f on the closed unit interval define $\psi(f)$ to be $\sum_{i=1}^n a_i f(t_i)$. With this notation the theorem can be stated.

5.9 THEOREM. *A net $\{f_\alpha, \alpha \in D\}$ from the Banach space $C[0, 1]$ converges weakly to f_0 if and only if $\lim_\alpha \underline{\lim}_\beta |\psi_\beta(f_\alpha) - \psi_\beta(f_0)| = 0$ for every net $\{\psi_\beta, \beta \in \mathscr{B}\}$ in Ξ.*

A similar theorem is available for the Banach space L^1. Let $\mathscr{M}$ be the collection of all measurable sets from the measure space upon which the function of L^1 are defined. The norm $\|f\| = \sup_{M \in M}\left|\int_E f d\mu\right|$ is topologically equivalent to the usual norm for L^1. If L^1 is viewed as a collection of functions defined on $\mathscr{M}$ with its topology determined by the new norm, it satisfies the hypothesis of Theorem 5.6.

Let Φ be the collection of all sets consisting of a finite number of ordered pairs $\{(a_1, M_1), (a_2, M_2), \cdots, (a_n, M_n)\}$ where M_i $(i = 1, 2, \cdots, n)$ is a measurable set from the measure space and $(a_1, a_2, \cdots, a_n)$ is a set of possitive numbers with the property $\sum_{i=1}^n a_i = 1$. For each $\varphi \in \Phi$, $\varphi = \{(a_1 M_1), \cdots, (a_n, M_n)\}$, and for each $f \in L^1$ define $\varphi(f)$ to be $\sum_{i=1}^n a_i \int_{M_i} f d\mu$.

This notation makes it possible to state the following theorem.

5.10 THEOREM. *A net $\{f_\alpha, \alpha \in D\}$ from the Banach space L^1 converges weakly to an element of L^1 if and only if $\lim_\alpha \underline{\lim}_\beta |\varphi_\beta(f_\alpha) - \varphi_\beta(f_0| = 0$ for every net $\{\varphi_\beta, \beta \in \mathscr{B}\}$ in Φ.*

A related theorem for sequences can be found in reference [12; page 89].

BIBLIOGRAPHY

1. L. Alaoglu, *Weak topologies of normed linear spaces,* Ann. Math **41** (1940), 252-267.
2. Richard Arens, *Duality in linear spaces,* Duke Math. J., **14** (1947), 787-794.
3. C. Arzelà, *Intorno alla continuità della somma d'infinità di funzioni continue* Rend. Dell'Accad. di Bologna (1883-1884), 79-84.
4. ———, *Sulle funzioni di linee,* Mem. dell'Accad. di Bologna Ser. V, **5** (1885), 225-244.
5. S. Banach, *Théorie des opérations linéaires,* Warsaw, 1932.
6. R. G. Bartle, *On compactness in functional analysis,* Trans. Amer. Math. Soc. **79** (1955), 35-37.
7. N. Bourbaki, *Espaces vectoriels topologique,* Chap. I-II, Actualités Scientifiques et Industrielles, no. 1189, Paris, Hermann, 1953.
8. ———, *Espaces vectoriels topologique,* Chap. III. IV, V, Actualites Scientifiques et Industrielles, no. 1229, Paris, Hermann, 1955.
9. J. W. Brace, *Almost uniform convergence,* Portugalie Mathematica, **14** (1955), 99-104.
10. M. M. Day, *A property of Banach spaces,* Duke Math. J., **8** (1941), 763-770.
11. J. A. Dieudonné, *La dualité dans les espaces vectoriels topologiques,* Annales Scientifiques de L'Ecole Normale Supérieur, **59** (1942) 107-139.
12. ———, *Sur les espaces de Köthe,* Journal d'Analyse Mathématique, **1** (1951), 81-115.
13. H. H. Goldstine, *Weakly complete Banach spaces,* Duke Math. J., **4** (1938), 125-131.
14. A. Grothendieck, *Sur les applications linéaires faiblement compactes d'espace du type* $C(K)$, Can J. Math., **5** (1953), 129-173.
15. D. H. Hyers, *Linear topological spaces,* Bull. Amer. Math. Soc., **51** (1945), 1-21.
16. Samuel Kaplan, *Cartesian products of reals,* Amer. J. Math., **74** (1952), 936-954.
17. J. L. Kelley, *General topology,* Van Nostrand Co., New York, 1955.
18. G. W. Mackey, *On infinite dimensional linear spaces,* Trans. Amer. Math. Soc., **57** (1945), 155-207.
19. ———, *On convex topological linear spaces,* Trans. Amer. Math. Soc., **60** (1946), 519-537.
20. J. von Neumann, *On complete topological spaces,* Trans. Amer. Math. Soc., **37** (1935), 1-20.
21. G. Sirvint, *Weak compactness in Banach spaces,* Studia Mathematica, **6** (1957), 71-94.
22. J. V. Wehausen, *Transformations in linear topological spaces,* Duke., Math J., **4** (1938), 157-169.

UNIVERSITY OF MARYLAND

CHAINABLE CONTINUA AND INDECOMPOSABILITY

C. E. BURGESS

This paper includes a study of continua[1] which are both linearly chainable and circularly chainable. Since there exist indecomposable continua and 2 indecomposable continua which are linearly chainable, it follows from Theorem 7 that there exist indecomposable continua and decomposable continua which have both of these types of chainability.

A *linear chain* C is a finite collection of open sets $L_1, L_2, \cdots, L_n$ such that

(1) each element of C contains an open set that does not intersect any other element of C,

(2) $\rho(L_i, L_j) > 0$ if $|i - j| > 1$, and

(3) $L_i \cdot L_j \neq 0$ if $|i - j| \leq 1$. If this is modified so that $L_1 \cdot L_n \neq 0$, then C is called a *circular chain*. Each of the sets $L_1, L_2, \cdots, L_n$ is called a *link* of C, and C is sometimes denoted by $(L_1, L_2, \cdots, L_n)$ or $C(L_1, L_2, \cdots, L_n)$. If ε is a positive number and C is a linear chain such that each link of C has a diameter less than ε, then C is called a *linear* ε-chain. A *circular ε-chain* is defined similarly.

If C is either a linear chain or a circular chain and $H_1, H_2, \cdots, H_n$ are connected sets covered by C, then these sets are said to have the *order* $H_1, H_2, \cdots, H_n$ in C provided (1) no link of C intersects two of these n sets and (2) for each $i (i < n)$, there is a linear sub-chain in C which covers $H_i + H_{i+1}$ and which does not intersect any other of the sets $H_1, H_2, \cdots, H_n$.

A continuum M is said to be *linearly chainable*[2] if for every positive number ε, there is a linear ε-chain covering M. A continuum M is said to be *circularly chainable* if for every positive number ε, there is a circular ε-chain covering M.

A *tree* T is a finite coherent[3] collection of open sets such that

(1) each element of T contains an open set that does not intersect any other element of T,

(2) each two nonintersecting elements of T are a positive distance apart, and

(3) no subcollection of T consisting of more than two elements is a circular chain. If ε is a positive number and T is a tree such that

Presented to the American Mathematical Society, August 29, 1957; received by the editors December 5, 1958. This work was supported by the National Science Foundation under G-2574 and G-5880. Most of these results were obtained while the author was a visiting lecturer at the University of Wisconsin.

[1] Throughout this paper, a connected compact metric space is called a continuum.

[2] In some places in the literature, such continua have been said to be *chainable*.

[3] A collection G of sets is said to be *coherent* if for any two subcollections G_1 and G_2 of G such that $G_1 + G_2 = G$, some element of G_1 intersects some element of G_2.

each element of T has a diameter less than ε, then T is called an *ε-tree*. A continuum M is said to be *tree-like* if for every positive number ε, there is an ε-tree covering M.

A continuum M is said to be the *essential sum* of the elements of a collection G if the sum of the elements of G is M and no element of G is a subset of the sum of the other elements of G. If n is a positive integer and the continuum M is the essential sum of n continua and is not the essential sum of $n+1$ continua, then M is said to be *n-indecomposable*.[4]

A continuum M is said to be *unicoherent* if the intersection of each two continua having M as their sum is a continuum. A continuum M is said to be *bicoherent* if for any two proper subcontinua M_1 and M_2 having M as their sum, the set $M_1 \cdot M_2$ is the sum of two continua that do not intersect.

A continuum M is said to be a *triod* if M is the essential sum of three continua such that their intersection is a continuum which is the intersection of each two of them.

THEOREM 1. *If the continuum M is either linearly chainable or circularly chainable, then M does not contain a triod.*[5]

Proof. Since it is easy to see that every proper subcontinuum of M is linearly chainable, it will be sufficient to show that M is not a triod.

Suppose that M is a triod. Let M_1, M_2, and M_3 be three continua having M as their essential sum such that their intersection is a continuum H which is the intersection of each two of them. For each i $(i \leq 3)$, let p_i be a point of M_i that is not in either of the other two of the continua M_1, M_2, and M_3. Let ε be a positive number which is less than each of the numbers $\rho(p_1, M_2 + M_3)$, $\rho(p_2, M_1 + M_3)$, and $\rho(p_3, M_1 + M_2)$. Let C be either a linear ε-chain or a circular ε-chain which covers M. Since no link of C intersects two of the sets p_1, p_2, p_3, and H, consider the case in which these four sets are in C in the order named. This would involve the contradiction that M_2 intersects either the link of C that contains p_1 or the link of C that contains p_3. A similar contradiction results from supposing any other order of the sets p_1, p_2, p_3, and H in C.

THEOREM 2. *If the unicoherent continuum M is not a triod and M_1, M_2, M_3 are three continua having M as their essential sum, then*

---

[4] For any such continuum M, there is a unique collection consisting of n indecomposable continua having M as their essential sum [**4**].

[5] Bing [**2**] has used the fact that no linearly chainable continuum contains a triod, but for completeness a proof is given here for both types of chainability.

some two of these continua do not intersect and the other one intersects each of these two in a continuum.

Proof. Suppose that each two of the continua M_1, M_2, and M_3 intersect. It follows from the unicoherence of M that each of the sets $M_1 \cdot (M_2 + M_3)$ and $M_2 \cdot (M_1 + M_3)$ is a continuum and their sum is a continuum. Let $N = M_1 \cdot (M_2 + M_3) + M_2 \cdot (M_1 + M_3) = M_1 \cdot M_2 + M_1 \cdot M_3 + M_2 \cdot M_3$. Hence M is the essential sum of the three continua $M_1 + N$, $M_2 + N$, and $M_3 + N$ such that N is the intersection of each two of them and the intersection of all three of them. Since this is contrary to the hypothesis that M is not a triod, it follows that some two of the continua M_1, M_2, and M_3 do not intersect. Consider the case in which M_1 and M_3 do not intersect. Then M_2 intersects both M_1 and M_3, and since $M_1 \cdot M_2 = M_1 \cdot (M_2 + M_3)$ and $M_3 \cdot M_2 = M_3 \cdot (M_2 + M_1)$, it follows from the unicoherence of M that each of the sets $M_1 \cdot M_2$ and $M_3 \cdot M_2$ is a continuum.

THEOREM 3. *If the unicoherent continuum M is circularly chainable, then M is either indecomposable or 2-indecomposable.*

Proof. Suppose that M is the essential sum of three continua M_1, M_2, and M_3. By Theorem 1, M is not a triod. Hence by Theorem 2, one of these three continua, say M_2, intersects each of the other two such that $M_1 \cdot M_2$ and $M_2 \cdot M_3$ are continua and M_1 does not intersect M_3. For each i $(i \leq 3)$, let p_i be a point of M_i which is not in either of the other two of the continua M_1, M_2, and M_3. Let ε be a positive number which is less than each of the numbers $\rho(p_1, M_2 + M_3)$, $\rho(p_2, M_1 + M_3)$, $\rho(p_3, M_1 + M_2)$, and $\rho(M_1, M_3)$. Let C be a circular ε-chain which covers M. A contradiction can be obtained as follows for each of the three types of order in C for the five sets p_1, p_2, p_3, $M_2 \cdot M_1$, and $M_2 \cdot M_3$.

Case 1. If these five sets have the order $p_i, p_j, p_k, M_2 \cdot M_1, M_2 \cdot M_3$ in C, then M_j would intersect a link of C that contains one of the points p_i and p_k, contrary to the choice of ε.

Case 2. If these five sets have the order $p_1, M_2 \cdot M_1, p_i, p_j, M_2 \cdot M_3$ in C, then M_2 would intersect a link of C that contains one of the points p_1 and p_3, contrary to the choice of ε.

Case 3. If these five sets have the order $p_2, M_2 \cdot M_1, p_i, p_j, M_2 \cdot M_3$ in C, then each link of one of the linear chains of C from p_1 to p_3 would lie in $M_1 + M_3$. This would involve the contradiction that some link of C intersects both M_1 and M_3.

THEOREM 4. *If the circularly chainable continuum M is separated* *by one of its subcontinua, then M is linearly chainable.*

Proof. Let K be a subcontinuum of M which separates M. Then M is the sum of two continua M_1 and M_2 such that K is their intersection. Let p_1 and p_2 be points of $M_1 - K$ and $M_2 - K$, respectively, let ε be a positive number less than each of the numbers $\rho(p_1, M_2)$ and $\rho(p_2, M_1)$, and let C be a circular ε-chain covering M. Then each link of one of the linear chains in C from p_1 to p_2 is a subset of $M - K$. Let $L_1, L_2, \cdots, L_n$ be the links of C such that L_1 contains p_1 and there is a positive integer r such that L_r contains p_2 and no link of the linear chain $(L_1, L_2, \cdots, L_r)$ intersects K. There exist integers i and j such that L_i is the first link of $(L_1, L_2, \cdots, L_r)$ which intersects M_2 and L_j is the last link of $(L_1, L_2, \cdots, L_r)$ which intersects M_1. Then $(M_2 \cdot L_i, M_2 \cdot L_{i+1}, \cdots, M_2 \cdot L_r, L_{r+1}, \cdots, L_n, M_1 \cdot L_1, M_1 \cdot L_2, \cdots, M_1 \cdot L_j)$ is a linear ε-chain covering M.

THEOREM 5. *Every circularly chainable continuum M is either unicoherent or bicoherent. Furthermore, M is unicoherent provided some subcontinuum of M separates M, and M is bicoherent provided no subcontinuum of M separates M.*

Proof. Suppose that M is the sum of two continua H and K such that $H \cdot K$ is the sum of three mutually separated sets Y_1, Y_2, and Y_3. There exist three open sets D_1, D_2, and D_3 containing Y_1, Y_2, and Y_3, respectively, such that the closures of D_1, D_2, and D_3 are disjoint. For each i ($i \leq 3$), there exists a subcontinuum K_i of K irreducible from Y_i to $M - D_i$. The continuum $H + K_1 + K_2 + K_3$ is a triod, and this is contrary to Theorem 1. Hence it follows that if M_1 and M_2 are two continua having M as their sum, then the set $M_1 \cdot M_2$ is either a continuum or the sum of two continua.

It follows from Theorem 4 that M is linearly chainable, and hence unicoherent [3], provided some subcontinuum of M separates M. From this and the argument in the previous paragraph, it follows that M is bicoherent provided no subcontinuum of M separates M.

THEOREM 6. *If the circularly chainable continuum M is irreducible about some finite set consisting of n points, then there is a positive integer k not greater than n such that M is k-indecomposable.*

Proof. By Theorem 5, M is either unicoherent or bicoherent. If M is unicoherent, it follows from Theorem 3 that M is either indecomposable or 2-indecomposable. If M is bicoherent, it follows from Corollary 6.1 of [5] that there is a positive integer k not greater than n such that M is k-indecomposable.

THEOREM 7. *If the continuum M is linearly chainable, then in order that M should be circularly chainable, it is necessary and sufficient that M be either indecomposable of 2-indecomposable.*

Proof of necessity. Since every lineary chainable continuum is unicoherent [3], it follows from Theorem 3 that M is either indecomposable or 2-indecomposable.

Proof of sufficiency. The case where M is indecomposable and the case where M is 2-indecomposable will be considered separately.

Case 1. Suppose M is indecomposable, and let $C(L_1, L_2, \cdots, L_n)$ be a linear ε-chain covering M. There exist two disjoint continua K_1 and K_2 of M such that each of them intersects each of the sets $L_1 - cl(L_2)$ and $L_n - cl(L_{n-1})$. If follows that there exist a positive number ε', a linear ε'-chain C' covering M, and two subchains C_1 and C_2 of C' such that

(1) each link of C' is a subset of some link of C,

(2) C_1 and C_2 have no common link, and

(3) each of the chains C_1 and C_2 has one end link in $L_1 - cl(L_2)$ and the other end link in $L_n - cl(L_{n-1})$. Let W_1 denote the set of all points of M that are covered by C_1 and let W_2 denote $M - W_1$. Then $(L_1, W_1 \cdot L_2, W_1 \cdot L_3, \cdots, W_1 \cdot L_{n-1}, L_n, W_2 \cdot L_{n-1}, W_2 \cdot L_{n-2}, \cdots, W_2 \cdot L_2)$ is a circular ε-chain covering M.

Case 2. If M is 2-indecomposable, there exist two indecomposable continua M_1 and M_2 such that M is their essential sum and $M_1 \cdot M_2$ is a continuum. Let ε be a positive number. There exists a linear ε-chain C covering M such that M_1 intersects $L_1 - cl(L_2)$ and M_2 intersects $L_n - cl(L_{n-1})$. Since each composant of $M_i (i = 1, 2)$ is everywhere dense in M_i, it follows that for each i $(i = 1, 2)$ there exist two disjoint subcontinua K_i and H_i of M_i such that

(1) each of them intersects each link of C that intersects M_i,

(2) H_i contains $M_1 \cdot M_2$,

(3) each of the continua H_1 and K_1 intersects $L_1 - cl(L_2)$, and

(4) each of the continua H_2 and H_2 intersects $L_n - cl(L_{n-1})$. Hence there exist a positive number ε', a linear ε'-chain C' covering M, and three subchains C_1, C_2, and C_3 of C' such that

(1) each link of C' is a subset of a link of C,

(2) no two of the chains C_1, C_2, and C_3 have a common link,

(3) one end link of C_1 is in $L_1 - cl(L_2)$,

(4) one end link of C_2 is in $L_n - cl(L_{n-1})$,

(5) some link of C contains a link of C_1 and a link of C_2, and

(6) C_3 has one end link in $L_1 - cl(L_2)$ and the other end link in $L_n - cl(L_{n-1})$. Let W denote the set of all points of M that are covered by C_3, and let Y denote $M - W$. Then $(L_1, W \cdot L_2, W \cdot L_3, \cdots, W \cdot L_{n-1}, L_n, Y \cdot L_{n-1}, Y \cdot L_{n-2}, \cdots, Y \cdot L_2)$ is a circular ε-chain covering M.

THEOREM 8. *If n is a positive integer and for each proper subcontinuum H of the continuum M there is a positive integer r not greater than n such that H is r-indecomposable, then there is a positive integer k not greater than n such that M is k-indecomposable.*

Proof. Suppose that M is the essential sum of $n + 1$ continua $M_1, M_2, \cdots, M_{n+1}$. Some n of these continua have a connected sum, so consider the case in which $M_2 + M_3 \cdots + M_{n+1}$ is connected. There is an open set D which intersects M_1 such that the closure of D does not intersect $M_2 + M_3 + \cdots + M_{n+1}$. There is a subcontinuum M_1' of M_1 irreducible from the closure of D to $M_2 + M_3 + \cdots + M_{n+1}$. This involves the contradiction that $M_1' + M_2 + M_3 + \cdots + M_{n+1}$ is a proper subcontinuum of M and is the essential sum of $n + 1$ continua.

THEOREM 9. *If every proper subcontinuum of the continuum M is circularly chainable, then every subcontinuum of M is either indecomposable or 2-indecomposable.*

Proof. Since each proper subcontinuum of M is a proper subcontinuum of another proper subcontinuum of M, it follows that every proper subcontinuum of M is linearly chainable. Hence by Theorem 7, every proper subcontinuum of M is either indecomposable or 2-indecomposable. Consequently, it follows from Theorem 8 that M itself is either indecomposable or 2-indecomposable.

EXAMPLES. A pseudo-arc [1; 6] is an example of an indecomposable continuum which satisfies the hypothesis of Theorem 9, and a continuum which is the sum of two pseudo-arcs with a point as their intersection is an example of a 2-indecomposable continuum which satisfies this hypothesis.

THEOREM 10. *If the tree-like continuum M is circularly chainable, then M is linearly chainable.*

Pooof. Let ε be a positive number, and let $C(L_1, L_2, \cdots, L_n)$ be a circular $\varepsilon/3$-chain covering M. Then M is covered by a tree T such that

(1) each element of T is a subset of a link of C,

(2) some element K_0 of T intersects only one element of C, and

(3) no element of T intersects three elements of C. A function f will be defined as follows over T. For each element K of T, there is only one linear chain $(K_0, K_1, \cdots, K_m = K)$ from K_0 to K in T. Let $f(K_0) = 0$, and suppose that for some integer i $(0 \leq i \leq m)$, $f(K_i)$ has been defined. Then define $f(K_{i+1})$ as follows:

(1) let $f(K_{i+1}) = f(K_i) + 1$ provided K_i lies in some element L_j of C and K_{i+1} intersects $L_{j+1,\text{mod}\,n}$ but K_i does not intersect this set,

(2) Let $f(K_{i+1}) = f(K_i) - 1$ provided K_{i+1} lies in some element L_j of C and K_i intersects $L_{j+1,\text{mod}\,n} - L_j$ but K_{i+1} does not intersect this set, and

(3) let $f(K_{i+1}) = f(K_i)$ provided neither (1) nor (2) is satisfied. The range of f is an increasing finite sequence of consecutive integers $n_1, n_2, \cdots, n_r$. For each t $(1 \leq t \leq r)$, let M_t denote the sum of all elements X of T such that $f(X) = n_t$. Then $(M_1, M_2, \cdots, M_r)$ is a linear ε-chain covering M.

REFERENCES

1. R. H. Bing, *Concerning hereditarily indecomposable continua*, Pacific, J. Math. **1** (1951), 43-51.

2. ———, *Snake-like continua*, Duke Math. J. **18** (1951), 653-663.

3. ———, and F. B. Jones, *Another homogeneous plane continuum* Trans. Amer. Math. Soc. **90** (1959), 171-192.

4. C. E. Burgess, *Continua which are the sum of a finite number of indecomposable continua*, Proc. Amer. Math. Soc. **4** (1953), 234-239.

5. ———, *Separation properties and n-indecomposable continua*, Duke Math. J. **24** (1956), 595-600.

6. E. E. Moise, *An indecomposable plane continuum which is homeomorphic to each of its nondegenerate subcontinua*, Trans. Amer. Math. Soc. **63** (1948), 581-594.

UNIVERSITY OF UTAH

MULTIPLICATION FORMULAS FOR PRODUCTS OF BERNOULLI AND EULER POLYNOMIALS

L. CARLITZ

1. Put

$$(1.1)\qquad \frac{te^{xt}}{e^t - 1} = \sum_{n=0}^{\infty} B_n(x)\frac{t^n}{n!},\ \frac{2e^{xt}}{e^t + 1} = \sum_{n=0}^{\infty} E_n(x)\frac{t^n}{n!}\ .$$

The following multiplication formulas are familiar [5, pp. 18, 24]:

$$(1.2)\qquad B_m(kx) = k^{m-1}\sum_{r=0}^{k-1} B_m\left(x + \frac{k}{r}\right),$$

$$(1.3)\qquad E_m(kx) = k^m \sum_{r=0}^{k-1} (-1)^r E_m\left(x + \frac{k}{r}\right) \qquad (k \text{ odd})\ .$$

Let $\overline{B}_m(x)$, $\overline{E}_m(x)$ denote, respectively, the Bernoulli and Euler functions defined by

$$\overline{B}_m(x) = B_m(x)(0 \le x < 1),\ \overline{B}_m(x + 1) = \overline{B}_m(x)\ ,$$
$$\overline{E}_m(x) = E_m(x)(0 \le x < 1),\ \overline{E}_m(x + 1) = -\overline{E}_m(x)\ , \quad (m \ge 1)\ .$$

Then $\overline{B}_m(x)$ and $\overline{E}_m(x)$ also satisfy the multiplication formulas (1.2), (1.3).

In this note we obtain some generalizations of (1.2) and (1.3) suggested by a recent result of Mordell [4]. In extending some results of Mikolás [3], Mordell proves the following theorem. Let $f_1(x), \cdots, f_n(x)$ denote functions of x of period 1 that satisfy the relations

$$(1.4)\qquad \sum_{r=0}^{k-1} f_i\left(r + \frac{r}{k}\right) = C_i^{(k)} f_i(kx) \qquad (i = 1, \cdots, n)\ ,$$

where $C_i^{(k)}$ is independent of x. Let $a_1, \cdots, a_n$ be positive integers that are relatively prime in pairs. Then if the integrals exist and $A = a_1 a_2 \cdots a_n$,

$$(1.5)\qquad \int_0^A f_1\left(\frac{x}{a_1}\right) f_2\left(\frac{x}{a_2}\right) \cdots f_n\left(\frac{x}{a_n}\right) dx$$
$$= A\int_0^1 f_1\left(\frac{Ax}{a_1}\right) f_2\left(\frac{Ax}{a_2}\right) \cdots f_n\left(\frac{Ax}{a_n}\right) dx$$
$$= C_1^{(a_1)} C_2^{(a_2)} \cdots C_n^{(a_n)} \int_0^1 f_1(x) f_2(x) \cdots f_n(x) dx\ .$$

Received November 24, 1958.

2. We first prove

THEOREM 1. *Let* $n \geq 1$; $m_1, \cdots, m_n \geq 1$; $a_1, a_2, \cdots, a_n$ *positive integers that are relative prime in pairs*; $A = a_1, a_2, \cdots, a_n$. *Then*

$$\sum_{r=0}^{kA-1} \overline{B}_{m_1}\left(x_1 + \frac{r}{a_1 k}\right)\overline{B}_{m_2}\left(x_2 + \frac{r}{a_2 k}\right) \cdots \overline{B}_{m_n}\left(x_n + \frac{r}{a_n k}\right) \tag{2.1}$$
$$= C \sum_{r=0}^{k-1} \overline{B}_{m_1}\left(a_1 x_1 + \frac{r}{k}\right)\overline{B}_{m_2}\left(a_2 x_2 + \frac{r}{k}\right) \cdots \overline{B}_{m_n}\left(a_n x_n + \frac{r}{k}\right),$$

where

$$C = a_1^{1-m_1} a_2^{1-m_2} \cdots a_n^{1-m_n}. \tag{2.2}$$

In the first place for $n = 1$ it follows from (1.2) for arbitrary $a \geq 1$ that

$$\sum_{r=0}^{ka-1} \overline{B}_m\left(x + \frac{r}{ak}\right) = \sum_{r=0}^{k-1}\sum_{s=0}^{a-1} \overline{B}_m\left(r + \frac{s}{a} + \frac{r}{ak}\right)$$
$$= \sum_{r=0}^{k-1} \overline{B}_m\left(ax + \frac{r}{k}\right),$$

which agrees with (2.1).

For the general case, let S denote the left member of (2.1). Put

$$A_s = a_1 a_2 \cdots a_s \qquad (1 \leq s \leq n)$$

and replace r by $skA_{n-1} + r$. Then

$$S = \sum_{r=0}^{kA_{n-1}-1} \overline{B}_{m_1}\left(x_1 + \frac{r}{a_1 k}\right) \cdots \overline{B}_{m_{n-1}}\left(x_{n-1} + \frac{r}{a_{n-1}k}\right)$$
$$\cdot \sum_{s=0}^{a_n-1} \overline{B}_{m_n}\left(x_n + \frac{A_{n-1}s}{a_n} + \frac{r}{a_n k}\right)$$
$$= \sum_{r=0}^{kA_{n-1}-1} \overline{B}_{m_1}\left(x_1 + \frac{r}{a_1 k}\right) \cdots \overline{B}_{m_{n-1}}\left(x_{n-1} + \frac{r}{a_{n-1}k}\right)$$
$$\cdot \sum_{s=0}^{a_n-1} \overline{B}_{m_n}\left(x_n + \frac{s}{a_n} + \frac{r}{a_n k}\right)$$
$$= a_n^{1-m_n} \sum_{r=0}^{kA_{n-1}-1} \overline{B}_{m_1}\left(x_1 + \frac{r}{a_1 k}\right) \cdots \overline{B}_{m_{n-1}}\left(x_{n-1} + \frac{r}{a_{n-1}k}\right)$$
$$\cdot \overline{B}_{m_n}\left(a_n x_n + \frac{r}{k}\right).$$

Continuing in this way we get

$$S = a_{n-1}^{1-m_{n-1}} a_n^{1-m_n} \sum_{r=0}^{kA_{n-2}-1} \overline{B}_{m_1}\left(x_1 + \frac{r}{a_1 k}\right) \cdots \overline{B}_{m_{n-2}}\left(x_{n-2} + \frac{r}{a_{n-2}k}\right)$$

$$\cdot \overline{B}_{m_{n-1}}\left(a_{n-1}x_{n-1} + \frac{r}{k}\right)\overline{B}_{m_n}\left(a_n x_n + \frac{r}{k}\right)$$

$$= a_1^{1-m_1} \cdots a_n^{1-m_n} \sum_{r=0}^{k-1} \overline{B}_{m_1}\left(a_1 x_1 + \frac{r}{k}\right)\overline{B}_2\left(a_2 x_2 + \frac{r}{k}\right)$$

$$\cdots \overline{B}_{m_n}\left(a_n x_n + \frac{r}{k}\right) .$$

For $k = 1$, (2,1) reduces to

$$\sum_{r=0}^{A-1} \overline{B}_{m_1}\left(x_1 + \frac{r}{a_1}\right)\overline{B}_2\left(x_2 + \frac{r}{a_2}\right) \cdots \overline{B}_n\left(x_n + \frac{r}{a_n}\right) \tag{2.3}$$

$$= C \cdot \overline{B}_{m_1}(a_1 x_1)\overline{B}_{m_2}(a_2 x_2) \cdots \overline{B}_{m_n}(a_n x_n) ,$$

where C is defined by (2.2); (2.3) may be considered a direct generalization of (1.2).

We remark that a formula like (2.1) holds for any set of functions satisfying (1.4).

We note also that the formula (2.2) can be proved by means of the Chinese remainder theorem. This remarks applies also to formulas (3.4) and (4.8) below.

3. In the next place we have

THEOREM 2. *Let n be odd and ≥ 1; $m_1, \cdots, m_n \geq 1$; $a_1, a_2, \cdots, a_n$ positive odd integers that are relatively prime in pairs; $A = a_1 a_2 \cdots a_n$; k odd ≥ 1. Then*

$$\sum_{r=0}^{kA-1} (-1)^r \overline{E}_{m_1}\left(x_1 + \frac{r}{a_1 k}\right) \cdots \overline{E}_{m_n}\left(x_n + \frac{r}{a_n k}\right) \tag{3.1}$$

$$= C' \sum_{r=0}^{k-1} (-1)^r \overline{E}_{m_1}\left(a_1 x_1 + \frac{r}{k}\right) \cdots \overline{E}_{m_n}\left(a_n x_n + \frac{r}{k}\right) ,$$

where

$$C' = a_1^{-m_1} a_2^{-m_2} \cdots a_n^{-m_n} . \tag{3.2}$$

The proof is similar to that of Theorem 1, but makes use of (1.3) in place of (1.2); also the formula

$$\overline{E}_m(x + r) = (-1)^r \overline{E}_m(x) \qquad (m \geq 1) \tag{3.3}$$

is needed.

For $n = 1$ and a odd, we have

$$\sum_{r=0}^{ka-1}(-1)^r\overline{E}_{m_1}\left(x + \frac{r}{ak}\right) = \sum_{r=0}^{k-1}(-1)^{sk}\overline{E}_m\left(x + \frac{s}{a} + \frac{r}{ak}\right)$$
$$= a^{-m}\sum_{r=0}^{k-1}(-1)^r\overline{E}_m\left(ax + \frac{r}{k}\right),$$

which agrees with (3.1). For the general case let S' denote the left member of (3.1). Then

$$S' = \sum_{r=0}^{kA_{n-1}-1}\sum_{s=0}^{a_n-1}(-1)^{r+s}\overline{E}_{m_1}\left(x_1 + \frac{sA_{n-1}}{a_1} + \frac{r}{a_1k}\right)\cdots$$
$$\cdot\ \overline{E}_{m_{n-1}}\left(x_{n-1} + \frac{sA_{n-1}}{a_{n-1}} + \frac{r}{a_{n-1}k}\right)$$
$$\cdot\ \overline{E}_{m_n}\left(x_n + \frac{sA_{n-1}}{a_n} + \frac{r}{a_nk}\right).$$

If we put

$$sA_{n-1} = qa_n + t \qquad (0 \le t < a_n),$$

then $s \equiv q + t \pmod 2$, so that

$$\overline{E}_{m_n}\left(x_n + \frac{sA_{n-1}}{a_n} + \frac{r}{a_nk}\right) = (-1)^q\overline{E}_{m_n}\left(x_n + \frac{t}{a_n} + \frac{r}{a_nk}\right).$$

Since n is odd we therefore get

$$S' = \sum_{r=0}^{kA_{n-1}-1}(-1)^r\overline{E}_{m_1}\left(x_1 + \frac{r}{a_1k}\right)\cdots\overline{E}_{m_{n-1}}\left(x_{n-1} + \frac{r}{a_{n-1}k}\right)$$
$$\cdot\ \sum_{t=0}^{a_n-1}(-1)^t\overline{E}_{m_n}\left(x_n + \frac{t}{a_n} + \frac{r}{a_nk}\right)$$
$$= \sum_{r=0}^{kA_{n-1}-1}(-1)^r\overline{E}_{m_1}\left(x_1 + \frac{r}{a_1k}\right)\cdots\overline{E}_{m_{n-1}}\left(x_{n-1} + \frac{r}{a_{n-1}k}\right)$$
$$\cdot\ a_n^{-m_n}\,\overline{E}_{m_n}\left(a_nx_n + \frac{r}{k}\right).$$

Continuing in this way we ultimately reach (3.1).

For $k = 1$, (3.1) becomes

$$\sum_{r=0}^{A-1}(-1)^r\overline{E}_{m_1}\left(x_1 + \frac{r}{a_1}\right)\cdots\overline{E}_{m_n}\left(x_n + \frac{r}{a_n}\right) \tag{3.4}$$
$$= C'E_{m_1}(a_1x_1)\cdots E_{m_n}(a_nx_n),$$

subject to the conditions of the theorem.

4. **Theorem 2 can be extended further** by introducing the "Eulerian" polynomial [2] $\phi_m(x, \rho)$ defined by

$$\frac{1-\rho}{1-\rho e^t} e^{xt} = \sum_{m=0}^{\infty} \phi_m(x, \rho) \frac{t^m}{m!} \qquad (\rho \neq 1) . \tag{4.1}$$

In particular $\phi_m(x, -1) = E_m(x)$.

We shall assume that the parameter ρ is an fth root of unity. It follows easily from (4.1) that

$$\phi_{m-1}(kx, \rho) = \frac{(\rho-1)f^{m-1}}{m} \sum_{r=0}^{f-1} \rho^r B_m\left(x + \frac{r}{f}\right) . \tag{4.2}$$

We accordingly define the function $\overline{\phi}_n(x, \rho)$ *by means of*

$$\overline{\phi}_{m-1}(kx, \rho) = \frac{(\rho-1)e^{m-1}}{m} \sum_{r=0}^{f-1} \rho^r \overline{B}_m\left(x + \frac{r}{f}\right) . \tag{4.3}$$

It follows from (4.3) that

$$\overline{\phi}_n(x+1, \rho) = \rho^{-1}\overline{\phi}_n(x, \rho) , \tag{4.4}$$

so that if ρ is a primitive fth root of unity, $\overline{\phi}_n(x, \rho)$ has period f. Also by means of (4.1) we readily obtain the multiplication theorem [1] valid for $k \equiv 1 \pmod{f}$

$$\sum_{r=0}^{k-1} \rho^r \phi_m\left(x + \frac{r}{k}, \rho\right) = k^{-m}\phi_m(kx, \rho) \tag{4.5}$$

and consequently

$$\sum_{r=0}^{k-1} \rho^r \overline{\phi}_m\left(x + \frac{r}{k}, \rho\right) = k^{-m}\overline{\phi}_m(kx, \rho) . \tag{4.6}$$

We may now state

THEOREM 3. *Let $f > 1$, $n \equiv 1 \pmod{f}$; $m_1, \cdots, m_n \geq 1$, $a_1, a_2, \cdots, a_n$ positive integers that are relatively prime in pairs and such that $a_i \equiv 1 \pmod{f}$ for $i = 1, \cdots, n$; also let $k \equiv 1 \pmod{f}$. Then if $A = a_1 a_2 \cdots a_n$, we have*

$$\sum_{r=0}^{kA-1} \rho^r \overline{\phi}_{m_1}\left(x_1 + \frac{r}{a_1 k}, \rho\right) \cdots \overline{\phi}_{m_n}\left(x_n + \frac{r}{a_n k}, \rho\right) \tag{4.7}$$

$$= C' \sum_{r=0}^{k-1} \rho^r \overline{\phi}_{m_1}\left(a_1 x_1 + \frac{r}{k}, \rho\right) \cdots \overline{\phi}_{m_n}\left(a_n x_n + \frac{r}{k}, \rho\right),$$

where C' is defined by (3.2).

The proof is very much like that of Theorem 2 and will be omitted. We remark that for $k = 1$, (4.7) becomes

$$\sum_{r=0}^{A-1} \rho^r \overline{\phi}_{m_1}\left(x_1 + \frac{r}{a_1}, \rho\right) \cdots \overline{\phi}_{m_n}\left(x_n + \frac{r}{a_n}, \rho\right) \tag{4.8}$$
$$= C' \overline{\phi}_{m_1}(a_1 x_1, \rho) \cdots \overline{\phi}_{m_n}(a_n x_n, \rho) .$$

REFERENCES

1. L. Carlitz, *The multiplication formulas for the Bernoulli and Euler polynomials'* Mathematics Magazine, **27** (1953), 59-64.
2. G. Frobenius, *Uber die Bernoulli'schen Zahlen und die Euler'schen Polynome*, Sitzungsberichte der Preussischen Akademie der Wissenschaften (1910), 809-847.
3. M. Mikolás, *Integral formulas of arithmetical characteristics relating to the zeta-function of Hurwitz*, Publicationes Mathematicae, **5** (1957), 44-53.
4. L. J. Mordell, *Integral formulas of arithmetical character*, Journal of the London Mathematical Society, **33** (1957), 371-375.
5. N. E. Nörlund, *Vorlesungen über Differenzenrechnung*, Berlin, 1924.

DUKE UNIVERSITY

A CLASS OF RESIDUE SYSTEMS (mod r) AND RELATED ARITHMETICAL FUNCTIONS, II. HIGHER DIMENSIONAL ANALOGUES

ECKFORD COHEN

1. Introduction. In an earlier paper [3] with a similar name (to be referred to as I) we introduced the idea of a direct factor set (P-set) and the residue system (mod n) associated with such a set. We first review briefly these concepts. Two non-vacuous subsets P, Q of the positive integers Z are said to form a conjugate pair of direct factor sets provided the following two conditions are satisfied:

(i) an integer $n > 0$ is in P (or Q) if and only if, for each factorization, $n = n_1 n_2$, $(n_1, n_2) = 1$, n_1 and n_2 are also in P (or Q),

(ii) every positive integer n possesses a unique factorization of the form, $n = ab$ such that $a \in P$, $b \in Q$. A set of integers $a (\mathrm{mod}\, n)$ such that $(a, n) \in P$ is said to form a P-reduced residue system (mod n), or P-system (mod n), and the number of elements in such a system is denoted by $\phi_P(n)$. The fundamental result of I was a generalization of the Möbius inversion formula to conjugate pairs of direct factor sets. This result is reformulated in § 2 of the present paper.

In this paper we extend the notion of a P-system (mod n) from the set of integers X to t-dimensional vectors over X (briefly, X_t-vectors), $t \geqq 1$. The one dimensional case ($t = 1$) is the case already investigated in I. Two X_t-vectors, $A = \{a_i\}$, $B = \{b_i\}$, are said to be congruent (mod t, n), written $A \equiv B (\mathrm{mod}\, t, n)$, provided $a_i \equiv b_i (\mathrm{mod}\, n)$, $i = 1, \cdots, t$. Moreover, we place $(a_i) = (a_1, \cdots, a_t)$, using the convention, $(0, \cdots, 0) = 0$, and define vector sums and scalar multiples in the usual way. A P-reduced residue system (mod t, n), or P-system (mod t, n), is defined to be a maximal set of mutually incongruent X_t-vectors (mod t, n), $\{a_i\}$, satisfying $((a_i), n) \in P$. The number of elements in such a system depends only on t and n, and is denoted $J_{t,P}(n)$ and called the (t, P)-totient of n. In case P is the unit set 1, $J_{t,P}(n)$ reduces to the ordinary Jordan totient, $J_{t,1}(n) = J_t(n)$. A P-system with $P = Z$ is called a complete residue system (mod t, n); clearly $J_{t,Z}(n) = n^t$.

REMARK 1.1. An X_t-vector whose components are in Z will be called a Z_t-vector, and a P-system (mod t, n) consisting of elements of Z_t alone will be called a *positive* P-system (mod t, n).

We summarize now the salient points of the paper. In § 2 an enumerative principle for X_t-vectors (Theorem 2.1) is formulated, generalizing a result proved in [3, § 3] in the case $t = 1$. This result is used,

Received January 30, 1959.

in conjunction with the inversion principle of I, to obtain an evaluation of $J_{t,P}(n)$. A function $\phi_{\alpha,P}(n)$, formally generalizing $J_t(n)$, is also introduced, along with a generalized divisor function $\sigma_{\alpha,P}(n)$. Certain closely related functions, $\phi^*_{\alpha,P}(n)$ and $\sigma^*_{\alpha,P}(n)$ are also defined in § 2.

In § 3 we introduce the zeta function $\zeta_P(s)$ associated with a direct factor set P. In case $P = Z$, $\zeta_P(s)$ is the ordinary ζ-function, $\zeta(s)$. Employing the generalized inversion function $\mu_P(n)$ of I we also define "reciprocal" ζ-functions $\tilde{\zeta}_P(s)$ and obtain in (3.8) a generalization ($P = 1$, $Q = Z$) of the familiar fact,

$$\sum_{n=1}^{\infty} \frac{\mu(n)}{n^s} = \zeta^{-1}(s) , \qquad s > 1 , \tag{1.1}$$

where $\mu(n)$ denotes the Möbius function. Broad generalizations of other basic identities involving ζ-functions are also deduced.

In § 4 we obtain mean value estimates for the functions $\phi_{\alpha,P}(n)$ and $\sigma_{\alpha,P}(n)$, valid for *arbitrary* direct factor sets P, extending basic properties of $\phi(n)$ and $\sigma(n) = \sigma_{1,Z}(n)$. For example, (4.5) reduces in case $\alpha = 1$, $P = 1$, to the celebrated result [1, Theorem 330] of Mertens for the Euler ϕ-function,

$$\sum_{n \leqq x} \phi(n) = \frac{3x^2}{\pi^2} + O(x \log x) . \tag{1.2}$$

Using results of § 4, we obtain in § 5 (Theorem 5.1) for $t \geqq 2$, the asymptotic density of Z_t-vectors $\{a_i\}$, such that $(a_i) \in P$. Numerous special cases are considered (Corollary 5.2). We mention that Corollary 5.3, in case $t = 2$, yields a result of Kronecker asserting that the density of the integral pairs with a fixed greatest common divisor r is $6/\pi^2 r^2$.

In § 6 we generalize the so-called "second Möbius inversion formula" to conjugate sets P, Q (Theorem 6.1). Application of this extended inversion relation yields in (6.3) a generalization of broad scope of Meissel's well known identity,

$$\sum_{1 \leqq n \leqq x} \mu(n) \left[\frac{x}{n} \right] = 1 . \tag{1.3}$$

We also evaluate in § 6 a generalization to P-sets of Legendre's totient function $\phi(x, n)$, defined to be the number of integers a such that $1 \leqq a \leqq x$, $(a, n) = 1$.

REMARK 1.2. It is noted that many of the results of this paper are valid, not merely for direct factor sets, but for quite arbitrary sets of integers P. For example, this is true in the case of Corollary 5.1. Moreover, a number of the remaining results can be reformulated in such a manner as to be valid for arbitrary sets P. We shall restrict our attention, however, to direct factor sets, reserving the treatment of more general sets for a later paper, to be based on other methods. The advantage of a separate treatment of direct factor sets arises from the applicability of the generalized inversion theorem.

2. Generalized totient and divisor functions. Let P and Q denote an arbitrary conjugate pair of direct factor sets, and define, as in I,

$$\rho_P(n) = \begin{cases} 1 & (n \in P) \\ 0 & (n \notin P) , \end{cases} \tag{2.1}$$

$$\mu_P(n) = \sum_{d\delta=n} \rho_P(d)\mu(\delta) . \tag{2.2}$$

The functions $\rho_P(n)$ and $\mu_P(n)$ are termed, respectively the *characteristic* function and *inversion* function of the set P. The inversion formula of I can be restated in the form,

$$f(n) = \sum_{d\delta=n} \rho_Q(d)g(\delta) \rightleftarrows g(n) = \sum_{d\delta=n} \mu_P(d)f(\delta) . \tag{2.3}$$

This principle is a direct consequence of the relation,

$$\sum_{d\delta=n} \mu_P(d)\rho_Q(\delta) = \rho(n) , \tag{2.4}$$

where $\rho(n) = \rho_1(n)$ (that is, $\rho(n) = 1$ or 0 according as $n = 1$ or $n > 1$). Note that $\mu_P(n)$ reduces to $\mu(n)$ when $P = 1$.

In order to evaluate $J_{t,P}(n)$, we shall need the following results generalizing Theorem 4 of I to t dimensional vectors.

THEOREM 2.1. *If d ranges over the divisors of n contained in Q, and for each d, x ranges over the elements of a P-system* (mod t, δ), *$d\delta = n$, then the set dx constitutes a complete residue system* (mod t, n).

We omit the proof, which is analogous to the proof in case $t = 1$. On the basis of this result it follows immediately that

$$\sum_{d\delta=n} \rho_Q(d)J_{t,P}(\delta) = n^t . \tag{2.5}$$

Application of (2.3) to (2.5) yields

THEOREM 2.2.

$$J_{t,P}(n) = \sum_{d\delta=n} d^t\mu_P(\delta) . \tag{2.6}$$

Define now for α an arbitrary real number, the generalized totient,

$$\phi_{\alpha,P}(n) = \sum_{d\delta=n} d^\alpha\mu_P(\delta) , \tag{2.7}$$

so that $\phi_{\alpha,P} = J_{t,P}(n)$ in case $\alpha = t$ is a positive integer. We also define analogously a generalized divisor function by placing

$$\sigma_{\alpha,P}(n) = \sum_{d\delta=n} d^\alpha \rho_P(\delta) = \sum_{\substack{d\delta=n \\ \delta\in P}} d^\alpha . \tag{2.8}$$

Corresponding to the functions $\phi_{\alpha,P}(n)$, $\sigma_{\alpha,P}(n)$ we define related functions,

$$\phi^*_{\alpha,P}(n) = \sum_{d|n} d^\alpha \mu_P(d) \tag{2.9}$$

$$\sigma^*_{\alpha,P}(n) = \sum_{d|n} d^\alpha \rho_P(d) = \sum_{\substack{d|n \\ d\in P}} d^\alpha . \tag{2.10}$$

The following simple relations are noted.

$$\phi^*_{-\alpha,P}(n) = \frac{\phi_{\alpha,P}(n)}{n^\alpha} , \tag{2.11a}$$

$$\sigma^*_{-\alpha,P}(n) = \frac{\sigma_{\alpha,P}(n)}{n^\alpha} . \tag{2.11b}$$

Corresponding to the case $P = 1$, we place $\phi_{\alpha,1}(n) = \phi_\alpha(n)$, $\phi^*_{\alpha,1} = \phi^*_\alpha(n)$, and corresponding to the case $P = Z$, we write $\sigma_{\alpha,Z}(n) = \sigma_\alpha(n) = \sigma^*_{\alpha,Z}(n)$.

The following result is a generalization of [3, Theorem 8, $\alpha = 1$] and can be proved similarly.

THEOREM 2.3.

$$\phi_{\alpha,P}(n) = \sum_{d\delta=n} \phi_\alpha(d)\rho_P(\delta) . \tag{2.12}$$

We also note, by inversion of (2.7), the following generalization of (2.5).

$$\sum_{d\delta=n} \rho_Q(d)\phi_{\alpha,P}(\delta) = n^\alpha . \tag{2.13}$$

3. The zeta-functions of a P-set.

REMARK 3.1. In the definitions and general results of this section, s is assumed to be limited to values for which all occuring series converge absolutely.

First we define for real s,

$$\zeta_P(s) = \sum_{n=1}^{\infty} \frac{\rho_P(n)}{n^s} = \sum_{\substack{n=1 \\ n\in P}}^{\infty} \frac{1}{n^s} . \tag{3.1}$$

The function $\zeta_P(s)$ will be called the *zeta-function* of the direct factor set P. Note that $\zeta_Z(s) = \zeta(s)$, $\zeta_1(s) = 1$. We define the *reciprocal* zeta-function of P by

$$\tilde{\zeta}_P(s) = \sum_{n=1}^{\infty} \frac{\mu_P(n)}{n^s} ; \tag{3.2}$$

the function $\zeta_Q(s)$ will be designated the *conjugate* zeta-function of P.

By (1.1) it follows that $\tilde{\zeta}(s) \equiv \tilde{\zeta}_1(s) = 1/\zeta(s)$. We mention that Diricelet series of the form (3.1), (3.2) were discussed by Wintner [10, Chapter II] in case P is a semigroup generated by a set of primes.

First we prove two relations analogous to (2.4).

LEMMA 3.1.

$$\sum_{d\delta=n} \rho_P(d)\rho_Q(\delta) = 1 . \tag{3.3}$$

Proof. This is an immediate consequence of property (ii) of the conjugate pair P, Q.

LEMMA 3.2.

$$\sum_{d\delta=n} \mu_P(d)\mu_Q(\delta) = \mu(n) . \tag{3.4}$$

Proof. By the definition of $\mu_P(n)$, we have, with the left member of (3.4) denoted by $S(n)$,

$$\begin{aligned} S(n) &= \sum_{d\delta=n} \sum_{\substack{DD'=d \\ D'\in P}} \mu(D) \sum_{\substack{EE'=\delta \\ E'\in Q}} \mu(E) = \sum_{\substack{DD'EE'=n \\ D'\in P, E'\in Q}} \mu(D)\mu(E) \\ &= \sum_{DE|n} \mu(D)\mu(E) \sum_{\substack{D'E'=n/DE \\ D'\in P, E'\in Q}} 1 . \end{aligned}$$

By property (ii), it follows then that

$$S(n) = \sum_{DE|n} \mu(D)\mu(E) = \sum_{D|E} \mu(D) \sum_{E|(n/D)} \mu(E) ,$$

and (3.4) results by the fundamental property of $\mu(n)$, ((2.4) with $P = 1$, $Q = Z$).

The following relations are basic.

THEOREM 3.1.

$$\zeta_P(s)\zeta_Q(s) = \zeta(s) , \tag{3.5}$$

$$\tilde{\zeta}_P(s)\tilde{\zeta}_Q(s) = \zeta^{-1}(s) , \tag{3.6}$$

$$\zeta_P(s)\tilde{\zeta}_Q(s) = 1 . \tag{3.7}$$

Proof. By the nature of the Dirichlet product, (3.5), (3.6), and (3.7) follow, respectively, from (3.3), (3.4), and (2.4).

By Theorem 3.1 one obtains the following generalization of (1.1):

COROLLARY 3.1.

$$\tilde{\zeta}_P(s) = \frac{\zeta_P(s)}{\zeta(s)} = \frac{1}{\zeta_Q(s)} . \tag{3.8}$$

The equality of the first two expressions in (3.8) is equivalent to the fact [3, (4.6)],

$$\sum_{d|n} \mu_P(d) = \rho_P(n) \; . \tag{3.9}$$

The following identities can be verified by Dirichlet multiplication, in connection with (3.8), (2.13), and (2.11a).

THEOREM 3.2.

$$\sum_{n=1}^{\infty} \frac{\phi_{\alpha,P}(n)}{n^s} = \frac{\zeta(s-\alpha)}{\zeta_Q(s)} = \frac{\zeta(s-\alpha)\zeta_P(s)}{\zeta(s)} \; ; \tag{3.10}$$

$$\sum_{n=1} \frac{\phi^*_{\alpha,P}(n)}{n^s} = \frac{\zeta(s)}{\zeta_Q(s-\alpha)} = \frac{\zeta(s)\zeta_P(s-\alpha)}{\zeta(s-\alpha)} \; . \tag{3.11}$$

THEOREM 3.3.

$$\sum_{n=1}^{\infty} \frac{\sigma_{\alpha,P}(n)}{n^s} = \zeta(s-\alpha)\zeta_P(s) \; ; \tag{3.12}$$

$$\sum_{n=1}^{\infty} \frac{\sigma^*_{\alpha,P}(n)}{n^s} = \zeta(s)\zeta_P(s-\alpha) \; . \tag{3.13}$$

Note that in case $P = Z$, both (3.12) and (3.13) reduce to [7, Theorem 291].
It is also noted, on the basis of (3.12) and (3.8), that

COROLLARY 3.3.

$$\zeta_Q(s) \sum_{n=1}^{\infty} \frac{\sigma_{\alpha,P}(n)}{n^s} = \sum_{n=1}^{\infty} \frac{\sigma_\alpha(n)}{n^s} \; . \tag{3.14}$$

Multiplying (3.14) by $\zeta_P(s)$ and comparing coefficients, one obtains the arithmetical relation.

COROLLARY 3.4.

$$\sigma_{\alpha,P}(n) = \sum_{d\delta=n} \sigma_\alpha(d)\mu_P(\delta) \; . \tag{3.15}$$

This analogue of (2.12) can also be proved arithmetically on the basis of (3.9) and the definition of $\sigma_{\alpha,P}(n)$.

In the remainder of this section, we list for later reference, explicit evaluations of $\zeta_P(s)$ for various sets P. Let k and r denote fixed positive integers and p a fixed prime. We define direct factor sets $P = A_k$, B_k, C_p, D_r, E_r as follows: A_k (the set of kth powers), B_k (the set of k-free integers), C_p (the non-negative powers of p), D_r (the divisors of r), E_r (the complete divisors of r). A divisor d of r is said to be complete if $(d, r/d) = 1$.

We have the following representations.

$$\zeta_{A_k}(s) = \zeta(ks) \qquad (ks > 1), \tag{3.16}$$

$$\zeta_{B_k}(s) = \frac{\zeta(s)}{\zeta(ks)} \qquad (s > 1), \tag{3.17}$$

$$\zeta_{C_p}(s) = \frac{p^s}{p^s - 1} \qquad (s > 0), \tag{3.18}$$

$$\zeta_{D_r}(s) = \frac{\sigma_s(r)}{r^s} = \sigma_{-s}(r) \,, \tag{3.19}$$

$$\zeta_{E_r}(s) = \frac{\sigma'_s(r)}{r^s} = \sigma'_{-s}(r) \,, \tag{3.20}$$

where $\sigma'_s(r)$ denotes the sum of the sth powers of the complete divisors of r. For a proof of (3.17) we refer to [7, Theorem 303]; (3.18) results on summing a geometric series.

We mention the following special cases of (3.10) and (3.12), which result on the basis of (3.16) and (3.17), respectively.

$$\sum_{n=1}^{\infty} \frac{\phi_{\alpha,A_k}(n)}{n^s} = \frac{\zeta(s-\alpha)\zeta(ks)}{\zeta(s)} \qquad (s > \alpha,\ s > 1), \tag{3.12}$$

$$\sum_{n=1}^{\infty} \frac{\sigma_{\alpha,B_k}(n)}{n^s} = \frac{\zeta(s-\alpha)\zeta(s)}{\zeta(ks)} \qquad (s > \alpha,\ s > 1). \tag{3.22}$$

4. Mean values of totient and divisor functions. In this section we prove, along classical lines, some simple estimates for the functions introduced in § 2. We require no more than the following elementary facts:

$$\sum_{n \leqq x} \frac{1}{n^\alpha} = \begin{cases} O(1) & \text{if } \alpha > 1\,, \\ O(\log x) & \text{if } \alpha = 1\,, \\ O(x^{1-\alpha}) & \text{if } \alpha < 1\,; \end{cases} \tag{4.1}$$

$$\sum_{n \leqq x} n^\alpha = \frac{x^{\alpha+1}}{\alpha+1} + \begin{cases} O(x^\alpha) & \text{if } \alpha \geqq 0\,, \\ O(1) & \text{if } -1 < \alpha < 0\,; \end{cases} \tag{4.2}$$

$$\sum_{n > x} \frac{1}{n^\alpha} = O\left(\frac{1}{x^{\alpha-1}}\right), \qquad \alpha > 1\,. \tag{4.3}$$

LEMMA 4.1. *For P an arbitrary direct factor set, $\mu_P(n)$ is bounded; in fact, for each $n > 0$, $\mu_P(n) = 1, -1,$ or 0.*

Proof. In view of the factorability [3, Theorem 1] of $\mu_P(n)$, it suffices to prove the lemma in case $n = p^h$, p prime, $h > 0$. We have then by (2.2),

$$\mu_P(p^h) = \rho_P(p^h) - \rho_P(p^{h-1}) \,,$$

so that

$$\mu_P(p^h) = \begin{cases} 1 & (p^h \in P,\ p^{h-1} \notin P) \\ -1 & (p_h \notin P,\ p^{h-1} \in P) \\ 0 & (\text{otherwise}). \end{cases} \tag{4.4}$$

The lemma is proved.

As a consequence of Lemma 4.1, one obtains

COROLLARY 4.1. *The series* (3.2) *is absolutely convergent for* $s > 1$.

In the following, x will be assumed > 1.

THEOREM 4.1. *For all* $\alpha > 0$

$$\sum_{n \leqq x} \phi_{\alpha,P}(n) = \left(\frac{x^{\alpha+1}}{\alpha+1}\right)\frac{1}{\zeta_Q(\alpha+1)} + O(e_\alpha(x)) , \tag{4.5}$$

$$\sum_{n \leqq x} \sigma_{\alpha,P}(n) = \left(\frac{x^{\alpha+1}}{\alpha+1}\right)\zeta_P(\alpha+1) + O(e_\alpha(x)) , \tag{4.6}$$

where

$$e_\alpha(x) = \begin{cases} x^\alpha & (\alpha > 1) \\ x \log x & (\alpha = 1) \\ x & (\alpha < 1). \end{cases}$$

Proof. We prove (4.5). By (2.7)

$$\begin{aligned} \Phi_{\alpha,P}(x) \equiv \sum_{n \leqq x} \phi_{\alpha,P}(n) &= \sum_{n \leqq x} \sum_{\substack{\delta \mid n \\ (d\delta = n)}} \delta^\alpha \mu_P\left(\frac{n}{\delta}\right) \\ &= \sum_{d\delta \leqq x} \delta^\alpha \mu_P(d) = \sum_{d \leqq x} \mu_P(d) \sum_{\delta \leqq x/d} \delta^\alpha . \end{aligned} \tag{4.7}$$

Hence by (4.2) and Lemma 4.1,

$$\begin{aligned} \Phi_{\alpha,P}(x) &= \sum_{d \leqq x} \mu_P(d)\left\{\frac{(x/d)^{\alpha+1}}{\alpha+1} + O\left(\left(\frac{x}{d}\right)^\alpha\right)\right\} \\ &= \frac{x^{\alpha+1}}{\alpha+1}\sum_{d \leqq x} \frac{\mu_P(d)}{d^{\alpha+1}} + O\left(x^\alpha \sum_{d \leqq x} \frac{1}{d^\alpha}\right) . \end{aligned}$$

By (4.1) and Corollary 4.1, one may write then

$$\Phi_{\alpha,P}(x) = \frac{x^{\alpha+1}}{\alpha+1}\left\{\tilde{\zeta}_P(\alpha+1) - \sum_{d > x} \frac{\mu_P(d)}{d^{\alpha+1}}\right\} + (e_\alpha(x)) . \tag{4.8}$$

But by Lemma 4.1 and (4.3), it follows that

$$\sum_{d > x} \frac{\mu_P(d)}{d^{\alpha+1}} = O\left(\sum_{d > x} \frac{1}{d^{\alpha+1}}\right) = O\left(\frac{1}{x^\alpha}\right) \tag{4.9}$$

for all $\alpha > 0$. By (4.8), (4.9), and (3.8) the proof of (4.5) is complete.

The proof of (4.6) is similar and the details will be omitted; likewise for the following result.

THEOREM 4.2. *For all* $\alpha > 0$

$$\sum_{n \leqq x} \phi^*_{-\alpha,P}(n) = \frac{x}{\zeta_Q(\alpha+1)} + O(e^*_\alpha(x)) , \tag{4.10}$$

$$\sum_{n \leqq x} \sigma^*_{-\alpha,P}(n) = x\zeta_P(\alpha+1) + O(e^*_\alpha(x)) , \tag{4.11}$$

were $e^*_\alpha(x) = x^{-\alpha}e_\alpha(x)$ *and* $e_\alpha(x)$ *is defined as in Theorem* 4.1.

5. Asymptotic density of vector sets. We shall refer to the greatest common divisor (a_i) of the components of a Z_t-vector $\{a_i\}$ as the *index factor* of the vector. Let S be a set of positive integers and let $N_t(x, S)$ denote the number of Z_t-vectors with components $a_i \leqq x$ $(i = 1, \cdots, t)$ and with index factor in S. Then place

$$\delta_t(S) = \lim_{x\to\infty} \frac{N_t(x, S)}{x^t} ,$$

(if this limit exists) and call $\delta_t(S)$ the asymptotic density of the set of Z_t-vectors with index factor in S. We now prove the principal result of this section.

THEOREM 5.1. *If t is an integer* $\geqq 2$, *then*

$$N_t(x, P) = \frac{x^t}{\zeta_Q(t)} + \begin{cases} O(x \log x) & \textit{if } t = 2, \\ O(x^{t-1}) & \textit{if } t > 2. \end{cases} \tag{5.1}$$

Proof. For positive integral r, $x \geqq 1$, place

$$\Phi_{r,P}(x) = \sum_{n \leqq x} J_{r,P}(n) = \sum_{n \leqq x} \phi_{r,P}(n) , \qquad \Phi_{0,P}(x) = 1 .$$

Let j be a fixed integer, $1 \leqq j \leqq t$, and let $i_1, \cdots, i_j$ be a set of distinct integers satisfying $1 \leqq i < \cdots < i_j \leqq t$. Consider all Z_t-vectors such that the components in the positions $i_1, \cdots, i_j$ have the same value n, the components in the remaining positions are $\leqq n$, and the index factor is in P. Denote by S_j the set of all such vectors, including repetitions, obtained by letting n range over the set, $1 \leqq n \leqq x$, and by choosing the set, $i_1, \cdots, i_j$, in every possible way. Then if $N(S_j)$ denotes the number of elements in S_j, it follows that

$$N(S_j) = \binom{t}{j} \Phi_{t-j,P}(x) . \tag{5.2}$$

Consider now a fixed Z_t-vector, $\beta_k \in S_k$, $1 \leqq k \leqq t$, with exactly k of its components equal to n and the remaining components $<n$. Then β_k appears $\binom{k}{j}$ times in S_j, it being understood that $\binom{k}{j} = 0$ if $j > k$. In view of the fact,

$$\sum_{j=1}^{t} (-1)^{j+1}\binom{k}{j} = 1 ,$$

it follows that β_k is contained exactly once in the set

$$\sum_{j=1}^{t} (-1)^{j+1} S_j .$$

Consequently

$$N_t(x, P) = \sum_{j=1}^{t} (-1)^{j+1} N(S_j) ;$$

hence by (5.2),

$$N_t(x, P) = \sum_{j=1}^{t} (-1)^{j+1}\binom{t}{j}\Phi_{t-j,P}(x) .$$

The theorem follows by (4.5) on taking limits.

As a corollary of Theorem 5.1 one obtains by (3.8),

COROLLARY 5.1 (cf. [2, p. 8]). *If $t \geqq 2$, then $\delta_t(P)$ exists and is given by*

$$\delta_t(P) = \frac{1}{\zeta_Q(t)} = \frac{\zeta_P(t)}{\zeta(t)} . \tag{5.3}$$

As in § 3 let r and k denote positive integers and p a positive prime. On the basis of the evaluations (3.16)—(3.20), we obtain the following special cases of Corollary 5.1.

COROLLARY 5.2. *The asymptotic density of the Z_t-vectors,, $t \geqq 2$,*
(i) *with index factor a* kth power is

$$\delta_t(A_k) = \frac{\zeta(kt)}{\zeta(t)} ; \tag{5.4}$$

(ii) *with k-free index factor is*

$$\delta_t(B_k) = \frac{1}{\zeta(kt)} ; \tag{5.5}$$

(iii) *with index factor a non-negative power of p is*

$$\delta_t(C_p) = \left(\frac{p^t}{p^t-1}\right)\frac{1}{\zeta(t)} ; \tag{5.6}$$

(iv) *with index factor a divisor of r is*

$$\delta_t(D_r) = \frac{\sigma_t(r)}{r^t\zeta(t)}\,; \tag{5.7}$$

(v) *with index factor a complete divisor of r is*

$$\delta_t(E_r) = \frac{\sigma'_t(r)}{r^t\zeta(t)} = \frac{\sigma'_{-t}(r)}{\zeta(t)}\,. \tag{5.8}$$

The results contained in (5.4) and (5.5) are due originally to Gegenbauer [5]. In case $k = 1$, (5.5) becomes $\delta_t(B_1) = 1/\zeta(t)$, $t \geqq 2$ [9, p. 156]. Further specialization of (5.5) to the case $k = 1$, $t = 2$ yields the classical result [7, Theorem 332] asserting that the probability that a pair of integers be relatively prime is $6/\pi^2$. By (5.4), with $k = 2$, $t = 2$, it follows that the density of the integral pairs whose greatest common divisor is a perfect square is $\pi^2/15$. The case $p = 2$, $t = 2$ in (5.6) shows that the density of the integral pairs with greatest common divisor a power of 2 is $8/\pi^2$. By (5.7) with $r = 8$, $t = 2$, it follows that the density of the pairs of integers whose greatest common divisor is a factor of 8 is $255/32\pi^2$.

COROLLARY 5.3. *If $t \geqq 2$ and r is a positive integer, then the asymptotic density of the Z_t-vectors with index factor r is*

$$\delta_t(r) = \frac{1}{r^t\zeta(t)}\,. \tag{5.9}$$

Sketch of proof. The corollary is true in case $r = 1$, as noted above on the basis of (5.5), or alternatively by (5.7) with $r = 1$. The proof can be completed for arbitrary r by induction on the number of distinct prime factors of r and application of (5.8). The details are omitted.

The preceding corollary is due to Kronecker in case $t = 2$ [8, p. 311]. It was proved in the general case by Cesàro [1, p. 293]; a further generalization was given by G. Daniloff [4, p. 587].

6. Generalization of the second Möbius inversion formula. In case $P = 1$, $Q = Z$, the following inversion relation reduces to a familiar analogue [7, Theorem 268] of the Möbius inversion formula.

THEOREM 6.1. *Let x denote a positive real variable; then*

$$f(x) = \sum_{n\leqq x}\rho_Q(n)g\left(\frac{x}{n}\right) \rightleftarrows g(x) = \sum_{n\leqq x}\mu_P(n)f\left(\frac{x}{n}\right)\,. \tag{6.1}$$

Proof. Let $g(x)$ be defined as on the right of (6.1). Then

$$\sum_{n\leqq x}\rho_Q(n)g\left(\frac{x}{n}\right) = \sum_{n\leqq x}\rho_Q(n)\sum_{\substack{d\leqq x/n\\(l=nd)}}\mu_P(d)f\left(\frac{x/n}{d}\right)$$

$$= \sum_{l\leqq x}f\left(\frac{x}{l}\right)\sum_{l=dn}\mu_P(d)\rho_Q(n) = f(x) ,$$

on the basis of (2.4). The converse is proved similarly.

We define $[x]_P$ to be the number of positive integers $\leqq x$ belonging to P. It is evident, by property (ii) of the conjugate pair P, Q, that

$$[x] = [x]_Z = \sum_{\substack{n\leqq x\\ n\in Q}}\left[\frac{x}{n}\right]_P = \sum_{n\leqq x}\left[\frac{x}{n}\right]_P\rho_Q(n) . \tag{6.2}$$

Applying the above inversion theorem to (6.), one obtains

THEOREM 6.2.

$$[x]_P = \sum_{n\leqq x}\mu_P(n)\left[\frac{x}{n}\right] . \tag{6.3}$$

We deduce two special cases of (6.3). Let A_k, B_k be the P-sets defined in § 3 and place (as in I), $\lambda_k(n) = \mu_{A_k}(n)$, $\mu_k(n) = \mu_{B_k}(n)$. Putting $[x]_k = [x]_{B_k}$ and nothing that $[\sqrt[k]{x}] = [x]_{A_k}$, one obtains

COROLLARY 6.1.

$$[x]_k = \sum_{n\leqq x}\mu_k(n)\left[\frac{x}{d^k}\right] = \sum_{d^k\leqq x}\mu(d)\left[\frac{x}{d^k}\right] , \tag{6.4}$$

$$[\sqrt[k]{x}] = \sum_{n\leqq x}\lambda_k(n)\left[\frac{x}{n}\right] . \tag{6.5}$$

These formulas are classical [6], [9, p. 35]. Note that (6.4) and (6.5) reduce to (1.3) in the cases $k = 1$ and $k = 0$, respectively.

It can be shown easily, on the basis of (6.4), that $\delta_1(B_k) = 1/\zeta(k)$, $k > 1$ (cf. [7, Theorem 333] in case $k=2$). In words, this states that the asymptotic density of the k-free integers ($k \geqq 2$) is $1/\zeta(k)$; in conjunction with (5.5) it therefore follows that

COROLLARY 6.2. *If $kt \geqq 2$, then the asymptotic density of the Z_t-vectors with k-free index factor is $1/\zeta(kt)$.*

Finally, we consider the function $\phi_P(x, n)$ defined to be the number of positive integers $a \leqq x$ such that $(a, n) \in P$. In case $P = 1$, $\phi_P(x, n)$ becomes Legendre's function $\phi(x, n)$. To deal with $\phi_P(x, n)$ we have the following extension of [3, Theorem 4] which can be proved in much the same way.

LEMMA 6.1. *Let d range over the divisors of n, $d \in Q$, and for* *each such d, let y range over the positive integers $a \leqq x/d$ such that $(a, n/d) \in P$. Then the set dy consists of the positive integers $\leqq x$*

An immediate consequence of this lemma is

THEOREM 6.3.

$$\sum_{d|n} \phi_P\left(\frac{x}{d}, \frac{n}{d}\right)\rho_Q(d) = [x] . \tag{6.6}$$

THEOREM 6.4.

$$\phi_P(x, n) = \sum_{d|n} \mu_P(d)\left[\frac{x}{d}\right] . \tag{6.7}$$

Theorem 6.4 can be deduced from (6.6) by a direct application of the following easily proved extension of (2.3).

THEOREM 6.5. *If $f(x, n)$ and $g(x, n)$ are functions of the real variable x and the positive integral variable n, then*

$$g(x, n) = \sum_{d|n} \rho_Q(d) f\left(\frac{x}{d}, \frac{n}{d}\right) \rightleftarrows f(x, n) = \sum_{d|n} \mu_P(d) g\left(\frac{x}{d}, \frac{n}{d}\right) . \tag{6.8}$$

The proof is omitted.

BIBLIOGRAPHY

1. Ernest Cesàro, *Sur le plus grand commun diviseur de plusieurs nombres*, Annali di Matematica, **13** (1885), II, 291-294.
2. ———, *Sur l'étude des événements arithmétiques*, Mémoires de l'académie royale des sciences de Belgique, **47** (1886), 13 pages.
3. Eckford Cohen, *A class of residue systems* (mod r) *and related arithmetical functions, I. A generalization of Möbius inversion*, Pacific J. Math., **9** (1959), 13-23.
4. Georges Daniloff, *Contribution à la théorie des fonctions arithmétiques*, Recueil de l'Académie des Sciences Bulgare, **35** (1941), 479-590. (Bulgarian, French summary).
5. Leopold Gegenbauer, *Einige asymptotische Gesetze der Zahlentheorie*, Sitzungsberichte der Akademie der Wissenschaften (Vienna), **92** (1885), 1290-1306.
6. ———, *Über die Anzahl der Primzahlen*, Sitzungsberichte der Akademie der Wissenschaften (Vienna), **95** (1887), 94-96.
7. G. H. Hardy and E. M. Wright, *Introduction to the Theory of Numbers* (3rd edition), Oxford, 1954.
8. Leopold Kronecker, *Vorlesungen über Zahlentheorie*, Leipzig, 1901.
9. I. M. Vinogradov, *Elements of Number Theory*, New York, 1954, (English translation).
10. Aurel Wintner, *The Theory of Measure in Arithmetical Semigroups*, Baltimore, 1944.

THE UNIVERSITY OF TENNESSEE

BOOLEAN ALGEBRAS OF PROJECTIONS OF FINITE MULTIPLICITY

S. R. FOGUEL

Introduction. The multiplicity theory in Banach spaces has been developed recently by Dieudonne [2] and Bade [1]. In [6] we studied the algebra of bounded operators, in a given Hilbert space, that commute with all projections of a given Boolean algebra of self adjoint projections. By using Bade's paper [1], we propose to generalize these results to Banach spaces. The notation of [1] will be used. Let X be a complex Banach space. Let the Boolean algebra of projections be given as follows:

On the compact Hausdorff space Ω, let a measure $E(\cdot)$ be defined for every Borel set, such that:

1. For every Borel set α, $E(\alpha)$ is a projection on X.
2. For every $x \in X$, the vector valued set function $E(\cdot)x$ is countable additive.
3. If α and β are Borel sets then

$$E(\alpha)E(\beta) = E(\alpha \cap \beta) \ .$$

4. There exists a constant M such that $|E(\alpha) \leqq M|$ for every Borel set α.
5. The Boolean algebra of projections $E(\cdot)$ is complete. (See [1] for definition of completeness.)

In [1] the space Ω was defined to be the Stone space of the Boolean algebra. In the above form it is easier to find examples. Bade's results remain true for this slightly generalized version.

Throughout the paper we assume that the Boolean algebra has uniform multiplicity $n, n < \infty$. (Definition 3.2 of [1]). Thus the following is proved in [1]:

There exist n vectors $x_1, x_2, \cdots, x_n$ and n bounded functionals $x_1^*, x_2^*, \cdots, x_n^*$ such that:

1. $$X = \bigvee_{i=1}^{n} sp(E(\alpha)x_i, \ \alpha \text{ a Borel set})$$
2. Let $x_i^* E(\cdot)x_i = \mu_i(\cdot)$. The measures $\mu_i i = 1, \cdots, n$ are equivalent.
3. For every Borel set e, $\mu_i(e) \geqq 0$ and $\mu_i(e) = 0$ and only if $E(e) = 0$.
4. If $i \neq j$ then $x_i^* E(e) x_j = 0$.
5. For every $x \in X$ there exists n functions $f_1(\omega), \cdots, f_n(\omega)$ defined on Ω such that:
   a. $f_i(\omega) \in L(\Omega, \mu_i)$.
   b. For every Borel set e,

Received July 21, 1958, and in revised form January 9, 1959.

$$x_i^* E(e)x = \int_e f_i(\omega)\mu_i(d\omega) \, .$$

c. Let $e_m = \Big\{\omega \big| |f_i(\omega)| \leqq m, \, i = 1, \cdots, n\Big\}$
then

$$x = \lim_{m \to \infty} \sum_{i=1}^{n} \int f_i(\omega) E(d\omega) x_i \, .$$

d. The transformation T from X to $\sum_{i=1}^{n} L(\mu_i)$ given by

$$Tx = \begin{pmatrix} f_1(\omega) \\ \cdot \\ \cdot \\ \cdot \\ f_n(\omega) \end{pmatrix}$$

is continuous. The functions $f_1(\omega), \cdots, f_n(\omega)$ are uniquely defined by x, up to sets of measure zero.

These results are proved in 5.1 and 5.2 of [1]. Instead of writing $Tx = \begin{pmatrix} f_1(\omega) \\ \vdots \\ f_n(\omega) \end{pmatrix}$ let us use the notation $x \sim \begin{pmatrix} f_1(\omega) \\ \vdots \\ f_n(\omega) \end{pmatrix}$.

Let $\mathfrak{A}$ be the algebra of bounded operators on X, which commute with all the projections $E(\alpha)$. The purpose of this paper is to study $\mathfrak{A}$.

Representation of the Algebra $\mathfrak{A}$

Let $A \in \mathfrak{A}$, and let

$$Ax_i \sim \begin{pmatrix} a_{1,i}(\omega) \\ \cdot \\ \cdot \\ \cdot \\ a_{n,i}(\omega) \end{pmatrix} \qquad i = 1, \cdots, n \, .$$

Denote this correspondence by $A \sim (a_{i,j}(\omega))$. The functions $a_{i,j}(\omega)$ satisfy by 5.b.

2.1 $$x_i^* E(e) A x_j = x_i^* A E(e) x_j = \int_e a_{i,j}(\omega) \mu_i(d\omega)$$

and

$$a_{i,j}(\omega) \in L(\mu_i) \, .$$

Equation 2.1 defines the functions $a_{i,j}(\omega)$ uuiquely (a.e.).

Now let $x \in X$ and $x \sim \begin{pmatrix} f_1(\omega) \\ \vdots \\ f_n(\omega) \end{pmatrix}$. If e is a Borel set on which the functions $f_i(\omega)$, $a_{i,j}(\omega)$ are bounded then:

$$E(e)x = \sum_{i=1}^{n} \int_e f_i(\omega) E(d\omega) x_i$$

and

$$\begin{aligned} AE(e)x = E(e)Ax &= \sum_{i=1}^{n} \int_e f_i(\omega) E(d\omega)(Ax_i) \\ &= \sum_{i=1}^{n} \int_e f_i(\omega) E(d\omega)(E(e)Ax_i) \ . \end{aligned}$$

But

$$E(e)Ax_i = \sum_{j=1}^{n} \int_e a_{j,i}(\omega) E(d\omega) x_j \ .$$

Hence

$$E(e)Ax = \sum_{1 \leqq i, j \leqq n} \int_e f_i(\omega) E(d\omega) \Big(\int_e a_{j,i}(\lambda) E(d\lambda) x_j \Big) \ .$$

From condition 3 of the introduction it follows that

$$\begin{aligned} E(e)Ax &= \sum_{1 \leqq i, j \leqq n} \int_e a_{j,i}(\omega) f_i(\omega) E(d\omega) x_j \\ &= \sum_{j=1}^{n} \int_e \Big(\sum_{i=1}^{n} a_{j,i}(\omega) f_i(\omega) \Big) E(d\omega) x_j \ . \end{aligned}$$

Therefore

$$x_j^* E(e)Ax = \int_e \Big(\sum_{i=1}^{n} a_{j,i}(\omega) f_i(\omega) \Big) \mu_j(d\omega) \ .$$

This equation means

$$Ax \sim (a_{i,j}(\omega)) \begin{pmatrix} f_1(\omega) \\ \cdot \\ \cdot \\ \cdot \\ f_n(\omega) \end{pmatrix} .$$

Remark. Equation 5.b. of the introduction was proved here, for only some Borel sets. But we know that

$$Ax \sim \begin{pmatrix} g_1(\omega) \\ \cdot \\ \cdot \\ \cdot \\ g_n(\omega) \end{pmatrix}$$

for some functions $g_1(\omega), \cdots, g_n(\omega)$. The above argument shows that

$$(a_{i,j}(\omega)) \begin{pmatrix} f_1(\omega) \\ \cdot \\ \cdot \\ \cdot \\ f_n(\omega) \end{pmatrix} = \begin{pmatrix} g_1(\omega) \\ \cdot \\ \cdot \\ \cdot \\ g_n(\omega) \end{pmatrix} \text{a.e.}$$

THEOREM 2.1. *For every operator $A \in \mathfrak{A}$ there corresponds a matrix of measurable functions $a_{i,j}(\omega)$, $1 \leqq i, j \leqq n$, such that:*

1. $a_{i,j}(\omega) \in L(\mu_i)$.
2. *If*

$$x \sim \begin{pmatrix} f_1(\omega) \\ \cdot \\ \cdot \\ \cdot \\ f_n(\omega) \end{pmatrix}$$

then

$$Ax \sim (a_{i,j}(\omega)) \begin{pmatrix} f_1(\omega) \\ \cdot \\ \cdot \\ \cdot \\ f_n(\omega) \end{pmatrix}.$$

3. *If a matrix of functions, $(b_{i,j}(\omega))$, satisfies condition 2 then $a_{i,j}(\omega) = b_{i,j}(\omega)$ a.e.*

The matrix of the sum or product of two operators is the sum or product of the matrices. If A^{-1} exists and is bounded then

$$A^{-1} \sim (a_{i,j}(\omega))^{-1}.$$

The functions $a_{i,j}(\omega)$ are determined by equation 2.1.

Proof. The existence of a representing matrix was proved above. The other parts of the theorem follow from the uniqueness assertion given in condition 3.

COROLLARY. *Let $A \in \mathfrak{A}$. If $B \in \mathfrak{A}$ and $AB = I(BA = I)$ then $BA = I(AB = I)$.*

Proof. If $AB = I$ then

$$(a_{i,j}(\omega))(b_{i,j}(\omega)) = (\delta_{i,j}) \text{ a.e.}$$

Hence

$$(b_{i,j}(\omega))(a_{i,j}(\omega)) = (\delta_{i,j}) \text{ a.e.}$$

Thus by Theorem 2.1 $BA = I$.

THEOREM 2.2. *Let A_m, $A \in \mathfrak{A}$. If the sequence $\{A_m\}$ converges strongly to A then sequence of functions $\{a_{i,j}^{(m)}(\omega)\}$ converges in measure to $a_{i,j}(\omega)$, for each $1 \leqq i, j \leqq n$. (It does not matter with respect to what measure, because the measures are finite and equivalent).*

Conversely, the sequence $\{A_m\}$ *converges strongly to* A *if:*

1. *The sequence* $\{a_{i,j}^{(m)}\}$ *converges in measure to* $a_{i,j}(\omega)$.
2. *The sequence* $\{|A_m|\}$ *is bounded.*
3. $\bigcup_{K=1}^{\infty} \{\omega \| a_{i,j}^{(m)}(\omega)| \leqq K. \quad 1 \leqq i, j \leqq n, \ m = 1, 2, \cdots\} = \Omega.$

Proof. If for each $x \in X$

$$\lim_{m\to\infty} A_m x = Ax$$

then for every Borel set e

$$\left|\int_e (a_{i,j}^{(m)}(\omega) - a_{i,j}(\omega))\mu_i(d\omega)\right| \\ = |x_i^* E(e)(A_m x_j - Ax_j)| \leqq M|A_m x_j - Ax_j| \to 0 \quad m \to \infty$$

where M does not depend on e. Thus the sequence $\{a_{i,j}^{(m)}(\omega)\}$ converges in measure to $a_{i,j}(\omega)$.

On the other hand, if conditions 1, 2 and 3 are satisfied and e is a Borel set, on which the functions $a_{i,j}^{(m)}(\omega)$ are uniformly bounded, then

$$A_m E(e)x_i = \sum_{j=1}^{n} \int_e a_{j,i}^{(m)}(\omega)E(d\omega)x_j$$

and by the Lebesque Theorem, [5] IV.10.10

$$\lim_{m\to\infty} A_m E(e)x_i = \sum_{j=1}^{n} \int_e a_{j,i}(w)E(dw)x_j\ .$$

Now, by condition 3, the set of linear combinations of $E(e)x_i$, $1 \leqq i \leqq n$ and e as defined above, is dense. Thus the sequence $\{A_m x\}$ has a limit for x in a dense subset of X, and by condition 2 it has a limit for every $x \in X$. Let A be the strong limit of $\{A_m\}$ then

$$AE(e)x_i = \sum_{j=1}^{n} \int_e a_{j,i}(\omega)E(d\omega)x_j\ .$$

Thus the matrix of A is $(a_{i,j}(\omega))$. (See Remark before Theorem 2.1).

In order to develop further the theory, let us borrow the following results from [6].

LEMMA 2.1. *Let* $(a_{i,j}(\omega))$ *be a matrix of measurable finite functions. There exists a decomposition of the form*

2.2
$$(a_{i,j}(\omega)) = \sum_{k=1}^{n} z_k(\omega)\varepsilon_k(\omega) + N(\omega)$$

where $z_1(\omega), \cdots, z_n(\omega)$ *are measurable functions and* $\varepsilon_1(\omega), \cdots, \varepsilon_n(\omega)$, $N(\omega)$ *are matrices of measurable functions satisfying:*

$$\varepsilon_i^2(\omega) = \varepsilon_i(\omega),\ \textit{if } i \neq j \textit{ then } \varepsilon_i(\omega)\varepsilon_j(\omega) = 0,\ \sum_{i=1}^{n} \varepsilon_i(\omega) = (\delta_{i,j}) .$$

Also

$$\varepsilon_i(\omega)N(\omega) = N(\omega)\varepsilon_i(\omega), \qquad (N(\omega))^n = 0 .$$

Moreover, there exist n Borel sets $\beta_1, \cdots, \beta_n$ *whose union is* Ω *such that on* β_i *the numbers* $z_1(\omega), \cdots, z_i(\omega)$ *are different while* $z_{i+1}(\omega) = \cdots = z_n(\omega) = 0$.

The proof is given in Lemma 3.1, 3.2 and Theorem 3.1 of [6].

THEOREM 2.3. *Let* $A \in \mathfrak{A}$. *There exists a sequence of Borel sets,* $\{\alpha_m\}$ *such that:*

1. *The sequence* $\{\alpha_m\}$ *increases to* Ω.
2. *The operator* $AE(\alpha_m)$ *is spectral. (For definition of spectral operators see* [3]).

Thus A is a strong limit of a sequence of spectral operators.

Proof. Let $A \sim (a_{i,j}(\omega)) = \sum_{k=1}^{n} z_k(\omega)\varepsilon_k(\omega) + N(\omega)$, where the right side of the equation is defined in Lemma 2.1. Let α be a Borel set such that

a. On the set α the functions $z_k(\omega)$ are bounded.

b. If $\chi_\alpha(\omega)$ is the characteristic function of α, then $\chi_\alpha(\omega)\varepsilon_k(\omega)$ and $\chi_\alpha(\omega)N(\omega)$ are representing matrices of the operators $E_{k,\alpha}$ and N_α respectively in $\mathfrak{A}$.

Then, by Theorem 2.1,

$$AE(\alpha) = \sum_{i=1}^{n} \Big(\int_\alpha z_i(\omega)E(d\omega) \Big) E_{i,\alpha} + N_\alpha \tag{2.3}$$

where $E_{i,\alpha}$ are disjoint projections and N_α is a nilpotent of order n commuting with them.

Thus, for such α, the operator $AE(\alpha)$ is spectral, and the resolution of the identity (see [3]) of A restricted to $E(\alpha)X$ is

$$\sum E(z_i^{-1}(\cdot))E_{i,\alpha} .$$

In order to prove the theorem, we have to find a sequence of Borel sets, satisfying conditions a, b and 1. Also with no loss of generality, we may study the operator A on $E(\beta_i)X$ (Lemma 2.1). Thus we may assume that at each point ω, the matrix $(a_{j,k}(\omega))$ has exactly i eigenvalues.

Define

$$\alpha_m = \Big\{ \omega \,\big|\, |z_j(\omega)| \leqq m \text{ and } |z_j(\omega) - z_k(\omega)| \geqq \frac{1}{m},\ 1 \leqq k < j \leqq i \Big\} .$$

On the set α_m the matrix $\varepsilon_1(\omega)$ can be calculated as follows:

Let $Q(z)$ be the polynomial

$$Q(z) = b_0 + b_1 z + \cdots + b_{i(n+1)-1} z$$

such that:

$$Q(z_1(\omega)) = 1, \quad Q(z_j(\omega)) = 0, \quad 2 \leqq j \leqq i,$$
$$Q^{(P)}(z_j(\omega)) = 0, \quad 1 \leqq j \leqq i, \quad 1 \leqq p \leqq n$$

then

$$Q(a_{i,j}(\omega)) = \varepsilon_1(\omega) \text{ [see [4] p. 188].}$$

These equations have a unique solution $b_j = b_j(\omega)$, which are measurable and bounded (on α_m) functions of ω. Thus

$$\chi_{\alpha_m}(\omega)\varepsilon_1(\omega)$$
$$= \chi_{\alpha_m}(\omega)[b_0(\omega) + b_1(\omega)(a_{i,j}(\omega)) + \cdots + b_{i(n+1)-1}(a_{i,j}(\omega))^{i(n+1)-1}]$$

and this matrix represents the operator $E_{1,m}$, in $\mathfrak{A}$, where

$$E_{1,m} = E(\alpha_m)\Big[\int b_0(\omega)E(d\omega)$$
$$+ A\int b_1(\omega)E(d\omega) + \cdots + A^{i(n+1)-1}\int b_{i(n+1)-1}(\omega)E(d\omega)\Big].$$

Similarly the matrices $\chi_{\alpha_m}(\omega)\varepsilon_j(\omega)$ represent the operators $E_{j,m}$ in $\mathfrak{A}$, and by equation 2.2 the matrix $\chi_{\alpha_m}(\omega)N(\omega)$ represents a nilpotent of order n, N_m, in $\mathfrak{A}$ where

$$AE(\alpha_m) = \sum_{j=1}^{i} E_{j,m}\int_{\alpha_m} z_j(\omega)E(d\omega) + N_m .$$

COROLLARY. *Let $A \in \mathfrak{A}$ be a generalized nilpotent (see [3] for definition) then*

$$A^n = 0 .$$

Proof. By equation 2.3 and Theorem 8 of [3]

$$AE(\alpha_m) = N_m .$$

Hence for every $x \in X$

$$E(\alpha_m)A^n x = 0$$

therefore

$$A^n x = 0 .$$

LEMMA 2.2. *Let $A \in \mathfrak{A}$. If $A \sim (a_{i,j}(\omega))$ and $z_k(\omega)$, $k = 1, 2, \cdots, n$ are the functions defined in equation 2.2 then*

$$|z_k(\omega)| \leqq |A| \text{ a.e.}$$

Proof. Let us assume, to the contrary, that for some i and $\varepsilon > 0$ the set

$$\gamma = \{\omega \| z_i(\omega)| \geqq |A| + \varepsilon\}$$

is not of measure zero. Let $\{\alpha_m\}$ be the sequence defined in Theorem 2.3, for some $mE(\gamma \cap \alpha_m) \neq 0$. Now, on $\gamma \cap \alpha_m$ $|z_i(\omega)| \geqq |A| + \varepsilon > 0$ hence $\varepsilon_i(\omega) \neq 0$. Thus $E(\gamma \cap \alpha_m)E_{i,\alpha_m} \neq 0$, where E_{i,α_m} is defined in Theorem 2.3. If the operator B is the restriction of A to $E(\gamma \cap \alpha_m)E_{i,\alpha_m}X$ then

$$B = \int_{\gamma \cap \alpha_m} z_i(\omega)E(d\omega) + M$$

where M is a nilpotent. Thus, if $|\mu| \leqq |A|$ then $|\mu| \leqq |z_1(\omega)| - \varepsilon$, $\omega \in \gamma \cap \alpha_m$, and $\mu \notin \sigma(B)$. Also, if $|\mu| > |A|$ then $|\mu| > |B|$ and $\mu \notin \sigma(B)$. This shows that $\sigma(B)$ is empty which is impossible.

THEOREM 2.4. *Let $(a_{i,j}(\omega))) \sim A \in \mathfrak{A}$. If the number $\lambda \notin \sigma(A)$ then for some $\varepsilon > 0$*

$$\text{dist}\,(\lambda, \sigma(a_{i,j}(\omega))) \geqq \varepsilon \text{ a.e.}$$

Proof. Let $\lambda \in \rho(A)$. The matrix of $(\lambda I - A)^{-1}$ has the form

$$\sum_{i=1}^{n}\left(\frac{\varepsilon_i(\omega)}{\lambda - z_i(\omega)} - \frac{N(\omega)}{(\lambda - z_i(\omega))^2} + \cdots + \frac{(-N(\omega))^{n-1}}{(\lambda - z_i(\omega))^n}\right).$$

Thus by Lemma 2.2

$$\frac{1}{\text{dist}\,(\lambda, \sigma(a_{i,j}(\omega)))} = \max_k \frac{1}{|\lambda - z_k(\omega)|} \leqq |(\lambda I - A)^{-1}| \text{ a.e}\,.$$

THEOREM 2.5. *Let $(a_{i,j}(\omega)) \sim A \in \mathfrak{A}$ and let $f(z)$ be regular in a neighborhood of $\sigma(A)$. Then the matrix $f((a_{i,j}(\omega)))$ exists a.e. and it is the matrix corresponding to $f(A)$.*

Proof. Let e be a Borel subset of Ω then

$$x_k^* E(e) f(A) x_j = x_k^* E(e) \frac{1}{2\pi i}\int_C f(\lambda) R(\lambda, A) x_j d\lambda$$

where C is a finite collection of Jordan curves surrounding $\sigma(A)$. Now $R(\lambda; A) \sim (r_{k,j}(\omega, \lambda)) = (\lambda\delta_{k,j} - a_{k,j}(\omega))^{-1}$ thus

$$x_k^* E(e) f(A) x_j = \frac{1}{2\pi i}\int_C f(\lambda)(x_k^* E(e) R(\lambda, A) x_j) d\lambda$$

$$= \frac{1}{2\pi i}\int_C f(\lambda)\left[\int_e r_{k,j}(\omega; \lambda)\mu_k(d\omega)\right] d\lambda$$

by equation 2.1. The functions $r_{k,j}(\omega, \lambda)$ can be computed by Cramer's rule. By Theorem 2.4 and the compactness of C there exists a positive constant δ such that if $\lambda \in C$ then

$$\operatorname{dist}(\lambda, \sigma(a_{k,j}(\omega))) \geqq \delta \text{ a.e.}$$

Now, if e is a Borel set on which the functions $a_{i,j}(\omega)$ are bounded, then the functions $r_{k,j}(\omega, \lambda)$ are measurable and bounded on $e \times C$. For such Borel sets e, we may use Fubini's theorem to conclude that

$$x_k^* E(e) f(A) x_j = \int_e \frac{1}{2\pi i}\left(\int_C f(\lambda) r_{k,j}(\omega, \lambda) d\lambda\right)\mu_k(d\omega) .$$

From this equation it follows that the components of the matrix of $f(A)$ are given by

$$(*) \qquad \frac{1}{2\pi i}\int_C f(\lambda) r_{k,j}(\omega, \lambda) d\lambda \text{ a.e.}$$

Now by the argument of Lemma 2.1 in [6] the matrix $f((a_{k,j}(\omega)))$ exists a.e. and its components are, thus, given by $(*)$.

THEOREM 2.6. *Let $A \in \mathfrak{A}$ be a compact operator. If $A \sim (a_{i,j}(\omega))$ and*

$$(a_{i,j}(\omega)) = \sum_{k=1}^{n} z_k(\omega)\varepsilon_k(\omega) + N(\omega)$$

is the decomposition given in Lemma 2.1, *then there exists a sequence $\{\omega_v\}$, of points in ω, such that:*

1. $E(\{\omega_v\}) \neq 0$
2. $z_k(\omega) = 0$ a.e. *for* $\omega \neq \omega_v$ $\qquad v = 1, 2, \cdots$
3. $\lim_{v\to\infty} z_k(\omega_v) = 0$.

