Pacific Journal of Mathematics

IN THIS ISSUE—

Vol. 14, No. 2 **June, 1964**

PACIFIC JOURNAL OF MATHEMATICS

Mathematical papers intended for publication in the *Pacific Journal of Mathematics* should by typewritten (double spaced), and on submission, must be accompanied by a separate author's résumé. 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. Effective with Volume 13 the price per volume (4 numbers) is $18.00; single issues, $5.00. Special price for current issues to individual faculty members of supporting institutions and to individual members of the American Mathematical Society: $8.00 per volume; single issues $2.50. Back numbers are available.

Subscriptions, orders for back numbers, and changes of address should be sent to Pacific Journal of Mathematics, 103 Highland Boulevard, Berkeley 8, 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

ON THE FUNCTIONAL EQUATION
$F(mn)F((m, n)) = F(m)F(n)f((m, n))$

TOM M. APOSTOL AND HERBERT S. ZUCKERMAN

1. Introduction. Let f be a completely multiplicative arithmetical function. That is, f is a complex-valued function defined on the positive integers such that

$$f(mn) = f(m)f(n)$$

for all m and n. We allow the possibility that $f(n) = 0$ for all n. (If f is not identically zero then we must have $f(1) = 1$.) Given such an f we wish to study the problem of characterizing all numerical functions F which satisfy the functional equation

$$F(mn)F((m, n)) = F(m)F(n)f((m, n)) , \tag{1}$$

where (m, n) denotes the greatest common divisor of m and n. When $f(n) = n$ for all n, Equation (1) is satisfied by the Euler ϕ function since we have

$$\phi(mn)\phi((m, n)) = \phi(m)\phi(n)(m, n) .$$

More generally, it is known (see [1], [2]) that an infinite class of solutions of (1) is given by the formula

$$F(n) = \sum_{d|n} f(d)\mu\left(\frac{n}{d}\right)g\left(\frac{n}{d}\right),$$

where μ is the Möbius function and g is any multiplicative function, that is,

$$g(mn) = g(m)g(n) \quad \text{whenever } (m, n) = 1 .$$

Some work on a special case of this problem has been done by P. Comment [2]. In the case $f(1) = 1$ he has investigated those solutions F of (1) which have $F(1) \neq 0$ and which satisfy an additional condition which he calls "property O": If there exists a prime p_0 such that $F(p_0) = 0$ then $F(p_0^\alpha) = 0$ for all $\alpha > 1$. Comment's principal theorem states that F is a solution of (1) with property O and with $F(1) \neq 0$ if, and only if, F satisfies the two equations

$$F(mn)F(1) = F(m)F(n) \quad \text{whenever } (m, n) = 1$$

and

Received May 28, 1963.

$$F(p^\alpha) = F(p)f(p)^{\alpha-1} \quad \text{for all primes } p \text{ and all } \alpha \geqq 1 .$$

In this paper we study the problem in its fullest generality. In the case of greatest interest, $F(1) \neq 0$, we obtain a complete classification of all solutions of (1).

2. The solutions of (1) with $f(1) = 0$. If the given f has $f(1) = 0$ then f is identically zero and Equation (1) reduces to

$$F(mn)F((m, n)) = 0 \tag{2}$$

for all m, n. To characterize the solutions of (2) we introduce the following concept.

DEFINITION 1. A (finite or infinite) set $A = \{a_1, a_2, a_3, \cdots\}$ of positive integers is said to have property P if no a_i is divisible by any a_j^2.

Two simple examples of sets with property P are the set of primes and the set of products of distinct primes. The solutions of (2) may now be characterized as follows:

THEOREM 1. *A numerical function F satisfies* (2) *if, and only if, there exists a set A with property P such that $F(n) = 0$ whenever $n \notin A$.*

Proof. Let $A = \{a_1, a_2, a_3, \cdots\}$ be a set with property P. Define $F(a_1), F(a_2), F(a_3), \cdots$, in an arbitrary fashion and define $F(n) = 0$ if $n \notin A$. We shall prove that F satisfies (2).

Choose two integers m and n and let $d = (m, n)$. If $d \notin A$ then $F(d) = 0$ and (2) holds. If $d \in A$ then $mn \notin A$ since $d^2 \mid mn$. In this case we have $F(mn) = 0$ and again (2) holds. Therefore F satisfies (2) in all cases.

To prove the converse, assume F satisfies (2) and let A be the set of integers n such that $F(n) \neq 0$. We shall prove that A has property P. Choose any element b in A. If b were divisible by k^2 for some k in A, say $b = qk^2$, then we could take $m = qk, n = k$ in (2) to obtain

$$F(b)F(k) = 0$$

which is impossible since both b and k are in A. Therefore A has property P and the proof of Theorem 1 is complete.

3. The solutions of (1) with $f(1) = F(1) = 1$. Since we have characterized all solutions of (1) when $f(1) = 0$ we assume from now on that $f(1) \neq 0$ which means $f(1) = 1$. We divide the discussion in two parts according as $F(1) \neq 0$ or $F(1) = 0$. In the first case we introduce $G(n) = F(n)/F(1)$ and we see that (1) is equivalent to

$$G(mn)G((m, n)) = G(m)G(n)f((m, n))$$

with $G(1) = 1$. This means that the case with $F(1) \neq 0$ reduces to the case $F(1) = 1$. In this case we make a preliminary reduction of the problem as follows.

THEOREM 2. *Assume $f(1) = 1$. A numerical function F satisfies (1) with $F(1) = 1$ if, and only if, F is multiplicative and satisfies the equation*

$$F(p^{a+b})F(p^b) = F(p^a)F(p^b)f(p^b) \tag{3}$$

for all primes p and all integers $a \geqq b \geqq 1$.

Proof. Assume F satisfies (1). Taking coprime m and n in (1) we find $F(mn) = F(m)F(n)$, so F is multiplicative. Taking $m = p^a$, $n = p^b$ in (1) we obtain (3).

To prove the converse, assume F is a multiplicative function satisfying (3) for primes p and $a \geqq b \geqq 1$. Choose two positive integers m and n. If $(m, n) = 1$, Equation (1) is satisfied because it simply states that F is multiplicative. Therefore, assume $(m, n) = d > 1$ and use the prime-power factorizations

$$m = \prod_{i=1}^{\infty} p_i^{a_i}\,, \qquad n = \prod_{i=1}^{\infty} p_i^{b_i}\,, \qquad d = \prod_{i=1}^{\infty} p_i^{c_i}$$

where $a_i \geqq 0$, $b_i \geqq 0$, $c_i = \min(a_i, b_i)$, the products being extended over all primes. Since F is multiplicative we have

$$\begin{aligned} F(mn)F(d) &= \prod_{i=1}^{\infty} F(p_i^{a_i+b_i})F(p_i^{c_i}) \\ &= \prod_{0\leqq b_i\leqq a_i} F(p_i^{a_i+b_i})F(p_i^{b_i})\cdot \prod_{0\leqq a_i<b_i} F(p_i^{a_i+b_i})F(p_i^{a_i})\,. \end{aligned}$$

The factors corresponding to $b_i = 0$ or $a_i = 0$ are

$$\prod_{0=b_i\leqq a_i} F(p_i^{a_i})\cdot \prod_{0=a_i<b_i} F(p_i^{b_i}) = \prod_{a_i b_i=0} F(p_i^{a_i})F(p_i^{b_i})f(p_i^{c_i})$$

since $F(1) = f(1) = 1$. For the remaining factors we apply (3) to each product and we obtain

$$\begin{aligned} F(mn)F(d) &= \prod_{0\leqq b_i\leqq a_i} F(p_i^{a_i})F(p_i^{b_i})f(p_i^{b_i})\cdot \prod_{0\leqq a_i<b_i} F(p_i^{a_i})F(p_i^{b_i})f(p_i^{a_i}) \\ &= \prod_{i=1}^{\infty} F(p_i^{a_i})F(p_i^{b_i})f(p_i^{c_i}) = F(m)F(n)f(d)\,. \end{aligned}$$

This completes the proof of Theorem 2.

We turn now to the problem of finding all solutions of (3). If p is a prime for which $f(p) = 0$, then for this prime (3) becomes

$$F(p^{a+b})F(p^b) = 0 \quad \text{whenever } a \geqq b \geqq 1 . \tag{4}$$

For a fixed p the solutions of (4) may be characterized as follows:

THEOREM 3. *An arithmetical function F satisfies* (4) *for a given prime p if, and only if, there exists an integer $c \geqq 1$ such that*

$$F(p^i) = 0 \quad \textit{for}^1 \ 1 \leqq i \leqq c - 1 \quad \textit{and for } i \geqq 2c . \tag{5}$$

Proof. Assume F satisfies (5) for some $c \geqq 1$. Choose two integers a and b with $a \geqq b \geqq 1$. If $b \leqq c - 1$ then (5) implies $F(p^b) = 0$ so (4) is satisfied. If $b \geqq c$ then $a + b \geqq 2b \geqq 2c$ so $F(p^{a+b}) = 0$ and (4) is again satisfied.

To prove the converse, assume F is an arithmetical function satisfying (4) for some prime p. If $F(p^t) = 0$ for all integers $t \geqq 1$ then (5) holds with $c = 1$. Otherwise, we let c be the smallest $t \geqq 1$ for which $F(p^t) \neq 0$. Then $F(p^i) = 0$ for all $i \leqq c - 1$. Now take any $i \geqq 2c$ and write $i = a + c$ where $a \geqq c$. Taking $b = c$ in (4) we find $F(p^i) = 0$ for $i \geqq 2c$. Therefore (5) is satisfied for this choice of c and the proof of Theorem 3 is complete.

We consider next those primes p for which $f(p) \neq 0$. For such p the problem of solving (3) may be reduced as follows:

THEOREM 4. *Let p be a prime for which $f(p) \neq 0$. An arithmetical function F satisfies* (3) *if, and only if, there exists an arithmetical function g* (which may depend on p) *such that*

$$F(p^a) = g(a)f(p)^a \quad \textit{for all } a \geqq 1 , \tag{6}$$

where g satisfies the functional equation

$$g(a + b)g(b) = g(a)g(b) \quad \textit{for all } a \geqq b \geqq 1 . \tag{7}$$

Proof. Assume there exists a function g satisfying (7) and let $F(p^a) = g(a)f(p)^a$. Then if $a \geqq b \geqq 1$ we have

$$F(p^{a+b})F(p^b) = g(a + b)f(p)^{a+b}g(b)f(p)^b$$

and

$$F(p^a)F(p^b)f(p^b) = g(a)f(p)^a g(b)f(p)^b f(p)^b .$$

[1] If $c = 1$ the inequality $1 \leqq i \leqq c - 1$ is vacuous; in this case it is understood that (5) is to hold for all $i \geqq 2$.

Using (7) we see that F satisfies (3).

To prove the converse, assume F satisfies (3) and let

$$g(a) = \frac{F(p^a)}{f(p)^a}$$

for $a \geqq 1$. From (3) we see at once that g satisfies (7), so the proof of Theorem 4 is complete.

Next we determine all the solutions of the functional equation (7).

THEOREM 5. *Assume g is an arithmetical function satisfying* (7). *Then there exists an integer $k \geqq 1$, a divisor d of k, and a complex number C such that*

$$(8)\quad g(n) = 0 \quad \textit{for } 1 \leqq n \leqq k-1\,, \quad \textit{and for } n \geqq k, n \not\equiv 0 \pmod d\,,$$

$$(9)\qquad g(n) = C \quad \textit{for } n \geqq k, n \equiv 0 \pmod d\,.$$

Conversely, choose any integer $k \geqq 1$, any divisor d of k, and any complex number C. For those n satisfying $n \geqq k$ and $n \equiv 0 \pmod d$ let $g(n) = C$, and let $g(n) = 0$ for all other n. Then this g satisfies (7).

Proof. Assume g satisfies (7). If g is identically zero then (8) and (9) hold with any choice of k and d and with $C = 0$. If g is not identically zero, let k be the smallest positive integer n for which $g(n) \neq 0$ and let $C = g(k)$. Then $g(n) = 0$ for $1 \leqq n \leqq k-1$. If $n \geqq 2k$ we may write $n = k + r$, $r \geqq k$, and use (7) with $a = r$, $b = k$ to obtain the periodicity relation

$$(10)\qquad g(k+r) = g(r) \quad \text{for } r \geqq k\,.$$

In particular, $g(2k) = g(k)$. Therefore, to completely determine g we need only consider $g(n)$ for n in the interval $k+1 \leqq n \leqq 2k-1$. If $g(n) = 0$ for all n in this interval then $g(n) = 0$ for all $n \not\equiv 0 \pmod k$ and (8) and (9) hold with $d = k$, $C = g(k)$. Suppose, then, that $g(n) \neq 0$ for some n in the interval $k+1 \leqq n \leqq 2k-1$ and let $k+d$ be the smallest such n. Then $1 \leqq d \leqq k-1$. We prove next that $d \mid k$, that $g(n) = 0$ if $n \not\equiv 0 \pmod d$, and that $g(n) = C$ if $n \equiv 0 \pmod d$.

For this purpose we define a new function h by the equation

$$h(n) = \frac{g(n+k)}{g(k)} \quad \text{for } n \geqq 0\,.$$

Then the periodicity property (10) implies

$$(11)\qquad h(n+k) = h(n) \quad \text{if } n \geqq 0\,.$$

We also have

$$h(0) = h(k) = 1, h(n) = 0 \quad \text{if } 1 \leqq n < d, h(d) \neq 0 . \tag{12}$$

Now for $n \geqq 0$ we have

$$h(n + d) = h(n + d + 2k) = \frac{g(n + d + 3k)}{g(k)} \quad \text{and} \quad h(d) = \frac{g(d + k)}{g(k)},$$

Since $n + 2k > d + k > 1$ we may use (7) with $a = n + 2k, b = d + k$, to obtain

$$\begin{aligned} h(n + d)h(d) &= \frac{g(n + d + 3k)g(d + k)}{g(k)^2} \\ &= \frac{g(n + 2k)g(d + k)}{g(k)^2} = h(n + k)h(d) = h(n)h(d) . \end{aligned}$$

Since $h(d) \neq 0$ this implies

$$h(n + d) = h(n) \quad \text{if } n \geqq 0 . \tag{13}$$

Using (13) along with (12) we find

$$h(n) = 0 \quad \text{if } n \not\equiv 0 \pmod d, h(n) = 1 \quad \text{if } n \equiv 0 \pmod d .$$

Also, $d \mid k$ since $h(k) = 1$. This implies that $g(n) = 0$ if $n \not\equiv 0 \pmod d$, and that $g(n) = g(k) = C$ if $n \equiv 0 \pmod d$.

Now we prove the converse. Given $k \geqq 1$, a divisor d of k, and a complex number C, define g as indicated in (8) and (9). We must prove that this g satisfies (7). Choose integers a and b with $a \geqq b \geqq 1$. If $a \leqq k - 1$ then $b \leqq k - 1$ and $g(a) = g(b) = 0$ so (7) is satisfied. Suppose, then, that $a \geqq k$. We consider two cases: (i) $a \not\equiv 0 \pmod d$, and (ii) $a \equiv 0 \pmod d$.

If $a \not\equiv 0 \pmod d$ we have $g(a) = 0$ and the right member of (7) vanishes. If $a + b \not\equiv 0 \pmod d$ then $g(a + b) = 0$. If $a + b \equiv 0 \pmod d$ then $b \not\equiv 0 \pmod d$ and $g(b) = 0$. Therefore we always have $g(a + b)g(b) = 0$ so the left member of (7) also vanishes. This settles case (i).

In case (ii), $a \equiv 0 \pmod d$, we again consider the two alternatives $a + b \not\equiv 0 \pmod d$, $a + b \equiv 0 \pmod d$. If $a + b \not\equiv 0 \pmod d$ then $b \not\equiv 0 \pmod d$ and both sides of (7) vanish. If $a + b \equiv 0 \pmod d$ then $b \equiv 0 \pmod d$ so $g(a) = g(b) = g(a + b) = C$ and Equation (7) is satisfied. This completes the proof of Theorem 5.

Theorems 2 through 5 give us a complete classification of all solutions of (1) in the case $f(1) = F(1) = 1$.

4. The case $f(1) = 1$, $F(1) = 0$. In this case any F which satisfies (1) must also satisfy

$$F(m)F(n) = 0 \quad \text{whenever } (m, n) = 1 . \tag{14}$$

These functions may be characterized by means of sets of integers with the following property.

DEFINITION 2. A (finite or infinite) set $S = \{k_1, k_2, k_3, \cdots\}$ of positive integers will be said to have property Q if $1 < k_i < k_{i+1}$ and $(k_i, k_j) > 1$ for all i and j.

For example, the set of all multiples of a given integer $k_1 > 1$ has property Q, but there are more complicated sets with this property.

THEOREM 6. *A numerical function F satisfies* (14) *if, and only if, there exists a set S with property Q such that $F(n) = 0$ whenever $n \notin S$, and $F(n) \neq 0$ whenever $n \in S$.*

Proof. Assume F satisfies (14). Then $F(1) = 0$. If F is identically zero the theorem holds with S the empty set. If F is not identically zero there is a smallest integer $k_1 > 1$ with $F(k_1) \neq 0$. The set $\{k_1\}$ has property Q. If $F(n) = 0$ for all $n > k_1$ we may take $S = \{k_1\}$. Otherwise there exists a smallest integer $k_2 > k_1$ with $F(k_2) \neq 0$. The set $\{k_1, k_2\}$ has property Q because (14) implies $(k_1, k_2) > 1$. If $F(n) = 0$ for all $n > k_2$ we may take $S = \{k_1, k_2\}$. If $F(n) \neq 0$ for some $n > k_2$ we let k_3 be the smallest such n. Then (14) implies $(k_1, k_3) > 1$ and $(k_2, k_3) > 1$ so the set $\{k_1, k_2, k_3\}$ has property Q. Continuing in this way we obtain a set $S = \{k_1, k_2, k_3, \cdots\}$ (finite or infinite) with the properties indicated in the theorem.

To prove the converse, choose any set S with property Q, assign arbitrary nonzero values to the elements of S and let $F(n) = 0$ if $n \notin S$. To show that F satisfies (14), choose integers m and n with $(m, n) = 1$. Both m and n cannot be in S since S has property Q. Therefore at least one of m or n is not in S so at least one of $F(m)$ or $F(n)$ is zero. This completes the proof of Theorem 6.

Since Theorem 6 characterizes all solution of (14), all solutions of the more general equation (1) with $F(1) = 0$ must be found among those described in Theorem 6. For those solutions F of (14) which also satisfy (1) more can be asserted about the set S on which F does not vanish. We shall treat only the case in which f is never zero. In this case, if we write $G(n) = F(n)/f(n)$, Equation (1) is equivalent to

$$G(mn)G((m, n)) = G(m)G(n) . \tag{15}$$

In other words, if f never vanishes the problem reduces to the case in which f is identically 1. Moreover, $G(n) = 0$ if, and only if, $F(n) = 0$ so the set S on which G does not vanish is the same as that on which F does not vanish. For those G satisfying (15) with $G(1) = 0$ we shall prove:

THEOREM 7. *Let G be a solution of* (15) *with* $G(1) = 0$ *and let* $S = \{k_1, k_2, \cdots\}$ *be a set with property* Q *such that* $G(n) \neq 0$ *if, and only if,* $n \in S$. *Then S contains mn and (m, n) whenever it contains m and n. Moreover, every element in S is a multiple of k_1. If* $tk_1^a \in S$ *for some* $t \geqq 1, a \geqq 1$, *then G is constant on the subset* $\{tk_1^a, tk_1^{a+1}, tk_1^{a+2}, \cdots\}$.

Proof. If $m \in S, n \in S$, then $G(m) \neq 0$ and $G(n) \neq 0$. Therefore Equation (15) implies $G(mn) \neq 0$ and $G((m, n)) \neq 0$, so S contains mn and (m, n). Let $d = (k_i, k_1)$. Then $d \in S$ so $d = k_1$ since k_1 is the smallest member of S. Therefore each k_i in S is a multiple of k_1, as asserted.

If $tk_1^a \in S$, let $S(t) = \{tk_1^a, tk_1^{a+1}, tk_1^{a+2}, \cdots\}$. This is a subset of S. Taking $m = k_1$ and $n = tk_1^{a+r}$ in Equation (15) we find $G(tk_1^{a+r+1}) = G(tk_1^{a+r})$ so G is constant on $S(t)$.

BIBLIOGRAPHY

1. Douglas R. Anderson and T. M. Apostol, *The evaluation of Ramanujan's sum and generalizations*, Duke Math. J., **20** (1953), 211-216.

2. P. Comment, *Sur L'èquation fonctionelle* $F(nm)F((n, m)) = F(n)F(m)f((m, n))$, Bull. Res. Council of Israel, Sect. F7F (1957/58) pp. 14-20.

CALIFORNIA INSTITUTE OF TECHNOLOGY AND
UNIVERSITY OF WASHINGTON

IRREDUCIBLE GROUPS OF AUTOMORPHISMS OF ABELIAN GROUPS

REINHOLD BAER

The group Γ of automorphisms of the abelian group A is termed irreducible, if 0 and A are the only Γ-admissible subgroups of A. It is our aim to investigate the influence of the structure of the abstract group Γ upon the structure of the pair A, Γ. In this respect we succeed in proving the following results:

If Γ is locally finite, then A is an elementary abelian p-group and the centralizer Δ of Γ within the ring of endomorphisms of A is a commutative, absolutely algebraic field of characteristic p. If we impose the stronger hypothesis that Γ possesses an abelian torsion subgroup of finite index, then the rank of [the vector space] A over Δ is finite and Γ is a group of finite rank. If we add the further hypothesis that the orders of the elements in Γ are bounded, then A and Γ are finite.

NOTATIONS

Locally finite group	= group whose finitely generated subgroups are finite.
Almost abelian group	= group possessing abelian subgroups of finite index
Group of finite rank	= group whose finitely generated subgroups may be generated by fewer than a fixed number of elements
m-group	= group by whose subgroups the minimum condition is satisfied.

Composition of the elements in the basic abelian group A is denoted by addition. The effect of the endomorphism σ of A upon the element a in A will usually be denoted by $a\sigma$ unless A is considered as a vector space over some field of scalars in which case the scalars may appear to the left of the vectors.

PROPOSITION. *If the irreducible group Γ of automorphisms of the abelian group $A\,[\neq 0]$ is locally finite, then*

(a) *the centralizer Δ of Γ [within the ring of endomorphisms of A] is a commutative, absolutely algebraic field of characteristic p, a prime,*

Received August 5, 1963.

(b) *A is an elementary abelian p-group and*

(c) *the ring Λ of endomorphisms, spanned by Γ, is locally finite.*

Terminological Notes. The group Γ of automorphisms of the abelian group $A\,[\neq 0]$ is *irreducible*, if no proper subgroup of A is Γ-admissible.—The ring Λ of endomorphisms, *spanned by Γ*, consists of all the endomorphisms of the form $\sum_i c_i\sigma_i$ with integral c_i and σ_i in Γ.—A group [ring] is *locally finite*, if its finite subsets are contained in finite subgroups [subrings].

Proof. It is an immediate consequence of Schur's Lemma—see e.g. Jacobson [p. 26, Theorem 2]—that

(1) the centralizer Δ of Γ is a [not necessarily commutative] field.

If t is an element in A, then $t\Lambda$ is a Γ-admissible subgroup of A, since the ring Λ of endomorphisms is spanned by Γ. Application of the irreducibility of Γ shows that

(2) $t\Lambda = A$ for every $t \neq 0$ in A.

Consider now some element σ in Δ and some element $t \neq 0$ in A. From (2) we deduce the existence of an element σ' in Λ [depending on σ and t] such that $t\sigma = t\sigma'$. Since Λ is centralized by Δ, we have $\sigma\sigma' = \sigma'\sigma$; and it follows by complete induction that

$$t\sigma^i = t\sigma'^i \text{ for every positive } i\,.$$

Since Λ is spanned by Γ, there exist [finitely many] automorphisms σ_i in Γ and integers c_i such that $\sigma' = \sum_{i=1}^n c_i\sigma_i$. The subgroup θ of Γ, generated by $\sigma_1, \cdots, \sigma_n$, is finite, since Γ is locally finite. The subgroup $S = \{t\theta\}$ of A is consequently finitely generated. By construction S contains all the elements $t\sigma'^j = t\sigma^j$; and the subgroup generated by them is as a subgroup of a finitely generated abelian group likewise finitely generated. Thus we have shown:

(3) If t is an element in A and σ an element in Δ, then the subgroup $\{t, t\sigma, \cdots, t\sigma^i, t\sigma^{i+1}, \cdots\}$ of A is finitely generated.

Assume by way of contradiction that A is torsionfree. Then [multiplication by] 2 is an element, not 0, in the field Δ [by (1)] so that multiplication by 2^{-1} is likewise an automorphism of A [which belongs to Δ]. Application of (3) shows that for every $t \neq 0$ in A the subgroup

$$T = \{t, 2^{-1}t, \cdots, 2^{-i}t, \cdots\}$$

is finitely generated and consequently a free abelian group, not 0, of finite rank. But such a group is not closed under multiplication by 2^{-1}. This is a contradiction showing that A is not torsionfree. Consequently there exists a prime p such that A contains elements of order p. The set of elements x in A with $px = 0$ is therefore a Γ-admissible subgroup, not 0, of A; and we deduce from the irreducibility of Γ that

(4) $pA = 0$ for some prime p.

Thus (b) is proved.

If $t \neq 0$ is an element in A and $\sigma \neq 0$ an element in Δ, then $T = \{t, t\sigma, \cdots, t\sigma^i, t\sigma^{i+1}, \cdots\}$ is finitely generated by (3) and hence finite by (4). From $T\sigma \subseteqq T$ and (1) we deduce now that $\sigma \neq 0$ induces an automorphism of positive [finite] order n in T. From $t(\sigma^n - 1) = 0$ and (1) we conclude that $\sigma^n = 1$. Consequently

(5) there exists to every σ in Δ a positive integer k with $\sigma = \sigma^k$.

Because of (1) and (5) we may apply a Theorem of Jacobson [p. 217, Theorem 1] showing that

(6) the field Δ is commutative.

From (4) we conclude that p is the characteristic of Δ; and it follows from (5) that Δ is absolutely algebraic. Thus we have verified (a).

If Ξ is a finite subset of the ring Λ, spanned by Γ, then there exists a finite subset Ξ^* of Γ such that every element in Ξ has the form $\Sigma c(\sigma)\sigma$ for σ in Ξ^* and integral $c(\sigma)$. Since Γ is locally finite, Ξ^* is part of a finite subgroup θ of Γ. It is a consequence of (4) that the subring of Λ, spanned by θ, is finite. Hence Ξ is contained in a finite subring of Λ, proving (c).

Remark 1. Assume that Δ is a commutative, absolutely algebraic field of prime number characteristic p and that V is a vector space over Δ.

If firstly the rank of V over Δ is finite, then it is well known and easily verified that the ring of linear transformations of V over Δ [and its group of units] is locally finite.

We assume secondly the infinity of the rank of V over Δ. Denote by Λ the ring of all linear transformations of V over Δ and by Γ the group of units in Λ [= group of regular linear transformations in Λ]. It is obvious that Γ is not even a torsion group so that we are quite far away from local finiteness. Cp. Corollary 1 below. We are going to construct various irreducible substructures of Λ and Γ.

Denote by Λ_0 the totality of linear transformations σ in Λ with the property:

(0) the subspace of vectors v in V with $v\sigma = 0$ has finite co-rank in V.

It is clear and well known that Λ_0 is an ideal in Λ, the minimal ideal, not 0. Since Δ is a subring of Λ, we may form the sum $\Lambda_0 + \Delta$ which is easily seen to be a locally finite subring of Λ. Its group of units is likewise locally finite and it is an irreducible group of automorphisms of the abelian group V.

If we denote by I the subring of the integral multiples of 1 in Λ, then I is the prime field of characteristic p. The sum $\Lambda_0 + I$ is again a locally finite subring of Λ; and its group of units is likewise locally finite and an irreducible group of automorphisms of the abelian group V. In general, $\Lambda_0 + I$ does not contain its centralizer Δ.