Proof. Let β_i and α_m be the sets defined in Lemma 2.1 and Theorem 2.3. It is enough to prove the theorem for points in β_i, thus we assume that the matrix $(a_{j,k}(\omega))$ has exactly i eigenvalues. Define

$$e_{m,p} = \alpha_m \cap \left\{\omega \,|\, |z_k(\omega)| \geqq \frac{1}{p},\ k = 1, \cdots, i\right\} .$$

The operator A restricted to $E(e_{m,p})X$ is compact and, by Theorem 2.3. has a bounded inverse. Thus the space $E(e_{m,p})X$ has a finite dimension. Therefore there exists a finite set of points, $\omega_1^{m,p}, \cdots, \omega_j^{m,p}$, such that

$$E(\{\omega_k^{m,p}\}) \neq 0$$

and

$$E(e_{m,p} - \{\omega_1^{m,p}, \cdots, \omega_j^{m,p}\}) = 0 \ .$$

By letting $m, p \to \infty$ we get a sequence ω_v satisfying conditions 1 and 2. In order to prove 3, let us assume that for some $\varepsilon > 0$ there are infinitely many points, ω_v such that

$$|z_{k_v}(\omega_v)| \geqq \varepsilon \ .$$

The operator A is compact, hence $\sigma(A)$ has only zero as a limit point. By theorem 2.4 $z_{k_v}(\omega_v) \in \sigma(A)$. Thus for some constant $b \neq 0$

$$z_{k_v}(\omega_v) = b$$

for infinitely many points, ω_v. Let

$$G(b, A) = \frac{1}{2\pi i}\int_C R(\lambda; A) d\lambda$$

where C is a circle around b which does not contain any other point of $\sigma(A)$. The operator $G(b; A)$ is a compact projection. The matrix of $G(b; A)$ is, according to Theorem 2.5,

$$G(b; (a_{i,j}(\omega))) \ .$$

Thus

$$G(b; A)E(\{\omega_v\}) \neq 0$$

whenever $z_{k_v}(\omega_v) = b$, because the matrix of the product is not zero at ω_v. This contradicts the fact that $G(b; A)$ is a projection into a finite dimensional space, and thus condition 3 is proved.

EXAMPLES. The following two examples are designed to show that some of the theorems, proved in [6] for Hilbert spaces, are false for Banach spaces. Notice that the examples are simble because there exist projections on

$$sp\{E(\alpha)x_i, \ \alpha \text{ a Borel set}\} \ .$$

1. Let μ be the Lebesque measure on (0, 1). Let f be a monotone increasing function such that

$$f(0) = 1, \qquad f(1) = \infty, \qquad f \in L(0, 1) \ .$$

Define

$$\mu_1(e) = \int_e f(t)\mu(dt) \ .$$

The Banach space X will be $L_1(\mu) \oplus L_1(\mu_1)$. Each $x \in X$ has the form

$$x = \begin{pmatrix} g_1(\omega) \\ g_2(\omega) \end{pmatrix}$$

$$|x| = \int |g_1| d\mu + \int |g_2| f d\mu .$$

Let

$$E(e)x = \begin{pmatrix} \chi_e(\omega) g_1(\omega) \\ \chi_e(\omega) g_2(\omega) \end{pmatrix} .$$

It follows that the Boolean algebra is complete and has uniform multiplicity 2. Let

$$x_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad x_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

$$x_1^* \begin{pmatrix} g_1(\omega) \\ g_2(\omega) \end{pmatrix} = \int g_1 d\mu, \; x_2^* \begin{pmatrix} g_1(\omega) \\ g_2(\omega) \end{pmatrix} = \int g_2 d\mu .$$

If $A \sim \mathfrak{A}$ then $A \sim \begin{pmatrix} a_{1,1}(\omega), & a_{1,2}(\omega) \\ a_{2,1}(\omega), & a_{2,2}(\omega) \end{pmatrix}$ and

$$|A| \int |g_2| f d\mu \geqq \left| A \begin{pmatrix} 0 \\ g_1 \end{pmatrix} \right| = \left| \begin{pmatrix} a_{1,2}(\omega) g_2(\omega) \\ a_{2,2}(\omega) g_2(\omega) \end{pmatrix} \right| \geqq \int |a_{1,2} g_2| d\mu$$

for every $g_2 \in L_1(\mu_1)$. Thus

$$\int_e |a_{1,2}(\omega)| d\mu \leqq |A| \int_e f d\mu .$$

Hence $|a_{1,2}(\omega)| \leqq |A| f(\omega)$ a.e., or

$$a_{1,2}(\omega) = b_{1,2}(\omega) f(\omega) \text{ and } |b_{1,2}(\omega)| \leqq |A| .$$

Similarly

$$|A| \int |g_1| d\mu \geqq \left| A \begin{pmatrix} g_1(\omega) \\ 0 \end{pmatrix} \right| = \left| \begin{pmatrix} a_{1,1}(\omega) g_1(\omega) \\ a_{2,1}(\omega) g_1(\omega) \end{pmatrix} \right| \geqq \int |a_{2,1} g_1| f d\mu$$

Hence

$$|a_{2,1}(\omega) f(\omega)| \leqq |A| \text{ a.e.}$$

or

$$a_{2,1}(\omega) = \frac{b_{2,1}(\omega)}{f(\omega)} \text{ and } |b_{2,1}(\omega)| \leqq |A| .$$

Every operator in $\mathfrak{A}$ is given, thus, by a matrix of the form:

$$\begin{pmatrix} b_{1,1}(\omega), & b_{1,2}(\omega) f(\omega) \\ \dfrac{b_{2,1}(\omega)}{f(\omega)}, & b_{2,2}(\omega) \end{pmatrix}$$

where the functions $b_{i,j}(\omega)$ are measurable and bounded. Also, every such matrix defines a bounded operator.

This example shows that Theorem 2.2 of [6] can not be generalized to Banach spaces:

The two topologies on $\mathfrak{A}$ given by the norms $|A|$ and

$$\max_{i,j} \operatorname{ess\,sup}_{\omega} |a_{i,j}(\omega)|$$

are not equivalent.

2. Let $X = C_0 \oplus l_1$. Every $x \in X$ has the form

$$x = (x_1, y_1, x_2, y_2. \cdots, x_n, y_n, \cdots)$$

where

$$\lim_{n\to\infty} x_n = 0, \; |x| = \max |x_i| + \sum_{i=1}^{\infty} |y_i| \; .$$

Define

$$E_n(x_1, y_1, \cdots, x_n, y_n, \cdots) = (0, \cdots, 0, x_n, y_n, 0 \cdots) \; .$$

The Boolean algebra, generated by E_n, has uniform multiplicity 2. Let the projection F be defined by

$$F(x_1, y_1, \cdots, x_n, y_n, \cdots) \\ = \frac{1}{2}(x_1 + y_1, x_1 + y_1, \cdots, x_n + y_n, x_n + y_n, \cdots) \; .$$

The projection F is not bounded but $|FE_n| = 1$. Let the operator B be defined by

$$B(x_1, y_1, \cdots, x_n, y_n, \cdots) = \left(\frac{x_1}{2}, \frac{y_1}{2}, \cdots, \frac{x_n}{2^n}, \frac{y_n}{2^n}, \cdots\right)$$

and let $A = BF$. The operator A is bounded and compact, for if $|x| = |(x_1, y_1, \cdots, x_n, y_n, \cdots)| \leqq 1$ then

$$\left|Ax - \frac{1}{2}\left(\frac{x_1+y_1}{2}, \frac{x_1+y_1}{2}, \cdots, \frac{x_n+y_n}{2^n}, \frac{x_n+y_n}{2^n}, 0 \cdots 0 \cdots\right)\right| \\ = \left|\frac{1}{2}\left(0, \cdots, 0, \frac{x_{n+1}+y_{n+1}}{2^{n+1}}, \frac{x_{n+1}+y_{n+1}}{2^{n+1}}, \cdots\right)\right| \\ \leqq \frac{1}{2^{n+1}}\left[\left(\frac{\sup|x_n| + \sup|y_n|}{2}\right) \right. \\ \left. + \sum_{i=1}^{\infty} \frac{1}{2^n} + \sum_{i=1}^{\infty} |y_i|\right] \leqq \frac{1}{2^{n+1}}(1 + 2 + 1) \to 0 \; .$$

Thus A is the uniform limit of compact operators. Now, $\sigma(B) = \left\{\frac{1}{2^n}, n = 1, 2, \cdots\right\}$. If $0 \neq \lambda \in \sigma(A)$ then for some $x \in X$, $\lambda x = Ax$. Hence $x = Fx$ and $\lambda x = \lambda Fx = BFx = Bx$. Therefore

$$\sigma(A) = \left\{0, \frac{1}{2^n}, n = 1, 2, \cdots\right\} .$$

Let us compute $G\left(\frac{1}{2^n}, A\right)x$ for $x \in X$.

$$G\left(\frac{1}{2^n}; A\right)x = \sum_{k=1}^{\infty} G\left(\frac{1}{2^n}; A\right)E_k x .$$

Now on $E_k X$, $\sigma(A) = \left\{0, \frac{1}{2^k}\right\}$, hence

$$G\left(\frac{1}{2^n}; A\right)E_k x = 0 \text{ for } k \neq n$$

and

$$G\left(\frac{1}{2^n}; A\right)E_n x = FE_n x .$$

Therefore

$$G\left(\frac{1}{2^n}; A\right)x = FE_n x$$

and

$$\sum_{v=1}^{n} G\left(\frac{1}{2^v}; A\right)x = F(E_1 + \cdots + E_n)x .$$

The last equation shows that A is not spectral, and the preceding equation shows that Theorem 4.4 of [6] is false for Banach spaces:

There exists a compact operator A in $\mathfrak{A}$ that is not spectral though the projections $G(\xi; A)$ are uniformly bounded.

References

1. W. G. Bade, *A multiplicity theory for Boolean algebras of projections in Banach spaces*, forthcoming.
2. J. Dieudonne, *Sur la theorie spectrale*, J. Math. Pures Appl. (9) **35** (1956), 6–10.
3. N. Dunford, *Spectral operators*, Pacific J. Math., **4** (1954), 321–354.
4. ———, *Spectral theory I, Convergence to projections*, Trans. Amer. Math. Soc., **54** (1934), 185–217.
5. N. Dunford, and J. Schwartz, *Linear operators*, forthcoming.
6. S. R. Foguel, *Normal operators of finite multiplicity*, forthcoming.

INSTITUTE OF MATHEMATICAL SCIENCES
NEW YORK

AVERAGES OF FOURIER COEFFICIENTS

RICHARD R. GOLDBERG

We shall say the sequence $a_n (n = 1, 2, \cdots)$ is a p-sequence $(1 \leqq p < \infty)$ if there is a function $f \in L^p(0, \pi)$ such that

$$a_n = \int_0^\pi f(t) \cos nt \, dt \qquad n = 1, 2, \cdots ;$$

(i.e. the a_n are Fourier cosine coefficients of an L^p function).

A famous theorem of Hardy [1] states that if a_n is a p-sequence $(1 \leqq p < \infty)$ and $b_n = \frac{1}{n}(a_1 + a_2 + \cdots + a_n)$, then b_n is also a p-sequence.

In this paper we shall prove the following generalization of Hardy's theorem:

THEOREM 1. *Let $\psi(x)$ be of bounded variation on $0 \leqq x \leqq 1$, and let $1 \leqq p < \infty$. Then, if a_n is a p-sequence and*

$$b_n = \frac{1}{n} \sum_{m=1}^{n} \psi\left(\frac{m}{n}\right) a_m ,$$

b_n is also a p-sequence.

Hardy's theorem is the special case $\psi(x) = 1$ for $0 \leqq x \leqq 1$.

If the conclusion of Theorem 1 holds for each of two functions ψ it clearly holds for their difference. Hence it is sufficient to prove Theorem 1 in the case where $\psi(x)$ is non-decreasing for $0 \leqq x \leqq 1$. Further, since any non-decreasing function may be written as the difference of two non-negative non-decreasing functions (the second of which is constant) to prove Theorem 1 it is sufficient to prove

THEOREM 1A. *Let $\psi(x)$ be non-negative and non-decreasing on $0 \leqq x \leqq 1$ and let $1 \leqq p < \infty$. Then, if a_n is a p-sequence and*

$$b_n = \frac{1}{n} \sum_{m=1}^{n} \psi\left(\frac{m}{n}\right) a_m ,$$

b_n is also a p-sequence.

The proof of Theorem 1A will follow a sequence of lemmas.

LEMMA 1. *Let $B_t(x) = \int_0^x \cos yt \, d(y - [y])$. Then there is an $M > 0$*

Received February 24, 1959. This research was supported in part by the United States Air Force under Contract No. AF 49 (638)-383 monitored by the AF Office of Scientific Research of the Air Research and Development Command.

such that

$$|B_t(x)| \leqq M \qquad 0 \leqq t \leqq \pi;\ 0 \leqq x < \infty\ .$$

The symbol $[y]$ denotes the greatest integer not exceeding y.

Proof. Let n be any non-negative integer. Then for $t > 0$

$$\int_0^n \cos yt\, dy = \frac{\sin nt}{t}$$

and

$$\int_0^n \cos yt\, d[y] = \sum_{m=1}^{n} \cos mt = \frac{\sin (n + 1/2)t}{2 \sin t/2} - \frac{1}{2}\ .$$

Hence

$$\begin{aligned} B_t(n) &= \frac{\sin nt}{t} - \frac{\sin (n+1/2)t}{2 \sin t/2} + \frac{1}{2} \\ &= \sin nt\left(\frac{1}{t} - \frac{1}{2}\cot\frac{t}{2}\right) - \frac{\cos nt}{2} + \frac{1}{2} \end{aligned}$$

and so

$$(1) \qquad |B_t(n)| \leqq \left|\frac{1}{t} - \frac{1}{2}\cot\frac{t}{2}\right| + 1 \qquad n = 0, 1, 2, \cdots$$

The right side of (1) is bounded for $0 < t \leqq \pi$. Thus for some $M \geqq 1$

$$(2) \qquad |B_t(n)| \leqq M - 1 \qquad n = 0, 1, 2, \cdots;\ 0 < t \leqq \pi\ .$$

Now take any $x \geqq 0$ and let $n = [x]$. Then

$$B_t(x) = B_t(n) + \int_n^x \cos yt d(y - [y])$$

so that from (2) we have for any $x \geqq 0$

$$|B_t(x)| \leqq M - 1 + \int_n^x |d(y - [y])| \leqq M - 1 + x - n \leqq M,\ 0 < t \leqq \pi$$

and the proof is complete since $B_0(x) = x - [x] \leqq 1 \leqq M$.

(Henceforth we assume $\psi(x) \geqq 0$ and $\psi(x)$ non-decreasing for $0 \leqq x \leqq 1$.)

LEMMA 2. *There is an $M > 0$ such that*

$$\left|\int_0^n \psi\left(\frac{x}{n}\right)\cos xt\, d(x - [x])\right| \leqq M \qquad 0 \leqq t \leqq \pi;\ n = 1, 2, \cdots$$

Proof. With $B_t(x)$ as in Lemma 1 we have

$$\int_0^n \psi\left(\frac{x}{n}\right)\cos xt\, d(x-[x]) = \int_0^n \psi\left(\frac{x}{n}\right) dB_t(x)$$
$$= \psi(1)B_t(n) - \int_0^n B_t(x) d\psi\left(\frac{x}{n}\right) .$$

Thus with M as in Lemma 1

$$\left|\int_0^n \psi\left(\frac{x}{n}\right)\cos xt\, d(x-[x])\right| \leqq M\psi(1) + M\int_0^n d\psi\left(\frac{x}{n}\right) \leqq 2M\psi(1) ,$$

and the lemma is prove (with $2M\psi(1)$ instead of M).

LEMMA 3. *Let $f \in L'(0, \pi)$ and let*

$$d_n = \frac{1}{n}\int_0^\pi f(t)dt \int_0^n \psi\left(\frac{x}{n}\right)\cos xt\, d(x-[x]) \qquad n = 1, 2, \cdots .$$

Then

$$d_n = O\left(\frac{1}{n}\right) \qquad n \to \infty \tag{3}$$

and hence d_n is a p-sequence for every $p \geqq 1$.

Proof. By Lemma 2 there is an $M > 0$ such that $|d_n| \leqq \frac{M}{n}\int_0^\pi |f(t)|dt$ from which (3) follows. From (3) it follows that $\sum_{n=1}^\infty |d_n|^q < \infty$, for every $q > 1$. By the Hausdorff-Young theorem and the fact that $L^p \subseteqq L^{p'}$ if $1 \leqq p' \leqq p$, this implies that d_n is a p-sequence for every $p \geqq 1$. (See [2].)

From now on we shall write $f \sim a_n$ as an abbreviation for $a_n = \int_0^\pi f(t)\cos nt\, dt$, $n = 1, 2, \cdots$.

LEMMA 4. *Let* $1 \leqq p < \infty$, $f \in L^p(0, \pi)$ *and* $a(x) = \int_0^\pi f(t)\cos xt\, dt$ so *that*

$$f \sim a_n = a(n) .$$

Let

$$g(x) = \int_x^\pi \frac{1}{t}\psi\left(\frac{x}{t}\right) f(t)dt \qquad c_n = \frac{1}{n}\int_0^n \psi\left(\frac{x}{n}\right) a(x)dx .$$

Then $g \in L^p(0, \pi)$ and

$$g \sim c_n .$$

Proof. Since $|g(x)| \leqq \psi(1)\int_x^\pi \frac{|f(t)|}{t} dt$ it follows from the proof in [1] that $g \in L^p$. Also

$$\int_0^\pi g(x)\cos nx\,dx = \int_0^\pi \cos nx\,dx \int_x^\pi \frac{1}{t}\psi\Big(\frac{x}{t}\Big) f(t)dt$$
$$= \int_0^\pi \frac{1}{t} f(t)dt \int_0^t \psi\Big(\frac{x}{t}\Big)\cos nx\,dx = \int_0^\pi f(t)dt \int_0^1 \psi(x)\cos nxt\,dt$$
$$= \frac{1}{n}\int_0^\pi f(t)dt \int_0^n \psi\Big(\frac{x}{n}\Big)\cos xt\,dt = \frac{1}{n}\int_0^n \psi\Big(\frac{x}{n}\Big)\int_0^\pi f(t)\cos xt\,dt = c_n\,.$$

The changes in order of integration are valid since

$$\int_0^\pi |f(t)|dt \int_0^1 |\psi(x)\cos nxt|dx \leqq \psi(1)\int_0^\pi |f(t)|dt < \infty\,.$$

(Note $f \in L'(0, \pi)$ since $f \in L^p(0, \pi)$.) Thus $g \sim c_n$, which is what we wished to show.

We can now establish our principal result.

Proof of Theorem 1*A*. Let $f \in L^p(0, \pi)$ be such that $f \sim a_n$ and let $a(x), g(x), c_n$ be as in Lemma 4. Then

$$b_n = \frac{1}{n}\sum_{m=1}^{n}\psi\Big(\frac{m}{n}\Big)a_m = \frac{1}{n}\int_0^n \psi\Big(\frac{x}{n}\Big)a(x)d[x]$$

so that

$$c_n - b_n = \frac{1}{n}\int_0^n \psi\Big(\frac{x}{n}\Big)a(x)d(x-[x]) = \frac{1}{n}\int_0^n \psi\Big(\frac{x}{n}\Big)d(x-[x])\int_0^\pi f(t)\cos xt\,dt$$
$$= \frac{1}{n}\int_0^\pi f(t)dt\int_0^n \psi\Big(\frac{x}{n}\Big)\cos xt\,d(x-[x])\,.$$

The last iterated integral clearly converges absolutely, justifying the change in order of integration. By Lemma 3 $c_n - b_n$ is a p-sequence. Also c_n is a p-sequence since, by Lemma 4, $g \in L^p(0, \pi)$ and $g \sim c_n$. Hence $b_n = c_n - (c_n - b_n)$ is a p-sequence and the theorem is proved.

REMARK. Note that except for the result of Lemma 1 the only properties of the cosine function used were its boundedness and the fact that $O\Big(\frac{1}{n}\Big)$ is a p-sequence for all $p \geqq 1$.

LEMMA 5. *Let* $C_t(x) = \int_0^x \sin yt\,d(y-[y])$. *Then there is an* $M > 0$ *such that*

$$|C_t(x)| \leqq M \qquad 0 \leqq t \leqq \pi;\ 0 \leqq x < \infty\,.$$

Proof. Let n be any non-negative integer. Then for $t > 0$

$$\int_0^n \sin yt\, dy = \frac{1}{t} - \frac{\cos nt}{t}$$

and

$$\int_0^n \sin yt\, d[y] = \sum_{k=1}^{n} \sin kt = \frac{\cos t/2 - \cos(n+1/2)t}{2\sin t/2}\,.$$

Hence

$$\begin{aligned} C_t(n) &= \frac{1}{t} - \frac{\cos nt}{t} - \frac{\cos t/2 - \cos(n+1/2)t}{2\sin t/2} \\ &= (1-\cos nt)\left(\frac{1}{t} - \frac{1}{2}\cot\frac{t}{2}\right) - \frac{\sin nt}{2}\,. \end{aligned}$$

The remainder of the proof follows as in Lemma 1.

In view of Lemma 5 and the remark preceding it the exact analogue of Theorem 1 for sine coefficients must hold. This we now state:

THEOREM 2. *Fix $p \geqq 1$. If, for some $f \in L^p$,*

$$a_n = \int_0^\pi f(t)\sin ntdt \qquad n = 1, 2, \cdots,$$

and if $b_n = \frac{1}{n}\sum_{m=1}^{n} \psi\left(\frac{m}{n}\right)a_m$ where $\psi(x)$ is of bounded variation on $0 \leqq x \leqq 1$ then there exists $g \in L^p$ such that

$$b_n = \int_0^\pi g(t)\sin ntdt \qquad n = 1\,2, \cdots$$

REFERENCES

1. G. H. Hardy, *Notes on some points in the integral calculus*, Messenger of Mathematics **58** (1929), 50-52.

2. A. Zygmund, Trigonometrical Series, Warsaw 1935 p. 190.

NORTHWESTERN UNIVERSITY

RANGES AND INVERSES OF PERTURBED LINEAR OPERATORS

SEYMOUR GOLDBERG

1. Introduction. Let X and Y denote normed linear spaces and let $T \neq 0$ be a linear operator with domain $D(T) \subset X$ and range $R(T) \supset Y$. In this paper, $D(T)$ is not required to be dense in X and T need not be continuous. Furthermore, X and Y shall be assumed complete only when necessary. Under these general conditions, we investigate some invariant properties of the range and inverse of T when T is perturbed by a bounded linear operator A. For example, it is shown that if the range of T is not dense in Y and T has a bounded inverse, then $T+A$ has the same properties provided that $D(A) \supset D(T)$ and the norm of A is sufficiently small. In addition, a theorem of Yood ([5], Th. 2.1) is generalized with some of the proofs simplified.

DEFINITION. Let $X_1 = \overline{D(T)} \subset X$. When X_1 is considered as a normed linear space, the conjugate transformation T' is defined as follows: Its domain $D(T')$ consists of the set of all y' in the conjugate space Y' for which $y'T$ is continuous on $D(T)$; for such a y' we define $T'y' = x'$ where x' is the unique bounded linear extension of $y'T$ to X_1; that is, x' is in the conjugate space X_1' of X_1.

The above notations shall be retained throughout the discussion.

2. Ranges and inverses of $T + A$.

LEMMA 1. *If T has a bounded inverse, then so does $T + A$ whenever $\|A\| < \|T^{-1}\|^{-1}$.*

Proof. $\|(T + A)x\| > (\|T^{-1}\|^{-1} - \|A\|)\|x\|$.

THEOREM 1. *If $\overline{R(T)} = Y$ and T has a bounded inverse, then $\overline{R(T + A)} = Y$ and $T + A$ has a bounded inverse whenever $\|A\| < \|T^{-1}\|^{-1}$ and $D(T) \subset D(A)$.*

Proof. By [4] Th. 1.4, $(T')^{-1} = (T^{-1})'$ exists and is continuous on X_1'. Hence from the lemma we conclude that $(T + A)' = T' + A'$ has a bounded inverse since $\|A'\| = \|A\| < \|T^{-1}\|^{-1} = \|(T')^{-1}\|^{-1}$. The theorem now follows from [4] Th. 1.2.

If for $X = Y$, the resolvent of a linear operator T is defined as the set of scalars λ such that $(T - \lambda I)^{-1}$ exists and is continuous on a

Received February 19, 1959.

domain dense in X, then the following corollary is an immediate result of the theorem.

COROLLARY. *The resolvent of a linear operator is open.*

DEFINITION. For each $z \neq 0$ in Y, let

$$m_z(T) = \sup \{k / \|z - Tx\| \geqq k \|Tx\|, x \in D(T)\} .$$

We define $m(T) = \sup_{0 \neq z \in Y} m_z(T)$.

REMARK. $m(T) \leqq 1$; This follows from the fact that for $Tx \neq 0$ and for each $z \in Y$, $\|z - T\alpha x\| / \|T\alpha x\| \leqq 1 + \|z\| / \|T\alpha x\| \to 1$ as $|\alpha| \to \infty$.

LEMMA 2. *Let Y be complete. Then $\overline{R(T)} = Y$ if and only if $m(T) = 0$.*

Proof. If $\overline{R(T)} = Y$, it is easy to see that $m(T) = 0$. Suppose there exists an element $y_0 \in Y$ which is not in $\overline{R(T)}$. The 1-dimensional linear manifold $[y_0]$ spanned by y_0 and the linear manifold $[y_0] + \overline{R(T)}$ are closed in Y; moreover, $[y_0] \cap \overline{R(T)} = (0)$. Hence by [2] Th. 2.1, there exists a $k > 0$ such that $\|y_0 - y\| \geqq k\|y\|$ for all $y \in R(T)$; that is, $m(T) > 0$.

THEOREM 2. *If $\overline{R(T)} \neq Y$ and T has a bounded inverse, then $\overline{R(T + A)} \neq Y$ and $T + A$ has a bounded inverse whenever $\|A\| < m(T)/\|T^{-1}\|$, and $D(T) \subset D(A)$.*

Proof. Clearly there is no loss of generality if the theorem is proved for the completion $\tilde{Y}$ of Y. Thus it may be assumed that Y is complete. We now simplify and apply an argument given by Yood [5, p. 489]. From Lemma 1, $T + A$ has a bounded inverse. By Lemma 2, there exists, for each $\varepsilon > 0$, an element $y_0 \in Y$ but not in $\overline{R(T)}$ such that

$$(1) \qquad \|y_0 - Tx\| \geqq (m(T) - \varepsilon)\|Tx\| \text{ for all } x \in D(T) .$$

Suppose that the theorem is not true. Then $y_0 \in \overline{R(T + A)} = Y$ and thus we may choose an element $x \in D(T)$ so that

$$\|(T + A)x - y_0\| < \min (\varepsilon d, \|y_0\|) ,$$

where d is the distance between y_0 and $\overline{R(T)}$. In particular,

$$(2) \qquad \|(T + A)x - y_0\| < \varepsilon d \leqq \varepsilon\|y_0 - Tx\| \text{ and } x \neq 0 .$$

From (1) and (2),

$$\begin{aligned}\|A\|\|x\| \geqq \|Ax\| \geqq \|Tx - y_0\| - \|y_0 - (T + A)x\| > (1 - \varepsilon)\|y_0 - Tx\| \\ \geqq (1 - \varepsilon)(m(T) - \varepsilon)\|Tx\| \geqq \|T^{-1}\|^{-1}(1 - \varepsilon)(m(T) - \varepsilon)\|x\| .\end{aligned}$$

Since $\varepsilon > 0$ was arbitrary, $\|A\| \geqq \|T^{-1}\|^{-1}(m(T))$ which is impossible.

LEMMA 3. *Suppose X and Y are complete. If T is a closed linear operator, then $R(T) = Y$ and T^{-1} does not exist if and only if $\overline{R(T')} \neq X_1'$ and T' has a bounded inverse.*

Proof. This follows from the "state diagram" for closed operators [1].

THEOREM 3. *Suppose X and Y are complete. If T is closed, $R(T) = Y$ and T^{-1} does not exist, then $R(T + A) = Y$ and $(T + A)^{-1}$ does not exist whenever $D(T) \subset D(A)$ and $A < m(T')/\|(T')^{-1}\|$.*

Proof. By Lemma 3, $\overline{R(T')} \neq X_1'$ and T' has a bounded inverse. Futhermore, $D(A') = Y' \supset D(T')$ and $T' \neq 0$ since $D(T')$ is total ([4] Th. 1.1). From Theorem 2, it is clear that $\overline{R(T' + A')} \neq X_1'$ and $T' + A'$ has a bounded inverse. Since $T' + A' = (T + A)'$ and $T + A$ is closed, the theorem follows from Lemma 3.

3. A generalization of a theorem. In ([5] Th. 2.1), Yood proves a theorem about the range of a bounded linear transformation T and its conjugate T', where T maps Banach Space X into Banach space Y. We now generalize the theorem by requiring instead that T be a closed linear operator on $D(T)$. The results are stated in a different but more precise form than in [5].

DEFINITION. If T has a bounded inverse, let $K(T) = \|T^{-1}\|$, otherwise let $K(T) = 0$. We now define a number $\alpha(T)$ as follows:

$$\begin{aligned}\alpha(T) &= \min\left((m(T), \frac{m(T)}{K(T)}\right) \text{ if } m(T) > 0 \\ &= \infty \text{ if } m(T) = 0 .\end{aligned}$$

$\alpha(T')$ shall be defined in a similar manner.

THEOREM 4. *Suppose X and Y are complete. Let T be a closed linear transformation and let A represent a bounded linear transform-*

ation such that $D(A) \supset D(T)$. Then the following statements concerning T are equivalent.

(1) *Either T has bounded inverse or $R(T) = Y$.*
(2) $\overline{R(T' + A')} \subset R(T')$ *if* $\|A\| < \alpha(T')$.
(3) $R(T' + A') \subset R(T')$ *if* $\|A\| < \alpha(T')$.
(4) $R(T')$ *is not a proper dense subset of* X_1' *and* $\|A\| < \alpha(T')$ *implies that* $\overline{R(T' + A')} \subset R(T')$.
(5) $R(T')$ *is not a proper dense of* X_1' *and* $\|A\| < \alpha(T')$ *implies that* $R(T' + A') \subset R(T')$.
(6) $\overline{R(T + A)} \subset R(T)$ *if* $\|A\| < \alpha(T)$.
(7) $R(T + A) \subset R(T)$ *if* $\|A\| < \alpha(T)$.
(8) $R(T)$ *is not a proper dense subset of* Y *and* $\|A\| < \alpha(T)$ *implies that* $\overline{R(T + A)} \subset R(T)$.
(9) $R(T)$ *is not a proper dense subset of* Y *and* $\|A\| < \alpha(T)$ *implies that* $R(T + A) \subset R(T)$.

Proof. (1) implies (2): (T need not be closed): If T has a bounded inverse, then by [1] $R(T') = X_1' \supset R(T' + A')$ for all A. If T has no bounded inverse, then $R(T) = Y$ so that $R(T') \neq X_1'$ and T' has a bounded inverse by [1]. Since T' is closed, it follows that $R(T')$ is closed; i.e. $m(T') > 0$. If (2) is false, there exists an $x_0' \in \overline{R(T' + A')}$ but at a positive distance d from $R(T')$. By the argument as in Theorem 2, $\|A\| = \|A'\| \geqq \dfrac{m(T')}{K(T')} \geqq \alpha(T') > \|A\|$ which is impossible.

(2) implies (3). Obvious

(3) implies (1): (cf. [5]): If $R(T) \neq Y$ and T has no bounded inverse, then we show that (3) fails to hold. By [1], $R(T') \neq X_1'$ and T' has no bounded inverse. Therefore, we may choose an element $x_0' \in X_1'$, $\|x_0'\| = 1$ and $x_0' \notin R(T')$. For each $\varepsilon > 0$, there exists an element $y_0' \in D(T')$ such that $\|y_0'\| = 1$, $\|T'y_0'\| < \varepsilon$ and an element y_0 such that $\|y_0\| = 1$, $y_0'y_0 = \beta$ is real and $1 \geqq \beta \geqq 1/2$. Let A be defined by $Ax = \varepsilon(x_0'x - (\varepsilon\beta)^{-1}T'y_0'x)\, y_0$ for $x \in D(T)$. Hence

$$A'y_0' = \varepsilon y_0'y_0(x_0' - (\varepsilon\beta)^{-1}T'y_0') = \varepsilon\beta x_0' - T'y_0' ,$$

so that

$$(T' + A')y_0' = \varepsilon\beta x_0' \notin R(T'). \quad \text{Moreover, } \|A\| \leqq \varepsilon\Big(1 + \frac{1}{\beta}\Big) \leqq 3\varepsilon .$$

Since $\varepsilon > 0$ was arbitrary, it follows that (3) does not hold.

(4) and (5) implies (1): Follows from the above argument.

(1) implies (4) and (5): (T need not be closed): This follows from the fact that

(1) implies that $R(T')$ is closed and also that (1) implies (2).

(1) implies (6): If $R(T) = Y$, then (6) is satisfied. Suppose $R(T) \neq Y$ but that T has a bounded inverse. Hence $R(T)$ is closed so that $m(T) > 0$. If (6) is false, there exists an element $y_0 \varepsilon Y = \overline{R(T + A)}$ but $y_0 \notin R(T)$. The remaining argument is now as in Theorem 2.

(6) implies (7): Obvious

(7) implies (1): If $R(T) \neq Y$ and T has no bounded inverse, then for $\varepsilon > 0$, there exists an element $x_0 \in D(T)$, $\|x_0\| = 1$ such that $\|Tx_0\| < \varepsilon$. An element $x_0' \in X_1'$ is chosen so that $\|x_0'\| = 1$ and $x_0'x_0 = 1$. Suppose that $y \notin R(T)$ and $\|y\| = 1$. We define A by the relation

$$Ax = \varepsilon x_0'x(y - \varepsilon^{-1}Tx_0),\ x \in D(T)\ .$$

Then $(T + A)x_0 = \varepsilon y \notin R(T)$. Moreover, $\|A\| < 2\varepsilon$. Since $\varepsilon > 0$ is arbitrary, (7) cannot hold. Thus the assertion is proved.

(8) and (9) are equivalent to (1): This is shown in the same way that (4) and (5) were shown equivalent to (1).

If there is no restriction put on the inverse but only on the range of T, we may still infer something about the range of $T + A$. In fact, A need not be continuous. The following theorem illustrates this.

THEOREM 5. *Suppose X and Y are complete. If T is a closed linear operator with a closed range, then there exists a $\rho > 0$ such that $T + A$ is also a closed linear operator with a closed range whenever A is a linear operator (not necessarily continuous) with $D(A) \supset D(T)$ and $\|Ax\| \leqq \rho(\|x\| + \|Tx\|)$ for every $x \in D(T)$.*

Proof. We introduce another norm $\|x\|_1$, on $D(T)$ by defining $\|x\|_1, = \|x\| + \|Tx\|$. D_1 shall denote $D(T)$ with this new norm. Since X and Y are complete and T is closed, it is easy to see that D_1 is a complete normed linear space. Moreover, T_1 as a transformation of D_1 into Y is bounded and has an inverse. Thus by the closed graph theorem, T^{-1} is bounded; i.e. there exists an $m > 0$ such that $\|Tx\| \geqq m(\|x\| + \|Tx\|)$ for $x \in D_1$. Choose $\rho > 0$ so that $1 > \rho$ and $m - \rho > 0$. Thus $\|(T + A)x\| \geqq (m - \rho)(\|x\| + \|Tx\|)$, whence $T + A$ has a bounded inverse from $R(T + A)$ onto D_1. Clearly $T + A$ is continuous on D_1. Since defining a new norm in $D(T)$ does not alter the situation in Y, it follows that $R(T + A)$ is closed. In [3], Nagy proves that $T + A$ is a closed operator from $D(T)$ into Y, which completes the proof of the theorem.

BIBLIOGRAPHY

1. S. Goldberg, *Linear operators and their conjugates*, Pacific J. Math., **9** (1959), 69-79.

2. E. R. Lorch, *On a calculus of operators in reflexive vector spaces*, Trans. Amer. Math Soc., **45** (1939), 217-234.
3. B. Sz. Nagy, *On the stability of the index of unbounded linear transformations*, Acta Math. Sci. Hungaricae, **3** (1952), 49-52.
4. R. S. Phillips, *The adjoint semi-group*, Pacific J. Math., **5** (1955), 269-282.
5. B. Yood, *Transformations between Banach Spaces in the uniform topology*, Ann. of Math., **50** (1949), 486-503.

HEBREW UNIVERSITY

ON FUNCTIONS REPRESENTABLE AS A DIFFERENCE OF CONVEX FUNCTIONS

PHILIP HARTMAN

1. Introduction. A function $f(x)$ defined on a convex x-set D will be called a d.c. function on D if there exists a pair of convex functions $F_1(x)$, $F_2(x)$ on D such that $f(x)$ is the difference

$$f(x) = F_1(x) - F_2(x) \, . \tag{1}$$

In this note, "convex" function means "continuous and convex" function. D.c. functions have been considered, for example, by Alexandroff [1]. E. G. Straus mentioned them in a lecture in Professor Beckenbach's seminar (and used the abbreviation "d.c.").

When x is a real variable, so that D is a (bounded or unbounded) interval, then $f(x)$ is a d.c. function if and only if f has left and right derivatives (where these are meaningful) and these derivatives are of bounded variation on every closed bounded interval interior to D. Straus remarked that this fact implies that if $f_1(x), f_2(x)$ are d.c. functions of a real variable, then so are the product $f_1(x)f_2(x)$, the quotient $f_1(x)/f_2(x)$ when $f_2(x) \neq 0$, and the composite $f_1(f_2(x))$ under suitable conditions on f_2. He raised the question whether or not this remark can be extended to cases where x is a variable on a more general space. The object of this note is to give an affirmative answer to this question if x is a point in a finite dimensional (Euclidean) space.

2. Local d.c. functions. Let $f(x)$ be defined on a convex x-set D. The function $f(x)$ will be said to be d.c. at a point x_0 of D if there exists a convex neighborhood U of x_0 such that $f(x)$ is d.c. on $U \cap D$. When $f(x)$ is d.c. at every point x of D, it will be said to be locally d.c. on D.

(I) *Let D be a convex set in an m-dimensional Euclidean x-space and let D be either open or closed. Let $f(x)$ be locally* d.c. *on D. Then $f(x)$ is* d.c. *on D.*

While the proof of (I) cannot be generalized to the case where the m-dimensional x-space is replaced by a more general linear space, it will be clear that (II), below remains valid if the Euclidean x-space (but not the y-space) is replaced by a more general space.

Received February 10, 1959. This research was supported by the United States Air Force through the Air Force Office of Scientific Research of the Air Research and Development Command under contract No. AF 18 (603)-41. Reproduction in whole or in part is permitted for any purpose of the United States Government.

(II) *Let* $x = (x^1, \cdots, x^m)$ *and* $y = (y^1, \cdots, y^n)$. *Let D and E be convex sets in the x- and y-spaces, respectively; let D be either open or closed and let E be open. Let $g(y)$ be a* d.c. *function on E and let $y^j = y^j(x)$ where $j = 1, \cdots, n$, be* d.c. *functions on D such that $y = y(x) \in E$ for $x \in D$. Then $f(x) = g(y(x))$ is locally* d.c. *on D.*

This theorem is false (even for $n = m = 1$) if the assumption that E is open is omitted. In order to see this, let x and y be scalars, $g(y) = 1 - y^{1/2}$ ($\leqq 1$) on E: $0 \leqq y < 1$ and let $y = y(x) = |x - \frac{1}{2}|$ on $D: 0 < x < 1$. Since $f(x) = g(y(x)) = 1 - |x - \frac{1}{2}|^{1/2}$ does not have finite left and right derivates at the interior point $x = \frac{1}{2}$, the function $f(x)$ is not d.c. at the point $x = \frac{1}{2}$.

It will be clear from the proof that (II) remains correct if the assumption that E is open is replaced by the following assumption on E and $g(y)$: if x_0 is any point of D and $y_0 = y(x_0)$, let there exists a convex y-neighborhood V of y_0 such that $g(y)$ satisfies a uniform Lipschitz condition on $V \cap E$. (This condition is always satisfied if y_0 is an interior point of E; cf., e.g., Lemma 3 below).

COROLLARY. *Let D be either an open or a closed convex set in the $(x^1, \cdots, x^m)$-space. Let $f_1(x), f_2(x)$ be* d.c. *functions on D. Then the product $f_1(x)f_2(x)$ and, if $f_1(x) \neq 0$, the quotient $f_2(x)/f_1(x)$ are* d.c. *functions on D.*

The assertion concerning the product follows from (I) and (II) by choosing y to be a binary vector $y = (y^1, y^2)$, $g(y) = y^1y^2$, E the (y^1, y^2)-plane and $y^1 = f_1(x)$, $y^2 = f_2(x)$. Thus $f(x) \equiv g(y(x)) = f_1(x)f_2(x)$. Note that $g(y) = \frac{1}{2}(y^1 + y^2)^2 - \frac{1}{2}((y^1)^2 + (y^2)^2)$ is a d.c. function on E.

In the assertion concerning the quotient, it can be supposed that $f_2(x) \equiv 1$ and that $f_1(x) > 0$. Let y be a scalar, $g(y) \equiv 1/y$ on $E: y > 0$ and $y = f_1(x)$ on D. Thus $g(y)$ is convex on E and $f(x) \equiv g(y(x)) = 1/f_1(x)$.

3. Preliminary lemmas.

It will be convenient to state some simple lemmas before proceeding to the proofs of (I) and (II). The proofs of these lemmas will be indicated for the sake of completeness.

In what follows, $x = (x^1, \cdots, x^m)$ is an m-dimensional Euclidean vector and $|x|$ is its length. D is a convex set in the x-space.

LEMMA 1. *Let D be either an open or a closed convex set having interior points. Let $x = x_0$ be a point of D and U a convex neighborhood of x_0. Let $F(x)$ be a convex function on $D \cap U$. Then there exists a neighborhood U_1 of x_0 and a function $F_1(x)$ defined and convex on D such that $F(x) \equiv F_1(x)$ on $D \cap U_1$.*

In order to see this, let U_2 be a small sphere $|x - x_0| < r$ such that $F(x)$ is bounded on the closure of $D \cap U_2$. Let $G(x) = K|x - x_0| +$

$F(x_0) - 1$, where K is a positive constant, chosen so large that $G(x) > F(x) + 1 > F(x)$ on the portion of the boundary of U_2 interior to D. Clearly $G(x) < F(x)$ holds for $x = x_0$, hence, for x on $D \cap U_1$ if U_1 is a suitably chosen neighborhood of x_0. If $x \in D$, define $F_1(x)$ to be $\max(F(x), G(x))$ or $G(x)$ according as x is or is not in $U_2 \cap D$. Since $\max(F(x), G(x))$ is convex on $U_2 \cap D$ and $\max(F(x), G(x)) \equiv G(x)$ for x in a vicinity (relative to D) of the boundary of U_2 in D, it follows that $F_1(x)$ is convex on D. Finally, $F_1(x) \equiv \max(F(x), G(x)) = F(x)$ for $x \in U_1 \cap D$.

LEMMA 2. *Let D be a closed, bounded convex set having $x = 0$ as an interior point. There exists a function $h(x)$ defined and convex for all x such that $h(x) \leqq 1$ or $h(x) > 1$ according as $x \in D$ or $x \notin D$.*

In fact, $h(x)$ can also be chosen so as to satisfy $h(x) > 0$ for $x \neq 0$ and $h(cx) = ch(x)$ for $c > 0$. This function is then the supporting function of the polar convex set of D; Minkowski, cf. [2], § 4. The function $h(x)$ is given by 0 or $|x|\rho^{-1}(x/|x|)$ according as $x = 0$ or $x \neq 0$, where, if u is a unit vector, $\rho(u)$ is the distance from $x = 0$ to the point where the ray $x = tu, t > 0$ meets the boundary of D.

LEMMA 3. *Let D be a closed, bounded convex set having interior points and D_1 a closed convex set interior to D. Let $F(x)$ be a convex function on D. Then $F(x)$ satisfies a uniform Lipschitz condition on D_1.*

In fact, if $d > 0$ is the distance between the boundaries of D and D_1 and if $|F(x)| \leqq M$ on D, then $|F(x_1) - F(x_2)| \leqq 2M|x_1 - x_2|/d$ for $x_1, x_2 \in D_1$. This inequality follows from the fact that $F(x)$ is convex on the intersection of D and the line through x_1 and x_2.

4. *Proof of* (I). The proof will be given for the case of an open convex set D. It will be clear from the proof and from Lemma 1 how the proof should be modified for the case of a closed D.

To every point x_0 of D, there is a neighborhood $U = U(x_0)$, say U: $|x - x_0| < r(x_0)$, contained in D such that $f(x)$ is d.c. on U; that is, there exists a convex function $F(x) = F(x, x_0)$ such that $f(x) + F(x, x_0)$ is a convex function of x on $U(x_0)$. In view of Lemma 1, it can be supposed (by decreasing $r(x_0)$, if necessary) that $F(x, x_0)$ is defined and convex on D (although, of course, $f + F$ is convex only on U).

Let D_1 be a compact, convex subset of D. Then D_1 can be covered by a finite number of the neighborhoods $U(x_1), \cdots, U(x_k)$. Put $F(x) = F(x, x_1) + \cdots + F(x, x_k)$, so that $F(x)$ is defined and convex on D. Since $f(x) + F(x, x_j)$ is convex on $U(x_j)$, so is $f(x) + F(x) = f(x) + F(x, x_j) +$

$\sum_{i\neq j} F(x, x_i)$. Hence $f + F$ is convex on D_1.

Thus there exists a sequence of open, bounded convex sets $D_1, D_2, \cdots$ with the properties that the closure of D_j is contained in D_{j+i}, $D = \bigcup D_j$, and to each D_j there corresponds a function $F_j(x)$ defined and convex on D such that $f(x) + F_j(x)$ is convex on D_j.

Introduce a sequence of closed convex sets $C^1, C^2, \cdots$ such that $C^1 \subset D_1 \subset C^2 \subset D_2 \subset \cdots$. In particular, $D = \bigcup C^j$.

It will be shown that there is a function $G_1(x)$ with the properties that

(i) $G_1(x)$ is defined and convex on D,

(ii) $f(x) + G_1(x)$ is convex on D_2, and

(iii) $G_1(x) \equiv F_1(x)$ on C^1.

If this is granted for the moment, the proof of (I) can be completed as follows: If $G_1, \cdots, G_{k-1}$ have been constructed, let G_k be a function defined and convex on D such that $f + G_k$ is convex on D_{k+1} and $G_k \equiv G_{k-1}$ on C^k. Then $F(x) = \lim G_k(x)$ exists uniformly on compact subsets of D; in fact, $F(x) \equiv G_j(x)$ on C^k for all $j \geqq k$. Hence, $F(x)$ is defined and convex on D. Since $f(x) + F(x)$ is convex on C^k, $k = 1, 2, \cdots$, it is convex on D; that is, f is a d.c. function on D.

Thus, in order to complete the proof of (I), it remains to construct a $G_1(x)$ with the properties (i) – (iii). Let $k > 0$ be a constant so large that $F_2(x) - k \leqq F_1(x)$ for $x \in C^1$. Without loss of generality, it can be supposed that $x = 0$ is an interior point of C^1. Let $h(x)$ be the function given by Lemma 2 when D there is replaced by C^1. Put $H(x) = 0$ or $H(x) = K[h(x) - 1]$ according as $x \in C^1$ or $x \notin C^1$, where $K > 0$ is a constant. Thus $H(x)$ is defined and convex for all x and $H(x) \equiv 0$ on C^1. In particular,

$$(2) \qquad F_2(x) - k + H(x) \leqq F_1(x) \text{ for } x \in C^1 .$$

Choose K so large that

$$(3) \qquad F_2(x) - k + H(x) > F_1(x) \text{ on } D_1' ,$$

the boundary of D_1. This is possible since $h(x) - 1 > 0$ for $x \notin C^1$.

Define $G_1(x)$ as follows:

$$(4) \qquad \begin{aligned} G_1(x) &= \max\,(F_1(x), F_2(x) - k + H(x)) \text{ for } x \in D_1 , \\ G_1(x) &= F_2(x) - k + H(x) \text{ for } x \in D - D_1 , \end{aligned}$$

where $D - D_1$ is the set of points in D, not in D_1.

Clearly, (2) and the first part of (4) imply property (iii),

$$(5) \qquad G_1(x) = F_1(x) \text{ if } x \in C^1 ,$$

and (3) implies that

(6) $$G_1(x) = F_2(x) - k + H(x) \text{ for } x \text{ on and near } D_1',$$

the boundary of D_1.

By the first part of (4), $G_1(x)$ is convex on D_1. By the last part of (4) and by (6), $G_1(x)$ is convex in a vicinity of every point of $D - D_1$. Hence, $G_1(x)$ has property (i), that is $G_1(x)$ is convex on D.

Since $f(x) + F_1(x)$ is convex on D_1 and $f(x) + F_2(x)$, hence $f(x) + F_2(x) - k + H(x)$ is convex on $D_2 \supset D_1$, it follows that, on D_1, the function

$$f(x) + G_1(x) = \max(f + F_1, f + F_2 - k + H)$$

is convex. It also follows from the last part of (4) and from (6) that $f + G_1$ is convex in a vicinity of every point of $D_2 - D_1$. Hence G_1 has property (ii), that is, $f + G_1$ is convex on D_2. This completes the proof of (I).

5. *Proof of* (II). Without loss of generality, it can be supposed that $g(y)$ is convex on E.

Since $y^j(x)$ is a d.c. function on D, there exists a convex function $F(x)$ on D such that

(7) $$\pm\, y^j(x) + F(x) \text{ are convex on } D.$$

The function $F(x) = F(x, j, \pm)$ can be assumed to be independent of j, where $j = 1, \cdots, n$, and of $\pm$; for otherwise it can be replaced by $\sum_j F(x, j, +) + \sum_j F(x, j, -)$.

Let $x = x_0$ be a point of D and $y_0 = y(x_0)$. Let V be a convex neighborhood of y_0 such that g satisfies a uniform Lipschitz condition

(8) $$|g(y_1) - g(y_2)| \leqq M|y_1 - y_2|$$

on V; cf. Lemma 3. Let U be a neighborhood of x_0 such that $y(x) \in V$ for $x \in U \cap D$. It will be shown that

(9) $$f(x) + 3nMF(x) \text{ is convex on } D \cap U,$$

so that f is d.c. at $x = x_0$.

It is clear that there is no loss of generality in assuming that $g(y)$ has continuous partial derivatives satisfying

(10) $$|\partial g(y)/\partial y^j| \leqq M \text{ for } j = 1, \cdots, n \text{ and } y \in V.$$

For otherwise, g can be approximated by such functions.

In what follows, only x in $D \cap U$ and $y \in V$ occur. Let $x = x(s)$, where s is a real variable on some interval, be an arc-length parametrization of a line segment in $D \cap U$. The assertion (9) follows if it is shown that $e(s) + 3nMF(x(s))$, where $e(s) = f(x(s))$ is a convex function of s. It is clear that $e(s)$ has left and right derivatives (whenever these are meaningful). Let $e'(s)$ denote a left or a right derivate of $e(s)$ and $F'(x(s)), y^{j'}(x(s))$ the corresponding derivates of $F(x(s)), y^j(x(s))$. Let $\Delta s > 0$, then $e(s) + 3nMF(x(s))$ is convex if and only if $\Delta e' + 3nM\Delta F' \geqq 0$, where $\Delta e' = e'(s + \Delta s) - e'(s)$ and $\Delta F' = F'(x(s + \Delta s)) - F'(x(s))$.

By the definition of e,

$$e' = \sum (\partial g / \partial y^j) y^{j'} . \tag{11}$$

Hence,

$$\Delta e'(s) = \sum \Delta(\partial g/\partial y^j) y_0^{j'} + \sum (\partial g/\partial y^j)_1 \Delta_y^{j'} , \tag{12}$$

where $y_0^{j'} = y^{j'}(x(s))$ and $(\partial g/\partial y^j)_1$ is the value of $\partial g/\partial y^j$ at $y = y(x(s + \Delta s))$.

The usual proofs of the mean value theorem of differential calculus (via Rolle's theorem) imply the existence of a $\theta = \theta_j, 0 < \theta_j < 1$, such that

$$\Delta_y^j / \Delta s = y_\theta^{j'} , \tag{13}$$

where $y_\theta^{j'}$ is a number between the left and right derivates of $y^j(x(s))$ at the s-point $s + \theta_j \Delta s$. By (13), the equation (12) can be written as

$$\begin{aligned} \Delta e' = {} & \sum (\Delta \partial g/\partial y^j)(\Delta y^j/\Delta s) + \sum (\Delta \partial g/\partial y^j)(y_0^{j'} - y_\theta^{j'}) \\ & + \sum (\partial g/\partial y^j)_1 \Delta y^{j'}) . \end{aligned} \tag{14}$$

By (7),

$$|\Delta y^{j'}| \leqq \Delta F' \text{ and } |y_0^{j'} - y_\theta^{j'}| \leqq F_\theta' - F_0' \leqq \Delta F' ,$$

where F_θ' is the right derivate of $F(x(s))$ at the s-point $s + \theta_j \Delta s$ $(< s + \Delta s)$. Since $g(y)$ is convex, the first term on the right of (14) in non-negative. Hence (10) and (11) give

$$\Delta e' \geqq 0 - 2nM\Delta F' - Mn\Delta F' , \tag{15}$$

so that $e(s) + 3nMF(x(s))$ is convex. This proves (II).

6. "Minimal" convex functions.

Let $f(x)$ be d.c. on the unit sphere $|x| < 1$, so that there exist functions $F(x)$ on $|x| < 1$ such that

$$F(x) \text{ and } f(x) + F(x) \text{ are convex on } |x| < 1 . \tag{16}$$

The function $F(x)$ can be chosen so as to satisfy the normalization

$$F(0) = 0 \text{ and } F(x) \geqq 0 . \tag{17}$$

If x is a real variable, there exists a "least" $F(x)$, say $F_m(x)$, satisfying (16), (17) in the sense that (16), (17) hold for $F = F_m$ and (16), (17) imply

$$F_m(x) \leqq F(x) \text{ on } |x| < 1 . \tag{18}$$

In fact, $F_m(x)$ can be obtained as follows: A (left or right) derivative $f'(x)$ of $f(x)$ is of bounded variation on every interval $|x| \leqq a < 1$ and so $f'(x)$ can be written as $f'(x) = P(x) - N(x)$, where $P(x)$, $N(x)$ are the positive, negative variation of f' on the interval between 0 and x, say, with the normalization $N(0) = 0$. In particular, P and N are non-decreasing on $|x| < 1$. In this case, $F_m(x)$ is given by

$$F_m(x) = \int_0^x N(x)dx .$$

On the other hand, if x is a vector, there need not exist a least $F = F_m(x)$. In order to see this, let x be a binary vector and write (x, y) instead of x. Let $f(x, y) = xy$. If $\varepsilon > 0$, $F(x, y) = \frac{1}{2}(\varepsilon x^2 + y^2/\varepsilon)$ satisfies (16), (17). If a least $F = F_m$ exists, then $0 \leqq F_m(x, y) \leqq \frac{1}{2}(\varepsilon x^2 + y^2/\varepsilon)$. In particular, $0 \leqq F_m(x, 0) \leqq \varepsilon x^2$, and, therefore, $F_m(x, 0) \equiv 0$. Similarly, $F_m(0, y) \equiv 0$. But since F_m is convex, it follows that $F_m \equiv 0$. This contradicts the case $F = F_m$ of (16) and so, a least $F = F_m$ does no exist.

Although a "least" F need not exist, it follows from Zorn's lemma that "minimal" F's do exist.

References

1. A. D. Alexandroff, *Surfaces represented by the difference of convex functions*, Doklady Akademii Nauk SSSR, (N.S.) **72** (1950), 613-616.

2. T. Bonnesen and W. Fenchel, *Theorie der konvexen Korper*, New York (1948).

THE JOHNS HOPKINS UNIVERSITY

ON CONDITIONAL EXPECTATION AND QUASI-RINGS

M. V. JOHNS, JR. AND RONALD PYKE

1. Introduction. Let $(\Omega, \mathscr{A}, P)$ denote a complete probability space in which Ω is an arbitrary point set $(\omega \in \Omega)$, $\mathscr{A}$ is a σ-algebra of subsets of Ω $(A \in \mathscr{A})$ and P is a probability measure on $\mathscr{A}$ with respect to which P is complete. Let X, Y, Z, with or without subscripts, denote real-valued $\mathscr{A}$-measurable random variables (r. v.) Let $\mathscr{E}$ denote the space of P-integrable r. v.'s. Define a linear operator E on $\mathscr{E}$ by

$$E \circ X = \int_\Omega X \, dP .$$

E is the expectation operator and $E \circ X$ is called the expectation of X. The P-integrability criterion is equivalent to specifying $E \circ |X| < \infty$. Let $\mathscr{F}$, with or without subscripts, denote a complete σ-algebra contained in $\mathscr{A}$, and let $\mathscr{B}_k$ denote the σ-algebra of Borel sets of k-dimensional Euclidean space. For r.v.'s. $i=1, X_i, \cdots, k$, define $\mathscr{B}(X_1, \cdots, X_k) \subset \mathscr{A}$ as the minimal complete σ-algebra containing all inverse images with respect to the vector $(X_1, \cdots, X_k)$ of sets in $\mathscr{B}_k$. For $A \in \mathscr{A}$, let $I_A \in \mathscr{E}$ denote the indicator function of the set A; that is, $I_A(\omega)=1$ or 0 according as $\omega \in A$ or $\omega \notin A$. For $X \in \mathscr{E}$, define the completely-additive set function $Q_X : \mathscr{A} \to R_1$ by $Q_X(A) = E \circ XI_A$.

By the Radon-Nikodym Theorem there exists for $X \in \mathscr{E}$ and $\mathscr{F} \subset \mathscr{A}$, an $\mathscr{F}$-measurable solution $Y \in \mathscr{E}$ to the system of equations

$$E \circ (X - Y)I_A = 0 \qquad (A \in \mathscr{F}) \tag{1}$$

or equivalently

$$Q_X(A) = E \circ YI_A \qquad (A \in \mathscr{F}) .$$

This solution is unique a. s. (relative to the restriction of P to $\mathscr{F}$). The equivalence class of all such solutions (or any representative thereof) is denoted by $E\{X|\mathscr{F}\}$ and called the conditional expectation of X given $\mathscr{F}$. For $X, Y \in \mathscr{E}$ the notation $E\{X|Y\} = E\{X|\mathscr{B}(Y)\}$ will also be used. This definition of conditional expectation, which is the standard one, makes it necessary when proving theorems about conditional expectations to show at some stage of the proof that a functional equation of the form (1) is valid for all subsets of a specified σ-algebra. That this can be a tedious task is demonstrated by the existing proofs of some of the applications in § 4 of the theorems which are proved below.

Received December 23, 1958. This work was supported by the Office of Naval Research under Contract Nonr-225 (21).

It is the purpose of this note to define conditional expectations in an apparently less restrictive way, by narrowing the class of subsets A for which (1) must hold. It is shown that this definition is, nevertheless, equivalent to that given in the above paragraph. In § 3, some general theorems on conditional expectations are proved using this second definition. The proofs of these theorems are seen to be simpler and shorter than would be possible with conventional techniques. Besides serving to demonstrate the convenience of this second definition, these theorems are important in themselves and several applications of them are given.

The main tool to be used is the concept of a quasi-ring to be introduced and studied in the following section.

2. Conditional expectation given a quasi-ring. Von Neumann [5] defines a half-ring as a family of subsets closed under finite intersections and satisfying a certain finite chain condition. This same concept is termed a semi-ring by Halmos [3]. The related concept of quasi-ring, which is now defined, entails a weaker chain condition. This chain condition, (ii) of Definition 1 below, seems to be much more adaptable than that of von Neumann to problems in conditional expectation, as is demonstrated in § 3.

DEFINITION 1. A collection, $\mathscr{S}$, of subsets of Ω is said to be a *quasi-ring* if and only if

(i) $A, B \in \mathscr{S}$ implies $A \cap B \in \mathscr{S}$;

(ii) $A, B \in \mathscr{S}$ and $A \subset B$ implies that there exists $\{C_j\}_{j=1}^{n} \subset \mathscr{S}$ satisfying $C_i \cap C_k = \phi$ for $i \neq k$ and $B - A = C_1 \cup C_2 \cup \cdots \cup C_n$;

(iii) there exists $\{A_j\}_{j=1}^{\infty} \subset \mathscr{S}$ such that $\Omega = \bigcup_{j=1}^{\infty} A_j$.

In von Neumann's definition of a half-ring, condition (ii) is strengthened by requiring further that $A \cup C_1 \cup \cdots \cup C_j \in \mathscr{S}$ for all $j = 1, 2, \cdots, n$.

Examples of quasi-rings are: any countable class of disjoint sets which include the null set ϕ; in particular, the collection of atoms in an atomic, or discrete, probability space; any algebra or σ-algebra; the class of all left-open, right-closed rectangles in R_n with Lebesgue measure less than or equal to 1. This last example is a quasi-ring which is not a half-ring. Bell makes use of the half-ring analogous to this quasi-ring in his recent paper [1]. A closure property of quasi-rings that will be used in the following sections is given by

LEMMA 1. *If $\mathscr{S}_1$ and $\mathscr{S}_2$ are quasi-rings on a common space Ω then*

$$(2) \qquad \mathscr{S} = \mathscr{S}_1 \overset{*}{\cap} \mathscr{S}_2 \equiv \{A \cap B;\ A \in \mathscr{S}_1, B \in \mathscr{S}_2\}$$

is also a quasi-ring. (In common terminology $\mathscr{S}$ is the family of constituents of $\mathscr{S}_1$ and $\mathscr{S}_2$.)

Proof. Clearly $\mathscr{S}$ satisfies (i) of Definition 1. Moreover, let $A_i \in \mathscr{S}_1$ and $B_i \in \mathscr{S}_2 (i = 1, 2)$. If $A_1 \cap B_1 \subset A_2 \cap B_2$, then

$$(3) \qquad S \equiv (A_2 \cap B_2) - (A_1 \cap B_1) = [(A_2 - A_1) \cap (B_2 \cap B_1)] \cup [(B_2 - B_1) \cap A_2] ,$$

the two terms of the union being disjoint. By hypothesis there exist sequences $\{C_j\}_{j=1}^{n} \in \mathscr{S}_1$, $\{D_k\}_{k=1}^{m} \in \mathscr{S}_2$ satisfying

$$A_2 - A_2 \cap A_1 = \bigcup_{j=1}^{n} C_j, \quad B_2 - B_2 \cap B_1 = \bigcup_{k=1}^{m} D_k$$

and hence by (3), S has the representation

$$S = \bigcup_{j=1}^{x^n} (C_j \cap [B_2 \cap B_1]) \cup \bigcup_{k=1}^{m} (D_k \cap A_2)$$

all terms being disjoint. That $\mathscr{S}$ satisfies condition (iii) is seen by considering the collection of all pairwise intersections between elements of the respective sequences for $\mathscr{S}_1$ and $\mathscr{S}_2$ which satisfy (iii). Q. e. d.

An extension theorem for measures defined on a quasi-ring will now be given. The proof of the theorem is analogous to those of the more classical extension theorems and so will be omitted (e. g., cf. [5]).

For an arbitrary class $\mathscr{C}$ of subsets of Ω let $\sigma(\mathscr{C})$ denote the minimal σ-algebra containing $\mathscr{C}$.

THEOREM 1. *Let μ be a σ-finite completely additive set function defined on a quasi-ring $\mathscr{S}$. There exists a unique completely additive set function μ^* defined on $\sigma(\mathscr{S})$ such that for all $A \in \mathscr{S}$, $\mu^*(A) = \mu(A)$.*

In the event that there exists a finite family satisfying (iii) of Definition 1, the minimal algebra containing $\mathscr{S}$ is the collection of all finite unions of members of $\mathscr{S}$. After extending μ to this minimal algebra, Theorem 1 reduces in this case to a well known extension theorem (cf. Doob [2], p. 605).

DEFINITION 2. Let $X \in \mathscr{E}$ and $\mathscr{S} \subset \mathscr{A}$ where $\mathscr{S}$ is a quasi-ring. The class (or any representative thereof) of all $\sigma(\mathscr{S})$-measurable $Y \in \mathscr{E}$ satisfying the system of equations

$$(4) \qquad E \circ (X - Y) I_A = 0 \qquad (A \in \mathscr{S})$$

will be denoted by $E\{X \mid \mathscr{S}\}$, and called the conditional expectation of X given $\mathscr{S}$.

As a corollary to Theorem 1, one immediately obtains

THEOREM 2. *For $X \in \mathscr{E}$ and $\mathscr{S} \subset \mathscr{A}$*

$$E\{X \mid \mathscr{S}\} = E\{X \mid \sigma(\mathscr{S})\} \qquad \text{a. s.}$$

3. Some general theorems on conditional expectation. The following definition will be used:

DEFINITION 3. Quasi-rings $\mathscr{S}_1$ and $\mathscr{S}_2$ are said to be *conditionally independent* given a quasi-ring $\mathscr{S}$ (to be abbreviated as c. i. $\mid \mathscr{S}$) if and only if for all $A \in \mathscr{S}_1, B \in \mathscr{S}_2$,

$$(5) \qquad E\{I_A I_B \mid \mathscr{S}\} = E\{I_A \mid \mathscr{S}\} E\{I_B \mid \mathscr{S}\} \qquad \text{a. s.}$$

X and Y are said to be c. i. $\mid \mathscr{S}$ if and only if $\mathscr{B}(X)$ and $\mathscr{B}(Y)$ are c. i. $\mid \mathscr{S}$ (cf. Loève [4], p. 351).

The obvious notational changes are made in defining conditional independence given a r. v. If $\mathscr{S}_1$ and $\mathscr{S}_2$ are c. i. $\mid \{\phi, \Omega\}$, they are of course, independent in the usual stochastic sense. The above definition of conditional independence is closely related to that for σ-algebras given in Loève [4], as is shown by the next lemma. For well known properties of conditional expectations used in the following proofs, the reader is referred to [4].

LEMMA 2. *For $\sigma(\mathscr{S}_1)$ and $\sigma(\mathscr{S}_2)$ to be* c. i. $\mid \sigma(\mathscr{S})$ *it is necessary and sufficient that $\mathscr{S}_1$ and $\mathscr{S}_2$ be* c. i. $\mid \mathscr{S}$.

Proof. The necessity of the condition is immediate. The proof of sufficiency is by transfinite induction. Let $\mathscr{L}_1$ denote the class of all countable unions of elements of $\mathscr{S}_1$. For all ordinals α less than or equal to the first uncountable ordinal, α_0 say, define recursively $\mathscr{L}_\alpha$ as the set of countable unions of differences of elements of $\mathscr{T}_\alpha \equiv \bigcup_{\beta<\alpha} \mathscr{L}_\beta$. It is well known that $\sigma(\mathscr{S}_1) = \mathscr{T}_{\alpha_0}$. By hypothesis the equality (5) holds for all $A \in \mathscr{S}_1$ and $B \in \mathscr{S}_2$. Since $\mathscr{S}_1$ is closed under finite intersections, any countable union of elements in $\mathscr{S}_1$, and hence by definition any element of $\mathscr{L}_1$ may be represented as a disjoint union of elements in $\mathscr{S}_1$. Therefore, since conditional expectations have (a. s.) the linear and limit properties of integrals, it follows that (5) holds for all $A \in \mathscr{L}_1$ Clearly $\mathscr{L}_1$ is also closed under finite intersections. For induction purposes, assume that for any ordinal $\alpha < \alpha_0$, $\mathscr{T}_\alpha$ satisfies (5) and is closed under finite intersections. It is clear that (5) holds for differences of elements in $\mathscr{T}_\alpha$. For if $C, D \in \mathscr{T}_\alpha$, $C - D = C - (C \cap D)$, and since by assumption $C \cap D \in \mathscr{T}_\alpha$, (5) follows by writing $I_{C-D} = I_C - I_{C\cap D}$. Moreover, countable unions of elements of $\mathscr{T}_\alpha$ may be shown to satisfy (5) in the same way as was used above for $\mathscr{L}_1$. Therefore (5) is satisfied for all elements of $\mathscr{L}_{\alpha+1}$ and hence of $\mathscr{T}_{\alpha+1}$. From the identity $(A - B) \cap (C - D) = (A \cap C) - (B \cup D)$, it follows that $\mathscr{L}_{\alpha+1}$ and hence

$\mathscr{T}_{\alpha+1}$ is closed under finite intersection. It therefore follows by transfinite induction that (5) holds for all $A \in \sigma(\mathscr{S}_1)$ and $B \in \mathscr{S}_2$. The lemma follows by a repetition of the above argument for $\mathscr{S}_2$.

It is remarked that if there exists a conditional probability distribution relative to $\sigma(\mathscr{S})$ in the sense of Doob [2], the conditional expectations of (5) may be considered as integrals with respect to the distribution. In this case one might be tempted to view Lemma 2 as a simple extension of measures, and hence as a corollary to Theorem 1. Closer examination shows this to be a false supposition.

LEMMA 3. *For $X, Y \in \mathscr{E}$, let X and Y be* c. i. $| \mathscr{F}$. *Then if $XY \in \mathscr{E}$*

$$E\{XY \mid \mathscr{F}\} = E\{X \mid \mathscr{F}\}E\{Y \mid \mathscr{F}\} \qquad \text{a. s.}$$

Proof. This result follows from (5) upon approximating X and Y by simple functions in the usual way. The assumption that $XY \in \mathscr{E}$ is certainly not a necessary one but has been postulated in keeping with Definition 2.

The main theorem of this paper is

THEOREM 3. *Let $X \in \mathscr{E}$ and $\mathscr{F}_i \subset \mathscr{A} (i = 1, 2)$ be given. If $\mathscr{B}(X)$ and $\mathscr{F}_2$ are* c. i. $| \mathscr{F}_1$ *then*

$$E\{X \mid \mathscr{F}_1 \overset{*}{\cap} \mathscr{F}_2\} = E\{X \mid \mathscr{F}_1\} \qquad \text{a. s.} \tag{6}$$

Proof. Define $\mathscr{S} = \mathscr{F}_1 \overset{*}{\cap} \mathscr{F}_2$. $\mathscr{S}$ is a quasi-ring by Lemma 1. From Theorem 2, (4), and the fact that $E\{X \mid \mathscr{F}_1\}$ is $\sigma(\mathscr{S})$-measurable, it follows that to prove (6) it suffices to show that

$$E \circ XI_S = E \circ E\{X \mid \mathscr{F}_1\}I_S \qquad \text{a. s.}$$

for all $S \in \mathscr{S}$. Let $S = A \cap B$ for $A \in \mathscr{F}_1, B \in \mathscr{F}_2$. Then

$$\begin{aligned} E \circ XI_{A\cap B} &= E \circ E\{XI_B \mid \mathscr{F}_1\}I_A && \text{a. s.} \\ &= E \circ E\{X \mid \mathscr{F}_1\}E\{I_B \mid \mathscr{F}_1\}I_A && \text{a. s.} \end{aligned}$$

since X and I_B are c. i. $| \mathscr{F}_1$. Therefore

$$\begin{aligned} E \circ XI_{A\cap B} &= E \circ E\{I_B E\{X \mid \mathscr{F}_1\} \mid \mathscr{F}_1\}I_A && \text{a. s.} \\ &= E \circ E\{X \mid \mathscr{F}_1\}I_{A\cap B} && \text{a. s.} \end{aligned}$$

by (1). *Q.* e. d.

COROLLARY 3.1. *Let $X \in \mathscr{E}$ and let X and Z be* c. i. $| Y$. *Then*

$$E\{X \mid Y, Z\} = E\{X \mid Y\} \qquad \text{a. s.} \tag{7}$$

It is of interest to state this result under the stronger but more common assumption of independence, viz.,

COROLLARY 3.2. *For* $X \in \mathscr{E}$, *let the random vector* (X, Y) *be independent of* Z. *Then* (7) *holds.*

Proof. This is *a* consequence of the fact that (X, Y) being independent of Z implies that X and Z are c. i. $| Y$. To see this, consider

$$\begin{aligned} E\{I_{A\cap B} \mid Y\} &= E\{E\{I_{A\cap B} \mid Y, X\} \mid Y\} = E\{I_A E\{I_B \mid Y, X\} \mid Y\} \quad \text{a. s.} \\ &= E\{I_A \mid Y\}E\{I_B\} = E\{I_A \mid Y\}E\{I_B \mid Y\} \quad \text{a. s.} \end{aligned}$$

where $A \in \mathscr{B}(X)$, $B \in \mathscr{B}(Z)$.

It should be noted that Corollaries 3.1 and 3.2 remain valid if the random variables Y and Z are replaced by random functions since the proofs depend only on the properties of the corresponding σ-algebras.

Before stating a generalization of Theorem 3, we prove the following lemma:

LEMMA 4. *If* $\mathscr{F}_2$ *and* $\mathscr{F}_3$ *are* c. i. $| \mathscr{F}_1$, *then* $\mathscr{F}_1 \overset{*}{\cap} \mathscr{F}_2$ *and* $\mathscr{F}_1 \overset{*}{\cap} \mathscr{F}_3$ *are* c. i. $| \mathscr{F}_1$.

Proof. Let $A_i \in \mathscr{F}_i (i = 1, 2, 3)$ and $B_1 \in \mathscr{F}_1$. Then

$$\begin{aligned} E\{I_{A_1\cap A_2} I_{B_1\cap A_3} \mid \mathscr{F}_1\} &= I_{A_1} I_{B_1} E\{I_{A_2} I_{A_3} \mid \mathscr{F}_1\} \quad \text{a. s.} \\ &= I_{A_1} E\{I_{A_2} \mid \mathscr{F}_1\} I_{B_1} E\{I_{A_3} \mid \mathscr{F}_1\} \quad \text{a. s.} \\ &= E\{I_{A_1\cap A_2} \mid \mathscr{F}_1\} E\{I_{B_1\cap A_3} \mid \mathscr{F}_1\} \quad \text{a. s.} \end{aligned}$$

by hypothesis and lemma follows.

THEOREM 4. *Let* $Y \in \mathscr{E}$ *and* $\mathscr{F}_i \subset \mathscr{A}$ $(i = 1, 2, 3)$ *be* given. *If* $\mathscr{B}(Y) \subset \sigma(\mathscr{F}_1 \cup \mathscr{F}_2)$ *and if* $\mathscr{F}_2$ *and* $\mathscr{F}_3$ *are* c. i. $| \mathscr{F}_1$, *then*

$$E\{Y \mid \mathscr{F}_1 \overset{*}{\cap} \mathscr{F}_3\} = E\{Y \mid \mathscr{F}_1\} \quad \text{a. s.} \tag{8}$$

Proof. By Lemma 4 it follows that $\mathscr{F}_1 \overset{*}{\cap} \mathscr{F}_2$ and $\mathscr{F}_3$ are c. i. $| \mathscr{F}_1$. Therefore, (8) becomes a consequence of Theorem 3 since $\mathscr{F}_1 \overset{*}{\cap} \mathscr{F}_2$ and $\mathscr{F}_3$ being c. i. $| \mathscr{F}_1$ implies that $\mathscr{B}(Y)$ and $\mathscr{F}_3$ are c. i. $| \mathscr{F}_1$.

Of particular importance is the following special case of the above theorem:

COROLLARY 4.1. *Let* $g : R_2 \to R_1$ *be a* $\mathscr{B}_2$*-measurable function, and* r. v.'s X, Y, Z *be such that* $g(X, Y) \in \mathscr{E}$, *and either* X *and* Z *are* c. i. $| Y$ *or the vector* (X, Y) *is independent of* Z. *Then*

$$E\{g(X, Y) \mid Y, Z\} = E\{g(X, Y) \mid Y\} \qquad \text{a. s.}$$

As before, this result remains valid if the random variables X, Y and Z are replaced by random functions.

It should be remarked that many of the foregoing results may be obtained by elementary means for cases where the random variables involved possess joint probability density functions with respect to some dominating measure. In many applications, however, the existence of such density functions cannot be postulated.

4. Applications. As a first application of the results of § 3, the following theorem shows the equivalence of certain characterizations of conditional indepedence :

THEOREM 5. *For* r. v.'s X, Y, Z, *the following statements are equivalent* :

(a) *Z and X are* c. i. $\mid Y$

(b) *Z-Y and X-Y are* c. i. $\mid Y$

(c) $P\{Z \leqq z \mid Y, X\} = P\{Z \leqq z \mid Y\}$ a. s. *for all* $z \in R_1$.

Proof. (Note first the standard definition $P\{A \mid \mathscr{F}\} \equiv E\{I_A \mid \mathscr{F}\}$ which has been presupposed in (c).) Lemma 4 shows that $(a) \rightarrow (b)$. Since $\mathscr{B}(Z) \subset \mathscr{B}(Y, Z - Y)$ and $\mathscr{B}(Y, X) = \mathscr{B}(Y, X - Y)$ it follows from Theorem 4 that (b) $\rightarrow$ (c). (c) implies that $E\{I_A \mid Y, X\} = E\{I_A \mid Y\}$ for all A of the form $\{z_1 < Z \leqq z_2\}$ with $z_1, z_2 \in R_1$. The collection of all such inverse images forms a quasi-ring, $\mathscr{S}$, say, such that $\sigma(\mathscr{S}) = \mathscr{B}(Z)$. It follows then that for $A \in \mathscr{S}$, $B \in \mathscr{B}(X)$,

$$E\{I_A I_B \mid Y\} = E\{I_B E\{I_A \mid Y, X\} \mid Y\} = E\{I_A \mid Y\} E\{I_B \mid Y\} \qquad \text{a. s.}$$

and (a) follows by Lemma 2. Q. e. d.

The equivalence of (a) and (c) has been proved in a different form by Doob ([2], pp. 83–85) for the more general case in which Z and X are allowed to be finite-dimensional random vectors. It should be pointed out that the restriction to one-dimensional r. v.'s was solely for presentation purposes throughout this paper, and that all of the above results carry through when the conditioning r. v.'s are replaced by arbitrary families of r.v.'s. This is true simply because all results involving r. v.'s have been stated in terms of their induced σ-algebras. Roughly speaking, in this more general context, the implication (c)$\rightarrow$(a) of Theorem 5 states that for a Markov process the past and future are c. i. given the present.

A second application is in proving the statement that a stochastic process $\{X_t : t \in T\}$ with independent increments is a Markov process. Indeed this statement is a simple corollary of Theorem 4. For $t_1 < t_2 <$

$\cdots < t_n$, consider

$$P\{X_{t_n} \leqq x \mid X_{t_1}, X_{t_2}, \cdots, X_{t_{n-1}}\} = P\{(x_{t_n} - X_{t_{n-1}}) + X_{t_{n-1}} \leqq x \mid X_{t_{n-1}}, (X_t, \cdots, X_{t_{n-2}})\} \quad \text{a. s.}$$

$$= P\{X_{t_n} \leqq x \mid X_{t_{n-1}}\} \quad \text{a. s.}$$

The last equality is a consequence of the remark following Corollary 4.1, since $X_{t_n} - X_{t_{n-1}}$ and $(X_{t_1}, \cdots, X_{t_{n-2}})$ are independent. A proof of this fact, using only the standard theorems of conditional expectation, is lengthy and rather unattractive (cf., Doob [2], p. 85).

REFERENCES

1. C. B. Bell, *On the structure of stochastic independence*, Illinois J. Math., **2** (1958), 415-424.
2. J. L. Doob, *Stochastic processes*, J. Wiley and Sons, 1953.
3. P. Halmos, *Measure theory*, Van Nostrand, 1950.
4. M. Loève, *Probability theory*, Van Nostrand, 1955.
5. J. von Neumann, *Functional operators*, Vol. 1, Ann. of Math. Studies No. 21, Princeton University Press, 1950.

STANFORD UNIVERSITY

ARCS IN PARTIALLY ORDERED SPACES

R. J. KOCH

We present here a theorem on the existence of arcs in partially ordered spaces, and several applications to topological semigroups. The hypotheses are motivated by the structure of the partially ordered set of principal ideals of a compact connected topological semigroup with unit. Noteworthy among the applications are (1)[1] a compact, connected topological semigroup with unit contains an arc. (2) A compact, connected topological semigroup with zero, each of whose elements is idempotent, is arc-wise connected. Throughout the paper, arc is used in the sense of "continuum irreducibly connected between two points". We do not assume metricity of the spaces, but all spaces are assumed to be Hausdorff. Simple non-metric examples of the theorems are furnished by the "long line", i.e. the ordinals up to and including Ω, filled in with intervals, the operation being $a \cdot b = \min(a, b)$. The author is indebted to R. D. Anderson, R. P. Hunter, and W. Strother for useful suggestions.

We recall the following definitions: [10] $(X, \leq)$ is a partially ordered space if X is a space, and $\leq$ is a reflexive, antisymmetric, transitive binary relation on X. A chain in X is an ordered subset of $(X, \leq)$. We denote by Graph $(\leq)$ the set of pairs (x, y) with $x \leq y$. We denote by $A \backslash B$ the complement of B in A; closure is denoted by*, $F(A)$ denote the boundary of A, and $\square$ denotes the empty set.

The following result of the author [3] is presented here in detail because of its relation with Theorem 2.

THEOREM 1. *Let $(X, \leq)$ be a compact partially ordered space and let W be an open set in X. If*

(1) *For each $x \in X$, $\{y \mid y \leq x\}$ is closed, and*

(2) *For any $x \in W$, each open set about x contains an element y with $y < x$,*

then if C is any component of W, $C^* \cap F(W) \neq \square$.

Proof. We show first that if V is open and $V \subset W$, then $F(V) \neq \square$. Let M be a maximal chain in V^*. Then M is compact [10], hence M has a minimal element $m \in V^*$. If $m \in V$, then by hypothesis (2) above, the chain M can be extended, contrary to the maximality. Now

Received September 4, 1958. This research was supported by the United States Air Force through the Air Force Office of Scientific Research, Air Research and Development Command, under Contract AF 18 (603)-89.

* [1]This settles a question raised by D. Montogomery. The author learned of the question through A. D. Wallace.

let C be a component of W and suppose $C^* \cap F(W) = \square$. Then by standard arguments [5; p. 110] there is an open and closed set N with $C \subset N \subset W$. Hence N is an open subset of W with $F(N) = \square$, a contradiction.

In the next theorem we use the following topology for the space $S(X)$ of non empty closed subsets of a compact space X, which is an extension of the Hausdorff metric topology [1]. For open sets U and V of X, let $N(U, V) = \{A \mid A \in S(X), A \subset U, A \cap V \neq \square\}$. Take $\{N(U, V) \mid U, V$ open$\}$ for a subbasis for the open sets of $S(X)$. It is known that if X is compact Hausdorff, so is $S(X)$.

THEOREM 2. *Let $(X, \leq)$ be a compact partially ordered space, and let W be an open set in X. If*

(1) *Graph $(\leq)$ is closed and*

(2) *For any $x \in W$, each open set about x contains an element y with $y < x$,*

then any element x of W belongs to a (compact) connected chain C with $C \cap F(W) \neq \square$ and $x = \sup C$.

Proof. Let W be as above, and fix $a \in W$. Since $W^* \cap \{y \mid y \leq a\}$ is a compact partially ordered space and contains the relatively open set $W \cap \{y \mid y \leq a\}$ satisfying the above hypotheses, we may assume that $X = \{y \mid y \leq a\}$ and that W is an open set in X with $a \in W \subset W^* = X$. Let $\mathscr{C}$ denote the collection of all closed chains in X with $a \in C$ and $C \cap F(W) \neq \square$. By Theorem 1, $F(W) \neq \square$; hence if $z \in F(W)$, the elements a and z constitute an element of $\mathscr{C}$, so that $\mathscr{C} \neq \square$.

(i) *$\mathscr{C}$ is closed in $S(X)$.* We show that $S(X)\backslash\mathscr{C}$ is open. Let A be closed in X, with $A \notin \mathscr{C}$. If A is not a chain, then there are elements x and y of A with $x \not\leq y$ and $y \not\leq x$. By hypothesis (1) there exist open sets U and V about x and y with the property that $x' \in U$, $y' \in V$ imply $x' \not\leq y'$ and $y' \not\leq x'$. Then $N(X, U) \cap N(X, V)$ is an open set about A, and misses $\mathscr{C}$. If A is a chain but $a \notin A$, then $(X\backslash a, X\backslash a)$ is an open set about A which misses $\mathscr{C}$, and the case $F(W) \cap A = \square$ goes in a similar way.

Define $L(x) = \{y \mid y \leq x\}$ and $M(x) = \{y \mid x \leq y\}$. Also define $(x, y) = \{z \mid x < z < y\}$. Let δ be an open cover of X, and define a subset M_δ of $S(X)$ by: $C \in M_\delta$ iff C is a closed chain in X, and for any x and y in C with $x < y$ and $(x, y) \cap C = \square$, there exists $V \in \delta$ such that V^* meets both $L(x) \cap C$ and $M(y) \cap C$.

(ii) *$M_\delta \cap \mathscr{C} \neq \square$, for any open cover δ.* Let δ be an open cover of X, let $\mathscr{D}$ be the collection of all closed chains C with $C \subset W$, $C \in M_\delta$, and $a \in C$. Let $\mathscr{J}$ be a maximal tower in $\mathscr{D}$, and let $T = \bigcup \mathscr{J}$. Note that T is a chain containing a, hence by hypothesis (1), T^* is again a chain. Also it is easily seen that $\inf T = \inf T^*$. We will show that $T^* \in M_\delta$. Let $x, y \in T^*$ with $x < y$ and $(x, y) \cap T^* = \square$. Suppose $x = \inf T$; then if $\inf T \notin T$ it follows that $T \subset M(y)$, hence $x \in T^* \subset M(y)$, a contradiction. Hence $x = \inf T \in T$, so there exists $T_1 \in \mathscr{J}$ such that $x \in T_1$. Since $T_1 \in M_\delta$, there is $V \in \delta$ with V^* meeting both $L(x) \cap T_1$ and $M(y) \cap T_1$. Hence V^* meets both $L(x) \cap T^*$ and $M(y) \cap T^*$, and it follows that $T^* \in M_\delta$. Therefore we may suppose $\inf T < x < y$, so there exists t_0 with $\inf T < t_0 \leq x < y, t_0 \in T_0 \in \mathscr{J}$. Since $T_0 \in M_\delta$, it again follows that $T^* \in M_\delta$. This establishes that $T^* \in M_\delta$ and it remains to show $T^* \in \mathscr{C}$. If $T^* \subset W$, then $T^* \in \mathscr{D}$, so by the maximality of T, $T^* = T$. Hence $\inf T \in T \subset W$, so there exists $V \in \delta$ with $\inf T \in V$. There is an element $y \in V \cap W$ with $y < x$. By an easy argument, $T \cup y \in M_\delta$, so $T \cup y \in \mathscr{D}$, $T \cup y = T$, and $y \in T$, a contradiction. We conclude that $T^* \cap F(W) \neq \square$, and $T^* \in M_\delta \cap \mathscr{C}$.

(iii) *$M_\delta \cap \mathscr{C}$ is closed for each finite open cover δ of X.* Let δ be a finite open cover of X. We show that $S(X)\backslash M_\delta \cap \mathscr{C}$ is open. Let A be a closed set, $A \notin M_\delta \cap \mathscr{C}$. Then either $A \notin \mathscr{C}$, or $A \in \mathscr{C}$ but $A \notin M_\delta$. If $A \notin \mathscr{C}$, then by part (i) there is an open set about A which misses $\mathscr{C}$. Suppose $A \in \mathscr{C}$ and $A \notin M_\delta$. Then for some $x, y \in A$ we have $(x, y) \cap A = \square$, and for each $V \in \delta$, V^* misses either $L(x) \cap A$ or $M(y) \cap A$. By an easy argument which makes use of hypothesis (1) and compactness, there are open sets U_1 and W_1 about $L(x) \cap A$ and $M(y) \cap A$ resp. with the property that $x' \in U_1$ and $y' \in W_1$ imply $y' \not\leq x'$. Let $U_2 = \bigcap \{X\backslash V^* \mid V \in \delta, V^* \cap L(x) \cap A = \square\}$, $W_2 = \bigcap \{X\backslash V^* \mid V \in \delta, V^* \cap M(y) \cap A = \square\}$. Then, since δ is finite, U_2 is open and contains $L(x) \cap A$, W_2 is open and contains $M(y) \cap A$, and for each $V \in \delta$, V^* misses either U_2 or V_2. Let $U' = U_1 \cap U_2$, $W' = W_1 \cap W_2$. Now $N(U' \cup W', U' \cup W')$ is open, contains A, and as we next show, misses $M_\delta \cap \mathscr{C}$. Suppose $C \in \mathscr{C} \cap N(U' \cup W', U' \cup W')$. Let $x' = \sup(C \cap U')$, $y' = \inf(C \cap W')$; then $(x', y') \cap C = \square$. Also $L(x') \cap C \subset U'$ and $M(y') \cap C \subset W'$, and since for each $V \in \delta$, V^* misses either U' or W', we conclude that $C \notin M_\delta$. This completes (iii).

For any finite open cover δ, put $P_\delta = M_\delta \cap \mathscr{C}$, and let $\mathscr{P} = \{P_\delta\}$. Note that if $P_\alpha, P_\beta \in \mathscr{P}$, then there is a finite open cover γ which refines both α and β, and hence $P_\gamma \subset P_\alpha \cap P_\beta$. Therefore $\bigcap \mathscr{P} \neq \square$. Let $C \in \bigcap \mathscr{P}$, and we show next that C is order dense. Let $x, y \in C$ with $x < y$. Then $L(x) \cap M(y) \cap C = \square$, so by normality there are open sets U and V about $L(x) \cap C$ and $M(y) \cap C$ resp., with $U^* \cap V^* = \square$.

Let α be the finite open cover consisting of $\{X\backslash U^*, X\backslash V^*\}$. Since $C \in P_\alpha$ and it is false that the closure of each member of α meets both

$L(x)$ and $M(y)$, we conclude that $(x, y) \cap C \neq \square$, and C is order-dense. Hence C is a compact order dense chain from a to $F(W)$, and is therefore an arc. The proof is complete.

COROLLARY 1. *Let* $(X, \leq)$ *be a compact partially ordered space with unique minimal element* 0. *If*

(1) *Graph* $(\leq)$ *is closed in* $X \times X$ *and*

(2) $L(x)$ *is connected for each* $x \in X$, *then* X *is arcwise connected.*

Proof. Let $W = X \backslash 0$; then from the connectedness of each $L(x)$, we see that W satisfies the hypotheses of Theorem 2. Hence each element of X can be joined to 0 by a compact connected chain.

We note that Corollary 1 contains a result of Wallace [8].

Let S be a compact topological semigroup, and for $a \in S$, let $J(a) = a \cup Sa \cup aS \cup SaS$, and let $J = \{(x, y) \mid J(x) = J(y)\}$. Endow $\frac{S}{J}$ with the quotient topology, and let $\varphi: S \to \frac{S}{J}$ be the natural mapping. From the compactness of S it follows that $\frac{S}{J}$ is compact Hausdorff.

We denote by E the set of idempotents of S.

COROLLARY 2. *Let S be a compact connected topological semigroup, with* $S = ES \cup SE$: *then* $\frac{S}{J}$ *is arcwise connected.*

Proof. We define a partial order in $\frac{S}{J}$ by: $\varphi(a) \leq \varphi(b)$ iff $J(a) \subset J(b)$. From the compactness of S it follows that Graph $(\leq)$ is closed in $S \times S$. Further, if K denotes the minimal ideal of S, then $\varphi(K)$ is the unique minimal element of $\frac{S}{J}$. Note that $L(\varphi(x)) = J(x)$; since $S = ES \cup SE$, it follows that $J(x)$ is connected. Hence Corollary 1 applies, and the proof is complete.

REMARK. By a similar argument it can be seen that if L is a continuous monotone reflexive struct [4.7] on a continuum X, all of whose minimal elements are related, then $\frac{X}{L}$ is arcwise connected.

COROLLARY. 3. *Let S be a compact connected topological semigroup with unit u, and let V be an open set about u; then V contains an arc.*

Proof. If S is a group, the result is known [2], hence we may suppose that $u \notin K$, where K is the minimal ideal of S [9]. Let E denote the set of idempotents of S, and partially order E by: $e \leq f$ iff $e \in fSf$. It follows from the compactness of S that Graph $(\leq)$ is closed in $E \times E$. If there exists an open set W about u such that $W \cap E$ satisfies (2) of Theorem 2 (taking $X = E$), then $W \cap V \cap E$ also satisfies (2), so there is an arc in V. If for each open set W about u, $W \cap E$ fails to satisfy (2), then there exists $e \in E \cap V \cap (S \backslash K)$ such that $eSe \cap E \cap V = \{e\}$. Since $e \notin K$, eSe is non-degenerate. Hence eSe is a compact connected semigroup with unit e in which there is an open set containing e but no other idempotent. By a theorem of Mostert and Shields [6] there is a local one-parameter semigroup in $eSe \cap V$ and the proof is complete.

A compact connected semigroup with unit may fail to contain an arc which contains the unit. This is illustrated in Example 2 below, and is due to R. P. Hunter (unpublished).

EXAMPLE 1. Let R_+ denote the non-negative reals under addition, and let C by the unit disc in the complex plane: $C = \{z : |z| \leq 1\}$ Let $W = \{(z, t); z = \exp(2\pi i s), t = e^{-s}, s \in R_+\}$, and set $S = (C \times \{0\}) \cup W$. Then S is a compact connected semigroup with zero and unit, but does not contain a standard thread joining the two. We may describe S by saying it is a two-cell with an arc winding on its boundary.

EXAMPLE 2. Let D_i be the graph of $x^2 + y^2 \leq \frac{1}{i}$; $z = 1 - \frac{1}{i}$; $i = 1, 2, 3, \cdots$. The D_i then converge to a point u. From the center of D_{i+1} we start an arc A_i which winds upon the boundary of D_i as in Example 1. Let $S_i = A_i \cup D_i$; S_i is then a compact connected semigroup with zero 0_i and unit u_i. Repeat this construction for each positive integer i, and let $S = \bigcup_i S_i$. We define multiplication in S as follows: if $x, y \in S$ and both x and y belong to the same S_i, let xy be the product given in S_i. If $x \in S_i$ and $y \in S_j$ with $i < j$, define $xy = yx = x$. It is easy to see that S^* becomes a compact connected semigroup with zero 0_1 and unit u. Moreover, no arc in S^* contains u.

COROLLARY 4. *Let S be a locally compact connected topological semigroup with zero* (0), *each of whose elements is idempotent*; *then S is arcwise connected.*

Proof. Note that $e \leq f$ iff $e \in fSf$ defines a partial order on S, and that Graph $(\leq)$ is closed in $S \times S$. The conclusion is now immediate from Corollary 1.

We conjecture that compactness can be replaced by local compactness in Corollary 4, and further, that a locally compact connected semilattice is arcwise connected.

REFERENCES

1. O. Frink, *Topology in lattices*, Trans. Amer. Math. Soc., **51** (1942), 569-582.
2. A. M. Gleason, *Arcs in locally compact groups*, Proc. Nat. Acad. Sci., **36** (1950), 663-667.
3. R. J. Koch, *Note on weak cutpoints in clans*, Duke Math. J., **24** (1957), 611-616.
4. I. S. Krule, *Structs on the 1-sphere*, Duke Math. J., **24** (1957), 405-414.
5. C. Kuratowski, *Topologie*, II, Warsaw, 1950.
6. P. S. Mostert and A. L. Shields, *One parmeter semigroups in a semigroup*, (to appear).
7. A. D. Wallace, *Struct ideals*, Proc. Amer. Math. Soc., **6** (1955), 634-638.
8. ———, *Partial order and indecomposability*, Proc. Amer. Math. Soc., **5** (1954), 780-781.
9. ———, *The structure of topological semigroups*, Bull. Amer. Math. Soc., **61** (1955), 95-112.
10. L. E. Ward, Jr. *Partially ordered topological spaces*, Proc. Amer. Math. Soc., **5** (1954), 144-161.

LOUISIANA STATE UNIVERSITY

A SPACE OF MULTIPLIERS OF TYPE $L^p(-\infty, \infty)$

GREGERS L. KRABBE

1. Introduction. Let $V(G)$ denote the set of all functions having finite variation on G. Set $G = (-\infty, \infty) = \hat{G}$, and let $V_\infty(G)$ be the Banach space of all functions in $V(G)$ which vanish at infinity. If $f \in V_\infty(G)$, then there exists a bounded linear operator $(t_p f)$ on $L^p(\hat{G})$ such that

$$(\mathrm{i}_0) \qquad (\text{Fourier transform of } (t_p f)x) = (\text{Fourier transform of } x) \cdot f$$

for all x in $L^p(\hat{G})$. This will be shown in 7.2. In the terminology of Hille [3, p. 566], functions f having property (i_0) are called "factor functions for Fourier transforms of type (L_p, L_p)".

Suppose $1 < p < \infty$. When $f \in L^1(G) \cap V(G) \subset V_\infty(G)$, then $(t_p f)$ is a singular integral operator: for all x in $L^p(\hat{G})$ it is found that $(t_p f)x$ has the form

$$[(t_p f)x]_\lambda = \frac{1}{2\pi i}\int_{-\infty}^{\infty} x(\theta)\frac{F(\theta-\lambda)}{\theta-\lambda}d\theta \qquad (\lambda \in \hat{G}),$$

where the integral is taken in the Cauchy principal value sense.

In 6.2 will be defined a set $\blacktriangle(L^p(\hat{G}))$ which contains all factor functions for Fourier transforms of type (L_p, L_p); the set $\blacktriangle(L^p(\hat{G}))$ is a slight extension of what Mihlin [6] calls "multipliers of Fourier integrals". We will find a number N_p such that

(i) *if* $f \in V_\infty(G)$ *then* $f \in \blacktriangle(L^p(\hat{G}))$ *and* $\|(t_p f)\| \leqq N_p \cdot \|f\|_v$,

where $\|f\|_v$ denotes the total variation on G of the function f. Let F_* be the mapping $\{x \to x * F\}$, where $x * F$ is the convolution of the functions x and F;

$$[x * F]_\lambda = \int_{-\infty}^{\infty} x(\theta) \cdot F(\theta - \lambda)d\theta \qquad (\lambda \in \hat{G}).$$

Let (Yf) denote the Fourier transform of the function f:

(ii) *if* $f \in L^1(G) \cap V(G)$, *then the transformation* $(Yf)_*$ *is a densely defined bounded operator, and* $(t_p f)$ *is its continuous linear extension to the whole space* $L^p(\hat{G})$.

Let us for a moment call $G = \{0, \pm 1, \pm 2, \cdots\}$ and $\hat{G} = [0, 1]$. In

Received December 8, 1958, and in revised form February 11, 1959. This research was supported by the United States Air Force through the Air Force Office of Scientific Research of the Air Research and Development Command under contract No. AF 49(638)-505.

a sense, the following relations are duals of (i) and (ii), respectively:

(i′) *if* $F \in V(\hat{G})$ *then* $(\dot{Y}F) \in \blacktriangle(L^p(\hat{G}))$ *and* $\|t_p(YF)\| \leqq k_p \cdot \|F\|_v$

(ii′) *if* $F \in V(\hat{G})$ *then* $F_* = t_p(YF)$ *is a bounded operator on* $L^p(\hat{G})$.

When $\hat{G} = [0, 1]$ these properties are easily verified (see 8.1). We will not[1] prove (i′)-(ii′) for other choices of G.

When $G = [0, 1]$, then (ii) is seen to be a theorem due to Stečkin [10]; by means of appropriate definitions, it could be shown that (i) also holds for this particular choice of G.

2. **Applications.** If f belongs to the class S of members of $L^1(G) \cap V(G)$ such that $(Yf) \in L^1(\hat{G})$, then $(Yf)_* = (t_p f)$ is a bounded operator defined on all of $L^p(\hat{G})$; it is interesting to compare this result with the conclusion $F_* = t_p(YF)$ of (ii′). All the classical convolution operators (Poisson, Picard, Weierstrass, Stieltjes, Dirichlet, Fejér,..etc. [7]) are of the form $(t_p f)$, where $f \in S$. See § 8.

3. **Preliminaries.** We assume $1 < p < \infty$ throughout, and write $G = (-\infty, \infty)$. Denote by L^0 the set of step functions with compact support. Let V be the set of all functions a defined on G and such that $\|a\|_v \neq \infty$, where $\|a\|_v$ denotes the total variation on G.

3.1 Definitions. Let V_∞ be the set of all functions a in V such that $\lim a(\theta) = 0$ whenever $|\theta| \to \infty$. We will write L^p instead of $L^p(G)$. If $\iota = 0, 1$ and $f \in L^1$, then the Fourier transforms $[{}_\iota Yf]$ are the functions g_ι defined by

$$(1) \qquad [{}_\iota Yf]_\lambda = g_\iota(\lambda) = \int_{-\infty}^{\infty} \exp(2\pi i \lambda(-1)^\iota \theta) \cdot f(\theta) d\theta \qquad (\lambda \in G) .$$

To lighten the notation, we will write Yf for $[{}_1 Yf]$ and Ψf for $[{}_0 Yf]$.

3.2 Lemma. *If* $a \in L^1 \cap V$, *then* $a \in V_\infty$ *and*

$$(2) \qquad \int_{-\infty}^{\infty} e^{-2\pi i \theta t} da(t) = 2\pi i \theta \cdot [Ya]_\theta \qquad (\theta \in G) .$$

Proof. Since $a \in V$, the limits $a(\pm\infty) = \lim a(\theta)$ (when $\theta \to \pm\infty$) exist. Since $\|a\|_1 < \infty$ we have

$$(3) \qquad \lim_{\theta \to \pm\infty} \int_\theta^{\theta+1} |a| = 0 .$$

The eventuality $a(\pm\infty) \neq 0$ implies a contradiction of (3). Therefore

[1] It would be of interest to determine the validity of (i)-(ii) and (i′)-(ii′) in the general case where G is a connected locally compact abelian group with dual group $\hat{G}$. It is mainly in order to suggest such an investigation that (i′)-(ii′) are mentioned here.

$a(\pm\infty) = 0$, which permits the integration of (1) by parts to obtain (2).

3.3 DEFINITIONS. Let $\delta_* = (-\infty, -\delta] \cup [\delta, \infty)$ and let $(T_\delta a)x$ be the function defined by

$$[(T_\delta a)x]_\lambda = \int_{\delta_*} d\theta \frac{x(\lambda - \theta)}{2\pi i\theta} \int_{-\infty}^{\infty} e^{-2\pi i\theta t} da(t) \tag{4}$$

for all λ in G. We denote by V_1 the set of all members a of V such that, for all x in L^0, the limit

$$[(Ta)x]_\lambda = \lim_{\delta \to 0+} [(T_\delta a)x]_\lambda$$

exists almost-everywhere on G. Let Ta be the operator $\{x \to (Ta)x\}$ defined on L^0.

3.4 LEMMA. *If $h(\theta) = i\theta/|\theta|$, then $h \in V_1$ and Th is the restriction to L^0 of the Hilbert transformation. Moreover $\|(T_\delta h)x\|_p \leqq c_p \cdot \|x\|_p$, where c_p is the norm of Th.*

Proof. This follows from the statement in [8, p. 241] that $\|(T_\delta h)x\|_p \leqq \|(Th)x\|_p$. Theorem G in [1, p. 251] yields a less precise result.

3.5. LEMMA. *If $a \in L^1 \cap V$ then $a \in V_1$ and $x * [Ya] = (Ta)x$ whenever $x \in L^0$.*

Proof. Suppose $\delta > 0$. By definition

$$(x * [Ya])_\lambda = \int_{-\infty}^{\infty} d\theta \cdot x(\lambda - \theta) \cdot [Ya]_\theta = E^\delta(\lambda) + G^\delta(\lambda) ,$$

where

$$G^\delta(\lambda) = \int_{\delta_*} d\theta \cdot x(\lambda - \theta) \cdot [Ya]_\theta \qquad (\lambda \in G) ,$$

while $E^\delta(\lambda)$ is the same integral over the interval $(-\delta, \delta)$. It is clear that $\lim E^\delta(\lambda) = 0$ when $\delta \to 0+$. On the other hand, $G^\delta = (T_\delta a)x$ follows immediately from (2) and (4). This concludes the proof.

3.6 LEMMA. *Suppose $a \in V_1$ and $x \in L^0$. If there exists a number k_p such that $\|(T_\delta a)x\|_p \leqq k_p$ for all $\delta > 0$, then $\|(Ta)x\|_p \leqq k_p$.*

Proof. Set $q = p/(p-1)$. Observe first that

$$\|g\|_p = \sup\left\{\left|\int g \cdot \varphi\right| : \varphi \in L^q \text{ and } \|\varphi\|_q \leqq 1\right\} . \tag{5}$$

Next, we infer from a theorem of F. Riesz ([8], p. 227 footnote 10) that the uniform boundedness of $\|(T_\delta a)x\|_p$ implies that, for all φ in L^q with $\|\varphi\|_q \leqq 1$:

$$\int [(Ta)x]\cdot\varphi = \lim_{\delta\to 0+}\int [T_\delta a)x]\cdot\varphi \ . \tag{6}$$

By (5) we have $\left|\int [(T_\delta a)x]\cdot\varphi\right| \leqq k_p$; this enables us to use (6) to deduce $\left|\int [(Ta)x]\cdot\varphi\right| \leqq k_p$. The conclusion is reached by another application of (5).

3.7 LEMMA. *If $a\in L^1\cap V$ and $x\in L^0$, then*

$$\|(Ta)x\|_p \leqq 2^{-1}c_p\|a\|_v\|x\|_p \ .$$

Proof. Suppose $\delta > 0$. Apply Fubini's theorem to (4):

$$[(T_\delta a)x]_\lambda = \int_{-\infty}^{\infty} da(t)e^{-2\pi i\lambda t}\int_{\delta_*} d\theta\frac{x(\lambda-\theta)}{2\pi i\theta}e^{2\pi i t(\lambda-\theta)} \ .$$

Set $x^t(\beta) = x(\beta)\exp(2\pi i t\beta)$. Keeping both (4) and 3.4 in mind, we can therefore write

$$[(T_\delta a)x]_\lambda = (2i)^{-1}\int_{-\infty}^{\infty} da(t)\{e^{-2\pi i\lambda t}[(T_\delta h)x^t]_\lambda\} \ . \tag{7}$$

This implies

$$\|(T_\delta a)x\|_p \leqq 2^{-1}\|a\|_v \sup_{t\in G}\|(T_\delta h)x^t\|_p \ . \tag{8}$$

The derivation of (8) from (7) is obtained by a standard procedure (e.g. as in [3, Lemma 21.2.1]); it rests upon (5) and requires a single application of the Fubini theorem. On the other hand, 3.4 implies that

$$\|(T_\delta h)x^t\|_p \leqq c_p\cdot\|x^t\|_p \leqq c_p\cdot\|x\|_p \ .$$

In view of (8) therefore: $\|(T_\delta a)x\|_p \leqq 2^{-1}c_p\|a\|_v\|x\|_p$. Use now 3.6 to reach the conclusion.

4. The Banach space V_∞. Let V_s denote the set of all functions in V which have compact support. The norm $\{a\to\|a\|_v\}$ makes the set $\{a\in V\colon a(-\infty)=0\}$ into a Banach space V_0. Note that $V_s\subset V_\infty\subset V_0$. Henceforth V_∞ will be given the topology of V_0. We will write $\|a\|_\infty = \sup\{|a(\theta)|\colon \theta\in G\}$; it is easily checked that

$$\|a\|_\infty \leqq \|a\|_v \qquad \text{(when } a\in V_0) \ . \tag{9}$$

Let χ_n denote the characteristic function of the interval $(-n, n)$, and set $a_n = \chi_n\cdot a$.

4.1 LEMMA. *If* $a \in V_\infty$, *then* $\lim_{n\to\infty} \|a - a_n\|_v = 0$.

Proof. Suppose $f \in V$. Using the notation δ_* of 3.3, we have

$$\text{(iii)} \qquad \|f\|_v = v(f; [-\delta, \delta]) + v(f; \delta_*) ,$$

where $v(f; I)$ denotes the total variation over I. Set $\delta = n$ and $h_n = a - a_n$; therefore $v(h_n; [-\delta, \delta]) = |a(-\delta)| + |a(\delta)|$ and $v(h_n; \delta_*) = v(a; \delta_*)$. From (iii) therefore $\|h_n\|_v = |a(-\delta)| + |a(\delta)| + v(a; \delta_*)$, and the conclusion follows by letting $\delta \to \infty$.

4.2 REMARK. The set V_s is dense in V_∞ (since 4.1 and the fact that $a_n \in V_s$).

4.3 THEOREM. *The set* V_∞ *is a Banach space.*

Proof. Since V_∞ is a metric subspace of the Banach space V_0, it will suffice to show that V_∞ is complete. To that effect, consider a Cauchy sequence $\{g_k\}$ in V_∞; since $\{g_k\}$ is also in V_0, it will converge to some function f in V_0; therefore $f(-\infty) = 0$ and we need only establish that $f(\infty) = 0$. From (9) we see that

$$|f(\theta) - g_k(\theta)| \leqq \|f - g_k\|_v \qquad (\theta \in G) .$$

In view of $g_k(\infty) = 0$, the conclusion is obtained by letting $\theta \to \infty$ and $k \to \infty$.

5. The bilinear operator B_p. From 3.2 results that $V_s \subset L^1 \cap V \subset V_\infty$; it follows from 4.2 that $L^1 \cap V$ is dense in V_∞. Consider the bilinear operator $B = \{(x, a) \to (Ta)x\}$ which maps $L^0 \times (L^1 \cap V)$ into L^p. From 3.7 we see that B is a continuous bilinear mapping of $L^0 \times (L^1 \cap V)$ into L^p. Since L^0 and $L^1 \cap V$ are dense in L^p and V_∞, respectively, it follows that B has a (unique) continuous extension B_p to $L^p \times V_\infty$. Accordingly, if $a \in V_\infty$, then

$$\text{(10)} \qquad \|B_p(x, a)\|_p \leqq 2^{-1}c_p\|a\|_v\|x\|_p \qquad (\text{if } x \in L^p) .$$

If $a \in L^1 \cap V$, then (from 3.5)

$$\text{(11)} \qquad B_p(x, a) = x * Ya \qquad (\text{if } x \in L^0) .$$

5.1 NOTATION. We henceforth identify functions equal almost-everywhere on G. If the sequence $\{f_n\}$ converges in the mean of order p (i.e., in the topology of L^p), then its limit will be denoted $(L^p)\lim f_n$.

5.2 LEMMA. *Let* $\bar{\chi}_n$ *be the function defined by*

$$\bar{\chi}_n(\theta) = (\sin 2\pi n\theta)/\pi\theta \qquad (\theta \in G) .$$

If $f \in L^p$, *then* $f = (L^p)\lim f * \bar{\chi}_n$ *as* $n \to \infty$.

Proof. Observe that Dunford's proof [2, p. 51, Lemma 3] for the case $p = 2$ holds without alteration whenever $1 < p < \infty$.

6. The main result. Suppose $\iota = 0, 1$. When f is a locally integrable function, we set

$$[({}_\iota Y_p)f] = (L^p)\lim_{n\to\infty} [{}_\iota Y(\chi_n \cdot f)] . \tag{12}$$

As in 3.1, we lighten the notation by writing $Y_p f = [({}_1 Y_p)f]$ and $\Psi_p f = [({}_0 Y_p)f]$.

6.1 REMARK. If $f \in L^1$ then $[({}_\iota Y_p)f] = [{}_\iota Yf]$. The following definition is an extension of the one used by Mihlin ("Multipliers of Fourier integrals"[2]).

6.2 DEFINITION. A locally integrable function a is called a *"multiplier of type L^p"* if both the following conditions hold:

$$\begin{cases} \text{the transform } Y_p(a\cdot[\Psi x]) \text{ exists and belongs to } L^p \text{ whenever } x \in L^0 \\ \infty \neq \sup\{\|Y_p(a\cdot[\Psi x])\|_p \colon x \in L^0 \text{ and } \|x\|_p \leqq 1\} . \end{cases}$$

Let $\blacktriangle(L^p)$ denote the set of all multipliers of type L^p. When $a \in \blacktriangle(L^p)$, then $(t_p a)$ is defined as the continuous extension to all of L^p of the transformation $\{x \to Y_p(a\cdot[\Psi x])\}$ defined on L^0.

6.3 THEOREM. *If* $a \in V_\infty$, *then* $a \in \blacktriangle(L^p)$ *and* $(t_p a)x = B_p(x, a)$ *for all* x *in* L^p.

Proof. Note first that $a_n = (\chi_n \cdot a) \in L^1 \cap V$. Suppose $x \in L^0$. From (11) we see that

$$[B_p(x, a_n)]_\lambda = \int d\theta \cdot x(\theta) \int dt \cdot e^{-2\pi i(\lambda-\theta)t} a_n(t) \qquad (\text{when } \lambda \in G) .$$

By Fubini's theorem

$$[B_p(x, a_n)]_\lambda = \int dt \cdot a_n(t) e^{-2\pi i\lambda t} [\Psi x]_t \qquad (\text{for all } \lambda \text{ in } G) .$$

Or, equivalently

$$B_p(x, a_n) = Y(\chi_n \cdot a \cdot [\Psi x]) .$$

[2] See [6]; in that article, Mihlin gives a condition which ensures that a differentiable function be in $\blacktriangle(L^p)$.

From (10) and 4.1 we can now infer that

$$B_p(x, a) = (L^p) \lim_{n\to\infty} Y(\chi_n \cdot \{a \cdot [\Psi x]\}) .$$

From the definition (12) now results that $B_p(x, a) = Y_p(a \cdot [\Psi x])$ for all x in L^0. This completes the proof, in view of (10) and 6.2.

7. **Hille's definition.** Set $q = p/(p-1)$. The following definition is found in [3, p. 566]: a function a is said to be a *factor function for Fourier transforms of type* (L_p, L_p) if and only if

$$a \cdot [\Psi_q x] \in \{\Psi_q z : z \in L^p\}$$

wherever $x \in L^p$. This definition seems to require the restriction $p \leqq 2$, since $[\Psi_q x]$ need not exist otherwise.

7.1 THEOREM. *Suppose* $1 < p \leqq 2$. *If* a *is a factor function for Fourier transforms of type* (L_p, L_p), *then* $a \in \blacktriangle(L^p)$.

Proof. If a is such a factor function, there exists a bounded linear mapping $(t'_p a)$ of $L^p(G)$ into itself (see [3, Theorem 21.2.1]); this operator is defined by

$$a \cdot [\Psi_q x] = \Psi_q((t'_p a)x) \qquad \text{for all } x \text{ in } L^p .$$

In view of [11, 5.17], this implies

$$\text{(13)} \qquad Y_p(a \cdot [\Psi_q x]) = (t'_p a)x \qquad \text{for all } x \text{ in } L^p .$$

The conclusion follows from 6.1 and 6.2.

7.2 THEOREM. *Suppose* $1 < p \leqq 2$ *and* $a \in V_\infty$. *Then* a *is a factor function for Fourier transforms of type* (L_p, L_p); *moreover,*

$$\text{(14)} \qquad \Psi_q(B_p(x, a)) = a \cdot [\Psi_q x] \qquad (\text{when } x \in L^p) .$$

Proof. Since $B_p(x, a) \in L^p$ when $x \in L^p$ (see §4), it will suffice to prove (14). Consider first the case $(x, a) \in L^0 \times V_s$. From (12) we see that

$$\text{(15)} \qquad \Psi_q(B_p(x, a)) = (L^q) \lim_{n\to\infty} g_n ,$$

where $g_n = \Psi[\chi_n \cdot B_p(x, a)]$. From (11):

$$g_n(\lambda) = \int_{-n}^{n} d\theta \cdot e^{2\pi i\lambda\theta} \int d\alpha \cdot x(\alpha)[Ya]_{\theta-\alpha} \qquad (\text{when } \lambda \in G) .$$

A repeated application of the Fubini theorem yields

$$g_n(\lambda) = \int dt \cdot a(t)[\Psi x]_t \int_{-n}^{n} d\theta \cdot e^{2\pi i(\lambda - t)\theta} \qquad (\text{when } \lambda \in G) .$$

In the notation of 5.2 we accordingly have

$$g_n = \{a \cdot [\Psi x]\} * \bar{\chi}_n .$$

Since $a \cdot [\Psi x]$ is in L^q, it can be inferred from 5.2 and (15) that

$$\Psi_q(B_p(x, a)) = (L^q) \lim_{n \to \infty} (\{a \cdot [\Psi x]\} * \bar{\chi}_n) = a \cdot [\Psi x] .$$

Keeping $\Psi x = \Psi_q x$ in mind (see 6.1), it is clear that (14) is now proved in the case $(x, a) \in L^0 \times V_s$. Consider the bilinear operator $R = \{(x, a) \to a \cdot \Psi_q x\}$ defined on $L^p \times V_\infty$; since $||\Psi_q z||_q \leqq ||z||_p$, it follows that $||R(x, a)||_q \leqq ||x||_p ||a||_\infty$, and from (9) results that R is a bounded bilinear mapping of $L^p \times V_\infty$ into L^q. In view of (10), this remark also shows that the bilinear operator $J = \{(x, a) \to \Psi_q(B_p(x, a))\}$ is a bounded bilinear mapping of $L^p \times V_\infty$ into L^q.

Having shown that $R(x, a) = J(x, a)$ whenever $(x, a) \in L^0 \times V_s$, the desired conclusion $R = J$ can now be inferred from the denseness of L^0 and V_s in L^p and V_∞, respectively (see 4.2).

8. **Concluding remarks.** From 6.3, 3.2 and 3.5 follows that, if $f \in L^1 \cap V$ and $x \in L^p$, then $(t_p f)x = B_p(x, f) = Tf$; hence, if F is the Fourier-Stieltjes transform of f, we have (from 3.3) the relation

$$[(t_p f)x]_\lambda = \frac{1}{2\pi i} \int_{-\infty}^{\infty} x(\theta) \frac{F(\theta - \lambda)}{\theta - \lambda} d\theta \qquad (\lambda \in G)$$

which was announced in the introduction. Property (ii) of the introduction follows from (11) and 6.3. If $A \in L^1$ we denote by A_{*p} the bounded operator $\{x \to x * A\}$ defined on L^p. Let S be the set of all a in $L^1 \cap V$ such that $Ya \in L^1$, and observe that $(Ya)_{*p} = (t_p a)$ when $a \in S$. Again if $a \in S$, then $A = Ya \in L^1$ and $a = \Psi A$; from [4] it is seen that the spectrum of $(t_p a)$ is the closure of the range of a.

8.1 Remark. Set $\hat{G} = [0, 1]$ and $G = \{0, \pm 1, \pm 2, \cdots\}$. We will now sketch a proof of the properties (i′)-(ii′) described in §1. Denote by $||A||_v$ the total variation of A on $\hat{G}$, and suppose $||A||_v \neq \infty$. Observe that, since $A \in L^1(\hat{G})$, we may borrow from [5, p. 10] the following conclusion: $a = YA \in \blacktriangle(L^p(\hat{G}))$ and $t_p(YA) = A_*$ is a bounded linear operator on $L^p(\hat{G})$.

This is all of (i′)-(ii′) except for the inequality. The main result of [5] can be stated as follows[3]:

[3] The definition of $V_\sigma(a)$ is given in [5, p. 8].

$$(16) \qquad \|t_p(a)\| \leqq 2k_p \cdot V_\sigma(a) \; .$$

Note also that $|[YA]_n| \leqq |2\pi n|^{-1}\|A\|_v$ when $n \in G$ (this is obtained by integrating by parts, as in 3.2); consequently $V_\sigma(a) = V_\sigma(YA) \leqq m_p\|A\|_v$. In view of (16), the proof of the inequality in (i′) is completed.

REFERENCES

1. A. P. Calderón and A. Zygmund, *Singular integrals and periodic functions*, Studia Math. **14** (1954), 249-271.
2. N. Dunford, *Spectral theory in abstract spaces and Banach algebras*, Proceedings of the symposium on spectral theory and differential problems, Stillwater, Oklahoma, (1955), 1-65.
3. E. Hille and R. S. Phillips, *Functional analysis and semi-groups*, Amer. Math. Soc. Colloquium Publications, XXXI (1957).
4. G. L. Krabbe, *Spectral invariance of convolution operators on* $L^p(-\infty, \infty)$, Duke Math. J. **25** (1958), 131-142.
5. G. L. Krabbe, *Spectra of convolution operators on* L^p *and rings of factor-sequences*, Quart. J. Math. Oxford Ser. 2, **8** (1957), 1-12.
6. S. G. Mihlin, *On the multipliers of Fourier integrals*, Dokl. Akad. Nauk SSSR (N.S.) **109** (1956), 701-703.
7. H. Pollard, *Integral transforms*, Duke Math. J. **13** (1946), 307-330.
8. M. Riesz, *Sur les fonctions conjuguées*, Math. Zeit. **27** (1928), 218-244.
9. ———, *Sur les maxima des formes bilinéaires et sur les fonctionnelles linéaires*, Acta Math. **49** (1926), 465-497.
10. S. B. Stečkin, *On bilinear forms*, Doklady Akad. Nauk SSSR (N. S.) **71** (1950), 237-240.
11. E. C. Titchmarsh, *Introduction to the theory of Fourier integrals*, Oxford, 1948.

PURDUE UNIVERSITY

CHAINS OF INFINITE ORDER AND THEIR APPLICATION TO LEARNING THEORY

JOHN LAMPERTI AND PATRICK SUPPES

1. Introduction. The purpose of this paper is to study the asymptotic behavior of a large class of stochastic processes which have been used as models of learning experiments. We will do this by applying a theory of so-called "chains of infinite order" or "chaînes à liaisons complètes." Namely, we shall employ certain limit theorems for stochastic processes whose transition probabilities depend on the entire past history of the process, but only slightly on the remote past. Such theorems were given by Doeblin and Fortet [3] in a form close to that we employ; however, in order to accomodate certain cases of learning models we found it necessary to relax somewhat their hypotheses. A self-contained discussion of these and some additional results is the content of §2.

We should emphasize that this section is included to serve as preparation for the theorems of § 4, and it is original with us only in some details and extensions. In addition to [3], papers by Harris [7] and Karlin [8] contain very closely related results and arguments, but not quite in the form we require.

The processes which we shall study with these tools are called "linear earning models." From a psychological standpoint these models are very simple. A subject is presented a series of *trials*, and on each trial he makes a *response*, which consists of a choice from a finite set of possible actions. This response is followed by a *reinforcement* (again one of a finite number). The assumption of the model is that the subject's response probabilities on the next trial are linear functions of the probabilities on the present trial, where the form of the functions depends upon which reinforcement has occurred. Many results about such models may be found in Bush and Mosteller [2], Estes [4], and Estes and Suppes [6]. We will also study here models constructed along similar lines for experiments involving two or more subjects and a type of interaction between them [6, Section 9] and Atkinson and Suppes [1]. Precise definitions of these processes are given below in §3.

The references mentioned above do not, except in very special cases, give a thorough treatment of asymptotic properties. We shall prove that under general conditions linear learning models exhibit "ergodic" behavior; that is, that after much time has passed these processes become approximately stationary and the influence of the initial distributions

Received November 20, 1958. This research was supported by the Group Psychology and Statistics Branches of the Office of Naval Research under contracts with Stanford University.

goes to zero. This is not the case for all models which have been used in experimental work, but it seems as if ergodic behavior can be proved by our method in almost all the cases in which one might expect it. Our theorems to this effect, their proofs and some corollaries are given in §4.

The major work so far on limiting behavior of learning models is Karlin [8], who obtains detailed limit theorems for certain classes of models. However, the results and even the techniques of Karlin's paper do not apply to many cases of interest. His starting point is a representation of the linear model as a Markov process whose states are the response probabilities. Two typical situations when such a representation is impractical arise (i) when the probabilities with which the reinforcement is selected depend on two or more previous responses, and (ii) in the many-person situations mentioned above. Both these situations can (and will) be studied using infinite order chains, and ergodic behavior established under mild restrictions. On the other hand, Karlin's work treats interesting non-ergodic cases outside the scope of our approach. For example, consider a T-maze experiment in which the subject (a rat, say) is reinforced (rewarded) on each trial regardless of whether he goes left or right. In the appropriate linear model, the probability of a left turn eventually is either nearly 0 or nearly 1, and which it is depends upon the rat's initial response probabilities. The model of this experiment has been thoroughly studied in [8, Section 2], and these results have been generalized by Kennedy [9].

In conclusion we comment that both more detailed results and other applications seem possible using the ideas of "infinite order chains." We hope to contribute further to this development in the future.

2. Chains of infinite order. In this section we present a theory of non-Markov stochastic processes where the transition probabilities are influenced only slightly by the remote past. The original convergence theorems for this type of process are due to Doeblin and Fortet [3]; they are given here in a generalized form (Theorems 2.1 and 2.2). The weaker hypotheses make the proof of Lemma 2.1 more complicated than it is in [3], but the other proofs are not much affected. T. E. Harris has also studied these chains; we shall not use his results but remark that his paper [7] gives additional references and background on the subject. Finally we point out that the restriction to a finite number of states is not essential, and the theorems can be extended to the denumerable case without much change of methods.

Let I consist of the integers from 1 to N (to represent the states of the chain); we shall use the notation x for a finite sequence $i_0, i_1, \cdots$ of integers from I. The subscript "m" on x_m merely adds the specification that the sequence has m terms; the "sum" $x_m + x'$ will be the combined sequence $i_0, \cdots, i_{m-1}, i'_0, i'_1 \cdots$. The starting point for the theory will be a set of functions $p_i(x)$ defined for all $i \in I$ and all sequences x (including the sequence φ of length zero) and having the properties

$$p_i(x) \geqq 0,\ \sum_i p_i(x) = 1 . \tag{2.1}$$

The function $p_i(x)$ will be interpreted as the conditional probability that a path function of the random process will go next to state i, having just occupied state i_0, previously i_1, etc. With this interpretation in mind we define inductively the "higher transition probabilities":

$$p_i^{(n)}(x) = \sum_{j \in I} p_j(x) p_i^{(n-1)}(j + x) , \tag{2.2}$$

where of course $p_i^{(1)}(x) = p_i(x)$, the given function. It is easy to see that these higher probabilities also satisfy condition (2.1). The functions $p_i^{(n)}(x)$ are the analogues of the terms of the matrix P^n for a Markov chain with transition matrix P; the theorems we shall give generalize the convergence properties of the matrices P^n.

We shall first impose a positivity condition on the transition probabilities; specifically we assume that for some state j_0, some positive integer n_0, and some $\delta > 0$,

$$p_{j_0}^{(n_0)}(x) > \delta \text{ for every } x . \tag{2.3}$$

We also need to make precise the "slight" dependence of these probabilities on the remote past; indeed, this is the crux of the whole theory. Define

$$\varepsilon_m = \sup | p_i(x + x') - p_i(x + x'') | \tag{2.4}$$

where the sup is taken over all states i, all sequences x' and x'', and all sequences x which contain the state j_0 at least m times. We shall use the postulate

$$\sum_{m=0}^{\infty} \varepsilon_m < \infty . \tag{2.5}$$

(In [3], ε_m is defined in the same way except that the sup is taken over all x of length at least m. Since this results in larger $\varepsilon'_m s$, and since it is also assumed there that $\sum \varepsilon_m < \infty$, our hypotheses are strictly weaker.) Throughout this section, (2.3) and (2.5) will be assumed.

LEMMA 2.1.

$$\lim_{m \to \infty} [\sup | p_i^{(n)}(x + x') - p_i^{(n)}(x + x'') |] = 0 , \tag{2.6}$$

where the sup *is the same as in* (2.4) (i.e., *x contains j_0 at least m times); the convergence is uniform in n.*

Proof. We define quantites $\varepsilon_m^{(k)}$ by using $p_i^{(k)}$ instead of p_i in (2.4); then of course $\varepsilon_m^{(1)} = \varepsilon_m$, and the conclusion of the lemma is equivalent to $\varepsilon_m^{(k)} \to 0$ uniformly in k as $m \to \infty$. Now

$$\begin{aligned} & |p_i^{(k)}(x + x') - p_i^{(k)}(x + x'')| \\ & = |\sum_j \{p_i^{(k-1)}(j + x + x')p_j(x + x') - p_i^{(k-1)}(j + x + x'')p_j(x + x'')\}| \\ & \leqq \sum_j p_j(x + x')|p_i^{(k-1)}(j + x + x') - p_i^{(k-1)}(j + x + x'')| \\ & + \sum_j |p_j(x + x') - p_j(x + x'')|p_i^{(k-1)}(j + x + x'') . \end{aligned}$$

Suppose that x contains j_0 m times. Then the second term of the above estimate is less than $N\varepsilon_m$. The absolute value in the first term is less than $\varepsilon_m^{(k-1)}$, but if $j = j_0$ this can be improved to $\varepsilon_{m+1}^{(k-1)}$. Taking account of (2.3) and assuming that $n_0 = 1$, we obtain the estimate

$$\varepsilon_m^{(k)} \leqq N\varepsilon_m + \delta\varepsilon_{m+1}^{(k-1)} + (1 - \delta)\varepsilon_m^{(k-1)} . \tag{2.7}$$

(In case $n_0 > 1$, the same idea can be carried out; the details are more cumbersome and will not be given.)

Now (2.7) can be iterated to obtain an estimate of $\varepsilon_m^{(k)}$ in terms of ε_m. After some computation the result is

$$\begin{aligned} \varepsilon_m^{(k)} \leqq & N\varepsilon_m \sum_{i=0}^{k-1} (1 - \delta)^i + N\varepsilon_{m+1}\delta \sum_{i=0}^{k-2} (i + 1)(1 - \delta)^i \\ & + \cdots + N\varepsilon_{m+l}\delta^l \sum_{i=0}^{k-l-1} \binom{i + l}{i}(1 - \delta)^i + \cdots + N\delta^{k-1}\varepsilon_{m+k-1} . \end{aligned}$$

If the series are extended to infinity, the inequality remains true; calling these (infinite) series $A_0, A_1, \cdots, A_{k-1}$ we have

$$\varepsilon_m^{(k)} \leqq N \sum_{i=0}^{k-1} \varepsilon_{m+i}\delta^i A_i .$$

But it can be shown without much difficulty that

$$A_{l+1} - A_l = (1 - \delta)A_{l+1} ,$$

or $A_{l+1} = A_l/\delta$. Since $A_0 = \delta^{-1}$ we obtain $A_l = \delta^{-(l+1)}$, and hence

$$\varepsilon_m^{(k)} \leqq \delta^{-1} \sum_{i=0}^{k-1} \varepsilon_{m+i} . \tag{2.8}$$

Recalling hypothesis (2.5), the uniform convergence of $\varepsilon_m^{(k)}$ follows from (2.8).

LEMMA 2.2.

$$\lim_{n\to\infty} |p_i^{(n)}(x') - p_i^{(n)}(x'')| = 0 \tag{2.9}$$

and the convergence is uniform in x' and x''.

Proof. For clarity we shall use probabilistic arguments, although a purely analytic rephrasing is not hard. Consider two stochastic processes operating independently with transition probabilities $p_i(x)$, one with the sequence x' for its past history up to time 0 and the other with x''. In view of Lemma 2.1, for any $\varepsilon > 0$ there is an m such that if the two processes have occupied the same states for a period which includes j_0 at least m times and ends sometime before time n, then their probabilities of being in state i at time n differ by at most $\varepsilon/2$. But it follows from condition (2.3) that with probability one, there will sometime be a period of length m during which both processes remain in state j_0. We can take n large enough so that this simultaneous "run" of state j_0 will occur before time n with probability not less than $1 - \varepsilon/2$. For this and all greater values of n, therefore, the two processes have probabilities of occupying state i at time n which differ by at most ε, and this proves (2.9). It is also easy to see from (2.3) and Lemma 2.1 that n can be chosen uniformly in x' and x''.

With this much preparation we shall now prove the first theorem:

THEOREM 2.1. *The quantities*

$$\lim_{n\to\infty} p_i^{(n)}(x) = \pi_i \tag{2.10}$$

exist, are independent of x, *and satisfy* $\sum_I \pi_i = 1$; *the convergence is uniform in* x.

Proof. Applying (2.2) repeatedly, we have

$$p_i^{(n+m)}(x) = \sum_{x_m} p_{i_{m-1}}(x)p_{i_{m-2}}(i_{m-1}+x)\cdots p_{i_0}(i_1+\cdots+i_{m-1}+x)p_i^{(n)}(x_m+x)$$

where $x_m = i_0, i_1, \cdots, i_{m-1}$. Therefore

$$|p_i^{(n+m)}(x) - p_i^{(n)}(x)| \leqq \sum_{x_m} p_{i_{m-1}}(x)\cdots p_{i_0}(i_1+\cdots+i_{m-1}+x)|p_i^{(n)}(x_m+x) - p_i^{(n)}(x)|$$

and by Lemma 2.2, for any ε there is an n such that each term within absolute value signs on the right is less than ε. Since the weights $p_{i_{m-1}}(x)\cdots p_{i_0}(i_1+\cdots+i_{m-1}+x)$ sum to one, we have

$$|p_i^{(n+m)}(x) - p_i^{(n)}(x)| < \varepsilon ,$$

and so $p_i^{(n)}(x)$ has a (uniform in x) limit π_i. Since there are a finite number of states,

$$\sum_i \pi_i = \sum_i \lim_{n\to\infty} p_i^{(n)}(x) = \lim_{n\to\infty} \sum p_i^{(n)}(x) = 1 ,$$

and this completes the proof.

Next we shall define joint probabilities. If x_m is $i_0, i_1, \cdots, i_{m-1}$, let

$$(2.11) \qquad p_{x_m}(x') = p^{(1)}_{x_m}(x') = p_{i_{m-1}}(x')p_{i_{m-2}}(i_{m-1} + x') \cdots p_{i_0}(i_1 + \cdots + i_{m-1} + x) .$$

This is, of course, the probabilitity of executing the sequence of states x_m starting with past history x'. We can define also the higher joint probabilities:

$$(2.12) \qquad p^{(n)}_{x_m}(x') = \sum_{j \in I} p_j(x')p^{(n-1)}_{x_m}(j + x') .$$

Analogues of Lemmas 2.1 and 2.2 can be proved for these quantities by the same arguments used already; in this way it is not difficult to prove

THEOREM 2.2. *The quantities*

$$(2.13) \qquad \lim_{n \to \infty} p^{(n)}_{x_m}(x') = \pi_{x_m}$$

exist, are independent of x', *and satisfy* $\sum_{i_0 \cdots i_{m-1}} \pi_{x_m} = 1$; *the convergence is uniform in* x'.

REMARK. These two theorems imply the existence of a stationary stochastic process with the $p_i(x)$ for transition probabilities. The idea is that the quantities π_{x_m} can be used to define a probability measure on the "cylinder sets" in the space of infinite sequences of members of I, and this measure can then be extended. This stationary process need not concern us further here.

Finally we will prove convergence theorems for certain "moments" which are useful in studying experimental data. The idea is that if we have a stochastic process with the functions $p_i(x)$ for transition probabilities, the probability $p_i(x_m)$ that the state at time m is i given the past history x_m is itself a random variable, and so it makes sense to study $E(p_i^\nu(x_m))$. More formally, define

$$(2.14) \qquad \alpha_i^\nu(m, x) = \sum_{i_0, \cdots, i_{m-1}} p_i^\nu(x_m + x)p_{x_m}(x)$$

where $p_{x_m}(x)$ is defined by (2.11). Thus $\alpha_i^1(m, x)$ is the same as $p_i^{(m)}(x)$. Theorem 2.1 states that $\lim_{m \to \infty} \alpha_i^1(m, x) = \pi_i$ exists. We shall now prove

THEOREM 2.3. *The quantities*

$$(2.15) \qquad \lim_{m \to \infty} \alpha_i^\nu(m, x) = \alpha_i^\nu$$

exist for every positive integer ν; *convergence is uniform in* x *and the limit is independent of* x.

Proof. We use a simple estimate to show that $\alpha_i^\nu(m, x)$ is a Cauchy sequence:

$$\begin{aligned}
&|\alpha_i^\nu(m+k+h, x) - \alpha_i^\nu(m+k, x)| \\
&= |\sum_{x_{m+k+h}} p_i^\nu(x_{m+k+h} + x)p_{x_{m+k+h}}(x) - \sum_{x_{m+k}} p_i^\nu(x_{m+k} + x)p_{x_{m+k}}(x)| \\
&\leqq \sum_{x_{m+k+h}} |p_i^\nu(x_{m+k+h} + x) - p_i^\nu(x_m + x)|p_{x_{m+k+h}}(x) \\
&+ \sum_{x_{m+k}} |p_i^\nu(x_{m+k} + x) - p_i^\nu(x_m + x)|p_{x_{m+k}}(x) \\
&+ |\sum_{x_{m+k+h}} p_i^\nu(x_m + x)p_{x_{m+k+h}}(x) - \sum_{x_{m+k}} p_i^\nu(x_m + x)p_{x_{m+k}}(x)| \,.
\end{aligned}$$

If m is chosen large enough, the first two terms will be arbitrarily small; this involves nothing more than the conditions (resulting from (2.3) and (2.5)) that $\varepsilon_m \to 0$, and that a long sequence x contains j_0 many times with high probability. The last term may be rewritten by carrying out the summation over all the indices except those in x_m; this yields

$$|\sum_{x_m} p_i^\nu(x_m + x)(p_{x_m}^{(k+h)}(x) - p_{x_m}^{(k)}(x))| \leqq \sum_{x_m} |p_{x_m}^{(k+h)}(x) - p_{x_m}^{(k)}(x)|$$

which is small for all h (and for all x) if k is large enough, by Theorem 2.2. Thus if $n = m + k$, $|\alpha_i^\nu(n+h, x) - \alpha_i^\nu(n, x)|$ is small for all h, and this proves that the limit (2.15) must exist; the limit is uniform in x since $\alpha_i^\nu(m, x)$ is uniformly Cauchy. Another estimate along much the same line can be made to show that for any $\varepsilon > 0$,

$$|\alpha_i^\nu(m+k, x) - \alpha_i^\nu(m+k, x')| \leqq \varepsilon$$

provided m and k are large. Since the limit of $\alpha_i^\nu(m+k, x)$ exists as $m + k \to \infty$, we can conclude that the limit is the same for all x.

It is also desirable to consider some additional "cross" moments involving $p_i(x_m)$ for several states at once; accordingly we define

$$\alpha_{j_1 j_2 \cdots j_k}^{\nu_1 \nu_2 \cdots \nu_k}(m, x) = \sum_{x_m} p_{j_1}^{\nu_1}(x_m + x)p_{j_2}^{\nu_2}(x_m + x) \cdots p_{j_k}^{\nu_k}(x_m + x)p_{x_m}(x) \,. \tag{2.16}$$

The following theorem is then a generalization of Theorem 2.3, which treats the case $k = 1$:

THEOREM 2.4. *The quantities*

$$\lim_{m\to\infty} \alpha_{j_1 \cdots j_k}^{\nu_1 \cdots \nu_k}(m, x) = \alpha_{j_1 \cdots \nu_k}^{\nu_1 \cdots \nu_k} \tag{2.17}$$

exist uniformly in x for all non-negative integers $\nu_1 \cdots \nu_k$ and all $j_1 \cdots j_k \in I$, and the limits are independent of x.

The argument used in proving Theorem 2.3 works in this case also with only trivial changes, and need not be repeated. Finally we remark that moments involving several values of n can be considered, and it can be shown that their limits exist also. This provides a generalization of Theorem 2.2.

3. Definition of linear learning models. The models we consider apply to an experimental situation which consists of a sequence of trials. On each trial the subject of the experiment makes a response, which is followed by a reinforcing event. Thus an experiment may be represented by a sequence $(A_1, E_1, A_2, E_2, \cdots A_n, E_n, \cdots)$ of random variables, where the choice of letters follows conventions established in the literature: the value of the random variable A_n is a number j representing the actual response on trial n, and the value of E_n is a number k representing the reinforcing event on trial n. The relevant data on each trial may then be represented by an ordered pair (j, k) of integers with $1 \leqq j \leqq r$, and $0 \leqq k \leqq t$, that is, we envisage in general r responses and $t+1$ reinforcing events. Any sequence of these pairs of integers is a sequence of values of the random variables and thus represents a possible experimental outcome. The general aim of the theory is to predict the probability distribution of the response random variable when a particular distribution, or class of distributions, is imposed on the reinforcement random variable.

In dealing with the general linear model with r responses and $t+1$ reinforcing events we are following the formulation in Chapter 1 of Bush and Mosteller [2], although our notation is somewhat different, being closer to Estes [4] and Estes and Suppes [6].

The theory is formulated for the probability of a response on trial $n+1$ *given* the entire preceding sequence of responses and reinforcements. For this preceding sequence we use the notation x_n. Thus

$$x_n = (k_n, j_n, k_{n-1}, j_{n-1}, \cdots, k_1, j_1)\ .$$

(It is convenient to write these sequences in this order, but note that the numbering here is from past to present, not the reverse as in §2.) Our single axiom is the following linearity assumption:

Axiom L. If $E_n = k$ *and* $P(x_n) > 0$ *then*

$$P(A_{n+1} = j | x_n) = (1-\theta_k)P(A_n = j | x_{n-1}) + \theta_k \lambda_{jk}\ , \tag{3.1}$$

where $0 \leqq \theta_k,\ \lambda_{jk} \leqq 1$ *and* $\sum_j \lambda_{jk} = 1$.

We obtain the linear model studied intensitively in [6] by setting:

$$\begin{cases} \theta_k = \theta & \text{for } k \neq 0 \\ \theta_k = 0 & \text{for } k = 0 \\ \lambda_{jj} = 1 & \\ \lambda_{jk} = 0 & \text{for } j \neq k \\ t = r\ . & \end{cases} \tag{3.2}$$

A linear model satisfying (3.2) we shall term an *Estes Model*, and for such models (3.1) may be replaced by the simpler condition:

$$(3.3)\quad P(A_{n+1} = j \mid x_n) = \begin{cases} (1-\theta)P(A_n = j \mid x_{n-1}) + \theta & \text{if } E_n = j \\ (1-\theta)P(A_n = j \mid x_{n-1}) & \text{if } E_n = k,\ k \neq 0,\ k \neq j \\ P(A_n = j \mid x_{n-1}) & \text{if } E_n = 0\ . \end{cases}$$

Axiom L satisfies the combining classes condition of Bush and Mosteller. Upon replacing θ by $1-\alpha$ in (3.1) essentially their general formulation of the linear model is obtained, although they do not explicitly indicate dependence on the sequence x_n.

We also define here certain moments which are of experimental interest and whose asymptotic properties we investigate subsequently. The *moments* $\alpha^{\nu}_{j,n}$ of the response probabilities at trial n are:

$$(3.4)\qquad \alpha^{\nu}_{j,n} = \sum_{x_{n-1}} P^{\nu}(A_n = j \mid x_{n-1}) P(x_{n-1})\ .$$

And if the appropriate limits exist, we define

$$(3.5)\qquad \alpha^{\nu}_{j} = \lim_{n\to\infty} \alpha^{\nu}_{j,n}\ .$$

The moments (3.4) are formed in an unsymmetrical way; however, they enter in a natural way in the expression of quantities which are easily observed experimentally—for instance, the joint probability $P(A_{n+1} = j, A_n = j)$. (For other examples, see [6].)

We are also interested in studying extensions of the linear model to multiperson situations. We may suppose that we have s subjects in a situation such that the probability of a particular reinforcing event for any one subject will depend in general on preceding responses and reinforcements of the other $s-1$ subjects as well as on his own prior responses and reinforcements. The data on each trial may then be represented by an ordered $2s$-tuple $(j_1, k_1, \cdots, j_s, k_s)$ of integers with $1 \leqq j_i \leqq r_i$, $0 \leqq k_i \leqq t_i$, for $i = 1, \cdots, s$, and any sequence of such tuples represents a possible experimental outcome. Let $A^{(i)}_n$ and $E^{(i)}_n$ be the response and reinforcement random variables for the ith subject on trial n. We may then generalize Axiom L to:

Axiom M. For $1 \leqq i \leqq s$, *if* $E^{(i)}_n = k$ *and* $P(x_n) > 0$ *then*

$$(3.6)\qquad P(A^{(i)}_{n+1} = j \mid x_n) = (1 - \theta^{(i)}_{\nu}) P(A^{(i)}_n = j \mid x_{n-1}) + \theta^{(i)}_k \lambda^{(i)}_{jk}\ ,$$

where $0 \leqq \theta^{(i)}_k,\ \lambda^{(i)}_{jk} \leqq 1$ *and* $\sum_j \lambda^{(i)}_{jk} = 1$.

Experimental tests of Axiom M for two-person situations are reported in Estes [5] and in Atkinson and Suppes [1]. Let $x^{(i)}_{n-1}$ be just the sequence of first $n-1$ responses and reinforcements of subject i. It is a consequence[1] of Axiom M that

$$P(A_n^{(i)} = j \mid x_{n-1}^{(i)}) = P(A_n^{(i)} = j \mid x_{n-1}) ,$$

and it is in terms of $x_{n-1}^{(i)}$ that we define moments $\alpha_{\nu,j,n}^{(i)}$ exactly analogous to (3.4). We shall also be interested in the joint moments

$$\gamma_{j_1,\cdots j_s,n}^{\nu} = \sum_{x_{n-1}} P(A_n^{(1)} = j_1, \cdots, A_n^{(s)} = j_s \mid x_{n-1}) P(x_{n-1}) , \tag{3.7}$$

and their asymptotes $\gamma_{j_1,\cdots,j_s}^{\nu}$ if they exist. To work with these latter moments in terms of Axiom M we need the additional reasonable assumption that when all the $n-1$ preceding responses and reinforcements are given, the s responses on trial n are statistically independent:

Axiom I. If $P(x_{n-1}) > 0$ *then*

$$P(A_n^{(1)} = j_1, \cdots, A_n^{(s)} = j_s \mid x_{n-1}) = \prod_{i=1}^{s} P(A_n^{(i)} = j_i \mid x_{n-1}) .$$

The experimental restriction implied by Axiom I has been satisfied in the multiperson studies employing the linear model.

4. Asymptotic theorems for learning models. After dealing with some matters of notation, we state general theorems on the existence of asymptotic moments. The hypotheses of the theorems give some broad conditions which guarantee ergodic behavior. We begin with the one-person models satisfying Axiom L.

In this section it will be convenient to use some of the notation of §2. Thus we may write $P(A_n = j \mid x_m + x')$ in place of $P(A_n = j \mid x_{n-1})$ to indicate we are interested in the last m terms of x_{n-1}. The "sum" $x_m + x'$ is just the combined sequence x_{n-1}. We reserve the subscript m for counting back m trials from a given trial n.

To clarify the general theorem it is desirable to define in an exact way the notion of the conditional probability of a reinforcing event depending on only a finite number m of past trial outcomes and independent of the trial number.

DEFINITION. A linear model has a *reinforcement schedule with past dependence of length* m if, and only if, for all k, n and n' with $n, n' > m$ and all x_m, x' and x''

$$P(E_n = k \mid x_m + x') = P(E_{n'} = k \mid x_m + x'') . \tag{4.1}$$

(It is understood that x_m includes the response $A_{j,n}$ which precedes $E_{k,n}$ on trial n.) It is to be noticed that the use of n on one side and n' on the other side of (4.1) yields independence of trial number. The term

[1] Proof of this fact is analogous to that of Theorem 4.8 of [**6**].

reinforcement schedule has been used because of its frequent occurrence with approximately this meaning in the experimental literature. For the conditional probabilities of (4.1) we shall use the notation

$$\pi_{k,x_m} = P(E_n = k \mid x_m + x) \; . \tag{4.2}$$

We may now state the first general theorem.

THEOREM 4.1. *Let $\mathscr{L}$ be a linear model such that*

(i) *$\mathscr{L}$ has a reinforcement schedule with past dependence of length m^*,*

(ii) *there is an integer k^* such that*

(a) $\theta_{k^*} \neq 0$

(b) *there is a δ^* and an m_0 such that for all sequences x and all integers n*

$$P(E_{n+m_0} = k^* \mid x_n) \geqq \delta^* > 0 \; .$$

Then the asymptotic moments α_j^ν of $\mathscr{L}$ all exist and are independent of the initial distribution of responses.

Proof. The central task is to characterize $\mathscr{L}$ as a chain of infinite order and show that satisfaction of the hypotheses of the theorem implies satisfaction of conditions (2.3) and (2.5). With this accomplished the asymptotic theorems of §2 may be applied to $\mathscr{L}$. It is most convenient to take as states of the chain the ordered pairs (j, k), where j is the response on trial n, say, and k is the reinforcement on the *preceding* trial. Consider now the reinforcement k^* of the hypothesis of the theorem. Let j^* be a response such that $\lambda_{j^*k^*} \neq 0$. (There is at least one such j^* since $\sum_j \lambda_{jk} = 1$; in the Estes model $j^* = k^*$.) With the pair (j^*, k^*) as the state j_0 of the infinite order chain, we shall establish (2.3) and (2.5).

To verify (2.3), we use (ii)b of the hypothesis and the following equalities and inequalities, which hold for all x and n:

$$\begin{aligned} P(A_{n+m_0+1} &= j^*, E_{n+m_0} = k^* \mid x_n) \\ &= \sum_{x_{m_0-1}} P(A_{n+m_0+1} = j^* \mid E_{n+m_0} = k^*, x_{m_0-1} + x_n) \\ &\qquad \cdot P(E_{n+m_0} \mid x_{m_0-1} + x_n) P(x_{m_0-1} \mid x_n) \; . \end{aligned}$$

Applying Axiom, L, the right-hand side becomes:

$$\begin{aligned} &= \sum_{x_{m_0-1}} [(1 - \theta_{k^*}) P(A_{n+m_0} = j^* \mid x_{m_0-1} + x_n) + \theta_{k^*}\lambda_{j^*k^*}] \\ &\qquad \cdot P(E_{n+m_0} = k^* \mid x_{m_0-1} + x_n) \cdot P(x_{m_0-1} \mid x_n) \\ &\geqq \theta_{k^*}\lambda_{j^*k^*} \sum_{x_{m_0-1}} P(E_{n+m_0} = k^* \mid x_{m_0-1} + x_n) P(x_{m_0-1} \mid x_n) \\ &\geqq \theta_{k^*}\lambda_{j^*k^*} P(E_{n+m_0} = k^* \mid x_n) \\ &\geqq \theta_{k^*}\lambda_{j^*k^*}\delta^* \qquad \text{by (ii)b} \; . \end{aligned}$$

To establish (2.5), consider the following equalities and inequalities:

$$(4.3)\quad |P(A_{n'+1} = j,\, E_{n'} = k \,|\, x + x') - P(A_{n''+1} = j,\, E_{n''} = k \,|\, x + x'')|$$
$$= \pi_{k, x_{m^*}} |P(A_{n'+1} = j \,|\, E_{n'} = k,\, x + x') - P(A_{n''+1} = j \,|\, E_{n''} = k,\, x + x'')|\,,$$

where x_{m^*} means the last m^* terms of x, and where the sequence x contains at least m occurrences of k^*, with $m > m^*$. The equality follows from (i) of the hypothesis, for by virtue of (i)

$$\pi_{k, x_{m^*}} = P(E_{n'} = k \,|\, x + x') = P(E_{n''} = k \,|\, x + x'')\ .$$

Applying Axiom L once to the right-hand side of (4.3) we get, ignoring $\pi_{k, x_{m^*}}$:

$$|P(A_{n'+1} = j \,|\, E_{n'} = k,\, x + x') - P(A_{n''+1} = j \,|\, E_{n''} = k,\, x + x'')|$$
$$= (1 - \theta_k) |P(A_{n'} = j \,|\, x + x') - P(A_{n''} = j \,|\, x + x'')|\ .$$

We do not know that $\theta_k \neq 0$, but as we apply Axiom L repeatedly, we obtain the factor $(1 - \theta_{k^*})$ at least m times, so that

$$(4.4)\quad |P(A_{n'+1} = j,\, E_{n'} = k \,|\, x + x') - P(A_{n''+1} = j,\, E_{n''} = k \,|\, x + x'')|$$
$$\leqq (1 - \theta_{k^*})^m |P(A_{n'-h} = j \,|\, x') - P(A_{n''-h} \,|\, x'')|\,,$$

where h is the length of x[2]. The difference term on the right of this inequality is not more than 1, so that from (4.4) we obtain the estimate for $m > m^*$

$$\varepsilon_m \leqq (1 - \theta_{k^*})^m\,,$$

whence

$$\sum_{m=0}^{\infty} \varepsilon_m < \infty\,,$$

which is (2.5).

On the basis of (2.3) and (2.5) we know from Theorem 2.4 that the asymptotic cross-moments of $\mathscr{L}$ exist and are independent of the initial distribution of responses. But

$$P(A_n = j \,|\, x_{n-1}) = \sum_k P(A_n = j,\, E_{n-1} = k \,|\, x_{n-1})\,,$$

and so the moments $\alpha^{\nu}_{j,n}$ can be expressed as sums of the cross-moments for the infinite order chain $\mathscr{L}$, which insures the existence of the limitting moments (3.5) and that they do not depend upon initial conditions.

There are several remarks to be made about the theorem just

[2] If all $\theta_k \neq 0$, the original condition given in [**3**] would be satisfied; our weaker condition (2.5) allows inclusion of cases where some of the θ_k are 0 (i.e. where there can be trials without a reinforcement).

proved. First, we observe that a simple sufficient (but not necessary) condition for (ii)b is

$$\min_{x_{m^*}} \pi_{k^*, x_{m^*}} \neq 0 \ . \tag{4.5}$$

The interpretation of (4.5) is that the reinforcing event k^* has positive probability on every trial no matter what sequence x_{m^*} of responses and reinforcements preceded. A number of interesting experimental cases of the linear model can be described in terms of (4.5), (i) and (ii)a of Theorem 4.1.

I. *Contingent case with lag v.* In the Estes model let $P(E_n = k \mid A_{n-v} = j, x) = \pi_{kj}(v)$, for all x such that $P(A_{n-v} = j, x) > 0$. To satisfy (4.5), we need only that for some k, $\pi_{jk}(v) \neq 0$ for all j. Experimental data for $v = 0, 1, 2$ are given in Estes [5].

II. *Double contingent case.* Let

$$P(E_n = k \mid A_n = j, A_{n-1} = j', x) = \pi_{k,jj'} \ ,$$

for all x such that $P(A_n = j, A_{n-1} = j', x) > 0$.

Then (i) of Theorem (4.1) is immediately satisfied, and for (ii)a and (4.5) we need a k such that $\theta_k \neq 0$ and for all j and j', $\pi_{k,jj'} \neq 0$.

An interesting fact about (I) and (II) is that although they are simple to test experimentally and their asymptotic response moments exist on the basis of Theorem 4.1, there is no known constructive method for computing the actual asymptotes. (The Estes [5] test of (I) excludes non-reinforced trials which cause the computational difficulties.) It may also be noted that the convergence theorems in Karlin [8] do not in general apply to (II), and apply to (I) only if $v = 0$.

On the basis of the proof of Theorem 4.1 we may, by virtue of Theorem 2.2, conclude that the asymptotic joint probabilities of successive responses also exist:

COROLLARY 1. *If the hypothesis of Theorem* 4.1 *is satisfied, then for every m the limit as* $n \to \infty$ *of*

$$P(A_{n+m} = j_m, A_{n+m-1} = j_{m-1}, \cdots, A_n = j_0)$$

exists.

We may regard the quantities $P(A_n = j \mid x_{n-1})$, for $1 \leqq j \leqq r$ as a random probability vector with an arbitrary joint distribution F_1 on trial 1, and distribution F_n on trial n. The following corollary is a consequence of the existence of the moments α_j^{ν} independent of the initial response probabilities.

COROLLARY 2. *If the hypothesis of Theorem* 4.1 *is satisfied, then there is a unique asymptotic distribution* F_∞, *independent of* F_1 *to which the distributions* F_n *converge.*

For the multiperson situation characterized by Axioms I and M, we have a theorem analogous to Theorem 4.1. For use in the hypothesis of this theorem we define the notion of *reinforcement schedule with past dependence of length* m, exactly as we did in (4.1), namely, we have such a schedule if for all k, $1 \leqq i \leqq s$, all n and n' with $n, n' > m$ and all x_m, x' and x''

$$\begin{aligned}\pi_{k^{(1)},\cdots,k^{(s)},x_m} &= P(\underline{E}_n^{(1)} = k^{(1)}, \cdots, \underline{E}_n^{(s)} = k^{(s)} | x_m + x') \\ &= P(E_n^{(1)} = k^{(1)}, \cdots, E_n^{(s)} = k^{(s)} | x_m + x'') \,.\end{aligned}$$

THEOREM 4.2. *Let* $\mathscr{M}$ *be an s-person linear model such that*

(i) $\mathscr{M}$ *has a reinforcement schedule with past dependence of length* m^*,

(ii) *there are integers* $k^{(i)*}$, *for* $1 \leqq i \leqq s$, *such that*

(a) $\theta^{(i)}_{k^{(i)*}} \neq 0$,

(b) *there is a* δ^* *and an* m_0 *such that for all sequences* x *and all integers* n

$$P(E^{(1)}_{n+m_0} = k^{(1)*}, \cdots, E^{(s)}_{n+m_0} = k^{(s)*} | x_n) \geqq \delta^* > 0 \,.$$

Then the asymptotic moments $\gamma^{\nu}_{j^{(1)},j^{(2)},\cdots,j^{(s)}}$ *of* $\mathscr{M}$ *all exist and are independent of the initial distribution of responses.*

Proof. The states of the chain are now defined as $2s$-tuples $(i^{(1)}, \cdots, j^{(s)}, k^{(1)}, \cdots, k^{(s)})$, where $j^{(i)}$ is the response made by the ith subject and $k^{(i)}$ is the reinforcement for that subject on the preceding trial. Using the reinforcements $k^{(i)*}$ of the hypothesis, let $j^{(i)*}$ be such that $\lambda^{(i)}_{j^{(i)*}k^{(i)*}} \neq 0$. We take $(j^{(1)*}, \cdots, j^{(s)*}, k^{(1)*}, \cdots, k^{(s)*})$ as the state j_0 for which we establish (2.3) and (2.5). To simplify notation, it is convenient to define:

$$\begin{aligned}&P_{n+1}(j, k | x) = P(A^{(1)}_{n+1} = j^{(1)}, \cdots, A^{(s)}_{n+1} = j^{(s)}, E^{(1)}_n = k^{(1)}, \cdots, E^{(s)}_n = k^{(s)} | x), \\ &p_{n+1}(j^{(i)} | k, x) = P(A^{(i)}_{n+1} = j^{(i)} | E^{(1)}_n = k^{(1)}, \cdots, E^{(s)}_n = k^{(s)}, x)\,, \\ &\pi_{k,m^*} = \pi_{k^{(1)},\cdots,k^{(s)*},x_{m^*}} \,.\end{aligned}$$

Moreover, we omit the superscript notation from θ and λ.

To verify (2.3) we proceed exactly as in the proof of Theorem 4.1, applying now Axioms I and M instead of L, and we obtain that

$$p_{n+m_0+1}(j, k | x_n) \geqq \prod_{i=1}^{s} \theta_{k^{(i)*}} \lambda_{j^{(i)*}k^{(i)*}} \delta^* \,.$$

For (2.5), we first observe that by virtue of (i) of the hypothesis and Axiom I

$$|p_{n'+1}(j, k|x + x') - p_{n''+1}(j, k|x + x'')|$$
$$= \pi_{k,m^*}|\prod_{i=1}^{s} p_{n'+1}(j^{(i)}|k, x + x') - \prod_{i=1}^{s} p_{n''+1}(j^{(i)}|k, x + x'')| .$$

We notice next that the right-hand side is

$$\leqq \pi_{k,m^*}\Big\{ p_{n'+1}(j^{(1)}|k, x + x')|\prod_{i=2}^{s} p_{n'+1}(j^{(i)}|k, x + x')$$
$$- \prod_{i=2}^{s} p_{n''+1}(j^{(i)}|k, x + x'')|$$
$$+ \prod_{i=2}^{s} p_{n''+1}(j^{(i)}|k, x+x'')|p_{n'+1}(j^{(1)}|k, x + x') - p_{n''+1}(j^{(1)}|k, x + x'')| .$$

Continuing this same development, we obtain:

$$\leqq \sum_{i=1}^{s} |p_{n'+1}(j^{(i)}|k, x + x') - p_{n''+1}(j^{(i)}|k, x + x'')| .$$

And by the line of reasoning used in the proof of Theorem 4.1, if the sequence x contains state $(j^{(1)*}, \cdots, k^{(s)*})$ at least m times the last quantity is

$$\leqq \sum_{i=1}^{s} (1 - \theta_{k^{(i)*}})^m .$$

Provided $m > m^*$ this inequality yields an estimate of ε_m from which we conclude that (2.5) holds. The existence of the asymptotic moments then follows from the theory of §2 as in the case of Theorem 4.1. Q.E.D.

A pair of corollaries follow from the theorem just proved which are exactly like the two given after Theorem 4.1.

Finally, we want to remark that Axiom L involves linear functions which are distance diminishing, i.e., have slope less than one. The asymptotic results of this section apply to many learning models in which these linear functions are replaced by non-linear functions having this property.

References

1. Richard C. Atkinson, and Patrick Suppes, *An analysis of two-person game situations in terms of statistical learning theory*, J. of Experimental Psychology, **55** (1958), 369-378.
2. Robert R. Bush, and Frederick Mosteller, *Stochastic Models for Learning*, New York, 1955.
3. W. Doeblin, and R. Fortet, *Sur des chaînes à liaisons complètes*, Bull. Soc. Math. France, **65** (1937), 132-148.
4. W. K. Estes, *Theory of learning with constant, variable, or contingent probabilities of reinforcement*, Psychometrika, **22** (1957), 113-132.
5. ———, *Of models and men*, Amer. Psychologist, **12** (1957), 609-617.

6. W. K. Estes, and Patrick Suppes, *Foundations of Statistical Learning Theory*, I. *The Linear Model for Simple Learning*, Technical Report No. 16, Contract Nonr 225(17), Applied Mathematics and Statistics Laboratory, Stanford University, 1957. An abridged version appears as Chapter 8 of *Studies in Mathematical Learning Theory*, edited by R. R. Bush and W. K. Estes, Stanford University Press, 1959.
7. T. E., Harris, *On chains of of infinite order*, Pacific J. Math., **5** (1955), 707–724.
8. Samuel Karlin, *Some random walks arising in learning models* I, Pacific J. Math., **3** (1953), 725–756.
9. Maurice Kennedy, *A convergence theorem for a certain class of Markoff processes*, Pacific J. Math., **7** (1957), 1107-1124.

STANFORD UNIVERSITY

ON RADICALS AND CONTINUITY OF HOMOMORPHISMS INTO BANACH ALGEBRAS

EDITH H. LUCHINS

1. Introduction.

All Banach algebras considered are over the real field and all homomorphisms considered are algebraic (real-linear). An algebra is called semi-simple, strongly semi-simple, or strictly semi-simple, if its Jacobson radical [5], Segal radical [10], or strict radical [8], respectively, is the zero ideal; that is, if its regular maximal right ideals, its regular maximal two-sided ideals, or those of its two-sided ideals which are regular maximal right ideals, intersect in the zero ideal. Rickart [9, Corollary 6.3] proved that a semi-simple commutative Banach algebra has the property that every homomorphism of a Banach algebra into it is continuous. Call an algebra with this property an *absolute* algebra. Yood [12, Theorem 3.5] proved that every homomorphism of a Banach algebra onto a dense subset of a strongly semi-simple Banach algebra is continuous. Thus a strongly semi-simple Banach algebra is "almost" absolute. The question arose: Is a (noncommutative) semi-simple or strongly semi-simple Banach algebra necessarily absolute? A negative answer is furnished in the present note. It is shown that in order for a Banach algebra to be absolute it is sufficient that it be strictly semi-simple and necessary that it have zero as its only nilpotent element. The latter condition is shown to be sufficient for some special Banach algebras to be absolute.

2. Necessary condition for a Banach algebra to be absolute.

THEOREM 1. *An absolute Banach algebra has no nonzero nilpotent elements.*

Proof. Suppose the Banach algebra B contains a nonzero nilpotent element. Then there exists a nonzero $v \in B$ such that $v^2 = 0$. Let A be an infinite dimensional Banach algebra such that $A^2 = (0)$. Since A is an infinite dimensional complete vector space, there exists a discontinuous linear functional on A; denote it by $f(x)$. Let $\pi(x) = f(x)v$. Since $f(x)$ is linear and $v^2 = 0$, π is seen to be a homomorphism of A into B.

Let $\|y\|$ be a Banach norm for B. Then $\|\pi(x)\| = |f(x)|\ \|v\|$ since $f(x)$ is a scalar. Since $f(x)$ is discontinuous $|f(x)|$ is not bounded and

Received February 5, 1959. This report was prepared while the author held the New York State Fellowship, 1957-8, of the American Association of University Women. It is based in part on the author's doctoral dissertation written at the University of Oregon under the sponsorship of Professor Bertram Yood. Thanks are offered to the referee for helpful suggestions.

hence π is not bounded. Thus π is discontinuous so that B is not absolute.

It is known that if a Banach algebra B is semi-simple and has a unique norm then a homomorphism of a Banach algebra *onto* B is necessarily continuous [9, Theorem 6.2]; that the resulting proposition is false if the word *into* is substituted for *onto* follows from our Theorem 1. Indeed, this theorem shows that for a Banach algebra to be absolute it is not sufficient that it have the properties of being simple, semi-simple, strongly semi-simple and having an identity and a unique norm. Thus, these properties are possessed by the algebra of all 2 by 2 matrices over the reals, under a Banach norm, and yet, since this algebra has nonzero nilpotent elements, Theorem 1 shows that it is not absolute.

3. Sufficient condition for a Banach algebra to be absolute. In [8] there was introduced the concept of a strictly semi-simple algebra. It was shown [8, Theorems 2 and 3] that a Banach algebra B is strictly semi-simple if and only if it is isomorphic to a subalgebra of $C(X, Q)$, the algebra of quaternion-valued functions continuous, and vanishing at ∞, on a locally compact Hausdorff space X.

THEOREM 2. *A strictly semi-simple Banach algebra B is absolute.*

Proof. Let A be a Banach algebra with a homomorphism T into $B \subset C(X, Q)$. Let $T_x(a) = T(a)(x)$. The kernel of T_x is closed since Q is simple, and therefore T_x is continuous, whence T_x is of bound 1. That T is continuous can now be shown by the 6-line argument of Loomis [7, p. 77]. One could also use [12, Theorem 3.5].

COROLLARY 1 (Rickart). *A semi-simple commutative Banach algebra is absolute.*

4. Concerning some special Banach algebras. For each subset S of a Banach algebra B, let $S_L(S_R)$ denote the set of all left (right) annihilators of S. B is called an annihilator algebra [3] if $B_R = 0 = B_L$ and if $I_L \neq 0$ $(I_R \neq 0)$ for each proper closed right (left) ideal I, where 0 denotes the zero ideal.

Lemma 1 is due to Forsythe and McCoy [4, p. 524].

LEMMA 1. *In a ring without nonzero nilpotent elements every idempotent is in the center.*

THEOREM 3. *That a Banach algebra B have zero as its only nilpotent element is both a necessary and a sufficient condition for B to be either strictly semi-simple or absolute, provided any of the following conditions is satisfied:*

(a) *B is finite-dimensional.*
(b) *B satisfies the descending chain condition on right ideals.*
(c) *B is a semi-simple annihilator algebra.*

Proof. If B is strictly semi-simple, then $B \subset C(X, Q)$ by [8] and hence has only zero as a nilpotent element. If B is absolute, then zero is its only nilpotent element by Theorem 1. Conversely, suppose B has no nonzero nilpotent elements.

Suppose condition (a) or (b) holds. Then B has a nilpotent radical and therefore is semi-simple; also B is then a direct sum of division algebras and therefore has the property that every left (or right) ideal is two-sided [2, p. 463]. Thus B is strictly semi-simple and therefore absolute by Theorem 2.

Suppose condition (c) holds. Let M be any regular maximal right ideal in B. Bonsall and Goldie [3, pp. 155-6] show that for any semi-simple annihilator algebra B, $M_L = Be$ where e is a nonzero idempotent, B is a minimal (closed) left ideal, eB a minimal (closed) right ideal, $(eB)_L$ a maximal left ideal, and $(Be)_R = M$.

If B has no nonzero nilpotent elements, then e is in the center by Lemma 1 so that $Be = eB$ is a two-sided ideal. But the left and right annihilators of a closed two-sided ideal are identical [3, p. 159] so that $(eB)_L = (Be)_R = M$.

Since $(eB)_L$ is a left ideal, M, which was any regular maximal right ideal in B, has been shown to be a left ideal. Thus B is strictly semi-simple since it is semi-simple by hypothesis, and therefore absolute by Theorem 2.

COROLLARY 2. *An H^* algebra B is commutative if and only if any of the following properties is satisfied:*
(a) *B has no nonzero nilpotent elements.*
(b) *B is strictly semi-simple.*
(c) *B is absolute.*

Proof. An H^* algebra is the closure of the direct sum of matrix algebras M_σ [1, pp. 379–380]. If condition (a) holds, then each M_σ must have zero as its only nilpotent element and therefore must be one-dimensional. Hence each M_σ is generated by an idempotent e_σ which, by Lemma 1, is in the center. For $u, v \in \Sigma M_\sigma$, $u = \Sigma r_k e_k$, $v = \Sigma s_t e_t$, r_k, s_t scalars, $uv = vu$ so that ΣM_σ is commutative and therefore so is its closure, B. Thus condition (a) implies that B is commutative.

Suppose B is commutative. Since an H^* algebra is semi-simple, if commutative it is strictly semi-simple and $\subset C(X, Q)$ by [8], so that zero is its only nilpotent element. Hence condition (a) prevails.

The remainder of the corollary follows immediately from Theorem 3 since an H^* algebra is a semi-simple annihilator algebra [6, p. 697].

BIBLIOGRAPHY

1. W. Ambrose, *Structure theory for a special class of Banach algebras*, Trans. Amer. Math. Soc. **57** (1945), 364-386.

2. R. Arens and I. Kaplansky, *Topological representation of algebras*, Trans. Amer. Math. Soc. **63** (1948), 457-481.

3. F. F. Bonsall and A. W. Goldie, *Annihilator algebras*, Proc. London Math. Soc. **5** (1954), 154-167.

4. A. Forsythe and N. H. McCoy, *On the commutativity of certain rings*, Bull. Amer. Math. Soc. **52** (1946), 523-526.

5. N. Jacobson, *The radical and semi-simplicity for arbitrary rings*, Amer. J. Math. **67** (1945), 300-320.

6. I. Kaplansky, *Dual rings*, Ann. of Math. **49** (1948), 689-701.

7. L. H. Loomis, *An introduction to abstract harmonic analysis*, New York: Van Nostrand, 1953.

8. E. H. Luchins, *On strictly semi-simple Banach algebras*, Pacific J. Math. (to appear.)

9. C. E. Rickart, *The uniqueness of norm problem in Banach algebras*, Ann. of Math. **51** (1950), 615-628.

10. I. E. Segal, *The group algebra of a locally compact group*, Trans. Amer. Math. Soc. **61** (1947), 69-105.

11. B. Yood, *Topological properties of homomorphisms between Banach algebras*, Amer. J. Math. **76** (1954), 155-167.

12. ———, *Homomorphisms on normed algebras*, Pacific J. Math. **8** (1958), 373-381.

UNIVERSITY OF OREGON
AND
UNIVERSITY OF MIAMI

MULTIPLICATION FORMULAE FOR THE *E*-FUNCTIONS REGARDED AS FUNCTIONS OF THEIR PARAMETERS

T. M. MACROBERT

1. Introduction. The formulae to be proved are

$$\sum_{i,-i} \frac{1}{i} E(p; m\alpha_r : q; m\rho_s : ze^{i\pi})$$

$$= (2\pi)^{-\frac{1}{2}(m-1)(p-q-1)} m^{m(\Sigma\alpha_r - \Sigma\rho_s) - \frac{1}{2}(p-q-1)}$$

(1)
$$\times \sum_{i,-i} \frac{1}{i} E \left\{ \begin{matrix} \alpha_1, \alpha_1 + \frac{1}{m}, \cdots, \alpha_1 + \frac{m-1}{m}, \cdots, \alpha_p + \frac{m-1}{m} : \\ \frac{1}{m}, \frac{2}{m}, \cdots, \frac{m-1}{m}, \rho_1, \cdots, \rho_q + \frac{m-1}{m} : \end{matrix} \left(\frac{z}{m^{p-q-1}}\right)^m e^{i\pi} \right\},$$

where m is a positive integer, $p > q + 1$, and $|\mathrm{amp}\, z| < 1/2(p - q - 1)\pi$. If $p \leq q + 1$, both sides vanish identically.

For all values of p and q

$$E(p; m\alpha_r : q; m\rho_s : ze^{\pm i\pi})$$

$$= (2\pi)^{-\frac{1}{2}(m-1)(p-q-1)} m^{m(\Sigma\alpha_r - \Sigma\rho_s) - \frac{1}{2}(p-q+1)}$$

(2)
$$\times \sum_{n=0}^{m-1} \left(\frac{m^{p-q-1}}{z}\right)^n E \left\{ \begin{matrix} \alpha_1 + \frac{n}{m}, \cdots, \alpha_1 + \frac{n+m-1}{m}, \cdots, \alpha_p + \frac{n+m-1}{m} : \\ \frac{n+1}{m}, \frac{n+2}{m}, \cdots * \cdots, \frac{n+m}{m}, \rho_1 + \frac{n}{m}, \cdots, \end{matrix} \right.$$

$$\left. \rho_q + \frac{n+m-1}{m} : \left(\frac{z}{m^{p-q-1}}\right)^m e^{\pm i\pi} \right\},$$

the asterisk indicating that the parameter m/m is omitted.

The proof of (1) is based on the formula ([1], p. 374)

(3)
$$E(p; \alpha_r : q; \rho_s : z) = \frac{1}{2\pi i} \int \frac{\Gamma(\zeta) \Pi \Gamma(\alpha_r - \zeta)}{\Pi \Gamma(\rho_s - \zeta)} z^{\zeta} d\zeta\ ,$$

where the integral is taken up the η -axis, with loops, if necessary, to ensure that the pole at the origin lies to the left and the poles at

Received January 7, 1959.

$\alpha_1, \alpha_2, \cdots, \alpha_p$ to the right of the contour. Zero and negative integral values of the α's and ρ's are excluded, and the α's must not differ by integral values. The contour must be modified if $p < q+1$; and if $p = q+1, |z| < 1$; but we are here concerned only with the case $p > q+1$. Then z must satisfy the condition $|\text{amp } z| < 1/2(p-q+1)\pi$.

From (3) it follows that, if $p > q+1, |\text{amp } z| < 1/2(p-q-1)\pi$,

$$(4) \qquad \sum_{i,-1} \frac{1}{i} E(p; \alpha_r: q; \rho_s: ze^{i\pi}) = \frac{1}{i} \int \frac{\Pi\Gamma(\alpha_r - \zeta)}{\Gamma(1-\zeta)\Pi\Gamma(\rho_s - \zeta)} z^\zeta d\zeta .$$

For, on substituting on the left from (3), a factor $(e^{i\pi\zeta} - e^{-i\pi\zeta})$ appears in the integral, and

$$\Gamma(\zeta) \sin \pi\zeta = \pi/\Gamma(1-\zeta) .$$

The three following formulae ([1], pp. 154, 406, 407) are also required.

If m is a positive integer,

$$(5) \qquad \Gamma(mz) = (2\pi)^{\frac{1}{2}-\frac{1}{2}m} m^{mz-\frac{1}{2}} \Gamma(z)\Gamma\left(z + \frac{1}{m}\right) \cdots \Gamma\left(z + \frac{m-1}{m}\right);$$

$$(6) \qquad \begin{aligned} &\int_0^\infty e^{-\lambda}\lambda^{k-1} E(p; \alpha_r: q; \rho_s: z/\lambda^m) d\lambda \\ &= (2\pi)^{\frac{1}{2}-\frac{1}{2}m} m^{k-\frac{1}{2}} E(p+m; \alpha_r: q; \rho_s: z/m^m) , \end{aligned}$$

where $R(k) > 0, \alpha_{p+1+\nu} = (k+\nu)/m, \nu = 0, 1, 2, \cdots, m-1$;

$$(7) \qquad \begin{aligned} &\frac{1}{2\pi i} \int e^\zeta \zeta^{-\rho} E(p; \alpha_r: q; \rho_s: \zeta^m z) d\zeta \\ &= (2\pi)^{\frac{1}{2}m-\frac{1}{2}} m^{\frac{1}{2}-\rho} E(p; \alpha_r: q+m; \rho_s: zm^m) , \end{aligned}$$

where the contour of integration starts from $-\infty$ on the ξ-axis, passes round the origin in the positive direction, and ends at $-\infty$ on the ξ-axis, amp ζ being $-\pi$ initially, and $\rho_{q+1+\nu} = (\rho+\nu)/m, \nu = 0, 1, 2, \cdots, m-1$.

2. Proofs of the formulae. On applying (4) on the left of (1) and replacing ζ by $m\zeta$ the left hand side becomes

$$\frac{m}{i} \int \frac{\pi\Gamma(m\alpha_r - m\zeta)}{\Gamma(1-m\zeta)\pi\Gamma(m\rho_s - m\zeta)} z^{m\zeta} d\zeta .$$

Here apply (5) and get

$$(2\pi)^{-\frac{1}{2}(m-1)(p-q-1)} m^{m(\Sigma\alpha_r - \Sigma\rho_s)-\frac{1}{2}(p-q-1)}$$

$$\times \frac{1}{i}\int \frac{\Pi\left\{\Gamma(\alpha_r-\zeta)\Gamma\left(\alpha_r+\frac{1}{m}-\zeta\right)\cdots\Gamma\left(\alpha_r+\frac{m-1}{m}-\zeta\right)\right\}}{\Gamma(1-\zeta)\Gamma\left(\frac{1}{m}-\zeta\right)\cdots\Gamma\left(\frac{m-1}{m}-\zeta\right)\Pi\left\{\Gamma(\rho_s-\zeta)\cdots\Gamma\left(\rho_s+\frac{m-1}{m}-\zeta\right)\right\}}$$

$$\times\left(\frac{z}{m^{p-q-1}}\right)^{m\zeta} d\zeta\ ,$$

and from (4), this is equal to the right hand side of (1).

Formula (2) can be obtained by showing that

$$E(:\ :e^{\pm i\pi}z) = e^{1/z}$$

$$= \sum_{n=0}^{m-1} \frac{(1/z)^n}{n!} F\left\{;\frac{n+1}{m},\cdots * \cdots,\frac{n+m}{m};(mz)^{-m}\right\}$$

$$= (2\pi)^{\frac{1}{2}m-\frac{1}{2}} m^{-\frac{1}{2}} \sum_{n=0}^{m-1}\left(\frac{1}{mz}\right)^n E\left\{:\frac{n+1}{m},\cdots * \cdots,\frac{n+m}{m}: e^{\pm i\pi}(mz)^m\right\}\ ,$$

and then generalizing by employing (6) and (7).

Note 1. Ragab's formula [2]

$$\sum_{i,-i}\frac{1}{i}\int_0^\infty e^{-pt}E\left(\alpha,\alpha+\frac{1}{m},\cdots,\alpha+\frac{m-1}{m}:\ :e^{i\pi}zm^{-m}/t\right)dt$$

(8)

$$= (2\pi)^{\frac{1}{2}+\frac{1}{2}m} m^{-m\alpha-\frac{1}{2}} p^{\alpha-1} z^\alpha \exp(-\,p^{1/m}z^{1/m})\ ,$$

where m is a positive integer greater than 1, p is positive, $|\text{amp}\ z| < 1/2(m-1)\pi$, can be derived by substituting on the left from (4), changing the order of integration, evaluating the inner integral, applying (5), replacing ζ by $\alpha - \zeta/m$, and applying (3).

Note 2. It has been pointed out by a referee that there seems to be some connection between the formulae of this paper and certain formulae of Meijer's for the G-function which are reproduced on pages 209, 210 of the first volume of Higher Transcendental Functions [McGraw Hill Book Co., 1953].

References

1. T. M. MacRobert, *Functions, of a complex variable* (4th edition, London, 1954).
2. F. M. Ragab, *The inverse Laplace transform of an exponential function*, New York University, Institute of Mathematical Sciences, Astia Document No. AD 133670.

THE UNIVERSITY
GLASGOW

CLASSES OF MINIMAL AND REPRESENTATIVE DOMAINS AND THEIR KERNEL FUNCTIONS

MICHAEL MASCHLER

1. Introduction. In connection with the problem of obtaining classes of conformally equivalent domains in the space of one or several complex variables, S. Bergman [3] introduced two kinds of canonical domains named *minimal domains* and *representative domains*. Since the mapping functions onto these domains were expressed in a closed form by using the Bergman kernel function and its derivatives, it was possible to deduce interesting properties of the kernel function which, in turn, provided more information about the canonical domains. (See S. Bergman [1] [3], M. Schiffer [9], M. Maschler [7]).

The object of this paper is to discuss "minimal domains" and "representative domains" with respect to certain subclasses of analytic functions, and to deduce solutions to some extremal problems. In addition, differential equations are obtained for the kernel function, which are valid for various classes of domains. The methods we use apply to the theory of functions of several complex variables as well, but first, the case of one complex variable should be clarified.

Let D be a plane domain having a boundary of positive capacity. We consider the class of analytic functions $w = f(z)$ which have single-valued, regular derivatives in D, and which possess developments of the form

$$(1.1) \qquad w = (z - t) + a_{m+1}(z - t)^{m+1} + a_{m+2}(z - t)^{m+2} + \cdots$$

in the neighborhood of a point t in D. There exists one function in this class which maps D onto a domain having the smallest area[1]. This latter domain will be called an *m-minimal domain* with the origin as center. For $m = 1$ we obtain the ordinary minimal domains.

As $w = f(z)$ may be multivalued and non-univalent, one has to extend the theory of the kernel function to domains on a Riemann surface, which may have "identified points", (That is, points which correspond to a single point of a univalent domain, under a conformal mapping). The ideas of this extension are not new and are treated here for the

Received January 7, 1959. This research was sponsored in part by the Office of Scientific Research of the Air Research and Development Command, United States Air Force, through its European Office, under Contract No. AF 61 (052)-04. Some of the results of this paper were presented in: (1) "Domain functions and conformal mappings with applications to extremal problems", a thesis for the Ph. D. degree at the Hebrew University, Jerusalem. (2) The Bull. of the Research council of Israel. Vol 7F (1957) p. 42. (Abstract).

[1] The area is defined by $\iint_D |f'(z)|^2 \, d\omega$, where $d\omega$ is the area element.

sake of completeness. We then treat Bergman's problem of minimizing the integral

$$\iint_D |f(z)|^2 \, d\omega \tag{1.2}$$

where the functions $f(z)$ are single-valued, analytic, regular in the domain D and satisfy, at a non-branch point t in D, the condition

$$f^{(\nu)}(t) = X_\nu,\ X_\nu \text{ are constants, } \nu = 0, 1, 2, \cdots, m\ . \tag{1.3}$$

(See S. Bergman [2].) We prove that these minimizing functions are transformed, under a conformal mapping which is locally univalent at t, onto similar minimizing functions for the image domain, multiplied by the derivative of the mapping function. (The constants (1.3) are transformed linearly).

The mapping function onto an m-minimal domain can be expressed in a closed form in terms of the kernel function and its derivatives [Section 3]. This leads to a local condition for the kernel function, satisfied if and only if the kernel function belongs to an m-minimal domain.

Simply-connected m-minimal domains are always images of a 1-minimal simply-connected domain (i. e., a circle), under a mapping function which is a polynomial of degree at most m, and vice versa [Section 4]. This is no longer true, in general, for the case of multiply-connected m-minimal domains [Section 7]; however, each choice for the values of the first m derivatives of the mapping functions at the center of a 1-minimal domain, determines a mapping onto an m-minimal domain with the same center [Section 4].

The shape of the doubly-connected 1-minimal domains is studied in Section 5. It is shown that the 1-minimal doubly-connected domain always has identified points, provided that no boundary component is reduced to a single point. Therefore, these minimal domains are different from those studied by P. Kufareff [6], which he obtained by restricting attention only to single-valued mapping functions.

Let $M_D(z, t)$ be a minimizing function of (1.2), for functions satisfying in (1.3) the values

$$X_0 = 1,\ X_1 = X_2 = \cdots = \cdots X_{m-1} = 0\ , \qquad m \geqq 1\ . \tag{1.4}$$

Let $M_D^*(z, t)$ be a similar function for the case

$$X_0 = 0\ ,\ X_1 = 1\ ,\ X_2 = X_3 = \cdots = X_m = 0\ , \qquad m \geqq 1\ . \tag{1.5}$$

The function $[M_D^*(z, t)/M_D(z, t)]$ satisfies (1.1) and remains invariant under a conformal mapping which satisfies (1.1) [Section 6]. This function is said to map D onto an *m-representative domain* with the origin as center. In general, it is different from the m-minimal domain with the same center, but if both domains coincide and have the same center, say at the origin, then the minimizing functions for the m-minimal and m-representative domain Δ satisfy the differential equation

$$\frac{d}{d\zeta}\left(\frac{M_\Delta^*(\zeta, 0)}{M_\Delta(\zeta, 0)}\right) = M_\Delta(\zeta, 0) \tag{1.6}$$

for $\zeta \in \Delta$ [Section 6]. The interest in this relation is that it remains invariant under each transformation $w = f(\zeta)$ which satisfies $f'(0) = 1$, $f^{(\nu)}(0) = 0$, $\nu = 2, 3, \cdots, m$. Thus, this relation holds for a general class of conformally equivalent domains.

2. The Bergman kernel function for generalized domains. Various extremal problems in conformal mappings yield, as solutions, a mapping function which may be meromorphic, and/or many-valued. In order to treat such problems, it is desirable to extend the concept of a domain and its Bergman kernel function. Making use of known ideas (see e. g., S. Bergman [4], p. 33, and R. Nevanlinna [8]), we proceed as follows:[2]

Let D be a univalent domain in the z-plane, where boundary has a positive capacity. Let $w=w(z)$ be a function of z, defined for $z \in D$. We demand that $w'(z)$ exists, that it is a *single-valued*, meromorphic function for $z \in D$, and does not vanish identically.

Among the set of points: $\{w(z) \mid z \in D\}$, we identify all the images of the same point z. $w(z_1)$ and $w(z_2)$ are said to be different points if $z_1 \neq z_2$. The obtained set is called: a generalized domain.

EXAMPLES.

a. If D_1 is the unit circle $|z| < 1$, and $w(z) = z^2$, then the generalized domain Δ_1 consists of two coverings of the unit circle with a branch point $w = 0$.

b. If D_2 is the ring $1 < |z| < e$, and $w(z) = \log z$, then the generalized domain Δ_2 is the strip $0 < \Re w < 1$, where points w_1 and w_2, which satisfy

$$\Re w_1 = \Re w_2 , \ \Im w_1 \equiv \Im w_2 \pmod{2\pi} , \tag{2.1}$$

are identified.

DEFINITION. Let Δ be a generalized domain, obtained from a (classical) domain D, according to the above procedure. We say that a function $F(w)$ belongs to the class $\mathscr{L}^2(\Delta)$ if there exists a function $f(z)$, belonging to the class[3] $\mathscr{L}^2(D)$ such that

[2] The proofs are simple and we omit them.

$$F(w) = f(z(w)) \cdot \frac{dz}{dw}\,, \quad z \in D\,. \tag{2.2}$$

It is clear that $F(w)$ may have a pole at a branch point of $\varDelta$. Integration over the domain $\varDelta$ is defined by the relation

$$\iint_{\varDelta} G(w)\, d\omega = \iint_{D} G(w(z)) \cdot |\, w'(z)\,|^2\, d\omega_z\,, \tag{2.3}$$

whenever the right-hand side exists. Here $d\omega_z$ denotes the area element in the z-plane.

Let D be a univalent, positively bounded domain, without identified points, which generates, as described above, two generalized domains $\varDelta$ and $\varDelta^*$, then the mappings of D onto $\varDelta$ and $\varDelta^*$ determine a one-to-one mapping:

$$w^* = w^*(w)\,, \quad w \in \varDelta\,, \tag{2.4}$$

from $\varDelta$ onto $\varDelta^*$. This mapping will be called a conformal mapping. All these definitions do not depend on the particular choice of D.

From now on, unless otherwise stated, we shall use the term "domain" to mean a generalized domain. All such domains have the property that they can be mapped conformally onto univalent domains with a boundary of positive capacity and without identified points.

Let D be a (generalized) domain. Introducing a scalar multiplication

$$(f; \bar{g}) = \iint_D f \cdot \bar{g}\, d\omega\,, \tag{2.5}$$

makes $\mathscr{L}^2(D)$ a Hilbert space which possesses a Bergman kernel function

$$K_D(z, \bar{t}) = \sum_{\nu=1}^{\infty} \varphi_\nu(z)\overline{\varphi_\nu(t)}\,, \tag{2.6}$$

where $\varphi_\nu(z)$, $\nu = 1, 2, \cdots$ is a complete orthonormal system.

The kernel function depends only on the domain D and not on the particular choice of the complete orthonormal system. As a function of z, for each fixed t which is not a branch point, the kernel function belongs to the class $\mathscr{L}^2(D)$. It may or may not be singular if t is a branch point. $K_D(t, \bar{t}) > 0$ and takes the value infinity if t is a branch point.

If a domain D can be mapped conformally onto a domain $\varDelta$, by a mapping function $w(z)$, $z \in D$, then the kernel functions of D and $\varDelta$ satisfy the relation:

$$K_D(z, \bar{t}) = K_\varDelta(w(z), \overline{w(t)}) \cdot w'(z) \cdot \overline{w'(t)}, \; z, t \in D\,. \tag{2.7}$$

[3] A function $f(z)$ is said to belong to the class $\mathscr{L}^2(D)$ if it is single-values and regular in the domain D and satisfies:

$$\iint_D |\, f(z)\,|^2\, dx\, dy < \infty, \; z = x + iy$$

The integration is in the Lebesgue sense.