The preceding constructions show that such vector spaces V over Δ always "arise" from locally finite irreducible groups of automorphisms of the abelian group V. On the other hand it is impossible to prove that an irreducible group of automorphisms of the abelian group V which is contained in Λ always contains a locally finite, irreducible group of automorphisms of V. This may be seen from the following construction:

the vector space V over Δ is the direct sum $V = \sum_\nu S(\nu)$ of subspaces $S(\nu)$ of rank 1. Their [infinite] cardinal is the rank of V over Δ. Denote by θ the group of all [regular] linear transformations σ in Γ with the properties:

$$(+)\qquad \begin{cases} S(\nu)\sigma = S(\nu) \text{ for every } \nu \text{ and} \\ \sigma \text{ induces the identity in almost all } S(\nu)\,. \end{cases}$$

If we denote by $\theta(\nu)$ the subgroup of all those σ in θ which induce the identity in every $S(\mu)$ with $\nu \neq \mu$, then it is easily seen that θ is the direct product of the $\theta(\nu)$ and that every $\theta(\nu)$ is isomorphic to the multiplicative group of Δ and hence an abelian torsion group of rank 1 without elements of order p.

The groups $\theta(\nu)$ and θ are equal to 1 if, and only if, Δ is the prime field of characteristic 2; and this possibility we exclude in the sequel.

Every infinite set possesses a torsionfree, simply transitive permutation group, as follows from the existence of torsionfree groups of any preassigned infinite cardinality. Consequently there exists a subgroup θ^* of Γ which is torsionfree and induces faithfully a simply transitive group of permutations on the set of the $S(\nu)$. Naturally θ is normalized by θ^* so that the product $\theta\theta^*$ is a subgroup of Γ.

Every torsion subgroup of $\theta\theta^*$ is a subgroup of θ so that none of the torsion subgroups of $\theta\theta^*$ is an irreducible group of automorphisms of the abelian group V.

Consider a $\theta\theta^*$-admissible subgroup $T \neq 0$ of V. Then there exists an element $t \neq 0$ in T; and this element t has the form

$$t = \sum_{\nu} t(\nu) \quad \text{with } t(\nu) \text{ in } S(\nu)$$

where almost all $t(\nu)$ are 0. But $t \neq 0$ implies the existence of some μ with $t(\mu) \neq 0$. There exists a linear transformation $\sigma \neq 1$ in $\theta(\mu)$; and the element

$$t - t\sigma = t(\mu) - t(\mu)\sigma$$

is an element, not 0, in $T \cap S(\mu)$. Consequently $[T \cap S(\mu)]\theta(\mu) = S(\mu)$ is part of T; and now it is clear that $T = T\theta\theta^* = V$. Thus $\theta\theta^*$ is in irreducible group of automorphisms, as we wanted to show.

LEMMA 1. *The following properties of the group G are equivalent*:

(i) *G is locally finite.*

(ii) *If the epimorphic image H of G is not locally finite, then there exists a minimal normal subgroup of H and there exists a normal subgroup $N \neq 1$ of H such that H induces in N a locally finite group of automorphisms.*

(iii) *Every epimorphic image $H \neq 1$ of G possesses a locally finite normal subgroup $N \neq 1$.*

(iv)
(a) *If J is the intersection of all the normal subgroups X of G with locally finite G/X, then G/J is locally finite.*
(b) *If the normal subgroup X of G with locally finite G/X is itself not locally finite, then there exists a normal subgroup Y of G with $Y \subset X$ such that X/Y is locally finite or nilpotent.*
(c) *If an epimorphic image of G is not locally finite, then it possesses a minimal normal subgroup.*

Terminological Reminder. A group is *nilpotent*, if every epimorphic image, not 1, has a center, not 1.

NOTES I. If the minimum condition is satisfied by the normal subgroups of G, then every epimorphic image, not 1, of G possesses a minimal normal subgroup. Furthermore there exists among the normal subgroups X of G with locally finite G/X a minimal one, say M. If K is a normal subgroup of G with locally finite G/K, then $M \cap K$ is a normal subgroup of G and $G/(M \cap K)$ is isomorphic to a subgroup of the direct product of the locally finite groups G/M and G/K. Hence $G/(M \cap K)$ is likewise locally finite; and we deduce $M = M \cap K \subseteqq K$ from the minimality of M. It follows that M is the intersection J of all the normal subgroups X of G with locally finite G/X, showing the local finiteness of G/J. Thus we have seen that in the presence of the minimum condition for normal subgroups of G condition (iv.a) and the first half of condition (ii) may be omitted.

II. Dr. Karl Gruenberg (London) has pointed out to me that the group G is—as a consequence of our Proposition—locally finite, if the minimum condition is satisfied by the normal subgroups of G and if a finite term of the derived series of G equals 1. This fact is, by Note I, an obvious special case of Lemma 1: the equivalence of (i) and (iv). It was this suggestion of Dr. Gruenberg which led us to the present Lemma 1.

Proof. For future use in this proof we restate first two well known properties of local finiteness:

(1) An extension of a locally finite group by a locally finite group is a locally finite group.

(2) Products of locally finite normal subgroups are locally finite normal subgroups.

For the proofs of (1) and (2) see Specht [p. 141, Satz 40*].

Since epimorphic images of locally finite groups are locally finite, condition (ii) is an immediate consequence of (i). -Assume next the validity of (ii) and consider a homomorphic image $H \neq 1$ of G. We want to show then the existence of a locally finite normal subgroup, not 1, of H. This is certainly the case, if H itself is locally finite. Hence we assume next that H is not locally finite. Then there exists, by (ii), a normal subgroup $N \neq 1$ of H such that H induces in N a locally finite group of automorphisms. If N happens to be locally finite, then we have again reached our goal; and thus we assume next that N is not locally finite. There exist normal subgroups X of H with $N \cap X = 1$; and among these there exists a maximal one, say L [Maximum Principle of Set Theory]. Then $N \cap L = 1$ so that $NL/L \cong N$. Hence NL/L is a non locally finite normal subgroup of the epimorphic image H/L of H and G. Thus H/L itself is not locally finite; and a second application of condition (ii) shows the existence of a minimal normal subgroup K/L of H/L. From the maximality of L we deduce that the normal subgroup $N \cap K$ of H is different from 1. If X is a normal subgorup of H with $1 \subset X \subseteqq N \cap K$, then $X \cap L = 1$ so that $L \subset LX \subseteqq K$. From the minimality of K/L we deduce $K = LX$; and from Dedekind's modular law it follows that $N \cap K = X(N \cap K \cap L) = X$. Hence $N \cap K$ is a minimal normal subgroup of H. Denote by C the centralizer of $N \cap K$ in H. Then C is a normal subgroup of H and H/C is essentially the same as the group of automorphisms, induced in

$N \cap K$ by H. This latter group is an epimorphic image of the group of automorphisms, induced in N by H, which is locally finite. Thus H/C is locally finite. If $N \cap K \cap C = 1$, then $N \cap K$ is isomorphic to the subgroup $C(N \cap K)/C$ of the locally finite group H/C; and we have found a desired locally finite normal subgroup, not 1, of H. If $N \cap K \cap C \neq 1$, then we deduce $N \cap K = N \cap K \cap C \subseteqq C$ from the minimality of $N \cap K$; and this implies the commutativity of $N \cap K$. The minimality of the normal subgroup $N \cap K$ of H is equivalent with the irreducibility of the group of automorphisms, induced in $N \cap K$ by H. Since $N \cap K$ is abelian, and since the induced irreducible group of automorphisms is locally finite,we may apply our Proposition (b). Thus $N \cap K$ is an elementary abelian p-group; and as such it is locally finite. This completes the derivation of (iii) from (ii).

Assume next the validity of (iii). Form the product P of all the locally finite normal subgroups of G. This is by (2) a locally finite characteristic subgroup of G. If $P \neq G$, then we could deduce from (iii) the existence of a locally finite normal subgroup $Q/P \neq 1$ of G/P. Then Q is an extension of the locally finite group P by the locally finite group Q/P; and we deduce from (1) the local finiteness of Q. Hence $Q \subseteqq P \subseteqq Q$ by the construction of P so that $P = Q$ and $Q/P = 1$, a contradiction. Consequently $G = P$ is locally finite; and we have proved the equivalence of (i)–(iii).

If G is locally finite, then the intersection J, occurring in (iv.a), is equal to 1 and $G/J = G$ is locally finite; and (iv.b) is satisfied because of the absence of normal subgroups of G which are not locally finite. Likewise (iv.c) is satisfied by default because of the absence of epimorphic images which are not locally finite. Thus (iv) is a consequence of (i).

Assume finally the validity of (iv). Then G/J is locally finite by (iv.a), if J is the intersection of all normal subgroups X with locally finite G/X. Assume by way of contradiction that J is not locally finite. Then there exists by (iv.b) a normal subgroup N of G with $N \subset J$ such that J/N is locally finite or nilpotent. If we let $G^* = G/N$ and $J^* = J/N$, then J^* is a normal subgroup of G^* with locally finite G^*/J^* and J^* is locally finite or nilpotent.

Assume first that J^* is nilpotent. Since every epimorphic image of G^* is an epimorphic image of G, we deduce from (iv.c):

(+) If an epimorphic image of G^* is not locally finite, then it possesses a minimal normal subgroup.

Consider next an epimorphism σ of G^* upon some group which is not locally finite. Since G^*/J^* is locally finite, $G^{*\sigma}$ is not an epimorphic image of G^*/J^* so that $J^{*\sigma} \neq 1$. Since J^* is nilpotent, so is $J^{*\sigma}$. It follows that the center Z of $J^{*\sigma}$ is different from 1. Since Z is a characteristic subgroup of the normal subgroup $J^{*\sigma}$ of $G^{*\sigma}$, it is a normal subgroup of $G^{*\sigma}$. Since Z is centralized by $J^{*\sigma}$, the group of automorphisms, induced in Z by $G^{*\sigma}$, is an epimorphic image of $G^{*\sigma}/J^{*\sigma}$ and hence of the locally finite group G^*/J^*. Thus we have shown:

(++) If an epimorphic image of G^* is not locally finite, then it possesses a normal subgroup, not 1, in which it induces a locally finite group of automophisms.

Combining (+) and (++) we see that condition (ii) is satisfied by G^*. Hence $G^* = G/N$ is locally finite. Consequently $N \subset J \subseteqq N$ by the definition of J, a contradiction showing that J/N is locally finite. But then G/N is an extension of the locally finite group J/N by the locally finite group G/J so that G/N is by (1) locally finite. Again we obtain the impossible $N \subset J \subseteqq N$. This contradiction shows that J is locally finite. Hence G is, by (1), locally finite as an extension of the locally finite group J by the locally finite group G/J. Thus (i) is a consequence of (iv) and we have shown the equivalence of conditions (i)–(iv).

As usual we say that a group is *almost-abelian*, if it possesses an abelian subgroup of finite index. The principal more or less well known properties of almost-abelian groups that we are going to need are collected in the following

LEMMA 2. *Assume that G is an almost-abelian torsion group.*

(a) *G is locally finite and possesses an abelian normal subgroup of finite index in G.*

(b) *Every abelian normal subgroup of finite index in G is a product of finite abelian normal subgroups of G.*

(c) *G is an* **m**-*group if, and only if, the minimum condition is satisfied by the normal subgroups of G.*

Terminological Note. The group G is an ***m*-*group***, if the minimum condition is satisfied by the subgroups of G.

Proof. If A is an abelian subgroup of finite index in G, then A possesses but a finite number of conjugates in G, since the normalizer of A contains A and has therefore finite index in A. By Poincaré's Theorem the intersection B of the subgroups conjugate to A in G has finite index too and is, naturally, an abelian normal subgroup of G.— If U is a finitely generated subgroup of G, then $U/(U \cap B) \cong UB/B$ is finite. It follows that $U \cap B$ is finitely generated; cp. e.g. Baer [1; p. 396, (1.3)]. The finitely generated subgroup $U \cap B$ of the abelian torsion group B is finite as is $U/(U \cap B)$ and hence U is finite. This completes the proof of (a).

If K is an abelian normal subgroup of G with finite G/K, then G induces in K a finite group of automorphisms. Hence t^G is, for every element t in K, a finite class of conjugate elements. The finitely generated abelian torsion group $\{t^G\}$ is a finite abelian normal subgroup of G, proving (b).

Assume that the minimum condition is satisfied by the normal subgroups of G. There exists by (a) an abelian normal subgroup N of G with finite G/N. Thus G induces in N a group of automorphisms of finite order n. An immediate application of Baer [2; p. 4, Lemma 1] shows that N is an $\boldsymbol{m}$-group. Since G/N is finite and N an $\boldsymbol{m}$-group, G is an $\boldsymbol{m}$-group, proving (c).

THEOREM. *If the irreducible group Γ of automorphisms of the [non-trivial] abelian group A is an almost-abelian torsion group, then*

(A) *A is an elementary abelian p-group,*

(B) *the centralizer Δ of Γ [within the ring of endomorphisms of A] is a commutative, absolutely algebraic field of characteristic p,*

(C) *the rank of A over Δ is finite,*

(D) *Γ is of finite rank.*

Note on Terminology. The group X *is of finite rank,* if there exists a positive integer n, the rank of X, such that every finitely generated subgroup of X may be generated by n [or fewer] elements.

Note on Hypotheses. It is a consequence of Lemma 2, (a) that Γ is locally finite. But this hypothesis which sufficed for the Proposition is not sufficient under the present circumstances as may be seen from the following construction: Suppose that A is a countably infinite, elementary abelian p-group. Denote by Γ the set of all automorphisms σ of A with the property:

The subgroup of fixed elements of σ has finite index in A.

It is easy to see that this set Γ of automorphisms of A is a locally finite group and that it is an irreducible group of automorphisms of A. The centralizer Δ of Γ [within the ring of endomorphisms of A] is the prime field of characteristic p so that the rank of A over Δ is infinite. Hence (C) does not hold and (D) does not hold either.

Proof. Γ is by Lemma 2, (a) locally finite so that properties (A) and (B) are immediate consequences of the Proposition. Before effecting the general proof of (C) we treat the following

Special Case. Δ is algebraically closed.

We may consider A as a vector space over Δ and the Δ-admissible subgroups of A we may consequently term subspaces. If U is a subset of A, then we denote by $[U]$ the subspace $\sum_{u \in U} u\Delta$ of A spanned by U.

Let θ be a finite abelian normal subgroup of Γ and σ a homomorphism of θ into the multiplicative group of roots of unity in Δ. Then we term the subspace S of A a *σ-subspace* [or more precisely a θ-σ-subspace] of A, if

$$x\alpha = x\alpha^\sigma \text{ for every } x \text{ in } S \text{ and every } \alpha \text{ in } \theta .$$

There exists an element $v \neq 0$ in A and $v\theta$ is a finite subset of A, since θ is finite. Hence $[v\theta]$ is a subspace, not 0, of A which is of finite positive rank and θ-admissible. Consequently there exists among the θ-admissible subspaces of finite, positive rank one R of minimal rank. In R a finite abelian group θ^* of automorphisms [= linear transformations] is induced by θ. From the minimality of the rank of R we deduce

$$R = [r\theta] = [r\theta^*] \quad \text{for every } r \neq 0 \text{ in } R .$$

The ring Λ of endomorphisms of R which is spanned by θ^* and Δ is commutative, since θ^* and Δ are commutative and centralize each other; and 0 and R are the only Λ-admissible subgroups of R. Application of Schur's Lemma—cp. Jacobson [p. 26, Theorem 2]—shows that the centralizer Λ^* of Λ [within the ring of endomorphisms of R] is a field. From the commutativity of Λ we deduce $\Lambda \subseteqq \Lambda^*$. Hence Λ is part of some commutative field of characteristic p. All the roots of unity of this field are already contained in the algebraically closed subfield Δ. Since θ^* is finite and contained in Λ, the elements in θ^* are roots of unity and belong therefore to Δ. If σ is the homomorphism of θ which maps every element upon the automorphism it induces in R, then σ is the epimorphism of θ upon $\theta^* \subseteqq \Delta$ with $r\alpha = r\alpha^\sigma$ for every r in R and α in θ. Thus R is a σ-subspace of A; and from the minimality of the positive rank of R we deduce that the rank of R is 1. Thus we have shown:

(1) If θ is a finite abelian normal subgroup of Γ, then there exists a σ-subspace of rank 1 of A for some homomorphism σ of θ into Δ.

If σ is a homomorphism of the finite abelian normal subgroup θ of Γ into Δ, then there exist σ-subspaces of A [like 0] and *the sum $A(\sigma) = A(\theta, \sigma)$ of all the σ-subspaces of A* is again a σ-subspace of A.

If γ is an automorphism in Γ, then mapping the automorphism α in the finite abelian normal subgroup θ of Γ upon $\gamma\alpha\gamma^{-1} = \alpha^{\gamma^{-1}}$ is an automorphism of θ and mapping the automorphism α in θ upon $\alpha^{\gamma^{-1}\sigma}$ is a homomorphism $\gamma^{-1}\sigma$ of θ in Δ. Since the numbers in Δ commute with the automorphisms in Γ, we find that

$$x\gamma\alpha = x\alpha^{\gamma^{-1}}\gamma = x(\alpha^{\gamma^{-1}})^{\sigma}\gamma = x\gamma(\alpha^{\gamma^{-1}})^{\sigma} = x\gamma\alpha^{\gamma^{-1}\sigma}$$

for every x in $A(\sigma)$ and α in θ. This proves:

(2) If θ is a finite abelian normal subgroup of Γ, if σ is a homomorphism of θ into Δ and γ is an automorphism in Γ, then

$$A(\sigma)\gamma = A(\gamma^{-1}\sigma) .$$

Consider again a finite abelian normal subgroup θ of Γ and a finite set Ξ of homomorphisms σ of θ into Δ with $A(\theta, \sigma) \neq 0$. If the sum of the $A(\sigma)$ with σ in Ξ were not their direct sum, then there would exist a minimal subset Ξ' of Ξ such that the sum of the $A(\sigma)$ with σ in Ξ' is not their direct sum. It is clear that Ξ' contains at least two homomorphisms. Hence we may number the homomorphisms in Ξ' as follows: $\sigma(0), \sigma(1), \cdots, \sigma(k)$ with $0 < k$. Because of the minimality of Ξ' we have: $S = \sum_{i=1}^{k} A[\sigma(i)]$ is the direct sum of the $A[\sigma(i)]$, but $\sum_{i=0}^{k} A[\sigma(i)]$ is not. Hence

$$A[\sigma(0)] \cap S \neq 0 .$$

Consequently there exists an element $s \neq 0$ in $A[\sigma(0)] \cap S$. Clearly

$$s = \sum_{i=1}^{k} s_i \quad \text{with } s_i \text{ in } A[\sigma(i)] .$$

If some s_i were 0, then s would belong to $A[\sigma(0)] \cap \sum_{j\neq i} A[\sigma(j)]$ so that this intersection were not 0 and the sum of the $A[\sigma(j)]$ with $j \neq i$ were not their direct sum contradicting the minimality of Ξ'. Thus $s_i \neq 0$ for $i = 1, \cdots, k$. If α is an automorphism in θ, then

$$\sum_{i=1}^{k} (s_i\alpha^{\sigma(0)}) = s\alpha^{\sigma(0)} = s\alpha = \sum_{i=1}^{k} (s_i\alpha) = \sum_{i=1}^{k} (s_i\alpha^{\sigma(i)}) .$$

Since S is the direct sum of the $A[\sigma(i)]$ with $i = 1, \cdots, k$, we deduce $s_i\alpha^{\sigma(0)} = s_i\alpha^{\sigma(i)}$ for $i = 1, \cdots, k$ from this equation. Since every $s_i \neq 0$ and every α^{σ} is a number in the field Δ, we conclude $\alpha^{\sigma(0)} = \alpha^{\sigma(i)}$ for every i. Since this last equation is true for every α in θ, we have shown $\sigma(0) = \sigma(i)$ for $i = 1, \cdots, k$; and this is impossible. Thus we have shown that the sum of the $A(\sigma)$ with σ in Ξ is their direct sum.

Let now B be the sum of all the $A(\sigma) \neq 0$. Then the result of the preceding paragraph of our proof shows that B is the direct sum of the $A(\sigma)$. If γ is an automorphism in Γ, then we deduce $B = B\gamma$ from (2). Since $B \neq 0$ is therefore a Γ-admissible subspace of A, and since Γ is an irreducible group of automorphisms, we have $B = A$. If finally n is the exponent of θ [so that $\theta^n = 1$], then θ is mapped by every homomorphism σ of θ into Δ into the finite group of nth roots of unity in the field Δ. Hence there exists only a finite number of homomorphisms of θ into Δ. We summarize these results as follows:

(3) If θ is a finite abelian normal subgroup of Γ, then A is the direct sum of the finitely many $A(\theta, \sigma) \neq 0$ [with σ a homomorphism of θ into Δ].

Next we recall that Γ is an almost-abelian torsion group. Application of Lemma 2 shows the existence of an abelian normal subgroup Λ of Γ with finite Γ/Λ; and Λ is the product of finite normal subgroups of Γ. Let $h = [\Gamma : \Lambda]$.

Consider next a finite abelian normal subgroup θ of Γ with $\theta \subseteqq \Lambda$. If λ is an automorphism in Λ, then λ induces the identity automorphism in θ [since Λ is abelian]. If σ is a homomorphism of θ into Δ, then $\sigma = \lambda\sigma$; and consequently there exist at most h distinct homomorphisms of the form $\gamma\sigma$ for γ in Γ. Assume now that the homomorphism σ of θ into Δ has the additional property $A(\sigma) \neq 0$. Denote the distinct homomorphisms of the form $\gamma\sigma$ for γ in Γ by $\sigma(1), \cdots, \sigma(k)$. Then $k \leqq h$. The subspace $S = \sum_{i=1}^{k} A[\sigma(i)]$ is different from 0; and S is Γ-admissible because of (2). Application of the irreducibility of Γ shows $S = A$. From (3) we deduce now that S is the direct sum of the $A[\sigma(i)]$ and that to every homomorphism σ' of θ into Δ with $A(\sigma') \neq 0$ there exists an i with $\sigma' = \sigma(i)$. If we say now that σ is a relevant homomorphism of θ, if σ is a homomorphism of θ into Δ with $A(\theta, \sigma) \neq 0$, then we may express our results as follows:

(4) If the finite abelian normal subgroup θ of Γ is part of Λ, then there exist at most h relevant homomorphisms σ of θ; and if σ', σ'' are relevant homomorphisms, then there exists an automorphism γ in Γ with $\sigma' = \gamma\sigma''$.

Because of (4) there exists among the finite abelian normal subgroups of Γ which are contained in Λ one, say Λ^*, with a maximum number of relevant homomorphisms.

Suppose now that θ is a finite abelian normal subgroup of Γ with $\Lambda^* \subseteqq \theta \subseteqq \Lambda$. If σ is a relevant homomorphism of θ and σ^* is the restriction of σ to Λ^*, then

$$x\alpha^{\sigma^*} = x\alpha^{\sigma} = x\alpha \text{ for } x \text{ in } A(\theta, \sigma) \text{ and } \alpha \text{ in } \Lambda^* .$$

It follows that $A(\theta, \sigma) \subseteqq A(\Lambda^*, \sigma^*)$. Hence σ^* is a relevant homomorphism of Λ^*. But A is by (3) both the direct sum of all the $A(\theta, \sigma)$

with relevant σ and all the $A(\Lambda^*, \tau)$ with relevant τ. This implies in particular that $A(\Lambda^*, \tau)$ is for every relevant τ the direct sum of all the $A(\theta, \sigma)$ with relevant σ such that $\sigma^* = \tau$, since we have $A(\theta, \sigma) \subseteqq A(\Lambda^*, \sigma^*) = A(\Lambda^*, \tau)$. Hence the mapping $\sigma \rightarrow \sigma^*$ is a single valued mapping of the set of all the relevant homomorphisms of θ upon the full set of all the relevant homomorphisms of Λ^*. Because of the maximality of the number of relevant homomorphisms of Λ^* it follows that this mapping is actually one-to-one. This implies in particular that $A(\theta, \sigma) = A(\Lambda^*, \sigma^*)$ for every relevant homomorphism σ of θ. Thus we have shown the following facts:

(5) If σ is a relevant homomorphism of Λ^*, if θ is a finite abelian normal subgroup of Γ with $\Lambda^* \subseteqq \theta \subseteqq \Lambda$, then there exists one and only one relevant homomorphism σ' of θ with

$$A(\Lambda^*, \sigma) = A(\theta, \sigma') \text{ and}$$
$$x\alpha^{\sigma} = x\alpha^{\sigma'} = x\alpha \text{ for every } x \text{ in } A(\Lambda^*, \sigma) \text{ and } \alpha \text{ in } \Lambda^* .$$

By (1) there exists at least one relevant homomorphism σ of Λ^*. Let $S = A(\Lambda^*, \sigma)$. If λ is an automorphism in Λ, then there exists a finite abelian normal subgroup θ of Γ which contains Λ^* and λ and which in turn is contained in Λ, since Λ is the product of finite abelian normal subgroups of Γ. By (5) there exists one and only one relevant homomorphism σ' of θ with $S = A(\theta, \sigma')$. It follows in particular that $S = S\lambda$ and that the automorphism induced in S by λ is just the multiplication by the number $\lambda^{\sigma'}$ in Δ. Thus we have shown:

(6) $S = S\Lambda$ and the automorphisms induced in S by elements in Λ are multiplications by numbers in Δ.

If $t \neq 0$ is some element in S, then $T = t\Delta$ is a subspace of rank 1. By (6) we have $T = T\Lambda \subseteqq S$. The automorphisms in Γ map T upon subspaces of rank 1; and because of $T = T\Lambda$ and $[\Gamma : \Lambda] = h$ the number of these subspaces is finite and does not exceed h. The sum of these subspaces $T\gamma$ for γ in Γ is different from 0 and it is Γ-admissible. Because of the irreducibility of Γ it is A; and thus we have shown that A is the sum of finitely many subspaces of rank 1. Since their number does not exceed h, we have shown:

(7) the rank of A over Δ does not exceed h.
By (7) we have proven (C) in the special case.

Reduction of the general case to the special case. We note first that Δ is by (B) a commutative, absolutely algebraic field of characteristic p. Let Δ^* be the algebraic closure of Δ. Then there exists a vector space V over Δ^* with the following properties:

(i) A is a subgroup of V and the subfield Δ of Δ^* operates on A in the preassigned way.

(ii) The Δ^*-vector space V is spanned by its subset A.

(iii) There exists a group Γ^* of Δ^*-automorphisms of V such that A is Γ^*-admissible and mapping every automorphism in Γ^* upon its restriction on A effects an isomorphism of Γ^* upon Γ.

Denote by Γ^{**} the group of automorphisms of the additive group V, generated by Γ^* and the multiplications by elements, not 0, in Δ^*. The isomorphic groups Γ and Γ^* are almost-abelian torsion groups. Since Δ^* is an absolutely algebraic field of characteristic p, its multiplicative group is an abelian torsion group. Since Γ^* and Δ^* centralize each other, Γ^{**} is an almost-abelian torsion group.

We don't claim that Γ^{**} is an irreducible group of automorphisms of V. Hence we consider the Γ^{**}-admissible subgroups X of V [these are exactly the Γ^*-admissible Δ^*-subspaces of V] which satisfy $A \cap X = 0$. There exist such subgroups X as for instance $X = 0$. Application of the maximum principle of set theory shows that among these subgroups X there exists a maximal one, say M. We may form the Δ^*-vector space V/M and Γ^{**} induces on V/M a group of linear transformations Ξ which, as an epimorphic image of Γ^{**}, is an almost-abelian torsion group.