(This relation is to be understood in the sense that the ratio between the two sides approaches 1 if z, or t, or both variables, approach a singularity point of the kernel function.) We shall end this section by stating an important theorem of S. Bergman for the generalized domains. (See [4], p. 26):

THEOREM 1. *Let D be a (generalized) domain, and t a fixed point in D, which is not a branch point. Consider the functions $f(z)$, $z \in D$, which belong to the class $\mathscr{L}^2(D)$ and satisfy:*

$$f^{(\nu)}(t) = X_\nu\,,\ \nu = 0, 1, 2, \cdots, m\,; \tag{2.8}$$

where X_ν are fixed complex numbers and $f^{(\nu)}(t)$ is the derivative of the order ν of $f(z)$ at the point t; then there exists among them one and only one function $M_D^{X_0, X_1, \cdots, X_m}(z, t)$ which minimizes the integral $\iint_D |f(z)|^2 d\omega$. This function can be represented in a closed form by using the kernel function and its derivatives:

$$M_D^{X_0, X_1, \cdots, X_m}(z, t) = -\frac{\begin{vmatrix} 0 & K_{0\bar{0}}(z, \bar{t}) & K_{0\bar{1}}(z, \bar{t}) & \cdots & K_{0\bar{m}}(z, \bar{t}) \\ X_0 & K_{0\bar{0}} & K_{0\bar{1}} & \cdots & K_{0\bar{m}} \\ X_1 & K_{1\bar{0}} & K_{1\bar{1}} & \cdots & K_{1\bar{m}} \\ \vdots & \cdots & \cdots & \cdots & \cdots \\ X_m & K_{m\bar{0}} & K_{m\bar{1}} & \cdots & K_{m\bar{m}} \end{vmatrix}}{\begin{vmatrix} K_{0\bar{0}} & K_{0\bar{1}} & \cdots & K_{0\bar{m}} \\ K_{0\bar{1}} & K_{1\bar{1}} & \cdots & K_{1\bar{m}} \\ \cdots & \cdots & \cdots & \cdots \\ K_{m\bar{0}} & K_{m\bar{1}} & \cdots & K_{m\bar{m}} \end{vmatrix}}. \tag{2.9}$$

The value of the minimum is:

$$\lambda_D^{X_0, X_1, \cdots, X_m}(t) = -\begin{vmatrix} 0 & \bar{X}_0 & \bar{X}_1 & \cdots & \bar{X}_m \\ X_0 & K_{0\bar{0}} & K_{0\bar{1}} & \cdots & K_{0\bar{m}} \\ X_1 & K_{1\bar{0}} & K_{1\bar{1}} & \cdots & K_{1\bar{m}} \\ \vdots & \cdots & \cdots & \cdots & \cdots \\ X_m & K_{m\bar{0}} & K_{m1} & \cdots & K_{m\bar{m}} \end{vmatrix} \Big/ J_m\,, \tag{2.10}$$

where J_m denotes the denominator which appears in (2.9).

Here

$$(2.11)\qquad K_{0\bar{j}}(z,\bar{t})=\left[\frac{\partial^j}{\partial\bar{\zeta}^j}(K_D(z,\bar{\zeta})\right]_{\bar{\zeta}=\bar{t}}\qquad (j=0,1,\cdots,m)\,;$$

$$(2.12)\qquad K_{i\bar{j}}=\left[\frac{\partial^{i+j}}{\partial z^i\partial\bar{\zeta}^j}(K_D(z,\bar{\zeta}))\right]_{z=t,\,\bar{\zeta}=\bar{t}}\,,\qquad (i,j=0,1,\cdots,m)\,.$$

($i=0$, *means that one should not differentiate with respect to* z. *Similarly, if* $j=0$.)

Proof. This theorem was proved by S. Bergman for univalent, domains without identified points.[4] ([4], pp. 26-27.) One can use the same proof, (which is based on the method of Lagrange multipliers), provided that one shows first that the minimum problem has a solution. This is done as follows: If D is a generalized domain, then it can be obtained from a univalent, domain, without identified points D^*, by a conformal mapping $z=z(z^*)$, $z^*\in D^*$. If t^* is the inverse image of t, then

$$(2.13)\qquad \left.\frac{dz}{dz^*}\right|_{z^*=t^*}\neq 0\,,\ \infty\,,$$

because t is not a branch point, and D^* is univalent. Therefore, the inverse function $z^*=z^*(z)$ is regular at t and $(dz^*/dz)\,|_{z=t}=c\neq 0$.

To each function $f(z)$ of the class $\mathscr{L}^2(D)$, corresponds one function $f^*(z^*)$ of the class $\mathscr{L}^2(D^*)$ such that

$$(2.14)\qquad f(z)=f^*(z^*(z))\cdot\frac{dz^*}{dz}\,,$$

hence,

$$(2.15)\qquad \iint_D|f|^2\,d\omega=\iint_{D^*}|f^*|^2\,d\omega_{z^*}\,.$$

Thus, there is a one-to-one mapping between the family of functions considered in the theorem and the family of functions $f^*(z^*)$ of the class $\mathscr{L}^2(D^*)$ which satisfy:

$$(2.16)\qquad f^{*(\nu)}(t^*)=Y_\nu\,,\qquad \nu=0,1,2,\cdots,m\,,$$

where Y_ν are complex numbers satisfying the system of equations:

$$(2.17)\qquad \left[\frac{d^\nu}{dz^\nu}\left[f^*(z^*(z))\frac{dz^*}{dz}\right]\right]_{z=t}=X_\nu\,,\qquad \nu=0,1,2,\cdots,m.$$

(See (2.8) and (2.14).) This system has one and only one solution because $c\neq 0$. Bergman's theorem ensures the existence of a unique func-

[4] There always exist functions satisfying (2.8), if the boundary has a positive capacity. (See also K. I. Virtanen [**11**].)

tion which minimizes the right integral of (2.15) under the conditions (2.16); hence it follows from (2.14), (2.15) and (2.16) that this function multiplied by dz^*/dz is the solution of the original problem, and that it is unique.

REMARK 1. Incidentally, we have proved that *if D and D^* are two (generalized) domains, and D^* is mapped conformally onto D by the function $z = z(z^*)$, $z^* \in D$, then, if $z^* = z^*(z)$ is the inverse mapping,*

$$M_D^{X_0, X_1, \cdots X_m}(z, t) = M_{D^*}^{Y_0, Y_1, \cdots, Y_m}(z^*(z), z^*(t)) \cdot \frac{dz^*}{dz} \tag{2.18}$$

$$\lambda_D^{X_0, X_1, \cdots, X_m}(t) = \lambda_{D^*}^{Y_0, Y_1, \cdots, Y_m}(z^*(t)) , \tag{2.19}$$

where Y_ν, $\nu = 0, 1, 2, \cdots, m$, satisfy (2.16), (2.17); *provided that t and $t^* = z^*(t)$ are not branch points of D and D^*, respectively.*

REMARK 2. It is possible to solve a similar extremal problem when t is a branch point, but the solution depends on the type of the branch point at t. This solution will not be considered in this paper.

REMARK 3. *If t is not a branch point, then the denominator of the right-hand side of* (2.9) *is finite and positive.*

Proof. Denote this denominator by J_m. It follows from (2.6) that $0 < J_0 = K_{0\bar{0}} < \infty$, provided t is not a branch point. Substituting $X_0 = X_1 = \cdots = X_{m-1} = 0$, $X_m = 1$ in (2.10), we obtain:

$$\lambda_D^{0,0,\cdots,0,1}(t) = \frac{J_{m-1}}{J_m} , \quad m \geqq 1 . \tag{2.20}$$

But, by definition, and because of Theorem 1, $0 < \lambda_D^{0,0,\cdots,0,1}(t) < \infty$, hence, by induction, $0 < J_m < \infty$.

3. Minimal domains with respect to almost identity mappings. In this section we shall be concerned with "almost identity" mapping-functions, i. e., functions of the class $\mathscr{E}_{m,t}(D)$ defined as follows:

DEFINITION. Let D be a domain containing a point t which is not a branch point. A function $f(z)$ is said to belong to the class $\mathscr{E}_{m,t}(D)$, if it satisfies the following conditions:

$$f'(z) \in \mathscr{L}^2(D) \quad \text{(see definition in Section 2)}, \tag{3.1}$$

$$f(t) = 0, f'(t) = 1, f''(t) = f^{(3)}(t) = \cdots = f^{(m)}(t) = 0 ; \qquad m \geqq 1. \tag{3.2}$$

This class has the following property:

LEMMA. *If $\zeta = f(z)$ belongs to the class $\mathscr{E}_{m,t}(D)$ and maps D onto a domain Δ, and if $w = \varphi(\zeta)$ belongs to the class $\mathscr{E}_{m,0}(\Delta)$, then $w = W(z) \equiv \varphi(\zeta(z))$ belongs to the class $\mathscr{E}_{m,t}(D)$.*

The proof is obvious.

The functions of the class $\mathscr{E}_{m,t}(D)$ map D onto various domains, among which we look for that domain which has the least area[5], and we try to determine the mapping function of D onto this domain. It follows from the Lemma that the domain having the least area is an "m-minimal domain" with center at the origin, in the following sense:

DEFINITION. A domain Δ is called an m-minimal domain having a center at a point τ, if $\tau \in \Delta$ and is not a branch point of Δ, and if any conformal mapping $w^* = w^*(w)$, $w \in \Delta$, which satisfies $w^*(w) \in \mathscr{E}_{m,\tau}(\Delta)$, maps Δ onto a domain whose area is not smaller than the area of Δ.

REMARK. It is clear that *a translation $w^* = w + a$, $w \in \Delta$, maps an m-minimal domain with center at the point τ onto an m-minimal domain with center at the point $\tau + a$.*

Denote by $M \equiv M_D(z, t)$, the function $M_D^{X_0, X_1, \cdots, X_{m-1}}(z, t)$ (sec (2.9)), for the special case:

$$X_0 = 1,\ X_1 = X_2 = \cdots = X_{m-1} = 0\ , \tag{3.3}$$

From Theorem 1 we obtain immediately:

THEOREM 2. *Let D be a (generalized) domain and t a point in it, which is not a branch point, then there exists a unique function $f(z)$ satisfying the condition*

$$f(z) \in \mathscr{E}_{m,t}(D)\ , \tag{3.4}$$

which maps D onto an m-minimal domain Δ with center at the origin. This mapping-function is given by

$$f(t) = 0\ ,\ f'(z) = M_D(z, t) \tag{3.5}$$

(see (3.3)).

COROLLARY 1. *If D itself is an m-minimal domain with center at the point t, then $f(z) = z - t$ is the mapping-function required in Theorem 2. Therefore, in this case, $M_D(z, t) = 1$. Hence, by* (2.9) *and* (3.3), *if D is an m-minimal domain with center at the point t, then*

---

[5] The area of a domain Δ is defined as $\iint_\Delta d\omega$ (see (2.3)). Thus, different coverings are counted separately and, when identified points exist, only the fundamental domain is counted.

(3.6) $$P_{m-1}(z, \bar{t}) = \text{constant} \neq 0 \ .$$

Here $P_{m-1}(z, \bar{t})$ *denotes the domain function which is formed by crossing out the first and last coloumns and the second and last rows in the determinant which appears in the numerator of* (2.9). $K = K_D$, *(see* (2.11), (2.12))[6].

COROLLARY 2. *The converse is also true. Indeed, if* (3.6) *is satisfied*[5], *then it follows from Theorem* 2 *that a mapping function of* D *onto an* m*-minimal domain* Δ *with center at the origin is obtained by a translation* $f(z) = z - t$; *therefore,* D *itself is an* m*-minimal domain with center at the point* t.

The area S_Δ of the m-minimal domain Δ can be calculated from (2.10). Indeed, by (2.10),

(3.7) $$S_\Delta = Q_{m-1}/J_{m-1} \ .$$

Here Q_{m-1} is the determinant which is formed by crossing out the first, second and last rows and the first, second and last columns in the numerator of (2.10). J_{m-1} was defined in Remark 3 of Section 2. (In the case $m = 1$, we define $Q_0 = 1$).

THEOREM 3. *A domain* D *containing a point* t, *which is not a branch point, is an* m*-minimal domain with center at the point* t, *if and only if the right-hand side of* (3.7) *is equal to the area* S_D *of* D.

Proof. By Theorem 2, D can be mapped onto an m-minimal domain Δ having an area given by (3.7). If $S_\Delta = S_D$, D itself is an m-minimal domain. If D is an m-minimal domain then $S_\Delta = S_D$.

REMARK. Observe that the right-hand side of (3.7) depends only on the kernel function and its derivatives at the single point t.

1-minimal domains were introduced by S. Bergman and their definition was later extended to domains in the space of n-complex variables. Some properties of 1-minimal domains were studied by S. Bergman [3] [4], by M. Schiffer [9] and by the present author [7]. We shall see that many properties of 1-minimal domains can be extended to properties of m-minimal domains, and that these new properties yield information about the behaviour of the kernel function as well as distortion theorems for certain classes of domains.

4. Simply-connected m-minimal domains. It is known that a simply-connected 1-minimal domain can only be a circle, the center of which is

[6] The constant on the right-hand side of (3.6) cannot be zero, because $M_D(t, t) = 1$ and the denominator of the right-hand side of (2.9) is finite and positive (see Remark 3 of Section 2).

the center in the sense of the definition of a minimal domain. We shall show, in this section, that the class of all simply-connected m-minimal domains can be obtained from a circle by mapping-functions which are polynomials of degree (at most) m. First, we consider the mapping of any domain D (not necessarily simply-connected) onto an m-minimal domain.

THEOREM 4. *Let D be an m-minimal domain ($m \geqq 1$), having a center at the origin. Let D^* be a domain, containing the origin and being locally univalent there, which can be mapped conformally onto D by a transformation $z = z(z^*)$, $z^* \in D^*$, which satisfies $z(0) = 0$; then*

$$
(4.1)\qquad \begin{vmatrix} 0 & K^*_{0\bar{0}}(z^*, 0) & K^*_{0\bar{1}}(z^*, 0) & \cdots & K^*_{0\,\overline{m-1}}(z^*, 0) \\ \left.\dfrac{dz}{dz^*}\right|_{z^*=0} & K^*_{0\bar{0}} & K^*_{0\bar{1}} & \cdots & K^*_{0\,\overline{m-1}} \\ \left.\dfrac{d^2z}{dz^{*2}}\right|_{z^*=0} & K^*_{1\bar{0}} & K^*_{1\bar{1}} & \cdots & K^*_{1\,\overline{m-1}} \\ \vdots & \cdots & \cdots & \cdots & \cdots \\ \left.\dfrac{d^mz}{dz^{*m}}\right|_{z^*=0} & K^*_{m-1\,\bar{0}} & K^*_{m-1\,\bar{1}} & \cdots & K^*_{m-1\,\overline{m-1}} \end{vmatrix} \cdot \frac{dz^*}{dz} = \text{constant} \neq 0,
$$

$z^* \in D^*$. Here $K^*(z^*, \bar{\zeta}^*) \equiv K_{D^*}(z^*, \bar{\zeta}^*)$; $dz/dz^* |_{z^*=0} \neq 0$; $t^* = 0$. (See (2.11)(2.12).) *Conversely: if (4.1) is satisfied, and $dz/dz^* \neq 0$, then $z = z(z^*)$ maps D^* onto an m-minimal domain.*

Proof. If D is an m-minimal domain with center at the origin, then $M_D(z, 0) = \text{constant}$. (See Section 3, Corollary 1.)[7] Choosing $f^*(z^*) = dz/dz^*$ and substituting it in (2.16), one observes that (2.17) is satisfied (since $X_0 = 1, X_1 = X_2 = \cdots = X_{m-1} = 0$). Therefore, by (2.18),

$$
(4.2)\qquad M_D(z, 0) = M_{D^*}^{Y_0, Y_1, \ldots, Y_{m-1}}(z^*(z), 0) \cdot \frac{dz^*}{dz},
$$

where

$$
Y_\nu = \left.\frac{d^{\nu+1}z}{dz^{*\nu+1}}\right|_{z^*=0}, \quad \nu = 0, 1, \cdots, m-1 .
$$

Thus the relation (4.1) has been established. The converse statement is obtained by reversing the order of these arguments, and by using Corollary 2 of Section 3.

If D is a simply-connected m-minimal domain, we can assume that D^* is the unit circle. Since the kernel function of the unit circle is

[7] For the definition of $M_D(z, 0)$, see (3.3).

$$K_{D^*}(z^*, \bar{\zeta}^*) = \frac{1}{\pi} \frac{1}{(1 - z^*\bar{\zeta}^*)^2}, \tag{4.3}$$

(see [4] p. 9), it follows that

$$K^*_{0\bar{j}}(z^*, 0) = \frac{1}{\pi}(j+1)!\, z^{*j}, \qquad j = 0, 1, 2, \cdots, \tag{4.4}$$

$$\pi K^*_{i\bar{j}} = \begin{cases} 0 & , i \neq j \\ (j+1)!\, j! & , i = j \end{cases} \qquad , i, j = 0, 1, 2, \cdots. \tag{4.5}$$

Let $z = z(z^*) = a_1 z^* + (a_2 z^{*2})/2! + (a_3 z^{*3})/3! + \cdots$ in the neighbourhood of the origin, $a_1 \neq 0$; then, substitution of (4.4) and (4.5) in (4.1) yields, after some trivial calculations,

$$a_1 + \frac{a_2}{1!} z^* + \frac{a_3}{2!} z^{*2} + \cdots + \frac{a_m}{(m-1)!} z^{*m-1} = C \cdot \frac{dz}{dz^*}. \tag{4.6}$$

Therefore, the constant C is equal to 1 and all the derivatives of $z(z^*)$ of an order greater than m vanish at the origin. Thus we have proved:

THEOREM 5. *Any polynomial of degree m having a non-zero derivative at the origin maps the circle about the origin onto an m-minimal domain whose center is the image of the center of the circle. And conversely, any simply-connected m-minimal domain can be obtained from a circle by a mapping whose function is a polynomial of degree (at most) m, the derivative of which is not zero at the origin.*

Theorem 5 suggests that perhaps all m-minimal multiply-connected domains are images of 1-minimal domains under polynomial mappings. This however, is not true in general, as we shall see later (see Section 7). Nevertheless, each p-connected 1-minimal domain generates a class of p-connected m-minimal domains conformally equivalent to it; this class has $m + 1$ complex degrees of freedom. Indeed, since any domain can be mapped onto a 1-minimal domain such that a non-branch fixed point corresponds to its center, we can assume that the domain D^* of Theorem 4 is a 1-minimal domain having a center at the origin. A necessary and sufficient condition for D^* to be such a domain is: $K_{D^*}(z^*, 0) = \text{constant} \neq 0$ (see (3.7); see also [7]); therefore, $K^*_{i\bar{0}} = K^*_{0\bar{i}} = 0$, $i = 1, 2, 3, \cdots$, and (4.1) reduces to:

$$\begin{vmatrix} -\left.\dfrac{dz}{dz^*}\right|_{z^*=0} & K^*_{0\bar{1}}(z^*, 0) & \cdots & K^*_{0\,\overline{m-1}}(z^*, 0) \\ \left.\dfrac{d^2z}{dz^{*2}}\right|_{z^*=0} & K^*_{1\bar{1}} & \cdots & K^*_{1\,\overline{m-1}} \\ \vdots & \cdots & \cdots & \cdots \\ \left.\dfrac{d^mz}{dz^{*m}}\right|_{z^*=0} & K^*_{m-1\,\bar{1}} & & K^*_{m-1\,\overline{m-1}} \end{vmatrix} = C \cdot \frac{dz}{dz^*}. \tag{4.7}$$

Here, $K^*(z, \bar{\zeta}^*) = K_{D^*}(z^*, \bar{\zeta}^*)$, $dz/dz^*|_{z^*=0} \neq 0$ and $C = \text{constant} \neq 0$ must be equal to minus the minor of the element $-(dz)/dz^*|_{z^*=0}$. Thus, the conformal mapping $z = z(z^*)$ maps the 1-minimal domain D^* with center at the origin onto an m-minimal domain, $m \geqq 1$ such that the centers correspond, if and only if $z = z(z^*)$ satisfies (4.7).

Let us choose arbitrary constants $-c_1, c_2, c_3, \cdots, c_m, c_1 \neq 0$, for the elements of the first column of the determinant in (4.7); then dz/dz^*, thus defined, will indeed satisfy $(d^kz)/dz^{*k}|_{z^*=0} = c_k$[8].
Since, moreover, a translation carries an m-minimal domain onto an m-minimal domain, we have arrived at

THEOREM 6. *Let D be a 1-minimal domain having a center at the origin; then, to each choice of $m+1$ constants, $c_0, c_1, c_2, c_3, c_m, c_1 \neq 0$, there exists one and only one function.*

(4.8)

$$z = c_0 + c_1 z^* + \frac{1}{2!} c_2 z^{*2} + \cdots + \frac{1}{m!} c_m z^{*m} + d_{m+1} z^{*m+1} + d_{m+2} z^{*m+2} + \cdots ,$$

which maps D^* onto an m-minimal domain with center at the point c_0. The mapping function is given by (4.7) where $(d^kz)/dz^{*k}|_{z^*=0}$ are replaced by c_k.

COROLLARY. *In general, d_j, $j = m+1, m+2, \cdots$, depend on the choice of $c_2, c_3 \cdots, c_m$, but they do not depend on the choice of c_0 and c_1. Indeed, in order to obtain d_j, one has to differentiate j times the left-hand side of (4.7) and to put $z^* = 0$. The resulting expression does not contain either c_0 or c_1.*

5. Doubly-connected 1-minimal domains. There is an unpublished result of *M.* Schiffer stating that univalent finitely-connected domains which do not possess identified points cannot be 1-minimal domains unless they are circles punctured at isolated points (the center of the circles is not punctured).

P. Kuffareff [6] studied the normalized conformal mapping of a ring onto a domain having a minimal area, restricting the mapping-function to be single-valued, and he found out that the minimal domain thus obtained lies on a double-sheeted Riemann surface. It seems natural to ask whether the use of a wider class of mapping-functions, i. e. integrals of functions of the class $\mathscr{L}^2$, yields different minimal domains. We shall show that this is indeed the case: a 1-minimal domain which is conformally equivalent to the ring always possesses identified points; hence, the mapping-function from the ring onto it is multi-valued; in

[8] This is shown by differentiating (4.7) k times and putting $z^* = 0$.

some cases the minimal domain is even a univalent domain (with identified points).

Let T be the ring $0 < r < |z| < 1$ in the z-plane. It is known (see [4], p. 10) that its kernel function is

$$K_T(z, \bar{t}) = \frac{1}{\pi z\bar{t}}\left[\wp\{\log(z\bar{t})\} + \frac{\eta_1}{\omega_1}\right]. \tag{5.1}$$

Here $\wp(v)$ is the Weierstrass elliptic function having the periods $2\omega_1 = -2\log r$, $2\omega_3 = 2\pi i$, $\eta_1 = \zeta(\omega_1)$, where $\zeta(v)$ is the corresponding Weierstrass zeta-function. Let t be a fixed real point in the ring T, then the mapping

$$u = \log z + \log t - 2\log r \tag{5.2}$$

will map the ring onto a domain S which is an infinite strip

$$\log t - \log r < \Re u < \log t - 2\log r\ , \tag{5.3}$$

in which the points $u \pm 2k\pi i$, $k = 0, 1, 2, \cdots$ are identified. Let $ABCD$ be the fundamental rectangle of S: $A \equiv \log t - \log r$, $B \equiv \log t - 2\log r$, A, B are real; $C \equiv B + 2\pi i$, $D \equiv A + 2\pi i$. It follows from (2.7) that the kernel function of S satisfies:

$$K_S(u, \bar{\tau}) = \frac{1}{\pi}\left[\wp(u) + \frac{\eta_1}{\omega_1}\right], \qquad \tau = 2\log t - 2\log r. \tag{5.4}$$

Our aim is to map T onto a 1-minimal domain in such a way that the point t will correspond to its center. (From symmetry considerations it follows that the generality of the mapping is not affected by the fact that t is required to be real). In order to achieve this, we first map T onto S and then map S onto a 1-minimal domain (τ corresponds to its center). This last mapping is produced by the function

$$w = -\pi\int_{\tau}^{u} K_S(v, \bar{\tau})dv + \zeta(\tau) - \frac{\eta_1}{\omega_1}\tau = \zeta(u) - \frac{\eta_1}{\omega_1}u \tag{5.5}$$

(see Theorem 2, (2.9); $m = 1$).
(All other 1-minimal domains whose centers correspond to t are obtained from this one by the mapping $W = c_0 + c_1(w - \zeta(\tau) + (\eta_1/\omega_1)\tau$, $c_1 \neq 0$; see Corollary in Section 4).

Let Δ be the minimal domain obtained from S by the mapping (5.5). It follows from the quasi-periodicity of the zeta function that

$$w(u \pm 2k\pi i) - w(u) = \pm k\frac{\pi i}{\log r}\ , \qquad k = 0, 1, 2, \cdots; \tag{5.6}$$

hence, *the points* $w \pm k\pi i/\log r$ *are identified in* Δ.

It remains to find the fundamental domain of Δ which is the image of the rectangle $ABCD$ under the mapping (5.5). For this purpose we first determine the image under the same mapping of the rectangle $EFGH$, where $E \equiv 0$, $F \equiv \omega_1 = -\log r$ $G = -\omega_2 = -\log r + \pi i$ $H \equiv \omega_3 = \pi i$ together with its reflections $FE'H'G$ $(E' = 2\omega_1, H' = 2\omega_1 + \pi i)$, $HGF''E''$ $(F'' = \omega_1 + 2\pi i, E'' = 2\pi i)$ and $GH'E'''F''$ $(E''' = 2\omega_1 + 2\pi i)$. (See Figure 1.)

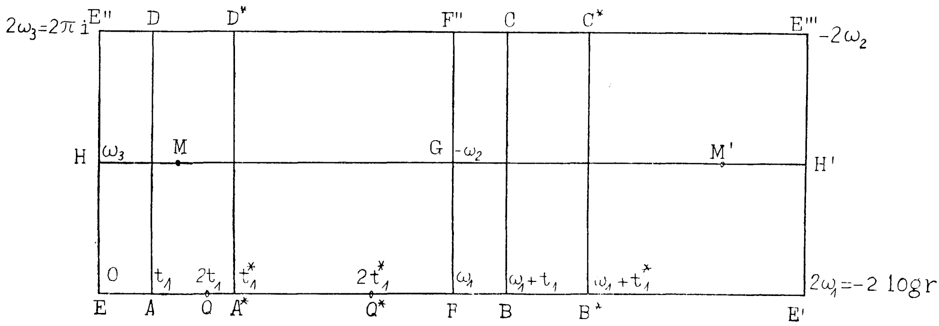


Figure 1. u – *plane* $t_1 = \log t - \log r$

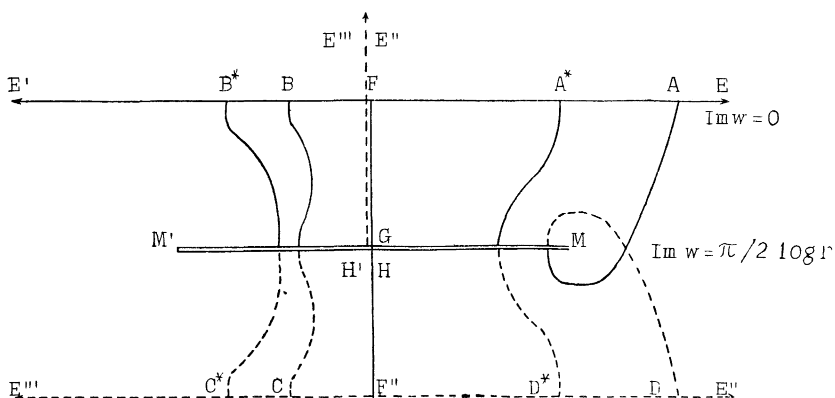


Figure 2. w–*plane*

The image can be deduced from [5] p. 190 : there are two and only two points, M on HG and its symmetric point M' on GH', where $-w'(u) = \wp(u) + \eta_1/\omega_1 = 0$. These points correspond to the two branch points of the first order in the image which are on the line $\Im w = (\pi)/2 \log r$. The image of the four rectangles lies on a two-sheeted Riemann surface as shown in Figure 2. (We denote points in the w-plane which correspond to points in the u-plane by the same letter. Lines on one sheet are traced by a dotted line.)

In order to obtain more information about the image of the rectangle $ABCD$, we shall prove :

LEMMA. *The distance $\overline{MH}$ in the u-plane is smaller than the distance $\overline{MG}$.*

Proof. $\wp(u) + \eta_1/\omega_1$ is real and a monotone function and it takes all the values from $-\infty$ to $+\infty$ as u traces the line $EHGFE$. (The function $\wp(u)$ has these properties and (see [10] p. 184)

$$\eta_1 = \frac{\pi^2}{2\omega_1}\left[1/6 + \sum_{n=1}^{\infty} 1/\sin^2 \pi \frac{n\omega_3}{\omega_1}\right]$$

is real because each term is real; so also is ω_1). This function vanishes at the point M on HG; therefore, it is negative at the point H. It remains to show that

$$\wp(\omega_3 + \tfrac{1}{2}\omega_1) + \frac{\eta_1}{\omega_1} > 0.$$

Indeed,

$$\wp(u) + \frac{\eta_1}{\omega_1} = \frac{\pi^2}{4\omega_1^2} \cdot \sum_{n=-\infty}^{\infty} 1/\sin^2\left(\pi \frac{u - 2n\omega_3}{2\omega_1}\right) \tag{5.7}$$

(see [10] p. 184); it is real for $u = \omega_3 + \frac{1}{2}\omega_1$, hence it is sufficient to consider the real part of each term of the series. It is easy to verify that, for $u = \omega_3 + \frac{1}{2}\omega_1$, the real part of each term of the right-hand side of (5.7) is positive; hence M lies between ω_3 and $\omega_3 + \frac{1}{2}\omega_1$.

From this lemma it follows that three possibilities can occur (taking various values for t):

(i) *The rectangle ABCD in the u-plane contains M but does not contain M'.*

(ii) *This rectangle, which we now denote by $A^*B^*C^*D^*$, contains neither M nor M'.*

(iii) *This rectangle, which we now denote by $A^{**}B^{**}C^{**}D^{**}$, contains M' but does not contain M.*

In the first and third case, the minimal domain Δ will contain one branch point and will thus lie on a two-sheeted Riemann surface (and it will, of course, possess identified points). In the second case, only one sheet is required for the minimal domain (which, however, still possesses identified points).

The figure shows only the fundamental domain (for the cases (i) and (ii)). The images of the lines AD, BC, A^*D^*, B^*C^* are not exact. The center of the minimal domain lies on the real axis of the w-plane and it is the image of the point $u = 2\log t - 2\log r$ (see 5.4).

6. m-representative domains. Attempts to generalize the Riemann mapping theorem to the case of domains in the n complex dimensional space lead to various other classes of canonical domains. A well known class is the class of the "representative domains, introduced by S. Bergman (see e. g. [1] [3]).

In this section we shall limit ourselves to the case of a plane domain and generalize the concept of the representative domains so that the mapping functions onto these new canonical domains will satisfy the relations (3.2).

DEFINITION. Let $M_D^*(z, t)$ be equal to $M_D^{X_0, X_1, \ldots X_m}(z, t)$ (see (2.9)), for

$$X_0 = 0,\ X_1 = 1,\ X_2 = X_3 = \cdots = X_m = 0 \tag{6.1}$$

Let $M_D(z, t)$ be defined again by (3.3). Here $t \in D$ and is not a branch point. The function

$$f(z) = \frac{M_D^*(z, t)}{M_D(z, t)}\ ;\ m \geqq 1 \tag{6.2}$$

satisfies the relation

$$f(t) = 0,\ f'(t) = 1,\ f^{\nu}(t) = 0\ ;\ 2 \leqq \nu \leqq m\ , \tag{6.3}$$

and maps D onto a domain Δ. The domain Δ will be called an *m-representative domain* with center at the origin. An m-representative domain with a different center is obtained by a translation.

These m-representative domains are indeed canonical domains in the sense of the following.

THEOREM 7. *If a domain D in the z-plane is mapped onto a domain D^* in the ζ-plane by a function $\zeta = \zeta(z)$ which satisfies*

$$\zeta(t) = 0,\ \zeta'(t) = 1,\ \zeta^{(\nu)}(t) = 0\ ;\ 2 \leqq \nu \leqq m\ , \tag{6.4}$$

and t is a non-branch point in D, then

$$\frac{M_D^*(z, t)}{M_D(z, t)} = \frac{M_{D^*}^*(\zeta(z), 0)}{M_{D^*}(\zeta(z), 0)}\ . \tag{6.5}$$

Thus D and D^* generate the same m-representative domain.

Proof. Replacing m by $m-1$ and z by ζ in (2.16) and (2.17), we see that

$$X_0 = 1,\ X_1 = X_2 = \cdots = X_{m-1} = 0 \tag{6.6}$$

$$Y_0 = 1,\ Y_1 = Y_2 = \cdots = Y_{m-1} = 0 \tag{6.7}$$

satisfy the equations (2.16) and (2.17); therefore

$$M_D(z, t) = M_{D^*}(\zeta(z), 0) \cdot \frac{d\zeta}{dz} . \tag{6.8}$$

(See Remark 1 in Section 2.) Similarly,

$$M_D^*(z, t) = M_{D^*}^*(\zeta(z), 0)\frac{d\zeta}{dz} . \tag{6.9}$$

The relation (6.5) now follows from (6.8) and (6.9).

In general, m-minimal domains are different from m-representative domains. (See Section 7.) It is, therefore, interesting to look for properties of domains which are simultanuously m-minimal and m-representative, with the same center.

THEOREM 8. *If D is an m-minimal domain and also an m-representative domain, with the same center t then*

$$M_D^*(z, t) = z - t ; \qquad z \in D . \tag{6.10}$$

Proof. On the one hand D is an m-minimal domain, therefore the mapping

$$w = f(z) = z - t \tag{6.11}$$

maps it onto an m-minimal domain Δ with the origin as center; hence it is implied from (3.5) that

$$M_D(z, t) = 1 , \qquad z \in D . \tag{6.12}$$

On the other hand, Δ is also an m-representative domain with a center at the origin, hence (6.2) and (6.12) imply

$$w = M_D^*(w, 0) ; \tag{6.13}$$

therefore, by (6.9) we obtain the relation (6.10).

By reversing the arguments of the proof we obtain immediately the converse theorem:

THEOREM 9. *If a domain D is an m-minimal domain with t as center, and its kernel function with its derivatives satisfy the relation* (6.10), *then D is also an m-representative domain with the same center.*

Proof. It follows from (6.12) (6.2), and (6.10) that $w = z - t$ maps D onto an m-representative domain with the origin as center, therefore, D itself is an m-representative domain with t as center.

Using the transformation formulas for $M_D(z, t)$ and $M_D^*(z, t)$, under conformal mappings, one can now obtain a differential equation for the kernel functions of the class of all domains which are obtained from the domain D of the previous theorem by a mapping satisfying (6.4).

THEOREM 10. *For each domain D^* which is conformally equivalent to the domain D of Theorem* 8, *and for which the mapping function $\zeta = \zeta(z)$ satisfies* (6.4), *there exists a differential equation for $K_{D^*}(\zeta, \bar{\tau})$. This equation can be put in the form*

$$\frac{d}{d\zeta}\left(\frac{M^*_{D^*}(\zeta, 0)}{M_{D^*}(\zeta, 0)}\right) = M_{D^*}(\zeta, 0) . \tag{6.14}$$

Proof. Formulas (6.5), (6.10), (6.12) imply

$$z - t = M^*_{D^*}(\zeta, 0)/M_{D^*}(\zeta, 0) . \tag{6.15}$$

(6.8) and (6.12) imply

$$\frac{dz}{d\zeta} = M_{D^*}(\zeta, 0) . \tag{6.16}$$

Equation (6.14) is obtained now by differentiating (6.15) with respect to ζ.

REMARK. For the case $m = 1$, one has

$$M^*_{D^*}(\zeta, 0) = -\frac{K^*_{0\bar{1}} K^*_{0\bar{1}}(\zeta, 0) - K^*_{0\bar{0}} K^*_{0\bar{1}}(\zeta, 0)}{K^*_{0\bar{0}} K^*_{1\bar{1}} - K^*_{0\bar{1}} K^*_{1\bar{0}}} , \tag{6.17}$$

$$M_{D^*}(\zeta, 0) = K^*_{0\bar{0}}(\zeta, 0)/K^*_{0\bar{0}} ; \tag{6.18}$$

where $K^* = K_{D^*}$. Inserting this in (6.14), one obtains, after some calculations the relation

$$\frac{1}{[K_{D^*}(\zeta, 0)]^3} \begin{vmatrix} K_{D^*}(\zeta, 0) & \dfrac{\partial K_{D^*}(\zeta, 0)}{\partial \zeta} \\ \dfrac{\partial K^*_{D}(\zeta, 0)}{\partial \bar{\tau}} & \dfrac{\partial^2 K_{D^*}(\zeta, 0)}{\partial \zeta \partial \bar{\tau}} \end{vmatrix} = \text{const.} \tag{6.19}$$

This relation and its generalization to the case of domains in the n complex dimensional space was proved in [7].

7. A counter example. It is interesting to note that Theorem 5 no longer holds, in general, if we replace the circle by a 1-minimal multiply-connected domain D. A counter-example is an obvious deduction from the following theorem.

THEOREM 11. *If a* 1-*minimal domain D, with the origin as center, is mapped onto a* 2-*minimal domain with the origin as center, and the mapping function is a polynomial*

$$w = a_1 z + \tfrac{1}{2} a_2 z^2, \quad a_1, a_2 \neq 0 \ , \tag{7.1}$$

then D is also a 1-*representative domain with the origin as center.*

Proof. D is a 1-minimal domain with the origin as center; therefore, $K(z, 0) \equiv K_D(z, 0) =$ constant, for $z \in D$. (See Corollary 1, Section 3.) This implies that $K_{1\bar{0}} = K_{0\bar{1}} = 0$. Hence, it follows from Theorem 4 that

$$-a_1 K_{0\bar{0}} K_{1\bar{1}} - a_2 K_{0\bar{1}}(z, 0) K_{0\bar{0}} = c \cdot (a_1 + a_2 z) \ . \tag{7.2}$$

Thus, the value of the constant c is $-K_{0\bar{0}} K_{1\bar{1}}$ and

$$K_{0\bar{1}}(z, 0) = K_{1\bar{1}} z \ . \tag{7.3}$$

The last relation is equivalent to the relation (6.10), for $m = 1, t = 0$, hence, by Theorem 9, D is also a 1-representative domain with the origin as center.

COROLLARY. *The relation* (6.19) *is a consequence of* (7.3) *for any domain which is conformally equivalent to the domain D of Theorem* 11. *As there are domains for which* (6.19) *does not hold*, e. g., *a ring, for which* (6.19) *can be proved incorrect by a direct calculation* (see [4] *p.* 10), *one arrives at the conclusion that not all minimal domains are also representative domains with the same center, and that Theorem* 5 *does not hold if one replaces the circle by a general* 1-*minimal domain.*

REFERENCES

1. S. Bergman, *Ueber die Existenz von Repräsentantenbereichen*, Math. Ann., **102** (1929), 430-446.

2. ———, *Sur les fonctions orthogonales de plusieurs variables complexes avec les applications a la théorie des functions analytiques*, Intersc. Pub., (1941) and Mém. des Sc. Math., **106**, Paris (1947).

3. ———, *Sur la fonction-noyan d'un domaine et ses applications dans la théorie des transformations pseudo-conformes*, Mém. des Sc. Math., **108**, Paris (1948).

4. ———, *The kernel function and conformal mapping*, Amer. Math. Soc., New-York (1950).

5. H. Kober, *Dictionary of conformal representations*, Dover Pub. (1952).

6. P. Kuffareff, *Ueber das Zweifachzusammenhängende Minimalgebiet*, Bull. Inst. Math. et Mec., Univ. de Tomsk, **1** (1935-1937), 228-236.

7. M. Maschler, *Minimal domains and their Bergman kernel function*, Pacific J. Math., **6** (1956), 501-516.

8. R. Nevanlinna, *Uniformisierung*, Springer-Verlag, Berlin, 1953

9. M. Schiffer, *Sur les domaines minima dans la théorie des transformations pseudo-conformes*, C. R. Acad. Sci., Paris, **207** (1938), 112-115.
10. J. Tannery, J. Molk, *Éléments de la théorie des functions elliptiques*, **1**, Paris (1893).
11. K. I. Virtanen, *Über Extremalfunktionen auf offenen Riemannschen Flächen*, Ann. Acad. Scient. Fenn. Ser. A. I. **141** (1952).

THE HEBREW UNIVERSITY, JERUSALEM

ON THE IMBEDDABILITY OF THE REAL PROJECTIVE SPACES IN EUCLIDEAN SPACE[1]

W. S. MASSEY

1. Introduction. Let P_n denote n-dimensional real projective space. This paper is concerned with the following question: What is the lowest dimensional Euclidean space in which P_n can be imbedded topologically or differentiably? Among previous results along this line, we may mention the following;

(a) If n is even, then P_n is a non-orientable manifold, and hence cannot be imbedded topologically in $(n+1)$-dimensional Euclidean space, R^{n+1}.

(b) For any integer $n > 1$, P_n cannot be imbedded topologically in R^{n+1}, because its mod 2 cohomology algebra, $H^*(P_n, Z_2)$, does not satisfy a certain condition given by R. Thom (see [6], Theorem V, 15).

(c) If $2^{k-1} \leqq n < 2^k$ then P_n cannot be imbedded topologically in Euclidean space of dimension $2^k - 1$. This result follows from knowledge of the Stiefel-Whitney classes of P_n (see Thom, loc. cit., Theorem III. 16 and E. Stiefel, [5]; also [4]).

In the present paper, we prove the following result: If $m = 2^k, k > 0$, then P_{3m-1} cannot be imbedded differentiably in R^{4m}. For example P_5 cannot be imbedded differentiably in R^8, nor can P_{11} be imbedded in R^{16}. Of course if $n > m$, P_n cannot, *a fortiori*, be imbedded differentiably in R^{4m}. Thus for many values of n our theorem is an improvement over previous results on this subject.[1]

The proof of this theorem depends on certain general results on the cohomology mod 2 of sphere bundles. These general results are formulated in § 2, and in § 3 the proof of the theorem is given. Finally in § 4 some open problems are discussed.

The author acknowledges with gratitude that he has benefited from several stimulating conversations with F. P. Peterson on this topic.

2. Steenrod squares in a sphere bundle with vanishing characteristic. Let $p: E \rightarrow B$ be a locally trivial fibre space (in the sense of

Received December 29, 1958. During the preparation of this paper, the author was partially supported by a grant from the National Science Foundation. An abstract announcing the main result of this paper was submitted to the American Mathematical Society in December, 1958; see the *Notices of the American Mathematical Society*, 6 (1959), 61.

[1] This result partially solves a problem proposed by S. S. Chern (see Ann. Math., **60** (1954), p. 222). It follows that P_n cannot be imbedded in R^{n+2} for $n > 3$ except possibly in case $n = 2^k - 1, k > 2$. The case $n = 2^k - 1$ is still open. The importance of this problem, and some of its implications, were emphasized by H. Hopf in his address at the International Congress of Mathematicians held in Cambridge, Massachusetts in 1950. This address is published in volume I of the Proceedings of the Congress (see pp. 193-202).

[6]) with fibre a sphere of dimension $k-1$. We will use the following notation (due to Thom) for the Gysin sequence with the integers mod 2, Z_2, as coefficients:

$$\cdots \longrightarrow H^{q-k}(B) \xrightarrow{\mu} H^q(B) \xrightarrow{p^*} H^q(E) \xrightarrow{\psi} H^{q-k+1}(B) \longrightarrow \cdots$$

Recall that the homomorphism μ is multiplication by the mod 2 characteristic class: $\mu(x) = x \cdot w_k$ for any $x \in H^{q-k}(B)$. In case $w_k = 0$, then $\mu = 0$, and the Gysin sequence splits up into pieces of length three as follows:

$$0 \longrightarrow H^q(B) \xrightarrow{p^*} H^q(E) \xrightarrow{\psi} H^{q-k+1}(B) \longrightarrow$$

Moreover, if we choose an element $a \in H^{k-1}(E)$ such that $\psi(a) = 1 \in H^0(B)$, then any element $x \in H^q(E)$ may be expressed uniquely in the form

$$x = p^*(u) + a \cdot p^*(v) \tag{1}$$

where $u \in H^q(B)$ and $v \in H^{q-k+1}(B)$ (the proof is the same as that given in § 8 of [3] except that here we are using Z_2 for coefficients). It is clear from this formula that the Steenrod squares and cup products in $H^*(E)$ are completely determined provided we know the Steenrod squares and products in $H^*(B)$, and provided we can express $Sq^i(a)$ for $1 \leqq i \leqq k-1$ in the form (1), i.e., provided the cohomology classes $\alpha_i \in H^{k-1+i}(B)$ and $\beta_i \in H^i(B)$ in the expression

$$Sq^i(a) = p^*(\alpha_i) + a \cdot p^*(\beta_i) \tag{2}$$

are known for $1 \leqq i \leqq k-1$. Of course the classes a and α_i are not uniquely determined. If a' is any other element of $H^{k-1}(E)$ such that $\psi(a') = 1$, then by exactness of the Gysin sequence there exists an unique element $b \in H^{k-1}(B)$ such that

$$a' = a + p^*(b) .$$

Corresponding to formula (2) there is an analogous formula

$$Sq^i(a') = p^*(\alpha_i') + a' \cdot p^*(\beta_i') . \tag{2'}$$

An easy computation shows that

$$\begin{cases} \alpha_i' = \alpha_i + b\beta_i + Sq^i(b) , \\ \beta_i' = \beta_i . \end{cases} \tag{3}$$

Thus β_i is unique; it is an invariant of the given fibre space. However, only the coset of α_i modulo the subgroup $\{\beta_i b + Sq^i b : b \in H^{k-1}(B)\}$ is unique. This coset is also an invariant of the given fibre space.

LEMMA. $\beta_i = w_i$, *the* ith *Stiefel-Whitney class* mod 2 *of the fibre space.*

This lemma is due essentially to Liao, [2]. See also an analogous proof in Massey, [3], § 9.

Thus β_i is identified with a standard invariant of sphere spaces. On the other hand, the coset of α_i does not seem to be related to any standard invariants. It may be thought of as a sort of "secondary characteristic class", defined for all sphere spaces for which the mod 2 characteristic class w_k vanishes.

In view of this lemma we may write the above equations in the following form:

$$(4) \qquad Sq^i(a) = p^*(\alpha_i) + a \cdot p^*(w_i)$$

$$(5) \qquad \alpha'_i = \alpha_i + bw_i + Sq^i(b)\ , \quad b \in H^{k-1}(B)\ .$$

3. Application to the problem of imbedding manifolds in Euclidean space. Our method of applying the results of the preceding section to prove that certain manifolds cannot be imbedded differentiably in r-dimensional Euclidean space is essentially the same as that used in our earlier paper [3]. To save the reader the trouble of referring to this earlier paper, we give a brief summary of the essential points of this method.

Let M^n be a compact connected differentiable manifold which is imbedded differentiably in an $(n + k)$-sphere, S^{n+k}. We assume that S^{n+k} has been given a Riemannian metric. Choose a positive number ε so small that given any point $a \in S^{n+k}$ of distance $\leqq \varepsilon$ from M^n, there exists a unique geodesic segment through a of length $\leqq \varepsilon$ normal to M^n. Let N denote the set of all points $a \in S^{n+k}$ whose distance from M^n is $< \varepsilon$. N is an open tubular neighborhood of M^n in S^{n+k}. Let E denote the boundary of N, and $p: E \to M^n$ the projection defined by assigning to each point $a \in E$ the point $p(a) \in M^n$ where the unique geodesic segment through a of length ε normal to M^n meets M^n. Then (E, p, M^n) is a realization of the normal $(k - 1)$-sphere bundle of M^n for the given imbedding, and E is a hypersurface in S^{n+k}. Let V denote the complement of N is S^{n+k}, and let $j: E \to V$ denote the inclusion map.

For convenience we introduce the following notation: A^q denotes the image of the homomorphism $j^*: H^q(V, G) \to H^q(E, G)$, where G is the coefficient ring, and $A^* = \sum_q A^q$. Then A^* and $p^*[H^*(M^n, G)]$ are both sub-rings of $H^*(E, G)$, and they are obviously closed under any cohomology operations such as Steenrod squares and reduced pth powers. Even more, A^* and $p^*[H^*(M^n, G)]$ are "permissible sub-algebras" of $H^*(E, G)$ in the terminology of Thom, [6], p. 177. The sub-ring A^* must satisfy the following conditions:

$$A^\circ = H^\circ(E, G) \tag{6}$$

$$H^q(E, G) = A^q + p^* H^q(M^n, G) \qquad (0 < q < n+1, \text{ direct sum}) \tag{7}$$

$$A^q = 0 \qquad \text{for } q \geqq n + k - 1 . \tag{8}$$

The proof that conditions (6), (7), and (8) hold is based on Theorem V.14 of Thom [6]; see also § 14 of [3]. The existence of the sub-algebra A^* satisfying these conditions is a rather stringent requirement on $H^*(E, G)$.

Our program for trying to prove that a certain manifold M^n cannot be imbedded differentiably in S^{n+k} (or equivalently, in Euclidean $(n+k)$-space, R^{n+k}) may be briefly outlined as follows. Assume that such an imbedding is possible, and let $p: E \to M^n$ denote the normal $(k-1)$-sphere bundle of this imbedding. By a well-known theorem of Seifert and Whitney, the characteristic class of the normal bundle vanishes, hence the Gysin sequence splits up into pieces of length three as described in the preceding section. Then if one can determine the structure of the cohomology ring of E and perhaps determine some other cohomology operations in E, it may be possible to prove that $H^*(E, G)$ does not admit any sub-ring A^* satisfying the conditions stated in the preceding paragraph. But this is a contradiction.

Using this method with $G = Z_2$, we will now prove our main result:

THEOREM. *If $m = 2^k$, $k > 0$, then $P_{3m-1}(R)$ cannot be imbedded differentiably in R^{4m}.*

Proof. Let x be the generator or $H^1(P_{3m-1}, Z_2)$. As is well known, the cohomology algebra $H^*(P_{3m-1}, Z_2)$ is the truncated polynomial algebra generated by x and subject to the sole relation $x^{3m} = 0$. According to a result of E. Stiefel [5], the total Stiefel-Whitney class $w = \sum_{i \geqq 0} w_i$ of the tangent bundle of P_{3m-1} is given by the formula

$$w = (1 + x)^{3m} .$$

(This may be proved directly by the method of Wu [7].) Using the Whitney duality theorem, one sees that the total Stiefel-Whitney class $\overline{w} = \sum_{i \geqq 0} \overline{w}_i$ of the normal bundle is given by

$$\overline{w} = (1 + x)^m$$

since

$$w\overline{w} = (1 + x)^{3m}(1 + x)^m = (1 + x)^{4m} = 1 + x^{4m} = 1$$

(Recall that $m = 2^k$). It follows that $\overline{w}_1 = 0$ and $\overline{w}_m = x^m$. Now assume that the differentiable imbedding of P_{3m-1} in S^{4m} is possible and let $p: E \to P_{3m-1}$ denote the normal bundle, whose fibre is an m-sphere.

The characteristic class $\overline{w}_{m+1} = 0$, hence we can apply the method of § 1. Choose $a \in H^m(E, Z_2)$ such that $\psi(a) = 1$. Then equations (4) of § 2 applied to this case give

$$Sq^1(a) = p^*\alpha_1$$
$$a^2 = Sq^m(a) = p^*\alpha_m + a \cdot p^*(x^m) .$$

If $a' \in H^m(E, Z_2)$, $\psi(a') = 1$, and $a' \neq a$, then of necessity $a' = a + p^*(b)$ with $b = x^m$. Therefore equation (5) of § 2 gives

$$\alpha'_1 = \alpha_1 + Sq^1(x^m) = \alpha_1$$
$$\alpha'_m = \alpha_m + x^m \cdot x^m + Sq^m x^m = \alpha_m .$$

Hence α_1 and α_m are invariants, independent of the choice of a. Since $H^m(E, Z_2)$ admits the direct sum decomposition

$$H^m(E, Z_2) = A^m + p^*[H^m(P_{3m-1}, Z_2)]$$

by (7), it follows that we may choose the cohomology class a so that it belongs to A^m. From now on we assume such a choice has been made. Note also that it follows directly from the Gysin sequence that for $q > 0$.

$$\text{rank } A^q = \text{rank}\, H^{q-m}(P_{3m-1}, Z_2)$$

where A^q and $H^{q-m}(P_{3m-1})$ are considered as vector spaces over the field Z_2. Thus rank $A^q = 0$ or 1 for all values of q; it follows that A^q has at most one non-zero element.

First, we consider the invariant α_1. Two cases are possible: $\alpha_1 = 0$, or $\alpha_1 = p^*(x^{m+1})$. If $\alpha_1 = p^*(x^{m+1})$, then the sub-ring A^* is not closed under the operation Sq^1, which is already a contradiction. For the remainder of the proof we will assume that $\alpha_1 = 0$, i.e. $Sq^1 a = 0$, and show that this also leads to a contradiction.

Next we consider the invariant α_m. Here again two cases are possible, $\alpha_m = 0$ or $\alpha_m = x^{2m}$. First let us consider the case where $\alpha_m = 0$, i.e.., $a^2 = a \cdot p^*(x^m)$. Since $a^2 \neq 0$, it must be the unique non-zero element of A^{2m}. Let y denote the unique non-zero element of A^{2m-1}. It is clear that either

$$y = p^*(x^{2m-1}) + a \cdot p^*(x^{m-1}) \text{ or } y = a \cdot p^*(x^{m-1}) .$$

Now $a^2 y \in A^{4m-1}$, therefore $a^2 \cdot y = 0$ by equation (8). An easy calculation shows that $a^2(a \cdot p^* x^{m-1}) = a \cdot p^*(x^{3m-1}) \neq 0$. It follows that $y = p^*(x^{2m-1}) + a \cdot p^*(x^{m-1})$. Next, a computation shows that

$$Sq^1(y) = Sq^1[p^* x^{2m-1} + a \cdot p^* x^{m-1}] = p^*(x^{2m}) + a \cdot p^*(x^m) .$$

Thus $Sq^1(y)$ and a^2 are distinct non-zero elements of A^{2m} which is obviously impossible. Thus we see that the assumption $\alpha_m = 0$ leads to a contradiction.

Next, consider the case where $\alpha_m = x^{2m}$, i.e. $a^2 = p^*(x^{2m}) + a \cdot p^*(x^m)$. The pattern of the proof in this case is the same as in the preceding paragraph. Let y be the unique non-zero element of A^{2m-1} as before, then either $y = p^*(x^{2m-1}) + a \cdot p^*(x^{m-1})$ or $y = a \cdot p^*(x^{m-1})$, and we must have $a^2 y = 0$ exactly as before. Once again an easy calculation shows that $a^2(p^* x^{2m-1} + a \cdot p^* x^{m-1}) \neq 0$, hence we must have $y = a \cdot p^* x^{m-1}$. Again, one finds that $Sq^1 y = a \cdot p^* x^m \neq a^2$, which is a contradiction. Thus we have shown that the assumption that P_{3m-1} can be imbedded differentiably in S^{4m} leads to a contradiction.

4. Some open problems. H. Hopf has proved in [1] that P_n can be imbedded differentiably in R^{2n-1} or R^{2n}, according as n is odd or even, i.e., according as P_n is orientable or not. This result, together with our main theorem and the previous results mentioned in the introduction, enables one to settle definitely the question of imbedding P_n in the lowest possible dimensional Euclidean space for $n \leqq 5$: for $n \leqq 5$, Hopf's result is the best possible. The first undecided case is P_6. It follows from Hopf's result that it can be imbedded in R^{12}, and from our result that it cannot be imbedded in R^8. Can it be imbedded in R^9?

The invariants α_i introduced in § 2 raise many interesting questions. Are these invariants of the normal bundle the same for any imbedding of a manifold in Euclidean space? Or, is it possible to give an example of different imbeddings of a manifold in Euclidean space which give rise to different sets of invariants α_i for the corresponding normal bundles?[2] In any case, it seems reasonable to hope that a further investigation of their properties may furnish new tools for proving non-imbeddability theorems for manifolds.

One may also carry out an analogous investigation using the integers mad p, Z_p, for any odd prime p as coefficients, and using Steenrod pth powers instead of Steenrod squares.

References

1. H. Hopf, *Systeme Symmetrischer Bilinearformen und Euklidische Modelle der Projektiven Räume*, Vierteljschr. Naturf. Ges. Zürich, **85** (1940), (32), 165–177.

2. S. D. Liao, *On the theory of obstructions of fiber bundles*, Ann. of Math., **60** (1946), 146–191.

3. W. S. Massey, *On the cohomology ring of a sphere bundle*, Jour. Math. and Mech., **7** (1958), 265–290.

4. F. P. Peterson, *Some non-embedding problems*, Bol. Soc. Mat. Mex., **2** (1957), 9–15.

[2] This question is related to the following: It is possible to give an example of an n-manifold M^n which admits two different imbeddings in R^{n+k} such that the normal bundles of these imbeddings are non-equivalent? As far as I know, the answer to this question is not known.

5. E. Stiefel, *Über Richtungsfelder in den Projektiven Räumen und einen Satz aus der Reelen Algebra,* Comm. Math. Helv., **13** (1940), 201-218.
6. R. Thom, *Espaces fibrés eu Spheres et Carrés de Steenrod,* Ann. Ecole Norm. Sup., **69** (1952), 109-182.
7. W. T. Wu, *Classes charactéristiques et i-carrés d'une variété,* C. Λ. Acad. Sci. Paris, **230** (1950), 508-511.

BROWN UNIVERSITY

SEMI-GROUPS OF CLASS (C_0) IN L_p DETERMINED BY PARABOLIC DIFFERENTIAL EQUATIONS

THOMAS W. MULLIKIN

1. Introduction. This paper treats mixed Cauchy problems for the parabolic partial differential equation in one space variable;

$$u = p(x)u_{xx} + q(x)u_x + r(x)u \ . \tag{1.1}$$

Our results are for non-singular equations, that is, the variable x is restricted to a finite interval $[a, b]$, and the function p is real-valued with $p(x) > 0$ on $[a, b]$. The functions q and r may be complex-valued. We require that p, q and r be in $L_\infty[a, b]$ and that p, p' and q be absolutely continuous with p', p'' and q' in $L_\infty\,[a, b]$.

We impose usual boundary conditions $\pi(u) = 0$ by

$$M_{i1}u(a) + N_{i1}u(b) + M_{i2}u'(a) + N_{i2}u'(b) = 0,\ i = 1, 2 \ . \tag{1.2}$$

The constants M_{ij}, N_{ij} are real or complex and the matrix $(M_{ij}; N_{ij})$ has rank two.

With Equation (1.1) is associated the ordinary differential operator

$$A = p(x)D^2 + q(x)D + r(x)I,\ D = \frac{d}{dx} \ . \tag{1.3}$$

With the above restrictions on the coefficients, A is defined in $L_p[a, b]$[1], $1 \leq p < \infty$, as a closed operator with dense domain, $D(A)$, given by

$$D(A) = \{u \in L_p \,|\, u \text{ and } u' \text{ are absolutely continuous and } u, u', u'' \in L_p\} \ . \tag{1.4}$$

The boundary conditions define restrictions A_π of A to subdomains,

$$D(A_\pi) = \{u \in L_p \,|\, u \text{ and } u' \text{ are absolutely continuous, } \pi[u] = 0, \text{ and } u, u', u'' \in L_p\} \ . \tag{1.5}$$

Our problem is to determine those A_π which generate *semi-groups of class* (C_0) in $L_p[a, b]$ (see Hille and Phillips [1], p. 320). Our main result is

Received September 24, 1958, and in revised form January 9, 1959. This paper presents some of the results of a Doctoral Dissertation at Harvard University. This research was supported in part by the Office of Naval Research under contract N 50 ri-07634 at Harvard University and by The RAND Corporation of Santa Monica, California.

[1] We denote by $L_p[a, b]$, $1 \leq p < \infty$ the complex Lebesgue space defined by Lebesgue measure on $[a, b]$. Any Lebesgue space defined by a different measure μ will be denoted by $([a, b], \mu)$.

THEOREM 4. *If π is regular*[2], *the operator A_π is the infinitesimal generator of a semi-group of class (C_0) in $L_p[a, b]$, $1 \leq p < \infty$.*

The theory of adjoint semi-groups (Hille and Phillips [10], p. 426) can be used to extend the results of Theorem 4 to the Banach space $L_\infty[a, b]$. However, these results apply only in proper closed subspaces of L_∞, and for brevity we do not include them.

In § 6 we investigate the necessity of regularity of π for the generation of a semi-group of class (C_0) by the special operators $\Omega_\pi = D^2$ in $L_p[0, 1]$. We have the partial result

THEOREM 5. *Let π and π^+ be adjoint boundary conditions relative to the operator D^2. If both Ω_π and Ω_{π^+} generate semi-groups of class (C_0) in any $L_p[0, 1]$, $1 < p < \infty$, then π and π^+ are regular.*

We also show that for certain non-regular π the operator $\Omega_\pi = D^2$ can be defined either in $L_1([0, 1], dx^2)$ or in $L_1([0, 1], d(1-x)^2)$ as the generator of a semi-group of class (C_0). These operators can be shown to be equivalent to singular operators in $L_1[0, 1]$.

We give, what seems to be, the first application of the Feller-Phillips-Miyadera Theorem (Hille and Phillips [10], p. 360); other applications have been of its corollary, the Hille-Yosida Theorem. Probably Theorem 2, where this theorem is applied, can also be proved by an appropriate use of spectral resolutions of the operators $\Omega_\pi = D^2$ in L_1 and L_2, however, we use spectral resolutions in only one instance. In any case, the eigenfunctions of the A_π can be used to give analytic representations of the semi-groups. In essence, we simply establish in L_p a certain type of behavior near $t = 0$ of solutions to the heat equation with a variety of boundary conditions.

Extensive application of semi-group theory to parabolic differential equations have been made by W. Feller ([4], [6], [7], [8]) and E. Hille [9]. Their papers contain our results for those real differential equation and real boundary conditions which determine positivity preserving semi-groups in L_1 and in L_2.

We plan in a later paper to present a study which we have made of the hyperbolic equation

$$u_{tt} + a(x)u_t = p(x)u_{xx} + q(x)u_x + r(x)u \;. \tag{1.6}$$

2. Equivalent semi-group.

We make considerable use of the following notions. If $\{T_t\}$ is a semi-group of class (C_0) on a Banach space U and if H is a linear homeomorphism of U onto another Banach space V, then it is easily shown that $\{S_t\}$ defined by

$$S_t = HT_tH^{-1} \tag{2.1}$$

[2] See G. D. Birkhoff [1], p. 383; J. D. Tamarkin [12]; or Coddington and Levinson [2], pp. 299-305.

is a semi-group of class (C_0) on V. We say that $\{T_t\}$ and $\{S_s\}$ are *homeomorphically equivalent.*

If ω is a constant and α a real positive constant, and if $\{T_t\}$ is a semi-group of class (C_0), then $\{S_t\}$ defined by

$$S_t = e^{\omega t} T_{\alpha t} \tag{2.2}$$

is a semi-group of class (C_0).[1]

We make the following

DEFINITION 1. Let $\{T_t\}$ and $\{S_t\}$ be semi-groups of class (C_0) defined respectively on Banach spaces U and V. Then $\{T_t\}$ and $\{S_t\}$ are said to be *equivalent* provided there exist constants ω and α, α real and $\alpha < 0$, such that $\{T_t\}$ and $e^{\omega t}S_{\alpha t}$ are homeomorphically equivalent.

For our applications we need the following theorem, which is easily verified.[2]

THEOREM 1. *Let $\{T_t\}$ and $\{S_t\}$ be equivalent semi-groups of class (C_0) defined respectively in Banach spaces U and V,* i.e.

$$S_t = H(e^{\omega t} T_{\alpha t}) H^{-1} . \tag{2.3}$$

The infinitesimal generators A and B are related by

$$B = (\omega I + \alpha H A H^{-1}) , \quad D(B) = HD(A) . \tag{2.4}$$

The resolvents of A and B are related by

$$R(\lambda ; B) = HR(\lambda - \omega ; \alpha A) H^{-1} . \tag{2.5}$$

We make now

DEFINITION 2. Let A and B be closed operators defined respectively in Banach spaces U and V with dense domains. Then A and B are said to be *equivalent* provided there exists a linear homeomorphism H of U onto V such that (i) $D(B) = HD(A)$ and (ii) $B = (\omega I + \alpha H A H^{-1})$ for some constants ω and α, α real and $\alpha > 0$.

3. Boundary conditions. The linear forms in (1.2) define a two dimensional sub-space of a four dimensional complex vector space. It is convenient for our discussion to specify such subspaces by Grassman coordinates, which are defined by

[1, 2] See Hille and Phillips [**10**], Theorem 12.2.2 and Theorem 13.6.1.

$$A = \begin{vmatrix} M_{11} & N_{11} \\ M_{21} & N_{21} \end{vmatrix} B = \begin{vmatrix} N_{11} & M_{12} \\ N_{21} & M_{22} \end{vmatrix} C = \begin{vmatrix} N_{11} & N_{12} \\ N_{21} & N_{22} \end{vmatrix}$$

(3.1)

$$D = \begin{vmatrix} M_{11} & N_{12} \\ M_{21} & N_{22} \end{vmatrix} E = \begin{vmatrix} M_{12} & N_{12} \\ M_{22} & N_{22} \end{vmatrix} F = \begin{vmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{vmatrix}$$

These coordinates satisfy the quadratic relationship

(3.2) $$FC - BD = AE ,$$

and they are unique to within a constant of proportionality. Also, any constants, not all zero, which satisfy (3.2) define by (3.1) a set of conditions π of rank 2 (Hodge and Pedoe [11], p. 312).

We now define, for brevity in the sequel, four types of boundary conditions by the following sets:

(3.3)
$$\begin{aligned} \tau_1 &= \{\pi \mid E = B + D = 0\} \\ \tau_2 &= \{\pi \mid E \neq 0, \text{ or } E = 0 \text{ and } B + D \neq 0, \text{ or } A \neq 0 \text{ and} \\ &\qquad B = C = D = E = F = 0\} , \\ \tau_3 &= \{\pi \mid F = C = 0 \text{ and one and only one of } A, B, D, E \neq 0\} , \\ \tau_4 &= \{\pi \mid A = E = 0, B = D = 1 \text{ and } FC = 1\} . \end{aligned}$$

Sets τ_1 and τ_2 have only the absorbing boundary conditions in common, i.e. $u(a) = u(b) = 0$. Sets τ_3 and τ_4 are disjoint subsets of τ_2. The set τ_3 contains only separated endpoint boundary conditions. Representatives of these types in the form of (1.2) are easily determined by imposing the defining conditions in (3.1).

It is a simple matter to check that all boundary conditions in the set τ_2 are regular in the sense of G. D. Birkhoff. With one exception, $u(a) = u(b) = 0$, all π in the set τ_1 are non-regular.

4. $\Omega_\pi = D^2$ in $L_1[0, 1]$ and $L_2[0, 1]$. For the special operator $\Omega_\pi = D^2$ on $[0, 1]$ we need

LEMMA 1. *Ω_π in $L_p[0, 1], 1 \leq p < \infty$, is a closed operator with dense domain. Except for those non-regular π given by*

(4.1)
$$\begin{aligned} &\alpha u(0) + u(1) = 0 \\ &\alpha u'(0) - u'(1) = 0 , \quad \alpha^2 = 1 , \end{aligned}$$

the resolvent $R(\lambda; \Omega_\pi)$ exists for all $\lambda, \Re(\lambda) > \omega_0 \geq 0$ for some ω_0, and $R(\lambda; \Omega_\pi)$ is expressed in all $L_p, 1 \leq p < \infty$, by a Green's function as

(4.2) $$R(\lambda; \Omega_\pi)[u](.) = \int_0^1 G(., t, \lambda)u(t)dt .$$

The proof of Lemma 1 is easy and is omitted.[3] We, however, shall refer to the explicit expression for $G(x, t, \lambda)$ which is

$$G(x,t,\lambda) =$$

$$(4.3)\quad \frac{1}{\Delta}\begin{cases} -F\sqrt{\lambda}\,sh\sqrt{\lambda}(x-t) + sh\sqrt{\lambda}t[Ash\sqrt{\lambda}\,(1-x) \\ \qquad + D\sqrt{\lambda}\,ch\sqrt{\lambda}(1-x)] + ch\sqrt{\lambda}t[B\sqrt{\lambda}\,sh\sqrt{\lambda}(1-x) \\ \qquad - E\lambda ch\sqrt{\lambda}\,(1-x)]\,, \\ \text{for } t \le x, \text{ and} \\ -C\sqrt{\lambda}\,sh\sqrt{\lambda}(t-x) + (\text{above with } x \text{ and } t \text{ interchanged}), \\ \text{for } t \ge x\,. \end{cases}$$

The function $\Delta(\lambda)$ is given in terms of (3.1) by

$$(4.4)\quad \Delta(\lambda) = (F+C)\lambda + A\sqrt{\lambda}\,sh\sqrt{\lambda} + (B+D)\lambda ch\sqrt{\lambda} - E\lambda^{3/2}sh\sqrt{\lambda}$$

where the principle value of $\sqrt{\lambda}$ is chosen for $\Re(\lambda) \ge 0$.

In § 5 it will be shown that our main result, Theorem 4, follows easily from the rather difficult

THEOREM 2. *If π is regular, then $\Omega_\pi = D^2$ generates a semi-group of class (C_0) in $L_1[0, 1]$ and in $L_2[0, 1]$.*

We prove Theorem 2 by a series of lemmas. Our method of proof amounts to proving this theorem for the subsets τ_3 and τ_4 of the set τ_2 of regular π. These results are then used to define a factorization of $R(\lambda; \Omega_\pi)$ for any regular π by which we reduce estimates on $\|[R(\lambda; \Omega_\pi)]^n\|$, $n = 1, 2, \cdots$, which are needed for an application of the Feller-Phillips-Miyadera Theorem, to estimates on certain functions of the complex parameter λ.

The necessity for estimating $\|[R(\lambda; \Omega_\pi)]^n\|$ for $n > 1$ results when Ω_π generates a semi-group $\{T_t\}$ for which $\|T_t\|$ is not bounded by $e^{\omega t}$ for any ω. Whether or not $\|T_t\| \le e^{\omega t}$ for a semi-group of class (C_0) in a Banach space.[4] In one instance, part (b) of Lemma 3, we are able to guess an equivalent norm for $L_1[0, 1]$ so that the Hille-Yosida Theorem can be applied, whereas in the L_1 norm this does not seem to be the case.

We have the easy

LEMMA 2. *For π in the set τ_3, Ω_π generates a semi-group of class (C_0) both in $L_1[0, 1]$ and in $L_2[0, 1]$.*

[3] See Coddington and Levinson [**2**], pp. 300-305.

[4] See Feller [**5**] where it is shown that if $\{T_t\}$ is a semi-group of class (C_0) in a Banach space, then an equivalent norm can always be defined *by the semi-group* so that in this norm $\|T_t\| < e^{\omega t}$.

Proof. (a) For $L_2[0, 1]$ all such Ω_π are self-adjoint with negative spectrum and a set of eigenfunctions which are a basis for $L_2[0, 1]$. It follows easily that such Ω_π generate semi-groups of contracting operators in $L_2[0, 1]$. (b) In $L_1[0, 1]$ we have by Fubini's Theorem, since $G(x, t, \lambda)$ is continuous,

$$\|R(\lambda; \Omega_\pi)u\| \leq \int_0^1 \int_0^1 |G(x, t, \lambda)| |u(t)| dt dx$$
$$\leq \|u\|_1 \max_{0 \leq t \leq 1} \int_0^1 |G(x, t, \lambda)| dx .$$

From (4.3) for these special π one gets easily

$$\|R(\lambda; \Omega_\pi)\| \leq \frac{1}{\lambda} . \tag{4.6}$$

By the Hille-Yosida Theorem, Ω_π generates a semi-group of contracting operators. This completes the proof.

The proof is not so easy for

LEMMA 3. *For π in the set τ_4, Ω_π generates in $L_1[0, 1]$ and in $L_2[0, 1]$ a semi-group of class (C_0).*

Proof. Any π in the set τ_4 is given by

$$\begin{aligned} au(0) + u(1) &= 0 \\ au'(0) + u'(1) &= 0 \end{aligned} \qquad a \neq 0 . \tag{4.7}$$

We note that if the complex constant a in (4.7) is such that $|a| = 1$, then the conditions π are self-adjoint relative to the operator D^2.

(a) We set $\sigma = \log |a|$ and define a linear homeomorphism H of $L_2[0, 1]$ onto $L_2[0, 1]$ by

$$H[u](x) = e^{-\sigma x} u(x) . \tag{4.8}$$

The operator $\tilde{\Omega}_\pi$ equivalent to Ω_π is

$$\tilde{\Omega}_\pi = D^2 + 2\sigma D + \sigma^2 I . \tag{4.9}$$

Now $\tilde{\Omega}_\pi$ is a perturbation by the unbounded operator

$$B = 2\sigma D + \sigma^2 I \tag{4.10}$$

of the operator $\Omega_{\tilde{\pi}}$, where $\tilde{\pi}$ is given by

$$\begin{aligned} \alpha u(0) + u(1) &= 0 \\ \alpha u'(0) + u'(1) &= 0 , \quad \alpha = \frac{a}{|a|} = e^{i\theta} . \end{aligned} \tag{4.11}$$

The domain $D(B)$ of B is the same as $D(\tilde{\Omega}_\pi) = D(\Omega_{\tilde{\pi}})$.

Now $\Omega_{\tilde{\pi}}$ is self-adjoint in $L_2[0, 1]$ with eigenvalues $\lambda_n = -(\theta + (2n+1)\pi)^2$, $n=0, \pm 1, \cdots$, and eigenfunctions $\phi_n(x) = \exp[i(\theta+(2n+1)\pi)x]$, which are a basis for $L_2[0, 1]$. Then $\Omega_{\tilde{\pi}}$ generates a contraction semi-group given by

$$T_t[u] = \sum_{n=-\infty}^{\infty} a_n e^{\lambda_n t}\phi_n(x),\ a_n = (u, \phi_n)\ . \tag{4.12}$$

We want to establish that B is in the perturbing class $\mathfrak{P}(\Omega_{\tilde{\pi}})$ of $\Omega_{\tilde{\pi}}$ (Hille and Phillips [10], p. 394). Since $D(B) = D(\Omega_{\tilde{\pi}})$ we must establish that

(4.13)
(i) $BR(\lambda; \Omega_{\tilde{\pi}})$ is bounded for some λ,
(ii) BT_t on $D(\Omega_{\tilde{\pi}})$ is bounded for all $t > 0$, and therefore extensible to $\overline{BT_t}$ on $L_2[0, 1]$, and
(iii) $\int_0^1 \|\overline{BT_t}\| dt < \infty$.

Now (i) of (4.13) follows immediately from (4.2). For (ii) of (4.13) we compute for $u \in D(\Omega_{\tilde{\pi}})$,

$$\begin{aligned} \frac{1}{2}\|BT_t(u)\|_2^2 &\le 4\sigma^2(DT_t(u), DT_t(u)) + \sigma^4\|T_t(u)\|_2^2 \\ &= 4\sigma^2 T_t(u)DT_t(u)\big|_0^1 - 4\sigma^2(T_t(u), D^2T_t(u)) \\ &\qquad + \sigma^4\|T_t(u)\|_2^2\ . \end{aligned} \tag{4.14}$$

Using the facts that $\pi(T_t(u)) = 0$, $\|T_t(u)\|_2 \le \|u\|_2$, and $\lambda_n \le 0$, we get

$$\frac{1}{2}\|BT_t(u)\|_2^2 \le \sigma^4\|u\|_2^2 + 4\sigma^2\|u\|_2^2\{\max_{-\infty\le n\le\infty} - \lambda_n e^{2\lambda_n t}\}\ . \tag{4.15}$$

Therefore, since $\lambda e^{-\lambda t}$ has on $[0, \infty)$ the maximum $1/2et$,

$$\|BT_t(u)\|_2 \le 2\sigma\Big(\sigma^2 + \frac{2}{et}\Big)^{1/2}\|u\|_2\ . \tag{4.16}$$

This proves (ii) in (4.13) as well as (iii)

Since $B \in \mathfrak{P}(\Omega_{\tilde{\pi}})$, the operator $\tilde{\Omega}_\pi$ generates a semi-group of class (C_0) (Hille and Phillips [10], p. 400). Since $\tilde{\Omega}_\pi$ is equivalent to Ω_π, this proves our lemma for $L_2(0, 1)$.

(b) In $L_1[0, 1]$ we do not use a perturbation argument as in $L_2[0, 1]$ because of the difficulty in proving (ii) of (4.13) without using orthogonality relations.

Again let $\sigma = \log|a|$ and introduce in $L_1[0, 1]$ an equivalent norm by

$$\|f\|_0 = \int_0^1 |f(x)|e^{-\sigma x}dx\ . \tag{4.17}$$

The identity mapping of $L_1[0,1]$ under these two norms is a linear homeomorphism and Ω_π is equivalent to itself.

We get by Fubini's Theorem

$$(4.18) \qquad ||R(\lambda;\Omega_\pi)u||_0 \le \int_0^1 |u(t)| \int_0^1 |G(x,t,\lambda)| e^{-\sigma x} dx dt .$$

The Grassman coordinates for (4.7) are $A = E = 0, B = D = a, C = 1$, and $F = a^2$, and from (4.3) for real $\lambda, \lambda > \sigma^2 (\sigma = \log|a|)$,

$$(4.19) \quad |G(x,t,\lambda)| \le \frac{\begin{cases} |a|^2 sh\sqrt{\lambda}(x-t) + |a| sh\sqrt{\lambda}(1+t-x), & t \le x \\ sh\sqrt{\lambda}(t-x) + |a| sh\sqrt{\lambda}(1+x-t), & t \ge x \end{cases}}{\lambda(-1 - |a|^2 + 2|a| ch\sqrt{\lambda})}$$

We recognize the right-hand side of (4.19) as the Green's function, $G_1(x,t,\lambda)$, for d^2/dx^2 and the real boundary conditions π_1 given by

$$(4.20) \qquad \begin{aligned} -|a|u(0) + u(1) &= 0 \\ -|a|u'(0) + u'(1) &= 0 \end{aligned}$$

for which $A = E = 0, B = D = |a|, C = -1$, and $F = -|a|^2$.

Now the function $e^{-\sigma x}$ is an eigenfunction of the operator $\Omega_{\pi_1^+}$ for the eigenvalue σ^2, where π_1^+ is the adjoint of π_1, which is represented by (4.20) if $|a|$ is replaced by $|a|^{-1}$. Since these are real boundary conditions, $G_1(x,t,\lambda)$, for real λ, defines the Green's function for $\Omega_{\pi_1^+}$ if integration is done with respect to the variable x. Therefore for (4.18) we have with λ real

$$(4.21) \qquad \begin{aligned} ||R(\lambda;\Omega_\pi)u||_0 &\le \int_0^1 \frac{|u(t)|e^{-\sigma t}}{\lambda - \sigma^2} dt \\ &\le \frac{||u||_0}{\lambda - \sigma^2}, \ \lambda > \sigma^2 . \end{aligned}$$

This proves that Ω_π generates a semi-group of class (C_0) in L_1 normed by $||u||_0$, and therefore in L_1 with the usual norm. This completes the proof of our lemma.

The extension to all π in the set τ_2 is based on

LEMMA 4. *Let π be in the set τ_2. Then*

$$(4.22) \qquad R(\lambda;\Omega_\pi) = \sum_{i=1}^{6} f_i(\lambda) R(\lambda;\Omega_{\pi_i}) ,$$

where π_1 and π_2 are in the set τ_4 and $\pi_3, \cdots, \pi_6$ are in the set τ_3. The functions $f_i(\lambda)$ are given by

(42.3) $$f_i(\lambda) = \alpha_i \frac{\Delta_i(\lambda)}{\Delta(\lambda)}, \qquad i = 1, 2, \cdots, 6,$$

where the α_i are constants and $\Delta(\lambda)$ for π and $\Delta_i(\lambda)$ for π_i are defined by (4.4).

Proof. We use the Grassmann coordinates to define the π_i as follows. By adding and subtracting constants we write π as $\sum_{i=1}^{6} \alpha_i \pi_i$ where

(4.24)
$$\begin{aligned}
\pi&: \ (A, B, C, D, E, F),\\
\pi_1&; \ (0, 1, C - X, 1, 0, F - \overline{X}),\\
\pi_2&: \ \left(0, 1, \frac{X}{|X|}, 1, 0, \frac{\overline{X}}{|X|}\right),\\
\pi_3&: \ (1, 0, 0, 0, 0, 0),\\
\pi_4&: \ (0, 1, 0, 0, 0, 0),\\
\pi_5&: \ (0, 0, 0, 1, 0, 0),\\
\pi_6&: \ (0, 0, 0, 0, 1, 0),
\end{aligned}$$

$\alpha_1 = 1, \alpha_2 = |X|, \alpha_3 = A, \alpha_4 = B - 1 - |X|, \alpha_5 = D - 1 - |X|$ and $\alpha_6 = E$. Now X has to be chosen so that the coordinates of π_1 satisfy (3.2). X is given by

(4.25)
$$X = C - \rho e^{i\theta}, \ \theta = \arg(C - \overline{F}) \text{ and}$$
$$\rho = \frac{|C - \overline{F}| + \sqrt{|C - \overline{F}|^2 + 2}}{2}$$

Using the linearity of the numerator of the Green's function (4.3) in the constants A, B, C, D, E, and F, we get the expression (4.23).

We shall apply to the functions $f_i(\lambda)$ of Lemma 6 the following:

THEOREM 3. *Let $f(\lambda)$ be analytic in a half plane $\Re(\lambda) > \alpha$. Let $f(\lambda)$ satisfy either of the following conditions:*

(i) *$f(\lambda)$ is real for real λ and $(-1)^k f^{(k)}(\lambda) \geq 0$ (or ≤ 0) for all real $\lambda, \lambda > \alpha, k = 0, 1, \cdots$, i.e., f is completely monotonic in $(\alpha, +\infty)$.*

(ii) (a) *$\int_{-\infty}^{\infty} |f(\sigma + i\tau)| d\tau < M < +\infty, \sigma > \alpha$, M independent of σ.*

(b) *$\lim_{|\tau| \to \infty} f(\sigma + i\tau) = 0$ uniformly in every closed subinterval of $\alpha < \sigma < +\infty$.*

Then there exist real numbers $K > 0$ and ω such that

(4.26) $$\sum_{k=0}^{n} \frac{|f^{(k)}(\lambda)|}{k!} (\lambda - \omega)^k < K, \text{ for } n = 0, 1, \cdots,$$

and λ real, $\lambda > \omega$.

Proof. Suppose that (i) holds and that $f(\lambda) > 0$ for real λ (otherwise replace f by $-f$). Then $|f_i^{(k)}(\lambda)| = (-1)^k f_i^k(\lambda)$ and with $\omega = \alpha + 1$

$$\sum_{k=0}^{\infty} \frac{|f_i^{(k)}(\lambda)|}{k!}(\lambda - \omega)^k = \sum_{k=0}^{\infty} \frac{f_i^k(\lambda)}{k!}(\omega - \lambda)^k = f(\omega),\ \lambda \geq \omega\ , \tag{4.27}$$

since f is analytic in the region $\Re(\lambda) > \alpha$. Then (4.26) follows with $K = |f(\alpha + 1)|$ and $\omega = \alpha + 1$.

Suppose that condition (ii) holds. Then f is the Laplace transform (Widder [13], p. 265) of a function $\phi(t)$ for which $\phi(t) = 0, t < 0$ and $|\phi(t)| \leq Me^{\sigma - t}, \sigma > \alpha$. We have (Widder [13], p. 57)

$$f^{(k)}(\lambda) = \int_0^{\infty} (-t)^k e^{-\lambda t}\phi(t)dt \qquad \Re(\lambda) > \alpha\ . \tag{4.28}$$

So with $\omega = \alpha + 2$ and real $\lambda, \lambda > \omega$,

$$\sum_{k=0}^{n} \frac{|f^{(k)}(\lambda)|}{k!}(\lambda - \omega)^k \leq \int_0^{\infty} e^{-\omega t}|\phi(t)|dt \leq M\ . \tag{4.39}$$

Therefore (4.26) follows with $K = M$ and $\omega = \alpha + 2$.

We finally come to

Proof of Theorem 2. We shall establish the existence of real constants M and $\omega > 0$ such that in both L_1 and L_2 for real λ

$$\|[R(\lambda; \Omega_\pi)]^{n+1}\| \leq \frac{M}{(\lambda - \omega)^n}\ , \quad \lambda > \omega, n = 1, 2, \cdots\ . \tag{4.30}$$

By the Feller-Phillips-Miyadera Theorem this will prove our theorem.

In the representation (4.22) for $R(\lambda; \Omega_\pi)$, each Ω_{π_i} generates a semigroup of class (C_0) in L_1 and in L_2, either by Lemma 2 or by Lemma 3. Then for each $R(\lambda; \Omega_{\pi_i})$, $i = 1, 2, \cdots, 6$ (4.30) holds in L_p, $p = 1, 2,$ and M and $\omega > 0$ can be chosen independently of i and p.

Iterates of a resolvent can be computed by

$$[R(\lambda; \Omega_\pi)]^{n+1} = \frac{(-1)^n}{n!}\frac{\partial^n}{\partial \lambda^n}R(\lambda; \Omega_\pi)\ , \tag{4.31}$$

(Hille and Phillips [10], p. 184). Making use of (4.22), (4.31) and (4.30) for each $R(\lambda; \Omega_{\pi_i})$, we get

$$\|[R(\lambda; \Omega_\pi)]^{n+1}\| \leq \frac{M}{(\lambda - \omega)^{n+1}} \sum_{i=1}^{6} \sum_{k=0}^{n} \frac{|f_i^{(k)}(\lambda)|}{k!}(\lambda - \omega)^k\ , \tag{4.32}$$

real $\lambda, \lambda > \omega$, and $n = 0, 1, \cdots$.

We suppose now that π is such that either $E \neq 0$ or $B + D \neq 0$.

The only other regular π is in the set τ_3, and has been dealt with in Lemma 2. With this assumption, each of the functions $f_i(\lambda)$ of Lemma 4 can be written as

$$f_i(\lambda) = J_i + \frac{K_i}{\sqrt{\lambda}} + \frac{L_i}{\lambda} + F_i(\lambda) , \qquad i = 1, 2, \cdots, 6 \tag{4.33}$$

for uniquely determined constants and a unique analytic function $F_i(\lambda)$.

For $\Re(\lambda) > 0$ we have chosen a branch of $\lambda^{1/2}$, so that the first three functions in (4.33) are analytic and satisfy condition (i) of Theorem 3. The functions $F_i(\lambda)$ are analytic and can be shown to satisfy conditions (ii) of Theorem 3. Then (4.32) and (4.33) together with Theorem 3 give our desired result (4.30). This proves our theorem.

5. A_π in $L_p[a, b]$, $1 \leq p < \infty$. With the tedious work done in § 4, we now come to our main result

THEOREM 4. *If π is regular, the operator A_π is the infinitesimal generator of a semi-group of class (C_0) in $L_p[a, b]$, $1 \leq p < \infty$.*

Proof. The assumptions on the coefficients of A in (1.3) are such that standard changes of independent and dependent variables[5] can be made to show that A_π in $L_p[a, b]$ is equivalent in the sense of Definition 2 to $\tilde{A}_\pi$ in $L_p[0, 1]$, where

$$\tilde{A}_\pi = \Omega_{\tilde{\pi}} + r_1 I . \tag{5.1}$$

The conditions $\tilde{\pi}$ are as in (1.2) and can readily be shown to be regular if and only if conditions π are regular.

The function r_1 in (5.1) is in $L_\infty[0, 1]$, and therefore $r_1 I$ is a bounded operator in any L_p. So $\tilde{A}_\pi$ is obtained by perturbing $\Omega_{\tilde{\pi}}$ by a bounded operator. Perturbation theory shows that $\tilde{A}_\pi$ generates a semi-group of class (C_0) if and only if $\Omega_{\tilde{\pi}}$ does (see Hille and Phillips [10], Theorem 13.2.1).

This reduces our proof to that of showing that for regular π the operators $\Omega_\pi = D^2$ generate semi-groups of class (C_0) in any $L_p[0, 1]$, $1 \leq p < \infty$. This extension of Theorem 2 we shall now give.

Let π^+ denote the boundary conditions adjoint to π relative to the operator D^2 (Coddington and Levinson [2], pp. 288–293). It is readily checked that the Grassmann coordinates (A', B', C', D', E', F') of π^+ are obtained from those of π by interchanging F and C and taking complex conjugates. From (3.3) it follows that π^+ is in the set τ_2 if and only if π is.

[5] See Courant and Hilbert [3], p. 250.

Let π, and therefore π^+, be regular boundary conditions. Then by Lemma 1 the resolvent $R(\lambda; \Omega_\pi)$ exists for $\Re(\lambda)$ greater than some ω_0, and it is expressed by (4.2).

We denote the norm of a bounded linear operator T in L_p by $N_p\{T\}$. Then by Theorem 2 and the Feller-Phillips-Miyadera Theorem (Hille and Phillips [**10**], p. 360), we have

$$N_p\{[R(\lambda; \Omega_\pi)]^n\} \leq M_p(\lambda - \omega_0)^{-n}, \Re(\lambda) > \omega_0 , \tag{5.2}$$

$p = 1, 2$ and $n = 1, 2, \cdots$

Now $R(\lambda; \Omega_\pi)$ is defined by (4.2) on the space of continuous functions, which is dense in $L_p[0, 1]$, $1 \leq p < \infty$. If we let $M = \max(M_1, M_2)$ and apply the Riesz Convexity Theorem (Zygmund [**14**], p. 198), we obtain (5.2) for $1 \leq p \leq 2$. By the Feller-Phillips-Miyadera Theorem, this is sufficient for Ω_π to generate a semi-group of class (C_0) in L_p, $1 \leq p \leq 2$.

Also by Theorem 2 and the above argument, Ω_{π^+} generates a semi-group of class (C_0) in any $L_p\,[0, 1]$, $1 \leq p \leq 2$. It is readily shown that Ω_{π^+} in L_q and $\Omega_\pi +$ in L_p, $1/p + 1/q = 1, 1 < p \leq 2$, are adjoints of each other. The theory of adjoint semi-groups (Hille and Phillips [**10**], Chapter IV) shows that Ω_π in L_q generates a semi-group of class (C_0), since Ω_{π^+} does in L_p. This completes the proof of our theorem.

6. Non-regular π. One result relating to the necessity of regularity of π for A_π to generate a semi-group of class (C_0) in $L_p[a, b]$ is given in

LEMMA 5. *If A_π generates a semi-group of class (C_0) in $L_2[a, b]$, then π is regular.*

Proof. As we saw in the proof of Theorem 4, it is sufficient to prove this result for $\Omega_\pi = D^2$ in $L_2[0, 1]$.

Let π be a set of non-regular boundary conditions. It is simply a matter of computation to show that for the function $u(x) = 1, 0 \leq x \leq \frac{1}{2}$, and $u(x) = 0, \frac{1}{2} < x \leq 1$ we get in (4.2)

$$\|R(\lambda; \Omega_\pi)u\|_2 > C\lambda^{-3/4} \tag{6.1}$$

for all real λ sufficiently large and $C > 0$. Thus, by the Feller-Phillips-Miyadera Theorem, Ω_π does not generate a semi-group of class (C_0) in $L_2[0, 1]$.[6] This proves our result.

We now have[7]

[6] Indeed, this proves that Ω_π does not generate a semi-group of the more general class (A) in $L_2[0, 1]$ since it is not true that $\lambda R(\lambda; \Omega_\pi)u \to u$ as $\lambda \to +\infty$ (Hille and Phillips [**10**], p. 322).

[7] By a more careful analysis, the complete result can probably be proven that regularity of π is necessary for A_π to generate a semi-group of class (C_0) in $L_p[a, b]$.

THEOREM 5. *Let π and π^+ be adjoint boundary conditions relative to the operator D^2. If both Ω_π and Ω_{π^+} generate semi-groups of class (C_0) in any $L_p[0, 1], 1 < p < \infty$, then π and π^+ are regular.*

Proof. Suppose that Ω_π and Ω_{π^+} generate semi-groups of class (C_0) in some $L_p[0, 1]$. Then Ω_π generates a semi-group of class (C_0) in $L_q[0, 1]$, $1/p + 1/q = 1$. An application of the Riesz Convexity Theorem, as in Theorem 4, shows that Ω_π generates a semi-group of class (C_0) in $L_2[0, 1]$. By Lemma 5, π is regular, and therefore also π^+. This completes the proof.

For certain of the non-regular π, other Lebesgue spaces can be chosen in which operators Ω_π are defined and generate semigroups of class (C_0). The construction of these spaces is suggested by the method of proof used in part (b) of Lemma 3.

Suppose that conditions π are given by

$$\begin{aligned} u(0) &= au'(1) \\ u(1) &= 0 \end{aligned} \qquad |a| \geq 1 . \tag{6.2}$$

Then, if $G(x, \tau, \lambda)$ is the Green's function of Ω_π, it can be shown that $G_1(x, \tau, \lambda) \equiv |G(\tau, x, \lambda)|$ is the Green's function for Ω_{π_1}, where conditions π_1 are given by

$$\begin{aligned} u(0) &= 0 \\ u(1) &= |a| u'(0) . \end{aligned} \tag{6.3}$$

Also Ω_{π_1} has the real, non-negative eigenfunction $\phi(x) = \sigma^{-1} sh\sigma x$ where σ is the largest real root of $sh\sigma = |a|\sigma$. In a manner similar to that in part (b) of Lemma 3, one can show that Ω_π can be defined in the Lebesgue space $L_1([0, 1], \phi(x)dx)$ as the generator of a semi-group of class (C_0). This space is also norm equivalent to the space $L_1([0, 1], dx^2)$.

The linear homeomorphism of $L_1([0, 1], dx^2)$ onto $L_1([0, 1], d(1 - x)^2)$ defined by $u(x) \to u(1 - x)$, shows that $\Omega_{\tilde{\pi}}$ generates a semi-group of class (C_0) in $L_1([0, 1], d(1 - x)^2)$ where the conditions $\tilde{\pi}$ are given by

$$\begin{aligned} u(0) &= 0 \\ u(1) &= -\,au'(0) . \end{aligned} \tag{6.4}$$

In each of these spaces, $L_1[0, 1]$ can be shown to be a dense subspace. The operators Ω_π and $\Omega_{\tilde{\pi}}$ can be shown to be equivalent to singular operators in $L_1[0, 1]$.

We do not know whether similar results hold for other non-regular π.

Acknowledgment

The author wishes to express his gratitude to Professor G. Birkhoff for his encouragement and advice.

Bibliography

1. G. D. Birkhoff, "*Boundary value and expansion problems of ordinary linear differential equations* Trans. Amer. Math. Soc., **9** (1908), 373–395.
2. E. Coddington and N. Levinson, *Theory of Ordinary Differential Equations* McGraw-Hill, New York, 1953.
3. R. Courant and D. Hilbert, *Mathematische Physik*, Vol. I, Springer, Berlin, 1931.
4. W. Feller, *The parabolic differential equation and the associated semi-groups of transformations*, Ann. of Math. **55** (1952), 468-519.
5. ———, *On the generation of unbounded semi-groups of bounded linear operators*, Ann. of Math. **58** (1953), 166–174.
6. ———, *The general diffusion operator and positivity preserving semi-groups in one dimension* Ann. of Math. **60** (1954), 417-436.
7. ———, *On second order differential operators*, Ann. of Math. **61** (1955), 90–105.
8. ———, *Generalized second order differential operators and their lateral conditions*, Illinois J. Math., **1** (1957), 459-504.
9. E. Hille, *The abstract Cauchy problem and Cauchy's problem for the parabolic differential equation*, Jour. d'Analyse Math., **3** (1953/54), 81-198.
10. E. Hille and R. S. Phillips, *Functional Analysis and Semi-Groups*, Amer. Math. Soc. Publications, Vol. XXXI, 1957.
11. Hodge and Pedoe, *Methods of Algebraic Geometry*, Vol. I, University Press, Cambridge, 1947.
12. J. D. Tamarkin, *Some general problems of the theory of ordinary linear differential equations and expansion of an arbitrary function in a series of fundamental functions*, *Math. Zeit.*, **27** (1927), 1-54.
13. D. V. Widder, *The Laplace Transform*, Princeton University Press, 1946.
14. A. Zygmund, *Trigonometrical Series*, Chelsea, New York, 1952.

THE RAND CORPORATION

RECURRENT MARKOV CHAINS

S. OREY

In this paper Markov chains $\{X_i\}$, $i = 1, 2, \cdots$, which stationary transition probabilities are considered which take values in some measurable space $(S, \mathscr{B})$ and satisfy

(*) The Borel field $\mathscr{B}$ is separable and there exists a sigma finite measure m on $(S, \mathscr{B})$ such that P[entering E at some time $[X_0 = x] = 1$ for all $x \in S$ and all $E \in \mathscr{B}$ with $m(E) > 0$, where P is the underlying probability measure.