Suppose that S/M is a Ξ-admissible subspace, not 0, of V/M. Then S is a Γ^{**}-admissible subspace of V with $M \subset S$. Because of the maximality of M we have $S \cap A \neq 0$. Because of the irreducibility of Γ the subgroup of A which is spanned by the Γ-admissible subset $(S \cap A)\Gamma$ is A. It follows from (i) to (iii) that V is spanned by the subset $(S \cap A)\Gamma^{**}$. Hence $S = V$, proving that Ξ is an irreducible group of automorphisms of V/M. Its centralizer contains Δ^*. Since Δ^* is algebraically closed, and since the centralizer of the irreducible, almost-abelian torsion group Ξ of automorphisms is, by Lemma 2, (a) and the Proposition, absolutely algebraic, Δ^* is exactly the centralizer of Ξ [within the ring of endomorphisms of V/M]. Application of the Special Case shows that the rank of V/M over Δ^* is finite. From $M \cap A = 0$ and (i)-(iii) we deduce that the ranks of A over Δ and of V/M over Δ^* are equal. Hence the rank of A over Δ is finite, proving (C).

Since Δ and Λ centralize each other, we may form the product Φ of the abelian group Λ and the multiplicative subgroup of the, by (B), commutative field Δ. It is clear that Φ is an abelian group of automorphisms of A. The group Λ is a torsion group as a subgroup of the torsion group Γ. The field Δ is by (B) an absolutely algebraic field of prime characteristic so that its multiplicative group is a group of roots of unity and hence a torsion group. The abelian group ϕ is therefore a product of two torsion subgroups and hence is a torsion group. Next we note that a subgroup of A is ϕ-admissible if, and only if, it is a Λ-admissible Δ- subspace of A.

Since the rank of A over Δ is, by (C), finite, and since ϕ-admissible subgroups of A are subspaces of A, there exists among the ϕ-admissible subspaces, not 0, of A one D of minimal positive rank. The abelian group ϕ of automorphisms of A induces in D an abelian group ϕ^* of automorphisms. If $X \neq 0$ is a ϕ^*-admissible subgroup of D, then X is ϕ-admissible and hence a subspace of A. Because of the minimality of the rank of D, the subspaces X and D have the same rank; and this implies $X = D$ so that ϕ^* is an irreducible group of automorphisms of D. It is a consequence of Schur's Lemma—cp. Jacobson [p. 26, Theorem 2]—that the centralizer ϕ^{**} of ϕ^* within the ring of endomorphisms of D is a [not necessarily commutative] field. But ϕ^* is a subgroup of the center of ϕ^{**} which is a commutative field. Furthermore ϕ^* is a torsion group as an epimorphic image of ϕ. Thus we see that ϕ^* is a subgroup of the group of roots of unity of a commutative field and as such ϕ^* is an abelian torsion group of rank 1.

Since $D \neq 0$ is ϕ-admissible, D is a Λ-admissible subspace of A. Since Γ/Λ is finite, $D\Gamma$ is a finite set of subspaces of A. The sum $T = \sum_{\sigma \in \Gamma} D\sigma$ of these finitely many subspaces is a Γ-admissible subspace of A; and we deduce $T = A$ from the irreducibility of Γ. Denote by $D_1, \cdots, D_f$ the finitely many distinct subspaces of the form $D\sigma$ with σ in Γ; and denote by Λ_i the totality of automorphisms in Λ which induce the identity automorphism in D_i. If [as we may assume without loss in generality] $D = D_1$, then Λ/Λ_1 is essentially the same as a subgroup of ϕ^* showing that Λ/Λ_1 is of rank 1. If $D_i = D_1\sigma$, then $\Lambda_i = \sigma^{-1}\Lambda_1\sigma$ proving the isomorphy $\Lambda/\Lambda_i \simeq \Lambda/\Lambda_1$. Hence all the Λ/Λ_i are abelian torsion groups of rank 1. An automorphism in Λ belongs to $\Lambda_1 \cap \cdots \cap \Lambda_f$ if, and only if, it induces the identity automorphism in all the $D\sigma$ with σ in Γ and hence in their sum $T = A$. Thus $\Lambda_1 \cap \cdots \cap \Lambda_f = 1$; and this shows that Λ is isomorphic to a subgroup of the direct product of the f isomorphic groups Λ/Λ_i of rank 1. Hence Λ is an abelian group of rank not exceeding f. But Γ/Λ is finite [and Γ is locally finite by Lemma 2, (a)] proving that Γ too is of finite rank.

REMARK 2. The reader should notice that throughout the proof of Properties (C) and (D) we have not fully used the requirement that Γ be an irreducible group of automorphisms. All we used is the fact that Γ is an irreducible group of linear transformations of A over Δ.

REMARK 3. It is impossible to prove (D) in the stronger form:

$$\Gamma \text{ is an } \boldsymbol{m}\text{-group} .$$

For let K be the algebraically closed, absolutely algebraic field of characteristic a prime p. If we denote by A the additive group of K, then every element, not 0, in K induces in A an automorphism by multiplication. The group Λ of these automorphisms of A is clearly irreducible and an abelian torsion group of rank 1. But it is not an $\boldsymbol{m}$-group, since it contains elements of every positive order prime to p. For a similar construction see Duguid-McLain.

If Γ is a group of automorphisms of the abelian group A, then a ring Λ of endomorphisms of A is spanned by Γ. It is clear that Γ and Λ are centralized by the same automorphisms of A and that a subgroup of A is Γ-admissible if, and only if, it is Λ-admissible. In particular Γ is an irreducible group of automorphisms of A if, and only if, Λ is an irreducible ring of endomorphisms of A. Because of this relation between Γ and Λ it is possible to express our results more forcefully within the framework of the theory of endomorphism rings.

The following lemma puts a number of well known results in a form convenient for our applications.

LEMMA 3. *The following properties of the vector space V of rank > 1 over the* [*not necessarily commutative*] *field Δ are equivalent*:

(a) *The rank of V over Δ is finite.*

(b) *If Δ is the centralizer of the irreducible ring Λ of endomorphisms of the abelian group V, then Λ is the centralizer of Δ.*

(c) *There exists a finite irreducible group of linear transformations of the vector space V over Δ.*

Notational Remark. The set Ξ of endomorphisms of the abelian group A is *irreducible*, if 0 and A are the only Ξ-admissible subgroups of A; and the set Λ of linear transformations of the vector space V over Δ is *irreducible*, if 0 and V are the only Λ-admissible subspaces of V.—Groups of linear transformations consist of automorphisms.

Proof. Assume first that the rank of V over Δ be finite. If Δ is the centralizer of the irreducible ring Λ of endomorphisms of the abelian group V, then we apply the Density Theorem of Jacobson [p. 28] to see that Λ is the ring of all linear transformations of the vector space V over Δ. Hence (b) is a consequence of (a).

Assume next that the rank of V over Δ be infinite. Denote by Λ the ring of all the linear transformations σ of the vector space V over Δ with the property:

$$\text{The rank of the subspace } V\sigma \text{ of } V \text{ is finite}.$$

If S is a subspace of V of finite co-rank, then the transformations in Λ annihilating S induce in the vector space V/S over Δ the ring of all linear transformations. This implies in particular that Δ is the centralizer of Λ. If $v \neq 0$ is an element in V, then $v\Lambda = V$. Since the rank of V over Δ is infinite, Λ is not the ring of all linear transformations of the vector space V over Δ. Thus (b) is not satisfied by V and Δ; and this shows that (a) is a consequence of (b), proving the equivalence of (a) and (b).

Assume again that the rank n of V over Δ be finite. Then there exists a finite basis B of V over Δ. We distinguish two cases.

Case 1. The characteristic of Δ is not 2.

Denote by θ the set of all the linear transformations σ of V over Δ with the property:

σ *is an automorphism of* V *which maps every element in* B *upon an element of the form* $\pm b$ *for* b *in* B.

It is clear that θ is a group whose order is $2^n(n!)$. The ring of endomorphisms of V which is spanned by θ contains for every pair of elements x, y in B one and only one linear transformation λ of V over Δ such that $x\lambda = y$ and $b\lambda = 0$ for $b \neq x$ in B. These form a soc. system of matrix units over Δ, showing that θ is a finite irreducible group of linear transformations of the vector space V over Δ.

Case 2. The characteristic of Δ is 2.

Then B generates a subgroup C of the abelian group V whose order is exactly 2^n. Denote by θ the set of all the linear transformations σ of the vector space V over Δ with the property:

$$\sigma \text{ induces an automorphism in the abelian group } C\,.$$

Every σ in θ is an automorphism of V and θ is a group of automorphisms of V which is essentially the same as the group of all the automorphisms of C. Since C is finite, so is θ. The ring of endomorphisms of V which is spanned by θ contains for every pair x, y of elements in B one and only one linear transformation of V over Δ which maps x onto y and all the other elements in B upon 0, since θ is transitive on the elements, not 0, in C. These linear transformations form a soc. system of matrix units over Δ, showing again that θ is a finite irreducible group of linear transformations of the vector space V over Δ. Thus we have shown in both cases that (c) is a consequence of (a).

Assume finally the validity of (c). Then there exists a finite irreducible group θ of linear transformations of the vector space V over Δ. If $v \neq 0$ is an element in V, then $v\theta$ is a finite subset of V which spans a subspace $S \neq 0$ of the vector space V over Δ. Since

S is θ-admissible and θ is irreducible, we have $S = V$; and from the finiteness of $v\theta$ we deduce the finiteness of the rank of V. This completes the proof.

COROLLARY 1. *The following properties of the irreducible ring Λ of endomorphisms of the abelian group A and of the centralizer Δ of Λ [within the ring of endomorphisms of A] are equivalent provided the rank of A over Δ is >1:*

(i) *Λ is spanned by an almost abelian torsion group of automorphisms of A.*

(ii) *Δ is a commutative, absolutely algebraic field of prime number characteristic p [so that $pA = 0$] and the rank of A over Δ is finite.*

(iii) *The rank of A over Δ is finite and the group of automorphisms in Λ is locally finite.*

(iv) *Λ is the centralizer of Δ and the group of automorphisms in Λ is locally finite.*

Proof. It is the content of our Theorem that (ii) is a consequence of (i).—If (ii) is true, then we note that Λ is a ring of linear transformations of the vector space V over Δ; and it is well known [and easily verified] that because of (ii) finite subsets of Λ span finite subrings of Λ; cp. Proposition (c). Thus (iii) is a consequence of (ii).

Assume the validity of (iii). Because of the irreducibility of Λ we may deduce from Schur's Lemma—see Jacobson [p. 26, Theorem 2]—that Δ is a [not necessarily commutative] field. Since the rank of A over Δ is finite, we may apply Lemma 3. Consequently

(+) Λ is the ring of all linear transformations of the vector space A over the field Δ; and

(+ +) there exists a finite irreducible group θ of linear transformations of the vector space A over the field Δ.

Denote by Γ the group of all the automorphisms in Λ. Then we deduce from (+) easily [because of the finiteness of the rank] that Λ is spanned by Γ and that Γ is consequently an irreducible group of automorphisms of A. Since Γ is locally finite [by (iii)], it follows from our Proposition that Δ is a commutative field. Application of (+) shows that Δ is the center of Λ. Another application of (+) shows that the group θ [of (+ +)] is contained in Λ. If Δ^* is the multiplicative group of all the elements, not 0, in Δ, then $\Delta^*\theta$ is a subgroup of Γ which is clearly an almost abelian torsion group of automorphisms of A. The irreducibility of this group of automorphisms of A is a consequence of (+ +); and thus we have deduced (i) from (iii).

If the equivalent conditions (i)–(iii) are satisfied by Λ, then Λ is by Lemma 3 the centralizer of Δ so that (iv) is a consequence of (i) to (iii).—Assume conversely the validity of (iv). If the rank of A over Δ were infinite, then there would exist a linear transformation σ of the vector space A over Δ which induces a permutation of order 0 on some basis of A over Δ. Clearly σ is an automorphism of A; and σ belongs by (iv) to the group of automorphisms in Λ. This group is consequently not a torsion group; but, by (iv), it is locally finite. This contradiction shows the finiteness of the rank of A over Δ so that (iii) is a consequence of (iv), completing the proof.

REMARK 4. The reader should note that we have deduced (i) from (iii) in the following somewhat stronger form:

(i*) *Λ is spanned by a torsion group of automorphisms with finite central quotient group*.

Inspection of this proof shows that we have deduced from (iii) the following facts [which are clearly contained in the equivalent conditions (i)–(iv)]:

Δ is part and hence the center of Λ and Λ contains a finite group θ of automorphisms such that the ring Λ is spanned by Δ, θ.

It is then clear that every element in Λ has the form

$$\sum_{\sigma \in \theta} \delta(\sigma)\sigma \qquad \text{with } \delta(\sigma) \text{ in } \Delta .$$

As Λ may be considered as a vector space over its commutative subfield [its center] Δ, we may conclude now that

Λ has finite rank over its center Δ.

REMARK 5. Condition (iii) is contained in conditions (ii) and (iv); and it is an immediate consequence of Lemma 3 that (iv) may be deduced from (iii). Thus (iv) appears somewhat weaker than (iii) and (iii) does not tell anything that it is not contained in the other conditions. Observation of the proofs shows that the insertion of (iii) has been convenient for them.

REMARK 6. Suppose that Δ is a commutative, absolutely algebraic field of prime number characteristic p, that V is a vector space of finite rank over Δ and that Λ is the ring of all the linear transformations of V over Δ. Assume furthermore that Γ is a group of automorphisms in Λ spanning Λ. Then Γ and Λ are irreducible; and it is a consequence of Corollary 1 that

(*) Γ is locally finite;

and we deduce from Lemma 3 the existence of a finite irreducible group θ of linear transformations of the vector space V over Δ. Since the ring Δ is spanned by Γ, and since θ is finite, there exists a finite subset Σ of Γ such that θ is part of the subring Σ. By (*) a finite subgroup $\bar{\Sigma}$ of Γ is generated by Σ. Since Σ and $\bar{\Sigma}$ span the same subring Σ^* of Δ which contains the irreducible group θ of linear transformations, we have shown:

(**) Γ contains a finite irreducible group of linear transformations of V over Δ.

If it were known that the elements, not 0, in Δ are contained in Γ, then we would have shown that Γ contains an irreducible torsion group of automorphisms of V with finite central quotient group; and this would constitute a considerable improvement on the property (i*) of Remark 4. But we have not been able to decide whether or not Γ contains an irreducible, almost abelian torsion group of automorphisms of V.

COROLLARY 2. *The following properties of the irreducible, almost abelian torsion group Γ of automorphisms of the abelian group A are equivalent*:
(i) *A is finite.*
(ii) *A is of finite rank.*
(iii) *The orders of the elements in Γ are bounded.*
(iv) *Γ is finite.*
(v) *If the abelian torsion group θ of automorphisms of A centralizes Γ, then the orders of the elements of θ are bounded.*

Proof. It is fairly obvious that the conditions (ii)–(v) are consequences of the finiteness of A. To prove the converse we note first that as a consequence of the Theorem and of the general hypotheses of our Corollary the following properties of A, Γ are satisfied:
(1) A is an elementary abelian p-group.
(2) The centralizer Δ of Γ within the ring of endomorphisms of A is a commutative, absolutely algebraic field of characteristic p.
(3) The rank of A over Δ is finite.
(4) The rank of Γ is finite.

Elementary abelian p-groups of finite rank are finite, since the orders of its finite subgroups are bounded. Hence (i) is a consequence of (ii) and (1).

Assume next that the orders of the elements in Γ are bounded. There exists by hypothesis an abelian subgroup Σ of Γ whose index $[\Gamma : \Sigma]$ is finite. The rank of Σ is finite by (4) and Σ is a torsion group the orders of whose elements are bounded. Then Σ is the direct product of finitely many primary groups; and the primary components of Σ are finite, since they are of finite rank and the orders of their elements are bounded. Hence Σ itself is finite, implying the finiteness of Γ. Thus (iv) is a consequence of (iii).

If Γ is finite, then we deduce from (1) that every element in A is contained in a finite Γ-admissible subgroup of A. Thus the finiteness of A is a consequence of the irreducibility of Γ and we have verified the equivalence of (i)–(iv).

Assume finally the validity of (v). The elements not 0 in Δ form by (2) an abelian torsion group θ of rank 1 which by (v) is finite. Hence Δ is finite. Application of (3) shows the finiteness of A, completing the proof.

COROLLARY 3. *Assume that Δ is a commutative, absolutely algebraic field of characteristic p [a prime], that V is a vector space over Δ and that Γ is the group of all [regular] linear transformations of V over Δ. Then the rank of Γ is finite if, and only if, V is finite or the rank of V over Δ is [0 or] 1.*

Proof. It is clear that our conditions are sufficient for the finiteness of the rank of Γ; and thus we assume next that the rank of Γ is finite.

Assume first by way of contradiction that the rank of V over Δ is infinite. Then it is easy to construct a subgroup of Γ which induces a group of permutations in some preassigned basis of V over Δ and which is an infinite, elementary abelian 2-group. This subgroup is not of finite rank so that Γ itself is not of finite rank. This is a contradiction showing that

(+) the rank n of V over Δ is finite.

Assume next that $1 < n$. Then we may represent V in the form $V = S \oplus T$ where S and T are subspaces of V over Δ and where the rank of S is 2. Denote by a, b a basis of S. Consider the set θ of all the linear transformations of V over Δ with the properties:

$$a\sigma = a + \delta b \quad \text{with } \delta \text{ in } \Delta\,,$$
$$b\sigma = b \text{ and } x\sigma = x \text{ for } x \text{ in } T\,.$$

It is easily seen that θ is a subgroup of Δ which is isomorphic to the additive group of Δ. Hence θ is an elementary abelian p-group; and the rank of θ is finite if, and only if, θ is finite. But the rank of Γ is finite implying the finiteness of the rank of the subgroup θ of Γ. Consequently θ is finite; and this implies the finiteness of the field Δ. Application of (+) shows that V is finite; and thus we have shown:

$$\text{If } 1 < n, \text{ then } V \text{ is finite }.$$

But this is just the fact that we wanted to prove.

REMARK 7. One should compare Corollary 3 and the statement (D) of the Theorem.

BIBLIOGRAPHY

1. Reinhold Baer, *Nilpotent groups and their generalizations*, Trans. Amer. Math. Soc., **47** (1940), 393-434.
2. ———, *Groups with descending chain condition for normal subgroups*, Duke Math. J., **16** (1949), 1-22.
3. ———, *The hypercenter of a group*, Acta Math., **89** (1953), 165-208.
4. ———, *Gruppentheoretische Eigenschaften*, Math. Ann., **149** (1963), 181-210.
5. A. M. Duguid and D. H. McLain, *FC-nilpotent and FC-soluble groups*, Proc. Cambridge Phil. Soc., **52** (1956), 391-398.
6. N. Jacobson, *Structure of Rings*, Amer. Math. Soc. Coll. Publ. **37** (1956).
7. Wilhelm Specht, *Gruppentheorie*, Berlin-Göttingen-Heidelberg 1956.

MATHEMATISCHES SEMINAR DER
UNIVERSITÄT FRANKFURT A.M.

AN ABSTRACT POTENTIAL THEORY WITH CONTINUOUS KERNEL

H. S. BEAR

1. Introduction. In the study of complex function algebras, it is a standard technique to consider the functions as being defined on the spectrum (maximal ideal space) of the algebra. In other words, one routinely replaces a function algebra A by its Gelfand representation $\hat{A}$. Recall that the Gelfand representation of a Banach algebra is just the standard representation of any normed space A as a family of functionals on A^*. Each $x \in A$ is represented as the functional $\bar{x}$ on points $F \in A^*$ defined by $\bar{x}(F) = F(x)$. The Gelfand representation simply restricts the domain of $\bar{x}$ to the very small set consisting of those $F \in A^*$ which are multiplicative (i.e., to the homomorphisms of the algebra A). Of course this restriction is necessary if $\hat{A}$ is to be again an algebra. However, a fair amount of structure accrues to the representation by virtue of this restriction (cf. [17]).

To consider the standard example, let A be the algebra of continuous complex valued functions on the unit circle in the complex plane which have analytic extensions to the unit disc. Then the spectrum S_A is (homeomorphic to) the disc, and the representation $\bar{f}$ gives the analytic extension for each $f \in A$. Now consider the space C of all continuous real functions on the unit circle. These functions also have natural extensions, as harmonic functions, to the unit disc. It follows that the disc is embedded as a compact subset Σ of C^*, and that the harmonic extensions appear as functionals on C^* restricted to this set Σ. In this setting, the disc is not a *unique* set to which the functions extend "naturally," since the circle can be put on other Riemann surfaces on which the Dirichlet problem is solvable.

In this paper we present axioms for a subset Σ of C^*, where $C = C(\Gamma)$ for an arbitrary compact space Γ, so that the representation described above does give an effective generalization of the classical potential theory on the disc or sphere in n-space. The theory we develop in this way is quite different in intent from those developed in recent years by Bauer, Brelot, and others (cf. [1], [2], [7]). In particular, we start with assumptions which insure that a global Dirichlet problem is automatically solvable.

Our set Σ in C^* consists of positive continuous functions z on Γ weighting a given positive measure μ on Γ. That is, we restrict the canonical representation of C as functionals on C^* to a subspace of

Received May 31, 1963. This research was supported in part by the National Science Foundation, under grant NSF-GP-216.

C^* consisting of functionals of the form $z\mu$, where for each $u \in C$ we have $(z\mu)(u) = \int uzd\mu$. The functionals $z\mu$ are generalizations of the Poisson measures on the circle or sphere, and the representation $\bar{u}$ on Σ obtained for each $u \in C$ is a generalized solution of the Dirichlet problem with boundary value u. A surprising amount of the classical theory of harmonic functions on the disc or sphere turns out to depend on the purely topological assumptions we make.

2. Basic assumptions. We list here a set of assumptions and some notation which will be used throughout.

Let Γ be a compact Hausdorff space with topology $\mathscr{T}$.

Let $C = C(\Gamma)$ be the linear space of all continuous real-valued functions on Γ, with the topology $\mathscr{T}_u$ of uniform convergence. The uniform norm in C is denoted $\|u\|$.

Assume there is given a positive probability measure μ on the Baire sets of Γ. In addition, we are given a set Δ of strictly positive continuous functions z on Γ such that

$$\int z(\theta)d\mu(\theta) = 1 \tag{1}$$

for all $z \in \Delta$. The function identically one is assumed to be in Δ, and is denoted z_0: $z_0(\theta) \equiv 1$. Hence the measures $z\mu$, for $z \in \Delta$, are functionals of norm one in C^*, and include $\mu = z_0\mu$.

We want to extend the functions $u \in C$ to a compact set containing Γ, and consisting of Γ and the points represented by the continuous kernels $z \in \Delta$. We do this by representing C as a space of continuous functions on a subset of C^* consisting of evaluation functionals, and the functionals $z\mu$. Accordingly, define $\Gamma^* = \{e_\theta \colon \theta \in \Gamma\}$, where $e_\theta(u) = u(\theta)$ for all $u \in C$. Similarly, let $\Delta^* = \{z\mu \colon z \in \Delta\}$, where $z\mu(u) = \int uzd\mu$ for all $u \in C$. We let $\Sigma^* = \Gamma^* \cup \Delta^*$, and introduce the axioms below on Σ^*, Δ^*, Γ^* and μ. Unless otherwise specified, the topology in C^*, and subsets thereof, is the w^* topology, $\mathscr{T}^*$.

Axiom 1. *$\Sigma^* = \Delta^* \cup \Gamma^*$ is a compact set in C^*.*

Axiom 2. *Γ^* is the boundary of Δ^* in Σ^*.*

Axiom 3. *The mapping $z \to z\mu$ is a homeomorphism of Δ, $\mathscr{T}_u$ onto Δ^*, $\mathscr{T}^*$.*

The representation of C as functions on Σ^* is as follows: for each $u \in C$, we define $\bar{u}$ on Σ^* by

$$\bar{u}(\zeta) = \zeta(u) \qquad (\zeta \in \Sigma^*). \tag{2}$$

That is, $\bar{u}(e_\theta) = u(\theta)$ for $\theta \in \Gamma$, and $\bar{u}(z\mu) = \int uzd\mu$ for $z \in \Delta$. For simplicity we will denote the points of Δ^* as z rather than $z\mu$, and write

$$\bar{u}(z) = \int u(\theta)z(\theta)d\mu(\theta) \qquad (z \in \Delta) . \tag{3}$$

We let H denote the space of all functions $\bar{u}$ on Σ^*, for $u \in C$.

Axiom 1 merely expresses the fact that we want a compact extension of our given space Γ (or its homeomorphic image Γ^*). The second axiom makes it clear that the Silov boundary of our linear space H is in fact a bona fide topological boundary (cf. [5, p. 229], [2]). Although Axiom 3 appears to be quite strong, it turns out to be exactly the necessary assumption for a theory with jointly continuous kernel. Notice that the axioms above are satisfied in the classical case which we shall consider our model: Γ is the unit circle in the plane, μ is the normalized Lebesgue measure, and Δ is the set of Poisson kernels.

LEMMA 1. *Σ^* is Hausdorff. Γ, $\mathscr{T}$ is homeomorphic to Γ^*, $\mathscr{T}^*$. Each $\bar{u} \in H$ is continuous on Σ^*.*

Proof. The subspace Σ^* is Hausdorff since C^* is. If $\theta_n \rightarrow \theta$ in $\mathscr{T}$, then certainly $u(\theta_n) \rightarrow u(\theta)$ for all $u \in C$, or $e_{\theta_n} \rightarrow e_\theta$ in $\mathscr{T}^*$. The mapping $\theta \rightarrow e_\theta$ is therefore a continuous one-to-one mapping on a compact space to a Hausdorff space, and hence a homeomorphism. The w^* topology on C^* is by definition the weakest such that the functions $\bar{u}$ of (2) are continuous. Therefore the functions $\bar{u}$ are in particular continuous on the subset Σ^*.

LEMMA 2. *H is a uniformly closed linear subspace of $C(\Sigma^*)$ and H contains the constant functions.*

Proof. The functionals of Σ^* are all of norm one, by (1), and the restriction $H|\Gamma^*$ can be identified with C on Γ. Hence uniform convergence on Γ^* is equivalent to uniform convergence on all of Σ^*, and H is in fact isomorphic and isometric with C. The constant functions are in H since $\bar{c}(z) = c$ for all $z \in \Delta$, by assumption (1).

Our axioms are given in terms of Γ^* and Δ^* as subsets of C^* to facilitate the description of a topology on the union $\Gamma^* \cup \Delta^*$. However, the embedding $\Gamma \cup \Delta \rightarrow \Gamma^* \cup \Delta^*$ is one-to-one, as we shall show in Lemma 5. It follows that we can consider our assumptions as statements about a given compact set Γ and a distinguished subset Δ of $C(\Gamma)$. Accordingly, we will drop the stars from Γ^* and Δ^*, and regard $\Sigma = \Gamma \cup \Delta$ as the object under consideration. The points of Σ are the points θ of Γ, and the points (functions) z of Δ. The topology $\mathscr{T}^*$ on Σ coincides on Γ with the given compact topology $\mathscr{T}$, and on Δ with the uniform topology $\mathscr{T}_u$ of C relativized to Δ. We write $\bar{u}(\theta) = u(\theta)$ for $\theta \in \Gamma$, and $\bar{u}(z) = \int u(\theta)z(\theta)d\mu(\theta)$ for $z \in \Delta$.

LEMMA 3. *If* $\bar{u}(z) = 0$ *for all* $z \in \Delta$, *then* $u \equiv 0$. *If* $\bar{u}(z) \geqq 0$ *for all* $z \in \Delta$, *then* $u \geqq 0$.

Proof. Both of these statements are immediate from the facts that Δ is dense in Σ (Axiom 2), and the functions $\bar{u}$ are continuous on Σ.

LEMMA 4. *If* U *is a nonempty open set in* Γ, *then* $\mu(U) > 0$.

Proof. Assume that $\mu(U) = 0$ for some nonempty open set $U \subset \Gamma$. Let u be a function in C such that $u = 0$ outside U, and $u \not\equiv 0$. Then for every $z \in \Delta$, $\bar{u}(z) = \int uzd\mu = 0$, since $u = 0$ off U, and $\mu = 0$ on U. This contradicts Lemma 3, and proves the statement.