Such chains were introduced by Harris in [6], [7]. Let $P^n(x, E)$ be the n-step transition probability, $P^1(x, E) = P(x, E)$. In [7] it is proved that there exists a unique (up to constant factor) sigma finite measure Q which is *stationary* in the sense that $Q(E) = \int_S P(x, E)Q(dx)$.

Section 1 establishes some preliminary results. The relationship between (*) and Doeblin's condition is investigated. The results of Harris [6], [7] are summarized and extended. Note that many notational conventions used throughout the paper are introduced in § 1.

In § 2 it is shown that after the deletion of an inessential Q-null set the process splits up into a finite number, d, of disjoint cyclically moving classes.

Section 3 studies the asymptotic behavior of $P^n(x,\cdot)$ in case the stationary measure Q happens to be a probability measure. The approach is the "direct" approach of Markov and Doeblin and Doob [4]. It is shown that if $d = 1$, the total variation of $(P^n(x,\cdot)-Q)$ approaches 0 as n approaches ∞; for $d > 1$ the convergence statement must be modified in an obvious way. For the relationship of these results to those of [3] see the beginning of § 3.

Section 4 considers the asymptotic behavior of

$$U(n) = \sum_{N=1}^{\infty} P\left[\sum_{i=1}^{N} f(X_i) = n\right],$$

where f is a measurable function from S into the positive integers. If $\int_S f(x)Q(dx) < \infty$, $U(n)$ is for large n approximately a periodic function. The period depends both on the $\{X_i\}$ process and on f; this period may be greater than 1 even though the d associated with the $\{X_i\}$ process is 1 and $f(x) = 1$ for a set of x of positive Q-measure.

Section 5 is concerned with the behavior of normed sums,

Received September 11, 1958. This paper was prepared with the partial support of the Office of Ordnance Research, U. S. Army under Contract DA-04-200-ORD-171.

$$\frac{1}{B_n}\sum_{i=1}^{n} f(X_i) ,$$

where f is a real-valued measurable function on S. Neither the method of Doeblin (exploited and developed in [2]) nor that of Bernstein (used in [1]) is directly applicable to processes merely satisfying (*) and $Q(S) = 1$. Nevertheless, ideas from both those methods combined with the results of the previous sections make it possible to give conditions under which such normed sums obey the central limit theorem and even obey the Erdös-Kac-Donsker invariance principle. Results from [1] are made use of. As indicated at the relevant places, ideas of [2] and [8] are also used. The work of this section naturally leads to some problems related to recurrence times, and these are discussed in § 6.

1. Preliminary results. Let m be a measure satisfying (*). If ν is any measure on $(S, \mathscr{B})$ $p^n_\nu(x,\cdot)$, $P^n_{0\nu}(x,\cdot)$ are to denote, respectively, the absolutely continuous and singular component of P^n with respect to ν. Superscripts will be omitted when $n = 1$. It will be assumed that p^n_ν is measurable in (x, y) in the product space $S \times S$; this assumption is justified by Doob [4], p. 616. It will also be assumed that for all positive integers s and t

$$p^{s+t}_\nu(x, y) \geqq \int_S p^s_\nu(z, y)p^t_\nu(x, z)\nu(dz)$$

holds for all x and y; that such a choice of densities is always possible was shown in Doob [4], p. 146.

When the subscript ν is omitted it will be assumed in this section that ν is m; in all the other sections the omission of ν will mean $\nu = Q$.

If A is a Borel set, $m(A) > 0$, "the process on A" will have the same significance as in [7], i.e., if $X_{A0}, X_{A1}, \cdots$ are the successive members of the sequence $X_0, X_1, \cdots$ with values in A, $\{X_{Ai}\}$, $i = 0, 1, \cdots$ is the process on A. This is a Markov process with transition probability

$$(1.1) \qquad P_A(x, E) = P(x, E) + \int_{S-A} P(x, dy)P(y, E) + \int_{S-A}\int_{S-A} P(x, dy)P(y, dz)P(z, E) + \cdots$$

for every Borel subset E of A and $x \in A$.

The process on A will also satisfy (*). Notions defined for the original process can thus be relativized to A; notationally this is indicated by a subscript A, e.g., $p^n_A(x, y)$ is defined like $p^n(x, y)$ but using $P_A(x, E)$ in place of $P(x, E)$.

If $0 < m(A) < \infty$, v belongs to the open interval $(0, 1)$, and j is a positive integer, $K(A, v, j)$ is to be the set of all $x \in A$ such that

$$m\left\{y \in A: \sum_{i=1}^{j} p^i(x, y) > \frac{1}{j}\right\} > vm(A) .$$

By the convention explained above $K_C(A, v, j)$ is meaningful whenever C is a Borel set including A (the definition now involves p_C^i in place of p^i).

Harris observed that if $m(A) < \infty$,

$$A = \bigcup_{k=1}^{\infty} K(A, r, k)$$

easily follows from (*), for every $r \in (0, 1)$. Lemma 1.1 is a modification of a lemma in [7].

LEMMA 1.1. *Let A, B be Borel sets such that $m(B) > 0, m(A) < \infty$, $B \subseteqq K(A, v, k)$ for some $v \in (0, 1)$ and some positive integer k. Then $K(B, r, k) = B$ provided $(1 - r) \geqq (1 - v)m(A)/m(B)$, where $r \in (0, 1)$.*

Proof. For every $x \in B$,

$$m\left\{y \in B: \sum_{i=1}^{k} p^i(x, y) < \frac{1}{k}\right\} < (1 - v)m(A)$$

$$= \left[(1 - v) \frac{m(A)}{m(B)}\right] m(B) < (1 - r)m(B)$$

if the proviso of the lemma holds.

LEMMA 1.2. *Let $A = K(A, r, k)$, $r \in (0, 1)$, k a positive integer, A a Borel set of finite m-measure. Then there exists a probability measure φ on A and numbers $\alpha, \eta \in (0, 1)$ such that $P^{k+1}(x, E) \leqq 1 - \eta$ whenever $E \subseteqq A$, $E \in \mathscr{B}$, $\varphi(E) \leqq \alpha$ and $x \in A$.*

Proof. Let η be any number such that $0 < \eta < rm(A)k^{-3}$. Let x be some point in A. Define

$$C = \left\{z \in A: \sum_{i=1}^{k} p^i(x, z) > \frac{1}{k}\right\} , \quad C_i = \left\{z \in C: p^i(x, z) > \frac{1}{k^2}\right\} ,$$

Then

$$C \subseteqq \bigcup_{i=1}^{k} C_i \text{ and } m(C) > rm(A) .$$

Let E be a Borel subset of A, and suppose $P^{k+1}(x, E) > 1 - \eta$. Then

$$\eta > 1 - P^{k+1}(x, E) = P^{k+1}(x, S - E) \geqq \frac{1}{k} \sum_{i=1}^{k} \int_{C_i} p^i(x, y) P^{k+1-i}(y, S - E) m(dy)$$

$$\geqq \frac{1}{k^3}\sum_{i=1}^{k}\int_{C_i} P^{k+1-i}(y, S-E)m(dy) = \frac{1}{k^3}\sum_{i=1}^{k}\int_{C_i}(1 - P^{k+1-i}(y, E))m(dy)$$

$$\geqq \frac{1}{k^3}\left[r\cdot m(A) - \sum_{i=1}^{k}\int_{C_i} P^{k+1-i}(y, E)m(dy)\right]$$

$$\geqq \frac{1}{k^3}\left[r\cdot m(A) - k\int_A \sum_{i=1}^{k} P^i(y, E)m(dy)\right].$$

Let

$$\varphi(E) = \frac{1}{km(A)}\int_A \sum_{i=1}^{k} P^i(y, E)m(dy) \text{ and } \alpha = \frac{rm(A) - k^3\eta}{k^2 m(A)}$$

The inequality above yields $\varphi(E) > \alpha$, proving the lemma.

COROLLARY 1.2.A. *Let A, C be Borel sets, $m(C) < \infty$, $A \subseteqq C$ and $A = K_C(A, r, k)$, where $r \in (0, 1)$ and k is a positive integer. Then there exists a probability measure φ on A and numbers $\alpha, \eta \in (0, 1)$ such that $P_C^{k+1}(x, E) \leqq 1 - \eta$ whenever $x \in A$, and E is a Borel subset of A with $\varphi(E) \leqq \alpha$.*

Proof. Since the process on C also satisfies (*) the lemma applies to it, yielding the corollary.

COROLLARY 1.2B. *If A, C, r, k are as in Corollary 1.2A the process on A satisfies Doeblin's condition.*[1]

Proof. $A = K_A(A, r, k)$ since $K_C(A, r, k) \subseteqq K_A(A, r, k)$. So the conclusion of Corollary 1.2A applies with P_A for P_C; but this gives Doeblin's condition for the process on A.

Let $\mathscr{D}$ be the collection of $A \in \mathscr{B}$ such that the process on A satisfies Doeblin's condition.

It will be seen that $\mathscr{D}$ is an important collection. In [6] Harris announced a result which, slightly extended, asserts that when (*) holds one has for all $x, y \in S$

$$\frac{\sum_{n=0}^{N} P^n(x, E)}{\sum_{n=0}^{N} P^n(y, F)} \longrightarrow \frac{Q(E)}{Q(F)} \text{ and } N \longrightarrow \infty \tag{1.2}$$

for all Borel sets E, F with $E \subseteqq F$, $F \in \mathscr{D}$. In this connection see also Theorem 1. In [7] the question of more general validity of (1.2) was raised. If merely $E, F \in \mathscr{B}$, $0 < m(F) < \infty$ is assumed, (*) does not

[1] It is actually Doob's generalization of Doeblin's condition that is referred to. See [4], p. 192, Hypothesis (D).

imply (1.2) for all $x, y \in S$. It is easy to give examples of chain satisfying (*) and with denumerable state space where (1.2) is violated even in the special case where $x = y$ or where $E = F$. On the other hand, when S is denumerable one shows easily that, regardless of whether or not the process satisfies (*), one has for $x \neq y$,

$$P[X_i = y \text{ for some positive } i \mid X_0 = x] \leqq \liminf_{N\to\infty} \frac{\sum_{n=0}^{N} P^n(x, E)}{\sum_{n=0}^{N} P^n(y, E)}$$

$$\leqq \limsup_{N\to\infty} \frac{\sum_{n=0}^{N} P^n(x, E)}{\sum_{n=0}^{N} P^n(y, E)} \leqq \frac{1}{P[X_i = x \text{ for some positive } i \mid X_0 = y]} \cdot$$

It follows that (1.2) will hold for every pair of x and y belonging to the same recurrence class. In particular in case (*) holds and S is denumerable (1.2) will hold for all x and y outside a fixed Q-null set. It would be interesting if this could be shown to hold even when S is not denumerable.

Harris showed in [7] that if $A = K_A(A, r, k)$ the process on A has a stationary probability measure; this also follows from Corollary 1.2B. Whenever the process on some Borel set B has a stationary probability measure it will be denoted by Q_B.

LEMMA 1.3. *If* $A, B \in \mathscr{B}, B \subseteqq A, A \in \mathscr{D}$ *then* $B \in \mathscr{D}$.

Proof. Assume the hypotheses of the lemm. Then there is a positive function $\varepsilon(n)$ such that

$$\sum_{n=1}^{\infty} \varepsilon(n) < \infty$$

and $P_A^n(x, E) < Q_A(E) + \varepsilon(n)$ for every Borel subset E of A. Let N be an integer such that

$$\sum_{n=N}^{\infty} \varepsilon(n) < \frac{1}{4} .$$

Then there exists an integer M such that for all $x \in B$ and Borel subsets E of B,

$$P_B^N(x, E) = \sum_{i=N}^{\infty} P[X_{BN} \in E, X_{BN} = X_{Ai} \mid X_{A0} = x]$$
$$< \sum_{i=N}^{M} P[X_{BN} \in E, X_{BN} = X_{Ai}] + \frac{1}{4} < \sum_{i=N}^{M} [Q_A(E) + \varepsilon(i)] + \frac{1}{4} .$$

So when $Q_A(E) < 1/(4(M - N))$, $P_B^n(x, E) < 3/4$ proving the lemma.

LEMMA 1.4. *Let $0<r<1$, $A \in \mathscr{D}$ and $m(A)<\infty$. Then $A \subseteqq (A, r, k)$ for some k.*

Proof. If the lemma is false there is for every k an $x_k \in A$ such that if

$$E_k = \left\{y \in A : \sum_{i=1}^{k} p^i(x_k, y) \leqq \frac{1}{k}\right\},$$

$m(E_k) > (1-r)m(A)$. Let E'_k satisfy $E'_k \subseteqq E_k$, $m(E - E'_k) = 0$, and $P_0^j(x_k, E'_k) = 0$, $j = 1, \cdots, k$. So

$$\sum_{j=1}^{k} P^j(x_k, E'_k) \leqq \frac{1}{k} m(E'_k) \leqq \frac{1}{k} m(A),$$

and the last term approaches zero as k approaches ∞. Therefore, $Q_A(E_k)$ approaches zero. Since Q_A and m are finite on A and m is absolutely continuous with respect to Q_A, $m(E_k)$ must tend to zero, which results in a contradiction.

The restriction $m(A) > 0$ or $m(A) < \infty$ appeared frequently above. Note that there always exists finite measures q having the same null sets as Q and therefore satisfying (*). If such a q is chosen for m in the preceding lemma the hypothesis $m(A) < \infty$ may be dropped and the conclusion may be weakened to $A \subseteqq K(S, r, k)$ for some k. Letting $S_k = K(S, r, k)$, where r is fixed, $0 < r < 1$, the preceding sentence can be restated thus: $A \in \mathscr{D}$ implies $A \subseteqq S_k$ for some positive integer k. Clearly $S_k \subseteqq S_{k+1}$, $k = 1, 2, \cdots$. By the remark preceding Lemma 1.1 $S = \bigcup_{i=1}^{\infty} S_k$. Lemma 1.1 asserts that

$$K(S_k, r', k) = S_k \text{ if } (1 - r') \geqq (1 - r)q(S)/q(S_k).$$

For k sufficiently big such a choice of r' will be possible, since S_k approaches S. So then by Corollary 1.2B such S_k belong to $\mathscr{D}$. Now by Lemma 1.3 all S_k and all their Borel subsets belong to $\mathscr{D}$. This proves the following theorem.

THEOREM 1. *If (*) holds S can be represented as a union of Borel sets S_i, $i = 1, 2, \cdots$ such that $S_i \subseteqq S_{i+1}$ and a Borel set A belongs to $\mathscr{D}$ if and only if $A \subseteqq S_k$ for some k.*

Harris showed that if A is a Borel set such that the process on A has a stationary probability measure Q_A, Q_A can be extended to a stationary sigma finite measure on S, $\overline{Q}_A$. $\overline{Q}_A(E)$ is the expected number of visits to E up to and including the first return to A if the process starts with the initial distribution Q_A. Analytically

$$\overline{Q}_A(E) = \int_A Q_A(dx) P_A(x, E) \tag{1.3}$$

where $P_A(x, E)$ is defined by (1.1) (regardless of whether $E \subseteqq A$ or not).

The process $\{X_i\}$ is determined by the function $P(x, E)$ and the initial distribution. It will sometimes be convenient to indicate the initial distribution as a subscript on the expectation operator E, or on P.

Let $A \in \mathscr{B}$, and suppose $\{X_{Ai}\}$ possesses a stationary probability measure Q_A. Let V be the least positive integer such that $X_V \in A$. It is not hard to verify that if f is any measurable function from S into the positive integers one has

$$E_{Q_A}\left\{\sum_{i=1}^{V} f(X_i)\right\} = \int_S f(x)\overline{Q}_A(dx) \tag{1.4}$$

where both sides are infinite if either one is.[2]

2. Cyclic decomposition. In this section it is shown that the arguments applied by Doob in [4] to processes satisfying Doeblin's condition can be extended to the case where only (*) is assumed. In particular there exists a Borel set C with positive Q measure such that $\text{g.l.b.}_{x,y\in C}\, p^{\alpha}(x, y) > 0$ for some positive integer α. This leads at once to the desired decomposition. The only place where it is necessary to deviate from the treatment of [4] is in the proof of Lemma 2.1 below (Lemma 5.3, p. 200 of [4]).

LEMMA 2.1. *If* (*) *holds there exist* $A, B \in \mathscr{B}$ *and a positive integer* n *such that* $Q(A) > 0$, $Q(B) > 0$ *and* $\text{g.l.b.}_{\substack{x\in A\\ y\in B}}\, p^n(x, y) > 0$.

Proof. Let $D \in \mathscr{B}$ satisfy

$$Q\left\{y \in D: \sum_{i=1}^{k} p^i(x, y) > \frac{1}{k}\right\} > r > 0 ,$$

for all $x \in D$. By section 1 such r, D, k exist. Then there exists a $D_1 \in B$ and a positive integer n_1 such that $D_1 \subseteqq D$, $Q(D_1) > 0$, and

$$Q\left\{y: p^{n_1}(x, y) > \frac{1}{k^2}\right\} > \frac{r}{k}$$

for all $x \in D_1$. Also there must exist $D_2 \in B$, and a positive integer n_2 such that $D_2 \subseteqq D_1$, $Q(D_2) > 0$ and for all $x \in D_2$

$$Q\left\{y: p^{n_1}(x, y) > \frac{1}{k^2} \text{ and } Q\left\{z \in D: p^{n_2}(y, z) > \frac{1}{k^2}\right\} > \frac{r}{k}\right\} > \frac{r}{k^2} .$$

Let

[2] This is part of the assertion of Lemma 6. The discrete analogue of this formula is formula (A) of the appendix to [**2**]. Cf. also footnote 7.

$$\hat{H}_1 = \left\{(x, y): x \in D_2,\ p^{n_1}(x, y) > \frac{1}{k^2} \text{ and } Q\left\{z \in D: p^{n_2}(y, z) > \frac{1}{k^2}\right\} > \frac{r}{k}\right\},$$

and

$$\hat{H}_2 = \left\{(y, z) \in D \times D: \text{there exists an } x \text{ such that } (x, y) \in \hat{H}_1 \text{ and } p^{n_2}(y, z) > \frac{1}{k^2}\right\}.$$

Let $\hat{Q}$ be the product measure $Q \times Q$ in the product space $D \times D$. Given any $\hat{Q}$-null set N of $D \times D$ it is clear that it is always possible to choose two points (x_0, y_0), (x_1, y_1) in $(D \times D) - N$ such that $(x_0, y_0) \in \hat{H}_1$, $(x_1, y_1) \in \hat{H}_2$ and $y_0 = x_1$. From here on the proof follows that of [4], middle of p. 201. Since the ν of Doob corresponds sometimes to n_1 and sometimes to n_2 in an obvious way, the conclusion here will be that $p^{n_1+n_2}(x, y)$ is bounded away from zero for $x \in A$, $y \in B$. The lemma follows if $n = n_1 + n_2$.

THEOREM 2.1. *If* (*) *holds there is a* $C \in \mathscr{B}$, $Q(C) > 0$ *and a positive integer* α *such that* $\operatorname*{g.l.b.}_{x,y \in C} p^{\alpha}(x, y) > 0$.

Proof. This follows easily from the preceding lemma. For details see [4], Lemma 5.4.

Let C satisfy

$$\operatorname*{g.l.b.}_{x,y \in C} p^{\alpha}(x, y) > 0 .$$

It is known (cf. [4], p. 202) that if d is the greatest common divisor of

$$I(C) = \{\alpha : \operatorname*{g.l.b.}_{x,y \in C} p^{\alpha}(x, y) > 0\}$$

then all sufficiently large multiples of d belong to $I(C)$. With no more essential variations from the development as given in [4] one obtains:

THEOREM 2.2. *Suppose* (*) *holds. Then there exists a unique integer* d *such that whenever* $C \in \mathscr{B}$, $Q(C) > 0$ *and* $I(C)$ *non-void then* d *is the greatest common devisor of* $I(C)$. *There exists a partition of* S *into Borel sets* $C_0, C_1, \cdots, C_{d-1}, F$, *such that* $Q(F) = 0$ *and for* $x \in C_i$ $P(x, C_{i+1}) = 1$, *where the subscripts are integers modulo* d.

3. Convergence of $P^n(x, E)$ when Q is finite. In this section it is assumed that (*) holds and that Q is a probability measure. The basic method used here goes back to Markov and Doeblin; in detail, however, this presentation leans on Chapter V of Doob [4].

In [3] Doob investigated Markov chains possessing stationary probability measures. To see that under the assumptions of this section

Theorem 5 of [3] is applicable one needs only to check that $P_0^n(x, S) \to 0$ as $n \to \infty$ for all $x \in S$. For (cf. [3], p. 409) if for $m = 1, 2, \cdots, A_m$ is a Q-null set such that $P_0^m(x, S - A_m) = 0$,

$$P[X_n \in A_n | X_0 = x] = P[X_1 \in A_1, X_2 \in A_2, \cdots, X_n \in A_n | X_0 = x] .$$

So if $P[X_n \in A_n | X_0 = x] > \varepsilon > 0$ for all n,

$$P\left[\text{ever entering } S - \bigcup_{i=1}^{\infty} A_i | X_0 = x\right] < 1 - \varepsilon ,$$

contradicting (*). Theorem 3 below, however, shows somewhat more than would follow from an application of Theorem 5 of [3] since it proves, say when $d = 1$, that the variation of $(P^n(x, \cdot) - Q)$ approaches 0 when n approaches ∞ while [3] gives only $(P^n(x, E) - Q(E))$ approaches 0 as n approaches ∞ for every $E \in \mathscr{B}$.

If φ is a totally additive set function on S, $||\varphi||$ is to be the total variation of φ on S. $T\varphi$ $(T^n\varphi)$ will denote the measure

$$\int_S P(x, \cdot)\varphi(dx) \left(\int_S P^n(x, \cdot)\varphi(dx)\right) .$$

LEMMA 3.1. *Suppose there exists a Borel set C of positive Q-measure, an $\varepsilon > 0$ and for every $x, y \in S$ a positive integer $w(x, y)$ such that $p^w(x, z) > \varepsilon$, $p^w(y, z) > \varepsilon$ for all $z \in C$. Then for any two probability measures φ_1, φ_2 on S there is an n such that $||T^n\varphi_1 - T^n\varphi_2|| \leqq (1 - \varepsilon Q(C)/2)||\varphi_1 - \varphi_2||$.*

Proof. Note that $||T^n\varphi_1 - T^n\varphi_2||$ is nonincreasing in n. Let ρ_x be the measure such that $\rho_x(E) = 1$ (0) if $x \in E$ $(x \in S - E)$, $E \in \mathscr{B}$. Consider first the case $\varphi_2 = \rho_x$, and write φ for φ_1. There exists a unique real number α and measure ν such that $\varphi = \nu + \alpha\rho_x$ and $\nu(\{x\}) = 0$. Let $\rho = (1 - \alpha)\rho_x$. Let $A_i = \{y : w(x, y) = i\}$. So the A_i are disjoint and $\bigcup_1^\infty A_i = S$. Let

$$\varepsilon' = \frac{\varepsilon Q(C)(1 - \alpha)}{4}$$

and choose N so large that

$$\nu\left(\bigcup_{i=N+1}^{\infty} A_i\right) < \varepsilon' .$$

Define the measures ν_i, ρ_i by $\nu_i(E) = \nu(E \cap A_i)$ and $\rho_i = ||\nu_i||\rho_x$. Then

$$\left\|\nu - \sum_{i=1}^{N} \nu_i\right\| < \varepsilon', \left\|\rho - \sum_{i=1}^{N} \rho_i\right\| < \varepsilon' .$$

Now

$$\left\|T^N\Big(\sum_{i=1}^N \nu_i\Big) - T^N\Big(\sum_{i=1}^N \rho_i\Big)\right\| \leqq \sum_{i=1}^N \|T^N\nu_i - T^N\rho_i\| \leqq \sum_{i=1}^N \|T^i\nu_i - T^i\rho_i\| .$$

From the definition of ν_i and ρ_i it follows by an obvious argument that the densities of $T^i\nu_i$, $T^i\rho_i$ with respect to Q are equal at least to $\varepsilon\|\nu_i\|$ everywhere on C, from which it follows immediately that there is enough cancellation to insure $\|T^i\nu_i - T^i\rho_i\| \leqq 2\|\nu_i\|(1 - \varepsilon Q(C))$. So

$$T^N\nu - T^N\rho\| \leqq \left\|\sum_{i=1}^N T^N\nu_i - \sum_{i=1}^N T^N\rho_i\right\| + 2\varepsilon' \leqq 2\|\nu\|\Big(1 - \frac{\varepsilon Q(C)}{2}\Big) ,$$

proving the lemma for this special case.

Now let φ_1 and φ_2 be arbitrary probability measures. Let $\mu_1(\mu_2)$ be the positive (negative) variation of $\varphi_1 - \varphi_2$. So $\varphi_1 - \varphi_2 = \mu_1 - \mu_2$. Let α_i, ν_i be the unique real number and measure such that $\mu_i = \nu_i - \alpha_i\rho_x$, $\nu_i\{x\} = 0, i = 1, 2$. Let $\rho = \|\mu_1\|\rho_x$. For all big enough n and $i = 1, 2$ one has by the above:

$$\|T^n\mu_i - T^n\rho\| \leqq \|\mu_i - \rho\|\Big(1 - \frac{\varepsilon}{2}Q(C)\Big) \leqq 2\|\mu_i\|\Big(1 - \frac{\varepsilon Q(C)}{2}\Big) .$$

So

$$\begin{aligned}\|T^n(\varphi_1 - \varphi_2)\| &= \|T^n(\mu_1 - \mu_2)\| \leqq \|T^n\mu_1 - T^n\rho\| \\ &+ \|T^n\mu_2 - T^n\rho\| \leqq 2(\|\mu_1\| + \|\mu_2\|)\Big(1 - \frac{\varepsilon Q(C)}{2}\Big) \\ &= 2\|\varphi_1 - \varphi_2\|\Big(1 - \frac{\varepsilon Q(C)}{2}\Big) .\end{aligned}$$

LEMMA 3.2. *Assume (*), and Q a probability measure and the hypotheses of Lemma* 3.1. *Then $\|T^n\varphi - Q\|$ approaches* 0 *as n approaches ∞ for every probability measure φ.*

Proof. Assume the hypotheses of the lemma, and let φ be a probability measure. For every n one can find an m such that

$$\|T^m\varphi - T^mQ\| < \Big(1 - \frac{\varepsilon Q(C)}{2}\Big)^n ,$$

by repeated applications of the previous lemma. Since $\|T^k\varphi - T^kQ\|$ is nonincreasing in k and $T^kQ = Q$ for all k the lemma follows.

LEMMA 3.3. *Assume (*) and that $d = 1$ in the decomposition of Theorem* 3.2. *Then there is an $\varepsilon > 0$, a Borel set C of positive Q-measure,* *and for every* $x, y \in S$ *a positive integer* w *such that* $p^w(x, z) > \varepsilon$, $p^w(yz) > \varepsilon$ *for all* $z \in C$.

Proof. Assume the hypotheses of the lemma. Let C be a set satisfying Theorem 2.1, so that $\underset{x,y\in C}{\text{g.l.b.}}\ p(x, y) > 0, Q(C) > 0$. Suppose first that

(3.1) there exists a positive integer N, a $\delta > 0$, and for every $x, y \in S$ an integer γ such that

$$\sum_{i=\gamma+1}^{\gamma+N} P^i(x, C) > \delta,\ \sum_{i=\gamma+1}^{\gamma+N} P^i(y, C) > \delta\ .$$

Let M be a positive integer such that

$$\underset{x,y\in C}{\text{g.l.b.}}\ p^n(x, y) > 0$$

for all $n > M$. By (3.1) there exist positive integers $\beta(x)$, $\beta(y)$ such that $\beta(x) < N$ and $P^{\gamma+\beta(x)}(x, C) > \delta/N$ and $\beta(y) < N$ and $P^{\gamma+\beta(y)}(y, C) > \delta/N$. Then for $z \in \{x, y\}$, $u \in C$ one has

$$\begin{aligned} p^{\gamma+N+M}(z, u) &\geqq \int_C p^{N+M-\beta(z)}(v, u)P^{\gamma+\beta(z)}(z, dv) \\ &\geqq \frac{\delta}{N} \underset{v,u\in C}{\text{g.l.b.}}\ p^{N+M-\beta(z)}(v, u) \geqq \frac{\delta}{N} \min_{M\leqq n\leqq N+M} \underset{\substack{v\in C\\ u\in C}}{\text{g.l.b.}}\ p^n(v, u)\ . \end{aligned}$$

Thus to prove the lemma it suffices to prove (3.1).

Let δ' be a positive number less than 1. Let

$$A_n = \left\{z: \sum_{i=1}^{n} P^i(z, C) > \delta'\right\}\ .$$

Clearly $A_n \subseteqq A_{n+1}$ and by (*) $Q(A_n) \to 1$ as $n \to \infty$. Let $Q(A_N) > 3/4$. Under the present hypotheses the ergodic theorem shows that for *every* x and y there is an n_0 such that for $n > n_0$ P[number of visits to A_N in n steps $> 3n/4 \mid X_0 = z] > 3/4$ for $z \in \{x, y\}$. Let n be an integer greater than n_0 divisible by four. Let $a_{ik}\,(b_{ik}) = P$[entering A_N for the kth time at ith step$\mid X_0 = z]$, $z = x\,(z = y)$. For $k = 1, 2, \cdots, 3n/4$,

$$\sum_{i=1}^{n} a_{ik} > \frac{3}{4},\ \sum_{i=1}^{n} b_{ik} > \frac{3}{4}\ .$$

Let

$$a_i = \sum_{k=1}^{3/4n} a_{ik},\ b_i = \sum_{k=1}^{3/4n} b_{ik}\ .$$

Then

$$\sum_{i=1}^{n} a_i > \frac{9}{16}n,\ \sum_{i=1}^{n} b_i > \frac{9}{16}n\ .$$

Now $a_i\,(b_i) \leqq P[\text{entering } A_N \text{ at step } i \,|\, X_0 = z] \leqq 1,\ z = x\,(z = y)$. This together with

$$\sum_{i=1}^{n}(a_i + b_i) > \frac{9}{8}n$$

proves that for some $\gamma \leqq n$ both $a_\gamma > 1/8, b_\gamma > 1/8$ holds. So $P^\gamma(x, A_N) > 1/8$, $P^\gamma(y, A_N) > 1/8$. Then

$$\sum_{i=\gamma+1}^{\gamma+N} P^i(z, C) > \int_{A_N} \sum_{i=1}^{N} P^\gamma(z, dv)P^i(v, C) > \frac{1}{8}\delta' \text{ for } z \in \{x, y\}\ .$$

This proves (3.1) with $\delta = \delta'/8$.

THEOREM 3. *Assume* (*) *and that a stationary probability measure* Q *exists.*

If $d = 1$ *in Theorem* 2.2 $\|P^n(x,\cdot) - Q\|$ *approaches* 0 *as* n *approaches* ∞.

More generally, if $d \geqq 1$, *and* $Q_k(A) = Q(A \cap C_k)$, $k = 0, 1, \cdots d-1$, *for all* $A \in \mathscr{B}$, *and* d, C_k *as in theorem* 2.2, *and if for any initial distribution* φ *one defines* $\alpha_i(\varphi) = \lim_{n\to\infty} P_\varphi[X_{nd} \in C_i]$, $i = 0, \cdots, d-1$, *one has*

$$\left\| T^{nd+k}\varphi - d\sum_{i=0}^{d-1}(\alpha_i(\varphi)Q_{k+i}) \right\| \text{ approaches } 0 \text{ as } n \text{ approaches } \infty\ ,$$

where the subscripts are integers modulo $d, k = 0, 1, \cdots, d-1$.

Proof. The first assertion follows from Lemmas 3.2 and 3.3. The reduction of the second assertion to the first one is trivial.

Obviously $P^n(x,\cdot)$ cannot converge if Q is not finite. It may be conjectured that in this case $P^n(x, E) \to 0$ whenever $Q(E) < \infty$; such a result would be very useful. So far no proof of this conjecture has been found, not even under the additional hypothesis $E \in \mathscr{D}$.

4. A renewal theorem. Let $\{X_i\}$ satisfy (*) and assume $\{X_i\}$ has a stationary probability measuae Q. Let f be a measurable function from S to the positive integers. Let

$$U_\varphi(n) = \sum_{N=1}^{\infty} P_\varphi\left[\sum_{i=1}^{N} f(X_i) = n\right],$$

where φ is the initial probability distribution.

Let $\tilde{S} = \{(x, k): x \in S, k$ a positive integer$\}$. Let $\tilde{\mathscr{B}}$ be the smallest Borel field such that for every $E \in \mathscr{B}$ and every positive integer k $\{(x, k): x \in E\} \in \tilde{\mathscr{B}}$. Let $\{\tilde{X}_i\}$ be a Markov process with values in $(\tilde{S}, \tilde{\mathscr{B}})$ having the following transition probabilities:

$$\tilde{P}((x, i), \{(x, i+1)\}) = 1 \text{ for } i = 1, 2, \cdots, f(x) - 1 \text{ and } x \in S ,$$
$$\tilde{P}((x, f(x)), E'\} = P(x, E) \text{ if } E' = \{(z, 1): z \in E\}, E \in \mathscr{B} ,$$
$$\tilde{P}((x, i), \{(x, 1)\}) = 1 \text{ for } i > f(x) .$$

For $E \in \mathscr{B}$ let $E^0 = \{(z, f(z)): z \in E\}$. Let $\tilde{\varphi}$ be the probability measure on $(\tilde{S}, \tilde{\mathscr{B}})$ defined by $\tilde{\varphi}(E^0) = \varphi(E)$, $\tilde{\varphi}(\tilde{S} - S^0) = 0$.

Note that one has

$$U_\varphi(n) = \tilde{P}_{\tilde{\varphi}}[X_n \in S^0] . \tag{4.1}$$

Assume now that

$$\int_S f(x)Q(dx) < \infty .$$

Since $P(x, E) = \tilde{P}_{S^0}((x, f(x)), E^0)$ for all $x \in S$ and $E \in \mathscr{B}$, $\{X_{S^0_i}\}$ has a stationary probability measure $\tilde{Q}_{S^0}$. $\{\tilde{X}_i\}$ satisfies (*), so $\tilde{Q}_{S^0}$ can be extended to a stationary measure $\tilde{Q}$ for $\{\tilde{X}_i\}$. From (1.1), (1.3), (1.4) it follows that

$$\tilde{Q}(\tilde{S}) = \int_S f(x)Q(dx) ,$$

which was assumed finite. So

$$\overline{Q}(\cdot) = \frac{Q(\cdot)}{\int_S f(x)Q(dx)}$$

is a stationary probability measure for $\{\tilde{X}_i\}$. Apply Theorem 2.2 to $\{\tilde{X}_i\}$, obtaining an integer d and classes $C_1, \cdots, C_d, F$. Let

$$\alpha_i(\tilde{\varphi}) = \lim_{m\to\infty} P_{\tilde{\varphi}}[\tilde{X}_{md} \in C_i] .$$

Then

$$U(nd + j) \text{ approaches } d \sum_{i=0}^{d-1} \alpha_i(\tilde{\varphi})\overline{Q}(S^0 \cap C_{j+1})$$

by (4.1) and Theorem 3. Let $C_i' = \{x \in S; (x, f(x)) \in C_i\}$ and let $\tilde{\alpha}_i(\varphi) = \alpha_i(\tilde{\varphi})$ for $i = 0, 1, \cdots, d-1$. Then clearly

$$\tilde{\alpha}_i(\varphi) = \lim_{n\to\infty} P_\varphi\Big(\bigcup_{j=0}^{d-1}\Big[X_n \in C'_{i+j},\ \sum_{i=1}^{n} f(X_i) \equiv j (\mathrm{mod}\ d)\Big]\Big) \tag{4.2}$$

where the subscript on the C' is an integer modulo d.

These arguments establish the following theorem.

THEOREM 4. *Let $\{X_i\}$ satisfy (*) and assume there is a stationary probability measure Q for X_i. Assume $\int_S f(x)Q(dx) < \infty$. Then S can be partitioned into Borel sets $C'_1, \cdots, C'_d, F'$ such that $P(x, F') = 0$ for all $x \in S - F'$, $Q(F') = 0$, and for any initial distribution φ of the $\{X_i\}$ process one has*

$$\lim_{n\to\infty} U_\varphi(nd + j) = \frac{d \sum_{i=0}^{d-1} \tilde{\alpha}_i(\varphi) Q(C'_{j+1})}{\int_S f(x)Q(dx)},\ j = 0, 1, \cdots, d - 1\ ,$$

where $\tilde{\alpha}_i(\varphi)$ is defined by (4.2) and the subscripts are integers modulo d.

If the $\{X_i\}$ are independent (and then automatically identically distributed), $S = C'_0$ and the theorem yields

$$U_\varphi(nd) \to \frac{d}{\int_S f(x)Q(dx)} = \frac{d}{E\{f(X_i)\}}\ ;$$

in this case $U_\varphi(nd + j) \equiv 0$ for $j = 1, \cdots, d - 1$. This is a result of [5].

It seems plausible that $U_\varphi(n)$ approaches 0 as n approaches ∞ in case

$$\int_S f(x)Q(dx) = \infty\ .$$

This would follow from a proof of the conjecture made at the end of the previous section.

5. The invariance principle. Let f be a real-valued measurable function on the state space of the Markov chain $\{X_i\}$. Under certain conditions the sequence of sums

$$\sum_{i=1}^{n} f(X_i)$$

is known to behave like a sequence of partial sums of independent random variables, e.g., the central limit theorem holds for suitable norming constants. Two devices are available for proving results of this kind. The first method, which is due to S. Bernstein, uses the fact that in certain cases the dependence of X_{n+k} on $X_1, \cdots, X_n$ diminishes quickly as k increases; this method is applicable if $\{X_i\}$ satisfies Doeblin's condition. The second method is applicable to certain cases in which the state space is denumerable; the idea in this case, due to Doeblin, is that if V_i is defined to be the ith nonnegative integer n such that $X_n = x$, where x is a fixed state, the sums $f(X_{V_i+1}) + \cdots + f(X_{V_{i+1}})$, $i = 1, 2, \cdots$, are independent and identically distributed. If $\{X_i\}$ is merely assumed to satisfy (*) and have a stationary probability measure Q neither method applies. However, it will be shown below that a combination of the two methods may be used in this case.

Assume now that $\{X_i\}$ satisfies (*). Let $A \in \mathscr{B}$ have positive Q-measure. Let V_i (or $V_i(A)$) be the ith nonnegative integer v such that $X_v \in A$, i.e., $X_{V_i} = X_{A(i-1)}$, $i = 1, 2, \cdots$. Let Y_i (or $Y_i(A)$) be the vector $(X_{V_i+1}, X_{V_i+2}, \cdots, X_{V_{i+1}})$, $i=1,2,\cdots$. For $m=1,2,\cdots$ an m-tuple $(x_1, \cdots, x_m)$ with components in S will be called a *path*. On the set of all paths impose the smallest Borel field containing for $m = 1, 2, \cdots$, all m-dimensional cylinder sets with one dimensional base set in $\mathscr{B}$. Then $\{Y_i\}$ is a Markov process; indeed one obviously has

$$P[Y_{n+k} \in W \mid Y_n, \cdots, Y_1] = P[Y_{n+k} \in W \mid Y_n] = P[Y_{n+k} \in W \mid X_{V_n}]$$

for every Borel set W. $\{Y_i(A)\}$ will be called the *A-path process*. The property of path processes that will be exploited is the following: *if $A \in \mathscr{D}$, $\{Y_i(A)\}$ satisfies Doeblin's condition*; the proof is obvious.

If f is a real-valued function defined on S, f^* is defined on paths by the relation $f^*((x_1, \cdots, x_m)) = f(x_1) + \cdots + f(x_m)$, $m = 1, 2, \cdots$. U is to be the function identically equal to 1 on S. For $n = 1, 2, \cdots$, define L_n (or $L_n(A)$) to be the random variable such that L_n is the biggest w such that $V_w \leqq n$.

The reference in the hypotheses of the following two theorems to *some* A may seem unsatisfactory. This point will be discussed at the end of this section and the results of the next section are also relevant.

Before proceeding to the theorems it will be useful to state a lemma. This lemma will serve in the present context in place of Lemma 7.2, p. 224 of [4]; since the lemma follows easily from Theorem 3 and the argument is similar to the corresponding one in [4] no proof needs to be given. If $\{X_i\}$ satisfies (*) the process is *acyclic* (*cyclic*) if $d = 1$ ($d>1$) in Theorem 3.

LEMMA 5. *Assume $\{X_i\}$ satisfies (*) and has a stationary probability measure Q. Let $w(k)$, $k = 1, 2, \cdots$, be a sequence of positive integers diverging to infinity. Let M be a positive number and for $k = 1, 2, \cdots$, let F_k be a real-valued random variable measurable on $X_{w(k)}, X_{w(k)+1}, \cdots$, such that $|F_k|$ is bounded by M. Let T be the shift operator, i.e., $X_1(\omega) = X_0(T\omega)$, and let $TF_k(\omega) = F_k(T\omega)$, $k = 1, 2, \cdots$.*

If either $\{X_i\}$ is acyclic or for $k = 1, 2, \cdots$, and every $x \in S$

$$\lim_{k\to\infty} E\{F_k - TF_k | X_0 = x\} = 0 .$$

then

$$\lim_{k\to\infty} (E_\varphi\{F_k\} - E_Q\{F_k\}) = 0$$

for every initial probability distribution φ.

Since the process $\{Y_i\}$ is determined by the distribution of X_{V_1} it is natural to indicate the distribution of X_{V_1} by a subscript on the P or E when these operate on sets or random variables measurable on the $\{Y_i\}$ process.

THEOREM 5.1. *Let* $\{X_i\}$ *satisfy* (*) *and have a stationary probability measure* Q. *Let* $A \in \mathscr{D}$, $Q(A) > 0$, $Y_i = Y_i(A)$, $i = 1, 2, \cdots$. *Let* $\delta > 0$, f *a real-valued measurable function on* S. *Let* $E_{Q(A)}\{(f^*(Y_1))^{2+\delta}\} < \infty$, $m = E_{Q_A}\{f^*(Y_1)\}$. *Let* $\bar{f} = f - mQ(A)U$ *and* $\sigma^2 = E_{Q_A}\{(\bar{f}(Y_1))^2\}$. *Let*

$$S_n = \sum_{i=1}^{n} \bar{f}(X_i),\ B_n = \sigma\sqrt{n\cdot Q(A)} .$$

Then the distribution of S_n/B_n *approaches the normal with mean* 0, *variance* 1.

Proof. Let $V_i = V_i(A)$, $L_n = L_n(A)$, $Z_i = \bar{f}^*(Y_i)$, $1, 2, \cdots$. Let $[\alpha]$ denote the largest integer in α for $\alpha > 0$. The argument follows [2] and [8]. In particular the following decomposition is used:

$$\begin{aligned}(5.1)\qquad \frac{S_n}{B_n} = \frac{1}{B_n}\sum_{i=1}^{V_1} f(X_i) + \frac{1}{B_n}\sum_{i=V_{L_n}+1}^{n} f(X_i) + \frac{1}{B_n}\sum_{i=1}^{[nQ(A)]} Z_i \\ + \frac{1}{B_n}\sum_{i=[nQ(A)]+1}^{L_n-1} Z_i + \frac{1}{B_n} mQ(A)(n - V_{L_n} + V_1) .\end{aligned}$$

The distribution of the third term on the right tends to the desired normal by the central limit theorem for Markov chains satisfying Doeblin's condition (see [4], p. 228 or [1]).[3] So it suffices to show that each of the other terms approaches zero in probability. The corresponding facts were shown in [2] and [8], but some new arguments are needed in the present case; on the other hand, much of the following argument is due to [2] and [8].

The first term on the right causes no difficulty. That the second term approaches zero in probability follows from (5.2), which will be proved.

[3] These references consider only the acyclic case, but an easy modification works in the general case.

(5.2) $$\lim_{w\to\infty} P[(n - V_{L_n}) > w] = 0 \text{ uniformly in } n.$$

Suppose first that $\{X_i\}$ is acyclic. By Theorem 3 there exists a function $\delta(w)$ tending to zero as w approaches infinity and such that one has for $h = 0, 1, \cdots$:

(5.3) $$\sum_{k=0}^{h} P[n - V_{L_n} = k] = \sum_{k=0}^{h} \int_A \sum_{j=k+1}^{\infty} P[V_2 = j \mid X_0 = x] P[X_{n-k} \in dx]$$

$$= \sum_{k=0}^{h} \int_A \sum_{j=k+1}^{\infty} P[V_2 = j \mid X_0 = x] Q(dx) + \sum_{k=0}^{h} \delta(n = k)$$

$$= Q(A) \sum_{k=0}^{h} \int_A \sum_{j=k+1}^{\infty} P[V_2 = j \mid X_0 = x] Q_A(dx) + \sum_{k=0}^{h} \delta(n - k) .$$

Now

$$\frac{1}{Q(A)} = Q_A(S) = \sum_{k=0}^{\infty} \int_A \sum_{j=k+1}^{\infty} P[V_2 = j \mid X_0 = x] Q_A(dx) .$$

Let $\varepsilon > 0$; there then exist h_ε such that the first term of the last member of (5.3) exceeds $1 - \varepsilon/2$ for $h = h_\varepsilon$. Choose n_ε so that

$$\sum_{k=0}^{h_\varepsilon} \delta(n - k) < \frac{\varepsilon}{2}$$

for $n \geqq n_\varepsilon$. Then

$$\sum_{k=0}^{h_\varepsilon} P[n - V_{L_n} = k] > 1 - \varepsilon$$

for $n > n_\varepsilon$. Clearly one can find h'_ε such that

$$\sum_{k=0}^{h'_\varepsilon} P[n - V_{L_n} = k] > 1 - \varepsilon$$

for all n. So for $h \geqq h'_\varepsilon$ $P[n - V_{L_n} > h] < \varepsilon$ for all n, as had to be shown. If $\{X_i\}$ is cyclic a simple variation of the above argument can be used to prove (5.2) provided A is included in one of the cyclic classes. Clearly (5.2) for arbitrary A with positive A-measure follows.

Obviously (5.2) also shows that the last term in (5.1) approaches zero in probability.

Note that

$$\frac{n}{L_n - 1} = \frac{V_{L_n} - V_1}{L_n - 1} + \frac{n - V_{L_n} + V_1}{L_n - 1}$$

$$= \frac{\sum_{i=1}^{L_n-1} U^*(Y_i)}{L_n - 1} + \frac{n - V_{L_n} + V_1}{L_n - 1} + \frac{V_1}{L_n - 1} .$$

In the last member the middle term approaches zero in probability by (5.2), and the last term obviously tends to zero in probability. So the law of large numbers, valid for processes satisfying Doeblin's condition (see [1] or [4]),[3] applied to the first term leads to

$$\text{(5.4)} \qquad \frac{n}{L_n - 1} \text{ approaches } \frac{1}{Q(A)} \text{ in probability as } n \to \infty .$$

Write $[\alpha]$ for the largest integer in α for $\alpha > 0$ and define

$$\nu(n) = [Q(A)n(1 + \varepsilon)] \text{ and } \lambda(n) = [Q(A)n(1 - \varepsilon)] .$$

Let $\varepsilon > 0$. (5.4) shows that there must be an n_0 such that $P[\nu(n) \leqq L_n - 1 \leqq \lambda(n)] > 1 - \varepsilon$ for $n > n_0$. Thus to show the fourth term on the right in (5.1) approaches zero it needs only to be shown that

$$\text{(5.5)} \qquad \frac{1}{B_n} \max_{\nu(n) \leqq s \leqq \lambda(n)} \left| \sum_{i=\nu}^{s} Z_i \right| \text{ approaches zero in probability.}$$

To prove (5.5) assume temporarily that $\{Y_i\}$ is stationary, which will make $\{Z_i\}$ stationary. Then (5.5) is equivalent to

$$\text{(5.6)} \qquad \frac{1}{B_n} \max_{0 < s < \lambda(n) - \nu(n)} \left| \sum_{i=1}^{s} Z_i \right| \text{ approaches zero in probability.}$$

The expression in (5.6) equals

$$\frac{\sqrt{\lambda(n) - \nu(n)}}{B_n} \left\{ \max_{0 < s < \lambda(n) - \nu(n)} \frac{1}{\sqrt{\lambda(n) - \nu(n)}} \left| \sum_{i=1}^{s} Z_i \right| \right\} .$$

The distribution in the expression in braces approaches a limiting distribution by the Erdös-Kac-Donsker invariance principle, which is applicable here by [1],[3] and the corresponding fact for independent identically distributed random variables with normal distributions of mean 0. Since the quantity preceding the braces approaches zero (5.6), and hence (5.5), holds in this case. That (5.5) holds for any initial distribution follows from Lemma 5. So the theorem is proved.

In [1] Billingsley showed that the invariance principle of Erdös, Kac, and Donsker is applicable to certain sequences of dependent random variables. The following theorem extends these results to processes satisfying (*). The terminology is that of [1].

[3] These references consider only the acyclic case, but an easy modification works in the general case.

THEOREM 5.2.[4] *Under the conditions of Theorem* 5.1 *the invariance principle holds for the sequence* $\{S_n\}$ *with norming factors* $\{B_n\}$.

Proof. As in the proof of Theorem 4.1 of [1] it suffices to verify two conditions, (i) and (ii).

The verification of (i) in the present case is reduced to verifying the corresponding condition in the case where Doeblin's condition is satisfied in the same manner that the central limit theorem, Theorem 5.1, was reduced to the central limit theorem for processes satisfying Doeblin's condition. When Doeblin's condition is satisfied the argument of [1] applies.[3]

Verification of (ii) is carried out as in [1], except that Lemma 5 is used in place of Lemma 7.2, p. 224 of [4].[3] The fact that

$$\lim_{n\to\infty} E\left\{\frac{S_n^2}{B_n^2}\right\} \text{ exists and is finite}$$

is also needed; this is easily reduced to Lemma 7.3, p. 224 of [4] by using the decomposition (5.1) and the fact, proved above, that in the right member of (5.1) all terms other than the third one approach zero in probability.

As remarked above the hypotheses of Theorems 5.1, 5.2, have the unsatisfactory feature that they refer to some $A \in \mathscr{B}$. In [2] there are analogous hypotheses referring to some state of the denumerable state space; there, however, the hypotheses are proved invariant in the sense that if they hold for some state they hold for each state. In the present situation there exists no similar invariance. Indeed, it is very simple to give examples of a Markov process satisfying (*) and Doeblin's condition and of a function f such that the conditions of Theorem 5.1 are true for some $A \in \mathscr{B}$ but not for $A = S$.[5] Such examples show also that even when dealing with processes satisfying Doeblin's condition the theorems above may be applicable when the result of [1] is not.

Though the existence of moments of random variables of the form $f^*(Y_1(A))$ or $g(Y_1(A))$ does depend on the choice of A certain facts can be established. This is the subject of the following section.

6. Relations between path processes. In this section $\{X_i\}$ will satisfy (*), Q will be the stationary measure, $A, D \in \mathscr{B}$, $D \subseteqq A$, $0 < Q(D) < \infty$, and g will be a positive, real-valued, measurable function on state space.[6]

[3] These references consider only the acyclic case, but an easy modification works in the general case.

[4] It is clear that in the special case where S is denumerable and A has only one point as member the conditions $\delta > 0$ may be dropped, i.e., δ may be zero.

[5] Example 3 of [**2**] illustrates this.

[6] The condition that g be positive can be relaxed. See however footnote 7.

Let

$$G_1 = g(Y_1(A))$$

and for $i = 2, 3, \cdots$,

$$G_i = g(Y_i(A)) \text{ if } X_{V_j} \in A - D \text{ for } j = 2, 3, \cdots, i = 0 \text{ otherwise.}$$

If $\bar{x} = (x_1, \cdots, x_v), \bar{y} = (y_1, \cdots, y_w)$ are two paths $\bar{x} +' \bar{y}$ is to stand for the path $(x_1, \cdots, x_v, y_1, \cdots, y_w)$.

The conventions concerning measures appearing as subscripts made § 5 will be used here. For example, $E_{Q_D}\{G_i\}$ is the expected value of G_i when X_{V_1} has the measure Q_D associated with it. In case a probability measure concentrates all its weight on some point it will be convenient to use this point as a subscript; e.g., $E_x(G_i)$ is meaningful when $x \in A$.

LEMMA 6.[7]

$$E_{Q_D}\left\{\sum_{i=1}^{\infty} G_i\right\} = \int_A E_y\{G_1\}\overline{Q}_D(dy) \; (= E_{Q_A}\{G_1\}\overline{Q}_D(A) \text{ if } Q(A) < \infty) \, .$$

If $A = S$ *and* $g = f^*$ (1.4) *results.*

Proof.

$$\begin{aligned} E_{Q_D}\left\{\sum_{i=1}^{\infty} G_i\right\} &= \sum_{i=1}^{\infty} E_{Q_D}\{E\{G_i \mid X_0\}\} = \sum_{i=1}^{\infty}\int_D E_x\{G_i\} Q_D(dx) \\ &= \int_D E_x\{G_i\}Q_D(dx) + \int_D \Big(\int_{A-D} E_y\{G_1\}P_A(x, dy) \\ &\quad + \int_{A-D}\sum_{i=2}^{\infty}\int_{A-D} P_A(x, dy_1)\int_{A-D} P_A(y_1, dy_2)\cdots \\ &\quad \cdot \int_{A-D} P(y_{i-2}, dy_{i-1})E_y\{G_1\}P_A(y_{i-1}, dy)\Big) = \int_A E_y\{G_1\}\overline{Q}_D(dy) \, , \end{aligned}$$

the last equality following from (1.1), (1.3) with A for S, therefore P_A for P, and D for A. By [7] $\overline{Q}_A$ and $\overline{Q}_D$ differ only by a constant factor, if $Q(A) < \infty$. Then $\overline{Q}_A(A) = 1$ and $\overline{Q}_A = \overline{Q}_D \cdot Q_D(A)$. The equation in parenthesis follows.

If $g = f^*$ and $A = S$ one has

$$\int_S E_y\{G_1\}\overline{Q}_D(dy) = \int_S \left\{\int_S f(x)P(y, dx)\right\}\overline{Q}_D(dy)$$

[7] This lemma may be considered a generalization of (A) of the appendix to [**2**]. As in [**2**] the condition that g be positive can be weakened. Chung showed in Example 3 of [**2**] that even the special case (A) is false if no condition on g is assumed.

$$= \int_{x \in S} \left\{ f(x) \int_{y \in S} P(y, dx) \overline{Q}_D(dy) \right\} = \int_S f(x) \overline{Q}_D(dx) \, .$$

THEOREM 6.1. (a) *If* $g(\bar{x}) + g(\bar{y}) \leqq g(\bar{x} +' \bar{y})$ *for all paths* $\bar{x}$ *and* $\bar{y}$, *then*

$$\sum_{i=1}^{\infty} G_i < g(Y_1(D)) \text{ and } E_{Q_D}\{f(Y_1(D))\} \geqq \int_A E_y\{f(Y_1(A))\} \overline{Q}_D(dy)$$
$$(= E_{Q_A}\{f(Y_1(A))\} \overline{Q}_D(A) \text{ if } Q(A) < \infty) \, .$$

(b) *If* $g(\bar{x}) + g(\bar{y}) \geqq g(\bar{x} +' \bar{y})$ *for all paths* $\bar{x}$ *and* $\bar{y}$, *then*

$$\sum_{i=1}^{\infty} G_i \geqq f(Y_1(D)) \text{ and } E_{Q_D}\{f(Y_1(D))\} \leqq \int_A E_y\{f(Y_1(A))\} \overline{Q}_D(dy)$$
$$(= E_{Q_D}\{f(Y_1(A))\} \overline{Q}_D(A) \text{ if } Q(A) < \infty) \, .$$

Proof. The theorem follows immediately from Lemma 6.1.

Note that for $p \geqq 1$ $(0 < p \leqq 1)$ the function $f(\bar{x}) = (U^*(\bar{x}))^p$ satisfies condition (a) (condition (b)) of the theorem. Call $U^*(Y_1(B))$ the *recurrence time* to B when $B \in \mathscr{B}$; and if furthermore $Q(B) < \infty$, $p \geqq 0$ call $E_{Q_D}\{(U^*(Y_1(B)))^p\}$ the pth *stationary moment of the recurrence time.* It follows that when $Q(S) = 1$ $(Q(S) = \infty)$, $p > 0$, and $B \in \mathscr{B}$ ($B \in \mathscr{B}$ and $(B)) < \infty$), the pth stationary moment of the recurrence time to B is finite only if the same is true for every Borel superset (subset of positive Q-measure).

The hypotheses of parts (a) and (b) of Theorem 6.1 cannot both be satisfied by the same g unless g is a constant multiple of U^*. The theorem below, on the other hand, is such that for a wide class of functions both the hypotheses of (a) and (b) may be satisfied.

With reference to the hypotheses in Theorem 6.2 observe that if $B \in \mathscr{B}$ the three statements

(a) $E_x\{g(Y_1(B))\}$ is uniformly bounded for all $x \in B$,

(b) $E_\varphi\{g(Y_1(B))\}$ is uniformly bounded for all probability measures φ on B,

(c) $E_\varphi\{g(Y_1(B))\} < \infty$ for all probability measures φ on B,

are all equivalent.

THEOREM 6.2. *Let* $A \in \mathscr{D}$.

(a) *Suppose* (i) $c > 0$ *and* $f(\bar{x} +' \bar{y}) \leqq c(f(\bar{x}) + f(\bar{y}))$ *for all paths* $\bar{x}$ *and* $\bar{y}$, *and* (ii) $M > 0$ *and* $E_x\{f(Y_1(A))\} < M$ *for all* $x \in A$. *Then* $E_x\{f(Y_1(D))\}$ *is uniformly bounded for all* $x \in D$.

(b) *Suppose* (i) $c > 0$ *and* $f(\bar{x}) + f(\bar{y}) \leqq cf(\bar{x} +' \bar{y})$ *for all paths* $\bar{x}$ *and* $\bar{y}$, *and* (ii) $M > 0$ *and* $E_x\{f(Y_1(D))\} \leqq M$ *for all* $x \in D$. *Then* $E_{Q_A}\{f(Y_1(A))\} < \infty$.

Proof. Assume (i) and (ii) of (a) and let $x \in D$. Let ν be the first n such that $n \geqq 2$ and $X_{V_n(A)} \in D$, so that $G_i = 0$ for $i > \nu$. Using (i) of (a) repeatedly, and (ii) of (a), one has,

$$E_x\{f(Y_1(D))\} = \sum_{k=1}^{\infty} E_x\{f(Y_1(A) + 'Y_2(A) + \cdots + 'Y_k(A) \mid \nu = k\} \; P[\nu = k]$$

$$\leqq \sum_{k=1}^{\infty} c^{[\log_2 k]+1} E_x\left\{\sum_{i=0}^{k} G_i \mid \nu = k\right\} P[\nu = k] \leqq \sum_{k=1}^{\infty} c^{[\log_2 k]+1} kMP[\nu = k] \; .$$

Since $A \in \mathscr{D}$, $P[\nu = k]$ decreases exponentially and (a) follows.

Assume (i) and (ii) of (b). Let ν have the same significance as above. One has then, by Lemma 6 and repeated applications of (i) of (b),

$$\sum_{k=1}^{\infty} E_{Q_D}\{f(Y_1(A) +' \cdots + 'Y_k(A)) \mid \nu = k\} P[\nu = k] c^{[\log_2 k]+1}$$

$$\geqq E_{Q_D}\left\{\sum_{i=1}^{k} G_i \mid \nu = k\right\} P[\nu = k] = E_{Q_D}\left\{\sum_{i=1}^{\infty} G_i\right\} = E_{Q_D}\{f(Y_1(A))\} \bar{Q}_D(A) \; .$$

Note that when $\nu = k$ $Y_1(D) = Y_1(A) + '\cdots + 'Y_k(A)$; so assumption (ii) of (b) ensures that in the inequality above each of the expectations in the first member is at most M. Since $A \in \mathscr{D}$ $P[\nu = k]$ decreases exponentially. This proves (b).

As an application consider the following situation: there exist Borel sets D and C each of finite positive Q-measure and each containing only one point. Let A be the union of D and C. If g satisfies (i) of (a) in Theorem 6.2 and (i) of (b) in the same theorem, one has $E_{Q_D}\{g(Y_1(D))\} < \infty$ implies $E_{Q_A}\{g(Y_1(A))\} < \infty$, $E_{Q_A}\{g(Y_1(A))\} < \infty$ implies $E_{Q_C}\{g(Y_1(C))\} < \infty$, since (ii) of (a) and (b) are now automatically true. In particular, g will always satisfy (i) of (a) and (i) of (b) if $g(\bar{x}) = ((U^*(\bar{x}))^p$, $p > 0$. This gives again the result of Chung [2] that for two points, each of finite positive Q-measure, the pth moment of the recurrence time exists for both or neither.

REFERENCES

1. P. Billingsley, *The invariance principle for dependent random variables*, Trans. Amer. Math. Soc., **83** (1956), 250-268.
2. K. L. Chung, *Contributions to the theory of Markov chains* Trans. Amer. Math. Soc., **76** (1954), 397-419.
3. J. Doob, *Asymptotic properties of Markoff transition probabilities*, Trans. Amer. Math. Soc., **63** (1948), 393-438.
4. ———, *Stochastic processes*, New York, 1953.
5. P. Erdös, W. Feller, and H. Pollard, *A theorem on power series*, Bull. Amer. Math. Soc., **55** (1949), 201-204.
6. T. E. Harris *Recurrent Markov processes*, II (abstract). Ann. Math. Stat., **26** (1955), 152-153.
7. ———, *The existence of stationary measures for certain Markov processes*, Third

Berkeley Symposium on Mathematical Statistics and Probability, vol. II, Berkeley, 1956, 113-124.

8. D. G. Kendall, *A note on Doeblin's central limit theorem*, Proc. Amer. Math. Soc., **8** (1957), 1037-1039.

UNIVERSITY OF MINNESOTA AND
UNIVERSITY OF CALIFORNIA

ON QUADRUPLY TRANSITIVE GROUPS

E. T. PARKER

1. Introduction. A major unsolved problem in the theory of groups is whether there are any quadruply transitive finite groups other than the alternating and symmetric groups and the four Mathieu groups of degrees 11, 12, 23, and 24, respectively.

In this paper are proved two theorems which impose arithmetic restrictions on primes dividing the order of the subgroup fixing four letters of a finite quadruply transitive group, and on the degrees of Sylow subgroups thereof.

THEOREM 1 is stated, followed by a corollary, which is somewhat less general but of a more direct arithmetic form.

THEOREM 1. *If G is a quadruply transitive finite permutation group, H is the subgroup of G fixing four letters, P is a Sylow p-subgroup of H, P fixes $r \geqq 12$ letters and the normalizer in G of P has component A_r or S_r permuting the letters fixed by P, and P has no component of degree $\geqq p^3$ and no set of $r(r-1)/2$ permutation-isomorphic components, then G is alternating or symmetric.*

COROLLARY. *If G is a quadruply transitive permutation group of degree $n = kp + r$, with p prime, $k < p^2$, $k < r(r-1)/2$, $r \geqq 12$, and the subgroup of G fixing four letters has a Sylow p-subgroup P of degree kp, and the normalizer in G of P has component A_r or S_r permuting the letters fixed by P, then G is A_n or S_n.*

This corollary is a partial generalization of a theorem of G. A. Miller [6], which may be paraphrased to read like the above with "quadruply transitive" replaced by "primitive" and the inequalities replaced by "$k < p, k < r, r \geqq 5$" – "$r \geqq 3$" if the component of the normalizer of P is restricted to be S_r. Miller's theorem is proved for $r \geqq 5$ by showing first that the component A_r of the normalizer of P – or a subgroup of P of index 2 – splits off as a direct factor. The argument is completed by invoking the theorem of Netto [3, p. 207, Th. I] on primitive groups with primitive subgroups of lower degree. The proof of Theorem 1 makes use of the techniques in Miller's theorem; and in addition results on the structure of the automorphism groups of noncyclic groups of order p^2, on distribution of primes, and in particular

Received September 17, 1958. This paper is based on the author's dissertation, written at the Ohio State University in 1957, under the supervision of Professor Marshall Hall, Jr. The author expresses his gratitude to Professor Hall for guidance and encouragement.

a theorem of Bochert [1] giving a lower bound on the degrees of permutations in quadruply transitive groups.

A consequence of Miller's theorem, not mentioned in his paper, is that for infinitely many degrees, namely $p + 3$ and $2p + 3$ with p prime, quadruply transitive groups can be only the alternating or symmetric groups.

THEOREM 2. *If G is a quadruply transitive permutation group of degree n, and the subgroup of G fixing four letters is of order divisible by an odd prime p, with $5p > n - 4, 4p \neq n - 4$, then G is A_n or S_n.*

Some comments on terminology are in order: A_n and S_n designate respectively the alternating group and the symmetric group of degree n. A *component* of a permutation group is the permutation group induced on a transitive set of letters. "The subgroup fixing four letters" of a quadruply transitive group is the *largest* subgroup (unique to conjugacy) fixing four letters *individually*. (The phrase "fixed set-wise" is used explicitly when appropriate.)

Included in the author's dissertation is a self-contained proof that the only quadruply transitive groups of degrees $n \leqq 27$ are the alternating and symmetric groups, $6 \leqq n \leqq 27$ and $4 \leqq n \leqq 27$ respectively, and the Mathieu groups of degrees 11, 12, 23, and 24. As this result is in the literature (except perhaps for degree 27), though likely in no single source, these rather lengthy arguments are not included in this paper.

Section 2 contains three lemmas needed to prove Theorem 1. In §§ 3 and 4 are the proofs respectively of Theorem 1 and its corollary, and Theorem 2.

2. In this section are three lemmas.

LEMMA 1. *If B is a transitive permutation group of degree p or p^2 (p prime), and if B has a normal p-subgroup, then B has no composition factor[1] isomorphic with A_r, $r > 5$.*

Proof. If B is of degree p, then B is the metacyclic[2] group or a subgroup thereof, and hence is solvable.

If B is of degree p^2, then the normal p-subgroup T has an elementary subgroup (not necessarily proper) C normal in B. (For p-groups are solvable, and every minimal normal subgroup is a direct product of isomorphic simple groups.) Since B is transitive, C displaces all p^2

[1] A factor group of a composition series.

[2] The holomorph of the group of order p.

letters, and has all components permutation-isomorphic. C being abelian has all components regular. Thus C is either the regular elementary group of order p^2, or a subdirect product of p p-cycles.

If C is regular, then B is in the holomorph of the elementary group of order p^2. The only unsolvable [3, pp. 428-34] composition factor of this holomorph is $LF(2, p)$. The smallest alternating group of order divisible by p is A_p. $LF(2, p)$ is isomorphic with no alternating group whenever $p > 5$, since $p(p^2 - 1)/2 < p!/2$. The only unsolvable proper subgroup of $LF(2, p)$ is isomorphic with A_5 — $LF(2, p)$ is isomorphic with A_5 for $p = 5$, and has such a subgroup for $p \equiv \pm 1 \pmod 5$ [3, pp. 440-50].

If C is a subdirect product of p p-cycles, then each element of B must permute the transitive sets of C among themselves; these are sets of imprimitivity for B. Let K be the largest subgroup of B fixing the sets of imprimitivity; K is a normal subgroup of B. B/K is a permutation group on the transitive sets of K. Since T is transitive and K is intransitive, T is not a subgroup of K. Hence B/K has a normal p-subgroup. B/K of degree p is in the metacyclic group, and is therefore solvable. K is a subdirect product of metacyclic groups. Thus B is solvable.

LEMMA 2. *For $r \geqq 9$, A_r has no subgroup of index t, $r < t < r(r-1)/2$.*

Proof. Assume that A_r has a subgroup L of index t (r, t as above).

If L is intransitive, then L is in the group M_i of even permutations in the direct product of S_i and S_{r-i}, $0 < i < r$. M_i is of index $\binom{r}{i}$ in A_r. $\binom{r}{i} \geqq r(r-1)/2$ unless $i = 1$ or $r - 1$. For $i = 1$ or $r - 1$, M_i is A_{r-1}, of index r which fails to satisfy the strict inequality. A_{r-1} has no proper subgroup of index $< r - 1$; hence M_i has no proper subgroup of index $< r(r-1)/2$ in A_r.

There remains for consideration the case of L transitive.

Let q be a prime in the range $r/2 < q \leqq r$. If L is of order divisible by q, then an element of L is a q-cycle, and L is primitive [4, p. 162, Exercise 8]. If further $q \leqq r - 3$, then L is A_r [6]. Transitive L fulfilling the assumption must be of index in A_r divisible by each prime q, $r/2 < q \leqq r - 3$.

A theorem on distribution of primes will now be used [2, 7]: *If $x \geqq 25$, then there exists a prime q such that $x < q < 6x/5$.* A computation shows that for all $r \geqq 50$ there exist primes q_1, q_2, q_3 satisfying $r/2 < q_1 < q_2 < q_3 \leqq r - 3$. The existence of a triple of primes for each r in the range $20 \leqq r < 50$ is verified by inspection. For any $r \geqq 20$, $q_1 q_2 q_3 > (r/2)^3 > r(r-1)/2$.

Degrees $9 \leqq r \leqq 19$ remain to be considered. A primitive proper subgroup of A_r has no element a q-cycle, with prime $q \leqq r - 3$. If L is primitive, then for each odd prime $q \leqq r - 3$ a Sylow q-subgroup of L must be a proper subgroup of that of A_r. Thus the index of L in A_r must be divisible by each prime q, $2 < q \leqq r - 3$. This inequality is satisfied by primes $q = 3, 5, 7, 11$ when $14 \leqq r \leqq 19$; and $3 \cdot 5 \cdot 7 \cdot 11 > r(r-1)/2$ for these values of r. Similarly $q = 3, 5, 7$ for $10 \leqq r \leqq 13$, and $3 \cdot 5 \cdot 7 > r(r-1)/2$. Primitive L does not exist for $10 \leqq r \leqq 19$.

Imprimitive L must be of index divisible by each prime q, $r/2 < q \leqq r$. For $17 \leqq r \leqq 19$, $q = 11, 13, 17$, and $11 \cdot 13 \cdot 17 > r(r-1)/2$. For $13 \leqq r \leqq 16$, $11 \cdot 13 > r(r-1)/2$. For $r = 11$ or 12, $7 \cdot 11 > r(r-1)/2$. The maximal imprimitive subgroups of A_{10} are of orders $\frac{1}{2} \cdot 2!(5!)^2$ and $\frac{1}{2} \cdot 5!(2!)^5$, both of index $> 10 \cdot 9/2$. Such a subgroup of A_9 has order $\frac{1}{2} \cdot 3!(3!)^3$; index $> 9 \cdot 8/2$.

One case remains, namely degree 9 with L primitive. L must be of index divisible by 3 and 5. Since $3 \cdot 3 \cdot 5 > 9 \cdot 8/2$, L can be of index only 15 or 30 in A_9. L, having a 7-cycle, is triply transitive. If L is of index 30 in A_9, then the subgroup of L fixing two letters has order 84; a group of order 84 has only one Sylow 7-subgroup. If L is of index 15, then the largest subgroup of L fixing two letters set-wise is of order 336; but a group of degree 7 and order divisible by 2^4 contains a transposition.

LEMMA 3. *If G is a transitive permutation group homomorphic onto K, and the kernel of the homomorphism is transitive on the letters permuted by G, then the subgroup of G fixing one letter is homomorphic onto K. Moreover, the two homomorphisms belong to the same many-to-one mapping.*

Proof. Let G_1 be the subgroup of G fixing the letter 1. For any $k \in K$, there exists $g_k \in G$ such that $g_k \to k$ in the homomorphism. Let i be the letter onto which 1 is mapped by g_k. Being transitive, the kernel has an element g'_k mapping i onto 1. Then $g_k g'_k$ maps 1 onto 1, and corresponds to k in the homomorphism. Since k is an arbitrary element of K, it follows that G_1 has an element corresponding to any element of K.

3. In this section will be established.

THEOREM 1. *If G is a quadruply transitive finite permutation* *group, H is the subgroup of G fixing four letters, P is a Sylow p-subgroup of H, P fixes $r \geqq 12$ letters and the normalizer in G of P has component A_r or S_r permuting the letters fixed by P, and P has no component of degree $\geqq p^3$ and no set of $r(r-1)/2$ permutation-isomorphic components, then G is alternating or symmetric*

Proof. Let $N(P)$ be the normalizer in G of P. Each element of $N(P)$ maps any component of P onto a permutation-isomorphic component. Let P' be the largest subgroup of $N(P)$ fixing set-wise all components of P. P' is a normal subgroup of P. The components of P will be called *points*. $N^* = N(P)/P'$ is a permutation group on points.

By hypothesis all points are of degrees $\leqq p^2$. By Lemma 1 no transitive group of degree p or p^2 with a normal p-subgroup has a composition factor isomorphic with A_r, $r > 5$. By hypothesis $N(P)$ has component A_r or S_r on the $r \geqq 12$ letters fixed by P. Thus $N(P)$, or a subgroup thereof of index 2, is homomorphic with A_r. It follows that N^* has a composition factor of A_r, since P' has none.

N^* has a subgroup N with properties:

1. N has an element permuting the r letters fixed by P according to a, where a is any even permutation.

2. No proper subgroup of N has property 1.

A subgroup (not necessarily proper) of N^* with property 1 exists, since N^* itself has property 1. N^*, having only finitely many subgroups, has a minimal subgroup with property 1. It is not asserted that N is unique.

By the minimality condition on N, each component of N is either a single point or is homomorphic with A_r, the image being represented on the letters fixed by P. Let J designate a component of N permuting more than one point (not letters fixed by P). J is homomorphic with A_r, with kernel J_0 the subgroup of J corresponding to the identity permutation of the letters fixed by P. Assume that the kernel J_0 is transitive on the points permuted by J. Then by Lemma 3 the subgroup of J fixing one point possesses the homomorphism onto A_r; this contradicts the minimality property of N. Thus each J has intransitive kernel.

For any J, the components of the kernel J_0 will be called *blocks*. Since J_0 is intransitive on the points permuted transitively by J, it follows that J permutes more than one block. In fact, J permutes blocks according to a group isomorphic with A_r, since J_0 is the kernel of the homomorphism of J onto A_r. By hypothesis P has no set of $r(r-1)/2$ permutation-isomorphic components; thus each J permutes fewer than $r(r-1)/2$ points, and *a fortiori* fewer than this number of blocks. By

Lemma 2 A_r for $r \geqq 9$ has no transitive permutation representation of degree strictly between r and $r(r-1)/2$ (by hypothesis $r \geqq 12$). Accordingly each J permutes exactly r blocks according to A_r.

For $r \neq 6$ the only automorphisms [3, p. 209] of A_r are conjugations by elements of S_r. Thus there is a natural one-to-one correspondence between the letters fixed by P and the blocks permuted by any J such that each element of N permutes the sets alike.

Select a set of $s = [r/2]$ letters among the r fixed by P. Let N_0 be a minimal subgroup of N inducing all even permutations on these s letters. (That N_0 exists is argued as for N.) As each J is of degree $< r(r-1)/2$, each block contains fewer than $(r-1)/2$ points, hence fewer than s. The points of a block cannot be permuted according to any group homomorphic with A_s. By the minimality of N_0, each block fixed set-wise by N_0 is fixed point-wise by N_0. This is the case because N_0 has a composition factor of A_s, while the group permuting the points of one block has none.

N_0 is a group of permutations of points (transitive sets of P') and letters fixed by P. Thus N_0 determines a subgroup M of $N(P)$ such that $N_0 = M/P'$. M permutes the chosen set of s letters fixed by P according to A_s. Let M_0 be a minimal subgroup of M with this property. Each component of P' (point) is transitive of degree p or p^2 and has a normal p-subgroup. Hence by Lemma 1, no component of M_0 containing a single component of P' has a composition factor of A_s ($s > 5$, since $r \geqq 12$). Since M_0 is a minimal group homomorphic with A_s, each component of P' fixed set-wise by M_0 is fixed letter-wise by M_0. Thus each element of M_0 displaces at most $s/r \leqq 1/2$ of the letters of any component of $N(P)$; that is, at most half the letters displaced by P. M_0 has an element m displacing exactly three letters fixed by P. As $r \geqq 12$, m displaces at most half as many letters as the degree of G, diminished by 3. The theorem of Bochert [1] asserts that quadruply transitive G, with a non-identical element displacing so few letters as does m, is alternating or symmetric.

COROLLARY. *If G is a quadruply transitive permutation group of degree $n = kp + r$, with p prime, $k < p^2$, $k < r(r-1)/2$, $r \geqq 12$, and the subgroup of G fixing four letters has a Sylow p-subgroup P of degree kp, and the normalizer in G of P has component A_r or S_r permuting the letters fixed by P, then G is A_n or S_n.*

Proof. Since $k < p^2$, P is of degree $kp < p^3$, so that P has no component of degree $\geqq p^3$. Since components of P are of degree at least p, the hypothesis $k < r(r-1)/2$ implies that P has no set of $r(r-1)/2$ permutation-isomorphic components. The hypothesis of Theorem

1 is fulfilled, so that G is A_n or S_n.

4. THEOREM 2. *If G is a quadruply transitive permutation group of degree n, and the subgroup of G fixing four letters is of order divisible by an odd prime p, with $5p > n - 4, 4p \neq n - 4$, then G is A_n or S_n.*

Proof. Let H be the subgroup of G fixing four letters, and P be a Sylow p-subgroup of H. Since $5p > n - 4$, P is of degree $kp \leqq 4p$.

If $kp = n - 4$, then $k \leqq 3$, since $4p \neq n - 4$ by hypothesis. The theorem of Miller [6] applies with $n = kp + 4$ except for $k = p = 3$. However, $13 = 2 \cdot 5 + 3$, so that Miller's theorem is applicable to this case.

If $kp < n - 4$, then at least five letters are fixed by P. $N_G(P)$, the normalizer in G of P, has component T permuting the letters fixed by P quadruply transitive [5, Lemma 2.2]. A normal subgroup (not the identity) of a quadruply transitive group (other than S_4) is triply transitive [3, p. 198, Th. XI]. Thus T is primitive and has no regular normal subgroup, hence is unsolvable. The final subgroup in the commutator series of T is triply transitive.

P has at most four transitive sets, which cannot be permuted by an unsolvable group. If $p > 3$, then each component of P is of degree p and has a solvable holomorph. For $p = 3$, a component of degree 9 has a solvable automorphism group. Accordingly, the final member of the commutator series of $N_G(P)$ is a primitive group permuting only letters fixed by P. By Bochert's theorem [1], P is of degree at least $(n - 2)/2$, so that at most $(n + 2)/2$ letters are fixed by P. Hence G is at least $n + 1 - (n + 2)/2 = n/2 -$ ply transitive [3, p. 207, Th. I.] But G can be at most $(n + 3)/3 -$ ply transitive, or contains A_n. [4, p. 148, Th. VI.] The theorem is established for $n > 6$, and is trivial for the smaller degrees.

REFERENCES

1. Alfred Bochert, *Ueber die Classe der Transitiven Substitutionen-gruppen*, Math. Annalen **40** (1892), 176-193.

2. Robert Breusch, *Zur Verallgemeinerung des Bertrandschen Postulates, dass zwischen x und 2x stets Primzahlen liegen*, Math. Zeits., **34** (1932), 505-526.

3. W. Burnside, *Theory of Groups of Finite Order*, 2nd ed., Cambridge Univ. Press, 1911; reprinted Dover, New York, 1955.

4. Robert D. Carmichael, *Introduction to the Theory of Groups of Finite Order*, Ginn, New York, 1937.

5. Marshall Hall, Jr., *On a theorem of Jordan*, Pacific J. Math. **4** (1954), 219-226.

6. G. A. Miller, *Limits of the degree of transitivity of substitution groups*, Bull. Amer.

Math. Soc. **22** (1915), 68-71. See also Marshall Hall, Jr., *The Theory of Groups*, Macmillan, New York, 1959, pp. 68-72.
7. J. Nagura, *On the interval containing at least one prime number*, Proc. Japan Acad. **28** (1952), 177-181.

REMINGTON RAND UNIVAC
DIVISION OF SPERRY RAND CORPORATION
ST. PAUL, MINNESOTA

ON TOEPLITZ MATRICES, ABSOLUTE CONTINUITY, AND UNITARY EQUIVALENCE

C. R. PUTNAM

1. Preliminaries. For $n = 0, \pm 1, \pm 2, \cdots$, let $\{c_n\}$ be real numbers satisfying

$$(1) \qquad c_0 = 0, c_{-n} = c_n \text{ and } \sum_1^\infty c_n^2 < \infty ,$$

and consider the associated real-valued, even function $f(\theta)$ of period 2 and of class $L^2[0, \pi]$ defined by

$$(2) \qquad f(\theta) \sim \sum_{-\infty}^{\infty} c_n e^{in\theta} = 2 \sum_1^\infty c_n \cos n\theta .$$

(Throughout this paper it will be assumed for the sake of convenience that $c_0 = 0$. If $c_0 \neq 0$, T (see below) is modified merely by the addition of a multiple of the unit matrix.) Let $A = (a_{ij})$, where $a_{ij} = c_{i-j}(= c_{j-i})$ or $a_{ij} = 0$ according as $i < j$ or $i \geq j(i, j = 1, 2, \cdots)$, and define the Toeplitz matrix T and the Hankel matrices H and K by

$$(3) \qquad T = (c_{i-j}) = A + A^*, \ H = (c_{i+j-1}) \text{ and } K = (c_{i+j}) .$$

The matrices T, H and K are real and Hermitian (symmetric).

Let J denote the matrix belonging to the quadratic form $2 \sum_1^\infty x_n x_{n+1}$. The differential of its spectral matrix is given by $d\rho_{ij}(\theta) = 2\pi^{-1} \sin i\theta \sin j\theta d\theta$ (cf. Hilbert [5], p. 155, Hellinger [8], pp. 148 ff.). A direct calculation (cf. [11], Appendix 2) shows that

$$(4) \qquad T = F + K ,$$

where T and K are defined by (3), and F is given by

$$(5) \qquad F = \left(\int_0^\pi f(\theta) d\rho_{ij}(\theta) \right) ,$$

with $f(\theta)$ defined by (2) and (1). In particular, if $c_1 = 1$ and $c_n = 0$ for $n > 1$, then $f(\theta) = 2 \cos\theta$ and (5) is the spectral resolution of J (with the usual parameter λ being given by $\lambda = 2\cos\theta$).

It should be noted that the L^2 assumption on the sequence $\{c_n\}$ in (1) does not imply the boundedness of the various matrices considered above, although of course, the existence, in the mean, of the integrals in (5) is assured. Moreover, all two factor products of the type A^2, AA^*, etc. surely exist and it can be verified that

Received August 26, 1958. This research was supported by the United States Air Force through the Air Force Office of Scientific Research of the Air Research and Development Command, under Contract No. AF 18 (603)-139. Reproduction in whole or in part is permitted for any purpose of the United States Government.

$$AT - TA = H^2 , \tag{6}$$

where A, T and H are defined by (3); cf. [11], p. 517.

It is known (Toeplitz) that T is bounded if and only if $f(\theta)$ is essentially bounded, so that $|f(\theta)| \leq$ const. almost everywhere on $[0, \pi]$; [3], p. 360. Moreover, if M and m denote the essential upper and lower bounds of f, then the spectrum of T consists of the interval $[m, M]$ and, unless all $c_n = 0$, is purely continuous (so that the point spectrum is empty); [3] and [4]. Furthermore, if T is not bounded, but is still self-adjoint, then the spectrum of T is again purely continuous and is again the (unbounded) interval $[m, M]$; [4], p. 878. (Actually the results of Hartman and Wintner mentioned above are not restricted to the case of real sequences $\{c_n\}$ as in the present paper.) For necessary and sufficient conditions in order that a Hankel matrix be bounded, see [9].

The matrix A is bounded if and only if $g(\theta) \sim \sum_1^\infty c_n e^{in\theta}$ is essentially bounded (Toeplitz; cf. [4], p. 880, [11], p. 517). Clearly, if A is bounded, so also is T. In addition, if T is bounded, then $f(\theta)$ of (2) is bounded and consequently K is bounded (Toeplitz; cf. [2], p. 223, also [3], p. 365). In view of the easily verified relation

$$\|Hx\|^2 = \|Kx\|^2 + \left(\sum_1^\infty c_n x_n\right)^2 \tag{7}$$

H is bounded (or completely continuous) if and only if K is bounded (or completely continuous). It is seen that if A is bounded then all of the other matrices considered above are bounded.

In § 2 there will be pointed out a few consequences of the relations (4) and (5) bearing on the nature of Fourier series and the spectra of Toeplitz matrices belonging to real valued, even functions defined by (2). In §§ 3 and 4, sufficient conditions guaranteeing that a Toeplitz matrix be absolutely continuous or unitarily equivalent to a certain function of J, will be obtained. Some applications to Hilbert matrices will be given in § 5.

Condition (1) on the real sequence $\{c_n\}$ will be assumed throughout the paper.

2. Toeplitz matrices and Fourier series.

First there will be proved:

(I) *If the Hankel matrix K is bounded, then necessarily T is self-adjoint.*

This is an obvious consequence of (4) if it is noted that F of (5) is always self-adjoint. Incidentally, it is seen that in this case the domain of T is identical with that of F.

By the essential range of $f(\theta)$ will be meant the (closed) set of values λ for which $|f(\theta) - \lambda| < \varepsilon$ holds on a set of positive measure on $[0, \pi]$ for every $\varepsilon > 0$. Then one has:

(II) *If* K *is bounded and satisfies* $\|K\| \leq C$ $(=\text{const.})$, *then any closed interval of length* $2C$ *contained in* $[m, M]$ *contains at least one point of the essential range of* $f(\theta)$.

This assertion also follows from (4). For, it is clear that the spectrum of F is exactly the essential range of $f(\theta)$ (the situation being analogous to the problem of Toeplitz of determining the spectrum of a Laurent matrix; cf. [3], pp. 359-360 and the references cited there). Consequently, since relation (4) shows that F is obtained as a perturbation of T by the operator $-K$, the assertion now follows from the fact that the spectrum of T is the entire interval $[m, M]$.

A theorem similar to (II) is

(III) *If* K *is completely continuous then the essential range of* $f(\theta)$ *is* $[m, M]$.

The proof follows from the observation that, by (the generalization of) Weyl's theorem, the essential (cluster) spectra of T and F are identical. Since, by (I), T is self-adjoint, its spectrum is $[m, M]$ ([4]) and it follows that the essential spectrum of F is also $[m, M]$. But the spectrum of F is always contained in $[m, M]$ and hence must be identical with this interval. This implies (III).

A curious corollary of (III) is the following assertion:

(IV) *If* $g(\theta) \sim \sum_1^\infty c_n \sin n\theta$ *is continuous on* $-\infty < \theta < \infty$ *then the essential range of* $h(\theta) \sim \sum_1^\infty c_n \cos n\theta$ *is an interval (possibly unbounded).*

The assertion follows from (III) if it is noted that the continuity of $g(\theta)$ implies the complete continuity of the operator K. Cf. [3], p. 365. It is stated there that $H = (c_{i+j-1})$ is completely continuous if either $g(\theta)$ or the function $\frac{1}{2} f(\theta) \sim \sum_1^\infty c_n \cos n\theta$ is continuous on $[0, \pi]$. The proof seems to indicate however that $K = (c_{i+j})$ (or H) is completely continuous if either $g(\theta)$ or $f(\theta)$ is continuous on $[-\pi, \pi]$ (and hence on $-\infty < \theta < \infty$). See [2], p. 223. The continuity of $g(\theta)$ on $[0, \pi]$ implies its continuity on $[-\pi, \pi]$ but the corresponding assertion for $f(\theta)$ is false.

Another easy consequence of (4) is the following:

(V) *If* T_1 *and* T_2 *are two bounded Toeplitz matrices with the representations*

$$(4_m) \qquad T_m = F_m + K_m$$

where

$$(5_m) \qquad F_m = \left(\int_0^\pi f_m(\theta)d\rho_{ij}(\theta)\right) \text{ and } f_m(\theta) \sim 2\sum_1^\infty c_{mn}\cos n\theta \ ,$$

and if K_1 *and* K_2 *are completely continuous, then*

$$(8) \qquad T_1T_2 = \left(\int_0^\pi f_1(\theta)f_2(\theta)d\rho_{ij}(\theta)\right) + C \ ,$$

where C *is completely continuous.*

A relation similar to (8) holds of course for products with more than two factors. The proof of (V) follows easily if it is noted that the product of a bounded operator and a completely continuous operator is completely continuous. In particular, it is seen from (8) that the essential spectrum of T_1T_2 is the essential range of $f_1(\theta)\ f_2(\theta)$. The situation is to be compared with that for Laurent matrices; cf. the remark following (II) above.

3. Absolute continuity. It follows from Theorem 13 of [11], p. 523, that if A is bounded, then (6) implies that T is absolutely continuous whenever 0 is not in the point spectrum of H. That is, this last condition is sufficient in order to guarantee that $\int_Z dE(\lambda) = 0$, where

$$(9) \qquad T = \int \lambda dE(\lambda)$$

is the spectral resolution of T and Z is any set of one-dimensional Lebesgue measure zero. However it is possible that T is absolutely continuous even if 0 does belong to the point spectrum of H. In fact each T_N, belonging to the sequence $\{c_n\}$ with $c_n = c_N \neq 0$ $(N > 0)$ if $n = \pm N$ and $c_n = 0$ otherwise, is absolutely continuous; cf. [11], pp. 519, 524. This result will be generalized in the following theorem:

(VI) *Let the real sequence* $\{c_n\}$, $n = 0, \pm 1, \pm 2, \cdots$ *satisfy* (1) *define the associated matrices* A, T *and* F *as in* § 1, *and suppose that* A *is bounded. Then* T *is absolutely continuous whenever* F *is absolutely continuous.*

As remarked above, the boundedness of A implies that of all other operators considered. It follows from the argument of [10] (cf. p. 1027, formula line (4)) when applied to (6) that

$$(10) \qquad H\int_Z dE(\lambda) = 0 \ ,$$

where Z denotes any set of one-dimensional Lebesgue measure zero and H is defined by (3). (The square root $H^{1/2}$ appearing in [10] *loc. cit.* can clearly be taken to be any self-adjoint square root of the non-negative self-adjoint operator H. The H appearing there corresponds to a positive multiple of the operator H^2 of the present paper.) Next let y be any element in the range of $\int_Z dE(\lambda)$ so that

$$y = \int_Z dE(\lambda)x\ , \qquad \|x\| < \infty\ . \tag{11}$$

Since, by (10), $Hy = 0$, it follows from (7) that $Ky = 0$. Consequently, by (4), $Ty = Fy$. For $n = 0, 1, 2, \cdots$, $T^n y = \int_Z dE(\lambda)T^n x$ is also in the range of $\int_Z dE(\lambda)$, and it follows that $T^{n+1}y = FT^n y$. Hence

$$T^n y = F^n y \quad (n = 0, 1, 2, \cdots;\ T^0 = F^0 = I)\ , \tag{12}$$

where y is defined by (11). But (12) implies $E(\lambda)y = F(\lambda)y$, where

$$F = \int \lambda dF(\lambda) \tag{13}$$

is the spectral resolution of F, and hence $\int_Z dE(\lambda)y = \int_Z dF(\lambda)y$. But, whenever F is absolutely continuous, $\int_Z dF(\lambda) = 0$ and so, by (11), $\int_Z dE(\lambda)x = 0$ for all x. That is, T is absolutely continuous and the proof of (VI) is now complete.

4. Unitary equivalence. It was shown in [11] that each T_N (see the beginning of § 3 above) is absolutely continuous and that moreover T_N is unitarily equivalent to the corresponding $F = F_N$. This result will be considerably refined in the following theorem:

(VII) *Let the real sequence* $\{c_n\}$ *satisfy* (1) *and the condition*

$$|c_n| \leq \text{const.}\ \alpha^n \qquad (n = 1, 2, \cdots) \tag{14}$$

for some constant α, $0 < \alpha < 1$. *Then the associated matrices* T *and* F *are unitarily equivalent; thus, there exists a unitary matrix* U *such that*

$$T = UFU^*\ . \tag{15}$$

The condition (14) easily assures $\sum|c_n| < \infty$ and hence the boundedness of A and therefore (cf. § 1 above) that of all other operators considered. If all $c_n = 0$, then T and F are both the zero operator (matrix) and (15) is trivial. Suppose then that not all c_n are 0. It will first be shown that F is absolutely continuous.

To this end, consider $f(z) = 2\sum_1^\infty c_n \cos nz$ for the complex variable

$z = x + iy$. It is clear that $|\cos nz| = \frac{1}{2}|e^{inz} + e^{-inz}| \leq \frac{1}{2}(e^{ny} + e^{-ny}) \leq e^{n|y|}$ and hence, by (14), $|c_n \cos nz| \leq \text{const.}\,(\alpha e^{|y|})^n$. Since $0 < \alpha < 1$, it follows that $\alpha e^{|y|} < 1$ for y sufficiently small and so $f(z) \not\equiv 0$ and is analytic in a strip containing the real axis. Consequently $df(\theta)/d\theta$ can be zero at most a finite number of times on $0 \leq \theta \leq \pi$ and it follows that the (possibly many-valued) inverse function of $f(\theta)$ on $[0, \pi]$ is absolutely continuous (more precisely, that each of the finite number of branches of the inverse of $f(\theta)$ on $0 \leq \theta \leq \pi$ is absolutely continuous). Moreover, if $\lambda = 2\cos\theta$, the operator F can be represented (cf. (5)) as $F = \int_{-2}^{2} h(\lambda)dE_1(\lambda)$ where $E_1(\lambda)$ is the resolution of the identity belonging to the matrix J. Since $h(\lambda) = f(\theta)$ via the substitution $\lambda = 2\cos\theta$ it is clear that $h(\lambda)$ has a (possibly many-valued) absolutely continuous inverse and it follows (cf. [11], pp. 521–522) that F is absolutely continuous, as was to be shown. In fact, if one considers the spectral resolution of F as given by (13), it is seen from a comparison with (5) that zero sets on the λ-interval $-2 \leq \lambda \leq 2$ correspond to zero on the θ-interval $0 \leq \theta \leq \pi$ via the mapping $\lambda = 2\cos\theta$ and that F is absolutely continuous if and only if the relation

$$\{\theta; f(\theta) \text{ in } Z\} \text{ is a zero set} \tag{16}$$

holds whenever Z is a zero set.

By (VI) it now follows that T also is absolutely continuous. Moreover, since by (14),

$$\sum_i \sum_j c_{i+j}^2 = \sum_n n c_{n+1}^2 < \infty ,$$

K is completely continuous. In order to complete the proof it will be shown that

$$tr\,|K| < \infty , \tag{17}$$

where $|K|$ denotes the non-negative square root of K^2. An application of a theorem of Rosenblum ([12], p. 998, will then yield the desired unitary equivalence relation (15). See also Kato [6].

There remains then to prove (17). Let $\{\phi_n\}$, $n = 1, 2, \cdots$, denote the complete orthonormal sequence of vectors for which the n-th component of ϕ_n is 1 and all others are 0. Then

$$\begin{aligned} tr\,|K| &= \sum (|K|\phi_n, \phi_n) \leq \sum \|\,|K|\phi_n\| = \sum \|K\phi_n\| \\ &= \sum_n (\sum_m c_{n+m}^2)^{1/2} \leq \sum_n \sum_m |c_{n+m}| \leq \sum n|c_{n+1}| < \infty , \end{aligned}$$

the last inequality by (14). Thus (17) is proved and, as remarked earlier, the proof of (VII) is complete.