LEMMA 5. *The mapping* $\Gamma \cup \Delta \rightarrow \Gamma^* \cup \Delta^* = \Sigma^*$ *is one-to-one.*

Proof. The representation of a functional in C^* as a measure on Γ is of course unique. The lemma asserts that the representation of this measure in the form $z\mu$ for continuous *positive* z, or the form e_θ (unit point mass at θ), is unique. This is clearly the case if (and only if) the support of μ is all of Γ.

Since $\Sigma = \Gamma \cup \Delta$ consists of distinct functionals in C^*, H is a separating linear subspace of $C(\Sigma)$. Such a subspace has a Silov boundary in Σ; i.e., a unique minimal closed set Y in Σ such that each $\bar{u} \in H$ attains its maximum on Y. ([2], or for an elementary proof, [4]). Since each functional $\zeta \in \Sigma$ has norm one, it is clear that each $\bar{u} \in H$ attains its maximum on Γ. Moreover, $H | \Gamma = C(\Gamma)$, so Γ is a minimal closed set with this property. We have proved the following:

THEOREM 1. *The Silov boundary for* H *in* Σ *is the topological boundary* Γ *of* Δ *in* Σ.

It is of course true by definition that a maximum principle holds for the functions in H and the Silov boundary Γ. The fact that μ is supported by all of Γ, which follows from the fact that Γ is the topological boundary of Δ, allows us to sharpen the maximum principle to strict inequality. This situation also occurs in some function algebras (cf. [3], [13]).

THEOREM 2. (Strict maximum principle) *If* $\bar{u}(z) = \| \bar{u} \|$ *for some* $z \in \Delta$, *then* $\bar{u}$ *is a constant.*

Proof. Assume that $\bar{u}(z) = \| \bar{u} \| = \| u \|$, and that u is non-constant and hence $\bar{u}$ is non-constant). Let $u(\theta) = \|u\| - v(\theta)$, where $v(\theta) \geqq 0$ and v is not identically zero. Let $v(\theta) \geqq \varepsilon > 0$ for all θ in some open set $U \subset \Gamma$. Then

$$\begin{aligned} u(z) &= \int u(\theta)z(\theta)d\mu(\theta) \\ &= \int [\|u\| - v(\theta)]z(\theta)d\mu(\theta) \\ &= \|u\| - \int v(\theta)z(\theta)d\mu(\theta) \\ &\leqq \|u\| - \varepsilon\mu(U)\min z \\ &< \|u\| . \end{aligned}$$

This contradicts the assumption that $\bar{u}(z) = \|u\|$. Hence u and $\bar{u}$ are constant.

COROLLARY. *If* $\bar{u} \geqq 0$ *on* Σ *and* $\bar{u}(z) = 0$ *for some* $z \in \Delta$, *then* $\bar{u} \equiv 0$.

Proof. If $v = \|u\| - u$, then $\bar{v} = \|u\| - \bar{u}$, and $\bar{v}$ assumes its maximum, $\|u\|$, at the point $z \in \Delta$. Hence $\bar{v}$ is a constant, and $\bar{u} \equiv 0$.

THEOREM 3. *Δ is closed in C if no singleton in Γ is open and closed. In particular, Δ is closed in C if Γ is connected.*

Proof. Let $\{z_n\}$ be a sequence of distinct functions in Δ which converges uniformly to $w \in C$. We must show that $w \in \Delta$. The uniform convergence of the z_n implies that the functionals $z_n\mu$ converge in $\mathscr{T}^*$ to $w\mu$. Since Σ is compact in the w^* topolgy, $w\mu \in \Sigma$. Thus either $w \in \Delta$ and we are done, or $w\mu$ is evaluation at some $\theta_0 \in \Gamma$, for all $u \in C$. For $w\mu$ to be unit point mass at θ_0, we must have $\mu\{\theta_0\} > 0$, $w(\theta_0) > 0$, and $w = 0$ on $\Gamma \sim \{\theta_0\}$. This implies that $\{\theta_0\}$ is open, since w is continuous; $\{\theta_0\}$ is automatically closed since Γ is Hausdorff.

The following example, which gives the natural "potential theory" in one dimension, shows that the hypothesis on Γ in Theorem 3 is necessary.

EXAMPLE. Let Γ consist of the two points -1 and 1, with the discrete topology. Let μ assign mass $1/2$ to each point. We denote functions u on Γ by pairs, $u = (a, b)$, where $a = u(-1)$, $b = u(1)$. The family Δ will consist of the functions $z_x = (1 - x, 1 + x)$, for $-1 < x < 1$. The function z_0 is identically one, and for each function z_n we have

$$\int z_x d\mu = (1 - x)\frac{1}{2} + (1 + x)\frac{1}{2} = 1 .$$

The family $\Gamma^* \cup \Delta^*$ is clearly homeomorphic to the compact interval $[-1, 1]$, and Γ^* is the boundary of Δ^*. Here Δ is not closed in $C(\Gamma)$, since the function $(0, 2)$ is the uniform limit of functions $(1-x, 1+x)$ as $x \to 1$. The functions $\bar{u}$ can be represented as follows: if $u = (a, b)$, then

$$\bar{u}(z_x) = \int u z_x d\mu$$
$$= a\left(\frac{1-x}{2}\right) + b\left(\frac{1+x}{2}\right).$$

Hence the graph of $\bar{u}$ is the line joining $(-1, a)$ and $(1, b)$, and $\bar{u}(z_x)$ is the point on this line above x.

3. The harmonic functions on Δ. In this section we extend our class H to a class of functions which are "harmonic" on Δ, without necessarily being continuously extendable to all of Σ. We show that the kernels $P(z, \theta) = z(\theta)$ are harmonic in z for each fixed θ, and that they are extreme points of certain compact convex sets of harmonic functions. With one additional assumption on Δ, which holds in the classical case, we show that the set of differences of positive harmonic functions is isomorphic *and homeomorphic* with C^*.

LEMMA 6. *If $P(z, \theta) = z(\theta)$ for all $z \in \Delta$, all $\theta \in \Gamma$, then P is jointly continuous on $\Delta \times \Gamma$ with the product topology.*

Proof. The statement of the lemma holds for any family (here Δ) of continuous functions on a compact space, with the uniform topology [14, p. 224].

In connection with the above lemma, it is worth noting that the uniform topology is the *weakest* such that P is jointly continuous. Thus Axiom 3 is necessary if we are to develope a theory based on the idea of a jointly continuous kernel.

With the above definition of P, the representation (3) for functions $\bar{u} \in H$ can be written in the familiar form

$$\bar{u}(z) = \int u(\theta) P(z, \theta) d\mu(\theta) . \tag{4}$$

DEFINITION. Let $\mathscr{U}$ be the topology of uniform convergence on compact subsets of Δ (the u.c.c. topology, or compact-open topology). Let $\mathscr{H}$ denote the $\mathscr{U}$-closure of $H|\Delta$. That is, $\mathscr{H}$ is the set of all u.c.c. limits on Δ of functions in H. The functions in $\mathscr{H}$ will be called *harmonic*. The set $\mathscr{H}$ forms a locally convex real linear topological space with the topology $\mathscr{U}$, since the basic neighborhoods of zero, $\{v: \sup_K |v(z)| < \varepsilon\}$, are convex.

We interrupt our development here to point out explicitly that the family $\mathscr{H}$ just defined is the set of all harmonic functions in the classical case.

PROPOSITION. *If $\Gamma = \{z: |z| = 1\}$, Δ is the set of Poisson kernels on Γ (or the open unit disc $\{z: |z| < 1\}$, and μ is normalized Lebesgue measure on Γ, then $\mathscr{H}$ is the set of all functions on Δ which are harmonic in the classical sense.*

Proof. The proposition is simply the observation that every harmonic function on the open unit disc is the u.c.c. limit of harmonic functions continuous on the closed disc. To see this, let v be harmonic on Δ, $v + iw$ be analytic on Δ, and $\{p_n\}$ be a sequence of polynomials in z which converge u.c.c. to $v + iw$ on Δ. Then the continuous harmonic functions $\{Re p_n\}$ converge u.c.c. to v.

LEMMA 7. *Δ is locally compact, and each harmonic function is continuous on Δ.*

Proof. For each $z \in \Delta$ there are disjoint neighborhoods U and V in Σ such that $z \in U$ and $\Gamma \subset V$. Hence U^- is compact, and each point of Δ has a compact neighborhood $U^- \subset \Delta$. Since a harmonic function is a uniform limit of continuous functions on some (compact) neighborhood of each $z \in \Delta$, each $v \in \mathscr{H}$ is continuous on Δ.

LEMMA 8. *If K is a compact subset of Δ, then K is an equicontinuous family of functions on Γ. The functions in K are uniformly bounded, and uniformly bounded away from zero.*

Proof. Since $K \subset \Delta$, the hypothesis is that K is compact in the uniform topology $\mathscr{T}_u$. K is therefore a bounded set in the norm $\| \ \|$ of $\mathscr{T}_u$, which means the functions $z \in K$ are uniformly bounded on Γ. If the functions in K were not uniformly bounded away from zero, then there would be a limit point $z \in K$, since K is compact, with minimum value zero. This minimum value would be attained on the compact set Γ, which contradicts the assumption that all $z \in \Delta$ are strictly positive. The set K is equicontinuous since the uniform topology $\mathscr{T}_u$ is jointly continuous, and K is compact in $\mathscr{T}_u$ [14, p. 233].

DEFINITION. We will let H^+ denote the nonnegative functions in H, and $\mathscr{H}^+$ the closure in $\mathscr{U}$ of $H^+ | \Delta$.

THEOREM 4. (Harnack's inequality—see e.g. [8, p. 153]) *If K is* *a compact subset of Δ, then there are positive numbers m and M such that for every $v \in \mathscr{H}^+$,*

$$mv(z_0) \leqq v(z) \leqq Mv(z_0) \tag{5}$$

for all $z \in K$.

Proof. Recall that the function z_0 in (5) in identically one: $z_0(\theta) \equiv 1$. We prove that the inequality (5) holds for every $\bar{u} \in H^+$, and then the theorem follows by taking uniform limits on the compact set $K \cup \{z_0\}$.

Assume that $u \geqq 0$ on Γ, and let $\min z = \min \{z(\theta)\colon \theta \in \Gamma\}$. We have

$$\begin{aligned} \min z\bar{u}(z_0) &= \min z \int u \cdot 1 d\mu \\ &\leqq \int uzd\mu \\ &= \bar{u}(z) \\ &\leqq \| z \| \int u \cdot 1 d\mu \\ &= \| z \| \, \bar{u}(z_0) \; . \end{aligned}$$

If m is a uniform lower bound for the functions $z \in K$, and M is a uniform upper bound, then we have

$$m\bar{u}(z_0) \leqq \bar{u}(z) \leqq M\bar{u}(z_0)$$

for all $z \in K$, all $\bar{u} \in H^+$.

COROLLARY. (Harnack's second convergence theorem) *If $\{\bar{u}_n\}$ is an increasing sequence of functions in H^+, and $\{\bar{u}_n(z)\}$ is bounded for any $z \in \Delta$, then $\{\bar{u}_n\}$ converges* u.c.c. *on Δ.*

Proof. Suppose that $\bar{u}_n(z) \leqq B$ for all n, so that the positive series $\sum [\bar{u}_n(z_1) - \bar{u}_{n-1}(z_1)]$ converges. Let K be any compact set in Δ and let m and M be the constants of Theorem 4 for the set $K \cup \{z_1\}$. Then from (5) we have

$$\bar{u}_n(z_0) - \bar{u}_{n-1}(z_0) \leqq \frac{1}{m} [\bar{u}_n(z_1) - \bar{u}_{n-1}(z_1)] \, ,$$

and hence for all $z \in K$,

$$\bar{u}_n(z) - \bar{u}_{n-1}(z) \leqq \frac{M}{m} [\bar{u}_n(z_1) - \bar{u}_{n-1}(z_1)] \; .$$

Therefore the series $\sum [\bar{u}_n(z) - \bar{u}_{n-1}(z)]$ converges uniformly on K.

Since K is arbitrary, this says that $\{\bar{u}_n\}$ converges u.c.c. on Δ.

Notice that the corollary is stated for H^+, rather than $\mathscr{H}^+$. This is because it is not clear that if $v, w \in \mathscr{H}^+$, and $v - w \geqq 0$, that $v - w \in \mathscr{H}^+$, as would be required in the above proof ($\mathscr{H}^+$ is defined as the set of limits of H^+, and not as the positive functions in $\mathscr{H}$).

Now we can prove that the kernel $P(z, \theta)$ is harmonic in z for each fixed $\theta \in \Gamma$, and moreover, that each $P(\cdot, \theta) \in \mathscr{H}^+$.

THEOREM 5. *If $\theta_0 \in \Gamma$, then $P(\cdot, \theta_0) \in \mathscr{H}^+$.*

Proof. Let K be a compact subset of Δ, and $\varepsilon > 0$. We must find $\bar{u} \in H^+$ such that

$$|\bar{u}(z) - P(z, \theta_0)| < \varepsilon$$

for all $z \in K$.

Since K is an equicontinuous family, there is a neighborhood U of θ_0 in Γ such that $|z(\theta) - z(\theta_0)| < \varepsilon$ for all $z \in K$ and all $\theta \in U$. Let u be a nonnegative continuous function on Γ such that $u = 0$ off U, and $\int u d\mu = 1$. For $z \in K$ we have

$$\begin{aligned} |\bar{u}(z) - P(z, \theta_0)| &= \left|\int u(\theta)z(\theta)d\mu(\theta) - z(\theta_0)\right| \\ &= \left|\int u(\theta)[z(\theta) - z(\theta_0)]\, d\mu(\theta)\right| \\ &\leqq \sup_{\theta \in U} |z(\theta) - z(\theta_0)| \int u(\theta)d\mu(\theta) \\ &< \varepsilon . \end{aligned}$$

Since K and ε are arbitrary, and $\bar{u} \geqq 0$, $P(\cdot, \theta_0) \in \mathscr{H}^+$.

The next two theorems are extensions to our abstract setting of classical results of Herglotz [11], Bray-Evans [6], Evans [9], and Martin [15, p. 153].

THEOREM 6. *A function v on Δ is in $\mathscr{H}^+$ if and only if there is a positive Baire measure ν on Γ such that for all $z \in \Delta$,*

$$v(z) = \int P(z, \theta)d\nu(\theta) . \tag{6}$$

Proof. Assume first that v is given by (6). The integral in (6) can be approximated at any finite number of points $z \in \Delta$ by a Riemann sum of the form

$$\sum P(z, \theta_i)\nu(E_i) . \tag{7}$$

Any function of the form (7) is in $\mathscr{H}^+$ by Theorem 5. The set of functions of the form (7) is equicontinuous on Δ, since

$$\begin{aligned} |\Sigma P(z, \theta_i)\nu(E_i) - \Sigma P(z_1, \theta_i)\nu(E_i)| &\leqq \Sigma\,|z(\theta_i) - z_1(\theta_i)|\,\nu(E_i) \\ &\leqq \|z - z_1\|\,\Sigma\nu(E_i) \\ &= \|z - z_1\|\,\nu(\Gamma)\ . \end{aligned}$$

That is, any function of the form (7) will vary by less than ε on the sphere of radius $\varepsilon/\nu(\Gamma)$ around z_1. Therefore pointwise convergence of sums (7) will be uniform on any compact set $K \subset \Delta$ [14, p. 232]. Hence v is in the $\mathscr{U}$-closure of the $\mathscr{U}$-closed set $\mathscr{H}^+$; i.e., $v \in \mathscr{H}^+$.

Now assume that $v \in \mathscr{H}^+$, and let $\{\bar{u}_\alpha\}$ be a net of functions in H^+ which converges uniformly on compact sets to v:

$$\begin{aligned} v(z) &= \lim \bar{u}_\alpha(z) \\ &= \lim \int P(z, \theta)u_\alpha(\theta)d\mu(\theta)\ . \end{aligned}$$

The measures $\{u_\alpha\mu\}$ are all in some closed ball of C^*, since

$$\|u_\alpha\mu\| = \int u_\alpha(\theta)d\mu(\theta) = u_\alpha(z_0) \rightarrow v(z_0)$$

(recall that $z_0(\theta) \equiv 1$). The closed balls in C^* are $\mathscr{T}^*$ compact, so there is a subnet of $\{u_\alpha\mu\}$ which converges w^* to a positive measure ν. For this subnet, also denoted $\{u_\alpha\mu\}$, and the continuous function $P(z, \cdot)$ on Γ, we have

$$\begin{aligned} v(z) &= \lim \int P(z, \theta)u_\alpha(\theta)d\mu(\theta) \\ &= \int P(z, \theta)d\nu(\theta)\ . \end{aligned}$$

COROLLARY. *$v \in \mathscr{H}^+ - \mathscr{H}^+$ if and only if $v = \int P d\nu$ for some signed Baire measure ν.*

DEFINITION. $H_M = \left\{\bar{u} \in H: \int |u|\, d\mu \leqq M\right\}$. *Let $\mathscr{H}_M$ be the $\mathscr{U}$-closure of $H_M | \Delta$.*

The hypothesis $v \in \mathscr{H}_M$ is our replacement of the classical Fatou condition for harmonic functions in the disc: $(1/2\pi)\int_0^{2\pi} |v(re^{i\theta})|\, d\theta \leqq M$ for all $r < 1$ (Fatou [10] or [16, p. 201]). If $\Gamma = \{z: |z| = 1\}$, Δ is the set of Poisson kernels, etc., so the classical situation obtains, then $\mathscr{H}_M$ is the set of harmonic functions v such that the functions v_r are uniformly bounded by M in the L_1 norm, where $v_r(e^{i\theta}) = v(re^{i\theta})$ (see

[12, pp. 33–39]). The families $\mathscr{H}_M$ are *compact* sets of harmonic functions (Theorem 8 below), and this compactness accounts for much of their tractability.

LEMMA 9. *For each* $\theta_0 \in \Gamma$, $P(\cdot, \theta_0) \in \mathscr{H}_1$.

Proof. In the proof of Theorem 5 (that $P(\cdot, \theta_0) \in \mathscr{H}^+$) we found for a given compact $K \subset \varDelta$ a function $\bar{u} \in H^+$ such that $|\bar{u}(z) - P(z, \theta_0)| < \varepsilon$ for all $z \in K$. This function $\bar{u}$ was in H_1, since $u \geqq 0$ and $\int u d\mu = 1$. Thus $P(\cdot, \theta_0)$ is the u.c.c. limit of functions in H_1, or $P(\cdot, \theta_0) \in \mathscr{H}_1$.

For the following theorem in the classical context, see [18, p. 143] or [12, p. 33].

THEOREM 7. *A function v is in $\mathscr{H}_M$ if and only if there is a signed measure ν on Γ, with $\|\nu\| \leqq M$, such that for all $z \in \varDelta$,*

$$v(z) = \int P(z, \theta) d\nu(\theta) . \tag{7}$$

Proof. Assume that $v \in \mathscr{H}_M$, and that $\{\bar{u}_\alpha\}$ is a net of functions in H_M which converges uniformly to v on compact subsets:

$$\begin{aligned} v(z) &= \lim \bar{u}_\alpha(z) \\ &= \lim \int P(z, \theta) u_\alpha(\theta) d\mu(\theta) . \end{aligned}$$

The measures $\{u_\alpha \mu\}$ are all in the M-ball of C^*, since by hypothesis

$$\| u_\alpha \mu \| = \int | u_\alpha | d\mu \leqq M .$$

As in the proof of Theorem 6, there is a w^* accumulation point ν of $\{u_\alpha \mu\}$, and $\|\nu\| \leqq M$. If $\{u_\alpha \mu\}$ is a subnet converging w^* to ν, then

$$\begin{aligned} v(z) &= \lim \int P(z, \theta) u_\alpha(\theta) d\mu(\theta) \\ &= \int P(z, \theta) d\nu(\theta) \end{aligned}$$

for each $z \in \varDelta$.

Now assume that $\|\nu\| \leqq M$ and v is given by (7). We showed in the proof of Theorem 6 that v can be uniformly approximated on any compact $K \subset \varDelta$ by a finite sum $\Sigma P(\cdot, \theta_i)\nu(E_i)$, where $\{E_i\}$ is a partition of Γ. By Lemma 9, each $P(\cdot, \theta_i)$ occuring in this sum can be uniformly approximated (within $\varepsilon/n \|\nu\|$, if there are n summands) by a function $\bar{u}_i \in H_1 \cap H^+$. Hence there is a sum $\bar{w} = \Sigma \nu(E_i)\bar{u}_i$ which is uniformly close to v on the given compact set K. Clearly $\bar{w} \in H$, and we have

$$\begin{aligned}\int |w|\,d\mu &\leqq \sum |\nu(E_i)| \left|\int \bar{u}_i d\mu\right. \\ &= \sum |\nu(E_i)| \\ &\leqq \|\nu\| \\ &= M\,.\end{aligned}$$

Thus $\bar{w} \in H_M$, and v can be uniformly approximated on K by functions in H_M, so $v \in \mathscr{H}_M$.

COROLLARY. $\mathscr{H}^+ - \mathscr{H}^+ = \bigcup\{\mathscr{H}_M\colon M = 1, 2, \cdots\}$.

Proof. This follows from the corollary of Theorem 6.

LEMMA 10. *H_M is equicontinuous on Δ.*

Proof. If $z, z_1 \in \Delta$ and $u \in H_M$, then

$$\begin{aligned}|\bar{u}(z) - \bar{u}(z_1)| &= \left|\int u(\theta)[z(\theta) - z_1(\theta)]\,d\mu(\theta)\right| \\ &\leqq \|z - z_1\| \int |u(\theta)|\,d\mu(\theta) \\ &\leqq M\|z - z_1\|\,.\end{aligned}$$

THEOREM 8. *$\mathscr{H}_M$ is compact in the topology $\mathscr{U}$.*

Proof. The pointwise closure (and, a fortiori, u.c.c. closure) of an equicontinuous family is equicontinuous, and hence $\mathscr{H}_M$ is equicontinuous. By Ascoli's theorem [14, p. 233], the subfamily $\mathscr{H}_M$ of $C(\Delta)$ is compact in the topology $\mathscr{U}$ if and only if $\mathscr{H}_M$ is closed, $\mathscr{H}_M$ is equicontinuous, and $\{v(z)\colon v \in \mathscr{H}_M\}$ is bounded for each $z \in \Delta$. We have only the last condition to check. For each $z \in \Delta$, $\{|v(z)|\colon v \in \mathscr{H}_M\}$ is bounded by $M\|z\|$, since for $\bar{u} \in H_M'$,

$$|\bar{u}(z)| \leqq \int |u(\theta)|\, z(\theta) d\mu(\theta) \leqq M\|z\|\,,$$

and this estimate carries over to $\mathscr{H}_M$ on the compact set $\{z\}$.

COROLLARY. *$\mathscr{H}^+ - \mathscr{H}^+$ is σ-compact.*

Proof. This follows from the corollary of Theorem 7.

In the classical case of the unit ball in Euclidean space, the correspondence between functions in $\mathscr{H}^+ - \mathscr{H}^+$ and measures is one-to-one. The proof uses the specific form of the Poisson kernels [18, p. 143, 144]. The uniqueness of a representing measure ν is of course equivalent to the non-existence of a nontrivial measure orthogonal to all the functions $z \in \Delta$. Restated, the measure corresponding to a function in $\mathscr{H}^+ - \mathscr{H}^+$ is unique if the linear span of Δ is uniformly dense in C. We incorporate this hypothesis in the next theorem to make the statement explicit for the classical case.

THEOREM 9. *If the linear span of Δ is uniformly dense in C, then the isomorphism $\nu \rightarrow \int P d\nu$ is a homeomorphism of C^* onto $\mathscr{H}^+ - \mathscr{H}^+$.*

Proof. As noted above the hypothesis contains the assumption that the mapping is one-to-one. Since this isomorphism maps the compact M-ball of C^* onto the compact set $\mathscr{H}_M$, it is sufficient to show the mapping is continuous in either direction. We will show the mapping $v \rightarrow \nu$ from $\mathscr{H}^+ - \mathscr{H}^+$ to C^* is continuous. Let $v_\alpha \rightarrow v$ in the topology $\mathscr{U}$, and let ν_α, ν be the corresponding measures. Then $\int z d\nu_\alpha \rightarrow \int z d\nu$ for all $z \in \Delta$. Since the linear span of Δ is uniformly dense in C, we have $\int g d\nu_\alpha \rightarrow \int g d\nu$ for every continuous g, or $\nu_\alpha \rightarrow \nu$ in $\mathscr{T}^*$.

COROLLARY. *The extreme points of $\mathscr{H}_1$ are the functions $\pm P(\cdot, \theta)$, for $\theta \in \Gamma$.*

Proof. These are the images under the isomorphism above of the unit point masses on Γ which are the extreme points of the unit ball of C^*. (The positive extreme points are the minimal positive harmonic functions of R. S. Martin [15].)

REFERENCES

1. Heinz Bauer, *Axiomatische Behandlung des Dirichletschen Problems für elliptische und parabolische Differentialgleichungen*, Math. Annalen, **146** (1962), 1-59.

2. ———, *Silovscher Rand und Dirichletsches Problem*, Ann. Inst. Fourier, Grenoble, **11** (1961), 89-136.

3. H. S. Bear, *A strong maximum modulus theorem for maximal function algebras*, T.A.M.S., **92** (1959), 465-469.

4. ———, *The Silov boundary for a linear space of continuous functions*, Amer. Math. Mo., **68** (1961), 483-485.

5. H. F. Bohnenblust and S. Karlin, *Geometrical properties of the unit sphere of Banach algebras*, Annals of Math., **62** (1955), 217-229.

6. H. E. Bray and G. C. Evans, *A class of functions harmonic within the sphere*, Amer. J. Math., **49** (1927), 153-180.

7. M. Brelot, *Axiomatique des fonctions harmoniques et surharmoniques dans un espace*

localement compact, Sémenaire de théorie du potentiel, **2** (1957–1958), 40 pp (Inst. H. Poincare, Paris).
8. C. Carathéodory, *Theory of functions of a complex variable*, vol. 1, New York (1954).
9. G. C. Evans, *The logarithmic potential* (A. M. S. Colloquium Publications, vol. VI) New York (1927).
10. P. Fatou, *Séries trigonométriques et séries de Taylor*, Acta Math., **30** (1906).
11. G. Herglotz, *Ueber Potenzreihen mit positivem reelem Teil im Einheitskreis*, Leipz. Ber., **63** (1911).
12. Kenneth Hoffman, *Banach spaces of analytic functions*, Englewood Cliffs, N. J. (1962).
13. Kenneth Hoffman and I. M. Singer, *Maximal algebras of continuous functions*, Acta Math., **103** (1960), 217–241.
14. J. L. Kelley, *General topology*, New York (1955).
15. R. S. Martin, *Minimal positive harmonic functions*, T.A.M.S., **49** (1941), 137–172.
16. R. Nevanlinna, *Eindeutige analytische funktionen*, Berlin (1955).
17. Hugo Rossi, *The local maximum modulus principle*, Annals of Math., **72** (1960), 1–11.
18. M. Tsuji, *Potential theory in modern function theory*, Tokyo (1959).

UNIVERSITY OF CALIFORNIA, SANTA BARBARA

SUPERADDITIVITY INEQUALITIES

E. F. BECKENBACH

1. Introduction. In the theory of analytic inequalities, a principal tool is the notion of convex function [6, 1]. A hierarchy of convexity conditions, useful in this theory, can be expressed as follows: Let $K^{\alpha}(a, b)$ denote the class of functions p that are positive and continuous on an interval $a \leqq x \leqq b$ and such that sign (x) $[p(x)]^{\alpha}$ is convex on $[a, b]$ if $\alpha \neq 0$, and $\log p(x)$ is convex on $[a, b]$ if $\alpha = 0$; then for all real α and β with $\beta > \alpha$ we have $K^{\alpha}(a, b) \subset K^{\beta}(a, b)$ [8].

A different sort of hierarchy has been established by Bruckner and Ostrow [3]. In the present paper we are concerned with an illustration and some applications of this latter hierarchy. To describe it, we need a few definitions.

2. Definitions. Let $K(b)$ be the class of real-valued functions f that are continuous and nonnegative on a given closed interval $0 \leqq x \leqq b$ and vanish at the origin, $f(0) = 0$.

The *average function* F of a function $f \in K(b)$ is the function $F \in K(b)$ defined by

$$F(x) \equiv \frac{1}{x}\int_0^x f(t)dt\ , \qquad 0 < x \leqq b\ ,$$
$$F(0) = 0\ .$$

The function $f \in K(b)$, with average function F, will be said to be of *class*

$K_1(b)$ if and only if f is *convex* on $[0, b]$, i.e., if and only if for every x and $y \in [0, b]$, and for every α, $0 \leqq \alpha \leqq 1$, we have

$$(1) \qquad f[\alpha x + (1 - \alpha)y] \leqq \alpha f(x) + (1 - \alpha)f(y)\ ;$$

$K_2(b)$ if and only if $F \in K_1(b)$;

$K_3(b)$ if and only if f is *starshaped* (with respect to the origin) on $[0, b]$, i.e., if and only if for every $x \in [0, b]$, and for every α, $0 \leqq \alpha \leqq 1$, we have

$$(2) \qquad f(\alpha x) \leqq \alpha f(x)\ ;$$

$K_4(b)$ if and only if f is *superadditive* on $[0, b]$, i.e., if and only if for every x and $y \in [0, b]$ such that also $(x + y) \in [0, b]$ we have

Received August 27, 1963. The preparation of this paper was sponsored by the Office of Naval Research. Reproduction in whole or in part is permitted for any purpose of the United States Government.

$$(3) \qquad f(x+y) \geqq f(x) + f(y) ;$$

$K_5(b)$ if and only if $F \in K_3(b)$;

$K_6(b)$ if and only if $F \in K_4(b)$.

If $f \in K_2(b)$, $K_5(b)$, or $K_6(b)$, then f is said [3] to be, respectively, convex, starshaped, or superadditive *on the average* on $[0, b]$.

If $f \in K_i(b_0)$, then clearly $f \in K_i(b)$ for all positive $b < b_0$.

3. The hierarchy. The following class-inclusion implications have been established by Bruckner and Ostrow [3]:

$$K_1(b) \subset K_2(b) \subset K_3(b) \subset K_4(b) \subset K_5(b) \subset K_6(b) .$$

They have further given examples to show that none of the reverse implications are valid; i.e., they have given examples showing that

$$(4) \quad K_6(b) \not\subset K_5(b),\ K_5(b) \not\subset K_4(b),\ K_4(b) \not\subset K_3(b),\ K_3(b) \not\subset K_2(b),\ K_2(b) \not\subset K_1(b) .$$

Thus they have pointed out that the function f defined on $[0, 1]$ by

$$f(x) \equiv x^2 - x^3$$

is convex on the average on $[0, 4/9]$ but convex only on $[0, 1/3]$, that the function g defined on $[0, \infty)$ by

$$g(x) \equiv \begin{cases} x^2, & 0 \leqq x \leqq 1 , \\ x, & x > 1 , \end{cases}$$

is starshaped on $[0, b]$ for an arbitrarily large value of b but convex on the average only on $[0, 1]$, that the function h defined on $[0, \infty)$ by

$$h(x) \equiv n + (x-n)^2 , \qquad n \leqq x < n+1 , \qquad n = 0, 1, 2, \cdots ,$$

is superadditive on $[0, b]$ for an arbitrarily larger value of b but starshaped only on $[0, 1]$, etc.

It is our purpose first to use a *single* illustrative function f and its average function F to establish the fact that none of the foregoing reverse implications hold, and secondly to derive some general inequalities for convex, starshaped, and superadditive functions and to apply them to our particular illustrative functions.

4. Example. From

$$F(x) = \frac{1}{x}\int_0^x f(t)dt , \qquad 0 < x \leqq b ,$$

we obtain

$$F'(x) = \frac{1}{x}f(x) - \frac{1}{x^2}\int_0^x f(t)dt = \frac{1}{x}[f(x) - F(x)] ,$$

whence

$$f(x) = F(x) + xF'(x) . \tag{5}$$

We might call f the *inverse average function* of F.

Let us consider the function F defined on $[0, \infty)$ by

$$\begin{aligned} &F(x) \equiv e^{-1/x} , \qquad 0 < x < \infty , \\ &F(0) = 0 . \end{aligned} \tag{6}$$

Then (5) gives

$$\begin{aligned} &f(x) \equiv \left(1 + \frac{1}{x}\right)e^{-1/x} , \qquad 0 < x < \infty , \\ &f(0) = 0 . \end{aligned} \tag{7}$$

In the following Sections 5–9 we establish the *maximum* values b_i such that the function f defined by (7) is of class $K_i(b_i)$, $i = 1, 2, \cdots, 6$.

5. Convexity. A function f of class C'' is convex on an interval if and only if we have $f''(x) \geqq 0$ throughout the interval.

For the function f given by (7), a computation yields

$$f''(x) = \frac{1}{x^5}(1 - 3x)e^{-1/x} .$$

Accordingly, $f \in K_1(b)$ for

$$b = b_1 = \frac{1}{3} , \tag{8}$$

but for no larger value of b. The function is concave on the interval $[1/3, \infty)$.

Similarly, for the function F given by (6), we have

$$F''(x) = \frac{1}{x^4}(1 - 2x)e^{-1/x} .$$

Thus the maximum interval of convexity of F is $[0, 1/2]$, and F is concave on the interval $[1/2, \infty)$. Therefore $f \in K_2(b)$ for

$$b = b_2 = \frac{1}{2} , \tag{9}$$

but for no larger value of b.

A function that is convex on a left-hand portion of its interval of definition, and concave on the complementary right-hand portion, is said to be *convexo-concave* [1]. Thus both the function f given by (7) and the function F given by (6) are convexo-concave on the interval $[0, \infty)$.

6. **Starshapedness.** A function f of class C', $f \in K(b)$, is starshaped on the interval $[0, b]$, i.e., $f \in K_3(b)$, if and only if [3]

$$f'(x) \geqq \frac{f(x)}{x} \text{ for all } x \in (0, b] .$$

For the function f given by (7), we have

$$f'(x) - \frac{f(x)}{x} = \frac{1}{x^3}(1 - x - x^2)e^{-1/x} ,$$

whence it follows that $f \in K_3(b)$ for

$$b = b_3 = \frac{\sqrt{5} - 1}{2} , \tag{10}$$

but for no larger value of b.

Similarly, for the function F given by (6) we obtain

$$F'(x) - \frac{F(x)}{x} = \frac{1}{x^2}(1 - x)e^{-1/x} ,$$

so that $f \in K_5(b)$ for

$$b = b_5 = 1 , \tag{11}$$

but for no larger value of b.

Thus it happens that the maximum interval of starshapedness of the function f forms a golden section [7] of the maximum interval of starshapedness of the function F.

7. **Superadditivity.** Tests for superadditivity appear to be difficult to establish, and more difficult to apply. None are given, for example, in the treatments [5] and [9] of superadditive functions. A few tests, however, have been advanced by Bruckner [2]; see also § 14, below. One of Bruckner's tests, which we shall use in order somewhat to shorten our determination of the maximum interval of superadditivity of the function f given by (7), and of the function F given by (6), can be stated as follows:

BRUCKNER'S TEST. *Let the function $f \in K(b)$ be convexo-concave.*

Then f is superadditive on $[0, b]$, *i.e.*, $f \in K_4(b)$, *if and only if*

$$f\left(\frac{b}{2}+x\right)+f\left(\frac{b}{2}-x\right) \leqq f(b) \text{ for all } x \in\left[0, \frac{b}{2}\right].$$

In §§ 8 and 9, below, we shall prove the following results:

THEOREM 1. *The function f, defined by*

$$f(x) \equiv\left(1+\frac{1}{x}\right) e^{-1/x}, \quad 0<x<\infty,$$
$$f(0)=0,$$

is superadditive on $[0, b]$ *for*

$$0<b \leqq b^*,$$

where b^* is the unique positive solution of the transcendental equation

$$b=\frac{1-4 e^{-1/b}}{2 e^{-1/b}-1}$$

(*approx.* $b^* = 0.8955$), *but for no larger value of b.*

That is, the function $f \in K_4(b)$ for

(12) $$b=b_4=b^* \doteq 0.8955,$$

but for no larger value of b.

THEOREM 2. *The function F, defined by*

$$F(x) \equiv e^{-1/x}, \quad 0<x<\infty,$$
$$F(0)=0,$$

is superadditive on $[0, b]$ *for*

$$0<b \leqq \frac{1}{\log 2},$$

but for no larger value of b.

That is, the function $f \in K_6(b)$ for

(13) $$b=b_6=\frac{1}{\log 2},$$

but for no larger value of b.

8. **Proof of Theorem 2.** The method of proof we shall use is largely the same for both theorems. Since the formulas are simpler and the details shorter for Theorem 2, we shall treat it first and

then follow substantially the same pattern for Theorem 1.

Relative to the function F given by (6), consider the function G defined for $b \in (0, \infty)$ and $x \in [0, b/2]$ by

$$(14)\qquad \begin{aligned} &G(x; b) \equiv e^{-1/(b/2+x)} + e^{-1/(b/2-x)} - e^{-1/b}\,, \quad x \neq \frac{b}{2}\,, \\ &G\left(\frac{b}{2}; b\right) = 0\,. \end{aligned}$$

In accordance with Bruckner's test, we shall establish the maximum interval $[0, b]$ of superadditivity of the function F by determining the maximum value b such that

$$G(x; b) \leqq 0 \text{ for all } x \in \left[0, \frac{b}{2}\right].$$

In particular, for F to be superadditive on $[0, b]$, it is necessary that we have

$$(15)\qquad G(0; b) \equiv 2e^{-2/b} - e^{-1/b} \leqq 0\,,$$

or

$$\log 2 - \frac{2}{b} \leqq -\frac{1}{b}\,,$$

whence

$$b \leqq b_6 = \frac{1}{\log 2}\,.$$

Hence the function F is not superadditive on $[0, b]$ for any $b > b_6$. We shall show, however, that F is superadditive on $[0, b_6]$ (and therefore, of course, on $[0, b]$ for every positive $b < b_6$). That is, we show that

$$(16)\qquad G(x; b_6) \leqq 0 \text{ for all } x \in \left[0, \frac{b_6}{2}\right].$$

By (14), we have

$$(17)\qquad G\left(\frac{b_6}{2}; b_6\right) = 0\,,$$

and by the choice of b_6 we have also

$$(18)\qquad G(0; b_6) = 2e^{-2/b_6} - e^{-1/b_6} = 0\,.$$

We shall prove somewhat more than is needed for what is claimed in Theorem 2; namely, we shall show that we have not merely (16) but actually the *strict* inequality

$$G(x; b_6) < 0 \text{ for all } x \in \left(0, \frac{b_6}{2}\right). \tag{19}$$

In § 11, below, we shall make essential use of the fact that this is a strict inequality.

If (19) did not hold, then, by (17) and (18), $G(x; b_6)$ would attain a nonnegative maximum value at some interior point x_0 of $(0, b_6/2)$. At x_0 we would have

$$\frac{dG(x: b_6)}{dx} \equiv \frac{1}{(b_6/2 + x)^2} e^{-1/(b_6/2+x)} - \frac{1}{(b_6/2 - x)^2} e^{-1/(b_6/2-x)} = 0 ,$$

and therefore we would have

$$G(x_0; b_6) = \Phi(x_0) \geqq 0 , \tag{20}$$

in which the function Φ is defined by

$$\Phi(x) \equiv \left[1 + \left(\frac{b_6/2 - x}{b_6/2 + x}\right)^2\right] e^{-1/(b_6/2+x)} - e^{-1/b_6} , \qquad x \in \left(\frac{b_6}{2}\right).$$

The function Φ is more tractable than the function G, in that it permits us rigorously to establish the transcendental inequality (19) by investigating only a quadratic function. We shall show that we have

$$\Phi(x) < 0 \text{ for all } x \in \left(0, \frac{b_6}{2}\right), \tag{21}$$

thus contradicting (20) and establishing the theorem.

A computation yields

$$\frac{d\Phi}{dx} = \frac{2e^{-1/(b_6/2+x)}}{(b_6/2 + x)^4} Q(x) , \tag{22}$$

where Q is the quadratic polynomial function defined by

$$Q(x) \equiv (1 + \log 2)x^2 - (1 - \log 2)\left(\frac{b_6}{2}\right)^2. \tag{23}$$

Since

$$Q(0) = -(1 - \log 2)\left(\frac{b_6}{2}\right) < 0 ,$$

$$Q\left(\frac{b_6}{2}\right) = 2 \log 2\left(\frac{b_6}{2}\right) > 0 ,$$

and the coefficient of x^2 in (23) is positive, it follows that $Q(x)$ has precisely one zero on $(0, b_6/2)$, actually at

$$x_0 = \frac{b_6}{2}\sqrt{\frac{1-\log 2}{1+\log 2}}\,,$$

being negative on $(0, x_0)$ and positive on $(x_0, b_6/2)$. Accordingly, by (22), $\Phi(x)$ is strictly decreasing on $(0, x_0)$ and strictly increasing on $(x_0, b_6/2)$, whence the desired inequality (21) follows from

$$\Phi(0) = \Phi\left(\frac{b_6}{2}\right) = 0\,.$$

9. Proof of Theorem 1. In place of the function G of § 8, relative to the function F given by (6), we now consider, relative to the function f given by (7), the function g defined for $b \in (0, \infty)$ and $x \in [0, b/2]$ by

$$\begin{aligned} g(x; b) &\equiv \left(1 + \frac{1}{b/2+x}\right)e^{-(b/2+x)} \\ &\quad + \left(1 + \frac{1}{b/2-x}\right)e^{-1/(b/2-x)} - \left(1+\frac{1}{b}\right)e^{-1/b}\,, \quad x \neq \frac{b}{2}\,, \\ g\left(\frac{b}{2}; b\right) &= 0\,. \end{aligned} \tag{24}$$

To prove the theorem, we shall show that the maximum value b such that

$$g(x; b) \leqq 0 \text{ for all } x \in \left[0, \frac{b}{2}\right]$$

is given by (12).

In particular, for f to be superadditive on $[0, b]$, it is necessary that we have

$$\begin{aligned} g(0; b) &\equiv 2\left(1+\frac{2}{b}\right)e^{-2/b} - \left(1+\frac{1}{b}\right)e^{-1/b} \\ &\equiv \frac{e^{-1/b}}{b}[b(2e^{-1/b}-1) - (1-4e^{-1/b})] \leqq 0\,. \end{aligned} \tag{25}$$

Now, as we see through differentiation, on $[0, \infty)$ the function α, defined by

$$\begin{aligned} \alpha(b) &\equiv b(2e^{-1/b} - 1)\,, \qquad 0 < b < \infty\,, \\ \alpha(0) &= 0\,, \end{aligned} \tag{26}$$

is convex; $\alpha(b)$ is strictly decreasing from the value 0 at $b = 0$ to a negative value at the root b_0 (approx. $b_0 = 0.60$) of the transcendental equation

$$b = \frac{2}{e^{1/b} - 2} ,$$

which expresses the relation $d\alpha/db = 0$, and then $\alpha(b)$ is strictly increasing on $[b_0, \infty)$.

On the other hand, the function β, defined on $[0, \infty)$ by

$$\text{(27)} \qquad \begin{aligned} &\beta(b) \equiv 1 - 4e^{-1/b} , \qquad 0 < b < \infty , \\ &\beta(0) = 1 , \end{aligned}$$

is strictly decreasing on its entire interval of definition.

Since

$$\alpha(b_0) < 0 \text{ and } \beta(b_0) = \frac{1 - b_0}{1 + b_0} > 0 ,$$

it therefore follows from (25), (26), and (27) that the equation

$$g(0; b) = 0$$

has a single root $b \in (0, \infty)$, namely, at the solution

$$b = b_4 = b^* \doteq 0.8955$$

of the transcendental equation

$$\alpha(b) = \beta(b) ,$$

and that further $g(0; b)$ satisfies the inequalities

$$\text{(28)} \qquad \begin{aligned} &g(0; b) < 0 , \qquad 0 < b < b_4 , \\ &g(0; b) > 0 , \qquad b > b_4 . \end{aligned}$$

By (28), the function f is not superadditive on $[0, b]$ for any $b > b_4$; it remains for us to show that f is superadditive on $[0, b_4]$.

For this, it is sufficient that we establish the inequality

$$\text{(29)} \qquad g(x; b_4) \leqq 0 \text{ for all } x \in \left[0, \frac{b_4}{2}\right] .$$

By (24), we have

$$\text{(30)} \qquad g\left(\frac{b_4}{2}; b_4\right) = 0 ,$$

and by the choice of b_4 we have also

$$\text{(31)} \qquad g(0; b_4) = \frac{e^{-1/b_4}}{b_4}[b_4(2e^{-1/b_4} - 1) - (1 - 4e^{-1/b_4})] = 0 .$$

We shall prove that

(32) $$g(x;\, b_4) < 0 \text{ for all } x \in \left(0, \frac{b_4}{2}\right),$$

thus establishing (29) and with it the validity of the theorem.

If (32) did not hold, then, by (30) and (31), $g(x;\, b_4)$ would attain a nonnegative maximum value at some interior point x_0 of $(0, b_4/2)$. At x_0 we would have

$$\frac{dg(x;\, b_4)}{dx} \equiv \frac{1}{(b_4/2 + x)^3} e^{-1/(b_4/2+x)} - \frac{1}{(b_4/2 - x)^3} e^{-1/(b_4/2-x)} = 0 ,$$

and therefore

(33) $$g(x_0;\, b_4) = \varphi(x_0) \geqq 0 ,$$

in which the function φ is defined by

$$\varphi(x) \equiv \left[1 + \frac{1}{b_4/2 + x} + \left(1 + \frac{1}{b_4/2 - x}\right)\left(\frac{b_4/2 - x}{b_4/2 + x}\right)^3\right] e^{-1/(b_4/2+x)} - \left(1 + \frac{1}{b_4}\right) e^{-1/b_4} , \qquad x \in \left(0, \frac{b_4}{2}\right) .$$

We shall show that we have

(34) $$\varphi(x) < 0 \text{ for all } x \in \left(0, \frac{b_4}{2}\right),$$

thus contradicting (33) and establishing the theorem.

A computation yields

(35) $$\frac{d\varphi}{dx} = \frac{2e^{-1/(b_4/2+x)}}{(b_4/2 + x)^5} q(x) ,$$

where q is the cubic polynomial function defined by

(36) $$\begin{aligned} q(x) \equiv & \left[-3\left(\frac{b_4}{2}\right) - 1\right] x^3 + \left[3\left(\frac{b_4}{2}\right)^2 + 4\left(\frac{b_4}{2}\right) + 1\right] x^2 \\ & + \left[3\left(\frac{b_4}{2}\right)^3 - \left(\frac{b_4}{2}\right)^2\right] x + \left[-3\left(\frac{b_4}{2}\right)^4 - 2\left(\frac{b_4}{2}\right)^3 + \left(\frac{b_4}{2}\right)^2\right] . \end{aligned}$$

Since

$$q(0) = \left[-3\left(\frac{b_4}{2}\right)^2 - 2\left(\frac{b_4}{2}\right) + 1\right]\left(\frac{b_4}{2}\right)^2 \doteq -0.497\left(\frac{b_4}{2}\right)^2 < 0 ,$$

$$q\left(\frac{b_4}{2}\right) = 2\left(\frac{b_4}{2}\right)^2 > 0 ,$$

and the coefficient of x^3 in (36) is negative, it follows that $q(x)$ has precisely one zero on $(0, b_4/2)$, say at $x = x_0$. Then $q(x)$ is negative on $(0, x_0)$ and positive on $(x_0, b_4/2)$. Accordingly, by (35), $\varphi(x)$ is strictly decreasing on $(0, x_0)$ and strictly increasing on $(x_0, b_4/2)$, whence the desired inequality (34) follows from

$$\varphi(0) = \varphi\left(\frac{b_4}{2}\right) = 0 .$$

10. The reverse implications. The numbers b_i, $i = 1, 2, \cdots, 6$, as given by (8)-(13), satisfy

$$b_{i-1} < b_i , \qquad i = 2, 3, \cdots, 6,$$

in accordance with the following table of approximations:

i	b_i
1	0.3333
2	0.5000
3	0.6180
4	0.8955
5	1.0000
6	1.4428

Accordingly, since b_i is the maximum of all numbers b such that the function f given by (7) is of class $K_i(b)$, it follows that $f \in K_i(b_i)$ but $f \notin K_{i-1}(b_i)$, whence

$$K_i(b_i) \not\subset K_{i-1}(b_i), \qquad i = 2, 3, \cdots, 6.$$

This establishes (4).

11. The sign of equality. In determining maximum intervals of superadditivity, we have established the following results, except for the specification of the conditions under which the sign of equality holds.

THEOREM 3. *With the notation of Theorem* 1, *we have*

$$f(x + y) \geqq f(x) + f(y) \tag{37}$$

for all nonnegative x and y satisfying

$$x + y \leqq b^* .$$

The sign of equality holds in (37) *if and only if either at most one of x and y is different from* 0 *or else*

$$x = y = \frac{b^*}{2} .$$

THEOREM 4. *With the notation of Theorem 2, we have*

$$F(x + y) \geqq F(x) + F(y) \tag{38}$$

for all nonnegative x and y satisfying

$$x + y \leqq \frac{1}{\log 2} .$$

The sign of equality holds in (38) *if and only if either at most one of x and y is different from* 0 *or else*

$$x = y = \frac{1}{2 \log 2} .$$

Proof. We have only to discuss the conditions under which the sign of equality holds in (37) and (38).

To establish the validity of Bruckner's test, which we have used in the proof of Theorems 1 and 2, we observe that if $f \in K(b)$ is convexo-concave, then the difference

$$[f(x) + f(y)] - f(x + y) \tag{39}$$

is either nonincreasing, or nondecreasing, or first nonincreasing and then nondecreasing, in each of its variables, in the triangular region

$$x \geqq 0 , \qquad y \geqq 0 , \qquad x + y \leqq b ,$$

and hence attains its maximum value either on the line $x + y = b$ or at the origin. For the functions with which we are dealing, however, the above difference is either strictly decreasing, or strictly increasing, or first strictly decreasing and then strictly increasing, in each of its variables, *except* when the other is 0.

Hence, in applying Bruckner's test, the only points we have bypassed at which the sign of equality might hold lie along the axes, and thus the difference attains it maximum value *only* on the triangular boundary.

The boundary consists of the segments $0 \leqq x \leqq b_i$, $0 \leqq y \leqq b_i$, and the portion of the line $x + y = b_i$ in the first quadrant, where $b_i = b_4 = b^*$ and $b_i = b_6 = 1/(\log 2)$ for Theorems 3 and 4, respectively. The difference (39), for the functions of Theorems 3 and 4, vanishes identically on the axes, whereas on the interior of the remaining side, by (18) and (19), and by (31) and (32), it vanishes at the midpoint and otherwise is negative, as specified in the statement of the two theorem.

We note, in passing, that to establish Theorems 2 and 4 without recourse to Bruckner's test, we might adjust the foregoing proofs as follows. For any b', $0 < b' \leqq b_6$, by (14) we have

$$G\left(\frac{b'}{2}; b'\right) = 0 ; \tag{40}$$

further, by (15), we have

$$G(0; b') \leqq 0 , \tag{41}$$

with equality if and only if $b' = b_6$. The proof of (19) can now be extended to give

$$G(x; b') < 0 \text{ for all } x \in \left(0, \frac{b'}{2}\right) . \tag{42}$$

The conclusion of Theorems 2 and 4, including the condition for equality, follows from (40), (41), and (42). Analogous remarks hold for Theorems 1 and 3.

12. Superadditivity inequalities. Let $f \in K_4(b)$. An immediate induction on (3) yields

$$\sum_{i=1}^{n} f(x_i) \leqq f\left(\sum_{i=1}^{n} x_i\right) , \qquad 0 \leqq x_i \leqq b , \qquad \sum_{i=1}^{n} x_i \leqq b . \tag{43}$$

Since, by definition, any function $f \in K_4(b)$ is nonnegative, it follows from (3) that f is nondecreasing. Therefore, by (43), we have

$$\sum_{i=1}^{n} f(x_i) \leqq f(b) , \qquad 0 \leqq x_i \leqq b , \qquad \sum_{i=1}^{n} x_i \leqq b . \tag{44}$$

Thus, for example, for positive numbers x_i, $i = 1, 2, \cdots, n$, $n \geqq 1$, such that

$$\sum_{i=1}^{n} x_i = x_0 \leqq b^* , \tag{45}$$

by Theorem 3 we have

$$\sum_{i=1}^{n} \left(1 + \frac{1}{x_i}\right) e^{-1/x_i} \leqq \left(1 + \frac{1}{x_0}\right) e^{-1/x_0} ,$$

with equality if and only if either (a) $n = 1$, or (b) $n = 2$ and $x_1 = x_2 = b^*/2$.

Also, for positive numbers x_i satisfying (45), we have the weaker inequality

$$\sum_{i=1}^{n} \left(1 + \frac{1}{x_i}\right) e^{-1/x_i} \leqq \left(1 + \frac{1}{b^*}\right) e^{-1/b^*} ,$$

with equality if and only if either (a) $n = 1$ and $x_1 = b^*$, or (*b*) $n = 2$ and $x_1 = x_2 = b^*/2$.

Similarly, for positive numbers x_i, $i = 1, 2, \cdots, n$, such that

$$\sum_{i=1}^{n} x_i = x_0 \leqq \frac{1}{\log 2} ,$$

we have

$$\sum_{i=1}^{n} e^{-1/x_i} \leqq e^{-1/x_0} \leqq e^{-\log 2} ,$$

with analogous conditions for the sign of equality to hold.

13. Whittaker's inequality. If, for any number $a > 1$, in the foregoing discussions we substitute $x/\log a$ for x, then we obtain the following results:

The function f_a, defined by

$$f_a(x) \equiv a^{-1/x}\left(1 + \frac{1}{x} \log a\right) , \qquad 0 < x < \infty ,$$
$$f_a(0) = 0 ,$$

is convex on the interval $[0, (1/3) \log a]$, starshaped on the interval $[0, (1/2)(\sqrt{5} - 1) \log a]$, and superadditive on the interval $[0, b^* \log a]$.

The function F_a, defined by

$$F_a(x) \equiv a^{-1/x} , \qquad 0 < x < \infty ,$$
$$F_a(0) = 0 ,$$

is convex on the interval $[0, (1/2) \log a]$, starshaped on the interval $[0, \log a]$, and superadditive on the interval $[0, \log a/\log 2]$.

In particular, the function F_2 is superadditive on the interval $[0, 1]$. Therefore, for positive numbers x_i, $i = 1, 2, \cdots, n$, $n \geqq 1$, such that

$$\sum_{i=1}^{n} x_i = x_0 \leqq 1 , \tag{46}$$

we have the inequality

$$\sum_{i=1}^{n} 2^{-1/x_i} \leqq 2^{-1/x_0}$$

and the weaker inequality

$$\sum_{i=1}^{n} 2^{-1/x_i} \leqq 2^{-1} . \tag{47}$$

Substituting $1/(y_i + 1)$ for x_i in (46) and (47), we obtain the following result:

If the nonnegative numbers y_i, $i = 1, 2, \cdots, n$, $n \geqq 1$, *are such that*

$$\sum_{i=1}^{n} \frac{1}{1+y_i} \leqq 1 ,$$

then we have

$$\sum_{i=1}^{n} 2^{-y_i} \leqq 1 , \tag{48}$$

with equality if and only if either (a) $n = 1$ *and* $y_1 = 0$, *or* (b) $n = 2$ *and* $y_1 = y_2 = 1$.

The relation (48) is Whittaker's inequality [10, 4].