The proof of (VII) makes clear the following assertion:

(VIII) *Let the real sequence* $\{c_n\}$ *satisfy* (1) *and suppose that* (16) *holds for every zero set* Z. *In addition, suppose that*

$$(18) \qquad \sum_n n|c_{n+1}| < \infty \ , \quad \text{or even} \ \sum_n (\sum_m c_{n+m}^2)^{1/2} < \infty \ .$$

Then (15) *holds.*

It is clear that (18) implies $\sum|c_n| < \infty$ and hence that A is bounded (cf. § 1 above). In addition (18) implies $\sum nc_{n+1}^2 < \infty$ and hence the complete continuity of K; as shown before, (18) implies (17). Moreover, unless T and F are both 0, it follows from (16) that F (hence, by (VI), also T) is absolutely continuous. Relation (15) now follows from Rosenblum's theorem as before.

It was shown in [4], p. 878, that whenever T is self-adjoint (not even necessarily bounded) it has no point spectrum. On the other hand, F has a point spectrum whenever $f(\theta)$ has an interval of constancy, or more generally, whenever $f(\theta) \equiv$ const. holds on a set of positive measure. This situation can of course easily obtain for non-trivial $f(\theta)$ ($f(\theta) \not\equiv$ const., i.e., since $c_0 = 0$, $f(\theta) \not\equiv 0$) possessing derivatives of arbitrarily high order (but, of course, for which $f(z)$ is not analytic). But if $f(\theta)$ is of class C^p, its Fourier coefficients are $O(n^{-p-2})$ and so it is clear that the hypothesis (14) of (VII) guaranteeing unitary equivalence cannot be weakened to, say, $|c_n| \leq \text{const.}\ n^{-m}$ ($n = 1, 2, \cdots$) for any positive constant m. Of course, as (VIII) implies, relation (14) is not necessary for (15).

5. Hilbert matrices. A case of special interest is afforded by the sequence $\{c_n\}$ defined by $c_0 = 0$, $c_n = n^{-1}$ if $n > 0$ and $c_{-n} = c_n$. This sequence is of the type considered at the beginning of this paper and moreover $T = (|i-j|^{-1})$, $H = ((i+j-1)^{-1})$ and $S = A - A^* = ((i-j)^{-1})$, with the understanding of course that the (i, i) elements of T and S are 0. The matrices S and H are known to be bounded (Hilbert; cf., e.g., [2], pp. 212-213, 223). Moreover, the spectrum of H is exactly the interval $[0, \pi]$ and, in fact, is purely continuous ([6]). The matrix T is known to be unbounded ([2]), p. 214). Concerning T, there will be proved the following theorem:

(IX) *The matrix* $T = (|i-j|^{-1})$ *is a self-adjoint operator and is absolutely continuous; thus if* $T = \int \lambda dE(\lambda)$ *is the spectral resolution of* T, *then* $\int_Z dE(\lambda) = 0$ *for every set* Z *of one-dimensional Lebesgue measure zero.*

That Z is self-adjoint follows from an application of a theorem of

Hartman and Wintner [4], p. 878, if it is noted that $\frac{1}{2}f(\theta) = -\log(2|\sin\frac{1}{2}\theta|) \sim \sum_1^\infty n^{-1}\cos n\theta$ on $(-\pi, \pi)$ is half-bounded. Another proof of the assertion follows from (I) if it is noted that H, hence also K (cf. (7)), is bounded, since the odd function $g(\theta)$ defined by $g(\theta) = \frac{1}{2}(\pi - \theta) \sim \sum_1^\infty n^{-1}\sin n\theta$ on $(0, \pi)$ is bounded.

It is easy to verify that

$$ST - TS = 2H^2 \,, \tag{19}$$

a relation similar to (6). Moreover, since 0 is not in the point spectrum of H (cf., e.g. [7], p. 699 and the reference there to [1]), Theorem 13 of [11] implies, at least formally, the absolute continuity of T. The trouble stems from the fact that boundedness restrictions were imposed in [10] and [11] and that, although S and H in (19) are bounded, T is not. As a consequence, equation (19), although a valid matrix equation, conceivably cannot be regarded as an operator equation in Hilbert space. More precisely, it is not clear that whenever x is in the domain of T, D_T, then (19) holds, so that

$$STx - TSx = 2H^2x \,, \tag{20}$$

with the understanding that STx and TSx of (20) should mean $S(Tx)$ and $T(Sx)$ respectively. (For operator equations the associative law is of course essentially a matter of definition.) It will be shown below that in fact (20) does hold as an operator equation valid at least for all x in D_T. Once this has been established, it is easy to carry out the same reasoning as in [10], cf. pp. 1027–1028 (where the boundedness of all operators was supposed) and to obtain the equation (10) above, corresponding to formula line (4) in [10]. The absolute continuity of T then follows (cf. Theorem 13 of [11]) from the fact that 0 is not in the point spectrum of H.

In order to complete the proof there remains to be shown that if x is in D_T then (20) holds. To this end, it will be shown that if x is in Hilbert space, that is if $\|x\| < \infty$, then each of the series

$$\sum_m \sum_n s_{im}t_{mn}x_n \text{ and } \sum_m \sum_n t_{im}s_{mn}x_n \tag{21}$$

is absolutely convergent for $i = 1, 2, \cdots$, where, for convenience, $T = (t_{ij})$ and $S = (s_{ij})$. Grant, for the moment, that this has been shown. Then, from the absolute convergence of the first series of (21), it follows that in the iterated series the orders of summation may be interchanged, and hence that, for x in Hilbert space, the corresponding components of the vectors $(ST)x$ and $S(Tx)$ are identical. Now, if it is assumed in addition that x is in D_T, then the vector $S(Tx)$ is in Hilbert space, since S is bounded. Consequently $(ST)x$ is in Hilbert space and, since H^2 is bounded, it follows from the (matrix) equation (19) that $(TS)x$ is in Hilbert space. The absolute convergence of the second series of (21) then implies that $(TS)x = T(Sx)$, so that $T(Sx)$ is in Hilbert space (that is, essentially, that Sx is in D_T). Moreover, it is now seen that (19) implies the validity of (20) as an operator equation valid at least for all vectors in D_T.

Thus, in order to complete the proof of (IV) there now remains to be shown that the series of (21) are absolutely convergent whenever $\|x\| < \infty$. Consider the series $S_i = \sum_m \sum_n |s_{im}t_{mn}x_n|$. Since $t_{ij} \geq 0$ for all i, j and $s_{ij} \geq 0$ or $s_{ij} < 0$ according as $i \geq j$ or $i < j$, it is clear that

$$S_i = -\sum_{n=1}^{\infty}\left(\sum_{m=1}^{i} s_{im}t_{mn}|x_n|\right) + \sum_{n=1}^{\infty}\left(\sum_{m=i+1}^{\infty} s_{im}t_{mn}|x_n|\right). \tag{22}$$

But the inside series of the first double series on the right of equation (22) is finite and, consequently, the orders of summation may be interchanged to obtain $-\sum_{m=1}^{i} s_{im}(\sum_{n=1}^{\infty} t_{mn}|x_n|)$. Since x and the rows of T are in Hilbert space, the inside summation of this last series is always convergent by the Schwarz inequality. Hence the first series of (21) is absolutely convergent if and only if the series

$$\sum_{n=1}^{\infty}\left(\sum_{m=1}^{\infty} s_{im}t_{mn}\right)|x_n|, \tag{23}$$

obtained through modification of S_i in (22) by changing the sign of the first series, is convergent. Now the inside summation of (23) is the (i, n) element of $ST = D = (d_{in})$. Since S is bounded and the columns of T are in Hilbert space, the columns of D are in Hilbert space, that is $\sum_i d_{in}^2 < \infty$. But the matrix equation (19) can be written as $D + D^* = 2H^2$; hence, since H^2 is bounded, the columns of D^* and therefore the rows of D, are also in Hilbert space. Hence $\sum_n d_{in}^2 < \infty$ and so (23) is convergent by the Schwarz inequality. It has now been proved that the first series of (21) is absolutely convergent (for $i = 1, 2, \cdots$). Using the fact that $|s_{ij}| = t_{ij}$ it is seen that the absolute convergence of the first series of (21), that is, the convergence of $\sum_m \sum_n t_{im}t_{mn}|x_n|$, whenever $\|x\| < \infty$, implies the absolute convergence of the second series of (21). Thus both series are convergent for all x in Hilbert space and (IX) follows as indicated above.

References

1. T. Carleman, *Sur les équations intégrales singulières a noyau réel et symétrique*, Uppsala, 1923.
2. G. H. Hardy, J. E. Littlewood, and G. Polya, *Inequalities*, Cambridge, 1934.
3. P. Hartman and A. Wintner, *On the spectra of Toeplitz's matrices*, Amer. J. Math., **72** (1950), 359-366.
4. P. Hartman and A. Wintner, *The spectra of Toeplitz's matrices*, Amer. J. Math., **76**

(1954), pp. 867-882.
5. D. Hilbert, *Grundzüge einer allgemeinen Theorie der linearen Integralgleichuugen*, Leipzig, 1912.
6. T. Kato, *Perturbation of continuous spectra by trace class operators*, Proc. Japan Academy, **33** (1957), 260-264.
7. W. Magnus, *On the spectrum of Hilbert's matrix*, Amer. J. Math., **72** (1950), 699-704.
8. *Mathematical Monographs*, Northwestern University, vol. **1**, 1941.
9. Z. Nehari, *On bounded bilinear forms*, Annals of Math., **65** (1957), 153-162.
10. C. R. Putnam *On commutators and Jacobi matrices* Proc. Amer. Math. Soc., **7** (1956), 1026-1030.
11. ———, *Commutators and absolutely continuous operators*, Trans. Amer. Math. Soc., **87** (1958), 513-525.
12. M. Rosenblum, *Perburbations of the continuous spectrum and unitary equivalence*, Pacific J. Math., **7** (1957), 997-1010.
13. O. Toeplitz, *Zur Theoric der quadratischen und bilinearen Formen von unendlichvielen Veränderlichen*, Göttinger Nachrichten (1910), 489-506.

PURDUE UNIVERSITY

ON NONLINEAR POSITIVE OPERATORS

H. H. SCHAEFER

Introduction. The purpose of the present paper is to apply the well known Schauder fixed point theorem, in its general from due to Tychonov [8], to the situation of nonlinear (or rather, not necessarily linear) maps defined on (or on a subset of) the " positive " cone in a partially ordered locally convex linear space. Throughout this paper, no use is made of possible linear properties of the maps under consideration. As far as the author is informed, there is little history to the study of such mappings; the only work done seems to be contained in papers by Krein-Rutman [2], Rothe [9] and Morgenstern [3]. In [2], the Schauder theorem is largely applied to linear maps (where it can be avoided) and a few non-linear cases[1]. In [4], the author paid attention mainly to the case of linear compact maps in general locally convex spaces. At the end of that paper, with a somewhat sketchy proof, a general nonlinear theorem[2] is stated which however seems to need some improvement.

In this paper, the essential proposition resulting from the fixed point theorem is stated in the form of three different theorems to throw some light on potential ways of argument. While Th. 1, depending on a special convexity argument, is of a different character, Th. 2 is almost a special case of Th. 3. But as Banach spaces with normal order cones (with which Th. 2 is concerned) seem to be the most important ones in non-linear analysis, it might be useful to have the theorem stated separately, a much simpler proof than that of Th. 3 going with it. Applications have been selected so as to furnish a non-trivial example to each of the three theorems, the one to Th. 1 showing that it is not always fruitful to restrict attention to normed topologies. It is understood that each example constitutes a new result in its respective field.

Preliminary material. In the present section, we are going to collect some theorems and definitions on which argumentation will be primarily based in the sections to follow. The main tool will be the

FIXED POINT THEOREM (Tychonov). *Let E be a locally convex linear space, M a convex compact subset of E. If T is a continuous map on M into M, then T has a fixed point $x_0 \in M$.*

Received September 29, 1958, and in revised form February 11, 1959.

[1] Also, considerations are restricted to Banach spaces.

[2] Satz 3.1. This is restated and proved in this paper as Th. 3. The additional assumption to be made in [4] may be any one of hypotheses α, β stated with Th. 3 of the present paper.

For the proof, see [8]. To make this theorem more easily applicable to mappings that carry sets not necessarily compact into compact ones, we state the following slightly more general

FIXED POINT THEOREM (2nd form). *Let M be a complete convex set in E. If T is continuous on M into M such that $T(M)$ is relatively compact, then T has a fixed point $x_0 \in M$.*

Proof. Let M_1 be the closed convex hull of $T(M)$. As M is complete, M_1 is compact (Bourbaki [1], p. 81) and since obviously $T(M_1) \subset M_1$, Tychonov's theorem yields the desired result.

Let E be a linear space over the real scalar field. A partial ordering of E is a binary relation " $<$ " such that

1 $$x < x \text{ for all } x \in E .$$

2 $$\{x < y \,\&\, y < z\} \Rightarrow x < z .$$

3 $$\{x < y \,\&\, y < x\} \Rightarrow x = y .$$

Such an ordering is said to be compatible with the linear structure of E if in addition

4 $$\{x > 0 \,\&\, \lambda \geqq 0\} \Rightarrow \lambda x > 0 .$$

5 $$x > y \Rightarrow x + z > y + z \qquad \text{for all } z \in E .$$

The set of all $x \in E$ such that $x > 0$ is a convex cone C which contains its vertex 0, and which is proper (i.e. $C \cap -C = \{0\}$). C will be referred to as the positive cone with respect to a given partial ordering of E[3]). Conversely, each cone in E with the listed properties defines a partial ordering satisfying axioms 1 through 5, $x < y$ meaning $y - x \in C$.

Let E be a linear space, partially ordered by some such cone C. If T is a mapping defined on a subset of C, we will say T is positive whenever the range of T is in C. If E is, moreover, a topological space, T will be called strictly positive if $T(x_n) \to 0$ implies $x_n \to 0$ for any sequence $\{x_n\}$ in the domain of T.

Examples.

1. Let E be Hilbert space L_2 (0, 1) in its natural order, i.e. the positive cone C consisting of all elements $f : f(t) \geqq 0$, $t \in [0, 1]$. The positive mapping, defined on all of E,

$$T(f) = \int_0^t f^2(\tau)d\tau$$

[3] In this paper, all orderings are understood to be compatible with the linear structure of the space involved. Also, we exclude the trivial case $C = \{0\}$.

is not strictly positive. Take $f_n = \sqrt{nt^{n-1}}$, then $\|f_n\| = 1$. Now as

$$T(f_n) = t^n,\ \|T(f_n)\| = \frac{1}{\sqrt{2n+1}} \to 0 .$$

2. Let E be the B-space of continuous functions on the unit interval, with its natural partial order. Let $K(t, \tau)$ be $\geqq 0$ (but $\not\equiv 0$) and continuous on the unit square. If $P(z)$ is a polynomial with non-negative coefficients,

$$T(f) = \int_0^1 K(t, \tau)P[f(\tau)]d\tau$$

is strictly positive if and only if the constant term in P is > 0.

3. Denote by $\{E_\alpha\}$ a collection of topological linear spaces, each E_α being partially ordered by some positive cone C_α. Then the product space $E = \prod_\alpha E_\alpha$ is ordered by $C = \prod_\alpha C_\alpha$. Let A_α be a positive map on E into E_α, and consider the map

$$A(x) = (\cdots, A_\alpha(x), \cdots)$$

on E into E. Then A is strictly positive on C if and only if to each α, there is a $\beta(\alpha)$ such that $A_\beta(x) \to 0$ (in E_β) implies $x_\alpha \to 0$ (in E_α). In particular, if $A_\alpha(x) = A_\alpha(x_\alpha)$, then A is strictly positive if and only if each A_α is.

I. Morgenstern's theorem. If E is the Banach space L_1, partially ordered by the positive cone $C = \{f: f(t) \geqq 0\}$, it turns out that the intersection of C with the unit sphere $S = \{f: \|f\| = 1\}$ is convex. This is true for any abstract L-space or, more generally, for any normed space in which the norm is additive on C. To this situation Morgenstern [3] applied Schauder's fixed point theorem. He obtained the following

THEOREM (Morgenstern)[4]. *Let E be a Banach space, partially ordered by a positive cone C which is closed and on which the norm is additive. Then if T is continuous and strictly positive on $C \cap \{\|x\| = c\}$, $c > 0$, mapping this set into a compact one, there is some $\lambda > 0$ and $x \in C$ such that $\lambda x = T(x)$, $\|x\| = c$.*

The proof is readily obtained by applying the fixed point theorem (2nd form) to the map $cT(x)/\|T(x)\|$ on the set $C \cap \{\|x\| = c\}$. However, it may be so arranged as to yield a much more general proposition.

[4] The theorem is stated in our terminology and a slightly more general form.

THEOREM 1. *Let E be a locally convex space, partially ordered by a complete cone C, and let T be a continuous, strictly positive transformation on C, mapping bounded sets into compact ones. Assume $H: f(x)=1$ is a closed hyperplane meeting C in a nonvoid, bounded set. Then to each $c > 0$, there is an $x \in C$ and $\lambda > 0$ with*

$$\lambda x = T(x)\ , \quad f(x) = c\ .$$

Proof. It follows from our assumptions that the continuous linear form $f(x)$ is > 0 at every non-zero point of C. For assume there is an $x_0 \in C$, $x_0 \neq 0$, with $f(x_0) = 0$. Then if $y_0 \in H \cap C$, we would have $f(y_0 + \mu x_0) = 1$ for all $\mu \geqq 0$ which contradicts the hypothesis that $H \cap C$ be bounded. It is now also clear that f cannot be < 0 on C. Applying the fixed point theorem (2nd form) to the map $cT(x)/f[T(x)]$ on the set $H \cap C$, we get the desired result letting $\lambda = f[T(x)].\, c^{-1}$.

REMARK. We should point out the relation between Morgenstern's theorem and Th. 1. If, under the assumptions of the former, the norm coincides on C with a continuous linear form, then Morgenstern's theorem is a corollary of Th. 1. (This is the case in L_1, e.g.). Assume then, still under the assumptions of Morgenstern's theorem, that there is no such linear form. Now $C \cap \{\|x\| = c\}$ is convex ($c > 0$), so there is a closed hyperplane H separating this set from a convex open neighborhood of 0. Obviously $H \cap C$ is bounded and Th. 1 can be applied provided T is compact, continuous and strictly positive on C.

II. Banach spaces with normal positive cones. We will now extend Morgenstern's theorem to ordered Banach spaces in which the norm is not necessarily additive on the positive cone C. This assumption will be replaced by the weaker hypothesis that C is normal. A convex cone of vertex 0 in a normed space E is normal [5] if the topology of E is generated by a norm which is monotone (with respect to the order induced in E by C) on C. In terms of the given norm on E, $x \to \|x\|$, this amounts to saying there is a constant $\gamma > 0$ such that

$$\|x + y\| \geqq \gamma \|y\| \qquad \text{for all } x \in C, y \in C\ .$$

It can easily be checked that for all classical Banach spaces, the positive cones pertaining to their natural partial orders are normal ([5], p. 130).

THEOREM 2. *Let E be a normed space, partially ordered by a complete normal cone C. Let T be a strictly positive transformation, which is continuous and maps bounded subsets of C into compact ones. Then to each $c > 0$, there is $x \in C$ and $\lambda > 0$ with*

$$\lambda x = T(x)\ , \quad \|x\| = c\ .$$

Proof. Consider the mapping

$$x \to S_c(x) = T(x) + |c - \|x\||y$$

for fixed $0 \neq y \in C$ and $c > 0$. This is a continuous map carrying bounded subsets of C into compact ones. T being strictly positive, we have $\inf \left\{\|T(x)\| : x \in C \,\&\, \|x\| \geqq \frac{1}{2}c\right\} = \varepsilon > 0$. Hence, C being a normal cone, we obtain

$$\inf_{x \in C} \|S_c(x)\| \geqq \gamma \cdot \inf_{x \in C} \sup(\|T(x)\|, |c - \|x\||\,\|y\|) \geqq \gamma \sup\left(\varepsilon, \frac{1}{2}c\|y\|\right) > 0 .$$

Thus $x \to R_c(x) = cS_c(x)\|S_c(x)\|^{-1}$ maps $C \cap \{\|x\| \leqq c\}$ into a compact subset. So by the fixed point theorem (2nd form) there is an x in this subset with $x = R_c(x)$. Clearly $\|x\| = c$, and letting $\lambda = c^{-1}\|S_c(x)\|$ we have $\lambda x = T(x)$. Since $c > 0$ is arbitrary, the proof is complete.

III. A third theorem. The theorem presented in this section weakens the assumption in Th. 2 that E be normed and removes the hypothesis that C be a normal cone. Instead, we require either one of conditions α, β of hypothesis H (s. below) to hold. As the conclusion is only established for *some* continuous semi-norm $x \to p(x)$ on E (which, however, may be assumed to generate the topology of E if E is normed), Th. 3 is not a generalization of Th. 1 or 2. We start out with a

LEMMA. *If E is a locally convex space, C a closed proper convex cone in E of vertex* 0, *then there exists a continuous linear form on E, non-negative on C and > 0 at a given non-zero element of C.*

Proof. Let $0 \neq y \in C$. Since C is proper and closed, there is a convex open neighborhood U of $-y$ such that C and $\bigcup_{\lambda > 0}\lambda U$ do not intersect. Hence there is a closed hyperplane H separating C and $\bigcup_{\lambda > 0}\lambda U$. Obviously H contains 0, so has an equation $f(x) = 0$. After a potential change of sign, f will meet the requirement.

Now let E be any locally convex space, partially ordered by a complete positive cone. A mapping T, defined on a neighborhood of 0 in C into C, will be called of type "P" if it satisfies:

1. T is continuous and strictly positive.
2. There is a neighborhood U of 0 such that the image under T of $U \cap C$ is relatively compact.

Consider

HYPOTHESIS H. We will say that hypothesis H is satisfied if one of the two following statements is true:

α. To each compact subset of C, there exists a continuous semi-norm which is > 0 at each non-zero point of that set.

β. T is positive-homogeneous of some degree $\sigma > 0$, i.e. $T(\lambda x) = \lambda^\sigma T(x)$ for $x \in C$ and $\lambda > 0$.

For instance, condition α is automatically fulfilled if there exists a continuous norm on E (or even on C)[5]. Condition β is of course satisfied if T is a linear map.

THEOREM 3. *Assume hypothesis* H *holds and* T *is a mapping of type* "P". *Then there exists a continuous semi-norm* p *such that for each* $0 < c \leqq 1$, *there are an* $x \in C$ *and* $\lambda > 0$ *satisfying*

$$\lambda x = T(x)\ , \quad p(x) = c.$$

Proof. Let $U = \{x : q_1(x) \leqq 1\}$ be a closed neighborhood of 0 such that $T(U \cap C)$ is relatively compact. Second, let q_2 be selected, according to which one of conditions α, β in H is satisfied, as follows:

Case α. Let q_2 be a continuous semi-norm strictly positive on $T(U \cap C)$.

Case β. Let $q_2 = q_1$.

Third, by the lemma, we may choose an $y \in C$ and a continuous linear form f such that $f \geqq 0$ on C while $f(y) > 0$. We may further suppose that $\sup\{q_1(y), q_2(y), f(y)\} = 1$.

Put $p = \sup\{q_1, q_2, |f|\}$ and consider the set $U_1 = \{x \in C : p(x) \leqq c\}$, c being any fixed real number between 0 and 1 (1 included). For any positive integer n, form the mappings

$$T_n(x) = T(x) + |1 - p(Tx)| \cdot n^{-1}y$$

and $S_n(x) = cT_n(x)/p[T_n(x)]$. Obviously, S_n is a transformation of type "P", mapping U_1 into itself, provided the denominator $p[T_n(x)]$ has a positive lower bound. To show that this is true, consider first all $x \in U_1$ such that $p(T(x)) \leqq \frac{3}{4}$. Then

$$p(T_n(x)) \geqq f[T_n(x)] \geqq \frac{1}{4n} f(y) \neq 0\ .$$

For the remaining elements $x \in U_1$ we have $p(T(x)) > \frac{3}{4}$, hence

$$p(T_n(x)) \geqq p(T(x)) - |1 - p(T(x))| n^{-1} p(y) > \frac{3}{4} - \frac{1}{4} = \frac{1}{2}\ .$$

---

[5] Condition α can be weakened so as to require the existence and continuity of the semi-norms involved only on C.

Applying the fixed point theorem (2nd form) to S_n, we are sure there is an $x_n \in U_1$ satisfying $x_n = S_n(x_n)$, $p(x_n) = c$. Letting $\lambda_n = p(T_n(x_n)) \cdot c^{-1}$, by definition of S_n we obtain

$$(*) \qquad \lambda_n x_n = T(x_n) + |1 - p(T(x_n))| n^{-1} y \ .$$

$\{T(x_n)\}$ being relatively compact, it follows that $\{\lambda_n\}$ is a bounded sequence. Assume, for the moment, that $\{\lambda_n\}$ has a positive lower bound. Then as the right-hand side of our last equation is relatively compact, so is $\{x_n\}$. (Here we may remark that for a convergent subsequence of $\{x_n\}$, the corresponding subsequence of λ's converges automatically to some $\lambda > 0$.) Hence for each limiting point of a subsequence of $\{\lambda_n x_n\}$, such that $\lambda_{n_k} \to \lambda$, $\lambda x = T(x)$ and, by continuity, $p(x) = c$.

All that remains to prove is that $\lambda_n > \eta > 0$ for all n. Suppose there were a subsequence $\{\lambda_k\}$ tending to zero. From this it would follow that $p(T(x_k)) \to 0$ which, by definition of p, in turn would imply $q_2(T(x_k)) \to 0$. On the other hand, T being strictly positive, 0 is no limiting point to the sequence $T(x_n)$ because of $p(x_n) = c$. Thus if α of H is satisfied, we arrive at a contradiction. Now assume H holds by virtue of condition β. Letting $z_n = \lambda_n x_n$, $\{z_n\}$ has a limiting point z, say. Because T is strictly positive, we must have $z \neq 0$. Multiplying equation (*) by λ_n^σ, we get

$$\lambda_n^\sigma z_n = T(z_n) + o\left(\frac{1}{n}\right) y \ .$$

Now if there were any subsequence $\{\lambda_k\}$ of $\{\lambda_n\}$ such that $\lambda_k \to 0$, we would obtain (as $\lambda_k^\sigma \to 0$) $T(z) = 0$ for some $z \in C$, $z \neq 0$. This again contradicts the hypothesis that T be strictly positive, and the proof is complete.

REMARK. Hypothesis H was needed to prove that $\{\lambda_n\}$ does not have 0 as a limiting point. The proof of Satz 3.1 in [4] is essentially the same as the one presented here, but is incorrect at the point where it says " $\lambda_0 > 0$ " (l.c., p. 329, line 3 f.b.).

Applications. The remainder of this paper is concerned with a number of applications to the preceding theorems.

1. Consider the linear space ω of all real sequences $x = (x_1, x_2, \cdots)$, partially ordered by the positive cone $C = \{x : x_i \geqq 0, i = 1, 2, \cdots\}$. In the product topology (i.e. considering ω the product of countably many real lines) ω is locally convex. Let $r > 0$ be a fixed integer and let $k = (k_1, k_2, \cdots)$ denote any sequence of non-negative integers such that

$\sum_{i=1}^{\infty} k_i = r$.[6] Since the set $\{k\}$ of all such k is countable, we may arrange it into a sequence; hence consider $\{k\}$ as ordered by the natural order of subscripts.

Further, denote by x^k the product $\prod_{k_i \neq 0} x_i^{k_i}$. Now if to each k and each positive integer i there corresponds a real number $a_{ik} \geqq 0$ such that

$$\sum_{(k)} a_{ik} < C ,$$

where C is independent of i, the equations

$$y_i = \sum_{(k)} a_{ik} x^k$$

define a mapping $y = A(x)$ on the subspace of all bounded sequences into itself such that each y_i is a homogeneous form of degree r in the variables $x_1, x_2, \cdots$ (If $r = 1$, then A defines a bounded linear map on the B-space (m)).

Consider the properties

α. There are n rows in A (the first n rows, say) such that

$$\sum_{i=1}^{n} a_{ik} \geqq a_{jk} \qquad \text{for all } k \text{ and all } j > n .$$

β. $\sum_{i=1}^{n} y_i \to 0$ implies $\sum_{i=1}^{n} x_i \to 0$[7].

We prove the following theorem:

If a mapping A of the above mentioned type satisfies α and β, there are a $\lambda > 0$ and an $x > 0$ for which

$$\lambda x = A(x) .$$

REMARK. If $r = 1$, then the point spectrum of the bounded map A on (m) contains a positive real number.

Proof. Consider in ω the cone $C_1 = C \cap \{x : \sum_{i=1}^{n} x_i \geqq x_j, j > n\}$. Owing to α, $A(x)$ is defined on the cone C_1 into itself. Since ω is complete and C_1 closed, C_1 is a complete cone in ω. Next we show that A, which is in general not defined but on a dense subset of ω, is continuous on C_1. Let $x_n \to x$ in C_1. It follows from the definition of C_1 that all coordinates of all the x_n are uniformly bounded, say by some

[6] A more general theorem results if we admit all k such that $\sum_{i=1}^{\infty} k_i \leqq r$.

[7] Cf. Example 3 in the preliminary section.

constant $M \geqq 1$. Given $\varepsilon > 0$ and any fixed subscript i, we can find a k_0 such that

$$\sum_{k>k_0} a_{ik} < \frac{\varepsilon}{3M^r} .$$

Then if $y_n = A(x_n)$ and $y = A(x)$, we obtain

$$|(y_n - y)_i| \leqq \sum_{k \leqq k_0} |a_{ik}x_n^k - a_{ik}x^k| + \sum_{k>k_0} a_{ik}x_n^k + \sum_{k>k_0} a_{ik}x^k .$$

The last two righthand terms are, by the choice of k_0, each less than $\frac{1}{3}\varepsilon$. The first righthand term will be less than $\frac{1}{3}\varepsilon$ for $n > n_0$ if n_0 is large enough, since there are only finitely many coordinates of both x_n and x involved. Thus we have $|(y_n - y)_i| < \varepsilon$ if $n > n_0$ and continuity is established.

Now $f(x) = \sum_{i=1}^{n} x_i$ is a continuous linear form on ω. The intersection of the hyperplane $f(x) = 1$ with C_1 is certainly bounded as $(0 \leqq) x_i \leqq 1, i = 1, 2, \cdots$, in that intersection. Moreover, a set $\{x\}$ is bounded in ω if and only if $|x_i| < M_i$ uniformly on $\{x\}$. If M_i can be chosen independently of i, then the set is relatively compact by the well known Tychonov theorem. Thus on C_1 closed bounded sets coincide with compact sets, and A transforms bounded sets into compact ones on C_1.

By hypothesis β, A is strictly positive on C_1. (Conditions more explicit than β may be obtained easily by applying the reasoning of Example 3, preliminary section.) Hence A meets all the requirements of Th. 1 and the proof is complete.

2. In a recent paper [7], Schmeidler proved the existence of an eigenvalue to the homogeneous algebraic integral equation of order n

$$(*) \qquad \mu^n y^n(s) - \sum_{\beta=0}^{n-1} \mu^\beta y^\beta(s) a_\beta(s, y) = 0 , \qquad (0 \leqq s \leqq 1)$$

where n is an odd integer > 0 and

$$a_\beta(s, y) = \int_0^1 \cdots \int_0^1 K_\beta(s, t_1, \cdots, t_\nu) y^{\beta+1}(t_1) \cdots y^{\beta+1}(t_\nu) dt_1 \cdots dt_\nu$$

are homogeneous integral forms with continuous kernels $K_\beta(s_1 t_1 \cdots t_\nu)$ such that $(\beta + 1)(\nu + 1) = n + 1$ and the K's are symmetric with respect to all their arguments. Schmeidler shows (*) to be the natural generalization of a linear Fredholm equation with continuous symmetric kernel. In an earlier paper [6], a theorem was stated by Schmeidler that generalizes the well known Jentzsch theorem on linear Fredholm integral equations with positive kernel. The proof of that theorem of

Schmeidler's, however, appears to be incorrect[8]. We are going to show that the theorem yet is correct and holds under weaker conditions than the ones stated in [6]. Let us call (*) an *algebraic integral equation with non-negative coefficients* if n is *any* positive integer and

$$a_\beta(s, y) = \sum_{\alpha_1+\cdots+\alpha_\nu=n-\beta} \int_0^1 \cdots \int_0^1 K_{\beta\alpha_1} \cdots \alpha_\nu(s, t_1, \cdots, t_\nu) y(t_1)^{\alpha_1} \cdots y(t_\nu)^{\alpha_\nu} dt_1 \cdots dt_\nu$$

are homogeneous integral forms of order $n - \beta$ with continuous kernels $K_{\beta\alpha_1\cdots\alpha_\nu}(s, t_1, \cdots, t_\nu) \geqq 0$. We will prove this theorem:

If some α_β contains a term $\int K(s, t_1, \cdots, t_\nu) y(t_1)^\alpha \cdots y(t_r)^\alpha dt_1 \cdots dt_\nu$ such that α is the highest power occurring in any α_β, and if

$$\int_0^1 K(s, t_1, \cdots, t_\nu) ds \geqq \delta > 0 \qquad \textit{for } (t_1, \cdots, t_\nu) \in [0, 1]^\nu;$$

then (*) *has an eigenvalue $\mu_0 > 0$ with eigenfunction $y_0(s) \geqq 0$.*

Proof. We first state a

LEMMA. *Consider the mapping $\varphi: (a_0, \cdots, a_{n-1}) \to z_0$ where z_0 is the greatest real root of*

$$z^n - \sum_{\beta=0}^{n-1} a_\beta z^\beta = 0 . \tag{1}$$

Then φ is defined and continuous on the set $\{a_\beta \geqq 0; 0 \leqq \beta \leqq n-1\} \subset E^n$.

It is clear that $z_0 = 0$ if and only if $a_0 = a_1 = \cdots = a_{n-1} = 0$, and φ is continuous at that point. At any other point, however, z_0 is a simple root which implies continuity of φ.

Recalling that α is the highest power of y in any a_β, we observe that $a_\beta(s, y)$ $(0 \leqq \beta \leqq n-1)$ exist for all $y(s) \in L_\alpha(0, 1)$. Moreover, each $a_\beta(s, y)$ is a continuous map on $[0, 1] \times L_\alpha$ into the space $C(0, 1)$ of continuous functions on $[0, 1]$. For

$$|a_\beta(s, y) - a_\beta(t, \tilde{y})| \leqq |a_\beta(s, y) - a_\beta(s, y)| + |a_\beta(s, \tilde{y}) - a_\beta(t, \tilde{y})| \tag{2}$$

where the first righthand term can be estimated by expressions of the form

$$I = \int K[y(t_1) - \tilde{y}(t_1)](y(t_1)^{\alpha_1 - 1} + \cdots \tilde{y}(t_1)^{\alpha_1 - 1}) y(t_2)^{\alpha_2} \cdots y(t_\nu)^{\alpha_\nu} dt_1 \cdots dt_\nu .$$

Using Hölder's inequality we arrive at an estimate

$$|I| \leqq \text{const} \cdot \|y - \tilde{y}\| P(\|y\|, \|\tilde{y}\|)$$

where $P(u, v)$ is a homogeneous polynomial of order $n - \beta - 1$ in u, v

[8] The treacherous point is that the mapping $\eta \to y$, **[6]** p. 252 above, is not continuous.

and $\|\ \|$ denotes the norm in L_α. Thus, $a_\beta(s, y) - a_\beta(s, \tilde{y}) \to 0$ uniformly in s as $\tilde{y} \to y$ in L_α. The second righthand term in (2) can be estimated by terms

$$II = \int [K(s, t_1 \cdots t_\nu) - K(t, t_1 \cdots t_\nu)]\tilde{y}(t_1)^{\alpha_1} \cdots \tilde{y}(t_\nu)^{\alpha_\nu} dt_1 \cdots dt_\nu ;$$

so if $|s - t| < \delta(\varepsilon)$, the K's being continuous, we obtain

$$|II| \leqq \varepsilon \int \tilde{y}(t_1)^{\alpha_1} dt_1 \cdots \int \tilde{y}(t_\nu)^{\alpha_\nu} dt_\nu \leqq \varepsilon \|\tilde{y}\|^{n-\beta}$$

remembering that $\alpha_1 + \cdots + \alpha_\nu = n - \beta$ and $\int_0^1 y^\gamma dt \leqq \left(\int_0^1 y^\alpha dt\right)^\alpha$ if $\gamma \leqq \alpha$. This proves continuity of a_β on $[0, 1] \times L_\alpha$[9]. But from the above reasoning it is obvious that $\{a_\beta(s, y)\}$ is an equicontinuous, bounded set of functions if y runs through any bounded set of L_α. Hence, by the lemma, we have established:

The mapping $y(s) \to z_0(s)$, z_0 being defined as the greatest real root of (1) *for each $s \in [0, 1]$, maps any bounded subset of the positive cone in L_α onto a set of equicontinuous, non-negative and uniformly bounded functions over* $[0, 1]$.

Thus the map $y \to z_0$ satisfies the assumptions of Th. 2 if we can show that it is strictly positive. For that end, let $\|z_0\| \to 0$. If K is the kernel mentioned in our present theorem, we get by (1)

$$z_0(s)^n \geqq z_0(s)^\beta \int_0^1 \cdots \int_0^1 K(s, t_1, \cdots, t_\nu) y(t_1)^\alpha \cdots y(t_\nu)^\alpha dt_1 \cdots dt_\nu$$

where $\nu\alpha + \beta = n$. Since z_0 is the greatest real root of (1), there follows

$$z_0(s)^{n-\beta} \geqq \int_0^1 \cdots \int_0^1 K(s, t_1, \cdots, t_\nu) y(t_1)^\alpha \cdots y(t_\nu)^\alpha dt_1 \cdots dt_\nu \ .$$

Integrating this last equation, we obtain

$$(3) \qquad \int_0^1 z_0(s)^{\nu\alpha} ds \geqq \delta\left[\int_0^1 y(t)^\alpha dt\right]^\nu = \delta \|y\|^{\nu\alpha} \ .$$

Now assume first that y is bounded in L_α as $z_0 \to 0$. Then z_0 runs through a uniformly bounded set of continuous functions and hence, as it converges to 0 in measure, $\int_9^1 z_0(s)^{\nu\alpha} ds \to 0$. By (3) this implies $\|y\| \to 0$. This excludes that $\|y\| \to \infty$ as $z_0 \to 0$, for division of z and y by $\|y\|$ (remember that (*) is homogeneous) would lead to $\|y\| = 1$ while $z_0 \to 0$.

[9] In general, α_β are not continuous for the weak topology on L_α.

Thus $y \to z_0$ is strictly positive and the application of Th. 2 ends the proof.

3. Let Ω be a compact region in Euclidean n-space, $C(\Omega)$ the B-space of continuous functions on Ω, and $D(\Omega)$ the space of continuously differentiable functions on Ω in the topology of uniform convergence of the function and its first order derivatives. Assume a kernel $K(s, t) \geqq 0$ is given on $\Omega \times \Omega$ and a real function $f(s; u, p_i)$ of $2n+1$ arguments (letting $s = (x_1, \cdots, x_n)$) such that these conditions are satified:

1°. $K(\psi) = \int_\Omega K(s, t)\psi(t)dt$ is a compact linear transformation on $C(\Omega)$ into $D(\Omega)$ which has an eigenvalue $\lambda_1 > 0$ with an adjoint eigenfunction $\varphi(s)$ that is $\geqq 0$ and bounded except on an Ω-subset of (Lebesgue) measure 0.

2°. f is a continuous real valued function, defined for $s \in \Omega$, $u \geqq 0$, $|p_i| < \infty$ $(i = 1, \cdots, n)$ and such that

$$f(s; u, p_i) \geqq \begin{cases} \alpha u \text{ if } 0 \leqq u \leqq \delta\ , \\ K \text{ if } \delta < u\ , \end{cases}$$

where α, δ, K are three suitably chosen positive constants. Then the following theorem holds:

Under conditions 1°., 2°. *the nonlinear integro-differential equation*

$$\lambda u(s) = \int_\Omega K(s, t) f\Big(t; u(t), \frac{\partial u}{\partial x_i}(t)\Big)dt$$

has for each $c > 0$ at least one solution $u \geqq 0$ with $\lambda = \lambda(u) > 0$ and $\|u\| = c$, where $\|\ \|$ denotes a suitably chosen norm of $D(\Omega)$. Moreover, $\lambda(u)$ satisfies the inequality $\lambda \geqq \lambda_1 \cdot \inf(\alpha, Kc^{-1})$.

Proof. $D(\Omega)$ is partially ordered by the positive cone $C = \{u : u \geqq 0 \text{ on } \Omega\}$ but we note that C is not a normal cone (cf. sec. II). For any $\varepsilon > 0$, the transformation[10]

$$T_\varepsilon(u) = \int_\Omega K(s, t)\Big[f\Big(t; u, \frac{\partial u}{\partial x_i}\Big) + \varepsilon\Big]dt$$

is, due to the continuity of f and condition 1°., compact and continuous on C into C. For all $u \in C$ we have

$$\int_\Omega T_\varepsilon(u)ds \geqq \iint K(s, t)\varepsilon dtds \geqq \varepsilon \iint K(s, t)\varphi(s)dtds = \varepsilon\lambda_1 \int \varphi(s)ds > 0$$

[10] The following proof shows how cases may be handled where strict positiveness of the map involved cannot be verified. (T_ε is not necessarily strictly positive by Example 2 of the preliminary section if $\varepsilon = 0$.)

if sup ess $\varphi(s) = 1$. Hence T_ε is strictly positive and we may apply Th. 3, hypothesis H being satisfied through condition α. Following the notation of the proof to Th. 3, choose for q_1 any norm generating the topology of $D(\Omega)$, let $q_2 = \max|u|$ and write $p = \|\ \|$. (Obviously p is a norm generating the topology of $D(\Omega)$; it is also evident that, in the present case, the statement of Th. 3 holds for all $c > 0$ instead of $0 < c \leqq 1$.) Thus, c being fixed, to each $\varepsilon > 0$ there is a $u_\varepsilon \geqq 0$ and $\lambda_\varepsilon > 0$ such that

$$\lambda_\varepsilon u_\varepsilon = T_\varepsilon(u_\varepsilon)\ ,\ \|u_\varepsilon\| = c\ . \tag{1}$$

We are going to show that λ_ε has a positive lower bound for $\varepsilon > 0$. Multiplying (1) by φ and integrating, we obtain

$$\begin{aligned}\lambda_\varepsilon\int_\Omega u_\varepsilon\varphi dt &= \iint K(s, t)\varphi(s)[f(u_\varepsilon(t)) + \varepsilon]dsdt \\ &= \lambda_1\int \varphi(t)[f(u_\varepsilon(t)) + \varepsilon]dt > \lambda_1\int \varphi(t)f[u_\varepsilon(t)]dt\ .\end{aligned} \tag{2}$$

Now Ω is the union of two measurable subsets Ω_1 and Ω_2 such that $u_\varepsilon \leqq \delta$ in Ω_1 whereas $u_\varepsilon > \delta$ in Ω_2. On account of condition 2°. we have

$$\int_{\Omega_1}\varphi f[u_\varepsilon]dt \geqq \alpha\int_{\Omega_1}\varphi u_\varepsilon dt$$

and

$$\int_{\Omega_2}\varphi f[u_\varepsilon]dt \geqq K\int_{\Omega_2}\varphi dt \geqq \frac{K}{c\gamma}\int_{\Omega_2}\varphi u_\varepsilon dt\ ,$$

where $\gamma \leqq 1$ is a constant such that $\max|u| \leqq \gamma\|u\|$ for all $u \in D(\Omega)$[11]. Hence for the last integral in (2),

$$\int_\Omega \varphi f[u_\varepsilon]dt \geqq \alpha\int_{\Omega_1}\varphi u_\varepsilon dt + \frac{K}{c\gamma}\int_{\Omega_2}\varphi u_\varepsilon dt \geqq \inf\left(\alpha_1\frac{K}{c\gamma}\right)\int_\Omega \varphi u_\varepsilon dt$$

and we finally obtain

$$\lambda_\varepsilon \geqq \inf\left(\alpha, \frac{K}{c\gamma}\right)\cdot\lambda_1 \qquad \text{for } \varepsilon > 0\ . \tag{3}$$

Now let $\varepsilon \to 0$ in (1). As $\|u_\varepsilon\| = c$ independently of ε, the righthand side of (1) is a relatively compact sequence and so is the corresponding sequence of u_ε by (3). Thus for a common convergent sequence of λ_ε and u_ε, the limit function u satisfies $\lambda u = T_0(u)$ and the proof is complete.

[11] $p = \sup\{q_1, q_2, |f|\}$ (Th. 3) implies $\gamma \leqq 1$. If f does not depend on any p_i, the proof may be carried via Th. 2 and we will have $\gamma = 1$, $\|u\| = \max|u|$.

REMARK. If $f(s; u, p_i)$ is such that δ may be chosen arbitrarily large (e.g., if $f = \alpha u + g(s; u, p_i)$ with $g \geqq 0$), then we will have Ω_2 void and λ will satisfy the inequality

$$\lambda(u) \geqq \alpha\lambda_1 .$$

We apply the preceding theorem to the following problem.

Let Ω be a compact region in 3-space such that Green's function for the first boundary problem of potential theory exists. It is then well known that this kernel $G(s, t)$ satisfies condition 1°. of our theorem. It $f(s; u, p_i)$ is Hölder continuous with respect to all variables, then the equation

$$\lambda u(s) = \int_\Omega G(s, t) f\left(t; u, \frac{\partial u}{\partial x_i}\right) dt$$

is equivalent to the boundary problem

$$(*) \qquad \Delta u + \lambda^{-1} f\left(s; u, \frac{\partial u}{\partial x_i}\right) = 0, u = 0 \text{ on } \partial\Omega .$$

Hence, if f satisfies condition 2°., we have:

The nonlinear boundary problem (*) *has, for suitable values* $\lambda > 0$, *solutions* $u \geqq 0$ *such that* max u *attains any positive real number. If, moreover,* $f = u + g(s; u, p_i)$ $(g \geqq 0)$, *then for each such* λ

$$\lambda(u) \geqq \lambda_1 ,$$

where λ_1 *is the largest eigenvalue of the corresponding linear problem.*

REFERENCES

1. N. Bourbaki, *Élém. de mathém.*: Espaces vectoriels topologiques Chap. I, II. Act. Sci. et ind. 1189 (Paris 1953).

2. M. G. Krein and M. A. Rutman, *Linear operators leaving invariant a cone in a Banach space* (Russian), Uspehi mat. Nauk (N.S.) 3, No. 1 (23), 3-95 (1943). English translation: Amer. Math. Soc. Translat. 26.

3. D. Morgenstern, *Thesis* (T.U. Berlin 1952, quoted in [**6**], p. 249).

4. H. Schaefer, *Positive Transformationen in halbgeordneten lokalkonvexen Vektorräumen*, Math. Annalen **129** (1955), 323-329.

5. ———, *Halbgeordnete lokalkonvexe Vektorräume*, Math. Annalen **135** (1958), 115-141.

6. W. Schmeidler, *Algebraische Intgralgleichungen II*, Math. Nachr. **10** (1953), 247-255.

7. ———, *Über symmetrische algebraische Integralgleichungen*, Annales Acad. Scient. Fennicae, Ser A, I (220) (1956).

8. A. Tychonov, *Ein Fixpunktsatz*, Math. Annalen **117** (1935), 767-776.

9. E. Rothe, *On non-negative functional transformations*, Amer. J. Math., **66** (1944), 245-254.

THE STATE COLLEGE OF WASHINGTON

SOME CONNECTIONS BETWEEN CONTINUED FRACTIONS AND CONVEX SETS

ROBERT SEALL AND MARION WETZEL

The purpose of this paper is to develop certain connections between the continued fraction solutions and the convex set solutions to some of the moment problems. In particular, we shall develop some relations between the work of Wall [3], [4] on continued fractions and the work of Karlin and Shapley [1] on convex sets. The paper is divided into two parts:

I. Stieltjes-type continued fractions and convex sets.

II. Jacobi-type continued fractions and convex sets.

Two characterizations of the moment problem for the interval (0, 1), one by Riesz [2] in terms of convex closures and one in term of Hankel forms, are well known. The work of Karlin and Shapley [1] shows the equivalence of these two characterizations. A third characterization in terms of a Stieltjes-type continued fraction has been given by Wall [3], [4]. In part I we give an interpretation of the parameters in this continued fraction in terms of "distances" in certain convex bodies. This interpretation, through the work of Karlin and Shapley, immediately shows the equivalence of all three characterizations.

Solutions of the moment problem for the interval $(-1, 1)$, in terms of the Riesz condition and Hankel forms, are also well known. In part II we give a third solution in terms of a Jacobi-type continued fraction. Again, through an interpretation of the parameters in this continued fraction in terms of "distances" in certain convex bodies and an extension of the work of Karlin and Shapley, the equivalence of the three characterizations is immediate.

I. STIELTJES-TYPE CONTINUED FRACTIONS AND CONVEX SETS

1. The monotone Hausdorff moment problem. A sequence of real numbers $\{c_n\}(n = 0, 1, 2, \cdots)$ is called a monotone Hausdorff moment sequence if there exists a monotone nondecreasing real function $\phi(u)$, $0 \leqq u \leqq 1$, such that

$$c_n = \int_0^1 u^n d\phi(u), \qquad n = 0, 1, 2, \cdots .$$

Received December 8, 1958. The authors wish to acknowledge several helpful suggestions from Professor Walter T. Scott of Northwestern University.

The problem of determining such a function $\phi(u)$ is known as the monotone Hausdorff moment problem. We shall assume throughout part I unless otherwise designated that $c_0 = 1$.

Wall [3], [4] has shown that a sequence $\{c_n\}$ is a monotone Hausdorff moment sequence if and only if the power series

$$P(z) = \sum_{n=0}^{\infty} c_n z^n$$

has a continued fraction expansion of the form

$$\frac{1}{1} - \frac{(1-g_0)g_1 z}{1} - \frac{(1-g_1)g_2 z}{1} - \cdots, \tag{1.1}$$

where $0 \leqq g_p \leqq 1$, $p = 0, 1, 2, \cdots$. We shall agree that the continued fraction terminates with the first identically vanishing partial quotient. The sequence $\{(1-g_{p-1})g_p\}$ $(p = 1, 2, 3, \cdots)$ is called a *chain sequence* and the numbers g_p are called the *parameters* of the chain sequence. In general the parameters are not uniquely determined and we designate the minimal set of parameters by m_p. In this case $m_0 = 0$ and (1.1) takes the form

$$\frac{1}{1} - \frac{m_1 z}{1} - \frac{(1-m_1)m_2 z}{1} - \frac{(1-m_2)m_3 z}{1} - \cdots. \tag{1.2}$$

Riesz [2], [1], [3] proved that a sequence $\{c_n\}$ is a monotone Hausdorff moment sequence if and only if the point $(c_1, c_2, \cdots, c_n)$, $n = 1, 2, 3, \cdots$, is in the convex closure of the arc whose parametric equations are

$$\begin{aligned} x_1 &= t, \\ x_2 &= t^2, \\ &\cdots \\ x_n &= t^n, \ 0 \leqq t \leqq 1. \end{aligned} \tag{1.3}$$

The geometry of these convex bodies is developed rather fully in the work of Karlin and Shapley [1].

2. The connecting theorem. Before stating the theorem which connects continued fractions with convex bodies it is necessary to indicate some special notations for the Hankel determinants. We set

$$\Delta_{2n} = \begin{vmatrix} 1 & c_1 & \cdots & c_n \\ c_1 & c_2 & \cdots & c_{n+1} \\ & & \cdots & \\ c_n & c_{n+1} & \cdots & c_{2n} \end{vmatrix}, \qquad n = 0, 1, 2, \cdots, \tag{2.1}$$

$$(2.2)\qquad \underline{\Delta}_{2n+1} = \begin{vmatrix} c_1 & c_2 & \cdots & c_{n+1} \\ c_2 & c_3 & \cdots & c_{n+2} \\ & & \cdots & \\ c_{n+1} & c_{n+2} & \cdots & c_{2n+1} \end{vmatrix}, \qquad n = 0, 1, 2, \cdots, \quad (\underline{\Delta}_{-1} = 1),$$

$$(2.3)\qquad \overline{\Delta}_{2n} = \begin{vmatrix} c_1 - c_2 & c_2 - c_3 & \cdots & c_n - c_{n+1} \\ c_2 - c_3 & c_3 - c_4 & \cdots & c_{n+1} - c_{n+2} \\ & & \cdots & \\ c_n - c_{n+1} & c_{n+1} - c_{n+2} & \cdots & c_{2n-1} - c_{2n} \end{vmatrix}, \qquad n = 1, 2, 3, \cdots, \quad (\overline{\Delta}_0 = 1), \text{ and}$$

$$(2.4)\qquad \overline{\Delta}_{2n+1} = \begin{vmatrix} 1 - c_1 & c_1 - c_2 & \cdots & c_n - c_{n+1} \\ c_1 - c_2 & c_2 - c_3 & \cdots & c_{n+1} - c_{n+2} \\ & & \cdots & \\ c_n - c_{n+1} & c_{n+1} - c_{n+2} & \cdots & c_{2n} - c_{2n+1} \end{vmatrix}, \qquad n = 0, 1, 2, \cdots, \quad (\overline{\Delta}_{-1} = 1).$$

It is well known that a sequence $\{c_n\}$ is a monotone Hausdorff moment sequence if and only if the Hankel forms

$$\sum_{i,j=0}^{n} c_{i+j}x_ix_j, \qquad \sum_{i,j=0}^{n} c_{i+j+1}x_ix_j,$$

$$\sum_{i,j=0}^{n-1} (c_{i+j+1} - c_{i+j+2})x_ix_j, \qquad \sum_{i,j=0}^{n} (c_{i+j} - c_{i+j+1})x_ix_j$$

are all positive semidefinite. In (2.1) replace c_{2n} by $\underline{c}_{2n}$, and in (2.2) replace c_{2n+1} by $\underline{c}_{2n+1}$. Setting $\underline{\Delta}_{2n}$ and $\underline{\Delta}_{2n+1}$ equal to zero, we have the single relation

$$(2.5)\qquad \underline{c}_n = c_n - \frac{\underline{\Delta}_n}{\underline{\Delta}_{n-2}}, \qquad n = 1, 2, 3, \cdots,$$

provided $\underline{\Delta}_{n-2} \neq 0$. Similarly, (2.3) and (2.4) yield

$$(2.6)\qquad \bar{c}_n = c_n + \frac{\overline{\Delta}_n}{\overline{\Delta}_{n-2}}, \qquad n = 1, 2, 3, \cdots,$$

provided $\overline{\Delta}_{n-2} \neq 0$. If the sequence $\{c_n\}$ is a monotone Hausdorff moment sequence, then the quantities $\underline{c}_n$ and $\bar{c}_n$ have been interpreted as the "downward" and "upward" projections, respectively, of c_n on the boundary of the corresponding convex body [1].

We can now state the following theorem:

THEOREM 2.1. *If the sequence $\{c_n\}$ is a monotone Hausdorff moment sequence, then the elements and the minimal parameters in the continued fraction* (1.2) *can be written in the forms*

$$(2.7) \qquad a_n = (1 - m_{n-1})m_n = \frac{c_n - \underline{c}_n}{c_{n-1} - \underline{c}_{n-1}}, \qquad n = 1, 2, 3, \cdots, (\underline{c}_0 = 0)\,,$$

and

$$(2.8) \qquad m_n = \frac{c_n - \underline{c}_n}{\bar{c}_n - \underline{c}_n}, \quad 1 - m_n = \frac{\bar{c}_n - c_n}{\bar{c}_n - \underline{c}_n}, \qquad n = 1, 2, 3, \cdots .$$

From the proof it will be clear that a more general theorem is true. If $\{c_n\}$, $(c_0 = 1)$, is an arbitrary sequence of real numbers and its corresponding Stieltjes-type continued fraction is written in the form (1.2), where no longer it is necessary that $0 \leqq m_n \leqq 1$, $n = 1, 2, 3, \cdots$, the relations (2.7) and (2.8) are still valid.

If $\{c_n\}$ is a monotone Hausdorff moment sequence, then the m_n can be interpreted as the ratio of the "distance" of c_n to the lower boundary to the "distance" between the upper and lower boundaries of the corresponding convex body. Similar interpretations can be given to the a_n and $(1 - m_n)$. By Theorem 2.1 the equivalence of the condition in terms of Hankel forms and Wall's characterization in terms of the continued fraction (1.2), for the existence of a monotone Hausdorff moment sequence, is apparent.

Proof. The proof depends upon the following lemma:

LEMMA 2.1. *The determinants in* (2.1), (2.2), (2.3), *and* (2.4) *satisfy the relation*

$$(2.9) \qquad \underline{\Delta}_k \bar{\Delta}_k = \bar{\Delta}_{k+1}\underline{\Delta}_{k-1} + \underline{\Delta}_{k+1}\bar{\Delta}_{k-1}, \qquad k = 1, 2, 3, \cdots .$$

We shall indicate two proofs to this lemma.

Proof (1). By a substitution and an equivalence transformation, we write the continued fraction (1.2) in the form

$$(2.10) \qquad \frac{1}{z} - \frac{a_1}{1} - \frac{a_2}{z} - \frac{a_3}{1} - \cdots,$$

where $a_k = (1 - m_{k-1})m_k$, $k = 1, 2, 3, \cdots$, $(m_0 = 0)$. The recurrence formulas for the denominators of the continued fraction (2.10) are given by

$$(2.11) \qquad B_{2k}(z) = B_{2k-1}(z) - a_{2k-1}B_{2k-2}(z), \qquad k = 1, 2, 3, \cdots, \quad (B_0(z) = 1)\,,$$

and

$$(2.12) \qquad B_{2k+1}(z) = zB_{2k}(z) - a_{2k}B_{2k-1}(z), \quad k = 0, 1, 2, \cdots, \quad (a_0 = 1,\ B_{-1}(z) = 0,\ B_0(z) = 1)\,.$$

Furthermore, we have

$$B_{2k}(z) = \frac{\Delta_{2k}(z)}{\Delta_{2k-2}}, \qquad k = 1, 2, 3, \cdots, \tag{2.13}$$

and

$$B_{2k+1}(z) = \frac{z\Delta_{2k+1}(z)}{\Delta_{2k-1}}, \qquad k = 0, 1, 2, \cdots, \tag{2.14}$$

where Δ_{2k-2} and Δ_{2k-1} are obtained from (2.1) and (2.2), respectively, and we define

$$\Delta_{2k}(z) = \begin{vmatrix} 1 & c_1 & \cdots & c_{k-1} & 1 \\ c_1 & c_2 & \cdots & c_k & z \\ c_2 & c_3 & \cdots & c_{k+1} & z^2 \\ & & \cdots & & \\ c_k & c_{k+1} & \cdots & c_{2k-1} & z^k \end{vmatrix}, \qquad k = 1, 2, 3, \cdots, \tag{2.15}$$

and

$$\Delta_{2k+1}(z) = \begin{vmatrix} c_1 & c_2 & \cdots & c_k & 1 \\ c_2 & c_3 & \cdots & c_{k+1} & z \\ c_3 & c_4 & \cdots & c_{k+2} & z^2 \\ & & \cdots & & \\ c_{k+1} & c_{k+2} & \cdots & c_{2k} & z^k \end{vmatrix}, \qquad k = 1, 2, 3, \cdots, \quad (\Delta_1(z) = 1). \tag{2.16}$$

By a sequence of elementary operations on $\Delta_{2k}(z)$ and $\Delta_{2k+1}(z)$ it is seen that $\Delta_k(1) = \bar{\Delta}_{k-1}$, $k = 1, 2, 3, \cdots$.
Substituting this result in (2.13) and (2.14) we have

$$B_k(1) = \frac{\bar{\Delta}_{k-1}}{\Delta_{k-2}}, \qquad k = 1, 2, 3, \cdots. \tag{2.17}$$

We also note that

$$a_k = \frac{\Delta_{k-3}\Delta_k}{\Delta_{k-2}\Delta_{k-1}}, \qquad k = 1, 2, 3, \cdots, \quad (\Delta_{-2} = 1). \tag{2.18}$$

Substituting the results of (2.17) and (2.18) in (2.11) and (2.12), the relation (2.9) follows immediately.

Proof (2). By Laplace's Development and a sequence of elementary operations, Lemma 2.1 can be established directly. We shall omit the details.

The proof of Theorem 2.1 now follows. Using (2.5) and (2.18), the relation (2.7) is immediate.

The relation (2.8) is established by induction. Assume that

$0 \leqq m_n < 1$, $n = 1, 2, 3, \cdots$. Using (2.5), (2.6), and Lemma 2.1 it is clear that

$$m_1 = a_1 = \frac{c_1 - \underline{c}_1}{c_0 - \underline{c}_0} = \frac{c_1 - \underline{c}_1}{\bar{c}_1 - \underline{c}_1} ,$$

where $c_0 = 1$ and we define $\underline{c}_0$ to be zero, and $m_2 = \dfrac{c_2 - \underline{c}_2}{\bar{c}_2 - \underline{c}_2}$. Now assume that $m_k = \dfrac{c_k - \underline{c}_k}{\bar{c}_k - \underline{c}_k}$. Again using (2.5), (2.6), and Lemma 2.1 in the relation $m_{k+1} = \dfrac{a_{k+1}}{1 - m_k}$, the definition for the minimal parameters in a chain sequence [3], the induction is completed. If $m_k = 1$ then m_{k+1} is defined to be zero. In this case the corresponding moments fall on the upper and lower boundaries of their respective convex bodies.

3. Some results from the theory of chain sequences.

Regarding the uniqueness of the parameters g_p in the continued fraction (1.1) and the location of the moments in the convex bodies we have the following theorem:

THEOREM 3.1. *Given a monotone Hausdorff moment sequence,* $\{c_n\}$, *let*

$$\lim \frac{c_k - \underline{c}_k}{\bar{c}_k - c_k} = q . \tag{3.1}$$

If $q > 1$ the parameters g_p in (1.1) are uniquely determined, and if $q < 1$ the parameters are not uniquely determined. In case $q = 1$ the parameters may or may not be unique.

Proof. Wall [3] proved that the parameters in a chain sequence are uniquely determined if and only if the series

$$1 + \sum_{k=1}^{\infty} \frac{m_1 m_2 \cdots m_k}{(1 - m_1)(1 - m_2) \cdots (1 - m_k)}$$

diverges. Making use of this result and Theorem 2.1 our proof is immediate.

We designate the maximal parameters of the chain sequence in the continued fraction (1.1) by M_p. The maximal parameters can be interpreted in terms of "distances" in the convex bodies by the following theorem:

THEOREM 3.2. *The maximal parameters* M_n *in the continued fraction* (1.1) *can be written in the form*

$$M_n = \frac{c_n - \underline{c}_n}{\bar{c}_n - \underline{c}_n} + \frac{\bar{c}_n - c_n}{\bar{c}_n - \underline{c}_n}(1/T_n), \qquad n = 1, 2, 3, \cdots, \tag{3.2}$$

where

$$T_n = 1 + \sum_{r=n+1}^{\infty} \frac{c_{n+1} - \underline{c}_{n+1}}{\bar{c}_{n+1} - c_{n+1}} \frac{c_{n+2} - \underline{c}_{n+2}}{\bar{c}_{n+2} - c_{n+2}} \cdots \frac{c_r - \underline{c}_r}{\bar{c}_r - c_r}, \tag{3.3}$$

in the case that the a_{n+r}, $r = 1, 2, 3, \cdots$, of (2.9) are positive. If $a_{n+1}, a_{n+2}, \cdots, a_{n+r}$ are positive, $a_{n+k+1} = 0$, $(k > 0)$, and $m_{n+k} < 1$, then the summation in (3.3) runs only to $n + k$.

Proof. Wall [3] introduced an expression of the form (3.3) in discussing maximal parameters. Using his results and Theorem 2.1 our proof is immediate.

II. JACOBI-TYPE CONTINUED FRACTIONS AND CONVEX SETS

4. The "extended" monotone Hausdorff moment problem. A sequence of real number $\{c_n\}(n = 0, 1, 2, \cdots)$ shall be referred to as an "extended" monotone Hausdorff moment sequence if there exists a monotone nondecreasing real function $\phi(u)$, $-1 \leqq u \leqq 1$, such that

$$c_n = \int_{-1}^{1} u^n d\phi(u), \qquad n = 0, 1, 2, \cdots.$$

The problem of determining such a function $\phi(u)$ shall be referred to as the "extended" monotone Hausdorff moment problem. Again we shall assume throughout part II unless otherwise designated that $c_0 = 1$.

The work of Riesz [2] can be applied to the "extended" monotone Hausdorff moment problem. A sequence $\{c_n\}$ is an "extended" monotone Hausdorff moment sequence if and only if the point $(c_1, c_2, \cdots, c_n)$, $n = 1, 2, 3, \cdots$, is in the convex closure of the arc whose parametric equations are given by (1.3) where $-1 \leqq t \leqq 1$.

Let

$$\frac{1}{b_1 z + 1} - \frac{a_1 z^2}{b_2 z + 1} - \frac{a_2 z^2}{b_3 z + 1} - \cdots \tag{4.1}$$

be the Jacobi-type continued fraction expansion of the power series

$$P(z) = \sum_{n=0}^{\infty} c_n z^n.$$

We shall agree that the continued fraction terminates with the first identically vanishing partial quotient. We shall show that if the sequence $\{c_n\}$ is an "extended" monotone Hausdorff moment sequence, then the

a_p and b_p of (4.1) have the form of a generalized chain sequence and the parameters can again be represented in terms of "distances" in certain convex bodies.

5. The connecting theorem. As in §2, it is necessary to indicate some special notations for the Hankel determinants corresponding to an "extended" monotone Hausdorff moment sequence. We set

$$\underline{\Delta}_{2n} = \begin{vmatrix} 1 & c_1 & \cdots c_n \\ c_1 & c_2 & \cdots c_{n+1} \\ & \cdots & \\ c_n & c_{n+1} & \cdots c_{2n} \end{vmatrix}, \qquad n = 0, 1, 2, \cdots , \tag{5.1}$$

$$\underline{\Delta}_{2n+1} = \begin{vmatrix} 1 + c_1 & c_1 + c_2 & \cdots c_n + c_{n+1} \\ c_1 + c_2 & c_2 + c_3 & \cdots c_{n+1} + c_{n+2} \\ & \cdots & \\ c_n + c_{n+1} & c_{n+1} + c_{n+2} & \cdots c_{2n} + c_{2n+1} \end{vmatrix}, \qquad n = 0, 1, 2, \cdots , \ (\underline{\Delta}_{-1} = 1) , \tag{5.2}$$

$$\overline{\Delta}_{2n} = \begin{vmatrix} 1 - c_2 & c_1 - c_3 & \cdots c_{n-1} - c_{n+1} \\ c_1 - c_3 & c_2 - c_4 & \cdots c_n - c_{n+2} \\ & \cdots & \\ c_{n-1} - c_{n+1} & c_n - c_{n+2} & \cdots c_{2n-2} - c_{2n} \end{vmatrix}, \qquad n = 1, 2, 3, \cdots , \ (\overline{\Delta}_0 = 1), \text{ and} \tag{5.3}$$

$$\overline{\Delta}_{2n+1} = \begin{vmatrix} 1 - c_1 & c_1 - c_2 & \cdots c_n - c_{n+1} \\ c_1 - c_2 & c_2 - c_3 & \cdots c_{n+1} - c_{n+2} \\ & \cdots & \\ c_n - c_{n+1} & c_{n+1} - c_{n+2} & \cdots c_{2n} - c_{2n+1} \end{vmatrix}, \qquad n = 0, 1, 2, \cdots , \ (\overline{\Delta}_{-1} = 1) . \tag{5.4}$$

The sequence $\{c_n\}$ is an "extended" monotone Hausdorff moment sequence if and only if the Hankel forms

$$\sum_{i,j=0}^{n} c_{i+j}x_ix_j, \quad \sum_{i,j=0}^{n} (c_{i+j} + c_{i+j+1})x_ix_j ,$$
$$\sum_{i,j=0}^{n-1} (c_{i+j} - c_{i+j+2})x_ix_j , \quad \sum_{i,j=0}^{n} (c_{i+j} - c_{i+j+1})x_ix_j$$

are all positive semidefinite. As in part I replace c_{2n} by $\underline{c}_{2n}$ in (5.1) and c_{2n+1} by $\underline{c}_{2n+1}$ in (5.2). Setting $\underline{\Delta}_{2n}$ and $\underline{\Delta}_{2n+1}$ equal to zero, we have the single relation

$$\underline{c}_n = c_n - \frac{\underline{\Delta}_n}{\underline{\Delta}_{n-2}} , \qquad n = 1, 2, 3, \cdots . \tag{5.5}$$

Similarly, (5.3) and (5.4) yield

$$\bar{c}_n = c_n + \frac{\bar{\Delta}_n}{\bar{\Delta}_{n-2}}, \qquad n = 1, 2, 3, \cdots . \tag{5.6}$$

The methods of Karlin and Shapley [1] can be applied so that if the sequence $\{c_n\}$ is an "extended" monotone Hausdorff moment sequence then the quantities $\underline{c}_n$ and $\bar{c}_n$ of (5.5) and (5.6) are again interpreted as the "downward" and "upward" projections, respectively, of c_n on the boundary of the corresponding convex body.

We can now state the following theorem:

THEOREM 5.1. *If the sequence* $\{c_n\}$ *is an "extended" monotone Hausdorff moment sequence, then the elements* a_n *and* b_n *in the continued fraction* (4.1) *can be written in the forms*

$$a_n = 4m_n(1 - m_{n-1})l_n(1 - l_n), \quad n = 1, 2, 3, \cdots, \quad 0 \leqq m_n \leqq 1, \ (m_0 = 0), \ 0 \leqq l_n \leqq 1 , \tag{5.7}$$

$$= \frac{c_{2n} - \underline{c}_{2n}}{c_{2n-2} - \underline{c}_{2n-2}}, \qquad n = 1, 2, 3, \cdots, (c_0 = 0) , \tag{5.8}$$

where

$$m_n = \frac{c_{2n} - \underline{c}_{2n}}{\bar{c}_{2n} - \underline{c}_{2n}} , \qquad l_n = \frac{c_{2n-1} - \underline{c}_{2n-1}}{\bar{c}_{2n-1} - \underline{c}_{2n-1}} , \tag{5.9}$$

$$1 - m_n = \frac{\bar{c}_{2n} - c_{2n}}{\bar{c}_{2n} - \underline{c}_{2n}}, \quad 1 - l_n = \frac{\bar{c}_{2n-1} - c_{2n-1}}{\bar{c}_{2n-1} - \underline{c}_{2n-1}} , \ n=1, 2, 3, \cdots ,$$

and

$$b_n = 1 - 2m_{n-1}(1 - l_{n-1}) - 2(1 - m_{n-1})l_n, \qquad n = 1, 2, 3, \cdots , \ (l_0 = m_0 = 0) , \tag{5.10}$$

$$= 1 - \frac{c_{2n-1} - \underline{c}_{2n-1}}{c_{2n-2} - \underline{c}_{2n-2}} - \frac{c_{2n-2} - \underline{c}_{2n-2}}{c_{2n-3} - \underline{c}_{2n-3}}, \qquad n = 2, 3, 4, \cdots . \tag{5.11}$$

As in part I it will be clear that a more general theorem is true. If $\{c_n\}$, $(c_0 = 1)$, is an arbitrary sequence of real numbers and its corresponding Jacobi-type continued fraction (4.1) is written in the form that the a_n and b_n are given by (5.7) and (5.10), respectively, where $l_0 = m_0 = 0$ but it is no longer necessary that $0 \leqq l_n \leqq 1$ and $0 \leqq m_n \leqq 1$, $n = 1, 2, 3, \cdots$, then the relations (5.8) and (5.11) with (5.9) holding are still valid.

If $\{c_n\}$ is an "extended" monotone Hausdorff moment sequence, the geometric interpretations of the a_n, b_n, l_n, and m_n are apparent.

Proof. The proof depends upon the following lemma.

LEMMA 5.1. *The determinants in* (5.1), (5.2), (5.3), *and* (5.4) *satisfy the relations*

$$(5.12)\qquad \begin{aligned} \underline{\Delta}_{2k+1}\bar{\Delta}_{2k+1} &= \bar{\Delta}_{2k+2}\underline{\Delta}_{2k} + \underline{\Delta}_{2k+2}\bar{\Delta}_{2k}, \\ 2\underline{\Delta}_{2k}\bar{\Delta}_{2k} &= \bar{\Delta}_{2k+1}\underline{\Delta}_{2k-1} + \underline{\Delta}_{2k+1}\bar{\Delta}_{2k-1}, \end{aligned} \qquad k = 0, 1, 2, \cdots .$$

Proof. By Laplace's Development and a sequence of elementary operations, Lemma 5.1 can be established directly. We shall omit the details.

The proof to the theorem now follows. A well known formula for the a_k is given by

$$(5.13)\qquad a_k = \frac{\underline{\Delta}_{2k}\underline{\Delta}_{2k-4}}{\underline{\Delta}_{2k-2}\underline{\Delta}_{2k-2}}, \qquad k = 2, 3, 4, \cdots, \left(a_1 = \frac{\underline{\Delta}_2}{\underline{\Delta}_0^2}\right).$$

The formulas (5.5) and (5.13) yield (5.8).

By a substitution and an equivalence transformation, we write the continued fraction (4.1) in the form

$$(5.14)\qquad \frac{1}{b_1 + z} - \frac{a_1}{b_2 + z} - \frac{a_2}{b_2 + z} - \cdots .$$

The recurrence formula for the denominators of the continued fraction (5.14) is given by

$$(5.15)\qquad \begin{aligned} B_k(z) = (b_k + z)B_{k-1}(z) - a_{k-1}B_{k-2}(z), \\ k = 1, 2, 3, \cdots, (a_0 = 1,\ B_{-1}(z) = 0,\ B_0(z) = 1) . \end{aligned}$$

Furthermore, we have

$$(5.16)\qquad B_k(z) = \frac{\underline{\Delta}_{2k}(z)}{\underline{\Delta}_{2k-2}}, \qquad k = 1, 2, 3, \cdots ,$$

where $\underline{\Delta}_{2k-2}$ is obtained from (5.1) and we define $\underline{\Delta}_{2k}(z)$ the same as in (2.15). By a sequence of elementary operations on $\underline{\Delta}_{2k}(z)$ it is seen that $\underline{\Delta}_{2k}(-1) = (-1)^k\underline{\Delta}_{2k-1}$, $k = 1, 2, 3, \cdots$. Substituting this result in (5.16) we have

$$(5.17)\qquad B_k(-1) = \frac{(-1)^k\underline{\Delta}_{2k-1}}{\underline{\Delta}_{2k-2}}, \qquad k = 1, 2, 3, \cdots .$$

Setting z equal to -1 in (5.15), using the formulas (5.13) and (5.17), we can solve for b_k and obtain (5.11). We note that if we had set z equal to 1 and followed a similar procedure, we would have obtained the formula

$$(5.18)\qquad b_n = \frac{\bar{c}_{2n-1} - c_{2n-1}}{c_{2n-2} - \underline{c}_{2n-2}} + \frac{c_{2n-2} - \underline{c}_{2n-2}}{\bar{c}_{2n-3} - c_{2n-3}} - 1, \qquad n = 2, 3, 4, \cdots .$$

Assume that $0 \leq m_n < 1$, $0 \leq l_n < 1$, $n = 1, 2, 3, \cdots$. Using (5.5), (5.6), and (5.9) it can be shown directly that $b_1 = 1 - 2l_1$, and $a_1 = 4m_1l_1(1 - l_1)$. Now by using (5.5), (5.6), (5.9), and Lemma 5.1, (5.11) reduces to (5.10) for $n = k$, $k = 2, 3, 4, \cdots$. A similar statement applies to (5.8). If $m_k = 1$ then m_{k+1} is defined to be zero. A similar statement applies to l_k. In either case the corresponding moments fall on the upper and lower boundaries of their respective convex bodies.

If (5.18) had been used in place of (5.11) we note that (5.10) would have been obtained in the form

$$(5.19) \qquad b_n = 2(1 - l_n)(1 - m_{n-1}) + 2m_{n-1}l_{n-1} - 1,$$
$$n = 1, 2, 3, \cdots, (l_0 = m_0 = 0) .$$

By Theorem 5.1 and the condition in terms of Hankel forms, we can now state a theorem which characterizes the existence of an "extended" monotone Hausdorff moment sequence in terms of continued fractions. This theorem is analogous to Wall's solution [3], [4] for the regular monotone Hausdorff moment sequence. By Theorem 5.1 and an extension of the work of Karlin and Shapley, the equivalence of the continued fraction solution and the condition in terms of Hankel forms, and hence convex bodies, is apparent.

THEOREM 5.2. *The sequence* $\{c_n\}$ *is an "extended" monotone Hausdorff moment sequence if and only if the power series*

$$P(z) = \sum_{n=0}^{\infty} c_n z^n$$

has a Jacobi-type continued fraction (4.1) *expansion where the* a_n *and* b_n *are given by* (5.7) *and* (5.10), *respectively, and* $l_0 = m_0 = 0$, *and* $0 \leqq l_n \leqq 1$, $0 \leqq m_n \leqq 1$, $n = 1, 2, 3, \cdots$.

It should be pointed out that $P(z) = \sum_{m=0}^{\infty} c_m z^m$ is a moment generating function for the "extended" monotone Hausdorff moment problem if and only if $Q(w) = (1 + z)P(z)$, where $w = \dfrac{2z}{1+z}$, is a moment generating function for the regular monotone Hausdorff moment problem. From these relations it is observed that the l_n and m_n of Theorem 5.1 are equal to m_{2n-1} and m_{2n}, $n = 1, 2, 3, \cdots$, respectively, of Theorem 2.1. These results are obtained by contraction.

It can also be noted that $\{c_n\}$ is an "extended" monotone Hausdorff moment sequence if and only if

$$\{d_n/2^n\}, \quad d_n = \sum_{j=0}^{n} \binom{n}{n-j} c_j ,$$

is a regular monotone Hausdorff moment sequence. This result can be obtained by comparing coefficients in $P(z)$ and $Q(w)$ under the indicated transformation.

6. The continued fraction of the first differences. We prove the following theorem:

THEOREM 6.1. *If*

$$(6.1)\qquad 1 + c_1 z + c_2 z^2 + \cdots \sim \frac{1}{b_1 z + 1} - \frac{a_1 z^2}{b_2 z + 1} - \frac{a_2 z^2}{b_3 z + 1} - \cdots ,$$

where

$$(6.2)\qquad b_n = 1 - 2m_{n-1}(1 - l_{n-1}) - 2(1 - m_{n-1})l_n,\ n = 1, 2, 3, \cdots ,$$

$$(6.3)\qquad a_n = 4m_n(1 - m_{n-1})l_n(1 - l_n),\ n = 1, 2, 3, \cdots, (l_0 = m_0 = 0) ,$$

then

$$(6.4)\qquad \Delta c_0 + \Delta c_1 z + \Delta c_2 z^2 + \cdots \sim \frac{a_0^*}{b_1^* z + 1} - \frac{a_1^* z^2}{b_2^* z + 1} - \frac{a_2^* z^2}{b_3^* z + 1} - \cdots ,$$

where $\Delta c_n = c_{n+1} - c_n$, $n = 1, 2, 3, \cdots$, $(\Delta c_0 = 1 - c_1)$, and

$$(6.5)\qquad \begin{aligned} b_1^* &= 1 - 2l_1(1 - m_1) , \\ b_n^* &= 1 - 2m_{n-1}(1 - l_n) - 2l_n(1 - m_n), \qquad n = 2, 3, 4, \cdots , \end{aligned}$$

$$(6.6)\qquad \begin{aligned} a_0^* &= 2(1 - l_1) , \\ a_n^* &= 4l_n(1 - l_{n+1})m_n(1 - m_n), \qquad n = 1, 2, 3, \cdots . \end{aligned}$$

Proof. In order to prove the theorem it is necessary to note some determinants for the sequence $\{\Delta c_n\}$ corresponding to $\underline{\Delta}_{2n}$ and $\underline{\Delta}_{2n+1}$ of (5.1) and (5.2), respectively, for the sequence $\{c_n\}$. Noting (5.3) and (5.4) we observe that

$$(6.7)\qquad \underline{\Delta}_{2k}^* = \bar{\Delta}_{2k+1},\ \underline{\Delta}_{2k+1}^* = \bar{\Delta}_{2k+2}, \qquad k = 0, 1, 2, \cdots .$$

We observe directly that

$$a_0^* = 1 - c_1 = \Delta c_1 = 2(1 - l_1) .$$

Using (5.13) and (6.7) we note that

$$(6.8)\qquad a_k^* = \frac{\underline{\Delta}_{2k}^* \ \underline{\Delta}_{2k-4}^*}{\underline{\Delta}_{2k-2}^* \underline{\Delta}_{2k-2}^*} = \frac{\bar{\Delta}_{2k+1}\Delta_{2k-3}}{\bar{\Delta}_{2k-1}\bar{\Delta}_{2k-1}} .$$

The relations in (6.6) can now be established by (5.5) (5.6), (5.9), and Lemma 5.1.

Now, by (5.5), (5.10), (5.11), and (6.7),

$$(6.9)\qquad b_k^* = 1 - \frac{\underline{\Delta}_{2k-1}^* \underline{\Delta}_{2k-4}^*}{\underline{\Delta}_{2k-2}^* \underline{\Delta}_{2k-3}^*} - \frac{\underline{\Delta}_{2k-2}^* \underline{\Delta}_{2k-5}^*}{\underline{\Delta}_{2k-3}^* \underline{\Delta}_{2k-4}^*}$$

$$= 1 - \frac{\bar{\Delta}_{2k} \quad \bar{\Delta}_{2k-3}}{\bar{\Delta}_{2k-1} \bar{\Delta}_{2k-2}} - \frac{\bar{\Delta}_{2k-1} \bar{\Delta}_{2k-4}}{\bar{\Delta}_{2k-2} \bar{\Delta}_{2k-3}} ,$$
$$k = 1, 2, 3, \cdots , \ (\underline{\Delta}^*_{-1} = \underline{\Delta}^*_{-2} = 1, \ \underline{\Delta}^*_{-3} = 0) .$$

Now again by (5.5), (5.6), (5.9), and Lemma 5.1, the relations in (6.5) follow.

We note that a similar proof could be given for the corresponding theorem for a regular monotone Hausdorff moment sequence, thereby giving another proof to this well known result [**4**].

REFERENCES

1. S. Karlin and L. S. Shapley, *Geometry of moment spaces*, Mem. Amer. Math. Soc. no. 12 (1953).
2. F. Riesz, *Sur certain systems singuliérs d'equations intégrales*, Ann. de l'Éc. Nor., (3) **28** (1911), 34–62.
3. H. S. Wall, *Analytic theory of continued fractions*, Van Nostrand, 1948.
4. H. S. Wall, *Continued fractions and totally monotone sequences*, Trans. Amer. Math. Soc., **48** (1940), 165-184.

THE OHIO STATE UNIVERSITY
DENISON UNIVERSITY

VARIATIONS ON A THEME OF CHEVALLEY

ROBERT STEINBERG

1. Introduction. In this paper we use the methods of C. Chevalley to construct some simple groups and to gain for them the structural theorems of [3]. Among the groups obtained there are two new families of finite simple groups[1], not to be found in the list of E. Artin [1]. Whether the infinite groups constructed are new has not been settled yet.

Section 5 contains statements of the main results of [3]. In §§ 2, 3, 4 and 7, we define analogues of certain real forms of the Lie groups of type A_l, D_l and E_6 (in the usual notation), and extend to them the structural properties of the groups of Chevalley. Sections 6 and 9 treat some identifications, and § 8 deals with the question of simplicity. In §§ 10 and 11, using the extra symmetry inherent in a Lie algebra of type D_4, we consider two modifications of the first construction which are, perhaps, of more interest since they produce groups which have no analogue in the classical complex-real case: in fact, a basic ingredient of each of these variants is a field automorphism of order 3. In Sections 12 and 13, it is proved that new finite simple groups are obtained[1], and their orders are given. Section 14 deals with an application to the theory of group representations, and § 15 with some concluding observations.

The notation is cumulative. We denote by $|S|$ the cardinality of the set S, by K^* the multiplicative group of the field K, and by C the complex field. An introduction to the standard Lie algebra terminology together with statements of the principal results in the classical theory can be found in [3, p. 15–19]. (Proofs are available in [8] or [10]).

2. Roots and reflections. We first introduce some notations. Relative to a Cartan decomposition of a simple complex Lie algebra of rank l, let E be the real space generated by the roots, made into an Euclidean space in the usual way, and normalized as in [3, p. 17–18]. Relative to an ordering $\prec$ of the additive group generated by the roots, let Π be the set of positive roots, and $a(1), a(2), \cdots, a(l)$ the fundamental roots. For each root $r = \Sigma z_i a(i)$, set $\Sigma z_i = ht\ r$, the *height* of r. The ordering $\prec$ can always be chosen so that $ht\ r < ht\ s$ implies $r \prec s$ (see [3, p. 20, *l*. 35–40]); suppose this is done. Assume now the existence of an automorphism σ of E of order 2 such that $\sigma\Pi = \Pi$. This restricts the type of algebra to A_l, D_l $(l \geqq 4)$ or E_6 (see [3, p. 18]), and hence

Received October 2, 1958, in revised form January 8, 1959.

[1] Since the preparation of this paper, the author has learned that these groups have also been discovered by D. Hertzig [**6**], who has shown that they complete the list of finite simple algebraic groups.

implies that all roots have the same length. We also denote σr by $\bar{r}$. Clearly σ permutes the fundamental roots. Thus $ht\,\bar{r} = ht\,r$ for each root r. Finally, let W be the Weyl group, W^1 the subgroup of elements commuting with σ, and for each $w \in W$ denote by $n(w)$ the number of roots r for which $r \succ 0$ and $wr \prec 0$.

Consider now subsets S of Π of the following three types:

(1) S consists of one root r, which is self-conjugate ($\bar{r} = r$), and which can not be written as a sum of a conjugate pair of roots;

(2) S consists of a conjugate pair r, $\bar{r}$ such that $r + \bar{r}$ is not a root;

(3) S consists of three roots of the form r, $\bar{r}$, $r + \bar{r}$.

Note that in case (2) one has $r \perp \bar{r}$ because $ht\,r = ht\,\bar{r}$ implies that $r - \bar{r}$ is not a root. Shortly we prove the important fact:

2.1 LEMMA. *If Π^1 denotes the collection of sets of types* (1), (2) *and* (3) *above, then Π^1 is a partition of Π.*

In any case, the *fundamental sets* of Π^1 - those which contain fundamental roots - are disjoint because the fundamental roots are linearly independent. If w_r denotes the reflection in the hyperplane orthogonal to r, we set $w_s = w_r$, $w_r w_{\bar{r}}$ or $w_{r+\bar{r}} (= w_r w_{\bar{r}} w_r)$ according as S is of type (1), (2) or (3) above. Note that $w_s \in W^1$.

2.2 LEMMA. *For each fundamental $S \in \Pi^1$, w_s maps S onto $-S$ and permutes the positive roots not in S. Hence $n(w_s) = |S|$.*

Proof. Since $n(w_a) = 1$ for each fundamental root a [8, p. 19-01, Lemma 1], and since w_s can be written as a product of $|S|$ such reflections, it follows that $n(w_s) \leqq |S|$. By direct verification one sees that $w_s S = -S$. Hence the lemma is proved.

2.3 LEMMA. *The group W^1 is generated by the w_s corresponding to fundamental $S \in \Pi^1$.*

Proof. Using induction on $n(w)$, we show that each $w \in W^1$ is a product of elements of the given form. If $n(w) = 0$, $w = 1$, the statement is clearly true. If $n(w) > 0$, $w \neq 1$, there is a fundamental root a such that $a \succ 0$ and $wa \prec 0$. Since $\bar{a} \succ 0$ and $w\bar{a} = \overline{wa} \prec 0$, it follows that $r \succ 0$, $wr \prec 0$ for each root r in the set $S \in \Pi^1$ which contains a. Hence $n(ww_s^{-1}) = n(w) - n(w_s)$ by 2.2, and the induction hypothesis can be applied to ww_s^{-1} to complete the proof.

2.4 LEMMA. *W is a normal subgroup of the group generated by W and σ.*

Proof. One has $\sigma w_r \sigma^{-1} = w_{\bar{r}}$ for each root r. Since σ permutes the roots, and the root reflections generate W, one gets $\sigma W \sigma^{-1} = W$, and hence 2.4.

2.5 LEMMA. *The element w_0 of W defined by $w_0 \Pi = - \Pi$ is in W^1.*

Proof. By 2.4, $\sigma w_0 \sigma^{-1} \in W$. Since $\sigma w_0 \sigma^{-1} \Pi = - \Pi$, one concludes that $\sigma w_0 \sigma^{-1} = w_0$ and that $w_0 \in W^1$.

2.6 LEMMA. *Each $S \in \Pi^1$ is congruent under W^1 to a fundamental set.*

Proof. Write the element w_0 of 2.5 in the form $w_0 = w_k \cdots w_2 w_1$ guaranteed by 2.3. Since $S \succ 0$ and $w_0 S \prec 0$, there is an index i such that $w_{i-1} \cdots w_1 S \succ 0$ and $w_i \cdots w_1 S \prec 0$. If $T \in \Pi^1$ corresponds to w_i, it follows from 2.2 that $w_{i-1} \cdots w_1 S \subseteq T$, and clearly equality must hold.

By using 2.6 and examinining the fundamental root systems for groups of type A_l, D_l and E_6 (see [3, p. 18] or [8, p. 13-08]), one sees that a set in Π^1 of type (3) can occur only in the case A_l (l even). This turns out to be the most troublesome case in the sequel. Note however that sets of types (1) and (3) do not occur simultaneously.

Proof of 2.1. This follows from 2.6 and the fact that the fundamental sets of Π^1 are non-overlapping.

We now associate with W^1 a reflection group. Let E^+ and E^- respectively denote the positive and negative subspaces of E under σ, and for each $w \in W^1$ let $\tilde{w}$ and $\tilde{W}^1$ denote the restrictions of w and W^1 to E^+. Also denote by $\tilde{S}$ the vector r, $r + \bar{r}$ or $r + \bar{r}$ in the respective cases (1), (2) or (3) of 2.1.

2.7 LEMMA. *The restriction of W^1 to $\tilde{W}^1$ is faithful. $\tilde{W}^1$ is a reflection group of type $C_{[(l+1)/2]}$, B_{l-1} or F_4 in the respective cases that W is of type A_l, D_l or E_6, and, to within a change of scale, $\{\tilde{S} \mid S \in \Pi^1\}$ is a corresponding system of positive root vectors.*

Proof. First if $w \in W^1$, $\tilde{w} = 1$, then w maps each positive root onto another one. Hence $w = 1$, and the restriction is faithful. Those $\tilde{S}$ which correspond to the fundamental $S \in \Pi^1$ form a new fundamental root system (to within a change of scale) of the listed type, as one sees by considering the separate cases (see [3, p. 18]). Becuse of 2.3 and 2.6, the proof is complete if it can be shown that, for each fundamental $S \in \Pi^1$, $\tilde{w}_s$ is the reflection in the hyperplane orthogonal to $\tilde{S}$. If

$|S| = 2$ and $S = \{a, \bar{a}\}$, then w_s has -1 as a characteristic value of multiplicity 2. Since $w_s(a + \bar{a}) = -(a + \bar{a})$, $w_s(a - \bar{a}) = -(a - \bar{a})$, $a + \bar{a} \in E^+$, and $a - \bar{a} \in E^-$, it follows that $\tilde{w}_s$ has -1 as a characteristic value of multiplicity 1, and then that $\tilde{w}_s$ is the required reflection. If $|S| = 1$ or $|S| = 3$, the result follows from the definitions.

2.8 COROLLARY. *Any two sets of the same type in the partition 2.1 are congruent under* W^1.

Proof. Since sets of types (1) and (3) do not occur simultaneously, and since $\tilde{W}^1$ is transitive on its root vectors of a given length, 2.8 follows from 2.7.

A new ordering $<$ of the positive roots is now introduced. First if $R, S \in \Pi^1$, then $R < S$ means that $\min r \in R \prec \min s \in S$. Then if $r, s \in \Pi$, define $r < s$ to mean that either r and s belong to distinct sets R and S of Π^1 and $R < S$, or r and s belong to the same set of Π^1 and $r \prec s$,

2.9 LEMMA. *The roots in each set S of Π^1 occur consecutively in the ordering of the roots of Π relative to $<$. If r, s and $r + s$ are positive roots, then* $r + s > \min(r, s)$.

Proof. The first statement follows from the definition. Since $\prec$ respects heights, the second assertion is true if $r + s$ has minimum height in the set S of Π^1 containing it. Thus one may assume that there is a root t such that $r + s = t + \bar{t}$, $r \neq t$, $r \neq \bar{t}$, and that W is of type A_l (l even). Then each positive root is a sum of a string of distinct fundamental roots, and the strings corresponding to r and s are necessarily of different lengths. Thus $ht\, t = ht\, \bar{t} > \min(ht\ r, ht\ s)$. Since $\prec$ respects heights, this implies that $r + s > \min(t, \bar{t}) > \min(r, s)$.

3. **Construction of an involution.** Suppose that $\mathfrak{g}$ is a simple complex Lie algebra with a generating system $(X_r, X_{-r}, H_r, r \in \Pi)$ chosen to satisfy the conditions of Theorem 1 of [3]. Assume also that $\mathfrak{g}$ is restricted to type A_l, D_l ($l \geqq 4$) or E_6 so that the results of § 2 can be applied. Set $r(H_s) = r(s)$. Then, all roots being of the same length, it follows that:

$$3.1 \qquad X_r X_s = N_{rs} X_{r+s}\,;\ N_{rs} = 0,\ \pm 1;\ r,\ s \in \Pi\ .$$

For the same reason $r(s) = s(r)$ and $r(r) = 2$. By the uniqueness theorem for a simple Lie algebra with a given root structure (see [8, p. 11-04] or [10, p. 94]), there exists an automorphism σ_C of $\mathfrak{g}$ such that $\sigma_C H_r = H_{\bar{r}}$ and $\sigma_C X_r = c_r X_{\bar{r}}$, $c_r \in C^*$, $r \in \Pi$ or $-\Pi$, with $c_a = 1$ for each fundamental root a. Then each $c_{-a} = 1$, and by induction on the height one gets each $c_r = \pm 1$. Next let K be a field on which an automorphism σ of order 2 acts, let K_0 be the fixed field, and write $\sigma k = \bar{k}$, $k \in K$. Then following the procedure of [3, p. 32], one can transfer the base field of $\mathfrak{g}$ from C to K, and thus gain a Lie algebra $\mathfrak{g}_K$ over K and a semi-automorphism σ of $\mathfrak{g}_K$ such that $\sigma(kH_r) = \bar{k}H_{\bar{r}}$ and $\sigma(kX_r) = \pm \bar{k}X_{\bar{r}}$, $k \in K$, $r \in \Pi$ or $-\Pi$. Note [3, p. 32] that the field is not transferred for roots (or weights) and that the expression $r(s)$ retains its original meaning.

3.2 LEMMA. *The order of σ is 2. By appropriate sign changes of the X_r one can arrange things so that in the equations $\sigma X_r = k_r X_{\bar{r}}$, $r \in \Pi$, one has:*

(a) $k_{\bar{r}} = k_r$;

(b) *if $\bar{r} \neq r$, then $k_r = 1$;*

(c) *if $\bar{r} = r$, then k_r is 1 or -1 according as r belongs to an $S \in \Pi^1$ of 1 or 3 elements.*

Proof. One has $\sigma^2 X_a = X_a$, $\sigma^2 X_{-a} = X_{-a}$ for each fundamental root a. Thus $\sigma^2 = 1$, and this implies (a). If r, $\bar{r}$ is a conjugate pair in Π, if $r < \bar{r}$, and if $k_r = -1$, replace X_r by $-X_r$. Then (b) holds. If $|S| = 3$ in (c), there is a root s such that $r = s + \bar{s}$, and one gets (c) by applying σ to the equation $X_s X_{\bar{s}} = kX_r$. If $|S| = 1$, assume $ht\ r > 1$. Then there is either a self-conjugate fundamental root a such that $r - a$ is a root, or a conjugate pair of orthogonal fundamental roots b, $\bar{b}$ such that $r - b$, $r - \bar{b}$ and $r - b - \bar{b}$ are all roots. One then applies σ to the equation $X_{r-a}X_a = k_1 X_r$ or $(X_{r-b-\bar{b}}X_b)X_{\bar{b}} = k_2 X_r$, respectively, and completes the proof of (c) by induction on the height.