14. The method of Boas. The following sufficient condition for superadditivity was suggested to the author by R. P. Boas in personal correspondence:

BOAS'S TEST. *If the function* $f \in K(b)$ *is of class* C', *and there are numbers* $a \leqq b/2$ *and* $c \leqq a$ *such that*

(i) f *is starshaped on* $[0, 2a]$,

(ii) f *is concave and satisfies* $f(x/2) \leqq (1/2)f(x)$ *on* $[c, b]$,

(iii) $f'(0) < f'(b)$,

(iv) $f'(x) - f'(b - x)$ *has at most one zero in* $(0, a)$,

then f *is superadditive on* $[0, b]$.

The validity of the test can be established by considering separately the following three cases:

(i) $0 \leqq x \leqq a, \quad 0 \leqq y \leqq a,$

(ii) $x \geqq a, \quad y \geqq a, \quad x + y \leqq b,$

(iii) $x < a < y < b, \quad x + y \leqq b.$

Boas has observed that his test applies to such convexo-concave functions, or functions having ogive-shaped graphs, as $e^{-1/x\alpha}$ for $0 < \alpha \leqq 1$, $\log(1 + x^\lambda)$, and arc tan x^λ, yielding intervals of superadditivity and consequent inequalities typified by the inequality of Whittaker given in § 13, above.

A systematic tabulation of maximum intervals of superadditivity of such functions, of their average functions, and of their inverse average functions, might well be desirable.

15. Combination inequalities. If the function f is convex for $x \in [a, b]$, then, by Jensen's inequality [6, 1], for any numbers

$$x_i \in [a, b] , \qquad i = 1, 2, \cdots, n ,$$

and any weights

$$\alpha_i, \alpha_i > 0, \sum_{i=1}^{n} \alpha_i = 1 ,$$

we have

$$f\left(\sum_{i=1}^{n} \alpha_i x_i\right) \leqq \sum_{i=1}^{n} \alpha_i f(x_i) . \tag{49}$$

This inequality is an extension of the defining inequality (1).

Analogues of the inequality (49), for functions of the sort treated in this paper, are given in the two theorems that follow.

THEOREM 5. *If the function $f \in K(b)$ is convex for $x \in [0, a]$ and starshaped for $x \in [0, b]$, $b > a$, then for any numbers*

$$x_i \in [0, b], \qquad i - 1, 2, \cdots, n ,$$

and any weights

$$\alpha_i,\ \alpha_i > 0,\ \sum_{i=1}^{n} \alpha_i = 1 ,$$

we have

$$f\left(\frac{a}{b} \sum_{i=1}^{n} \alpha_i x_i\right) \leqq \frac{a}{b} \sum_{i=1}^{n} \alpha_i f(x_i) . \tag{50}$$

Proof. Since the numbers x_i satisfy $0 \leqq x_i \leqq b$, we have $0 \leqq ax_i/b \leqq a$, so that Jensen's inequality (49) can be applied for the numbers ax_i/b, yielding

$$f\left(\sum_{i=1}^{n} \alpha_i \frac{a}{b} x_i\right) \leqq \sum_{i=1}^{n} \alpha_i f\left(\frac{a}{b} x_i\right) . \tag{51}$$

Now since $0 \leqq x_i \leqq b$, and $a/b < 1$, the defining inequality (2) for starshapedness gives

$$f\left(\frac{a}{b} x_i\right) \leqq \frac{a}{b} f(x_i) , \qquad i = 1, 2, \cdots, n , \tag{52}$$

and (50) follows from (51) and (52).

For example, for the function f defined by (7) we have $a = 1/2$, $b = 1$, so that for positive numbers $x_i \leqq 1$ and weights α_i we have

$$\exp \frac{-2}{\sum_{i=1}^{n} \alpha_i x_i} \leqq \frac{1}{2} \sum_{i=1}^{n} \alpha_i e^{-1/x_i} .$$

THEOREM 6. *If the function $f \in K(c)$ is convex for $x \in [0, a]$, starshaped for $x \in [0, b]$, and superadditive for $x \in [0, c]$, $c > b > a$, then for any numbers*

$$x_i \in [0, b] , \qquad i = 1, 2, \cdots, n ,$$

satisfying

$$\sum_{i=1}^{n} x_i = c_0 \leqq c \,, \tag{53}$$

we have

$$f\left(\frac{ac_0}{bn}\right) = f\left(\frac{a}{bn}\sum_{i=1}^{n} x_i\right) \leqq \frac{a}{bn}\sum_{i=1}^{n} f(x_i) \leqq \frac{a}{bn} f\left(\sum_{i=1}^{n} x_i\right) \tag{54}$$

$$= \frac{a}{bn} f(c_0) \leqq \frac{a}{bn} f(c) \,.$$

Proof. By (50), we have

$$f\left(\frac{a}{bn}\sum_{i=1}^{n} x_i\right) \leqq \frac{a}{bn}\sum_{i=1}^{n} f(x_i) \,, \tag{55}$$

and from (43) and (53) we obtain

$$\sum_{i=1}^{n} f(x_i) \leqq f\left(\sum_{i=1}^{n} x_i\right) , \tag{56}$$

whence (54) follows from (55), (56), and the fact that f is a non-decreasing function.

By way of illustration, for positive numbers $x_i \leqq 1$ satisfying

$$\sum_{i=1}^{n} x_i = c_0 \leqq \frac{1}{\log 2} \,,$$

we have both lower and upper bounds for

$$\sum_{i=1}^{n} e^{-1/x_i} \,,$$

given by

$$e^{-2n/c_0} = \exp \frac{-2n}{\sum_{i=1}^{n} x_i} \leqq \frac{1}{2n}\sum_{i=1}^{n} e^{-1/x_i} \leqq \frac{1}{2n} \exp \frac{-1}{\sum_{i=1}^{n} x_i}$$

$$= \frac{1}{2n} e^{-1/c_0} \leqq \frac{1}{2n} e^{-\log 2} = \frac{1}{4n} \,.$$

For a function having a relatively longer interval of superadditivity, a more useful inequality would result.

References

1. E. F. Beckenbach, *Convex functions*, Bull. Amer. Math. Soc. **54** (1948), 439–460.

2. A. M. Bruckner, *Tests for the superadditivity of functions*, Proc. Amer. Math. Soc. **13** (1962), 126–130.

3. ——— and E. Ostrow, *Some function classes related to the class of convex functions*, Pacific J. Math. **12** (1962), 1203–1215.

4. B. L. Foster and R. R. Phelps, *Related inequalities*, Amer. Math. Monthly **64**

(1957), 677-678.
5. E. Hille and R. S. Phillips, *Functional analysis and semi-groups*, Amer. Math. Soc. Colloquium Publications, vol XXXI, rev. ed., Ch. 7, American Mathematical Society, Providence, R. I., 1957.
6. J. L. W. V. Jensen, *Sur un nouvel et important théorème de la théorie des fonctions*, Aata Math. **22** (1898-1899), 359-364.
7. Pacific Journal of Mathematics **11** (1961), cover design.
8. T. Radó, *On convex functions*, Trans. Amer. Math. Soc. **37** (1935), 266-285.
9. R. A. Rosenbaum, *Subadditive functions*, Duke Math. J. **17** (1950), 227-247.
10. J. V. Whittaker, *Problem* 4712, Amer. Math. Monthly **63** (1956), 669.

UNIVERSITY OF CALIFORNIA, LOS ANGELES

THE SIMPLE CONNECTIVITY OF THE SUM OF TWO DISKS

R. H. BING

1. **Introduction.** The following question was called to the author's attention several years ago by Eldon Dyer.

Question. Is the sum of two disks simply connected if their intersection is connected?

Later, the author saw a communication in which an erroneous proof was given that Example 1 of the present paper is not simply connected. We show in § 2 that Example 1 is simply connected. However, we give some examples (Examples 2, 3, 4) in §§ 3, 4, 5 that are not simply connected.

A topological characterization is given in § 4 of intersections that will prevent closed curves which finitely oscillate between two disks from being shrunk. If the intersection is snake-like or arcwise connected, such finitely oscillating curves can always be shrunk but there are examples in which infinitely oscillating curves cannot. It is the topology of the intersection which prevents the sum of two disks from being simply connected rather than the embeddings of the intersection in the disks as shown in §§ 4 and 5. In fact, as pointed out in § 6, much of what we have learned about the sums of disks applies to the sums of continuous curves.

We use Example 1 in § 7 to construct a peculiar group and show that a certain relation kills it.

All sets treated in this paper are metric.

Let I^n denote an *n-cell* and *Bd* I^n its *boundary*. A set A is *n-connected* if each map (continuous transformation) f of *Bd* I^{n+1} into A can be extended to map I^{n+1} into A. We say that $f(Bd\ I^{n+1})$ *can be shrunk to a point* if the map can be extended. A set is called an *ε-set* if its diameter is less than or equal to ε. A set A is *n-ULC* if for each $\varepsilon > 0$ there is a $\delta > 0$ such that each map of *Bd* I^{n+1} onto a δ-subset of A can be shrunk to a point on an ε-subset. A compact continuum is called a *continuous curve* if it is 0-*ULC*. A set is *simply connected* if it is 1-connected. It is *uniformly locally simply connected* if it is 1-*ULC*. We shall not treat higher types of connectivity in this paper.

We find it convenient to consider an abstract disk D rather than the square I^2. A map of *Bd D* is a *closed curve*. If h is a homeomorphism, $h(BdD)$ is a *simple closed curve*.

Received March 6, 1963. Research supported by NSF grants NSF-G21514 and NSF-G11665.

We shall use cylindrical coordinates (ρ, θ, z) to describe examples in E^3 (Euclidean 3 space). If no z coordinate is given, it is understood that $z = 0$. When we use D alone without subscripts it is understood that we mean the unit disk ($\rho \leqq 1$) in the $z = 0$ plane.

Let f be a map of $Bd\ D$ into E^2 so that the ρ value of each point of $f(Bd\ D)$ is positive. Let $k(\theta)$ be a map of the reals into the reals such that $k(\theta)$ mod 2π is the θ value of $f(1, \theta)$. We say that f *circles the origin n times* if $k(2\pi) - k(0) = 2\pi n$.

2. A false example. Let a, b be fixed numbers with $0 < a < b < 1$ and K_1 be a spiral connecting the circles $\rho = a$ and $\rho = b$ as shown in Figure 1 and given by the following formula.

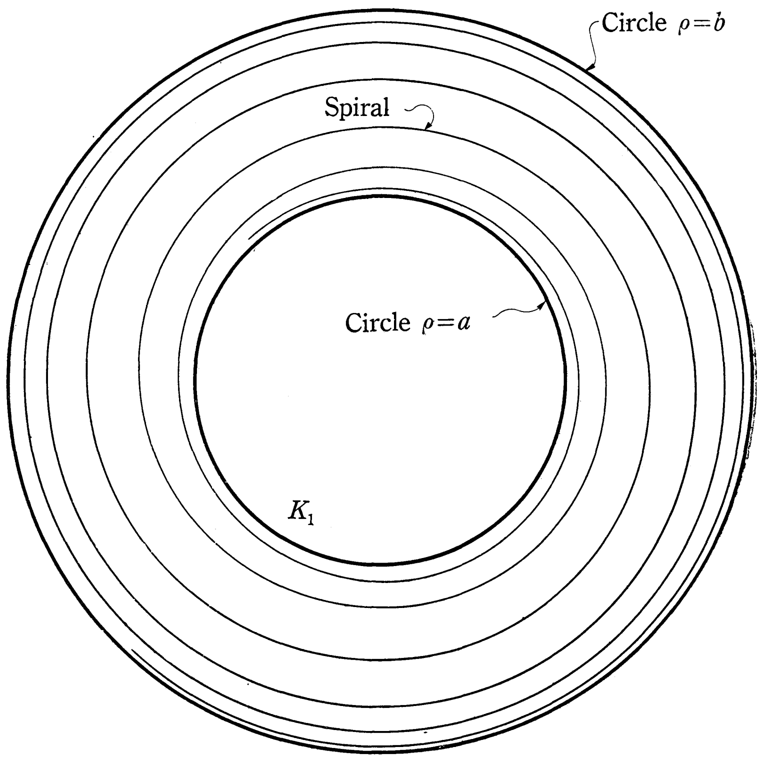


Figure 1

$$K_1 = \{(\rho, \theta)/\rho = a, b, \text{ or } (b + ae^\theta)/(1 + e^\theta)\} .$$

M. K. Fort showed [3] that any bounded plane continuum which has K_1 as a continuous image separates the plane.

EXAMPLE 1. Let D_1 be a disk in E^3 defined by

$$D_1 = \{(\rho, \theta, z)/\rho \leqq 1, z = \text{distance from } (\rho, \theta, 0) \text{ to } K_1\} .$$

Let D_2 be the reflection of D_1 through the $z = 0$ plane. Then $D_1 + D_2$ is the sum of two disks whose intersection is the connected set K_1.

THEOREM 1. *Example* 1 *is simply connected.*

Proof. Let f be a map of $Bd\ D$ into $D_1 + D_2$. We show that $D_1 + D_2$ is simply connected by showing that f can be extended to map D into $D_1 + D_2$.

With no loss of generality we suppose that the ρ value of each point of $f(Bd\ D) \geqq a$.

Special case. (The θ value of each point is fixed under f and the ρ value of each point of $f(Bd\ D) < b$.) In this special case we start by extending f to the circle $\rho = a$ by insisting that f is fixed on this circle.

For each point $q = (1, \theta_q)$ of $Bd\ D$ such that $f(q) \in K_1$, let S_q be the spiral from q about the circle $\rho = a$ described by the formulas $\rho \leqq 1$, $\rho = (2 - a + ae^{(\theta-\theta_q)})/(1 + e^{(\theta-\theta_q)})$, $\theta_q \leqq \theta$. Let f be extended to map S_q into K_1 so that f preserves the θ value of each point of S_q. This extension is made for each such spiral S_q for each point q of $Bd\ D$ such that $f(q) \in K_1$. Note that we have mapped a closed subset of D into K_1 and each component of $D - f^{-1}(K_1)$ other than the interior of $\rho = a$ intersects $Bd\ D$ in an open arc.

Let g_1 be the map of $f^{-1}(K_1 + D_1 \cdot f(Bd\ D)) = f^{-1}(D_1)$ into D_1 given by extended f. Then g_1 can be extended to take D into D_1. For convenience we also call this extended map g_1. Similarly there is a map g_2 of D into D_2 such that $g_2 = f$ on $f^{-1}(K_1 + D_2 \cdot f(Bd\ D)) = f^{-1}(D_2)$. Let g be the map of D into $D_1 + D_2$ given by g_1 on each component of $D - f^{-1}(K_1)$ which has an arc which goes into D_1 under f and $g = g_2$ on the rest of D.

Less special case. (f circles the origin once and the ρ value of each point of $f(Bd\ D)$ is less than b.) We show that there is a homotopy $h_t (0 \leqq t \leqq 1)$ of $Bd\ D$ into $D_1 + D_2$ such that $h_0 = f$, h_1 preserves the θ value of each point of $Bd\ D$ while the ρ value of each point of $h_1(Bd\ D)$ is less than b. The less special case then follows from the special case.

Let $k(\theta)$ be the function that shows that f circles the origin once. For convenience we suppose that $k(0) = 0$, $k(2\pi) = 2\pi$. Let $k_t(\theta) = t\theta + (1 - t)k(\theta)$, $(0 \leqq t \leqq 1)$. As t goes from 0 to 1, $k_t(\theta)$ goes from $k(\theta)$ to θ. For each point $p = (1, \theta_p)$ of $f^{-1}(K_1)$ we define $h_t(p)$ as a point in K_1 so that the θ value of $h_t(p)$ is $k_t(\theta_p)$. The ρ value of $h_t(p)$ is uniquely determined since the three arc components of K_1 are 1-manifolds almost normal to lines through the origin.

The homotopy h_t on $f^{-1}(K_1)$ is extended to $Bd\ D$ so that $h_t(p) \in D_i$ $(i=1, 2)$ if $f(p) \in D_i$, h_1 preserves the θ value of points of BdD and the value of each point of $h_1(Bd\ D)$ is less than b.

The following version of the less special case follows by a similar argument.

Alternative less special case. (f circles the origin once and the value of each point of $f(Bd\ D)$ is greater than a.)

General case. We suppose that $f(Bd\ D)$ intersects the spiral of K_1 in at least three points. Subdivide $Bd\ D$ into arcs $x_1x_2, x_2x_3, \cdots, x_nx_1 (n \geqq 3)$ so that no $f(x_ix_{i+1})$ (addition on subscripts is mod n so that $x_nx_{n+1} = x_nx_1$) intersects both circles in K_1 but each $f(x_i)$ is on the spiral of K_1. Let $x_iz_ix_{i+1}$ be the chord in D from x_i to x_{i+1}.

Extend the map f of $Bd\ D$ into $D_1 + D_2$ to map the chord $x_iz_ix_{i+1}$ into D_1 so that $f(x_iz_ix_i)$ misses the circles in K_1 and f on $x_ix_{i+1}+x_iz_ix_{i+1}$ circles the origin once. It follows from applications of the less special case and its alternative form that we can extend f to take the interiors of the $(x_ix_{i+1} + x_iz_ix_{i+1})$'s into $D_1 + D_2$. We can then extend f to the disk in D bounded by the chords into D_1.

3. A true example. Let C be a Cantor set on the numbers between 1/2 and 1. Let K_2 be the set in the plane consisting of the sum of circles in the plane with centers at the origin and radii in C and spirals joining adjacent circles as shown in Figure 2 and given by the following formula.

$$K_2 = \{(\rho, \theta)/\rho \in C \text{ or } \rho = (b + ae^{\theta})/(1 + e^{\theta})\}$$

where a, b are adjacent numbers of C with $a < b$.

EXAMPLE 2. Let E_1 be the disk in E^2 defined by

$$E_1 = \{(\rho, \theta, z)/\rho \leqq 1, z = \text{distance from } (\rho, \theta, 0) \text{ to } K_2\} .$$

Let E_2 be the reflection of E_1 through the plane $z = 0$. Then $E_1 + E_2$ is the sum of two disks whose sum is the connected set K_2.

Before proving that Example 2 is not simply connected we investigate an interesting property of K_2. M. K. Fort, Jr. showed [3] that any compact continuum in the plane separates the plane if it maps onto K_1. We modify his argument slightly to show the following.

THEOREM 2. *If f maps a closed bounded connected subset of the*

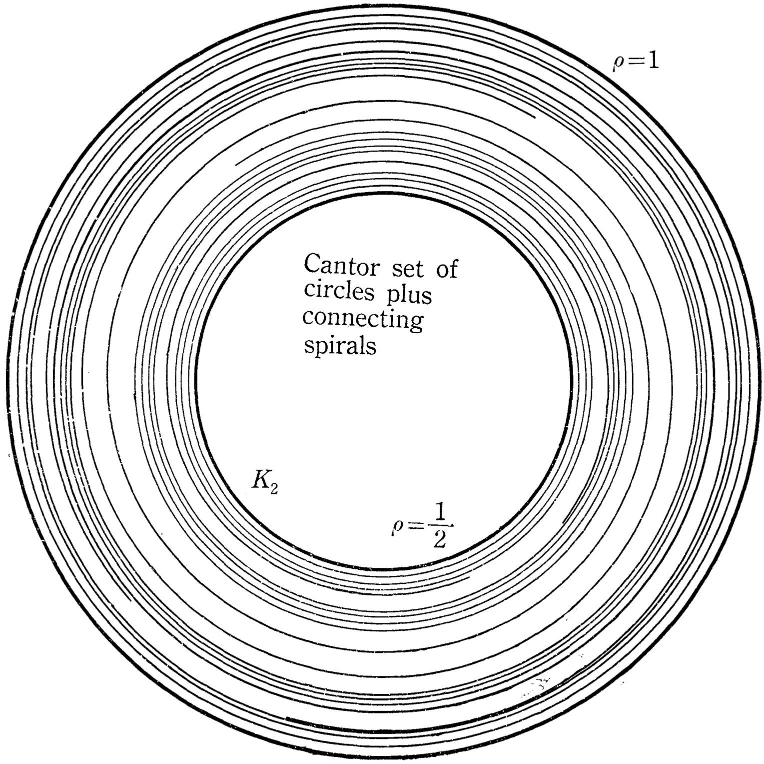


Figure 2

plane onto K_2 *then for each circle* J *in* K_2, *each component of* $f^{-1}(J)$ *separates the plane.*

Proof. Let S^1 be the circle $\rho = 1$ and define $g\colon K_2 \to S^1$ by $g(\rho, \theta) = (1, \theta)$. It is easy to verify that (K_2, S^1, g) is a locally trivial fiber space with totally disconnected fibers.

Suppose X is a component of $f^{-1}(J)$ that does not separate the plane. There is a homotopy pulling the map gf of X into S^1 to a constant map. Since S^1 is an ANR there is a neighborhood N of X such that the map gf of $\bar{N} \cdot f^{-1}(K_2)$ into S^1 is homotopic to a constant map. Take N so close to X that $\bar{N}$ does not cover $f^{-1}(K_2)$.

Let Y be a continuum in $f^{-1}(K_2)$ irreducible from $E^2 - N$ to X. Note that $Y \subset \bar{N}$, $Y \cdot X \neq 0$, and $Y \not\subset X$. It follows from the lemma on page 542 of [3] that $f(Y)$ is contained in an arc component of K_2. This violates the condition that the arc component of K_2 containing J does not intersect $K_2 - J$.

THEOREM 3. *Example* 2 *is not simply connected.*

Proof. Let x and y be points on the inner and outer circles in

K_2 and xz_iy be an arc in E_i from x to y. Let f be a map of $Bd\,D$ onto $xz_1y + xz_2y$ so that the upper half of $Bd\,D$ goes homeomorphically onto xz_1y and the lower half of $Bd\,D$ goes homeomorphically onto xz_2y. We show that Example 2 is not simply connected by showing that f cannot be extended to map D into $E_1 + E_2$.

Assume that f can be extended to send all of D into $E_1 + E_2$. We show that under this false assumption that $p = (1, 0, 0)$ and $q = (1, \pi, 0)$ belong to the same component of $f^{-1}(K_2)$. If they did not belong to the same component, it follows from Theorem 14 on page 171 of [6] (Theorem 10 on page 185 of 1932 edition) that there is a simple closed curve J in the plane $z = 0$ which misses $f^{-1}(K_2)$ and separates p from q in this plane. There would then be an arc A in $J \cdot D$ that intersects both the upper and lower halves of $Bd\,D$. This is impossible since f takes the upper half of $Bd\,D$ into E_1 and the lower half into E_2 but no point of A into $E_1 \cdot E_2 = K_2$.

Let Y be the component of $f^{-1}(K_2)$ containing p and q. Let Z be a subcontinuum of Y irreducible from p to q. Note that f maps Z onto K_2.

If F is a subcontinuum of Z which separates the plane E^2, no bounded component of $E^2 - F$ intersects Z since Z is irreducible from p to q and neither p nor q is in a bounded component of $E^2 - F$. Hence Z does not contain uncountably many mutually exclusive subcontinua each of which separates E^2. This contradicts Theorem 2 which says that for each circle J in K_2, $Z \cdot f^{-1}(J)$ separates E^2.

4. Finitely oscillating curves. A map of a simple closed curve J into the sum of two disks has only *finite oscillation* with respect to the two disks if J is the sum of a finite number of arcs such that the image of each lies in one of the disks. In some examples (Examples 3, 4, 5 to follow) finitely oscillating curves can be shrunk to points but some others cannot. The proof of Theorem 3 showed that Example 2 contained a finitely oscillating curve which could not be shrunk to a point.

We shall show that whether or not all finitely oscillating curves in the sum of two disks can be shrunk to points in the sum is dependent on whether or not the intersection has a certain extremal inverse property. A set X has the *extremal inverse property* with respect to its points p, q if there is a continuum Z in disk D with points p', q' on $Bd\,D$ and a map of Z into X that takes p', q' to p, q respectively.

Let K_3 be the sum of a triod T and a spiral S about T as shown in Figure 3 and given by the following equations.

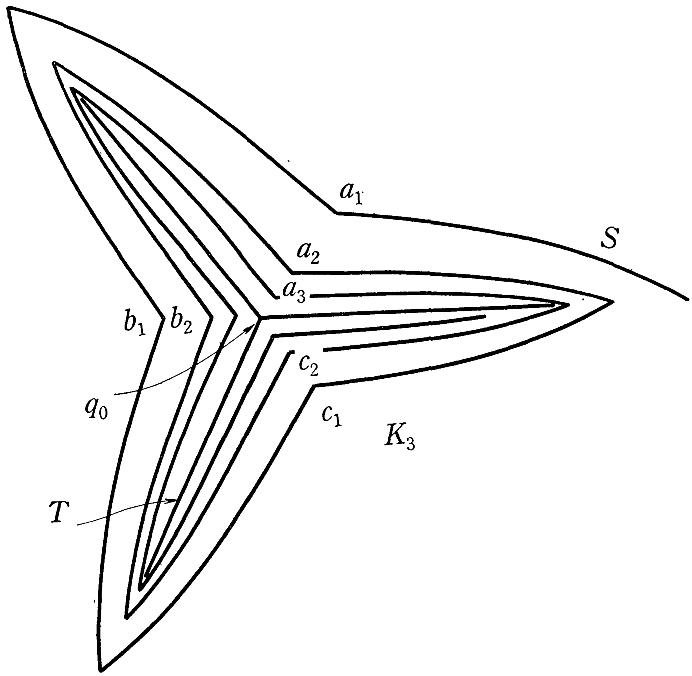


Figure 3

$$T = \{(\rho, \theta)/\rho \leqq 1, \theta = 0, 2\pi/3, \text{ or } 4\pi/3\} ,$$
$$S = \{(\rho, \theta)/\rho = |\cos 3\theta/2|^{\theta} + 1/\theta, \theta \geqq 2\pi\} .$$

EXAMPLE 3. Let D_1, D_2 be two disks whose intersection is K_3.

THEOREM 4. *K_3 has the extremal inverse property with respect to each pair of its points.*

Proof. We consider only the case where $p \in T$ and $q \in S$. Let S' be another spiral about T which misses S, f be a retraction of $S' + T$ onto T, and p' be a point of S' that maps onto p under f. Extend f to the identity on $S + T$. There is a disk containing $T + S + S'$ which has p' and q on its boundary.

THEOREM 5. *Each snake-like continuum has the extremal inverse property with respect to each pair of its points.*

Proof. Apply the following result to a subcontinuum of the snake-like continuum irreducible between the two points under consideration.

THEOREM 6. *Each snake-like continuum is the image of a pseudo-arc.*

Proof. This theorem has been proved by each of Fearnley [2], Lelek [4], and Mioduszewski [5] but we include a slightly different proof.

Let $D_1, D_2, \cdots$ be a sequence such that D_i is a $1/i$ chain properly covering snake-like continuum X and such that D_{i+1} is a refinement of D_i. It follows from Theorem 7 of [1] that if P is a pseudo-arc there is a sequence of proper open coverings $E_1, E_2, \cdots$ of P such that E_i has the same number of links as D_i and for the jth link of D_{i+1} there is an integer $n(i, j)$ such that the jth link of D_{i+1} lies in the $n(i, j)$th link of D_i and the jth link of E_{i+1} lies in the $n(i, j)$th link of E_i.

For each point p of P let $e(p, i)$ be the sum of the links of E_i containing p and $d(p, i)$ be the sum of the corresponding links of D_i. Note that $e(p, i+1) \subset e(p, i)$ and $d(p, i+1) \subset d(p, i)$. For each point p of P let $f(p)$ be the intersection of the closures of $d(p, i)$'s. Then f is a continuous transformation of P onto X.

THEOREM 7. *If a set has the extremal inverse property, so does each of its continuous images.*

THEOREM 8. *Each arcwise connected set has the extremal inverse property with respect to each pair of its points.*

Note that the following theorem applies to simply connected and uniformly locally simply connected continuous curves as well as merely to disks.

THEOREM 9. *Let A_1, A_2 be two compact sets each of which is* 0*-connected,* 1*-connected,* 0-ULC, *and* 1-ULC. *A necessary and sufficient condition that each finitely oscillating curve with respect to A_1, A_2 can be shrunk to a point in $A_1 + A_2$ is that $A_1 \cdot A_2$ has the extremal inverse property with respect to each pair of its points.*

Proof. If $A_1 \cdot A_2$ does not have the extremal inverse property with respect to point x, y of $A_1 \cdot A_2$, let f be a map of $Bd\,D$ into $A_1 + A_2$ such that the upper half of $Bd\,D$ goes into a path in A_1 from x to y and the lower half of $Bd\,D$ goes into a path in A_2 from x to y. It follows from the proof of Theorem 3 that $f(Bd\,D)$ cannot be shrunk to a point in $A_1 + A_2$.

To prove the sufficiency case consider a map f of $Bd\,D$ into $A_1 + A_2$ so that $Bd\,D$ is the sum of arcs $x_1x_2, x_2x_3, \cdots, x_nx_1$ so that each $f(x_i)$ lies on $A_1 \cdot A_2$ and each $f(x_ix_{i+1})$ lies in one of A_1, A_2. Just as we used chords in the general case of the proof of Theorem 1, we consider continua $Z_1, Z_2, \cdots, Z_n$ in D so that Z_i contains x_i and x_{i+1} and f can be extended to take $Z_1 + Z_2 + \cdots + Z_n$ into $A_1 \cdot A_2$. Then as in the proof of the General Case of the proof of Theorem 1 we extend f to take the components of $D-(Z_1+Z_2+\cdots+Z_n)$ which intersect $Bd\,D$ into the appropriate one of A_1, A_2 and then extend f to take the rest of D into A_1. To extend f to take D into A_i for example, one would add a null sequence of arcs in D to $Z_1+Z_2+\cdots+Z_n$ to get a set Z_0 so that $D-Z_0$ is a null sequence of open disks, use the fact that A_i is 0-connected and 0-ULC to extend f to Z_0, and finally use the fact that A_i is 1-connected and 1-ULC to extend f to take D into A_i.

THEOREM 10. *Suppose D_1, D_2 are two disks whose intersection is a continuum X and axb is an arc in D_1 that intersects X only at a and b. If f is a map of $Bd\,D$ into D_1+D_2 such that f takes the upper half of $Bd\,D$ onto axb and the lower half into D_2, then $f(Bd\,D)$ can be shrunk to a point in D_1+D_2.*

Proof. Since X has the extremal inverse property with respect to a and b as shown by its embedding in D_1, there is a continuum X' in D intersecting the inverses under f of a and b such that f may be extended to $X'+Bd\,D$. Then f is extended to take the part of D in component of $D-X'$ that contains upper arc of $Bd\,D$ into D_1 and the rest of D into D_2.

Question. The question suggests itself as to which continua have the extremal inverse property with respect to each pair of their points. Example 2 does not have it. Example 1 and 3 do. So do Examples 4 and 5 to be given in the next section. Perhaps Example 2 is unnecessarily complicated as an example of a continuum without the extremal inverse property in that it separates the plane into infinitely many pieces. Perhaps there is an example that does not separate the plane. Does each three branched tree-like plane continuum have the extremal inverse property with respect to each pair of its points? (A compact continuum is a three branched tree-like continuum if it is not snake-like but for each positive number ε it has an ε-cover whose 1-nerve is a triod.)

5. Infinite oscillation. We use rectilinear coordinates to define the two sets shown in Figures 4, 5.

$$K_4=\{(x,y)/(x=0,\ -2\leqq y\leqq 2),\ (y=1+\sin 1/x,\ 0<x\leqq 1)\,,\ \text{or}\ (y=-1+\sin 1/x,\ -1\leqq x<0)\}\,.$$

$K_5=$ sum of points of K_3 on or to the left of the vertical line through $(1,0)$ plus the interval from $(1,0)$ to $(1,1)$.

Theorems 5 and 8 show that K_4 and K_5 have the extremal

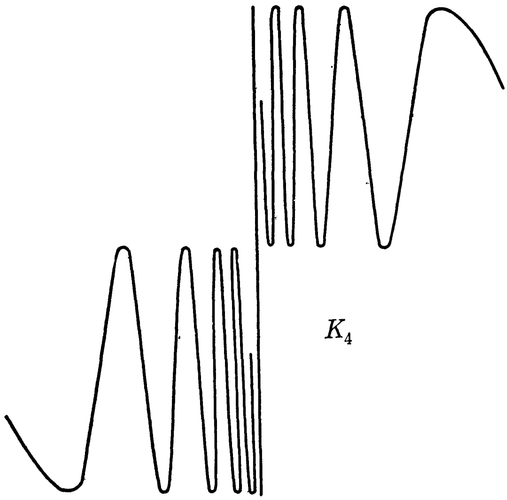


Figure 4

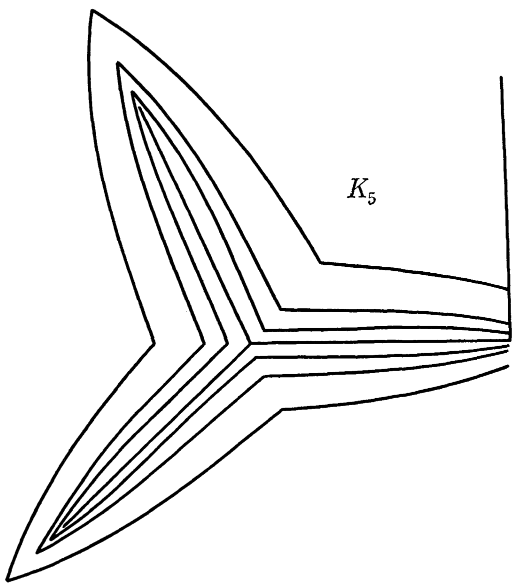


Figure 5

inverse property with respect to each pair of their points. It follows from Theorem 9 that finitely oscillating curves in the following two examples can be shrunk to points in the examples.

EXAMPLE 4. Two disks sewed together along K_4.

EXAMPLE 5. Two disks sewed together along K_5.

THEOREM 11. *Examples* 3, 4, 5 *are not simply connected.*

We only prove the first third of this theorem since the other parts are analogous. We suppose that the disks D_1, D_2 of Example 3 are obtained by pushing parts of a circular disk in the $z = 0$ plane up and down respectively as done in Examples 1 and 2. Theorem 13 shows that there is no loss of generality in supposing this. The disks would be larger than those in Examples 1 and 2 since K_3 is larger than those disks.

Proof that Example 3 *is not simply connected.* Let $a_1, a_2, \cdots$ be the points of K_3 on the open ray $\theta = \pi/3$ ordered inversely according to their distances from the origin q_0. Let $b_1, b_2, \cdots$ be the corresponding points of K_3 on the open ray $\theta = \pi$ and $c_1, c_2, \cdots$ be the corresponding points on the ray $\theta = 5\pi/3$.

Let p_i be the point of $Bd\, D$ whose θ value is $1/i$. Let p_0 be the point of $Bd\, D$ whose θ value is 0. Use p_ip_j to denote the arc on $Bd\, D$ in a clockwise direction from p_i to p_j.

Let f be a map of $Bd\, D$ into $D_1 + D_2$ satisfying the following conditions.

$$\begin{aligned} &f(p_j) = q_0 \text{ (origin) } (j = 0, 2, 4, 6, \cdots)\,, \\ &f(p_{6i-5}) = a_i\,, \\ &f(p_{6i-3}) = b_i\,, \\ &f(p_{6i-1}) = c_i\,, \\ &f(p_{2i-2}p_{2i-1}) \subset D_1\,, \\ &f(p_{2i-1}p_{2i}) \subset D_2\,. \end{aligned}$$

The θ value on each $f(p_ip_{i+1})$ is a constant and f takes the θ values of p_ip_{i+1} linearly onto the ρ values of $f(p_ip_{i+1})$.

Assume f can be extended to take D into $D_1 + D_2$. We call the extended map f. In this extension we suppose that no component of $f^{-1}(q_0)$ separates the $z = 0$ plane. (If a component X did separate the plane, we could modify f to map X plus each of its bounded complementary domains in $z = 0$ into q_0.)

Let F be the part of $D_1 + D_2$ whose ρ value is less than or equal to 1. A finite number of spanning arcs in D separates p_0 from $f^{-1}(D_1 + D_2 - F)$ in D so that no one of the arcs intersects $f^{-1}(q_0)$. Hence, we can cut down the disk D to a disk E such that E agrees with D in a neighborhood of p_0, $f(E) \subset F$, each point of $Bd\, E \cdot f^{-1}(q_0)$ lies on $Bd\, D$, and E agrees with D in a neighborhood of each such point. Let Z be the closure of $Bd\, E - Bd\, D$. Note that $q_0 \notin f(Z)$. We shall obtain a contradiction to the assumption that f can be extended to take D into $D_1 + D_2$ by showing that the map f of $Bd\, E$ into F cannot be extended to take E into F.

With no loss of generality we suppose that $p_0, p_1, p_2, \cdots$ all belong to $Bd\, E$. Note that no component of $E \cdot f^{-1}(K_3)$ intersects two p_i's unless perhaps they both have even subscripts since no two of $a_1, a_2, \cdots, b_1, b_2, \cdots, c_1, c_2, \cdots$ belong to the same component of $K_3 \cdot F$. For i odd, the component Y_i of $E \cdot f^{-1}(K_3)$ containing p_i separates $Bd\, E$ since $f(Bd\, E)$ crosses from D_1 to D_2 at $f(p_i)$. Since $Bd\, D$ intersects Y_i in at most a finite number of points and $f(Bd\, D)$ does not cross from D_1 to D_2 at the image of any of these points other than p_i, Y_i must intersect Z. Let q_i be a point of $Y_i \cdot Z$.

Note that since $a_1, a_2, \cdots, b_1, b_2, \cdots, c_1, c_2, \cdots$ belong to different components of $F \cdot K_3$, $Y_i \neq Y_j$ if i, j are different odd positive integers. Let q_∞ be a limit point of $q_1, q_3, q_5, \cdots$. Since for i sufficiently large Y_{i+2} separates Y_i from p_0 in E, $q_1, q_3, \cdots$ converges to q_∞ from the clockwise side. Since q_∞ is a limit point of each of ΣY_{6i-5}, ΣY_{6i-3}, ΣY_{6i-1}, then $f(q_\infty)$ is a limit point of each of $f(\Sigma Y_{6i-5})$, $f(\Sigma Y_{6i-3})$, $f(\Sigma Y_{6i-1})$. The only point common to the closures of these sets is the point q_0, so $f(q_\infty) = q_0$. However, $f(q_\infty) \neq q_0$ since $q_\infty \in Z$ and $q_0 \notin f(Z)$.

THEOREM 12. *The sum of two disks is simply connected if their intersection is connected and locally arcwise connected.*

Proof. Let f be a map of $Bd\, D$ into the sum of two disks D_1, D_2 such that $D_1 \cdot D_2$ is a continuous curve. For each arc ab of $Bd\, D$ which intersects $f^{-1}(D_1 \cdot D_2)$ only in its end points, extend f to map the chord acb of D into an arc in $D_1 \cdot D_2$ such that the diameter of $f(acb)$ is no more than twice the diameter of any other arc in $D_1 \cdot D_2$ from $f(a)$ to $f(b)$. Let f_i be a mapping of D into D_i that agrees with f on the part of D going into D_i under f. Then the extended f is f_1 on the components of D minus the chords which contain a point of $Bd\, D$ that f sends into $D_1 - D_2$ and is f_2 on the rest of D.

THEOREM 13. *The topology of the intersection of two disks determines whether or not their sum is simply connected.*

Proof. Suppose D_1, D_2, E_1, E_2 are disks and h is a homeomorphism of $D_1 \cdot D_2$ onto $E_1 \cdot E_2$. Let D be a circular disk and f a map of $Bd\, D$ into $D_1 + D_2$. We assume that $E_1 + E_2$ is simply connected and show that this assumption implies that f can be extended to map D into $D_1 + D_2$. We assume there are at least three points of $Bd\, D$ that f sends into $D_1 \cdot D_2$

Let g be a map of $f^{-1}(D_1 \cdot D_2 \cdot f(Bd\, D))$ into $E_1 \cdot E_2$ given by $g = hf$. For each arc ab of $Bd\, D$ which intersects $f^{-1}(D_1 \cdot D_2 \cdot f(Bd\, D))$ only in its end points, extend g to map the chord acb of D onto an arc in E_i if $f(ab) \subset D_i$ with the restriction that the diameter of

$g(acb)$ is not more than twice the diameter of any other arc in E_i from $g(a)$ to $g(b)$. Let E be the subdisk of D such that g has been defined to map $Bd\,E$ into $E_1 + E_2$.

Since $E_1 + E_2$ is simply connected, we extend g to map E into $E_1 + E_2$. Call the extension g. Consider $g^{-1}(E_1 \cdot E_2 \cdot g(E)) = X$. No two points of $Bd\,D$ can be joined by an arc in $D - X$ unless the points go into the same one of D_1, D_2 under f.

Define f on X to be $h^{-1}g$. Let f_i be the extended f restricted to $f^{-1}(D_i \cdot f(X + Bd\,D))$. Extend f_i to map D into D_i and call the extended map f_i. The extended map f is f_1 on each component of $D - X$ which has points of $Bd\,D$ which are sent by f into D_1 and is f_2 on the rest of D.

6. **Adding continuous curves.** What we have learned about the sum of disks partially applies to the sums of other continua. If the intersection of two disks is so bad as to make the sum of the disks not simply connected, it is bad enough to keep any two continuous curves whatever with the same intersection from being simply connected. The following example illustrated in Figure 6 shows that the converse is not true.

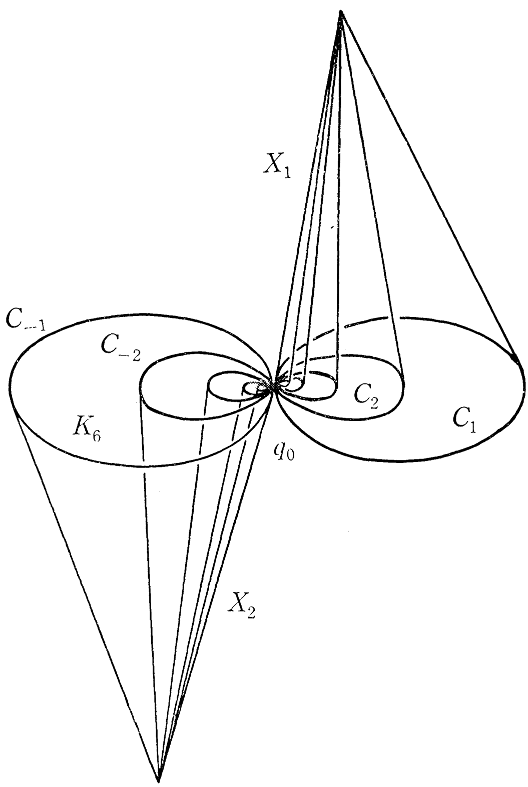


Figure 6

EXAMPLE 6. Let C_i $(i = 1, 2, 3, \cdots)$ be the circle in the x, y plane with equation $(x - 1/i)^2 + y^2 = (1/i)^2$. Denote the origin by q_0. Let $K_6 = C_1 + C_2 + \cdots + q_0 + C_{-1} + C_{-2} + \cdots$. The cone X_1 over $C_1 + C_2 + \cdots + q_0$ from a point above the xy plane is simply connected as is the cone X_2 over $q_0 + C_{-1} + C_{-2} + \cdots$ from a point below the xy plane. Although $X_1 \cdot X_2$ is a point, $X_1 + X_2$ is not simply connected.

THEOREM 14. *Suppose D_1, D_2 are two disks and F_1, F_2 are two continuous curves such that $D_1 \cdot D_2$ is homeomorphic with $F_1 \cdot F_2$. Then $D_1 + D_2$ is simply connected if $F_1 + F_2$ is.*

Proof. The proof is the same as the proof of Theorem 13 except that g maps $Bd\, E$ into $F_1 + F_2$ instead of into $E_1 + E_2$.

THEOREM 15. *Suppose G_1, G_2, G_3, G_4 are four simply connected and uniformly locally simply connected continuous such that $G_1 \cdot G_2$ is topologically equivalent to $G_3 \cdot G_4$. Then the fundamental group of $G_1 + G_2$ is isomorphic to the fundamental group of $G_3 + G_4$.*

Proof. Whether or not a loop in $G_1 + G_2$ can be shrunk to a point depends on how it crosses back and forth between G_1 and G_2. Suppose h is a homeomorphism of $G_1 \cdot G_2$ onto $G_3 \cdot G_4$ and x_0 is a point of $G_1 \cdot G_2$ that acts as a starting point of loops in $G_1 + G_2$ to determine the fundamental group of $G_1 + G_2$. We use $h(x_0)$ as a starting point for the loops in $G_3 + G_4$ to determine the fundamental group of $G_3 + G_4$.

Let $\{f\}$ be an element of the fundamental group of $G_1 + G_2$. It is an equivalence class of maps of the interval $[0, 1]$ into $G_1 + G_2$ such that the ends of $[0, 1]$ go into x_0. Let f be an element of $\{f\}$. Let f' be a map of $[0, 1]$ into $G_3 + G_4$ such that

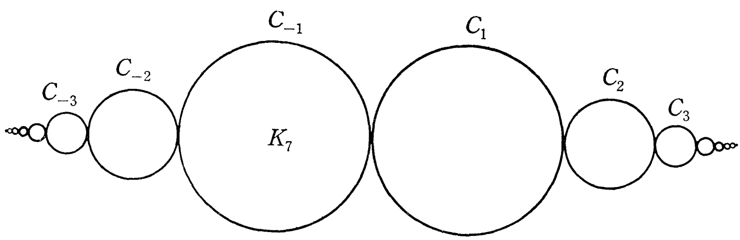
$$f'(x) = hf(x) \text{ if } f(x) \in G_1 \cdot G_2 ,$$
$$f'(x) \in G_{i+2} \text{ if } f(x) \in G_i .$$

Although these two conditions do not precisely define f', any two maps satisfying this condition are homotopic in $G_3 + G_4$. The element of the fundamental group of $G_3 + G_4$ corresponding to the element $\{f\}$ of the fundamental group of $G_1 + G_2$ is the equivalence class of loops containing f'.

Question. The preceding theorem suggests a topological invariant of compact closed sets. Two sets A, B are alike in a certain sense provided the sum of two Hilbert cubes sewed together along A have the same fundamental group as the sum of two Hilbert cubes sewed together along B. Is there a simpler characterization of this property?

7. An interesting group. One might attempt to compute the fundamental group of Example 1 by cutting it into two pieces with a vertical plane through the origin, fatten each piece to make them intersect in an open subset of their sum, find the fundamental group of each piece, and then apply Van Kampen's theorem to get the fundamental group of Example 1. We ignore the fattening since, being equivalent to taking the slice slightly to one side of the origin, it does not change the fundamental group of the pieces.

Each piece can be folded like a fan and deformed onto a set topologically equivalent to a set K_7 shown in Figure 7 and defined as follows.



Fignre 7

$$K_7 = \text{closure of } (C_1 + C_2 + \cdots + C_{-1} + C_{-2} + \cdots)$$

where C_i is the circle in the xy plane with $[(i-1)/i, \ i/(i+1)]$ as diameter and C_{-i} is the circle with $[(-(i-1)/i, \ -i/(i+1)]$ as diameter. (We used $[a, b]$ to denote the interval on the x axis from a to b.) The fundamental group of each piece into which we cut Example 1 is the same as the fundamental group of K_7.

Consider the origin as the starting point of loops in K_7 to determine its fundamental group $G(K_7)$. Then a loop is a map of the interval $[0, 1]$ into K_7 that sends the ends of the interval to the origin and an element of $G(K_7)$ is an equivalence class of loops. We can associate words with loops. If a loop goes across the top semicircle of C_i from left to right we write i; if it goes across this semicircle from right to left we write $\bar{i}$ and say i inverse. We call i and $\bar{i}$ letters and say that the letter is positive or negative according as i is positive or negative. Since we are permitting C_i's with negative subscripts, i inverse differs from $-i$. The inverse of $\bar{i}$ is i. A loop then corresponds to an ordered collection of letters (called a word) with the following restrictions.

a. No letter appears in any word more than a finite number of times.

b. There is not infinite oscillation between positive and negative letters.

Let us consider what words are equated if two pieces into which we divided Example 1 are joined together again. If a loop is slid from one piece to the other until it comes back to the first in one direction, each letter i (or $\bar{i}$) in it has been changed to $i+1$ (or $\overline{i+1}$) and if the loop is slid in the other direction, these are replaced by $i-1$ (or $\overline{i-1}$). Since we skipped 0 in putting subscripts on the C_i's we suppose $-1+1=1$ and $1-1=-1$. When we replace each i or $\bar{i}$ in a word W by $i+1$ or $\overline{i+1}$, we have produced a right shift and call the new word $R(W)$. We note that if $W_1 = R(W_2)$, then W_2 may be obtained from W_1 by a left shift and say $W_2 = \bar{R}(W_1)$.

Let us change the group G {equivalence classes of W_α's} by also putting words in the same equivalence class if they are equivalent after a shift. This shifting operation is to be permitted in equating words only a finite number of times as opposed to cancellation which was permitted infinitely often. We call the resulting group G {equivalence classes of W_α's$/R(W_\alpha) = W_\alpha$}. Since the fundamental group of Example 1 is trivial, it follows that this group is trivial.

The inverse of a word is obtained by reversing the order of the letters and replacing each letter with its inverse. If in a word there appears two adjacent subwords which are inverses of each other, the word obtained by canceling the subwords belongs to the same equivalence class with the original word. Infinite cancellation is permitted so that for example $(1, \bar{1}, 2, \bar{2}, 3, \bar{3}, \cdots)$ is equivalent to the trivial word.

Two words are equivalent if and only if they can be cancelled down to a common word. (We could have given more extensive rules but they boil down to this.) To multiply two words, we write one after the other. If $\{W_\alpha\}$ denotes the collection of words and G {equivalence classes of W_α's} denotes the group of equivalence classes of words, then

$$G(K_7) = G\{\text{equivalence classes of } W_\alpha\text{'s}\} .$$

To show algebraically that G {equivalence classes of W_α's$/R(W_\alpha) = W_\alpha$} is trivial, consider a word W. Since we did not permit infinite oscillation between positive and negative letters of W, we can express W as $W_1 W_2 \cdots W_n$ where each W_i has either all positive or all negative letters. We show that W is trivial by showing that each W_i is. We consider only the case where W_i consists of positive letters since the other case is analogous.

Consider $X = \bar{W}_i R(\bar{W}_i) R^2(\bar{W}_i) R^3(\bar{W}_i) \cdots$. It is a word since it only contains positive letters and none appears more than a finite number of times. Then

$$W_i = W_i X \bar{X} = R(\bar{W}_i) R^2(\bar{W}_i) \cdots \bar{X} = R(X)\bar{X} = X\bar{X} = 1 .$$

One might wonder what would have happened if we had not imposed the condition that there is not infinite oscillation between the positive and negative letters in words. This would have been equivalent to the fundamental group of K_7 after the sum of the bottom simicircles were shrunk to a point. Even after a shift, it seems that the group is not killed. After the shift we would have the fundamental group of Example 1 if the annulus in D_2 outside $\rho = a$ is shrunk to the circle $\rho = a$.

REFERENCES

1. R. H. Bing and F. B. Jones, *Another homogeneous plane continuum,* Trans. Amer. Math. Soc., **90** (1959), 171-192.

2. L. Fearnley, *Continuous mappings of the pseudo-arc,* Notices of the American Mathematical Society Abstract 564-171, **6** (1959), 841.

3. M. K. Fort, Jr., *Images of plane continua,* Amer. J. Math., **81** (1959), 541-546.

4. A. Lelek, *On weakly chainable continua,* Fundamenta Mathematicae, **51** (1962), 271-282.

5. J. Mioduszewski, *A functional conception of snake-like continua,* Fundamenta Mathematicae, **51** (1962), 179-189.

6. R. L. Moore, *Foundations of Point Set Theory,* Amer. Math. Soc. Colloq. Publ., vol. 13, revised edition, 1962.

THE UNIVERSITY OF WISCONSIN
THE INSTITUTE FOR ADVANCED STUDY

LENGTH-PRESERVING MAPS

HERBERT BUSEMANN

1. Introduction. If any two points of the metric space R can be connected by a rectifiable curve then a map of R into a metric space R' is *length-preserving or equilong*, if the length of any curve in R equals that of its image in R'. An equilong map of R means such a map of R into itself.

Folding a piece of paper repeatedly and in different ways exhibits a great variety of equilong maps of the euclidean plane. The original purpose of the present investigation was to determine all equilong maps which are not too pathological of the euclidean spaces and to find out whether other interesting[1] spaces admit length preserving maps which are not isometries.

However, equilong maps are connected with other important concepts. If the metric of R is intrinsic, i.e., if the distance of any two points equals the infimum of the lengths of all curves in R connecting these points, then an equilong map α of R into a metric space R' does not increase distance: $xy \geqq \alpha x \alpha y$. We denote as *shrinkage* any map of a metric space R into R' satisfying this inequality. Shrinkages which are not equilong enter significantly many branches of mathematics.[2] In fact, the linguistically preferable term "contraction" was avoided here, because it is widely used in functional analysis for the special shrinkages satisfying $xy \geqq k\alpha x \alpha y$ with $k > 1$ (see, for example, [5]). Therefore it seemed worthwhile to study the elementary properties of shrinkages as such.

On the other hand, isometries and local isometries are most important special maps (the latter in the theory of covering spaces) which preserve length. Our results on equilong maps will allow us to weaken the hypotheses in various theorems concerning (local) isometries. It often turns out that the axioms for a G-space (see [1]) need not all be satisfied and that a map can be proved to be onto where hitherto this had been assumed.

As to our original aims: we will *determine all locally finite equilong maps of the euclidean, hyperbolic, and spherical spaces.* The maximal open connected sets on which an equilong map is in-

Received May 22, 1963. This work was supported by a grant from the National Science Foundation.

1 "Interesting" is an essential qualification because there are many spaces with isolated equilong maps.

2 Among the less known applications, the shrinkages of cones on certain surfaces constructed by Rešetnyak [6] deserve special mention, because they yield elegant solutions of extremal problems in differential geometry.

jective are—in contrast to fundamental sets—uniquely determined. They are convex and their closures cover the space. Local finiteness refers to this covering. We thus obtain a division of the space into convex polyhedral regions $D_1, D_2, \cdots$, from which the equilong map is easily reconstructed.

A locally finite division of the space into convex polyhedral regions $D_1, D_2, \cdots$ belonging to an equilong map is *characterized* by the following property which is appealing through its simplicity (although the proof is not simple): *the number of* $(n-1)$*-dimensional faces of the* D_i *having a common* $(n-2)$*-face is even and if* $\delta_1, \cdots, \delta_{2k}$ *are the angles between these* $(n-1)$*-faces in cyclic order then*

$$\delta_1 + \delta_3 + \cdots + \delta_{2k-1} = \delta_2 + \delta_4 + \cdots + \delta_{2k} .$$

Because the existence of length preserving maps implies homogeneity properties of the space, the most interesting spaces from the point of view of these maps are those which possess large groups of motions. We will see that *neither general Minkowski spaces nor the hermitian and quarternion elliptic or hyperbolic spaces admit other locally finite equilong maps than motions.*

The initial stages of this work profited from discussions of the author with Professor G. Tallini in Rome.

2. Shrinkages. For purposes of comparison we define an *expansion* of one metric space R into another, R', as a map β satisfying $xy \leqq \beta x\beta y$. A shrinkage of R into R' is continuous but need not be injective. An expansion β is injective, the inverse map β^{-1} of $\beta(R)$ on R is a shrinkage and hence continuous, but β may be nowhere continuous. For $R = R'$ we speak of a shrinkage or an expansion of R.

The symbol (x, y, z) means that $x \neq y$, $y \neq z$ and $xy + yz = xz$. We begin with a simple observation concerning the *displacement* $x\alpha x$ of a point under a shrinkage. This function is continuous because

$$| x\alpha x - y\alpha y | \leqq xy + \alpha x\alpha y \leqq 2xy .$$

(1) *Let* α *be a shrinkage of* R. *If* $(p, x, \alpha p)$ *then* $x\alpha x \leqq p\alpha p$ *with equality only when* $px = \alpha p\alpha x$ *and* $(x, \alpha p, \alpha x)$.