We assume henceforth that the normalization indicated by 3.2 has been made and that the corresponding treatment has been given to the negative roots, so that one has once again the equations of structure of Theorem 1 of [3] (in particular, $X_r X_{-r} = H_r$).

4. **Some nilpotent groups.** As in [3], we set $x_r(t) = \exp(t\ ad\ X_r)$, $t \in K$, $r \in \Pi$, denote by $\mathfrak{X}_r$ the one-parameter group $\{x_r(t) \mid t \in K\}$, and by $\mathfrak{U}$ the group generated by all $\mathfrak{X}_r$, $r \in \Pi$.

4.1 LEMMA. *For r, $s \in \Pi$ and t_1, $t_2 \in K$, one has the commutator relation $(x_r(t_1), x_s(t_2)) = x_{r+s}(N_{rs}t_1t_2)$.*

Proof. This follows from [3, p. 33, *l*. 22] and the fact that all roots have the same length.

A straightforward computation yields:

4.2 $$\sigma \exp(t\ ad\ X_r)\,\sigma^{-1} = \exp(\bar{t}\ ad\ \sigma X_r)\ .$$

4.3 LEMMA. *Let Σ be a subset of Π satisfying the condition*

4.4 $$r,\ s \in \Sigma,\ r + s \in \Pi \ \text{imply} \ r + s \in \Sigma.$$

Then each $x \in \mathfrak{U}_\Sigma$, the group generated by all $\mathfrak{X}_r$, $r \in \Sigma$, can be written uniquely in the form $x = \prod x_r(t_r)$, the product being over the roots of Σ arranged in increasing order relative to $<$ (see § 2).

Proof. Using the formulas 4.1 repeatedly, one sees that the set of elements of the given form is closed under multiplication; thus each $x \in \mathfrak{U}_\Sigma$ has an expression of the given form. Uniqueness is proved by induction on $|\Sigma|$. If $|\Sigma| = 1$ and $\Sigma = \{r\}$, this follows from $x_r(t)X_{-r} = X_{-r} + tH_r - t^2X_r$ (see [3, p. 36, *l.* 15]). If $|\Sigma| > 1$, let r be the least element of Σ (relative to $<$), and set $\Sigma' = \Sigma - r$. Let $x \in \mathfrak{U}_\Sigma$ be written as $x = x_r(t_1)x_1$ and $x = x_r(t_2)x_2$ with $t_i \in K$ and $x_i \in \mathfrak{U}_{\Sigma'}$. Then $x_r(t_2 - t_1) = x_1x_2^{-1}$. Since $x_r(t_2 - t_1)X_{-r} = X_{-r} + (t_2 - t_1)H_r - (t_2 - t_1)^2X_r$, since $x_r(t_2 - t_1) \in \mathfrak{U}_{\Sigma'}$, and since r can not be written as a sum of roots larger than r by 2.9, it follows that the coefficient of H_r, namely $t_2 - t_1$, must be 0. Thus $x_1 = x_2$, and the induction hypothesis can be applied to Σ' to complete the proof.

The result 4.3 can be applied in the cases $\Sigma = \Pi$ and $\Sigma = S \in \Pi^1$. Because of 2.9, one gets:

4.5 COROLLARY. *Each $x \in \mathfrak{U}$ can be written uniquely in the form $x = \prod x_s$, $x_s \in \mathfrak{U}_s$, the product being over the sets S of Π^1 arranged in increasing order.*

Denote now by $\mathfrak{U}^1$, $\mathfrak{U}^1_s$, etc. the subgroups of elements of $\mathfrak{U}$, $\mathfrak{U}_s$, etc. commuting with σ.

4.6 LEMMA. *If $x \in \mathfrak{U}$ is written in the form 4.5, then $x \in \mathfrak{U}^1$ if and only if each $x_s \in \mathfrak{U}^1$. A necessary and sufficient condition for $x_s \in \mathfrak{U}_s$ to be in $\mathfrak{U}^1$ is that, in the cases (1), (2) or (3) of 2.1, x_s has the respective form (1) $x_r(t), \bar{t} = t$, (2) $x_r(t)x_{\bar{r}}(v)$, $v = \bar{t}$, or (3) $x_r(t)x_{\bar{r}}(v)x_{r+\bar{r}}(w)$, $v = \bar{t}$, $w + \bar{w} = N_{\bar{r}r}t\bar{t}$.*

Proof. If $x \in \mathfrak{U}^1$ commutes with σ, one has $x = \sigma x \sigma^{-1} = \prod(\sigma x_s \sigma^{-1})$. Since $\sigma x_s \sigma^{-1} \in \mathfrak{U}_s$ by 4.2, one gets $\sigma x_s \sigma^{-1} = x_s$ by the uniqueness in 4.5. Thus each $x_s \in \mathfrak{U}^1$. The converse is clear. In the cases listed in the second statement, one has

(1) $\sigma x_r(t)\sigma^{-1} = x_{\bar{r}}(\bar{t})$,

(2) $\sigma x_r(t)x_{\bar{r}}(v)\sigma^{-1} = x_r(\bar{v})x_{\bar{r}}(\bar{t})$, and

(3) $\sigma x_r(t)x_{\bar{r}}(v)x_{r+\bar{r}}(w)\sigma^{-1} = x_r(\bar{v})x_{\bar{r}}(\bar{t})x_{r+\bar{r}}(-\bar{w} + N_{r\bar{r}}\bar{t}\bar{v})$ by 3.2, 4.1 and 4.2. The required results now follow from 4.3.

4.7 LEMMA. *Let Π be the union of the disjoint sets Σ and Σ',* *each invariant under* σ, *and each satisfying* 4.4. *Then* $\mathfrak{U}^1 = \mathfrak{U}^1_\Sigma \mathfrak{U}^1_{\Sigma'}$ *and* $\mathfrak{U}^1_\Sigma \cap \mathfrak{U}^1_{\Sigma'} = 1$.

Proof. By [3, p. 41, Lemma 11], one can write $x \in \mathfrak{U}^1$ uniquely in the form $x = yy'$, $y \in \mathfrak{U}_\Sigma$, $y' \in \mathfrak{U}_{\Sigma'}$. The proof that y and y' are in $\mathfrak{U}^1$ is the same as that for the first part of 4.6.

If $\mathfrak{B}$ denotes the group generated by all $\mathfrak{X}_r$, $r < 0$, then one can define $\mathfrak{B}^1$, $\mathfrak{B}_s$, etc., and gain for these groups corresponding results.

5. Main results of Chevalley. For each simple complex Lie algebra $\mathfrak{g}$ (not necessarily one for which σ exists), consider the groups $\mathfrak{U}$ and $\mathfrak{B}$ and also the group G (denoted in [3] by G') which they generate. For each $w \in W$, if Σ consists of the roots r for which $r > 0$ and $wr < 0$, we set $\mathfrak{U}_\Sigma = \mathfrak{U}_w$ (denoted in [3] by $\mathfrak{U}''_w$). Let P_r and P, respectively, denote the additive groups generated by the roots and by the weights. Corresponding to each character χ of P_r into K^*, there is an automorphism $h = h(\chi)$ of $\mathfrak{g}_K$ defined by $hX_r = \chi(r)X_r$, $r \in \Pi$ or $-\Pi$. Let $\mathfrak{H}$ (denoted in [3] by $\mathfrak{H}'$) be the group generated by those automorphisms which correspond to characters which can be extended to P. For $h(\chi) \in \mathfrak{H}$, one has

5.1 $$hx_r(t)h^{-1} = x_r(\chi(r)t) .$$

The main results of [3] are as follows:

5.2 G contains $\mathfrak{H}$.

5.3 Corresponding to each $w \in W$ there is $\omega(w) \in G$ such that $\omega(w)X_r = c_rX_{wr}$, $\omega(w)H_r = H_{wr}$, $c_r \in K^*$, $r \in \Pi$ or $-\Pi$. The union of the sets $\mathfrak{H}\omega(w)$ is a group $\mathfrak{W}$ and the map $w \to \mathfrak{H}\omega(w)$ is an isomorphism of W on $\mathfrak{W}/\mathfrak{H}$.

Parenthetically, we remark that here one has:

5.4 $$\omega(w)\mathfrak{X}_r\omega(w)^{-1} = \mathfrak{X}_{wr} .$$

5.5 G is the union of the sets $\mathfrak{U}\mathfrak{H}\omega(w)\mathfrak{U}_w$, $w \in W$. These sets are disjoint and each element of G has a unique expression of the indicated form.

5.6 G is simple if one excludes the case (1) $|K| = 2$ and $\mathfrak{g}$ of type A_1, B_2 or G_2, and (2) $|K| = 3$ and $\mathfrak{g}$ of type A_1.

Before proving corresponding results for the group G^1 generated by $\mathfrak{U}^1$ and $\mathfrak{B}^1$, we identify G^1 in the case that $\mathfrak{g}$ is of type A_l.

6. Some unitary groups. Consider the form

6.1 $$f(\alpha, \beta) = \sum_1^{l+1} (-1)^i \alpha_i \bar{\beta}_i$$

on a space of l dimensions over K. Let $U_{l+1}(f)$ denote the corresponding unimodular unitary group and $C_{l+1}(f)$ its center. Then one has:

6.2 *If* $\mathfrak{g}$ *is of type* A_l, $G^1 \cong U_{l+1}(f)/C_{l+1}(f)$.

Proof. If $\mathfrak{g}$ *is of type* A_l, one can identify $\mathfrak{g}_K$ with $\mathfrak{sl}_{l+1}(K)$, the algebra of $(l+1)$th order matrices of trace 0, in such a way that, for each fundamental root $a(i)$, $X_{a(i)} \in \mathfrak{g}_K$ corresponds to $E_{i,i+1}$, the matrix with 1 in the $(i, i+1)$ position and 0 elsewhere [7, p. 393]. If $m = ((-1)^i\delta_{i,l+2-j})$ is the matrix corresponding to f, one can then verify that σ is the product of the transformations $Y \rightarrow mYm^{-1}$ (matrix multiplication) and $Y \rightarrow -\overline{Y}^t$ (t = transpose). According to a recent identification of R. Ree [7], $\mathfrak{U}$ and $\mathfrak{V}$, respectively, consist of the superdiagonal matrices (0 below and 1 on the diagonal) and the subdiagonal matrices, acting on $\mathfrak{sl}_{l+1}$ via inner automorphisms, so that the group G of Chevalley is in this case the projective unimodular group. Now it follows from material in [4, p. 66-69] that $U_{l+1}(f)$ is generated by its superdiagonal and subdiagonal elements and that $C_{l+1}(f)$ consists of scalar matrices. Thus to prove 6.2 it is enough to prove:

6.3 Let x be a superdiagonal matrix. Then $x \in \mathfrak{U}^1$ if and only if $x \in U_{l+1}(f)$.

A simple calculation using the concrete form of σ given above shows that $x\sigma = \sigma x$ if and only if $\bar{x}^t m^{-1} x m$ commutes with each $Y \in \mathfrak{sl}_{l+1}$. This is equivalent to $xm\bar{x}^t = km$, $k \in K$. If x is superdiagonal, k must be 1, because the $(1, l+1)$ entries of the matrices $xm\bar{x}^t$ and m are both -1. Thus 6.3 and 6.2 are proved.

It is to be observed that the form f has index $[(l+1)/2]$.

7. **Structure of G^1.** Recall that G^1 is the group generated by $\mathfrak{U}^1$ and $\mathfrak{V}^1$. For each $w \in W^1$, set $\mathfrak{U}^1_w = \mathfrak{U}^1 \cap \mathfrak{U}_w$. For each $S \in \Pi^1$, let G^1_S be the group generated by $\mathfrak{U}^1_S$ and $\mathfrak{V}^1_S$. Denote by X^1 the group of those characters of P_r into K^* which can be extended to characters χ of P which are selfconjugate in the sense that $\chi(\bar{a}) = \overline{\chi(a)}$ for all $a \in P$, and by $\mathfrak{H}^1$ the corresponding subgroup of $\mathfrak{H}$. For $S \in \Pi^1$, set $\mathfrak{H}^1_S = \mathfrak{H}^1 \cap G_S$. Finally, for each root r and each $k \in K^*$, denote by $\chi_{r,k}$ the character on P_r defined by $\chi_{r,k}(s) = k^{s(r)}$.

It is assumed until further notice that $\mathfrak{g}$ is not of type A_l (l even). We aim to prove:

7.1 Lemma. *For each* $w \in W^1$, $\mathfrak{H}\omega(w) \cap G^1$ *is not empty.*

Once this is established, it can (and will) be assumed that $\omega(w) \in G^1$ for each $w \in W^1$. Then:

7.2 Theorem. G^1 *is the union of the sets* $\mathfrak{U}^1\mathfrak{H}^1\omega(w)\mathfrak{U}^1_w$, $w \in W^1$. *The sets are disjoint and each element of* G^1 *has a unique expression of the indicated form.*

The steps of the proof are quite analogous to those in the proof of 5.5 in view of the following:

7.3 LEMMA. *Assume* $S \in \Pi^1$. *Then* (1) *if* $S = \{r\}$, *there is a homomorphism* φ_1 *of* $SL_2(K_0)$, *the unimodular group, onto* G_s^1 *such that*

$$\varphi_1\begin{pmatrix}1 & t\\ 0 & 1\end{pmatrix} = x_r(t),\ \varphi_1\begin{pmatrix}1 & 0\\ t & 1\end{pmatrix} = x_{-r}(t),\ \varphi_1\begin{pmatrix}k & 0\\ 0 & k^{-1}\end{pmatrix} = h(\chi_{r,k}) ,$$

and

$$\varphi_1\begin{pmatrix}0 & 1\\ -1 & 0\end{pmatrix} \equiv \omega(w_r) \pmod{\mathfrak{H}} ;$$

(2) *if* $S = \{r, \bar{r}\}$, *there is a homomorphism* φ_2 *of* $SL_2(K)$ *onto* G_s^1 *such that*

$$\varphi_2\begin{pmatrix}1 & t\\ 0 & 1\end{pmatrix} = x_r(t)x_{\bar{r}}(\bar{t}),\ \varphi_2\begin{pmatrix}1 & 0\\ t & 1\end{pmatrix} = x_{-r}(t)x_{-\bar{r}}(t),\ \varphi_2\begin{pmatrix}k & 0\\ 0 & k^{-1}\end{pmatrix} = h(\chi_{r,k}\chi_{\bar{r},\bar{k}}) ,$$

and

$$\varphi_2\begin{pmatrix}0 & 1\\ -1 & 0\end{pmatrix} \equiv \omega(w_r w_{\bar{r}}) \pmod{\mathfrak{H}} .$$

Proof. The existence of φ_1 is established in [3, p. 29, p. 36]. Since $\mathfrak{X}_r$ and $\mathfrak{X}_{-r}$ commute elementwise with $\mathfrak{X}_{\bar{r}}$ and $\mathfrak{X}_{-\bar{r}}$, it is clear that φ_2 also exists.

Proof of 7.1. By 7.3, $\mathfrak{H}\omega(w_s) \cap G^1$ is non-empty for each $S \in \Pi^1$. Thus 7.1 follows from 2.3.

Now we choose $\omega(w) \in G^1$ for each $w \in W^1$, and denote by $\mathfrak{W}^1$ the union of the sets $\mathfrak{H}^1\omega(w)$. Then the analogue of 5.3 holds.

7.4 LEMMA. G^1 *contains* $\mathfrak{H}^1$.

Proof. G^1 contains all $h(\chi) \in \mathfrak{H}^1$ such that χ is of the form $\chi_{a,k}$, $\bar{a} = a$, $\bar{k} = k$, or $\chi_{a,j}\chi_{\bar{a},\bar{j}}$ by 7.3. These characters generate X^1 (see [3, p. 48, Lemma 2]). Thus $G^1 \supset \mathfrak{H}^1$.

7.5 LEMMA. *For each* $S \in \Pi^1$, G_s^1 *is the the union of the sets* $\mathfrak{U}_s^1\mathfrak{H}_s^1$ *and* $\mathfrak{U}_s^1\mathfrak{H}_s^1\omega(w_s)\mathfrak{U}_s^1$.

Proof. Because of 7.3, this follows from the corresponding properties of the groups $SL_2(K_0)$ and $SL_2(K)$ (see [3, p. 34, Lemma 2]).

7.6 LEMMA. G^1 *is generated by the groups* $\mathfrak{U}_s^1$ *and* $\mathfrak{V}_s^1$ *which correspond to fundamental sets* $S \in \Pi^1$.

Proof. This follows from 7.1, 5.4 and 4.5.

7.7 LEMMA. $G^1 = \mathfrak{U}^1\mathfrak{W}^1\mathfrak{U}^1$.

Proof. This follows from 7.6, 7.5, 7.1 and 4.7 as in [3, p. 40, Lemma 10].

Proof of 7.2. That G^1 is the union of the given sets follows from 7.7 and 4.7 as in [3, p. 42, Theorem 2]. The disjointness and uniqueness follow from 5.5.

7.8 COROLLARY. $\mathfrak{H}^1 = \mathfrak{H} \cap G^1$.

Proof. Because of 7.2, this is clear.

7.9 COROLLARY. $\mathfrak{U}^1\mathfrak{H}^1$ *is the normalizer of* $\mathfrak{U}^1$ *in* G^1.

Proof. The normalizer contains $\mathfrak{U}^1\mathfrak{H}^1$ by 5.1, and equality follows from 7.2.

One also concludes from the preceding results:

7.10 COROLLARY. *The sets of* 7.2 *are the double cosets of* G^1 *relative to* $\mathfrak{U}^1\mathfrak{H}^1$.

7.11 COROLLARY. *If K is a finite field of characteristic p, then* $\mathfrak{U}^1$ *and* $\mathfrak{V}^1$ *are p-Sylow subgroups of* G^1.

In regard to 7.11, one sees from 4.5 and 4.6 that, if $|K| = q^2$ and $|\Pi| = N$, then $|\mathfrak{U}^1| = q^N$.

We now remove the restriction on $\mathfrak{g}$ and remark that the results of this section remain valid even if $\mathfrak{g}$ is of type A_l (l even). The key point here is that, if $S \in \Pi^1$ and $|S| = 3$, then there exists a homomorphism of $U_3(f)$ (see 6.1) onto G^1_S with properties like those of φ_1 and φ_2 in 7.3. We omit the proof which can be made to depend on the representation of G^1 by unitary matrices given in § 6.

8. Proof of simplicity.

Our aim here is to prove:

8.1 THEOREM. *If K_0 has at least* 5 *elements, then G^1 is simple.*

The simplicity of the group SL_2 over its center is assumed to be known. It is further assumed that $\mathfrak{g}$ is not of type A_l (l even) and that $l \geqq 3$. The proof to be given can be adapted with minor modifications to the missing groups, which are in any case adequately covered by 6.2 and [4, p. 70, Theorem 5].

8.2 LEMMA. *Assume R, $T \in \Pi^1$, $R \neq T$, and that r, t are elements of R, T, respectively. Then there is $\chi \in X^1$ such that $\chi(r) = 1$, $\chi(t) \neq 1$.*

Proof. Let $\tilde{R}$ (or more simply R) denote r or $r + \bar{r}$ in the cases $\bar{r} = r$ or $\bar{r} \neq r$, respectively, and then set $\chi_{R,k} = \chi_{r,k}$ or $\chi_{R,k} = \chi_{r,k}\chi_{\bar{r},\bar{k}}$ accordingly. Treat t and T similarly. If $R(T) = 0$, set $\chi = \chi_{T,k}$, $k \in K_0^*$, $k^2 \neq 1$. If $R(T) = \pm 1$, or if $R(T) = \pm 2$ and $|R| = |T| = 2$, set $\chi = \chi_{T,k^2}\chi_{R,k}^{-r(T)}$, $k \in K_0^*$, $k^3 \neq 1$. In the other cases of $R(T) = \pm 2$, set $\chi = \chi_{t,k}\chi_{\bar{t},k}\chi_{R,k}^{-r(t)}$, $k \in K_0^*$, $k^2 \neq 1$. Finally if $R(T) = \pm 4$, set $\chi = \chi_{t,k}\chi_{\bar{t},\bar{k}}$, $k = \bar{k}_1/k_1$, $k_1 \in K$, $\bar{k}_1 \neq \pm k_1$. One can check that these cases are exhaustive and that $\chi(r) = 1$ and $\chi(t) \neq 1$ in each case.

8.3 LEMMA. *If $w \in W^1$ and $w \neq 1$, there is $h \in \mathfrak{H}^1$ such that $\omega(w)h \neq h\omega(w)$.*

Proof. We first show that there exist $\chi \in X^1$ and $r \in \Pi$ such that $\chi(wr) \neq \chi(r)$. If there is an $R \in \Pi^1$ such that $wR \neq \pm R$, then χ and r exist by 8.2. If $wR = \pm R$ for all $R \in \Pi^1$, then, since $w \neq 1$, one has $wR = -R$ for all $R \in \Pi^1$. Since $l \geqq 3$, one can readily choose $r, t \in \Pi$ so that $r \perp \bar{r}$, $t = \bar{t}$ and $r(t) < 0$. If $k \in K_0^*$, $k^2 \neq 1$, then $\chi = \chi_{t,k}$ and r have the required property. If $h = h(\chi)$, a simple calculation now shows that X_r has different images under $\omega(w)h$ and $h\omega(w)$.

Assume now that H is a normal subgroup of G^1 and that $|H| > 1$.

8.4 LEMMA. $|H \cap \mathfrak{U}^1\mathfrak{H}^1| > 1$.

Proof. By 7.2 there is $x \in H$ such that $x \neq 1$ and $x = uh_1\omega(w)$ with $u \in \mathfrak{U}^1$, $h_1 \in \mathfrak{H}^1$ and $w \in W^1$. If $w \neq 1$, then by 8.3 there is $h \in \mathfrak{H}^1$ such that $\omega(w)h \neq h\omega(w)$. Then $y = hxh^{-1}x^{-1} \in H \cap \mathfrak{U}^1\mathfrak{H}^1$, and we assert that $y \neq 1$. Indeed, if $y = 1$, then

$$x = hxh^{-1} = huh^{-1}(hh_1\omega(w)h^{-1}\omega(w)^{-1})\omega(w) ,$$

and by 7.2 one gets $h\omega(w)h^{-1}\omega(w)^{-1} = 1$, a contradiction. Thus the assertion and the lemma are proved.

8.5 LEMMA. $|H \cap \mathfrak{U}^1| > 1$.

Proof. By 8.4, there is $x \in H \cap \mathfrak{U}^1\mathfrak{H}^1$ such that $x \neq 1$. Write $x = uh$, $u \in \mathfrak{U}^1$, $h \in \mathfrak{H}^1$, and suppose $h \neq 1$. Then there is a fundamental root r such that $hX_r = cX_r$, $c \in K$, $c \neq 1$. If $r \in S \in \Pi^1$, let y be the commutator of x with $x_r(1)$ or $x_r(1)x_{\bar{r}}(1)$ according as $|S| = 1$ or 2. Then $y \in H \cap \mathfrak{U}^1$, and it remains to show that $y \neq 1$. If $y = 1$, then, for the case $|S| = 1$, one has $x_r(1) = uhx_r(1)h^{-1}u^{-1} = ux_r(c)u^{-1}$. Now it follows easily from 4.1 that the subgroup $\mathfrak{U}_2$ of $\mathfrak{U}$ generated by those $\mathfrak{X}_r$ for which $ht\, r > 1$ contains the commutator subgroup of $\mathfrak{U}$. Thus $x_r(1-c) = x_r(1)x_r(c)^{-1} \in \mathfrak{U}_2$, whence $1 - c = 0$ by 4.3. This contradiction establishes $y \neq 1$. The case $|S| = 2$ can be treated similarly.

8.6 Lemma. *For some* $R \in \Pi^1$, $|H \cap U^1_R| > 1$.

Proof. Among all $x \in H \cap \mathfrak{U}^1$ with $x \neq 1$, choose one which maximizes the minimum $S \in \Pi^1$ for which $x_s \neq 1$ in the representation 4.5. If this minimum is R, we show $x = x_R$. Assuming the contrary, one can write $x = x_R x_T x_1$ with $x_R \neq 1$, $x_T \neq 1$, and x_1 denoting the remaining terms in 4.5. By 8.2, 5.1 and 4.6, there is $h \in \mathfrak{H}^1$ such that $hx_Rh^{-1} = x_R$ and $hx_Th^{-1} \neq x_T$. Thus $hxh^{-1} \neq x$ by 4.5. But then $y = x^{-1}hxh^{-1} \neq 1$, $y \in H \cap \mathfrak{U}^1$, and y provides a contradiction to the choice of x.

Using 8.6, one can deduce as in [3, p. 62, Lemma 15]:

8.7 Lemma. *If* $|H \cap \mathfrak{U}^1_R| > 1$ *for* $R \in \Pi^1$, *then* $H \supset \mathfrak{U}^1_R$.

Proof of 8.1. As in 8.3 choose (fundamental) roots r, t such that $r \perp \bar{r}$, $t = \bar{t}$ and $r(t) < 0$. Since $r \perp \bar{r}$, this implies that $r + t$, $\bar{r} + t$ and $r + \bar{r} + t$ are all roots. Set $R = \{r, \bar{r}\}$, $T = \{t\}$, $U = \{r + t, \bar{r} + t\}$, $V = \{r + \bar{r} + t\}$, $x_R(1) = x_r(1)x_{\bar{r}}(1)$, $x_T(1) = x_t(1)$, etc.. Then by 4.1 (used several times), one gets:

$$(x_R(1),\ x_T(1)) = x_U(N_{rt})x_V(N_{rt}N_{r,\bar{r}+t}) \,. \tag{8.8}$$

By 8.7, 5.4 and 2.8, either $x_R(1)$ or $x_T(1)$ is in H; hence so is their commutator. For the same reason one of the elements on the right of 8.8 is in H; hence so is the other. Thus, by 8.7, 5.4, 3.1 and 2.8, H contains all $\mathfrak{U}^1_S$, hence also $\mathfrak{U}^1$ by 4.5. Similarly H contains $\mathfrak{C}^1$, whence $H = G^1$. Thus G^1 is simple.

9. Some identifications. If $\mathfrak{g}$ is of type A_l, then G^1 has been identified in § 6 as a projective unitary group in $l + 1$ dimensions. Similarly, if $\mathfrak{g}$ is of type D_l ($l \geqq 4$), then using the representation of G given by Ree [7], one can show that G^1 is isomorphic to a projective orthogonal group corresponding to a form in $2l$ variables which has index $l - 1$ relative to K_0 and index l relative to K. The details in the complex-real case can be found in [2, p. 422]. If $\mathfrak{g}$ is of type E_6, then, again in the complex-real case, one can identify G^1 with a real form of E_6, the one characterized by Cartan [2, p. 493] by the fact that its Killing form, when written as a sum of real squares, contains a surplus of 2 positive terms. If $\mathfrak{g}$ is of type E_6 and K is finite, we show in § 12 that new groups are obtained[1], not isomorphic to any appearing in the list of finite simple groups given by Artin [1].

10. Second variation for D_4. A root system for D_4 has a fundamental basis consisting of roots a, b, c, d of the same length such that b, c, d are mutually orthogonal and each makes an angle of $2\pi/3$ with a. Let τ be the automorphism of order 3 of the underlying Euclidean space defined by $a, b, c, d \rightarrow a, c, d, b$, and let W^2 be the subgroup of elements of W commuting with τ. One can then obtain the analogues of the results of § 2 without essential change in the proofs. For example: W^2 is generated by the elements w_a and $w_b w_c w_d$, and is of type G_2. The roots are partitioned into sets of the types (1) $S = \{r\}$, $\tau r = r$, and (2) $S = \{r, \tau r, \tau^2 r\}$. Any 2 sets of the same type are congruent under W^2. One then introduces a field K on which an automorphism τ of order 3 acts, and defines a semi-automorphism τ of $\mathfrak{g}_K$ by $\tau(kX_r) = (\tau k)X_{\tau r}$. Then $\mathfrak{U}^2$ and $\mathfrak{V}^2$ are the subgroups of $\mathfrak{U}$ and $\mathfrak{V}$, respectively, made up of elements commuting with τ and G^2 is the group they generate. The whole previous developement goes through. It turns out that in the proof of simplicity it is enough to assume that the fixed field K_0 has at least 4 elements. In § 12, it is shown that once again new finite groups[1] are obtained.

11. Third variation for D_4. Assume now that K is a field admitting automorphisms σ and τ which are of orders 2 and 3 respectively, and which generate a group isomorphic to S_3, the symmetric group on 3 objects. Define corresponding semi-automorphisms σ and τ of the Lie algebra $\mathfrak{g}_K$ of type D_4 as in §§ 3 and 10. Then set $\mathfrak{U}^3 = \mathfrak{U}^1 \cap \mathfrak{U}^2$, $\mathfrak{V}^3 = \mathfrak{V}^1 \cap \mathfrak{V}^2$, and let G^3 be the group generated by $\mathfrak{U}^3$ and $\mathfrak{V}^3$. Again everything goes through. It need only be remarked that the present construction is possible only if K is infinite, and that all groups of type G^3 are simple.

12. Some new groups. The list L of known finite simple groups consists of the cyclic, alternating and Mathieu groups, and the "Lie groups", namely the groups G of Chevalley over A_l $(l \geqq 1)$, B_l $(l \geqq 2)$, C_l $(l \geqq 3)$, D_l $(l \geqq 4)$, E_6, E_7, E_8, F_4 and G_2, the groups G^1 over A_l $(l \geqq 2)$, D_l $(l \geqq 4)$ and E_6, and the groups G^2 over D_4, all constructed on a finite field. By the type of one of these latter groups we mean a combination consisting of the general mode of construction (G or G^1 or G^2), the underlying complex Lie algebra $\mathfrak{g}$, and the field K. We adopt the notation: $E_6^1(r)$ is the group of type G^1 over E_6 on a field of r elements. Our aim is to prove:

12.1 THEOREM. *If G is one of the groups $E_6^1(q^2)$ or $D_4^2(q^3)$, then $\hat{G}$*

is not isomorphic to a cyclic, alternating or Mathieu group, and two representations of $\hat{G}$ as Lie groups necessarily have the same type.

In other words the groups $E_6^1(q^2)$ and $D_4^2(q_1^3)$ are new[1] and distinct among themselves. We need some preliminary results. Let $\hat{G}$ be a Lie group over a field K of q, q^2 or q^3 elements in the cases G, G^1 or G^2, respectively, and set $\hat{W} = W$, W^1 or W^2 accordingly. The *Poincaré sequence* of $\hat{G}$ shall mean the list of numbers $q^{n(w)} (w \in \hat{W})$ arranged in non-decreasing order. Thus the first term is 1 and the last term is q^N, the integer N being the number of positive roots of $\mathfrak{g}$ (see 2.5, 4.5 and 4.6).

12.2 LEMMA. *The Poincaré sequence of $A_l^1(q^2)$, $D_l^1(q^2)$, $E_6^1(q^2)$ or $D_4^2(q^3)$ is obtained by writing the respective polynomial* $\prod_1^{l+1} \frac{t^i - (-1)^i}{t - (-1)^i}$,

$$(t^l + 1) \prod_2^{l-1} \frac{t^{2i} - 1}{t - 1}, \quad \frac{t^2 - 1}{t - 1} \cdot \frac{t^5 + 1}{t + 1} \cdot \frac{t^6 - 1}{t - 1} \cdot \frac{t^8 - 1}{t - 1} \cdot \frac{t^9 + 1}{t + 1} \cdot \frac{t^{12} - 1}{t - 1} \text{ or}$$

$(t + 1)(t^3 + 1)(t^8 + t^4 + 1)$ *as a sum of non-decreasing powers of t and then replacing t by q in the individual terms.*

To avoid interruption of the present development we give the proof in the next section. We also need the polynomials for the groups of Chevalley. As one sees from considerations in [3, p. 44, p. 64], these polynomials take the form $\prod[(t^{a(i)} - 1)/(t - 1)]$, the $a(i)$ being given in [3, p. 64]. Since $q^{n(w)} = |\mathfrak{U}_w^1|$ by 4.6 and 4.7, one can use 12.2 in conjunction with 7.2 and the definition of $\mathfrak{H}^1$ to compute $|G^1|$. In the same way, one can find $|G^2|$. Thus:

12.3 LEMMA. *If u is the g. c. d. of 3 and $q + 1$, the orders of $E_6^1(q^2)$ and $D_4^2(q^3)$ are $u^{-1}q^{36}(q^2 - 1)(q^5 + 1)(q^6 - 1)(q^8 - 1)(q^9 + 1)(q^{12} - 1)$ and $q^{12}(q^2 - 1)(q^6 - 1)(q^8 + q^4 + 1)$, respectively.*

The orders of the other Lie groups can be found in [1]. It is interesting to note that, if in the expressions in 12.2 and 12.3 which relate to the group $E_6^1(q^2)$ one replaces all plus signs by minus signs, then one obtains the corresponding properties of $E_6(q)$. A similar phenomenon occurs for each of the groups $A_l^1(q^2)$ and $D_l^1(q^2)$.

12.4 LEMMA. *The Poincaré sequence of a finite Lie group $\hat{G}$ is determined by the abstract group and the characteristic p of the base field K. The type of a finite Lie group is determined by its Poincaré sequence except that $B_l(q)$ and $C_l(q)$ have the same sequence, as do $A_1(q^3)$ and $A_2^1(q)$ also.*

Proof. If $\hat{G}$ is of type G, then, to within an inner automorphism, G and p determine $\mathfrak{U}$ as a p-Sylow subgroup, then $\mathfrak{U}\mathfrak{H}$ as the normalizer of $\mathfrak{U}$, and finally the numbers $|\mathfrak{U} \cap x\mathfrak{U}x^{-1}|$ as x runs through a system of representatives of the double coset decomposition of G relative to $\mathfrak{U}\mathfrak{H}$. These latter numbers are just the terms of the Poincaré sequence by the analogue of 7.10, since $|\mathfrak{U} \cap \omega(w)\mathfrak{U}\omega(w)^{-1}| = q^{n(w_0 w)}$ by 4.3. A similar proof of the first statement holds for groups of type G^1 or G^2. One proves the second statement by inspection of the Poincaré sequences for the various Lie groups.

By checking their orders, one sees that $A_1(q^3)$ and $A_2^1(q)$ can not be isomorphic. Thus the two statements of 12.4 can be combined to yield:

12.5. *The type of a finite Lie group is determined by the abstract group and the characteristic of the base field except that $B_l(q)$ and $C_l(q)$ may be isomorphic.*

This result has been obtained previously (for the previously known finite simple Lie groups) by Artin [1] and Dieudonné [5, p. 71–75] by different, more detailed methods. Artin actually draws the conclusion under the weak assumption that only $|\hat{G}|$ and p are known.

One also concludes from 12.4 the well-known fact that $A_2(4)$ and $A_3(2)$, both of order 20160, are not isomorphic.

An inspection of the results of 12.3 yields:

12.6 LEMMA. *Let $\hat{G}$ be either $E_6^1(q^2)$ or $D_4^2(q^3)$ over a field of characteristic p, and let Q be the largest power of p which divides $|\hat{G}|$. Let Q' be any prime power which divides $|\hat{G}|$. Then $Q^3 > |\hat{G}|$ and $Q \geqq Q'$.*

Proof of 12.1. Clearly $\hat{G}$ is not cyclic. Since $|\hat{G}| > 10^8$ and $Q^3 > |\hat{G}|$, it follows that $\hat{G}$ is not an alternating group (see [1]). $D_4^2(8)$ does not have the order of a Mathieu group and all other values of $|\hat{G}|$ are too large. $\hat{G}$ is not isomorphic to either of the groups $A_1(p_1)$ with $p_1 = 2^r - 1 = $ prime, or $A_1(2^s)$ with $2^s + 1 = $ prime, since in each case one has a prime p_2 such that p_2 divides $|\hat{G}|$ and $p_2^3 > |\hat{G}|$, and this is readily seen to be impossible by 12.3. But except for these two types, every simple finite Lie group verifies 12.6 (see [1] where the other groups are considered). Thus any representation of $\hat{G}$ as a Lie group must be over a field of characteristic p. An application of 12.4 completes the proof.

13. Proof of 12.2. By 2.2, 2.3 and 2.6, $n(w) = \sum |S|$, summed over those $S \in \Pi^1$ for which $wS < 0$. By 2.7, one can compute $n(w)$ within the framework of $\tilde{W}^1$ and its root system, but each root is to be counted with the right multiplicity (1, 2 or 3). Assume first that the group under consideration is $E_6^1(q^2)$. Then $\tilde{W}^1$ is of type F_4 and, in terms of coordinates relative to an orthonormal basis, its roots can be taken as $\pm x_i$, $(\pm x_1 \pm x_2 \pm x_3 \pm x_4)/2$, each of multiplicity 1, and $\pm x_i \pm x_j$ $(i \neq j)$, each of multiplicity 2 (see [8, p. 13-08]). The inequalities $x_1 - x_2 - x_3 - x_4 > 0$, $x_2 - x_3 > 0$, $x_3 - x_4 > 0$, and $x_4 > 0$ determine a fundamental region F of $\tilde{W}^1$ by [10, p. 160]. The last 3 inequalities determine a region L whose intersection with the unit sphere is lune-shaped with $(1, 0, 0, 0)$ as one of its vertices. The subgroup V of $\tilde{W}^1$ leaving $(1, 0, 0, 0)$ fixed is of type C_3 and has L as a fundamental region. Let $P(t)$ be the polynomial sought, let $P_1(t)$ be the corresponding polynomial for the group V, and let $P_2(t)$ be $\sum t^{n(w)}$, the sum being over those $w \in W^1$ for which $\tilde{w}F \subset L$. A simple geometric argument shows that $P = P_1P_2$. We next find P_2. The point $a = (16, 8, 4, 2)$ is in F. It has 24 transforms in L corresponding to the 24 elements $\tilde{w} \in \tilde{W}^1$ for which $\tilde{w}F \subset L$. These are a, $b = (15, 5, 3, 9)$, $c = (13, 11, 7, 1)$ and the points in L obtained from these by coordinate permutations. One can now find $n(w)$ for each of the 24 elements above. For example, if $\dot{w}$ maps a on b, then the roots positive at a and negative at b are $(x_1 - x_2 - x_3 - x_4)/2$, of multiplicity 1, and $x_2 - x_4$ and $x_3 - x_4$, each of multiplicity 2. Hence $n(w) = 5$. Thus P_2 is determined, and the original problem of rank 4 is reduced to one of rank 3. A similar reduction to rank 2 is possible, whence P can be determined. If one starts with $A_l^1(q^2)$ or $D_l^1(q^2)$ instead, the same inductive procedure can be carried through, and for $D_4^2(q^3)$ the polynomial P can be found rather quickly by enumerating $n(w)$ for the 12 elements of W^2. The results are those listed in 12.2.

14. Prime power representations. In [9], 14 assumptions on a finite group are made, and then some properties concerning the representations of the group are deduced. It is then verified that the groups of Chevalley satisfy the basic assumptions. The verification for G^1 or G^2 is virtually the same as for G because of the structure theorems of the present paper. Thus one gains the results of [9] (in particular Theorem 4) simultaneously for all known finite simple Lie groups.

15. Concluding remarks. We first note that it is possible to cover somewhat more ground than was indicated in the main development given here by allowing certain degeneracies to occur. For example, if σ on E is of order 2, if σ on K is of order 1, and if $\mathfrak{g}$ is of type A_{2l} or A_{2l-1}, then the construction of §§ 3, 4 and 5 yields a group of type B_l or C_l, respectively. Thus B_l, C_l and also A_m may be regarded as degenerate cases of A_m^1. Similarly D_l^1 degenerates to B_{l-1} and D_l; E_6^1 to F_4 and E_6; and D_4^3 to G_2, B_3, D_4, D_4^1 and D_4^2. It is easily verified that no other groups can be obtained by the present method of combining automorphisms of E and of K in various ways[1].

In regard to the construction given for G^1, it is to be noted that $\mathfrak{g}^1_K$, the set of fixed points of σ, is the Lie algebra (over K_0) of G^1 in many cases. We could have defined G^1 on $\mathfrak{g}^1_K$ in view of the easily proved facts that an automorphism x of $\mathfrak{g}_K$ commutes with σ if and only if $x\mathfrak{g}^1_K = \mathfrak{g}^1_K$, and that, in this case, the restriction of x to $\mathfrak{g}^1_K$ is 1 only if $x = 1$; but this would have led to a much more complicated development. It is also to be noted that one can not define G^1 as the subgroup G^σ of G made up of elements which commute with σ. The difference, roughly speaking, lies in $\mathfrak{H}$: a self-conjugate character on P_r may be extendable to a character on P but not to a self-conjugate one, as is proved by the following example. Let $\mathfrak{g}$ be of type A_1, and let w and $a = 2w$ be fundamental weight and root, respectively. Then χ defined by $\chi(a) = k^2$, $k^2 \in K_0^*$, $\bar{k} \neq k$, has the given property. One sees rather easily, however, that G^σ/G^1 is always isomorphic to a subgroup of P/P_r.

The proof of simplicity given in §8 is considerably shorter than the one given in [3], but this is at the expense of the assumption that K has enough elements: left open is the question of simplicity for the groups $E_6^1(q^2)$ with $q \leqq 4$, and $D_4^2(q^3)$ with $q \leqq 3$. The answer quite likely requires rather detailed methods such as those of [3].

More important, perhaps, and probably more difficult is the identification of the infinite groups constructed. An infinite analogue of 12.4 would go a long way in this direction. Finally, it seems likely that there is some sort of description of D_4^2 and D_4^3 by Cayley numbers.

REFERENCES

1. E. Artin, *Orders of classical simple groups*, Comm. Pure Appl. Math., **8** (1955), 455.
2. É. Cartan, *Oeuvres complètes*, Paris, 1952.
3. C. Chevalley, *Sur certains groupes simples*, Tôhoku Math. J., (2) **7** (1955), 14.
4. J. Dieudonné, *Sur les groupes classiques*, Paris, 1948.
5. ———, *On the automorphisms of the classical groups*, Mem. Amer. Math. Soc. 2, 1955.
6. D. Hertzig, *On simple algebraic groups*, Short communications, Int. Congress of Math., Edinburgh, 1958.
7. R. Ree, *On some simple groups defined by C. Chevalley*, Trans. Amer. Math. Soc. **84** (1957), 392.
8. Séminaire "*Sophus Lie*", Paris, 1954-1955.
9. R. Steinberg, *Prime power representations of finite linear groups* II, Canad. J. Math., **9** (1957), 347.
10. H. Weyl, *The structure and representation of continuous groups*, I.A.S. notes, Princeton, 1934-35.

UNIVERSITY OF CALIFORNIA, LOS ANGELES

ON THE SIMILARITY TRANSFORMATION BETWEEN A MATRIX AND ITS TRANSPOSE

OLGA TAUSSKY AND HANS ZASSENHAUS

It was observed by one of the authors that a matrix transforming a companion matrix into its transpose is symmetric. The following two questions arise:

I. Does there exist for every square matrix with coefficients in a field a non-singular symmetric matrix transforming it into its transpose?

II. Under which conditions is every matrix transforming a square matrix into its transpose symmetric?

The answer is provided by

THEOREM 1. *For every* $n \times n$ *matrix* $A = (\alpha_{ik})$ *with coefficients in a field* F *there is a non-singular symmetric matrix transforming* A *into its transpose* A^T.

THEOREM 2. *Every non-singular matrix transforming* A *into its transpose is symmetric if and only if the minimal polynomial of* A *is equal to its characteristic polynomial* i.e. *if* A *is similar to a companion matrix.*

Proof. Let $T = (t_{ik})$ be a solution matrix of the system $\sum(A)$ of the linear homogeneous equations.

$$TA - A^T T = 0 \tag{1}$$

$$T - T^T = 0\,. \tag{2}$$

The system $\sum(A)$ is equivalent to the system

$$TA - A^T T^T = 0 \tag{3}$$

$$T - T^T = 0 \tag{4}$$

which states that T and TA are symmetric. This system involves $n^2 - n$ equations and hence is of rank $n^2 - n$ at most. Thus there are at least n linearly independent solutions of $\sum(A)$.[1]

On the other hand it is well known that there is a non-singular matrix T_0 satisfying

$$T_0 A T_0^{-1} = A^T\,,$$

Received December 18, 1958.

[1] This part of the proof was provided by the referee. Our own argument was more lengthy.

From (1) we derive

$$T_0^{-1}TA = AT_0^{-1}T \tag{1a}$$

and conversely, (1a) implies (1) so that there is the linear isomorphism

$$T \to T_0^{-1}T$$

of the solution space of (1) onto the centralizer ring of the matrix A.

If the minimal polynomial of A is equal to the characteristic polynomial then the centralizer of A consists only of the polynomials in A with coefficients in F. In this case the solution space of (1) is of dimension n. A fortiori the solution space of $\sum(A)$ is at most of dimension n since the corresponding system involves more equations. Together with the inequality in the other direction it follows that the dimension of the solution space of $\sum(A)$ is exactly n. This implies that every solution matrix of (1) is symmetric.

If the square matrix A is arbitrary then we apply first a similarity (in the field F) which transforms it to the form

$$B = \begin{pmatrix} A_1 & & & & \\ & A_2 & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & \cdot \\ & & & & & A_r \end{pmatrix}$$

where A_i is a square matrix of the form

$$\begin{pmatrix} {}_pA & & & & & \\ L & {}_pA & & & & \\ & L & {}_pA & & & \\ & & & \cdot & & \\ & & & & \cdot & \\ & & & & & \cdot \\ & & & & L & {}_pA \end{pmatrix}$$

Here ${}_pA$ is the companion matrix of the irreducible polynomial p which is a factor of the characteristic polynomial of A and L is the matrix with 1 in the bottom left corner and 0 elsewhere, of appropriate size (Reference 1, p. 94). The matrix A is derogatory if two blocks A_i corresponding to the same p appear in B. Let A_1 and A_2 be two such blocks.

There is a non-singular matrix Y satisfying

$$Y{}_pA = {}_pA^TY .$$

The matrix of matrices V that has Y in the top left corner and 0 elsewhere, of appropriate size, satisfies

$$VA_2 = A_1^T V .$$

Consider then the matrix

$$\begin{pmatrix} S_1 & V & & & \\ & S_2 & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & \cdot \\ & & & & & S_r \end{pmatrix}$$

where S_i is a non-singular matrix transforming A_i into A_i^T. It is a non-singular non-symmetric matrix which transform B into its transpose. Thus Theorem 2 is proved.

REMARK. M. Newman pointed out to us that the product of two non-singular skew symmetric matrices B, C can always be transformed into its transpose by a non-symmmtric matrix, namely

$$B^{-1}BCB = (BC)^T = CB .$$

Theorem 2 shows that such a product BC must be derogatory.[2] This can also be shown directly in the following way:

Let λ be a characteristic root of BC and x a corresponding characteristic vector, then

$$BCx = \lambda x .$$

Since B is non-singular this implies

$$Cx = \lambda B^{-1}x$$

or

$$(C - \lambda B^{-1})x = 0 .$$

Since B is a non-singular skew symmetric matrix, it follows that the degree of B and hence the degree of $C-\lambda B^{-1}$ is even. Moreover, the skew symmetric matrix $C - \lambda B^{-1}$ has even rank.

[2] Although Newman's comment is only significant for fields of characteristic $\neq 2$ the remainder of this section holds generally if skew symmetric is understood to mean $T = - T^T$ and vanishing of the diagonal elements. We observe that this definition is invariant under the transformation $T \rightarrow X^TTX$. This is the transformation T undergoes when the matrix A in (1), (2) undergoes the similarity transformation $A \rightarrow X^{-1}AX$. Since this transformation preserves linear independence, we are permitted to apply it for the purpose of finding a non 'skew symmetric' solution of (1), (2). We now extend the field of reference to include the eigenvalues of A (from the theory of homogeneous linear equations it follows that the maximal number of linear independent solutions will remain the same). It can then be observed that for a block of the Jordan canonical form of a matrix any matrix with all coefficients zero excepting the first diagonal coefficient satisfies (1), (2). Therefore

It follows that another vector y exists such that also

$$(C - \lambda B^{-1})y = 0$$

and hence also

$$BCy = \lambda y \ .$$

This implies that λ is a characteristic root of multiplicity at least two and with at least two corresponding vectors. The product of two general non-singular skew symmetric matrices B, C has every characteristic root of multiplicity exactly 2. For, specialize to the case $B = C$. Then BC is a symmetric matrix whose characteristic roots are the squares of the roots of B, hence all exactly double for a general B. This shows that the general BC has all its characteristic roots double with two independent characteristic vectors. Such a matrix is derogatory and its characteristic polynomial is the square of its minimum polynomial.

REFERENCES

1. M. P. Drazin, *A note on skew-symmetric matrices*, Math. Gazette, **36** (1952), 253-255.
2. N. Jacobson, *Lectures on abstract algebra*, New York 1953.
3. ———, *An application of E. H. Moore's determinant of a Hermitian matrix*, Bull. Amer. Math. Soc., **45** (1939), 745-748.
4. K. Tatarkiewicz, *Sur l'orthogonalité généralisée des matrices propres*, Ann. Univ. Mariae Curie-Sklodowska, Sect. A, **9** (1955), 5-28.[3]
5. A. Voss, *Symmetrische und alternierende Lösungen der Gleichung $SX = XS'$*, S. ber, math. phys. Kl. K. B. Akademie München, **26** (1896), 273-281.

CALIFORNIA INSTITUTE OF TECHNOLOGY
MCGILL UNIVERSITY AND CALIFORNIA INSTITUTE OF TECHNOLOGY

for any matrix A we can find solutions of (1), (2) that are non 'skew-symmetric'.

[3] This paper which is related to our investigation was pointed out to us by the referee to whom we are indebted for other useful comments.

THE SUSPENSION OF THE GENERALIZED PONTRJAGIN COHOMOLOGY OPERATIONS

EMERY THOMAS

1. The main theorem. In a previous paper [9] I have defined a sequence of new cohomology operations, called the *generalized Pontrjagin operations*. These operations use as coefficient groups the summands of a certain type of graded ring: namely, *a ring with divided powers* (defined by H. Cartan in [1]), which is termed a Γ-ring in [9]. Let $A = \sum_k A_k$ be a ring with divided powers such that each summand A_k is a cyclic group of infinite or prime power order; we termed such rings *p-cyclic* in [9]. Then, the Pontrjagin operations are functions

$$\mathfrak{P}_t\colon H^{2n}(X;\ A_{2k}) \longrightarrow H^{2tn}(X;\ A_{2tk}) \quad (k,\ n > 0;\ t = 0,\ 1, \cdots)$$

where $H^q(Y,\ B;\ G)$ denotes the qth (singular) cohomology group of the pair $(Y,\ B)$ with coefficients in the group G.

Let C be a cohomology operation relative to integers r, s and coefficient groups G, Π. That is, C is a natural transformation

$$C\colon H^r(Y,\ B;\ G) \longrightarrow H^s(Y,\ B;\ \Pi)\ .$$

With each operation C we associate a second operation, $S(C)$, called the *suspension* of C. $S(C)$ is a natural transformation

$$H^{r-1}(Y,\ B;\ G) \longrightarrow H^{s-1}(Y,\ B;\ \Pi)\ ;$$

its definition is given in § 3.

The purpose of this note is to determine $S(\mathfrak{P}_t)$, where $\mathfrak{P}_t$ is the generalized Pontrjagin operation. In order to state our result concerning $S(\mathfrak{P}_t)$, we need an additional cohomology operation, the Postnikov square (see [3], [10]). This was defined in [9], but only for a restricted class of coefficient groups. In this paper we will define the Postnikov square as a cohomology operation

$$\mathfrak{p}\colon H^q(Y,\ B;\ A_{2k}) \longrightarrow H^{2q+1}(Y,\ B;\ A_{4k})\ , \qquad (q,\ k > 0)$$

where A_{2k} is an even summand of a p-cyclic ring with divided powers.

We now may state the main result of the paper.

THEOREM I. *For any cohomology operation C, let $S(C)$ denote the suspension of the operation C. Then,*

Received October 18, 1957, in revised form December 19, 1958. This research has been partly supported by U. S. Air Force contract AF 49 (638)-79.

(i) $$S(\mathfrak{P}_2) = \mathfrak{p}$$

(ii) $$S(\mathfrak{P}_t) = 0\ , \qquad (t > 2)$$

where 0 *denotes the zero cohomology operation.*

The proof of Theorem I is given in § 5. In § 2 we define the operation $\mathfrak{p}$, while in § 3 we give the definition of the suspension. In § 4 we discuss relative cohomology operations, while in § 6 we give some additional properties of the operation $\mathfrak{p}$. In particular, we show that $S(\mathfrak{p}) = 0$. Finally, the last section gives the theorem, $\delta S(C) = C\delta$, for any operation C.

I would like to thank Professor N. E. Steenrod for the valuable suggestions made to me at the time of revising the paper. In particular the definition of the suspension in § 3 and Theorem 7.1 are due to him.

2. The definition of the Postnikov square. The definition of the Postnikov square, $\mathfrak{p}$, is obtained by first defining a "model operation", p, which uses only a restricted category $\mathscr{C}$ of coefficient groups. The category $\mathscr{C}$ is defined as follows: let $Z_r = Z/rZ$ $(r = 0, 1, \cdots)$, where Z = integers = Z_0. Denote by $\mathscr{C}$ the category of all groups of the form Z_θ, where θ is zero or a power of a prime. For each group Z_θ in $\mathscr{C}$ we have defined a p-cyclic ring with divided powers,

$G(Z_\theta) = G_0(Z_\theta) + \cdots + G_t(Z_\theta) + \cdots$ (direct sum) (see [9; 1.17]).

In particular,

$$G_2(Z_\theta) = \begin{cases} Z_\theta\ , & \text{if } \theta \text{ is zero or odd} \\ Z_{2\theta}\ , & \text{if } \theta \text{ is a power of } 2. \end{cases}$$

We define a generator for $G_2(Z_\theta)$ by

$$g_2(1_\theta) = \begin{cases} 1_\theta\ , & \text{if } \theta \text{ is zero or odd} \\ 1_{2\theta}\ , & \text{if } \theta \text{ is a power of } 2 \end{cases}$$

where $1_r = 1 \bmod r$ $(r = 0, 1, \cdots)$. The group $G_2(Z_\theta)$ will be the coefficient domain for the operation $\mathfrak{p}$. As remarked in [9; § 2], once we have defined the operation $\mathfrak{p}$ for the category of *regular cell complexes*, the definition easily extends to the category of all topological spaces. Hence, in what follows we restrict attention to regular cell complexes, which we will simply term *complexes*.

Let K be a complex and L a subcomplex of K. Let Z_θ be a group in the category $\mathscr{C}$; that is, θ is zero or a power of a prime. We define an operation

$$p\colon H^q(K, L; Z_\theta) \longrightarrow H^{2q+1}(K, L; G_2(Z_\theta))$$

as follows. Let $u \in H^q(K, L; Z_\theta)$; let β be the homomorphism from Z_θ to $G_2(Z_\theta)$ given by $\beta(1_\theta) = \theta g_2(1_\theta)$. Define

(2.1) $$p(u) = \beta_*(u \cup \delta_* u) .$$

Here, δ_* is the Bockstein coboundary operator associated with the exact sequence

$$0 \longrightarrow Z \xrightarrow{\theta} Z \longrightarrow Z_\theta \longrightarrow 0 ,$$

and the cup-product is taken relative to the natural pairing $Z_\theta \otimes Z \approx Z_\theta$.

It is easily seen that this agrees with the usual definition of the operation p (see [3] and [10]). For let $\bar{u} \in C^q(K, L; Z)$ be a cochain representing u; that is, $\delta\bar{u} = \theta\bar{v}$, for some cochain $\bar{v} \in C^{q+1}(K, L; Z)$. Then, a cocycle representing $\beta_*(u \cup \delta_* u)$ is given by $\bar{u} \cup \delta\bar{u}$, which coincides with the definition given in [10].

In [9; 8.14] we defined a function w which goes from $H^q(K; Z_\theta)$ to $H^{2q+1}(K; Z_\theta)$. This function can be extended to the relative case, following the method given in § 4. When this is done it is easily shown that

(2.2) $$p(u) = \beta_* w(u) ,$$

a result we will need later.

The Postnikov square, $\mathfrak{p}$, is defined using the operation p as follows: let $u \in H^q(K, L; A_{2k})$, where A_{2k} is an even summand of a p-cyclic ring with divided powers. By hypothesis, A_{2k} is a cyclic group whose order is infinite or a power of a prime. Thus, there is an integer θ such that A_{2k} is isomorphic to Z_θ, where $Z_\theta \in \mathscr{C}$. Let ν be an isomorphism from A_{2k} to Z_θ. Then, by 3.1 in [9], for each non-negative integer r we have defined a homomorphism ζ_r mapping $G_r(Z_\theta)$ to A_{2rk}, which is an extension of ν^{-1}. We define the operation $\mathfrak{p}$ by

(2.3) $$\mathfrak{p}(u) = \zeta_2^* p \nu_*(u) ;$$

that is, $\mathfrak{p}$ is the composition of the following functions:

$$H^q(K, L; A_{2k}) \xrightarrow{\nu_*} H^q(K, L; Z_\theta) \xrightarrow{p}$$
$$H^{2q+1}(K, L; G_2(Z_\theta)) \xrightarrow{\zeta_2^*} H^{2q+1}(K, L; A_{4k}) .$$

We show the independence of this definition from the particular choice of the isomorphism ν (and hence ζ_2). This is a consequence of the fact that

(2.4) LEMMA. $$p\alpha_* = G_2(\alpha)_* p ,$$

where α is a homomorphism from Z_θ to a group Z_τ in $\mathscr{C}$, and $G_2(\alpha)$ is the homomorphism from $G_2(Z_\theta)$ to $G_2(Z_\tau)$ induced by the functor G (see [9; 1.23]).

Using 2.2, the proof of 2.4 is entirely similar to that given for 5.22 in [9] and is omitted here. From 2.4 the proof of the independence of the definition of $\mathfrak{p}$ follows along exactly the same lines as 3.5 and 3.6 in [9]; we omit the details.

3. Suspension of cohomology operations. The definition of the suspension used here is due to N. E. Steenrod[1]. Let I denote the unit interval, $[0, 1]$, and $\dot{I}$ the subspace $\{0\} \cup \{1\}$. The group $H^1(I, \dot{I}; Z)$ is cyclic infinite; let v be a fixed generator. For each space X and coefficient group G define a function ϕ from $H^q(X; G)$ to $H^{q+1}(I \times X, \dot{I} \times X; G)$ by

$$\phi(u) = v \times u \,. \tag{3.1}$$

We use singular cohomology for X, and the natural pairing $Z \otimes G \approx G$ for the cross-product. In § 7 we prove the following lemma.

(3.2) LEMMA. *The function ϕ is an isomorphism mapping $H^q(X; G)$ onto $H^{q+1}(I \times X, \dot{I} \times X; G)(q > 0)$.*

Consider now any cohomology operation C, which is defined on relative cohomology groups; say, C maps $H^r(X, A; G)$ to $H^s(X, A; \Pi)$ for each pair (X, A). Define an absolute cohomology operation, $S(C)$, which maps $H^{r-1}(Y; G)$ to $H^{s-1}(Y; \Pi)$, for each space Y, by

$$S(C)(u) = \phi^{-1} C\phi(u) \qquad (u \in H^{r-1}(Y; G)) \,. \tag{3.3}$$

Using the method described in § 4 we may extend $S(C)$ to an operation defined on relative cohomology groups, an operation which we continue to denote by $S(C)$. We wish to apply this construction to the operation $\mathfrak{P}_t$; as defined in [9], this is just an absolute operation. Thus, to use Definition 3.3 we must first extend the definition of $\mathfrak{P}_t$ to the relative case.

4. Relative cohomology operations. Let $O(q, r; G, \Pi)$ denote the set of absolute cohomology operations relative to dimensions q, r and coefficient groups G, Π; that is, if $C \in O(q, r; G, \Pi)$, then C: $H^q(X; G) \to H^r(X, \Pi)$ for each space X. As is well-known the set $O(q, r; G, \Pi)$ is in 1–1 correspondance with the group $H^r(K; \Pi)$, where K is an Eilenberg-MacLane space of type (G, q). The correspondance is obtained by assigning $C(\iota)$ to ι, where ι is the fundamental class in $H^q(K; G)$. Choose now a base point $e \in K$, and let α^*: $H^*(K, e; \Lambda) \approx H^*(K; \Lambda)$ be the isomorphism induced by the inclusion $K \subset (K, e)$. For any CW-complex X and subcomplex A, the homotopy classes of maps $(X, A) \to (K, e)$

[1] This definition has the advantage that it can be used in the case of cohomology with local coefficients.

are in one-to-one correspondance with $H^q(X, A; G)$. Thus we define a relative cohomology operation, C', associated with an absolute operation, C, as follows:

$$C'(u) = f^*\alpha^{*-1} C(\iota) , \tag{4.1}$$

where $u \in H^q(X, A; G)$ and f is a map $(X, A) \to (K, e)$ such that

$$f^*\alpha^{*-1}(\iota) = u .$$

With the operation C' defined, one is then interested in whether the properties of C extend to the operation C'. We now prove a general lemma which essentially asserts that all the properties of C' do carry over to C'.

Let $O(q_1, \cdots, q_n, r; G_1, \cdots, G_n, \Pi)$ denote the group of absolute cohomology operations, T, in n variables; that is, if $u_i \in H^{q_i}(X; G_i)$ $(i = 1, \cdots, n)$, then, $T(u_1, \cdots, u_n) \in H^r(X; \Pi)$. The operation T extends to a relative operation, T', using the method just given for operations of a single variable. Suppose now we are given absolute cohomology operations

$$C \in O(q_1, \cdots, q_n, r; G_1, \cdots, G_n, \Pi) ,$$
$$E \in O(s_1, \cdots, s_p, r; \Pi_1, \cdots, \Pi_p, \Pi) ,$$

and

$$D_i \in O(q_1, \cdots, q_n, s_i; G_1, \cdots, G_n, \Pi_i) \qquad (i = 1, 2, \cdots, p).$$

Let C', E' D'_i, be the corresponding relative operations.

(4.2) PROPOSITION. *Suppose that for each space X and cohomology classes $u_i \in H^{q_i}(X; G_i)(i = 1, \cdots, n)$, we have*

$$C(u_1, \cdots, u_n) = E(D_1(u_1, \cdots, u_n), \cdots, D_p(u_1, \cdots, u_n)) .$$

Then, for each pair (X, A) and classes $u'_i \in H^{q_i}(X, A; G_i)$ $(i = 1, \cdots, n)$, we have

$$C'(u'_1, \cdots, u'_n) = E'(D'_1(u'_1, \cdots, u'_n), \cdots, D'_p(u'_1, \cdots, u'_n)) .$$

We give the proof at the end of this section, first illustrating the theorem by giving several corollaries.

(4.3) COROLLARY 1. *Let $C \in O(q, s; R, S)$, $D_i \in O(q_i, s_i; R, S)$ $(i = 1, 2)$, where R, S are rings, $q = q_1 + q_2$, and $s = s_1 + s_2$. Suppose that*

$$C(u_1 \cup u_2) = D_1(u_1) \cup D_2(u_2)$$

for all classes $u_i \in H^{q_i}(X; R)$. Then,

$$C'(u'_1 \cup u'_2) = D'_1(u'_1) \cup D'_2(u'_2) ,$$

for all classes $u'_i \in (H^{q_i}(X, A; R)$.

Proof. Let $E_R \in O(q_1, q_2, q; R, R, R)$ and $E_S \in O(s_1, s_2, s; S, S, S)$ be the respective cup-products. Let F be the composite operation $C \circ E_R$. Using Proposition 4.2 we see that $F' = C' \circ E'_R$. But since $F(u_1, u_2) = E_S(D_1(u_1), D_2(u_2))$, again using 4.2 we see that

$$F'(u'_1, u'_2) = E'_S(D'_1(u_1), D_2(u'_2)) ;$$

that is,

$$C'(u'_1 \cup u'_2) = D'_1(u'_1) \cup D'_2(u'_2) ,$$

as was to be shown.

Let C, D_1, D_2 be the same operations as in Corollary 1. Then,

(4.4) COROLLARY 2. $C'(u'_1 \times u'_2) = D'_1(u'_1) \times D'_2(u'_2)$, *where* $u_i \in H^{q_i}(X_i, A_i; R)$ $(i = 1, 2)$.

Proof. Let $p_1\colon (X_1 \times X_2, A_1 \times X_2) \to (X_1, A_1)$, $p_2\colon (X_1 \times X_2, X_1 \times A_2) \to (X_2, A_2)$ be projections. Then,

$$u'_1 \times u'_2 = p_1^*(u'_1) \cup p_2^*(u'_2) .$$

Thus,

$$\begin{aligned} C'(u'_1 \times u'_2) &= C'(p_1^* u'_1 \cup p_2^* u'_2) = D'_1(p_1^* u'_1) \cup D'_2(p_2^* u'_2) \\ &= p_1^*(D'_1 u'_1) \cup p_2^*(D'_2 u'_2) = (D'_1 u'_1) \times (D'_2 u'_2) . \end{aligned}$$

Here we have used Corollary 1 and the naturality of the cohomology operations involved.

To apply this to the operations $\mathfrak{P}_t$, recall the way in which these operations were defined (see § 3 in [9]). We defined a set of "model operations", P_t, which used as coefficient groups only the groups of the category $\mathscr{C}$ (see § 2). The operations $\mathfrak{P}_t$ were then defined by composing the operation P_t with coefficient group homomorphisms; that is, precisely the same pattern as followed in Definition 2.3. Thus, the operations $\mathfrak{P}_t$ are defined in the relative case by simply applying the method given in this section to the operations P_t.

Let P'_t be the relative operation obtained from P_t. We note several facts needed later.

(4.5) LEMMA. *Let* $u_i \in H^{q_i}(X_i, A_i; Z_\theta)$ $(i = 1, 2)$, *where* $Z_\theta \in \mathscr{C}$. *Then*

(1) $$P'_t(u_1 \times u_2) = P'_t(u_1) \times P'_t(u_2) \quad (t \text{ odd}) .$$

If $t=2$ *and* θ *is a power of* 2, *then,*

$$(2) \qquad P_2'(u_1 \times u_2) = P_2'(u_1) \times P_2'(u_2) + \nu_*[Sq_1(u_1) \times \mu_* w(u_2) + \mu_* w(u_1) \times Sq_1(u_2)] .$$

Here, ν is the homomorphism of Z_2 to $G_2(Z_\theta)$ given by $\nu(1_2) = \theta g_2(1_\theta)$, and μ is the factor homomorphism $Z_\theta \to Z_2$. The functions Sq and w are defined respectively in 9.6 and 8.14 of [9].

Proof. The first statement is a consequence of Corollary 4.3 and the fact that the absolute operations P_t satisfy this formula[2]. Equation 4.5(2) was remarked in [9; § 13] for the absolute operations P_t, and the case dim u_i odd. But it follows from 8.12 in [9] that 4.5(2) holds in general. In fact Theorem 8.11 in [9] can be obtained at once from equation 4.5(2). The extension of the equation to the relative operation P_t', follows then from application of Proposition 4.2.

Combining Proposition 4.2 and 8.2 of [9] we also obtain

(4.6) LEMMA. *Let t be an an integer where $t = p_k \cdots p_1$ (p_i prime). Let $u \in H^{2q}(X, A; Z)$. ($Z \in \mathscr{C}$). Then,*

$$P_t'(u) = P_{p_k}' \circ \cdots \circ P_{p_1}'(u).$$

Since it is in fact the relative operation, P_t', we will work with, from now on we drop the prime, writing only P_t for both the relative and absolute operation.

Proof of Proposition 4.2. Let $Y = K(G_1, q_1) \times \cdots \times K(G_n, q_n)$, where each $K(G_i, q_i)$ is on Eilenberg-MacLane space of type (G_i, q_i). Let π_j: $Y \to K(G_j, q_j)$ $(j = 1, \cdots, n)$, be the projection map and set $\bar{\iota}_j = \pi_j^*(\iota_j)$, where ι_j is the characteristic class in $H^{q_j},(K(G_j, q_j); G_j)$. Let e_j be a base point in $K(G_j, q_j)$ and set $Y' = (K(G_1, q_1), e_1) \times \cdots \times (K(G_n, q_n), e_n)$. Let ι_j', $\bar{\iota}_j'$ be the equivalent of ι_j and $\bar{\iota}_j$. Then, Proposition 4.2 follows at once from the following three lemmas (we keep the same notation as used in Proposition 4.2)

$$(4.7) \qquad C(u_1, \cdots, u_n) = E(D_1(u_1, \cdots, u_n), \cdots, D_p(u_1, \cdots, u_n))$$

if and only if

$$C(\bar{\iota}_1, \cdots, \bar{\iota}_n) = E(D_1(\bar{\iota}_1, \cdots, \bar{\iota}_n), \cdots, D_p(\bar{\iota}, \cdots, \bar{\iota}_n)) .$$

$$(4.8) \qquad C'(u_1', \cdots, u_n') = E'(D_1'(u_1', \cdots, u_n'), \cdots, D_p'(u_1', \cdots, u_n'))$$

if and only if

$$C'(\bar{\iota}_1', \cdots, \bar{\iota}_n') = E'(D_1'(\bar{\iota}_1, \cdots, \bar{\iota}_n'), \cdots, D_p'(\bar{\iota}_1', \cdots, \bar{\iota}_n'))$$

[2] The operations $\mathfrak{P}_t$ are easily defined for odd dimensional classes: see [9; § 7].

(4.9) *If* $C(\bar{\iota}_1, \cdots, \bar{\iota}_n) = E(D_1(\bar{\iota}, \cdots, \bar{\iota}_n), \cdots, D_p(\bar{\iota}_1, \cdots, \bar{\iota}_n))$

then,

$$C'(\bar{\iota}'_1, \cdots, \bar{\iota}'_n) = E'(D'_1(\bar{\iota}'_1, \cdots, \bar{\iota}'_n), \cdots, D'_p(\bar{\iota}'_1, \cdots, \bar{\iota}'_n)) .$$

We give only the proof of Lemma 4.7, the others being entirely similar. Assume first we are given classes $u_i \in H^{q_i}(X; G_i)$ $(i = 1, \cdots, n)$. Let f_j: $X \to K(G_j;\ q_j)$ be mappings such that $f_j^*(\iota_j) = u_j$. Set $f = f_1 \times \cdots \times f_n$: $X \to Y$. Then, by naturality, one has

(4.10) (a) $\qquad C(u_1, \cdots, u_n) = f^*C(\bar{\iota}_1, \cdots, \bar{\iota}_n)$,

(b) $\qquad D_i(u_1, \cdots, u_n) = f^*D_i(\bar{\iota}_1, \cdots, \bar{\iota}_n) \qquad (i = 1, \cdots, p).$

Suppose now that

$$C(\bar{\iota}_1, \cdots, \bar{\iota}_n) = E(D_1(\bar{\iota}_1, \cdots, \bar{\iota}_n), \cdots, D_p(\bar{\iota}_1, \cdots, \bar{\iota}_n)) .$$

Then, by 4.10,

$$C(u_1, \cdots, u_n) = f^*E(D_1(\bar{\iota}_1, \cdots, \bar{\iota}_n), \cdots, D_p(\bar{\iota}_1, \cdots, \bar{\iota}_n)) .$$

But E is natural with respect to mappings. Therefore,

$$\begin{aligned} &f^*E(D_1(\bar{\iota}_1, \cdots, \bar{\iota}_n), \cdots, D_p(\bar{\iota}_1, \cdots, \bar{\iota}_n)) \\ &= E(f^*D_1(\bar{\iota}_1, \cdots, \bar{\iota}_n), \cdots, f^*D_p(\bar{\iota}_1, \cdots, \bar{\iota}_n)) \\ &= E(D_1(u_1, \cdots, u_n), \cdots, D_p(u_1, \cdots, u_n)) , \end{aligned}$$

again by 4.10, which completes the proof of this assertion. The proof in the opposite direction is trivial.

5. The proof of Theorem I. Recall that the operation $\mathfrak{P}_t$ is defined by means of the model operations P_t and coefficient group homomorphisms. But it is clear that the isomorphism ϕ, defined in 3.1, commutes with coefficient group homomorphisms. Thus, it suffices to prove Theorem I with $\mathfrak{P}_t$ replaced by P_t, the operation $\mathfrak{p}$ replaced by p, and the group A_{2k} taken to be a group in the category $\mathscr{C}$, say $A_{2k} = Z_\theta$.

Assume first that t is an odd prime p. Since ϕ is an isomorphism, the proof of Theorem I (ii) consists simply in showing

$$P_p\phi(u) = 0 , \qquad u \in H^r(X;\ Z_\theta).$$

But this is immediate; for

$$P_p\phi(u) = P_p(v \times u) = P_p(\bar{v} \times u) = P_p(\bar{v}) \times P_p(u) ,$$

by Lemma 4.5(1). Here, $\bar{v}$ is a generator of $H^1(I,\ \dot{I};\ Z_3)$. However, $P_p(\bar{v}) = 0$, by dimensionality considerations. Thus, $P_p\phi(u) = 0$; and hence, $S(P_p) = 0$.

Now, suppose that t is any integer > 1 which is not a power of 2; say, $t = mp$, where p is an odd prime. Then, by Lemma 4.6

$$P_t\phi(u) = P_m \circ P_p\phi(u) = P_m(0) = 0 \ .$$

Consequently,

$$S(P_t) = 0 \ .$$

Thus, we have proved Theorem I(ii) for the case t is not a power of 2. Before concluding the proof of part (ii), we must prove part (i). Let the classes u and v be as above, where u has coefficients in the group Z_θ. If θ is zero or odd, then by Proposition 7.4 in [9], we have

$$P_2(v \times u) = P_2(\bar{v} \times u) = (\bar{v} \times u)^2 = \pm \bar{v}^2 \times u^2 = 0 \ ,$$

since $\bar{v}^2 = 0$. Thus, in this case $S(P_2) = 0$. Suppose now that θ is a power of 2.

Let η be the factor map $Z \to Z_\theta$. Then, $v \times u = (\eta_* v) \times u$, where the right hand side uses the pairing $Z_\theta \otimes Z_\theta \approx Z_\theta$. Thus, using Lemma 4.5(2), we have

$$\begin{aligned} P_2(v \times u) = P_2(\eta_* v \times u) &= P_2(\eta_* v) \times P_2(u) \\ &+ \nu_*[Sq_1(\eta_* v) \times \mu_* w(u) + \mu_* w(\eta_* v) \times Sq_1(u)] \ . \end{aligned}$$

Now, $P_2(\eta_* v) = 0$, $w(\eta_* v) = 0$ by dimensionality considerations. Also, since $\eta_* v$ is a 1-dimensional class, $Sq_1(\eta_* v) = \xi_* v$, where ξ is the natural map $Z \to Z_2$ (see Steenrod [4; 12.6]). Thus,

$$P_2(v \times u) = \nu_*[\xi_* v \times \mu_* w(u)] \ . \tag{5.1}$$

Consider the following commutative diagram:

$$\begin{array}{ccc} Z \otimes Z_\theta & \xrightarrow{1 \otimes \beta} & Z \otimes G_2(Z_\theta) \\ {\scriptstyle \xi \otimes \mu} \downarrow & & \approx \downarrow {\scriptstyle \omega} \\ Z_2 \otimes Z_2 \overset{\omega'}{\approx} Z_2 & \xrightarrow{\nu} & G_2(Z_\theta) \ , \end{array}$$

where β is the homomorphism of Z_θ to $G_2(Z_\theta)$ given by $\beta(1_\theta) = \theta g_2(1_\theta)$ (see 2.1). Then, from 5.1,

$$\begin{aligned} P_2(v \times u) &= \nu_* \omega'_* (\zeta \otimes \mu)_* [v \otimes w(u)] \\ &= \omega_*(1 \otimes \beta)_*[v \otimes w(u)] \\ &= v \times \beta_* w(u) \\ &= v \times p(u) \ , \ \text{by } 2.2 \ . \end{aligned}$$

Therefore,

$$P_2\phi(u) = P_2(v \times u) = v \times p(u) = \phi p(u) .$$

That is,

$$S(P_2) = p .$$

This proves part (i) of Theorem I. To complete the proof of the theorem we must show that

$$P_{2^r}\phi(u) = 0 , \qquad (r > 1).$$

But by part (i) of Theorem I and Lemma 4.6, we have

$$\begin{aligned} P_{2^r}\phi(u) &= P_{2^{r-1}} P_2\phi(u) = P_{2^{r-1}} \phi p(u) \\ &= P_{2^{r-2}} P_2\phi p(u) = P_{2^{r-2}} \phi p(p(u)) = 0 . \end{aligned}$$

Here, we use property 6.6 of the function p, which is proved independently in the next section. This completes the proof of Theorem I.

6. The properties of the operation $\mathfrak{p}$. We give here the main properties of the Postnikov square, $\mathfrak{p}$.

(6.1) THEOREM. *Let X be a space, and let $A = \sum_k A_k$ be a p-cyclic ring with divided powers. Suppose that $u \in H^q(X; A_{2k})$ $(q, k > 0)$. Then,*[3]

(6.2) $\mathfrak{p}(u) = 0$, *if order A_{2k} is odd or infinite,*

(6.3) $2\mathfrak{p}(u) = 0$,

(6.4) $\mathfrak{p}$ *is a homomorphism,*

(6.5) *if order $A_{2k} = 2^i$ $(i > 1)$ and $2u = 0$, then $\mathfrak{p}(u) = 0$,*

(6.6) $\mathfrak{p}(\mathfrak{p}(u)) = 0$,

(6.7) $f^*\mathfrak{p}(u) = \mathfrak{p}f^*(u)$,

(6.8) $\alpha_*\mathfrak{p}(u) = \mathfrak{p}\alpha_*(u)$,

where f^* is induced by a map f from a space Y to X, and α_* is induced by a homomorphism α from A to a p-cyclic ring with divided powers A'.

The proof of Theorem 6.1 falls into 2 parts. Suppose first that we have proved 6.2 through 6.7 with the operation $\mathfrak{p}$ replaced by the operation p, and the coefficient group A_{2k} restricted to be a group in the category $\mathscr{C}$. Then, the proof of 6.2–6.7 for the general case of the

[3] With the exception of 6.5 and 6.6, these properties are noted by J. H. C. Whitehead in [**10**].

function $\mathfrak{p}$ follows at once, using definition 2.3; that is, $\mathfrak{p} = \zeta_2{}^* p \nu_*$. In particular, 6.2–6.5 are simple consequences of the fact that $\zeta_2{}^*$ and ν_* are homomorphisms; 6.6 follows from 6.3 and 6.5, and 6.7 follows from the fact that f^* commutes with all coefficient group homomorphisms. Finally, to prove 6.8 for the operation $\mathfrak{p}$, one uses 2.4 and exactly the same argument as that used to prove I(9) in § 4 of [9]. Thus, we are left with proving 6.2 through 6.7 for the operation p. Let $u \in H^q(K; Z\ )$, where $Z_\theta \in \mathscr{C}$. Then,

(i) $$p(u) = 0\ , \quad \textit{if } \theta \textit{ is zero or odd.}$$

This follows at once from 2.1. For if θ is zero or odd, the homomorphism β is zero.

(ii) $$2p(u) = 0$$

This again is immediate from 2.1; for it is always the case that $2\beta = 0$.

(iii) $$p \textit{ is a homomorphism}$$

In § 5 we showed that the operation p is the suspension of the operation P_2. But by 7.4 in [6], all operations which are suspensions are homomorphisms.

(iv) $$\textit{If } \theta = 2^i\ (i > 1), \textit{ and } 2u = 0, \textit{ then } p(u) = 0.$$

Since $2u = 0$, we may use Lemma 13.3 of [9]: namely, there are classes $x \in H^{q-1}(K;\ Z_2)$ and $y \in H^q(K;\ Z_2)$ such that

$$u = \lambda_* \delta_*(x) + \nu_*(y)\ ,$$

where δ_* is the coboundary associated with the exact sequence

$$0 \longrightarrow Z \xrightarrow{\ 2\ } Z \longrightarrow Z_2 \longrightarrow 0\ ,$$

λ is the natural factor map $Z \to Z_\theta$, and ν maps Z_2 to Z_θ by $\nu(1_2) = (\theta/2)1_\theta$ (recall that $\theta = 2^i,\ i > 1$). Hence, by (iii) above,

$$\begin{aligned} p(u) &= p\lambda_*\delta_*(x) + p\nu_*(y) \\ &= G_2(\lambda)_* p\delta_*(x) + G_2(\nu)_* p(y) \\ &= G_2(\nu)_* p(y)\ , \end{aligned}$$

by 2.4 and (i) above, since $\delta_*(u)$ has integer coefficients. Now,

$$G_2(\nu)_* p(y) = G_2(\nu)_* \beta_* w(u)\ ,$$

by 2.2. We show that $p(u) = 0$ by showing that

$$G_2(\nu)\beta = 0\ .$$

From Definition 2.1 we recall that β maps Z_2 to $G_2(Z_2)$ by $\beta(1_2) = 2g_2(1_2)$. Hence, using 1.21 and 1.24 in [9],

$$\begin{aligned} G_2(\nu)\beta(1_2) &= 2G_2(\nu)g_2(1_2) = 2g_2(\nu 1_2) \\ &= 2g_2((\theta/2)1_\theta) = 2(\theta^2/4)g_2(1\ ) = (\theta^2/2)1_{2\theta} = 0 . \end{aligned}$$

For, $\theta^2/2 = 2^{2i}/2 = 2^{2i-1}$; and, $2\theta = 2^{i+1}$. But by hypothesis, $i \geqq 2$; thus $2i - 1 \geqq i + 1$.

(v) $$p(p(u)) = 0$$

This follows at once from (ii) and (iv) above.

(vi) $$f^* p(u) = p f^*(u) .$$

This is simply a special case of Theorem 3.6 of [7]. This, then completes the proof of Theorem 6.1.

We consider one more property of the operation $\mathfrak{p}$: namely, its behaviour with respect to suspension. We continue to denote by $S(C)$ the suspension of a cohomology operation C.

(6.9) PROPOSITION. *$S(\mathfrak{p}) = 0$, where 0 denotes the trivial cohomology operation.*

Proof. By the same reasoning as given in § 5, it sufficies to prove Proposition 6.9 with $\mathfrak{p}$ replaced by the operation p, and the coefficient group A_{2k} taken to be a group in the category $\mathscr{C}$, say $A_{2k} = Z$. Thus, we need simply show that $p\phi(u) = 0$, where $u \in H^q(L; Z_\theta)$. Now by Nakaoka [2] we have[4]:

$$p(v_1 \times v_2) = P_2(v_1) \times p(v_2) + p(v_1) \times P_2(v_2) ,$$

for classes $v_i \in H^{q_i}(X_i, A_i; Z_\theta)$ $(i = 1, 2)$.
Thus,

$$p\phi(u) = p(\bar{v} \times u) = P_2(\bar{v}) \times p(u) + p(\bar{v}) \times P_2(u) = 0 ,$$

since $P_2(\bar{v}) = p(\bar{v}) = 0$ by dimensionality considerations. Here, $\bar{v}$ is the image of v in $H^1(I, \dot{I}; Z_\theta)$. Hence, $S(p) = 0$, as was to be proved.

7. The relation $\delta S(C) = C\delta$. We give here a theorem, whose proof is due to N. E. Steenrod.

(7.1) THEOREM. *Let C be a cohomology operation, and let δ be the relative cohomology coboundary operator. Then,*

[4] Nakaoka only proves this for the case dim v_1, v_2 even; but the result is true in general, as is easily shown using Definition 2.1.

$$\delta S(C) = C\delta ,$$

where $S(C)$ *is the suspension of* C.

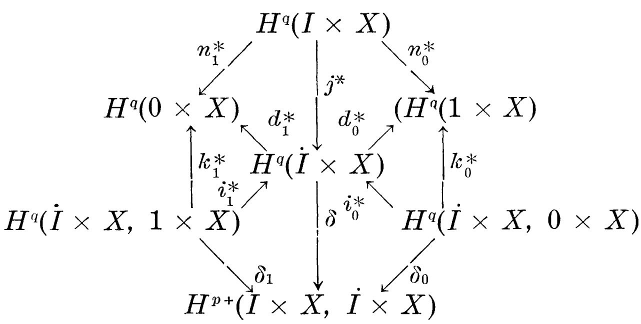
We sketch the proof; let X be a space and $A \subset X$ a subspace. Let X' denote the mapping cylinder of the inclusion map $A \subset X$. That is, unite $I \times A$ and X by identifying $1 \times A$ with A in X. Let $A' = 0 \times A$. The inclusions

$$(X', A') \longrightarrow (X', I \times A) \longleftarrow (X, A)$$

induce isomorphisms of the cohomology sequence of (X, A) and (X', A') with local cosfficients. Thus, we may discuss the behaviour of the coboundary δ in the cohomology sequence of the pair (X', A').

Consider the following hexagonal diagram (see [8], page 42):

(7.2)



Here all homomorphisms other than δ, δ_1, and δ_2 are induced by inclusions. Standard arguments, using exactness and homotopy equivalence, show that the arrows around the peripheries are isomorphisms. We agree to identify $H^q(X)$ with $H^q(0 \times X)$ by sending $u \to e \times u$, where e is the unit of $H^0(0; Z)$. At the end of this section we will use diagram 7.2 to prove the following lemma:

(7.3) LEMMA. *Let* ϕ *be the function defined in* 3.1. *Then,*

$$\phi = \delta_1 k_1^{*-1} ,$$

where k_1^*, δ_1 *are the functions defined in diagram* 7.2

Notice that this proves Lemma 3.2; for the functions δ_1, k_1^* are isomorphisms. Now let g^*: $H^{q+1}(X', A' \cup X) \to H^{q+1}(I \times A, \dot{I} \times A)$ be induced by the inclusion. Using the fact that $\dot{I}$ is a strong deformation retract of a neighborhood of $\dot{I}$ in I (see [8]; Chapter 1, 11.6), together with excision, one shows that g^* is an isomorhism onto.

(7.4) LEMMA. *The following diagram is commutative, where* f^* *is induced by the inclusion*

$$\begin{array}{ccc} H^{q+1}(I \times A,\ \dot{I} \times A) & \xrightarrow{g^{*-1}} & H^{q+1}(X',\ A' \cup X) \\ \uparrow \phi & & \downarrow f^* \\ H^q(A') & \xrightarrow{\ \ \delta\ \ } & H^{q+1}(X',\ A') . \end{array}$$

Thus $\delta = f^* g^{*-1} \phi$.

This is a consequence of Lemma 7.3 and commutativity relations in a slightly enlarged diagram. We omit the details.

The proof of Theorem 7.1 is an immediate consequence of Lemma 7.4. For let $u \in H^q(A')$. Then, by this lemma,

$$C\delta(u) = Cf^* g^{*-1} \phi(u) .$$

Using the naturality of the operation C, we have

$$Cf^* g^{*-1} \phi(u) = f^* g^{*-1} C\phi(u) .$$

But by Definition 3.1, $C\phi = \phi S(C)$.
Thus,

$$C\delta(u) = f^* g^{*-1} \phi S(C)(u) = \delta S(C)(u) ,$$

again using Lemma 7.4. This completes the proof of Theorem 7.1.

Proof of Lemma 7.3. We apply diagram 7.2 to the case $X = \varnothing$, $q = 0$, and coefficient group = integers. Then, the unit class of $H^0(\dot{I};\ Z)$ can be represented as a sum $v_0 + v_1$, where

$$v_0 = i_1^* k_1^{*-1} d_1^*(v_0 + v_1),\ \ v_1 = i_0^* k_0^{*-1} d_0^*(v_0 + v_1).$$

Thus,

$\delta(v_0) = -\ \delta(v_1) = v =$ a generator of $H^1(I,\ \dot{I};\ Z)$. Therefore, by Definition 3.1,

$$\phi(u) = v \times u = (\delta v_0) \times u .$$

But by the axioms for the cross-product, we may write

$$(\delta v_0) \times u = \delta(v_0 \times u) .$$

Furthermore, we have

$$v_0 = i_1^* k_1^{*-1}(e) ,$$

where $e = d_1^*(v_0 + v_1) =$ unit of $H^0(0;\ Z)$. Thus,

$$\begin{aligned} \delta(v_0 \times u) &= \delta(i_1^* k_1^{*-1}(e) \times u) \\ &= \delta i_1^* k_1^{*-1}(e \times u) = \delta_1 k_1^{*-1}(e \times u) . \end{aligned}$$

Here we have used the naturality of the cross-product and the commutativity of diagram 7.2. If we now identify $H^q(X)$ with $H^q(0 \times X)$ by sending $u \to e \times u$, we then have

$$\phi(u) = \delta(v_0 \times u) = \delta_1 k_1^{*-1}(u) ,$$

as was asserted.

REFERENCES

1. H. Cartan, *Seminaire* H. Cartan 1954/55, Paris (mimeographed).
2. M. Nakaoka, *Note on cohomological operations*, J. Inst. Polytech., Osaka City Univ., 4 (1953), 51-58.
3. M. Postnikov, *The classification of continuous mappings of a 3-dimensional polyhedron into a simply connected polyhedron of arbitrary dimension*, C. R. (Doklady) Acad. Sci. U. R. S. S., **64**(1949), 461-462.
4. N. Steenrod, *Products of cocycles and extensions of mappings*, Ann. Math., **48** (1947), 290-320.
5. ———, *Homology groups of symmetric groups and reduced operations*, Proc. Nat. Acad. Sci. U. S. A., **39** (1953), 213-223.
6. ———, *Cohomology operations*, Symposium Int. Topologie algebraica, Univ. de Mexico (1957), (to appear).
7. ———, *Cohomology operations derived from the symmetric group*, Comment. Math. Helv., **31** (1957), 195-218.
8. N. Steenrod and E. Eilenberg, *Foundations of Algebraic Topology*, Princeton University Press, 1952.
9. E. Thomas, *The generalized Pontrjagin cohomology operations and rings with divided powers*, Memoir Number 27 A. M. S. (1957).
10. J. H. C. Whitehead, *On the theory of obstructions*, Ann. Math., **54** (1951), 68-84.

UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA

ON TCHEBYCHEFF POLYNOMIALS

J. L. ULLMAN

1. Introduction. Let C be a closed bounded set having an infinite number of points. There is a unique polynomial $T_n(z)$ of degree n, and with one as coefficient of z^n, such that if $P_n(z)$ is any other polynomial with the same normalization,

$$M_n = \max_{z \in C} |T_n(z)| < \max_{z \in C} |P_n(z)| . \tag{1.1}$$

This is the Tchebycheff polynomial of degree n associated with C.

1.1. Assume that C has positive capacity, used throughout to mean logarithmic capacity, and a connected complement D. The conductor potential for such C is a real valued function $U(z)$ defined in D with the properties: (1.2) $U(z)$ is harmonic at finite points of D, (1.3) $U(z) - \log |z|$ is regular at infinity and zero there, (1.4) there is a number $\rho > -\infty$ such that $U(z) > \rho$ for z in D, (1.5) if $\{z_i\}$ is a convergent sequence of points with limit point on the boundary of D, then $\lim U(z_i) = \rho$, except perhaps when the limit point belongs to a subset of the boundary of capacity zero. The function $U(z)$ has a unique representation as a Lebesgue-Stieltjes integral

$$U(z) = \int \log |z - t| d\mu . \tag{1.6}$$

where μ is a completely additive, positive set function defined for Borel measurable sets, if it is specified that the carrier of μ consist of boundary points of D. [2].

1.2. Fejér [1] proved that the zeros of $T_n(z)$ lie in the convex hull H of C. The consequence

$$|z_{ni}| \leqq R , \tag{1.7}$$

where z_{ni} is a zero of $T_n(z)$, and R is a finite constant independent of n, will be sufficient for later reference. Let

$$\rho_n = \frac{1}{n} \log M_n . \tag{1.8}$$

Szegö [3] proved that

Received April 5, 1954, in revised form February 24, 1959. The author has received support, which he gratefully acknowledges, from the O.N.R. contract Nonr-330(00) administered by the University of Michigan. The results of this paper were presented in Abstract 515 appearing in the Bulletin of the American Mathematical Society, Vol. 60, July 1954.

(1.9) $$\lim \rho_n = \rho \ ,$$

where ρ is essentially defined for a set C of positive capacity in §1.1, and is taken as zero when C has zero capacity. If C does not have a connected complement, ρ is obtained by taking for D in §1.1 the unbounded component of the complement of C. The above results in conjunction with an argument due to R. Nevanlinna [2, p. 127], can be used to show that

(1.10) $$\lim \frac{1}{n} \log |T_n(z)| = U(z) \ ,$$

for z in the complement of H. The following results concern the extension of (1.10) to points of D in H.

1.3. Summary of results. Let C be a closed, bounded set of positive capacity, and with connected complement D. Let $\nu_n(S)$ be the total multiplicity of the zeros of $T_n(z)$ in the set S. If E is a closed subset of D, then

(I) $$\lim \frac{\nu_n(E)}{n} = 0 \ ,$$

and

(II) $$\lim \int_E \left| \frac{1}{n} \log |T_n(z)| - U(z) \right| dA = 0 \ .$$

If Γ is a continuously differentiable curve consisting of points of D, and with interior denoted by $I(\Gamma)$, then

(III) $$\lim \frac{\nu_n I((\Gamma))}{n} = \mu(I(\Gamma)) \ .$$

The set function μ is defined by (1.6). In the case D is bounded by a finite number of analytic, Jordan curves, then

(IV) $$\nu_n(E) < P \ ,$$

where P is a constant depending on E, but not on n. Also in this case

(V) $$\lim \frac{1}{n} \log |T_n(z)| = U(z) \ ,$$

for z in E, with the possible exception of a set of measure zero.

2. The results concerning the zeros of $T_n(z)$, namely (I) and (IV), are established first.

2.1. LEMMA 1. *Associated with D is a set of domains $\{D_n\}$,*

$n = 1, 2, \cdots$, *with the properties:*

(a) D_n *is an unbounded domain,*

(b) *the closure of* D_n *is contained in* D_{n+1}, *that is* $\bar{D}_n \subset D_{n+1}$,

(c) *each point of* D *is contained in some* D_n.

LEMMA 2. *Let* $u(z)$ *be harmonic at finite points of* D *and regular at infinity. Furthermore, if* $\{z_i\}$ *is a convergent sequence of points with limit point on the boundary of* D, *suppose that* $\liminf u(z_i) \geqq 0$, *except possibly if the limit point belongs to a subset of the boundary of capacity zero. If, in the exceptional cases,* $\liminf u(z_i) \geqq -\gamma$, $0 \leqq \gamma < \infty$, *then in fact* $\gamma = 0$, *and* $u(z) \geqq 0$, *for* z *in* D. [2].

2.2. The generalized Green's function of D with pole at w, $G(z, w)$, where the variable z and the parameter w are points of D, has the properties:

(2.1) $G(z, w) > 0$,

(2.2) $G(z, w)$ is harmonic in z, except if $z = w$, and is regular at infinity,

(2.3) $G(z, w) + \log |z - w|$ is regular when $z = w$,

(2.4) if $\{z_i\}$ is a convergent sequence of points with limit point on the boundary of D, then $\lim G(z_i, w)$ exists, and is equal to zero, except perhaps if the limit point belongs to a subset of the boundary of capacity zero, and

(2.5) at the exceptional points $\limsup G(z_i, w) \leqq M < \infty$, a constant depending on w, but not on $\{z_i\}$. When $w = \infty$,

(2.6) $G(z, \infty) = U(z) - \rho$, and

(2.7) for finite or infinite w, $G(z, w) = G(w, z)$.

2.3. LEMMA 3. *To each domain* D_k *there is a positive constant* m_k, *such that*

$$\rho_n - \rho \geqq m_k \frac{\nu_n(D_k)}{n} . \tag{2.8}$$

Proof. Let

$$u_n(z) = \frac{1}{n} \log |T_n(z)| , \tag{2.9}$$

and let $z_{n1}, \cdots, z_{nm}$, $m \leqq n$, be the zeros of $T_n(z)$ in D. The convention used in listing zeros will be to repeat multiple zeros according to their multiplicity. Consider the function

$$v_n(z) = (\rho_n - u_n(z)) + (U(z) - \rho) - \frac{1}{n}(G(z, z_{n1}) + \cdots + G(z, z_{nm})) , \tag{2.10}$$

$$= A_1(z) + A_2(z) - A_3(z) \,. \tag{2.11}$$

Let $\{z_i\}$ be a convergent sequence of points of D with limit point on the boundary. Now, $\lim A_1(z_i) \geqq 0$ by (1.1), (1.8) and (2.9), $\liminf A_2(z_i) \geqq 0$ by (1.4), and $\lim A_3(z_i) = 0$, except possibly if the limit point belongs to a subset of the boundary of capacity zero. In the exceptional case $\limsup A_3(z_i) \leqq M < \infty$, by (2.5). In addition $v_n(z)$ is harmonic in D and regular at infinity. The conditions of Lemma 2 are thus satisfied so that

$$v_n(z) \geqq 0 \,, \tag{2.12}$$

for z in D. Let $z_{n1}, \cdots, z_{np}$, $p \leqq m$, be the zeros of $T_n(z)$ in D_k. Then, by (2.1), (2.7), (2.10), (2.12),

$$\rho_n - \rho - (u_n(z) - U(z)) \geqq \frac{1}{n}(G(z_{n1}, z) + \cdots + G(z_{np}, z)) \,. \tag{2.13}$$

If m_k is the lower bound of $G(z, \infty)$ on D_k, then the value of (2.13) at $z = \infty$ yields (2.8).

2.4. *Proof of* (I). The set E will be contained in an element of $\{D_n\}$, say D_k. Hence by (2.8) and the definition of $\nu_n(S)$,

$$\frac{\nu_n(E)}{n} \leqq \frac{\nu_n(D_k)}{n} \leqq \frac{\rho_n - \rho}{m_k} \,. \tag{2.14}$$

The result then follows by (1.9).

2.5. *Proof of* (IV). Szegö [4] has shown, under the added restriction on D, that

$$\rho_n - \rho \leqq \frac{K}{n} \,, \tag{2.15}$$

where K is a constant not depending on n. This together with (2.8) yields

$$\nu_n(D_k) \leqq \frac{K}{m_k} \,. \tag{2.16}$$

Thus if D_k contains E, the assertion follows.

3. The next results proved are (II) and (V) concerning the mean convergence in the general case, and the point wise convergence in a special case, of the sequence $u_n(z) = 1/n \log |T_n(z)|$.

3.1. Let D_k again be a domain containing E. Assign to each point of E a circle centered at the point, lying in D_k, and with radius not exceeding 1/3. By the Heine-Borel theorem, a finite number of circles cover E. Hence it is sufficient to prove (II), replacing E by a circle in D_k with radius less than 1/3.

3.2. Let $s_{n1}, \cdots, s_{nn_1}$ be the zeros of $T_n(z)$ in the complement of D_{k+1} and let $r_{n1}, \cdots, r_{nn_2}$ be the zeros in D_{k+1}. By the convention of listing multiple zeros, $n_1 + n_2 = n$. Note that by (I),

$$\lim \frac{n_2}{n} = 0 \,. \tag{3.1}$$

Next define

$$S_n(z) = \prod_{i-1}^{n_1} (z - s_{ni}) \,, \tag{3.2}$$

and

$$R_n(z) = \prod_{i=1}^{n_2} (z - r_{ni}) \,. \tag{3.3}$$

Now

$$\left|\frac{1}{n}\log|T_n(z)| - U(z)\right| \tag{3.4}$$

$$= \left|\frac{n_1}{n}\frac{1}{n_1}\log|S_n(z)| + \frac{1}{n}\log|R_n(z)| - U(z)\right| \tag{3.5}$$

$$\leqq \frac{n_1}{n}\left|\frac{1}{n_1}\log|S_n(z)| - U(z)\right| + \frac{n_2}{n}|U(z)| + \frac{1}{n}\left|\log|R_n(z)|\right| \,. \tag{3.6}$$

It will be shown in §4.3 that the first term of (3.6) tends to zero uniformly in E. Also in E, $|U(z)|$ has a finite upper bound, so by (3.1), the second term also tends uniformly to zero in E.

3.3. *Proof of* (II). By the remarks of §§3.1 and 3.2, it is sufficient to prove

$$\lim \frac{1}{n}\int_{|z-a|<\delta} \log|R_n(z)|dA_z = 0 \,, \tag{3.7}$$

where $|z - a| < \delta$ is a subset of D_k and $\delta \leqq 1/3$. Let

$$\frac{1}{n}\log|R_n(z)| = \int \log|z - t|d\mu_n \,. \tag{3.8}$$

The integral in (3.7) then has the upper bound

$$\int_{|z-a|<\delta,}\left|\int_{|t-a|<2\delta} \log|z - t|d\mu_n\right|dA_z + \int_{|z-a|<\delta,}\left|\int_{|t-a|\geqq 2\delta} \log|z - t|d\mu_n\right|dA_z \,. \tag{3.9}$$

By (1.7) $\mu_n(S)$ is zero for any set S in the exterior of $|z| = R$. Hence the second integral in (3.9) is bounded by

$$\pi\delta^2 \frac{n_2}{n} \max\{|\log|R+\delta||, |\log|\delta||\} . \tag{3.10}$$

This tends to zero by (3.1). The first integral can be written

$$\int_{|z-a|<\delta,} \left(\int_{|t-a|<2\delta} \log \frac{1}{|z-t|} d\mu_n \right) dA_z , \tag{3.11}$$

since

$$|z-t| \leqq |z-a| + |t-a| < 3\delta \leqq 1 . \tag{3.12}$$

The order of integration can be changed, to yield

$$\int_{|t-a|<2\delta} \left(\int_{|z-a|<\delta} \log \frac{1}{|z-t|} dA_z \right) d\mu_n , \tag{3.13}$$

or

$$\int_{|t-a|<2\delta} g(t) d\mu_n , \tag{3.14}$$

where

$$g(t) = \begin{cases} \pi\delta^2 \log \dfrac{1}{|t-a|}, & \delta \leqq |t-a| < 2\delta , \\ \pi\delta^2 \log \dfrac{1}{\delta} + \dfrac{\pi}{2}(\delta^2 - |t-a|^2), & 0 \leqq |t-a| < \delta . \end{cases} \tag{3.15}$$

From this it follows that an upper bound for (3.11) is

$$\frac{n_2}{n} g(a) . \tag{3.16}$$

This tends to zero by (3.1).

3.4. *Proof of* (V). The contents of §3.2, in particular (3.6), reduce the proof to showing

$$\lim \frac{1}{n} \log |R_n(z)| = 0 , \tag{3.17}$$

for z in E, except possibly for a set of measure zero. Fy (IV) there are less than P zeros in E for each n, and each of these, by (1.7) is inside or on the circle $|z| = R$. Hence it is sufficient to show

$$\lim r_n(z) = \lim \frac{1}{n} |\log|z - a_n|| = 0 , \tag{3.18}$$

where $|a_n| \leqq R$, for $|z| < Q$, a disc covering E, with the possible exception of a set T of measure zero. For a fixed integer $k > 0$,

$$r_n(z) > \frac{1}{k} \tag{3.19}$$

either if

$$|z - a_n| > \exp\left(\frac{n}{k}\right), \tag{3.20}$$

or

$$|z - a_n| < \exp\left(-\frac{n}{k}\right). \tag{3.21}$$

Now (3.20) will ultimately fail to hold since $|z - a_n| \leqq R + Q$. Let $T(k)$ be the set of z for which (3.21) holds infinitely often, and let $T(k, p)$ be the set where (3.21) holds for some $n \geqq p$. It is clear that

$$T(k) \subset T(k, p). \tag{3.22}$$

Hence if $m_e(S)$ designates the exterior measure of a set S,

$$\begin{aligned} m_e(T(k)) \leqq m_e(T(k, p)) &\leqq \pi \sum_{n=p}^{\infty} \exp\left(\frac{-2n}{k}\right) \\ &= \exp\left(\frac{-2p}{k}\right)\left(1 - \exp\left(\frac{-2}{k}\right)\right)^{-1}. \end{aligned} \tag{3.23}$$

This bound holds for all values of p. Thus the exterior measure of $T(k)$, and hence its measure, is zero. Since T is the set where

$$\limsup r_n(z) > 0, \tag{3.24}$$

each point of T is contained in one of the sets $T(k)$. There are a denumerable number of the latter, each having measure zero. T thus has measure zero.

4. Let

$$s_n(z) = \frac{1}{n_1} \log |S_n(z)|. \tag{4.1}$$

It is first shown that

$$\lim s_n(z) = U(z), \tag{4.2}$$

for z in D_{k+1}, and that the convergence is uniform in D_k. This result completes the argument based on (3.6). The divergence theorem is then applied to (4.2) to yield the proof of (III).

4.1. LEMMA 4. If

$$\sigma_n = \max_{z \in C} s_n(z) , \tag{4.3}$$

then

$$\lim \sigma_n = \rho . \tag{4.4}$$

Proof. By (1.1), (1.8), (2.9), (4.3),

$$\sigma_n = \max s_n(z) \geqq \max u_{n_1}(z) = \rho_{n_1} . \tag{4.5}$$

Let z_1 be a point of C for which

$$\sigma_n = s_n(z_1) . \tag{4.6}$$

Then

$$\begin{aligned} \rho_n \geqq u_n(z_1) &= \frac{n_1}{n} s_n(z_1) + \frac{1}{n} \log |R_n(z_1)|, \\ &= \frac{n_1}{n} \sigma_n + \frac{1}{n} \log |R_n(z_1)| . \end{aligned} \tag{4.7}$$

Now z_1 is bounded from D_{k+1}, the domain containing the r_{ni}, and $|r_{ni}|$ has a bound independent of n by (1.7). Hence there are positive constants, a and b, such that

$$0 < a \leqq |z_1 - r_{ni}| \leqq b < \infty , \tag{4.8}$$

for all n and i. Combining this with (3.3) and (4.7) yields

$$\rho_n \geqq \frac{n_1}{n} \sigma_n - \frac{n_2}{n} K , \tag{4.9}$$

where $K = \max \{|\log a|, |\log b|\}$. From this and (4.5) it then follows that

$$\rho_{n_1} \leqq \sigma_n \leqq \frac{n}{n_1} \rho_n + \frac{n_2}{n_1} K . \tag{4.10}$$

The conclusion of the lemma now follows by (1.9), (3.1).

4.2. Form the function

$$w_n(z) = \sigma_n - s_n(z) - (\rho - U(z)) . \tag{4.11}$$

This can be treated like $v_n(z)$, (2.10), to show that it is positive in D.

LEMMA 5. *The functions $w_n(z)$ converge to zero in D_{k+1}, and uniformly in $\bar{D}_k$.*

Proof. Let the disc $|z - a| \leqq \gamma$ lie in D_{k+1}, and let $z_1 = a + r \exp(i\theta)$,

$r \leqq s < \gamma$. Since $w_n(z)$ is positive in D_{k+1}, and clearly harmonic there, the inequality

$$\frac{\gamma - s}{\gamma + s} w_n(a) \leqq w_n(z_1) \leqq \frac{\gamma + s}{\gamma - s} w_n(a) \tag{4.12}$$

holds. This shows that the convergence of $w_n(a)$ to zero implies the uniform convergence to zero in the circle $|z - a| = s$, and that if $w_n(a)$ does not converge to zero, the same will be true at each point of the circle. A similar relationship holds between the convergence of $w_n(\infty)$ and the convergence of $w_n(z)$ for $|z| \geqq s$, a domain lying in D_{k+1}. Thus the set of points of D_{k+1} where $\lim w_n(z) = 0$ is an open set, and the set where $\lim w_n(z) \neq 0$ is also an open set. Since D_{k+1} is open and connected, it cannot be expressed as the sum of two disjoint open sets, so that one of these sets must be a null set. Since $w_n(\infty) = \sigma_n - \rho$, a quantity tending to zero by Lemma 4, the non-null set is the one for which $\lim w_n(z) = 0$. By the Heine-Borel theorem, $\bar{D}_k$ can be covered by a finite number of circles lying in D_{k+1}, one of which will be of the form $|z| \geqq s$. The convergence will be uniform in each circle, and hence uniform in $\bar{D}_k$.

4.3. For application to (3.6), note that

$$\left| \frac{1}{n_1} \log S_n(z) - U(z) \right| \leqq |w_n(z)| + |\sigma_n - \rho| \,. \tag{4.13}$$

Thus by Lemmas 4 and 5, the left side converges uniformly to zero in $\bar{D}_k$, and hence in E.

4.4. *Proof of* (III). There is no loss in generality in assuming that Γ lies in D_k. If $z = a + r \exp(i\theta)$, $r \leqq s < \gamma$, then

$$|(w_n(z))_x| \leqq \frac{w_n(a)}{(\gamma - s)^2} \,; \tag{4.14}$$

where $(\ )_x$ denotes the partial derivative with respect to x. It is assumed that a is on Γ, and that $|z - a| \leqq s$ lies in D_{k+1}. The same inequality holds for the partial derivative with respect to y. The convergence of $w_n(a)$ to zero thus yields the uniform convergence to zero of the partial derivatives in the specified circles. An application of the Heine-Borel theorem then shows that the convergence is uniform on Γ. Thus

$$\lim \frac{1}{2\pi} \int_\Gamma (w_n(z))_x dy - (w_n(z))_y dx = 0 \,. \tag{4.15}$$

Using (4.11), it is seen that this is equivalent to

$$\lim \frac{1}{2\pi}\int_{\Gamma}(s_n(z))_x dy - (s_n(z))_y dx \tag{4.16}$$
$$= \frac{1}{2\pi}\int (U(z))_x dy - (U(z))_y dx \ .$$

Let $\lambda_n(S)$ be the total multiplicity of the zeros of $S_n(z)$ in the set S. Now both $U(z)$ and $s_n(z)$ are harmonic on Γ, and Γ is of sufficient smoothness for the application of the divergence theorem, so that the result

$$\lim \frac{\lambda_n(I(\Gamma))}{n} = \mu(I(\Gamma)) \tag{4.17}$$

is obtained. For any set S it follows from (3.2) that

$$\nu_n(S) - \lambda_n(S) \leqq \nu_n(D_{k+1}) = n_2 \ . \tag{4.18}$$

Thus, by (3.1) and (4.7) applied to

$$\frac{\lambda_n(I\Gamma))}{n} \leqq \frac{\nu_n(I(\Gamma))}{n} \leqq \frac{\lambda_n(I(\Gamma))}{n} + \frac{n_2}{n} \ , \tag{4.19}$$

the proof of (III) is completed.

5. Relationship to a paper by Walsh and Evans. The results (I) and (III) we obtained by other methods in [7], and another form of discussing the asymptotic behavior of $T_n(z)$ for z in the complement of C was used. The result (IV) is not found in [7], and we will discuss in more detail, and in a slightly more general context the significance of this and the other results.

Domain Polynomials. Besides the $T_n(z)$, there are other sets of polynomials which are associated with general sets C in the plane. We mention only the Carleman polynomials [3], $C_n(z)$, which require that C have connected complement, and Faber polynomials [5], $F_n(z)$, which require that the complement of C be simply connected. These are adequate to illustrate our remarks.

The Location Problem is an apt name to give to results relating to the location of zeros of domain polynomials, and known results suggest the further distinction of interior location and exterior location, corresponding to whether we refer to zeros on C or in the complement of C.

Results on Exterior Location. For sets with simply connected complements, and bounded by a simple analytic curve Γ, it has been shown by Johnston [3] and the author [5] that ultimately the zeros of $C_n(z)$ and $F_n(z)$, respectively, lie inside any simple interior level curve of Γ. It is not known whether this is true for $T_n(z)$, although (IV) shows that the zeros lie ultimately inside any exterior level curve.

A basic observation of this paper and [7] is that when C has a multiply connected complement, then zeros of $T_n(z)$ can lie in the complement of C and be uniformly bounded from C for arbitrarily large n. In the sense defined by (I) the number must be small in comparison with n, although they can exceed any finite bound. The refinement of (IV) states that if C is bounded by a finite number of analytic curves, then there is an absolute constant for any exterior level curve of C, which ultimately cannot be exceeded by the number of zeros of $T_n(z)$ exterior to this level curve. What has not been shown is whether a constant exists for the complement of C itself. Examples indicate that if there is such a constant, it cannot be less than $k - 1$, where k is the number of boundary components of C.

Interior Location. Formula (III) states that the proportion of zeros on any component of C, for $T_n(z)$, approaches the harmonic measure of the component. Where on the component the zeros accumulate is not known. The existant examples, namely $T_n(z)$ for the circle and ellipse, indicate that the limit points of the zeros, which can be called the center, have an interior location in the set. No precise characterization of the center for $T_n(z)$ has been found. In [6] a study is made of the center for $F_n(z)$. The indications are that the center will not be the same set for the different classes of domain polynomials.

BIBLIOGRAPHY

1. L. Fejer, *Über die Lage der Nullstellen von Polynomen, die aus Minimumforderungen gewisser Art entspringen*, Mathematische Annalen, **85**, (1922), 41-48.

2. Rolf. Nevanlinna, *Eindeutige Analytische Functionen*, Berlin, 1936.

3. P. C. Rosenbloom and S. E. Warschawski, *Approximation by polynomials, Lectures on functions of a complex variable*, Edited by Wilfred Kaplan, Ann Arbor, 1955, pp. 287-301.

4. G. Szegö, *Bemerkungen zu einer Arbeit von Herrn Fekete: Über die Verteilung der Wurzeln bei gewissen algebraischen Gleichungen mit ganzzahligen Koeffizienten*, Mathematische Zeitschrift, **21** (1924), 203-208.

5. J. L. Ullman, *Studies on faber polynomials*, Dissertation, Stanford University, 1949.

6. ———, *On faber polynomials*, submitted to the Transactions of the Amer. Math. Soc.

7. J. L. Walsh and J. P. Evans, *Note on the distribution of zeros of extremal polynomials*, Proc. Nat. Acad. Sci., May 1954, pp. 332-337.

THE UNIVERSITY OF MICHIGAN

ORDERINGS OF THE SUCCESSIVE OVERRELAXATION SCHEME

RICHARD S. VARGA

1. Introduction. One of the more frequently used iterative methods [11, 14, 18] in numerically solving self-adjoint partial difference equations of elliptic type:

$$\sum_{j=1}^{n} a_{i,j}x_j = k_i, \qquad a_{i,i} \neq 0, \qquad 1 \leqq i \leqq n \,, \tag{1}$$

is the Young-Frankel *successive overrelaxation scheme* [16, 4]. If superscripts denote the iteration indices, then the successive overrelaxation scheme is defined by

$$x_i^{(n+1)} = \omega\left\{\sum_{j=1}^{i-1} b_{i,j}x_j^{(n+1)} + \sum_{j=i+1}^{n} b_{i,j}x_j^{(n)} + g_i\right\} + (1-\omega)x_i^{(n)} \,, \tag{2}$$

where

$$b_{i,j} = \begin{cases} -a_{i,j}/a_{i,i}, & i \neq j \\ 0\,, & i = j \end{cases}; \; g_i = k_i/a_{i,i}, \qquad 1 \leqq i, j \leqq n \,. \tag{2$'$}$$

The parameter ω is the *relaxation factor.*

Since the introduction of this method, there has remained the question of the effect of different orderings of the equations of (1) on the rate of convergence of the overrelaxation scheme. Young [16] introduced the concept of a *consistent ordering* of the unknowns for a class of matrices satisfying his definition of *property* (A), and he conjectured [17] that, with certain additional assumptions, these consistent orderings were optimal[1] in the sense that, among all orderings, the consistent orderings give the fastest convergent iterative scheme for the case of $\omega = 1$ of (2).

The problem of the relationship between orderings and rates of convergence has been recently investigated by Heller [6], whose approach was combinatorial. Assuming the $n \times n$ matrix $A \equiv ||a_{i,j}||$ of (1) to be multi-diagonal, Heller concentrated on the problem of finding all orderings whose associated Gauss-Seidel iterative method, the special case of (2) with $\omega = 1$, had the same eigenvalues as the eigenvalues of the Gauss-Seidel method based on the "usual ordering."

Our approach to the question of orderings is based on the Perron-

Received October 3, 1958. Presented to the American Mathematical Society April 18, 1958.

[1] For some preliminary results on this conjecture for optimum orderings, see [**17**].

Frobenius theory of non-negative matrices.[2] Our main result (Theorem 4) contains as a special case a proof of Young's conjecture. On the other hand, while certain orderings may produce faster convregent iterative schemes than others, we prove (Theorem 5) that, for the case $\omega = 1$ of (2), different orderings have vanishingly small effect on the rate of convergence of the Gauss-Seidel iteration method for slowly convergent problems. This last result proves a conjecture by Shortley and Weller [10, p. 338] who observed this phenomenon in the numerical solution of the Dirichlet problem.

2. **Preliminary definitions.** We first define the class S of matrices. We shall later show in § 5 that the results, based on this class of matrices, hold for a large number of matrix problems (1) arising from the numerical solution of certain partial differential equations of elliptic type. We let B denote the square matrix of coefficients $b_{i,j}$ defined in (2′).

DEFINITION 1. The matrix $B \in S$ if and only if B satisfies the following conditions:

(i) $B = \|b_{i,j}\|$ is a non-negative $n \times n$ matrix, with zero diagonal entries, i.e., $b_{i,j} \geqq 0$ for $i \neq j$, and $b_{i,i} = 0$ for all $1 \leqq i, j \leqq n$.

(ii) B is *irreducible* [5, p. 458], i.e., there exists no permutation matrix Λ such that

$$\Lambda B \Lambda^{-1} = \begin{pmatrix} B_1 & B_2 \\ 0 & B_3 \end{pmatrix},$$

where B_1 and B_3 are square submatrices.

(iii) B is symmetric.

For any permutation, or ordering, ϕ of the integers $1 \leqq i \leqq n$, let Λ_ϕ denote the corresponding $n \times n$ permutation matrix and let $B_\phi \equiv \Lambda_\phi B \Lambda'_\phi = \Lambda_\phi B \Lambda_\phi^{-1}$, where in general A' denotes the transpose of the matrix A. For $B \in S$, B_ϕ is symmetric with zero diagonal entries, so that we can decompose B_ϕ into:

$$B_\phi = L_\phi + L'_\phi, \tag{3}$$

where L_ϕ is a strictly lower triangular matrix.[3] We define

$$M_\phi(\sigma) = \sigma L_\phi + L'_\phi, \qquad \sigma > 0. \tag{4}$$

It is clear that $M_\phi(\sigma)$ is a non-negative irreducible matrix for every $\sigma > 0$ and ϕ. Thus, by the Perron-Frobenius theory [8, 5] of non-negative matrices, $M_\phi(\sigma)$ possesses a positive simple eigenvalue, $m_\phi(\sigma)$, which

[2] A similar approach was employed Kahan [7′] in generalizing the results of Young [**16**]. Although Kahan was not directly concerned with the question of orderings, many of his results, stated without proof in [**7**], are nevertheless similar.

[3] An $n \times n$ matrix $L = \|l_{i,j}\|$ is strictly lower triangular if and only if $l_{i,j} = 0$ for $i \leqq j$, $1 \leqq i, j \leqq n$.

is greater than or equal in modulus to all other eigenvalues of $M_\phi(\sigma)$, and to $m_\phi(\sigma)$ can be associated an eigenvector with positive components. It can be shown, based on further results of the Perron-Frobenius theory, that $m_\phi(\sigma)$ has the following properties:

$$(5) \qquad \begin{cases} \text{(i) } m_\phi(\sigma) \text{ is a strictly increasing function of } \sigma \text{ [3, p. 598].} \\ \text{(ii) } m_\phi(\sigma) \text{ is an analytic function of } \sigma, \text{ for all } \sigma > 0.\text{[4]} \end{cases}$$

Before proceeding, we briefly state some of the terminology and conclusions of the Perron-Frobenius theory, which we shall frequently use. If C is an arbitrary non-negative irreducible $n \times n$ matrix, we say, following Frobenius [5], that C is *primitive* if the positive eigenvalue r given by the Perron-Frobenius theory is strictly greater in modulus than all other eigenvalues of C. If there are $k(>1)$ eigenvalues of C with modulus r, then C is said [9] to be *cyclic of index* k. In particular, if C is cyclic of index $k(>1)$, then [9] there exists a permutation matrix Λ such that

$$(6) \qquad \Lambda C \Lambda^{-1} = \begin{pmatrix} 0 & 0 & \cdots & 0 & C_1 \\ C_2 & 0 & \cdots & 0 & 0 \\ 0 & C_3 & \cdots & 0 & 0 \\ \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & & \cdot & \cdot \\ 0 & 0 & \cdots & C_k & 0 \end{pmatrix},$$

where the diagonal blocks of $\Lambda C \Lambda^{-1}$ are square submatrices with zero entries. For any matrix C, we shall let $\bar{\mu}[C]$ denote the *spectral radius* of C, i.e., $\bar{\mu}[C] = \max_j |\lambda_j|$, where λ_j is an eigenvalue of C.

3. Spectral radius as a function of ordering.

LEMMA 1. *If* $B \in S$, *then* $m_\phi(\sigma) = \bar{\mu}[B]\sigma^{1/2}h_\phi(ln\sigma)$, *where* $h_\phi(\alpha) = h_\phi(-\alpha)$ *for all real* α, *and* $h_\phi(0) = 1$.

Proof. For $\sigma > 0$, there exists an eigenvector $\boldsymbol{x}$ with positive components such that $M_\phi(\sigma)\boldsymbol{x} = m_\phi(\sigma)\boldsymbol{x}$. From definition,

$$M_\phi(\sigma) = \sigma L_\phi + L'_\phi = \sigma\Big(L_\phi + \frac{1}{\sigma}L'_\phi\Big) = \sigma M'_\phi\Big(\frac{1}{\sigma}\Big).$$

Thus, $M'_\phi\Big(\dfrac{1}{\sigma}\Big)\boldsymbol{x} = \dfrac{m_\phi(\sigma)}{\sigma}\boldsymbol{x}$. Since M_ϕ and M'_ϕ have the same eigenvalues, then

$$(7) \qquad \sigma m_\phi\Big(\frac{1}{\sigma}\Big) = m_\phi(\sigma), \ \sigma > 0 .$$

[4] Since $m_\phi(\sigma)$ is simple root of det $[M_\phi(\sigma) - \lambda I] = 0$, the analyticity of $m_\phi(\sigma)$ can be proved by means of the implicit function theorem.

If

$$h_\phi(ln\sigma) \equiv \frac{m_\phi(\sigma)}{\bar{\mu}[B]}\sigma^{-1/2},\ \sigma > 0\ ,$$

then equation (7) shows that $h_\phi(\alpha)$, $\alpha = ln\sigma$, is an even function of α. For $\sigma = 1$, $m_\phi(1) = \bar{\mu}[B]$ by definition, and thus $h_\phi(0) = 1$, which completes the proof.

From (5) and Lemma 1, it follows that $h_\phi(\alpha)$ is an analytic function of α for all real values of α.

LEMMA 2. *Let $A(\alpha) = e^\alpha L + D + e^{-\alpha}L'$, where L is a non-negative strictly lower triangular matrix, and D is any non-negative diagonal matrix. If $L + L'$ is irreducible, and $0 \leqq \alpha_1 \leqq \alpha_2$, then $\bar{\mu}[A(\alpha_1)] \leqq \bar{\mu}[A(\alpha_2)]$.*

Proof. If $C \equiv L + D + L' = ||c_{i,j}||$, then by assumptions stated in the lemma, C is non-negative and irreducible. Assume now that C is primitive, and consider any non-zero cycle ν of C of length $m \geqq 1$:

$$\nu = c_{i_0,i_1}c_{i_1,i_2}\cdots c_{i_{m-1},i_m=i_0},\ \text{where } c_{i_j,i_{j+1}} > 0,\ j = 0, \cdots, m-1\ .$$

It is clear that the corresponding cycle for $A(\alpha)$ is $t = e^{q\alpha}\nu$, where q is an integer. From the symmetry of C, there is another cycle t' of $A(\alpha)$ of the form: $t' = e^{-q\alpha}\nu$. Since t and t' are contained in the i_0-th diagonal entry of $A^m(\alpha)$, it follows that the trace of $A^m(\alpha)$ is composed of terms of the form: $2\nu\ \cosh(q\alpha)$. Using the monotonicity of $\cosh(x)$, we obtain, for $0 \leqq \alpha_1 \leqq \alpha_2$,

$$(8)\qquad tr[A^m(\alpha_1)] \leqq tr[A^m(\alpha_2)]\ ,$$

for all $m \geqq 1$. By assumption, C is primitive, which implies that $A(\alpha)$ is primitive for all real α. Since the trace of a matrix is equal to the sum of its eigenvalues, then

$$(9)\qquad tr[A^m(\alpha)] \sim (\bar{\mu}[A(\alpha)])^m,\ m \to \infty\ .$$

Combining the results of (8) and (9), and taking mth roots, we obtain the desired result, under the additional assumption that C is primitive. But if C is not primitive, then $\tilde{C} = C + \beta I$, $\beta > 0$, certainly is, and since

$$\bar{\mu}[\tilde{A}(\alpha) \equiv e^\alpha L + D + \beta I + e^{-\alpha}L'] = \bar{\mu}[A(\alpha)] + \beta\ ,$$

the desired result again follows.

THEOREM 1. *If $B \in S$, then $h_\phi(\alpha)$ is non-decreasing for $\alpha \geqq 0$. Moreover, for any $\alpha \neq 0$,*

$$(10)\qquad 1 \leqq h_\phi(\alpha) < \cosh(\alpha/2)\ .$$

Proof. For $\sigma > 0$, consider the matrix

$$P_\phi(\sigma) \equiv \frac{M_\phi(\sigma)}{\bar{\mu}[B]\sigma^{1/2}} = \frac{1}{\bar{\mu}[B]}\{\sigma^{1/2}L_\phi + \sigma^{-1/2}L'_\phi\} \; . \tag{11}$$

By definition, $\bar{\mu}[P_\phi(\sigma)] = h_\phi(ln\sigma)$. For any $\alpha_2 \geqq \alpha_1 \geqq 0$, $h_\phi(\alpha_2) \geqq h_\phi(\alpha_1)$ if and only if $\bar{\mu}[P_\phi(e^{\alpha_2})] \geqq \bar{\mu}[P_\phi(e^{\alpha_1})]$, and thus the first conclusion follows from Lemma 2, with D the null matrix.

To prove the second part of the theorem, we write $P_\phi(\sigma)$ in the form

$$P_\phi(e^\alpha) = \cosh(\alpha/2) \cdot T_\phi + \sinh(\alpha/2) \cdot K_\phi \; , \tag{12}$$

where

$$T_\phi \equiv \frac{1}{\bar{\mu}[B]}(L_\phi + L'_\phi); \quad K_\phi \equiv \frac{1}{\bar{\mu}[B]}(L_\phi - L'_\phi) \; . \tag{12'}$$

For any real α, $P_\phi(e^\alpha)$ is a non-negative, irreducible matrix. If $\boldsymbol{x}$ is the eigenvector of $P_\phi(e^\alpha)$ with positive components corresponding to the eigenvalue $h_\phi(\alpha)$, so normalized[5] that $(\boldsymbol{x}, \boldsymbol{x}) = 1$, then

$$(P_\phi(e^\alpha)\boldsymbol{x}, \boldsymbol{x}) = h_\phi(\alpha) = \cosh(\alpha/2) \cdot (T_\phi\boldsymbol{x}, \boldsymbol{x}) + \sinh(\alpha/2) \cdot (K_\phi\boldsymbol{x}, \boldsymbol{x}) \; .$$

Since K_ϕ is skew-symmetric, then $h_\phi(\alpha) = \cosh(\alpha/2) \cdot (T_\phi\boldsymbol{x}, \boldsymbol{x})$. But, T_ϕ is symmetric, non-negative, and irreducible, so that $(T_\phi\boldsymbol{x}, \boldsymbol{x}) \leqq \bar{\mu}[T_\phi] = 1$. Thus, from the first part of this theorem and Lemma 1, we have that $1 \leqq h_\phi(\alpha) \leqq \cosh(\alpha/2)$ for all real α. Assuming $\alpha \neq 0$, suppose that $(T_\phi\boldsymbol{x}, \boldsymbol{x}) = \bar{\mu}[T_\phi] = 1$. This is true only if $\boldsymbol{x}$ is also an eigenvector of T_ϕ, and thus, from (12), $\boldsymbol{x}$ is an eigenvector of K_ϕ. But since K_ϕ is a skew-symmetric matrix, the eigenvalues of K_ϕ are pure imaginary numbers. By the irreducibility of B, there exists at least one positive entry in the first row of L'_ϕ, and thus the first component of $K_\phi\boldsymbol{x}$ is a negative real number, which contradicts the fact that $\boldsymbol{x}$ is an eigenvector of K_ϕ. Thus, for $\alpha \neq 0$, $(T_\phi\boldsymbol{x}, \boldsymbol{x}) < 1$, and we have the inequality of (10), which completes the proof.

Since $h_\phi(\alpha)$ is analytic for all real α, we conclude the

COROLLARY. *If $B \in S$, then either $h_\phi(\alpha) \equiv 1$ for all real α, or $h_\phi(\alpha)$ is strictly increasing for $\alpha \geqq 0$.*

DEFINITION 2. If $B \in S$, then ϕ is an *h-consistent ordering* for B if and only if $h_\phi(\alpha) \equiv 1$ for all real α. Otherwise, ϕ is a *non-consistent ordering* for B.

We remark that the above definition of an h-consistent ordering generalizes for the class S the definitions of a consistent ordering given

[5] Here, $(\boldsymbol{x}, \boldsymbol{y})$ denotes, as usual, the scalar product of the vectors $\boldsymbol{x}$ and $\boldsymbol{y}$. If the components of $\boldsymbol{x}$ and $\boldsymbol{y}$ are x_i, y_i, respectively, then $(\boldsymbol{x}, \boldsymbol{y}) \equiv \sum_{i=1}^{n} x_i y_i$.

both by Young [16] and Arms, Gates, and Zondek [1]. To show this, assume that $B \in S$ satisfies Young's property (A), and that ψ is a consistent ordering for B in the sense of Young. Then, as shown by Young [16, p. 97], both $M_\psi(\sigma)$ and $\sigma^{1/2}B$ have the same characteristic polynomials, and hence the same eigenvalues. Thus, $m_\psi(\sigma) = \sigma^{1/2}\bar{\mu}[B]$, from which it follows that $h_\psi(\alpha) \equiv 1$, proving that ψ is also an h-consistent ordering in the sense of Definition 2. That consistent orderings in the sense of Arms, Gates, and Zondek for matrices $B \in S$ also satisfy Definition 2 can be proved in a similar manner.

THEOREM 2. *If $B \in S$, then there exists an h-consistent ordering ϕ for B if and only if B is cyclic of index 2.*

Proof. If B is cyclic of index 2, then by (6) there exists an ordering ψ and a permutation matrix Λ_ψ such that

$$\Lambda_\psi B \Lambda_\psi^{-1} \equiv B_\psi = \begin{pmatrix} 0 & B_1 \\ B_2 & 0 \end{pmatrix}, \tag{13}$$

where the diagonal blocks are square submatrices. Thus,

$$M_\psi(\sigma) = \begin{pmatrix} 0 & B_1 \\ \sigma B_2 & 0 \end{pmatrix},$$

and

$$M_\psi^2(\sigma) = \begin{pmatrix} \sigma B_1 B_2 & 0 \\ 0 & \sigma B_2 B_1 \end{pmatrix},$$

and thus $M_\psi^2(\sigma) = \sigma M_\psi^2(1)$. It follows then that $m_\psi(\sigma) = \bar{\mu}[B]\sigma^{1/2}$, and $h_\psi(\alpha) \equiv 1$, proving that ψ is an h-consistent ordering.

Since $B \in S$ implies that B is non-negative and irreducible, then B is either primitive or cyclic of index $k, k > 1$. Since B is moreover symmetric, it follows from (6) that B is either primitive or cyclic of index 2. We shall now that if B is primitive, *no* ordering of B is an h-consistent ordering. With B primitive, let ϕ be any ordering, and consider

$$A_\phi(\alpha) \equiv \frac{1}{\bar{\mu}[B]}\{e^{\alpha}L_\phi + e^{-\alpha}L'_\phi\}, \alpha \geqq 0 . \tag{14}$$

Following the notation of Lemma 2, suppose that every cycle of $A_\phi(\alpha)$ of length m has $q = 0$, for all $m \geqq 1$. This implies that every non-zero cycle of $A_\phi(\alpha)$ contains precisely the same number of terms from above the diagonal as from below the diagonal of $A_\phi(\alpha)$. Since $A_\phi(\alpha)$ has zero diagonal entries, then every non-zero cycle of $A_\phi(\alpha)$ has an even number of terms. Thus, the greatest common divisor γ of the lengths of these non-zero cycles is evidently 2. It is known [9] that $\gamma = 2$ if and only if $A_\phi(\alpha)$ is cyclic of index 2, and, for any real α, $A_\phi(\alpha)$ is cyclic of index

2 if and only if B is cyclic of index 2. This being a contradiction to the assumption that B is primitive, there than exists a positive integer m_0, and a positive integer q_0 such that the $tr[A_\phi^{m_0}(\alpha)]$ contains a term $\nu\cosh(q_0\alpha)$, $\nu > 0$, while $tr[A_\phi^{m_0}(0)]$ contains the corresponding term ν. As in the proof of Lemma 2, it follows that, for $\alpha \geqq 0$,

$$tr[{}_\phi^{m_0}(\alpha)] \geqq tr[A_\phi^{m_0}(0)] + \nu[\cosh(q_0\alpha) - 1]\ . \tag{15}$$

Since this particular cycle of length m_0 can be repeated cyclically, then

$$tr[A_\phi^{lm_0}(\alpha)] \geqq tr[A_\phi^{lm_0}(0)] + \nu^l[\cosh(q_0 l\alpha) - 1]\ . \tag{15'}$$

Since B is primitive, so is $A_\phi(\alpha)$ for all real α, and from (9) and the definition of $h_\phi(\alpha)$, we have

$$h_\phi(2\alpha) = \bar{\mu}[A_\phi(\alpha)] \sim (tr[A_\phi^m(\alpha)])^{1/m},\ m \to \infty\ . \tag{16}$$

For α sufficiently large so that $\nu e^{q_0\alpha} > 1$, we obtain from (15') and (16)

$$h_\phi(2\alpha) \geqq (\nu e^{q_0\alpha})^{1/m_0} > 1\ . \tag{17}$$

Thus, if B is primitive, no ordering ϕ of B is an h-consistent ordering, which completes the proof.

We finally remark that it has already been pointed out [2] that, in general, Young's property (A), on which Young's definition of consistent ordering depends, for the matrix of coefficients of (1) implies that the matrix B of (2) is *cyclic of index* 2. The same is true of its generalization [1] to property (A^π). This relationship to cyclic matrices has led to a further generalization [15] of the Young-Frankel overrelaxation scheme to matrices B of (2) which are cyclic of index p, $p \geqq 2$.

Returning to the successive overrelaxation scheme of (2), if $\boldsymbol{x}^{(n)}$ denotes the vector with components $x_i^{(n)}$, then for B symmetric, we can write (2) equivalently as

$$\boldsymbol{x}^{(n+1)} = \mathscr{L}_{\phi,\omega}\boldsymbol{x}^{(n)} + \boldsymbol{f} \tag{18}$$

where

$$\mathscr{L}_{\phi,\omega} \equiv (I - \omega L_\phi)^{-1}\{\omega L'_\phi + (1 - \omega)I\}\ , \tag{19}$$

and

$$\boldsymbol{f} = \omega(I - \omega L_\phi)^{-1}\boldsymbol{g}\ . \tag{19'}$$

Accordingly, we make the

DEFINITION 3. $\mathscr{L}_{\phi,\omega} \equiv (I - \omega L_\phi)^{-1}\{\omega L'_\phi + (1 - \omega)I\}$ is the *successive overrelaxation matrix*, corresponding to the matrix B and ordering ϕ. The quantity ω is the *relaxation factor*.

LEMMA 3. *Let $B \in S$. If, for $\omega > 0$, there exists a positive real τ for which*

$$m_\phi(\tau) = \left(\frac{\tau + \omega - 1}{\omega}\right),$$

then τ is an eigenvalue of $\mathscr{L}_{\phi,\omega}$. Moreover, if $0 < \omega \leqq 1$, $\bar{\mu}[\mathscr{L}_{\phi,\omega}]$ is the unique positive value of τ for which

$$m_\phi(\tau) = \left(\frac{\tau + \omega - 1}{\omega}\right).$$

Proof. It is known[6] that for $\omega > 0$, $\mathscr{L}_{\phi,\omega}\boldsymbol{v} = \lambda\boldsymbol{v}$ if and only if

$$(\lambda L_\phi + L'_\phi)\boldsymbol{v} = \left(\frac{\lambda + \omega - 1}{\omega}\right)\boldsymbol{v}, \tag{20}$$

from which the first part of the lemma follows. Since L_ϕ is a strictly lower triangular matrix, then $(I - \omega L_\phi)^{-1} = I + \omega L_\phi + \cdots + \omega^{n-1}L_\phi^{n-1}$. Clearly, lfor $0 < \omega < 1$, $\mathscr{L}_{\phi,\omega}$ is a non-negative irreducible matrix.[7] Thus, the argest in modulus eigenvalue of $\mathscr{L}_{\phi,\omega}$, $\bar{\mu}[\mathscr{L}_{\phi,\omega}]$, is positive, and its corresponding eigenvector $\boldsymbol{v}$ can be chosen to have positive components. From $\mathscr{L}_{\phi,\omega}\boldsymbol{v} = \bar{\mu}[\mathscr{L}_{\phi,\omega}]\boldsymbol{v}$, we have, by (20), that $m_\Phi(\sigma)$ and $\left(\frac{\sigma + \omega - 1}{\omega}\right)$ intersect in $\bar{\mu}[\mathscr{L}_{\phi,\omega}]$. By continuity, the result is true also for $\omega = 1$, which completes the proof.

We remark that $\frac{1}{\omega}\{\sigma + \omega - 1\}$, graphed against σ, defines a family of straight lines through the point (1, 1). Figure 1 illustrates the second part of Lemma 3.

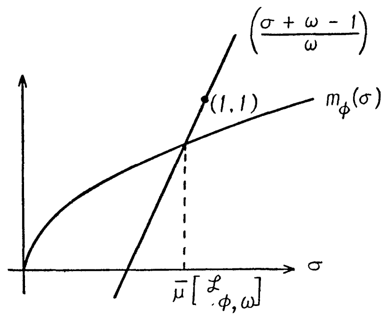


Figure 1

DEFINITION 4. If $B \in S$, and $0 < \omega < 1$, then $\xi(\bar{\mu}[B], \omega)$ is the unique positive value of σ for which $\bar{\mu}[B]\sigma^{1/2} = \left(\frac{\sigma + \omega - 1}{\omega}\right)$.

For the class of matrices S, the following theorem sharpens results due to Stein and Rosenberg [12], and Kahan [7,7′].

---

[6] See, for instance, [**16**, p. 99].

[7] It is, moreover, primitive.

THEOREM 3. *Let $B \in S$, and assume $0 < \omega \leqq 1$. If $\bar{\mu}[B] < 1$, then for ϕ a non-consistent ordering for B,*

$$\xi(\bar{\mu}[B], \omega) < \bar{\mu}[\mathscr{L}_{\phi,\omega}] < \left(\frac{2(1-\omega) + \omega\bar{\mu}[B]}{2 - \omega\bar{\mu}[B]}\right),$$

and for ϕ an h-consistent ordering for B, $\xi(\bar{\mu}[B], \omega) = \bar{\mu}[\mathscr{L}_{\phi,\omega}]$. If $\bar{\mu}[B] = 1$, then $[\mathscr{L}_{\phi,\omega}] = 1$. If $\bar{\mu}[B] > 1$, then for ϕ an non-consistent ordering of B, $\xi(\bar{\mu}[B], \omega) < \bar{\mu}[\mathscr{L}_{\phi,\omega}]$, and for ϕ an h-consistent ordering for B, $\xi(\bar{\mu}[B], \omega) = \bar{\mu}[\mathscr{L}_{\phi,\omega}]$.

Proof. We consider only the case when $\bar{\mu}[B] < 1$, since the other cases follow similarly. If ϕ is an h-consistent ordering for B, then $m_\phi(\sigma) = \bar{\mu}[B]\sigma^{1/2}$. From Definition 4 and Lemma 3, it follows that $\xi(\bar{\mu}[B], \omega) = \bar{\mu}[\mathscr{L}_{\phi,\omega}]$. If ϕ is a non-consistent ordering for B, then, from Theorem 1 and its corollary, $h_\phi(\alpha)$ is strictly increasing for $\alpha \geqq 0$, and $1 < h_\phi(\alpha) < \cosh(\alpha/2)$ for $\alpha \neq 0$, these inequalities giving directly

$$\bar{\mu}[B]\sigma^{1/2} < m_\phi(\sigma) < \bar{\mu}[B]\sigma^{1/2}\cosh\left(\frac{ln\sigma}{2}\right) = \bar{\mu}[B] \cdot \left(\frac{\sigma+1}{2}\right), \sigma \neq 1\,. \tag{21}$$

Consider the function $k_\phi(\sigma)$ defined by

$$k_\phi(\sigma) \equiv m_\phi(\sigma) - \left(\frac{\sigma + \omega - 1}{\omega}\right), \omega > 0\,. \tag{22}$$

For $\xi \equiv \xi(\bar{\mu}[B], \omega)$, it follows from Definition 4 and the first inequality of (21) that $k_\phi(\xi) > 0$. On the other hand, $k_\phi(1) < 0$ since $k_\phi(1) = \bar{\mu}[B]-1$. Thus, since $k_\phi(\sigma)$ is continuous in σ for all $\sigma \geqq 0$, there exists a τ with $\xi < \tau < 1$ for which $k_\phi(\tau) = 0$. By Lemma 3, $\bar{\mu}[\mathscr{L}_{\phi,\omega}] = \tau$, so that $\xi(\bar{\mu}[B], \omega) < \bar{\mu}[\mathscr{L}_{\phi,\omega}]$. Using the second inequality of (21), we have that

$$0 = k_\phi(\tau) = m_\phi(\tau) - \left(\frac{\tau + \omega - 1}{\omega}\right) < \bar{\mu}[B]\left(\frac{\tau+1}{2}\right) - \left(\frac{\tau + \omega - 1}{\omega}\right),$$

from which it follows that

$$\tau = \bar{\mu}[\mathscr{L}_{\phi,\omega}] < \left(\frac{2(1-\omega) + \omega\bar{\mu}[B]}{2 - \omega\bar{\mu}[B]}\right),$$

which completes the proof.

The special case $\omega = 1$ gives rise to inequalities like that of Stein and Rosenberg [12]. Since $\xi(\bar{\mu}[B], \omega = 1) = \bar{\mu}^2[B]$, we have the

COROLLARY.[8] *For the Gauss-Seidel method, $\omega = 1$ of (2), if $\bar{\mu}[B] < 1$, then*

[8] If $B \in S$ and $\bar{\mu}[B] < 1$, Young conjectured [**17**] that for ψ a consistent ordering of $B, \bar{\mu}[\mathscr{L}_{\psi,1}] \leqq \bar{\mu}[\mathscr{L}_{\phi,1}]$ for all orderings ϕ of B. Applying the first part of this corollary, we have a proof of this conjecture.

$$\bar{\mu}^2[B] \leqq \bar{\mu}[\mathscr{L}_{\phi,1}] < \left(\frac{\bar{\mu}[B]}{2 - \bar{\mu}[B]}\right),$$

equality holding if and only if ϕ is an h-consistent ordering for B. If $\bar{\mu}[B] = 1$, then $\bar{\mu}[\mathscr{L}_{\phi,1}] = 1$. If $\bar{\mu}[B] > 1$, then $\bar{\mu}^2[B] \leqq \bar{\mu}[\mathscr{L}_{\phi,1}]$, equality holding if and only if ϕ is an h-consistent ordering for B.

We now consider the subclass of matrices $B \in S$ for which $\bar{\mu}[B] < 1$. Following Young [16], we define the quantity:

$$(23) \qquad \omega_b = \frac{2}{1 + \sqrt{1 - \bar{\mu}^2[B]}} = 1 + \left[\frac{\bar{\mu}[B]}{1 + \sqrt{1 - \bar{\mu}^2[B]}}\right]^2,$$

so that[9] $1 < \omega_b < 2$. In Figure 1, it can be shown that ω_b is the unique value of the parameter ω, $0 \leqq \omega \leqq 2$, for which the straight line $\left(\frac{\sigma + \omega - 1}{\omega}\right)$ through the point $(1, 1)$ is tangent to the curve $\bar{\mu}[B]\sigma^{1/2}$. Thus, for $0 \leqq \omega \leqq \omega_b$, the quantity $\xi(\bar{\mu}[B], \omega)$ can be defined as the largest positive value of σ for which

$$\left(\frac{\sigma + \omega - 1}{\omega}\right) = \bar{\mu}[B]\sigma^{1/2}.$$

It is known [16] that if the matrix $B \in S$ satisfies Young's property (A), with $\bar{\mu}[B] < 1$ and ϕ a consistent ordering (in the sense of Young) for B, then ω_b is the overrelaxation factor which minimizes $\bar{\mu}[\mathscr{L}_{\phi,\omega}]$, and thus gives the fastest convergence in (2). A similar conclusion is obtained for the generalization of [1]. Thus, for certain matrices, ω_b is the optimum overrelaxation factor.

THEOREM 4.[10] *Let $B \in S$, and assume $\bar{\mu}[B] < 1$. Then $\xi(\bar{\mu}B, \omega) \leqq \bar{\mu}[\mathscr{L}_{\phi,\omega}]$ for $0 < \omega \leqq \omega_b$, with equality if and only if ϕ is an h-consistent ordering for B. For $\omega_b \leqq \omega < 2$, $\bar{\mu}[\mathscr{L}_{\phi,\omega}] \geqq \omega - 1$, with equality for all ω in this range if and only if ϕ is an h-consistent ordering for B.*

Proof. By Theorem 3, we need only consider the case $\omega \geqq 1$. If ϕ is a non-consistent ordering for B, then $h_\phi(\alpha) > 1$ for all real $\alpha \neq 0$. from this, it follows, as in the proof of Theorem 3, that the straight line $\left(\frac{\sigma + \omega - 1}{\omega}\right)$ intersects $m_\phi(\sigma)$ in a point whose abscissa is greater than $\xi(\bar{\mu}[B], \omega)$, for all ω with $1 \leqq \omega \leqq \omega_b$. Thus, by Lemma 3, $\mathscr{L}_{\phi,\omega}$ has at least one eigenvalue greater in modulus than $\xi(\bar{\mu}[B], \omega)$, so that

[9] Since $B \in S$, B is non-negative and irreducible, which implies that $\bar{\mu}[B] > 0$.

[10] Without the discussion of the case of equality, this result was stated in [**7**], and proved in [**7'**].

$\xi(\bar{\mu}[B], \omega) < \bar{\mu}[\mathscr{L}_{\phi,\omega}]$ for $1 \leqq \omega \leqq \omega_b$. If ϕ is an h-consistent ordering for B, it can be shown, using basically the proof of this as given originally in [16], that the following functional relationship

$$(\lambda + \omega - 1)^2 = \lambda\omega^2\mu^2 \tag{24}$$

holds, for $\omega \neq 0$, between the eigenvalues λ of $\mathscr{L}_{\phi,\omega}$ and the eigenvalues μ of B. From (24), it follows easily that $\xi(\bar{\mu}[B], \omega) = \bar{\mu}[\mathscr{L}_{\phi,\omega}]$ for $1 \leqq \omega \leqq \omega_b$, which completes the proof of the first part of the theorem.

For $\omega_b \leqq \omega \leqq 2$, we use a result of Kahan [7], which states that for any ordering ϕ and any real value of ω, $\bar{\mu}[\mathscr{L}_{\phi,\omega}] \geqq \omega| - 1|$. Thus, for the indicated range of ω, $\bar{\mu}[\mathscr{L}_{\phi,\omega}] \geqq \omega - 1$. If ϕ is an h-consistent ordering for B, it follows, using (24), that $\bar{\mu}[\mathscr{L}_{\phi,\omega}] = \omega - 1$ for $\omega_b \leqq \omega \leqq 2$. If ϕ is a non-consistent ordering for B, then by the first part of this theorem, $\bar{\mu}[\mathscr{L}_{\phi,\omega_b}] > \xi(\bar{\mu}[B], \omega_b) = \omega_b - 1$, the last equality following from (24) and the definitions of ξ and ω_b. Thus, if ϕ is a non-consistent ordering for B, then $\bar{\mu}[\mathscr{L}_{\phi,\omega}] \geqq \omega - 1$ for $\omega_b \leqq \omega < 2$, with strict inequality for $\omega = \omega_b$, which completes the proof.

COROLLARY. *If $B \in S$, and $\bar{\mu}[B] < 1$, then for all real ω and all orderings ϕ*

$$\min_{\phi}\{\min_{\omega} \bar{\mu}[\mathscr{L}_{\phi,\omega}]\} \geqq \omega_b - 1\ , \tag{25}$$

with equality if and only if B is cyclic of index 2.

Proof. For $\omega \geqq 0$, and $\omega > \omega_b$, $\bar{\mu}[\mathscr{L}_{\phi,\omega}] > \omega_b - 1$ for any ordering ϕ, by Kahan's result [7]. For $\bar{\mu}[B] < 1$, we have that $\xi(\bar{\mu}[B], \omega)$ is a decreasing function of ω for $0 < \omega \leqq \omega_b$. Since, by Theorem 2, there exists a consistent ordering for B if and only if B is cyclic of index 2, the result follows directly from Theorem 4.

4. Asymptotic rates of convergence. If $B \in S$ and $\bar{\mu}[B] < 1$, we define, as usual [16], the *rate of convergence* of the iterative scheme (2) as

$$R_{\phi,\omega} \equiv -ln\bar{\mu}[\mathscr{L}_{\phi,\omega}]\ . \tag{26}$$

In particular, we consider the Gauss-Seidel iterative scheme, the special case of (2) with $\omega = 1$. By the corollary to Theorem 3, in this case,

$$\bar{\mu}^2[B] \leqq \bar{\mu}[\mathscr{L}_{\phi,1}) < \left(\frac{\bar{\mu}[B]}{2 - \bar{\mu}[B]}\right)\ .$$

If $R \equiv -\ ln\bar{\mu}[B]$, we have

THEOREM 5. *If $B \in S$ and $\bar{\mu}[B] < 1$, then for all orderings ϕ*

$$(27)\qquad 1 \geqq \frac{R_{\phi,1}}{2R} > \frac{1}{2} + \frac{ln(2-\bar{\mu}[B])}{-2ln\bar{\mu}[B]} .$$

Thus,

$$(28)\qquad \lim_{\bar{\mu}[B]\uparrow 1} \frac{R_{\phi,1}}{2R} = 1 .$$

Proof. The inequalities of (27) follow directly from the discussion above. Applying L'Hospital's rule,

$$\lim_{\bar{\mu}[B]\uparrow 1} \frac{ln(2-\bar{\mu}[B])}{-2ln\bar{\mu}[B]} = 1/2 ,$$

from which (28) follows.

The above result contains as a special case a proof of a conjecture of Shortley and Weller [10], who observed, from numerical data, that for the numerical solution of the Dirichlet problem in a rectangle on a fine uniform mesh, the rate of convergence of the Gauss-Seidel iterative method is virtually independent of the order in which the points are swept. For illustration, we suppose, following Shortley and Weller, that we are solving numerically the Dirichlet problem in the unit square. Assuming that there are p equal intervals of subdivision in each coordinate direction, we let $u_{i,j}$ denote numerical approximation to $u(x, y)$, the analytic solution of the Dirichlet problem, where

$$x = \frac{i}{p}, y = \frac{j}{p}, 1 \leqq i, j \leqq (p-1) .$$

Making the well-known five-point approximation to Laplace's equation

$$(29)\qquad u_{i,j} = \frac{1}{4}\{u_{i+1,j} + u_{i-1,j} + u_{i,j+1} + u_{i,j-1}\}, 1 \leqq i, j \leqq (p-1) ,$$

where $u_{0,j}$, $u_{p,j}$, $u_{i,0}$, and $u_{i,p}$, determined by the given boundary values of the Dirichlet problem, are known, (29) is except for iteration superscript of the form (2) with $\omega = 1$. The corresponding $(p-1)^2 \times (p-1)^2$ matrix B_1, whose entries are one-fourth or zero, is obviously contained in S, and, as is easily shown, $\bar{\mu}[B_1] = \cos(\pi/p)$.

For completeness, we include also the well-known nine-point approximation to Laplace's equation,

$$(30)\qquad u_{i,j} = \frac{1}{5}\{u_{i+1,j} + u_{i-1,j} + u_{i,j+1} + u_{i,j-1}\} + \frac{1}{20}\{u_{i-1,j+1} + u_{i+1,j+1} + u_{i-1,j-1} + u_{i+1,j-1}\}, 1 \leqq i, j \leqq (p-1) ,$$

corresponding to a $(p-1)^2 \times (p-1)^2$ matrix B_2 which is also contained[11] in S. It can be shown that

$$\bar{\mu}[B_2] = \frac{\cos(\pi/p)}{5}\{4 + \cos(\pi/p)\} .$$

The following table gives information about the quantity

$$Q(\bar{\mu}[B]) \equiv \frac{1}{2}\left\{1 + \frac{ln(2 - \bar{\mu}[B])}{-ln\bar{\mu}[B]}\right\} . \tag{31}$$

p	$\bar{\mu}[B_1]$	$Q(\bar{\mu}[B_1])$	$\bar{\mu}[B_2]$	$Q(\bar{\mu}[B_2])$
10	.951 057	.976 103	.941 747	.971 595
25	.992 115	.996 073	.990 550	.995 065
50	.998 027	.999 014	.997 633	.998 818
100	.999 507	.999 753	.999 408	.999 704

TABLE 1

Thus, for either the five- or nine-point approximation, with $p = 25$ as an example, there is *less* than one-half of one percent difference in the rates of convergence of the Gauss-Seidel iterative scheme for all 576! orderings of the 576 unknowns.

5. Elliptic partial difference equations. We now show how the preceding results can be applied to the numerical solution of certain partial differential equations of elliptic type.

Given a closed bounded region Ω in Euclidean n space with interior R and boundary Γ, and given a function $g(x)$ defined on Γ, we seek a function $u(x)$ defined in Ω which is continuous in Ω, twice differentiable in R, which satisfies

$$\sum_{k=1}^{n} A_k(x)\frac{\partial^2 u}{\partial x_k^2} + F(x)u = G(x),\ x \in R , \tag{32}$$

and

$$u(x) = g(x), \qquad x \in \Gamma . \tag{33}$$

It is assumed[12] that the functions $F, G, A_1, \cdots, A_n$ are given functions of x which are continuous in Ω and twice-differentiable in R, and satisfy the conditions

$$A_k(x) > 0, \qquad F(x) \leqq 0, \qquad x \in \Omega, \qquad 1 \leqq k \leqq n . \tag{34}$$

After a cartesian mesh is laid over the closed region Ω, the above partial differential equation and boundary conditions are approximated [**16**, **14**] by the following system of N linear equations

[11] For $p \geqq 3$, the matrix B_1 is cyclic of index 2, while B_2 is primitive.

[12] For the numerical solution of (32) where $F, G, A_1, \cdots, A_n$ are only piecewise smooth, see for example [**14**].

$$\sum_{j=1}^{N} a_{i,j}x_j = k_i, \qquad 1 \leqq i \leqq N \,, \tag{35}$$

where N is the number of mesh points interior to Ω. If the mesh is sufficiently fine, the discrete approximation can be derived in such a way that the $N \times N$ matrix $A = ||a_{i,j}||$ satisfies the following properties:

(36)
(i) $A = ||a_{i,j}||$ is symmetric and irreducible.
(ii) $a_{i,j} \leqq 0$ for $i \neq j$, $1 \leqq i, j \leqq N$.
(iii) $\sum_{j=1}^{N} a_{i,j} \geqq 0$ for all i, $1 \leqq i \leqq N$, with strict inequality for some i.

The matrix A is thus positive definite [13]. If D is the $N \times N$ positive diagonal matrix with entries $a_{i,i}$, we may write (35) in the equivalent form:

$$(D^{-1/2}AD^{-1/2})D^{1/2} = \boldsymbol{x}\, D^{-1/2}\boldsymbol{k} \,, \tag{35'}$$

where $\boldsymbol{x}$ and $\boldsymbol{k}$ are column vectors with components x_i and k_i, $1 \leqq i \leqq N$, respectively. If $D^{1/2}\boldsymbol{x} \equiv \boldsymbol{y}$, $D^{-1/2}\boldsymbol{k} \equiv \boldsymbol{g}$, and $D^{-1/2}AD^{-1/2} \equiv \tilde{A}$, (35′) reduces to

$$\tilde{A}\boldsymbol{y} = \boldsymbol{g} \,. \tag{37}$$

Since $\tilde{A}$ has unit diagonal entries, we define the matrix $\tilde{B}$ as $\tilde{B} \equiv I - \tilde{A}$, and (37) can be written in the form

$$\boldsymbol{y} = \tilde{B}\boldsymbol{y} + \boldsymbol{g} \,. \tag{37'}$$

It follows from the definition of $\tilde{B}$ that $\tilde{B}$ is a non-negative irreducible and symmetric $N \times N$ matrix, which has zero diagonal entries. Thus, $\tilde{B} \in S$. Since A is positive definite, so is $\tilde{A}$, and from $\tilde{A} = I - \tilde{B}$, it follows that $\bar{\mu}[\tilde{B}] < 1$. Thus, the discrete numerical approximation to (32)-(33) can be reduced to the form (37′) where $\tilde{B} \in S$, and the results of the preceding sections are applicable.

BIBLIOGRAPHY

1. R. J. Arms, L. D. Gates, and B. Zondek, *A method of block iteration,* Journal Soc. Indust. Appl. Math., **4** (1956), 220-229.

2. Garrett Birkhoff and Richard S. Varga, *Reactor criticality and non-negative matrices,* Journal Soc. Indust. Appl. Math. **6** (1958), 354-377.

3. Gerard Debreu and I. N. Herstein, *Nonnegative square matrices,* Economentrica, **21** (1953), 597-607.

4. Stanley P. Frankel, *Convergence rates of iterative treatments of partial differential equations,* Math. Tables, and Other Aids to Computation, **4** (1950), 65-75.

5. Frobenius, *Über Matrizen aus nicht negativen Elementen,* Sitzungsberichte der Akademie der Wissenschaften zu Berlin, (1912), 456-477.

6. J. Heller, *Ordering properties of linear successive iteration schemes applied to multi-diagonal type linear systems*, Journal Soc. Indust. Appl. Math., **5** (1957), 238-243.

7. W. Kahan, *The rate of convergence of the extrapolated Gauss-Seidel iteration*, presented at the Conference on Matrix Computations, Wayne State University, September 4, 1957.

7′. W. Kahan, *Gauss-Seidel methods of solving large systems of linear equations*, Doctoral Thesis, University of Toronto, 1958.

8. O. Perron, *Zur Theorie der Matrices*, Math., **64** (1907), 259-263.

9. V. Romanovsky, *Recherches sur les chaines de Markoff*, Acta Math., **66** (1936), 147-251.

10. G. H. Shortley and R. Weller, *The numerical solution of Laplace's equation*, Journal Appl. Phys., **9** (1938), 334-344.

11. R. H. Stark, *Rates of convergence in numerical solution of the diffusion equation*, Journal Assoc. Computing Machinery, **3** (1956), 29-40.

12. P. Stein and R. L. Rosenberg, *On the solution of linear simultaneous equations by iteration*, Journal London Math. Soc., **23** (1948), 111-118.

13. O. Taussky, *A recurring theorem on determinants*, Amer. Math. Monthly, **56** (1949), 672-676.

14. R. S. Varga, *Numerical solution of the two-group diffusion equation in $x-y$ geometry*, IRE Trans. of the Professional Group on Nuclear Science, NS-4 (1957), 52-62.

15. Richard S. Varga, *p-cyclic matrices: a generalization of the Young-Frankel successive overrelaxation scheme*, Pacific J. Math., **9** (1959), 617-628.

16. David Young, *Iterative methods for solving partial difference equations of elliptic type*, Trans. Amer. Math. Soc., **76** (1954), 92-111.

17. David Yound, *Iterative methods for solving partial difference equations of elliptic type*, Doctoral Thesis, Harvard University, 1950.

18. David M. Young, *ORDVAC solutions of the Dirichlet problem*, Journal Assoc. Computing Machinery, **2** (1955), 137-161.

BETTIS ATOMIC POWER DIVISION
WESTINGHOUSE ELECTRIC CORP.

ON WEAK DIMENSION OF ALGEBRAS

ORLANDO E. VILLAMAYOR

1. Introduction. In this note we try to characterize algebras whose weak dimension is zero, i.e., algebras A which are flat A^e-modules.

In this direction, Theorems 1 and 2 give the corresponding results, for weak dimension, to known theorems for (strong) dimension. However, it seems to be more interesting to find relations between these two dimensions.

Theorem 3 gives such a relation for commutative algebras over a field. For the non-commutative case, only a weaker necessary condition is found in Theorem 5. However, in the case of algebras satisfying the descending chain condition for left ideals a complete picture of the 0-weak dimensional ones is given in Theorem 6.

Section 6 applies these results to group algebras. In [2] Auslander partially succeded in characterizing (von Neumann) regular group algebras. However, concerning the group, he only proved the necessity of the group being torsion and the sufficiency of the local finiteness. The difference seemed to be related to the Burnside problem. Theorem 8, then, fills the gap and the problem is now completely solved.

In the last section we study some relaitons between weak dimensionality and semisimplicity (in the sense of Jacobson) in tensor products of algebras.

After this paper was written we received a copy of a paper by Prof. Harada on the same subject [4]. However, there is no overlapping of the main results.

We would like to express our thanks and indebtedness to Professor Rosenberg for this helpful advice and criticism.

2. Notations and terminology. Throughout this note we use the homological notation and terminology of [2].

Since we are dealing with algebras over a (fixed) ground ring K, all tensor products are suppose to be taken over the ground ring K, unless otherwise specifically expressed, so, we shall use $\otimes$ for $\otimes_K$. Similarly, homological dimension of algebras are indicated by dim A or w.dim. A if they are considered over K, or R-dim A (resp. R-w.dim A) if they are considered over another ring R.

For a ring, simple and semisimple mean simple and semisimple with minimum condition for one-sided ideals. Regular will always mean regular in the sense of von Neumann.

Received October 10, 1958, and in revised form February 9, 1959. This work was partially supported by a J. S. Guggenheim fellowship and by the Office of Naval Research.

Semisimplicty in the sense of Jacobson is called J-semisimplicity.

3. Characterization of O-weak dimensional algebras. Let A be an algebra over a commutative ring K. The dimension (resp. weak dimension) of A as an algebra is, following the classical definitions, the dimension (resp. w.dim) of A as an A^e-module, where $A^e = A \otimes A^*$ (A^* the algebra anti-isomorphic to A). Since A is a cyclic A^e-module, we shall start with some considerations on cyclic flat modules (i.e. cyclic modules M with w.dim.$M = O$).

LEMMA 1. *Let R be any ring and A a cyclic left R-module. Then the following conditions are equivalent*:

(a) *A is R-flat.*

(b) $\mathrm{Tor}_1^R(R/I, A) = O$ *for every principal right ideal I in R.*

(c) *If $A = R/J$, a the image of 1 in A and $x \in J$, there exists $y \in R$ such that $xy = O$, $ya = a$.*

Proof. (a) $\Rightarrow$ (b) is obvious.
(b) $\Rightarrow$ (c). Let $x \in J$ and let I be the right ideal generated by x. According to ([3], VI, Ex. 19, p. 126) condition (b) implies $I \cap J = IJ$. Since $x \in I \cap J$, then $x \in IJ$, that is, there is a $z \in J$ such that $x = xz$, hence $za = O$, and $y = 1 - z$ verifies $xy = O$ and $ya = a$. (c) $\Rightarrow$ (a). Let $B = R/I$, for any right ideal I. If $x \in I \cap J$, condition (c) assures the existence of an element $z \in J$ such that $xz = x$, so that $x \in IJ$, hence $I \cap J = IJ$. That is, $\mathrm{Tor}_1^R(B, A) = O$ for every cyclic module B ([3], VI, Ex. 19), so A is flat ([5]).
As a consequence, we obtain

LEMMA 2. *Let $A = R/J$ be a cyclic flat left-R-module. If I is a finitely generated left ideal contained in J, there is a principal left ideal I' such that $I \subseteq I' \subseteq J$.*

Proof. We proceed by induction on the number of generators of I. If I has one generator, $I = I'$. Suppose the lemma is true if I has $n - 1$ generators, and suppose $x_1, \cdots, x_n$ generate I. Let us call a the image of 1 in A. If $x_1, x_2 \in I \subseteq J$, then $x_1 a = x_2 a = O$ and there is an element $y \in R$ such that $x_1 y = O$ and $ya = a$, hence $x_2 ya = O$, and there is a $z \in R$ such that $x_2 yz = O$ and $za = a$, so $yza = a$. If we call $r = 1 - yz$ then $x_1 r = x_1$, $x_2 r = x_2$ and $ra = O$. This last condition implies $r \in J$ and $I \subseteq I_1 \subseteq J$ where I_1 is the ideal generated by $r, x_3, \cdots, x_n$.

From these lemmas, the following well known result may be immediately proved:

COROLLARY 1. *If a cyclic left module $A = R/I$ is R-flat and I is* *finitely generated, then A is R-projective.*

In fact, Lemma 2 implies I is generated by a single element, say x, and Lemma 1 assures the existence of $y \in I$ with $xy = x$, hence $R \to Ry$ is a projection of R onto I and I is a direct summand in R, so A is projective.

Now, we shall apply these results to characterize O-weak dimensional algebras and the following theorem corresponds to that one given in [3] (IX, Proc. 7.7, p, 179) for dim. $A = O$.

THEOREM 1. *In order that* w.dim $A = 0$ *it is necessary and sufficient that, for every finite set* $\{a_1, \cdots, a_n\}$ *in* A, *there exists an element* e *in the two-sided* A*-module* $A \otimes A$ *such that* $a_i e = e a_i (1 \leq i \leq n)$ *and that, under the mapping* $x \otimes y \to xy$ *the image of* e *in* A *is* 1.

Proof. Let $a_1, \cdots, a_n \in A$. Suppose w.dim $A = O$, i.e., A is A^e-flat. The elements $1 \otimes a_i^* - a_i \otimes 1^*$ belong to $J = \mathrm{Ker}(A^e \to A)$, then they are contained in a principal left-ideal $I \subseteq J$. If z is the generator of I, $z.1 = O(1 \in A)$, then there is an element e such that $e.1 = 1$ and $ze = O$, hence $(1 \otimes a_i^* - a_i \otimes 1^*)e = O$ and the necessity of the conditions proved.

To prove the sufficiency, let us consider an element $z \in J$. Thus $z = \sum y_i (1 \otimes a_i^* - a_i \otimes 1)(y_i \in A^e, a_i \in A)$, so, there is an $e \in A^e$ such that $(1 \otimes a_i^* - a_i \otimes 1)e = O$, $e.1 = 1$, hence $ze = O$ and Lemma 1 implies A is A^e-flat.

As a consequence of Theorem 1 and [3], (IX, prop. 7.7) we obtain

COROLLARY 2. *If A is a finitely generated K-algebra, then* w.dim $A = 0$ *if and only if* dim. $A = 0$.

Of course, this result may also be obtained from Corollary 1 and the fact that $\mathrm{Ker}(A^e \to A)$ is a left ideal generated by the set $\{a_i \otimes 1^* - 1 \otimes a_i^*\}$, where the a_i's generate A as an algebra.

Now, following the same lines given by Rosenberg and Zelinsky ([9], Th. 1, p. 88) we prove

THEOREM 2. *Let A be a K-algebra which is free as a K-module. If* w.dim $A = O$, *then A is locally finite over K.*[1]

Proof. Let $\{x_i\}$ be a K-basis of A and $\{b_1, \cdots, b_n\}$ be a finite subset in A. If B is the subalgebra generated by the set $\{b_1, \cdots, b_n\}$, then, for every $z \in B$, $1 \otimes z^* - z \otimes 1 \in A^e$ is contained in the left-ideal generated by the set $\{1 \otimes b_i^* - b_i \otimes 1\}$.

[1] An algebra A over a ring K will be called locally finite if every finitely generated subalgebra is contained in a finitely generated free K-submodule of A.

Theorem 1 shows the existence of elements $y_1, \cdots, y_k$ such that $(1 \otimes b_i^* - b_i \otimes 1) \sum_j x_j \otimes y_j^* = O$ and $\sum_j x_j y_j = 1$. Thus

$$(1 \otimes z^* - z \otimes 1^*) \sum_j x_j \otimes y_j^* = O$$

for every $z \in B$, that is,

$$\sum z x_j \otimes y_j = \sum x_j \otimes y_j z \ .$$

If we write zx_j as a linear combination of the x_i's, this formula shows that $y_j z$ is a linear combination of the y's, that is,

$$y_j z = \sum_i k_{ij} y_i \ (k_{ij} \in K)$$

hence, $z = \sum_j x_j y_j z = \sum_{i,j} k_{ij} x_j y_i$, and, then, B is contained in the K submodule generated by the set $\{x_j y_i\}$

Finally, if we write the elements $x_j y_i$ in terms of a basis, say $\{x_k\}$, since only a finite number of x_k's appear in each $x_j y_i$, B is contained in a finitely generated K-free K-submodule of A.

4. Algebras over a field. In the case K is a field, then, trivially, A is a K-free K-module and, if w.dim $A = O$, the conditions of Theorems 1 and 2 must be satisfied.

The results of [3] (IX, 7.5 and 7.6), referred to weak dimension (i.e., starting from IX. 28 intead of IX. 2.8a) may be condensed, by using the equivalence between w.gl.dim $R = O$ and R being a (von Neumann) regular ring [5], in the following proposition:

PROPOSITION 1. *If A is a K-algebra over a regular ring K, then* w.dim $A = O$ *if and only if A^e is a regular ring.*

In the case of commutative algebras over a field a complete characterization of the case w.dim $A = O$ is obtained in the following result.

THEOREM 3. *Let A be a commutative algebra over a field K. Then, then following conditions are equivalent:*

(i) *A is locally separable*[2]

(ii) w. dim $A = O$

(iii) *$A \otimes F$ is regular for every field F containing K.*

Proof. (i) $\Rightarrow$ (ii). Obviously, since A is locally separable, it satisfies the conditions of Theorem 1.

(ii) $\Rightarrow$ (iii). This is a trivial consequence of the inequality

$$\text{w.gl.dim}\, A \otimes F \leq \text{w.gl.dim}\, F + \text{w.dim}\, A$$

obtained from the spectral sequences [3] (XVI, 5.5a p.347) and the equivalence between w.gl.dim $R = 0$ and regularity obtained in [5].

[2] An algebra A over a field K is called locally separable if every finitely generated subalgebra is contained in a (finitely generated) separable subalgebra.

(iii) $\Rightarrow$ (i). If A is commutative and it is not locally finite over K, then there is at least one element x which is transcendental over K, hence A contains a subalgebra isomorphic with the polynomial ring $K[x]$.

For every polynomial $p(x)$, let 0_p be the set of elements $y \in A$ such that $yp(x) = 0$. Let $I = \cup 0_p$, then, trivially, I is an ideal in A, and no element of $K(x)$ is in I; otherwise, if $q(x) \in I$, there is a $p(x)$ such that $q(x)p(x) = O$, contradicting the transcendency of x.

If A is regular, for every $p(x)$ there is an element z such that $z(p(x))^2 = p(x)$, hence $1 - zp(x) \in I$ and, in A/I, the images of all $p(x)$ have inverses, so A/I contains a subalgebra isomorphic with the field of rational functions $K(X)$.

Let us call $B = A/I$. If A is regular, then B is regular too, and, from the exactness of $A \to B \to 0$ we obtain $F \otimes A \to F \otimes B \to 0$ exact. Then, if $F \otimes A$ is regular, so is $F \otimes B$.

Since $B \supseteq K(X)$ and $K(X)$ is a field, B is the direct sum of $K(X)$-modules isomorphic with $K(X)$, hence, from the fact that $\otimes$ distributes on direct sums, $F \otimes B$ is a direct sum of $F \otimes K(X)$-modules isomorphic with $F \otimes K(X)$. Applying now ([2], Prop. 2, p. 659), we obtain w.gl.dim $F \otimes B \geq$ w.gl.dim $F \otimes K(X)$. Then, we must prove just that $F \otimes K(X)$ is not a regular ring. In fact, if F is any field containing properly K, then $F \otimes K(X)$ contains a subring isomorphic with $F \otimes K[X] \approx F[X]$, which is an integral domain, and $F \otimes K(X)$ is the set of rational functions $q(x)/p(x)$ with $q(x) \in F[X]$ and $p(x) \in K[X]$, hence, it is an integral domain but not a field because it has no inverse for $q(x) \in F[X]$ if $q(x) \notin K[X]$, thus $F \otimes K(X)$ is not a regular ring.

Thus, condition (iii) implies A is locally finite.

Since $A \otimes F$ is regular and commutative, $B \otimes F$ has to be semi-simple for every finitely generated subalgebra B and every field F containing K, hence B is separable and so A is locally separable.

The result of Corollary 2 can be extended, in the case of algebras over a field, by using the following result of Kaplansky ([7], Lemma 1).

LEMMA 3. *If I is a countably generated left-ideal in a regular ring R, then* $\dim_R I = O$.

A direct implication of this lemma is obtained in

THEOREM 4. *Let A be an algebra over a field K. If $[A:K] = \aleph_0$ and* w.dim $A = O$, *then* dim. $A = 1$.

Proof. Since w.dim $A = O$ implies $A \otimes A^*$ regular, and $\mathrm{Ker}(A^e \to A)$ is generated by the set $\{x_i \otimes 1 - 1 \otimes x_i^*\}$ (where the x_i's are generators of A) and this set is countable, then $\mathrm{Ker}(A^e \to A)$ is projective. Thus, $\dim A \leq 1$. Since $\dim A = O$ implies $[A; K]$ finite, then $[A:K] = \aleph_0$ implies $\dim A = 1$.

We shall say that an algebra A is locally one dimensional if every finite set of elements is contained in a subalgebra B such that $\dim B \leq 1$. The following theorem approximates the result obtained in Theorem 3 for commutative algebras.

THEOREM 5. *Let A be an algebra over a field K. If* $\text{w.dim}\, A = 0$, *then A is locally one dimensional.*

Proof. Let $\{a_1, \cdots, a_n)$ be a finite set of elements in A. Let B_0 be the subalgebra generated by this set. Since A^e is regular, there is an idempotent e_1 such that $(1 \otimes a_i^* - a_i \otimes 1)e_1 = 0$ and e_1 is mapped onto 1 by the natural map $\sigma: A^e \to A$ (In fact, the left ideal generated by the finite set $\{1 \otimes a_i^* - a_i \otimes 1\}$ is generated by an idempotent $1 - e_1$ which is mapped onto 0).

If $x \in B$, then $1 \otimes x^* - x \otimes 1^* = \sum y_i (1 \otimes a_i^* - a_i \otimes 1)(y_i \in A^e)$, thus $(1 \otimes x^* - x \otimes 1^*)e_1 = 0$ for every $x \in B_0$.

Let $\{b_1, \cdots, b_m\}$ be the set of elements of A appearing in e_1, and B_1 the subalgebra generated by $\{a_1, \cdots, a_n, b_1, \cdots, b_m\}$. Then, by the same arguments, there is an idempotent $e_2 \in A^e$ such that $(x \otimes 1 - 1 \otimes x^*)e_2 = 0$ for every $x \in B_1$, and $\sigma(e_2) = 1$.

By repeating the process we obtain a chain of subalgebras $B_0 \subseteq B_1 \subseteq B_2 \subseteq \cdots$ If we call $B = \cup B_i (i = 1, 2, \cdots)$ then Theorem 1 implies $\text{w.dim}\, B = 0$. In fact, for every finite subset $\{x_1, \cdots, x_h\}$ in B there is a finitely generated subalgebra B_k with $x_i \in B_k (1 \leq i \leq h)$, then e_{k+1} satisfies $(x_i \otimes 1^* - 1 \otimes x_i^*)e_{k+1} = 0$, $\sigma(e_{k+1}) = 1$, and $e_{k+1} \in B^e$.

Since B is, at most, countably generated, then Corollary 2 and Theorem 4 imply $\dim B \leq 1$.

REMARK 1. According to Proposition 1, if A is an algebra over a regular ring K, $\text{w.dim}\, A = 0$ implies A^e is regular. Then, in this case, Theorem 4 may be expressed in the following way:

THEOREM 4'. *Let A be an algebra over a regular ring K. If A is denumerably generated and* $\text{w.dim}\, A = 0$, *then* $\dim A \leq 1$.

Thus, Theorem 5 is valid for algebras over a regular ring K.

5. Algebras with descending chain condition. Theorem 3 shows that, for a commutative algebra over a field, $\text{w.dim}\, A = 0$ if and only if A is locally separable. We do not know whether this statement is true in the non commutative case.

In the case of algebras satisfying the descending chain condition for left-ideals, the following result, suggested to the author by Professor Rosenberg, characterizes completely the 0-w. dimensional case.

THEOREM 6. *Let A be an algebra over a field K satisfying the descending chain condition for left ideals. Then,* w.dim $A = 0$ *if and only if*:

(a) *A is semisimgle*
(b) *A is locally finite over K*
(c) *The center C of A is locally separable.*

Proof. If w.dim $A = 0$, condition (a) follows from the regularity of A, (b) from Theorem 1 and (c) from Theorem 3.

Suppose, now A satisfies (a), (b) and (c). Since A is semisimple, it is a direct sum of (a finite number of) simple algebras S_i satisfying conditions (b) and (c), and, because of the direct sum decomposition, w.dim $A = \max(\text{w.dim } S_i)$. Since each S_i is a matrix ring over a division algebra D_i satisfying (b) and (c) and w.dim $S_i =$ w.dim D_i, it will be enough to prove the sufficiency of these conditions for division algebras.

Let A be a division algebra. Condition (c) implies w.dim $C = O$. According to the sub-additivity of the dimension ([9], Th. 5, p.93) we have w.dim $A \leq$ w.dim. $C + C$-w.dim A, then, it is sufficient to prove that $A \otimes_C A^*$ is regular. This is so if $A \otimes_C S^*$ is regular for every finitely generated subalgebra S of A.

Since A^* is locally finite and S^* finitely generated, then $[S^* : C] < \infty$ and S^* is a division ring. Thus $A \otimes_Z S^*$ satisfies the descending chain condition. Since A is central simple and S^* simple (because now we are considering A and S as algebras over C), then $A \otimes_C S^*$ is simple, hence regular, and the theorem is proved.

6. Group algebras. In [2], Auslander studies necessary and sufficient conditions for a group G and a ring K to obtain (von Neumann) regular group algebras $K(G)$. He proved the necessity of G being a torsion group and the sufficiency of G being locally finite, besides the conditions on K.

In Theorem 8 we prove the necessity of the local finiteness, and then regular group algebras are completely characterized.

A similar difference existed between Theorem 3 and 4 in [8], but by direct aplication of Theorem 2 we fill the gap obtaining the following result.

THEOREM 7. *Let G be a group, S a subgroup contained in the center of G and K any commutative ring, then $K(S)$-w.dim $K(G) = O$ if and only if G/S is locally finite and K is uniquely divisible by the order of each element in G/S.*

In fact, the local finiteness of $K(G)$ as a $K(S)$-algebra implies the local finiteness of G/S.

THEOREM 8. *Let G be a group and K any commutative ring. Then $K(G)$ is regular if and only if G is locally finite and K is a regular ring uniquely divisible by the order of each element in G.*

Proof. A trivial modification in the proof of ([3] X. 6.1) may be used to prove

$$\text{w.dim}\, K(G) = \text{r.w.dim}_{K(G)} K\,.$$

Thus, $K(G)$ regular implies $\text{w.dim}\, K(G) = O$ and Theorem 2 implies G is locally finite.

The remaining part of the proof follows from Auslander's result. It also may be seen as a special case of Theorem 7.

REMARK 2. The proof of the necessity of the local finiteness of G for a group algebra $K(G)$ to be regular does not need all the homological machinery. In fact, it follows immediately from the following lemma:

LEMMA 4. *Let $K(G)$ be the group algebra generated by a group G over any commutative ring K and $g_1, \cdots, g_n$ be elements of G. Then the subgroup S generated by $\{g_1, \cdots, g_n\}$ is finite if and only if there is an element $x \in K(G)$ such that $(1 - g_i)x = 0\, (1 \leq i \leq n)$. If this is the case, $x = sy$, where s is the sum of all elements in S.*

Proof. If S is finite, the sum s of all elements in S satisfies the equations $(1 - g_i)x = 0$, and so every product sy.

Conversely, suppose $(1 - g_i)x = 0$ $(1 \leq i \leq n)$. Thus, $x = g_1x = \cdots = g_nx$. Since every $f \in S$ is a product of powers of the g_i's, then $fx = x$. Let $x = \sum_1^m k_jh_j (h_j \in G)$. For every $f \in S, fx = x$ implies x has a term k_1fh_1 and so all elements of S appear multiplied by k_1h_1, hence S is finite (because x is a finite sum). Furthermore, we obtained $x = k_1sh_1 + x'$, with $(1 - g_i)x' = 0$. By induction on the number of terms in x we obtain the last result.

A complete proof of Theorem 8 may be obtained as follows: Suppose $K(G)$ is regular. Then the ring homomorphism $\sigma : K(G) \to K$ defined by the group homomorphism $G \to \{1\}$ implies K is regular.

If $K(G)$ is regular, every finitely generated proper left ideal is a direct summand, hence it is annihilated, on the right, by a non-zero element $x \in K(G)$. Since all $1 - g$ are in $\text{Ker}\, \sigma$, every finite set generates a proper left ideal, and so the previous lemma implies G is locally finite.

Suppose $g \in G$ has order n. By Lemma 1 there is an element x with $\sigma(x) = 1$ and $(1 - g)x = O$, hence Lemma 4 implies $x = sy (s = \sum g^i)$ and so $\sigma(x) = \sigma(s)\sigma(y) = n\sigma(y) = 1$, hence n has an inverse in K

and the necessity of the conditions is proved.

Suppose, now, K and G satisfy the conditions of the theorem. Let $x = \sum k_i g_i \in \operatorname{Ker} \sigma$, so, $x = \sum k_i(g_i - 1)$. Let S be the subgroup generated by $g_1, \cdots, g_n$, m its order and s the sum of all elements in S. Since m has an inverse m^{-1} in K, then $y = m^{-1}s$ satisfies $xy = O$, $\sigma(y) = 1$, and Lemma 1 implies K is $K(G)$-flat. So $\text{w.dim}\, K(G) = \text{w.dim}_{K(G)} K = O$ ([3] X, 6.2) and $K(G)$ is regular.

7. Weak dimension and Jacobson semisimplicity. A ring will be called J-semisimple if its Jacobson radical is (O).

If T is a ring and M a left-T-module, then the ring $\operatorname{Hom}_T(M, M)$, with the operations defined in the classical way is a topological ring by defining the finite topology induced by M. ([6], Ch. IV).

If we are in the situation $S \subseteq R \subseteq \operatorname{Hom}_T(M, M)$, where S and R are rings, we shall say S *is dense in* R if it is so in the finite topology induced by $\operatorname{Hom}_T(M, M)$.

In this section we shall prove the following theorem:

THEOREM 9. *Let A be a K-projective K-algebra. If B is J-semisimple K-algebra and* $\text{w.dim}\, A = 0$, *then $A \otimes B$ is J-semisimple.*

Before proving the Theorem we shall state the following lemmas:

LEMMA 5. *Let T be any ring and M a left-T-module. If S, R are rings such that $S \subseteq R \subseteq \operatorname{Hom}_T(M, M)$, R is regular and S is dense in R, then S is J-semisimple.*[3]

Proof. Let $x \in S$. Since R is regular, there are elements $y, z \in R$, $z \neq O$, such that $z(1 - xy) = O$. Since $R \subseteq \operatorname{Hom}_T(M, M)$, there is at least one $m \in M$ such that $m^z \neq O$ and $m^{z(1-xy)} = O$, that is, there exists an $n \in M (n = m^z)$ such that $n \neq O$ and $n = n^{xy}$. Now, we have $x \in S$, $y \in R$, $(n^x)^y = n$, and S is dense in R, then there is an $u \in S$ such that $(n^x)^u = n$, that is, $n^{1-xu} = O$, and $1 - xu$ can not have an inverse in S, so xu is not quasi-regular, and S is J-semisimple.

LEMMA 6. *If A is a K-projective K-algebra and B is a K-algebra which is a subdirect sum of K-algebras P_i, then $A \otimes B$ is a subdirect sum of $A \otimes P_i$.*

Proof. B is a subdirect sum of P_i's if and only if the sequences $B \to P_i \to 0$ and $0 \to B \to \prod P_i$ are exact.

[3] The conditions of the lemma are, evidently, stronger that those which are really needed in the proof. In fact, we only need S to be 1-fold transitive in R and R J-semisimple in which, for every element x there is an y such that 1-xy has a left annihilator.

It may be seen that, if S is commutative, the conditions of Lemma 5 are necessary.

Now from the exact sequence $B \to P_i \to 0$ we obtain $A \otimes B \to A \otimes P_i \to 0$ exact. We need only to prove the exactness of $0 \to A \otimes B \to \prod (A \otimes P_i)$.

Since A is K-projective, we have, from $0 \to B \to \prod P_i$, the exact sequence $0 \to A \otimes B \to A \otimes \prod P_i$. We have ([3], Ex. II. 2, 31) a natural homomorphism

$$A \otimes \prod P_i \to \prod (A \otimes P_i)$$

which is, trivially, a monomorphism if $A = K$. Since $\otimes$ commutes with direct sums, it is a monomorphism if A is K-free and, a posteriori, if A is K-projective. Then the composite map gives the exact sequence

$$0 \to A \otimes B \to \prod (A \otimes P_i)$$

and the lemma is proved.

Proof of the theorem. Since B, being semisimple, is a subdirect sum of primitive rings P_i, then the previous lemma implies that $A \otimes B$ is a subdirect sum of rings $A \otimes P_i$; then, to prove the theorem it is sufficient to show that the rings $A \otimes P_i$ are J-semisimple. Now, since P_i is primitive, it is dense in a ring of linear transformations, that is, $P_i \subseteq R_i = \mathrm{Hom}_{F_i}(M_i, M_i)$ where the rings R_i are regular and the P_i's are dense in the R_i's. Since A is K-projective, we may apply the spectral sequences [3] (XVI, 5a, p. 347) and then, R_i regular and w.dim $A = O$ give $A \otimes R_i$ regular.

If we show the inclusion $A \otimes P_i \subseteq A \otimes R_i \subseteq \mathrm{Hom}_{A \otimes F_i}(A \otimes M_i, A \otimes M_i)$ and the density of $A \otimes P_i$ into $A \otimes R_i$, Lemma 5 completes the proof of the theorem.

Since A is K-projective, we have the exact sequence $0 \to A \otimes S_i \to A \otimes R_i = A \otimes \mathrm{Ho}_F\mathrm{m}(M_i, M_i)$.

If A is K-free, the natural mapping $A \otimes \mathrm{Hom}_F(M_i, M_i) \to \mathrm{Hom}_F(M, A \otimes M_i)$ is the natural mapping of a direct sum into a direct product, which is a monomorphism. Since A, being projective, is a direct summand of a free module, and since both $\otimes$ and Hom commute with finite direct sums, then the given mapping is also a monomorphism.

From the natural isomorphism $\mathrm{Hom}_F(M_i, A \otimes M_i) \approx \mathrm{Hom}_{A \otimes F}(A \otimes M_i, A \otimes M_i)$ we obtain the inclusions.

$$A \otimes S_i \subseteq A \otimes R_i \subseteq \mathrm{Hom}_{A \otimes F_i}(A \otimes M_i, A \otimes M_i)$$

Let $x \in A \otimes R_i$, then $x = \sum a_j \otimes r_j$, and $v_k = \sum_l b_{kl} \otimes m_{kl} (b_{kl} \in A, m_{kl} \in M_i)$ be a finite set of elements in $A \otimes M_i$. Then $x(v_i) = \sum_{l,j} a_j b_{kl} \otimes r_j(m_{kl})$. Since the set $\{m_{kl}\}$ is finite and B_i is dense in R_i, for each r_j there is an $s_j \in B_i$ such that $r_j(m_{kl}) = s_j(m_{kl})$, then $y = \sum a_j \otimes s_j \in A \otimes P_i$ and $y(v_k) = x(v_k)$, so $A \otimes P_i$ is dense in $A \otimes R_i$.

As a consequence of this theorem we can state the following corollary:

COROLLARY 3. *If A, B are algebras over a field K and* w.dim $B = O$, *Then* $J(A \otimes B) = J(A) \otimes B$. (We call $J(R)$ the Jacobson radical of a ring R).

In fact, since $A/J(A)$ is semisimple, from the exact sequence

$$0 \to J(A) \otimes B \to A \otimes B \to (A/J(A)) \otimes B \to 0$$

we obtain $J(A) \otimes B \supseteq J(A \otimes B)$.

From Theorem 2 and ([6], V. 14, Th. 1, p 123) it follows that every element in $J(A) \otimes B$ is quasi-regular, so it is a radical ideal in $A \otimes B$. Thus $J(A) \otimes B \subseteq J(A \otimes B)$ and the corollary is proved.

This result generalizes ([1], Th. 1).

REFERENCES

1. S. A. Amitsur, *The radical of field extensions*, Bull. of the Research Council of Israel, **7F** (1957), 1-10.
2. M. Auslander, *On regular group rings*, Proc. Amer. Math. Soc., **8** (1957), 658-664.
3. H. Cartan and S. Eilenberg, *Homological Algebra*, Princeton University Press, 1956.
4. M. Harada, *The weak dimension of algebras and its applications*, J. Inst. Polytechnics, Osaka City University **9** (1958), 47-58.
5. ———, *Note on the dimension of modules and algebras*, J. Inst. Polytechnics, Osaka City University **7** (1956), (17-27).
6. N. Jacobson, *Structure of rings*, Amer. Math. Soc., Coll. Pub., XXXVII, 1956.
7. I. Kaplansky, *On the dimension of modules and algebras*, Nagoya Math. Journal, **13** (1938), 85-89.
8. J. E. McLaughlin, *A note on regular group rings*, Mich. Math. Jour., **5** (1958), 127-128.
9. A. Rosenberg and D. Zelinsky, *Cohomology of infinite algebras*, Trans. Amer. Math. Soc., **82** (1956), 85-98.
10. O. E. Villamayor, *On the semisimplicity of group algebras*, Proc. Amer. Math. Soc., **9** (1958), 621-627.

UNIVERSIDAD NACIONAL DE LA PLATA
THE INSTITUTE FOR ADVANCED STUDY
NORTHWESTERN UNIVERSITY

Pacific Journal of Mathematics

Vol. 9, No. 3 July, 1959