If $(x, p, \alpha p)$ *then* $x\alpha x \geqq p\alpha p$ *with equality only when* $\alpha x\alpha p = xp$ *and* $(x, \alpha x, \alpha p)$.

For $(p, x, \alpha p)$ gives

$$x\alpha x \leqq x\alpha p + \alpha p\alpha x \leqq px + x\alpha p = p\alpha p$$

and $x\alpha x = p\alpha p$ implies first that $\alpha p\alpha x = px > 0$ and then $(x, \alpha p, \alpha x)$. From $(x, p, \alpha p)$ we conclude

$$x\alpha x \geqq x\alpha p - \alpha x\alpha p \geqq x\alpha p - xp = p\alpha p > 0 .$$

(2) *If β is an expansion of R and $(p, \beta p, x)$ then $x\beta x \geqq p\beta p$ with equality only when $\beta p\beta x = px$ and $(\beta p, x, \beta x)$.*

This follows from

$$x\beta x \geqq \beta p\beta x - \beta px \geqq px - \beta px = p\beta p > 0 .$$

The length of the curve C: $x(t)$ $(a \leqq t \leqq b)$ in a metric space (see [1] or [7]) is denoted by $L(C)$. Obviously,

(3) *If α is a shrinkage of R in R' and C: $x(t)$ is a rectifiable curve in R and αC: $\alpha x(t)$ is its image then $L(\alpha C) \leqq L(C)$.*

If C is the curve $x(t)$ $(a \leqq t \leqq b)$ and we put $x(a) = u$, $x(b) = v$ then $L(C) \geqq uv$. If the equality sign holds we call C a *segment* $T(u, v)$ from u to v, because $T(u, v)$ is isometric to an interval of length uv on the real axis ([1] or [7]|).

(4) *If α is a shrinkage of R in R' and $uv = \alpha u\alpha v$ then α maps a segment $T(u, v)$ isometrically on a segment $T(\alpha u, \alpha v)$.*

For if $z \in T(u, v)$ then

$$uv = uz + zv \geqq \alpha u\alpha z + \alpha z\alpha v \geqq \alpha u\alpha v = uv ,$$

whence $uz = \alpha u\alpha z$ and $zv = \alpha z\alpha v$. If w is a fourth point of $T(u, v)$, say on the subsegment $T(u, z)$, then it follows from what we just proved that $wz = \alpha w\alpha z$. This yields:

(5) *If p and q are fixed points of a shrinkage of R and if exactly one segment $T(p, q)$ exists then α leaves all points of $T(p, q)$ fixed.*

Thus the fixed points of a shrinkage of a euclidean or hyperbolic space form a set which is empty or convex and closed.

A ray in a metric space R is the isometric image of the non-negative real axis and hence may be represented in the form $p(t)$ $(t \geqq 0)$ with $p(t_1)p(t_2) = |t_1 - t_2|$. We prove:

(6) *Assume that for any two points x, y in R a segment $T(x, y)$ exists and that (w, x, y) and (x, y, z) imply (w, x, z). If α is a*

shrinkage of R and the displacement attains at p a positive minimum $\left(\text{i.e., } p\alpha p = \inf_{x \in R} x\alpha x > 0\right)$, *then p is the origin of a ray which* α *translates into itself.*

Conversely, if p is the origin of a ray which α *translates (properly) into itself then* $p\alpha p = \inf_{x \in R} x\alpha x$.

Let T be a segment from p to αp and x an interior point of T. Put $T_i = \alpha^i T$, $p_i = \alpha^i p$, $x_i = \alpha^i x$, $(i = 0, 1, \cdots)$, $p\alpha p = \rho$. Then (1) yields $x\alpha x = \rho$ and (x, p_1, x_1). Hence by hypothesis (p, p_1, x_1). Applying (1) to x and x_1 we obtain (p_1, x_1, p_2) hence (p, x_1, p_2) and $xx_1 = p_1p_2 = pp_1$. From (4) we conclude that T_1 is a segment from p_1 to p_2 and from (x, p_1, x_1) and (p, x_1, p_2) that $T \cup T_1$ is a $T(p, p_2)$. Continuing in this manner we prove that $\bigcup_{i=0}^{\infty} T_i$ is a ray $p(t)$ $(t \geqq 0)$ with $\alpha p(t) = p(t + \rho)$.

Conversely, if α induces the translation $\alpha p(t) = p(t + \rho)$ $(\rho > 0)$ of the ray $p(t)$, then for an arbitrary point x

$$px + nxx_1 + xp \geqq px + \sum_{i=1}^{n} x_{i-1}x_i + x_np_n \geqq pp_n = npp_1 .$$

Dividing by n and letting $n \to \infty$ we obtain $xx_1 \geqq pp_1$.

In order to see how (6) can be applied we prove

(7) *Let R be a convex subset of a Banach space with strictly convex spheres and* α *a shrinkage of R. The set S of the points where the displacement* $x\alpha x$ *takes its minimal value is either empty or convex. If the minimum is positive then S is the union of parallel rays and* α *coincides on S with a translation* δ *of the space.*

If the minimum is 0 we know from (5) that S is convex. If $p\alpha p = \inf x\alpha x = \rho > 0$ then, by (6) α induces a translation on a ray S_p with origin p. If also $q\alpha q = \rho$ then the says S_p and S_q are parallel because otherwise $\alpha^i p\alpha^i q \to \infty$ for $i \to \infty$ whereas $\alpha^i p\alpha^i q \leqq pq$. Therefore α coincides on $S_p \cup S_q$ with an ordinary translation δ of the space. This implies $pq = \alpha p\alpha q$ and it follows from (4) that α maps $T(p, q)$ isometrically on $T(\alpha p, \alpha q)$. (Segments are unique because the spheres are strictly convex) Therefore $x\alpha x = p\alpha p$ for $x \in T(p, q)$, whence $T(p, q) \subset S$ and $\alpha x = \delta x$ on S.

We add an observation which is of interest in connection with contraction maps (see Introduction).

(8) *For any two points* x, y *of R let a segment* $T(x, y)$ *exist and let* α *be a map of R in itself satisfying* $xy > \alpha x\alpha y$ *for* $x \neq y$. *If the displacement* $x\alpha x$ *attains a relative minimum at f, then f is*

a fixed point (so that the minimum is absolute) *and there is no other fixed point. If $x\alpha x$ attains at g a relative maximum then no point z with $(z, g, \alpha g)$ exists.*

If $f \neq \alpha f$ then by (1) any point x on a $T(f, \alpha f)$ would satisfy either $x\alpha x < f\alpha f$ or $fx = \alpha f \alpha x$. Both relations contradict the hypothesis. A second fixed point f' satisfies $ff' = \alpha f \alpha f'$ hence $f' = f$. If a point z with $(z, g, \alpha g)$ existed then a segment $T(z, g)$ would contain a point x with $(x, g, \alpha g)$ and arbitrarily small xg. But (1) would yield either $xg = \alpha x \alpha g$ or $x\alpha x > g\alpha g$.

If R is a differentiable manifold with a Riemann or Finsler metric then the nonexistence of z means that g lies on the so called cut locus or minimum point locus of αg. For compact R both f and g exist. If R is a spherical, or more generally, a spherelike ([1, p 128]) space, then g is also unique because αg is then the antipode to g and any two points and their antipodes have equal distance.

As mentioned in the introduction, the metric of R is called *intrinsic* if any two points x, y of R can be connected by a curve of finite length and

$$xy = \inf_{C_{xy}} L(C_{xy}) ,$$

where C_{xy} traverses all curves from x to y. If in such a space a curve of minimal length from x to y exists it is a segment $T(x, y)$.

Denote by $S(p, q)$ $(p > 0)$ the set of all points x for which $px < \rho$. If the metric of R is intrinsic and the closure $\bar{S}(p, 2\rho)$ of $S(p, 2\rho)$ is compact then a segment $T(x, y)$ exists for any two points x, y in $\bar{S}(p, \rho)$. In particular, if R is compact or finitely compact (which means that all $\bar{S}(p, \rho)$ are compact) then a $T(x, y)$ exists for arbitrary x, y. (These facts are implicit in the results of [1] and [7, p 142])

If the metric of R is intrinsic the following converse of (4) holds:

(9) *If α is a shrinkage of the space R with an intrinsic metric in the space R', moreover $uv = \alpha u \alpha v$ and the image αC of the curve C from u to v is a segment $T(\alpha u, \alpha v)$ then C is a $T(u, v)$ and α maps C isometrically on αC.*

For if C were not a $T(u, v)$, a curve C_0 from u to v with $L(C_0) < uv$ would exist and it would follow that

$$\alpha u \alpha v \leqq L(\alpha C_0) \leqq L(C_0) < uw .$$

(10) *Let R be a space with an intrinsic metric and α a con-*

tinuous map of R in R′. Then α is a shrinkage if and only if, $L(C) \geqq L(\alpha C)$ *for any curve C in R.*

The necessity follows from (3). Let C be a curve from x to y with $L(C) < xy + \varepsilon$; then

$$xy + \varepsilon > L(C) \geqq L(\alpha C) \geqq \alpha x \alpha y$$

proves the sufficiency.

An important corollary of (9) was already mentioned:

(11) *A length preserving map of a space R with an intrinsic metric into a space R′ is a shrinkage.*

It is clear that a shrinkage of a noncompact space onto itself or of a compact space into itself need not be an isometry. However Freudenthal and Hurewicz proved in [4]:

(12) *A shrinkage of a compact space onto itself is a motion.*[3]

In conjunction with (11) this yields:

(13) *A length preserving map of a compact space with an intrinsic metric onto itself is a motion.*

In particular, a locally isometric map of a compact G-space onto itself is a motion, a fact which the author proved in [1, p 172] without being aware of the paper [4]. A much more interesting generalization is given in (19).

It may be useful to emphasize that in (13) compactness cannot be replaced by finite compactness. If $(x_1, \cdots, x_{n-1}, z) = (x, z)$ are cartesian coordinates in E^n then the relations

$$\alpha(x, z) = \begin{cases} (x, z+1) \text{ for } z < 0 , \\ (x, 1-z) \text{ for } 0 \leqq z \leqq 1 , \\ (x, z-1) \text{ for } z > 1 , \end{cases}$$

define an equilong map of E^n on itself which is not a motion.

Since the inverse of an expansion is a shrinkage, (12) is valid also for expansions. However, according to [4] a stronger statement holds:

(14) *An expansion of a compact space into itself is a motion.*

[3] A motion of R is an isometry of R onto itself.

In particular:

(15) *An isometry β of a compact space R into itself is a motion.*

Since we will apply (15) we give a short proof: If $p \in R - \beta R$ existed then putting $p = p_0$, $p_i = \beta^i p$, $R = R_0$, $R_i = \beta^i R$ we would have (for $k \geqq 1$) $p_i \in R_i - R_{i+1}$, $p_i p_{i+k} = pR_k \geqq pR_1 > 0$, and the p_i would not have an accumulation point.

From (12) interesting results on special spaces may be obtained. For example:

(16) *Let α be a shrinkage of the spherical space S^n which is not a motion. Then α has at least one fixed point, maps at least one point on its antipode, and sends at least one pair of antipodes into the same point. $\alpha(S^n)$ lies in a closed hemisphere.*

The first two statements follow from well known topological facts, because by (12) the degree of the mapping α is zero. They can also be seen directly: If αx were never antipodal to x then the point x_t on $T(x, \alpha x)$ with $xx_t = t(x\alpha x)$ $(0 \leqq t \leqq 1)$ would be well defined and depend continuously on x, so that $x \to x_t$ would by (12) yield a continuous deformation of S^n into a proper subset. If δ is the antipodal map then $\delta\alpha$ is a shrinkage, hence maps some point u on its antipode, and $\alpha u = u$.

If α is not a motion then it may by (12) be regarded as a map of S^n into the n-dimensional euclidean space, and it follows from the Theorem of Borsuk and Ulam see [3, p. 337], that α sends at least one pair a, a' of antipodal points into the same point b.

If S^n has curvature 1, then any point x satisfies $\min(xa, xa') \leqq \pi/2$. Therefore $\alpha x b \leqq \pi/2$ for all x and $\alpha(S^n)$ lies in the hemisphere of S^n with center b.

3. Locally injective equilong maps. From now on we concentrate on length preserving maps. In particular we study regions in which these maps are injective. At each stage we will make it clear which properties on the spaces enter. The first is domain equivalence. We say:

Domain equivalence holds for two topological spaces R, R' if the topological image in R' (R) of an open set in $R(R')$ is open in $R'(R)$.

For $R = R'$ we follow the classical terminology and speak of *domain invariance* rather than equivalence. Brouwer's theorem states that domain equivalence holds for any two topological manifolds of the same dimension. Therefore all manifolds considered in differential geometry have the property of domain invariance. The finite dimensional, and probably all, G-spaces have this property.

Some simple *examples* will help to elucidate this concept and the facts which we are going to prove. First let R be the set in E^3 with cartesian coordinates x, y, z consisting of the plane $z = 0$ and the line $x = y = 0$. The metric of R is here and in the second example the intrinsic metric induced by the euclidean metric in E^3. The interval $1 < z < 3$, $x = y = 0$ is isometric to the interval $1 < y < 3$, $x = z = 0$ but the former is the sphere $S((0, 0, 2), 1)$ and is open, the second is not. Domain invariance does not hold in R. The map α of R in itself defined by

$$\alpha(x, y, 0) = (x, y, 0), \qquad \alpha(0, 0, t) = (0, t, 0)$$

preserves length and takes the first interval into the second.

Next take R as the set in E^3 consisting of the three coordinate axes. The interval $x = y = 0$, $1 < z < 3$, is isometric to the interval $x = y = 0$, $-1 < z < 1$. The former is open in R, the second is not. The map α which leaves all points on the x-axis, on the y-axis and the points $z \leqq 1$ on the z-axis fixed and maps $(0, 0, z)$ with $z > 1$ on $(0, 0, 1-z)$ preserves length and takes the first interval into the second.

For later purposes we point out that in both these spaces motions exist which are not the identity and leave $S((0, 0, 2), 1)$ pointwise fixed.

As third example we take the ray $t \geqq 0$ with the metric $|t_1 - t_2|$. The set $0 \leqq t < 1$ is isometric to $1 \leqq t < 2$, the first is open, the second is not. $t' = t + 1$ takes the first set into the second and is an isometry, but is not a motion because it does not map R onto itself.

Denoting the restriction of a map α to a set M by α_M we say that α is injective on M if α_M maps M bijectively on $\alpha(M)$. We prove

(17) *Let R, R' be locally compact spaces with intrinsic metrics and domain equivalence and α an equilong map of R in R' which is injective on the open set G. Then for every point p of G a positive ρ_p exists such that α_G (and hence α) maps $S(p, \rho_p)$ isometrically on $S(\alpha p, \rho_p)$.*

Choose $\delta > 0$ such that $\bar{S}(p, \delta)$ is compact and lies in G. Then α_G maps $\bar{S}(p, \delta)$ topologically on $\alpha_G\bar{S}(p, \delta) = \alpha\bar{S}(p, \delta)$. It follows from domain equivalence that $\alpha S(p, \delta)$ is open, hence $\rho = \rho_p > 0$ exists such that $\bar{S}(p', 2\rho)$ ($p' = \alpha p = \alpha_G p$) lies in $\alpha S(p, \delta)$ and is compact.

For any two points x', y' in $\bar{S}(p', \rho)$ there is a segment $T(x', y')$ which is contained in $\bar{S}(p, 2\rho)$. Since α_G is topological in $\bar{S}(p, \delta)$ and preserves length $\alpha_G^{-1}T(x', y')$ is a curve from $x = \alpha_G^{-1}x'$ to $y = \alpha_G^{-1}y'$ of length $L(T(x', y')) = x'y' \geqq xy$, hence by (11) $x'y' = xy$. In particular $p'x' = px$. Therefore $\alpha_G^{-1}\bar{S}(p' \rho)$ is isometric to $\bar{S}(p, \rho)$ and contained in $\bar{S}(p, \rho)$. It follows from (15) that $\bar{S}(p, \rho) = \alpha_G^{-1}\bar{S}(p', \rho)$, which proves (17). Our examples show that (17) is not valid without the hypothesis of domain equivalence.

A map β of R in R' is *locally injective* if every point of R has a neighborhood on which β is injective. Adhering to the terminology of [1] we do not use the strict analogue to define local isometries but require a little more: The map α of R in R' is *locally isometric* if for every point p of R a positive ρ_p exists such that α maps $S(p, \rho_p)$ isometrically on $S(\alpha p, \rho_p)$. We now prove the important fact

(18) THEOREM. *If R and R' are finitely compact spaces with intrinsic metrics and domain equivalence then a locally injective equilong map of R into R' is a local isometry of R onto R'.*

Our third example shows that (18) does not hold without the hypothesis of domain equivalence even if $R' = R$.

Proposition (17) yields the existence of a positive function σ_p defined in R such that α maps $\bar{S}(p, \rho_p)$ isometrically on $\bar{S}(\alpha p, \rho_p)$ and hence $S(p, \rho_p)$ on $S(\alpha p, \rho_p)$. Let $\rho(p)$ be the supremum of the numbers ε for which α maps $\bar{S}(p, \varepsilon)$ isometrically on $\bar{S}(\alpha p, \varepsilon)$. If $\rho(p) = \infty$ then α is an isometry of R on R' and $\rho(x) = \infty$ for all x in R. If $\rho(p) < \infty$ then $S(p, \varepsilon) \supset S(q, \varepsilon - pq)$ for $pq < \varepsilon$ shows that $|\rho(p) - \rho(q)| \leqq pq$. Therefore $\rho(p)$ is a positive continuous function which has a positive minimum on every $\bar{S}(x, \sigma)$ (which is compact by hypothesis).

Let q be any point of R and $q' = \alpha q$. We must prove that for a given point r' of R' a point r in R exists with $\alpha r = r'$. Because R' is finitely compact and has an intrinsic metric there is a segment T' from q' to r'. Let $\delta > 0$ be the minimum of $\rho(p)$ for $p \in \bar{S}(q, q'r')$ and choose $q'_0 = q', q'_2, \cdots, q'_n = r'$ on T' such that $(q'_{i-1}, q'_i, q'_{i+1})$ and $q'_{i-1}q'_i < \delta$.

Since α maps $\bar{S}(q, \delta)$ isometrically on $\bar{S}(q', \delta)$ there is a segment $T(q, q_1)$ in $S(q, \delta)$ which α maps isometrically on the subsegment $T(q', q'_1)$ of T'. For the same reason α maps a suitable segment $T(q_1, q_2)$ in $S(q_1, \delta)$ on the subsegment $T(q'_1, q'_2)$ of T'. Thus we arrive at a segment $T(q_{n-1}, q_n)$ mapped on the subsegment $T(q'_{n-1}, q'_n) =$

$T(q'_{n-1}, r')$ of T'. With $q_n = r$ we have $\alpha r = r'$. It follows, by the way, from (9) that $\bigcup_{i=1}^{n} T(q_{i-1}, q_i)$ is a $T(q, r)$.

Notice the following application of (13) and (18):

(19) THEOREM. *If R is a compact space with an intrinsic metric and domain invariance, then a locally injective length preserving map of R in itself is a motion.*[4]

Compactness in (19) cannot be replaced by finite compactness, see [1, p 173], but there are various conditions under which it can. Define $\rho(p)$ as in the proof of (18). We introduce the condition

$$(*) \qquad \inf_{p \in \alpha^{-1}(p')} \rho(p) > 0 \text{ for each } p' \in \alpha R,$$

which holds when both R and R' are G-spaces [1, p 171]. Under the hypotheses of (18) R and R' are arcwise and locally arcwise connected and (*) guarantees that R is a covering space of R'. Therefore we have (compare [1, p 174])

(20) *Let R and R' be finitely compact spaces with intrinsic metrics and domain equivalence. If the locally injective length preserving map α of R in R' satisfies (*) and the fundamental group of R is not isomorphic to a proper subgroup of that of R', then α is an isometry of R on R'*

Papers [8], [9] and the last part of [2] deal with conditions which can replace the requirement on the fundamental groups. From (18) and [2] it may be deduced, for example, that a locally injective equilong map of any (complete) locally Minkowskian space into itself is a motion.

4. Regions of injectivity. A *region of injectivity* of the map β of the space R into the space R' is a maximal open connected set on which β is injective. If β is a locally isometric map of R on R' then such a region is what is usually called a fundamental domain (or, depending on the terminology, its interior). We are here interested in regions of injectivity of equilong maps of spaces in themselves. No interesting statements are possible unless the space has special properties, in particular besides domain invariance, one or more of the following three:

---

[4] The following example shows that domain invariance is necessary: The space consists of the origin 0 and the circles $-2 \cdot 3^{-n} + 3^{-n}e^{i\theta}$ $(n = 0, 1, 2, \cdots)$ of the complex plane; the metric is given by arclength. Then α defined by $\alpha(0) = 0$ and $\alpha(-2 \cdot 3^{-n} + 3^{-n}e^{i\theta}) = -2 \cdot 3^{-n-1} + 3^{-n-1}e^{3i\theta}$ is locally injective and equilong but is neither onto nor isometric.

I. *A motion of the space which leaves all points of a nonempty open set fixed is the identity.*

All spaces considered in differential geometry and all G-spaces have this property, [1, p 178]. The first two examples in § 4 are spaces for which I does not hold.

II. *If* (x, y, z) *then the segment* $T(x, y)$ *is unique. Every point* p *has a neighborhood* $S(p, \delta)$ *such the segment* $T(x, y)$ *is unique for points* x, y *in* $S(p, \delta)$.

Again all the usual spaces of disfferential geometry and all G-spaces have this property. A Minkowski space of dimension greater than one does not have it unless its spheres are strictly convex.

III. *An isometry of a sphere* $S(p, \rho)$ *on a sphere* $S(q, \rho)$ $(\rho > 0)$ *can be extended to a motion of the space.*

This condition may be trivially satified, namely when no isometric spheres with distinct centers exist and the only isometry of $S(p, \rho)$ is the identy. Such a space satisfies I. If, in addition, domain invariance and property II hold, then our theory implies that its only locally finite (see below) equilong map is the identity.

All simply connected complete Riemann spaces with analytic metrics satisfy II. Particularly interesting among these are the elementary, i.e. the euclidean, hyperbolic, and spherical (dim >1) spaces, the hermitian or quaternion elliptic and hyperbolic spaces and the elliptic and hyperbolic Cayley planes. Apart from the elliptic spaces these are the only finitely compact G-spaces with *pairwise mobility*, which means: Given four points x, y, x', y' with $xy = x'y'$ then a motion exists which takes x into x' and y into y' (see [1] for the compact case and [10] for the general case.

Finally we mention that, because of the existence of dilations the Minkowski spaces also satisfy III.

In order not to interrupt our arguments later, we first prove a lemma:

(21) *Let* M *be a closed set with a nonempty interior* M_i *in a finitely compact space which has an intrinsic metric and satisfies II. If* M *contains with any two points* x, y *the segment* $T(x, y)$ *when it is unique, then* $M = \bar{M}_i$ *and with* x *and* y *the set* M *contains at least one, and* M_i *contains all,* $T(x, y)$.

Let $p \in M$, $q \in M_i$ and $T = T(p, q)$. We show first that $T - p \subset M_i$.

If $q' \in T$ is sufficiently close to, but different from, q then $T(q', q)$ lies in M_i, and $T(q', q)$, $T(q', p)$ are unique, hence are subsegments of T and lie in M. If a point of $T - p$ not in M_i existed then traversing $T(p, q')$ from q' towards p we would meet a first point $b \neq p$ not in M_i. Choose $\delta > 0$ such that $T(x, y)$ is unique for x, y in $S(b, \rho)$ and then u, v on T with (p, u, b), $bu < \rho$, (b, v, q') and $bv < \rho$. The segments $T(u, x)$ with $x \in M_i \cap S(b, \rho)$ lie in M, because they are unique and x, u lie in M. The set $\bigcup_x T(u, x) - u$ is open and contains b which would imply $b \in M_i$.

Thus $T - p \subset M_i$ whence $p \in \bar{M}_i$ and $M = \bar{M}_i$. Also, trivially, $T \subset M_i$ for $p \in M_i$, hence M_i contains all $T(p, q)$. If p, q are given points of M then sequences $p_1, p_2, \cdots$ and $q_1, q_2, \cdots$ in M_i exist tending to p and q respectively. Because $T(p_i, q_i) \subset M_i$ and $\{T(p_i, q_i)\}$ contains—by finite compactness—a subsequence tending to a segment $T(p, q)$, the latter lies in M. Next we observe:

(22) *Let R be a metric space which has a countable base and satisfies I and III. Then a bijective locally isometric map β of a connected open set G in R on an open set G' in R can be extended to a motion of R.*

By hypothesis each point p of G has a neighborhood $S(p, \rho_p)$ which β maps isometrically on $S(\beta p, \rho_p)$. Since G is connected and R has a countable base there is a sequence of points $p_1, p_2, \cdots$ in G such that β maps $S(p_i, \rho_i) = S(p_i, \rho_{p_i})$ isometrically on $S(\beta_{p_i}, \rho_i)$, $G = \bigcup_{i=1}^{\infty} S(p_i, \rho_i)$, and $S(p_{i+1}, \rho_{i+1}) \cap S_i \neq \phi$ where $S_i = \bigcup_{k=1}^{i} S(p_i, \rho_i)$.

It follows from III that the restriction of β to $S(p_i, \rho_i)$ is the restriction of a motion ν_i of R. It suffices to prove that $\beta = \nu_1$ on each S_i. The assertion is trivial for S_1. Assume it has been proved for S_n. Then $\nu_{n+1} = \beta = \nu_1$ on $S(p_{n+1}, \rho_{n+1}) \cap S_n$. Since this set is not empty it follows from I that $\nu_{n+1} = \nu_1$, in particular $\nu_{n+1} = \nu_1 = \beta$ on S_{n+1}. This result and (17) yield

(23) *Let R be a finitely compact space with an intrinsic metric and domain invariance satisfying I and III. If the equilong map α of R is injective on the connected open set G then α_G can be extended to a motion of R.*

We add the following:

(24) *If (under the hypotheses of (23)) G is a region of injectivity then no boundary point p of G has a neighborhood on which α is injective. Hence two distinct regions of injectivity are disjoint.*

For an indirect proof assume that α is injective on $S(p, \rho)$ $(\rho > 0)$. Put $S_n = S(p, \rho/n)$. For each n there would be a point $q_n \in S_n - G \cup S_n$ and a point $r_n \in G$ such that $\alpha r_n = \alpha q_n$. Otherwise α would be injective on $G \cup S_n$. Because α is injective on $S(p, \rho)$ we have $r_n p \geqq \rho$.

Let $q' \in S_n \cap G$. It follows from (22) that $\alpha q'_n \alpha r_n = q'_n r_n$. On the other hand

$$2\rho/n > q_n q'_n \geqq \alpha q_n \alpha q'_n \geqq \alpha q'_n \alpha r_n - \alpha q_n \alpha r_n = \alpha q'_n \alpha r_n = q'_n r_n$$
$$\geqq r_n p - p q'_n > \rho - \rho/n \,.$$

If two distinct regions G, D of injectivity were not disjoint then, since neither can be contained in the other, D would contain a boundary point p of G because it is connected.

We now come to an important fact which will enable us in the most interesting cases either to prove the nonexistence of equilong maps which are not motions or to construct all length preserving maps which are not too wild.

(25) THEOREM. *Let R be a finitely compact space with an intrinsic metric and domain invariance satisfying I, II and III. A region of injectivity D of an equilong map of R in itself contains with two points x, y all segments $T(x, y)$, its closure $\bar{D}$ therefore contains with x, y at least one $T(x, y)$. Moreover $\alpha_{\bar{D}}$ is injective and is the restriction of a motion of R to $\bar{D}$.*

By (22) there is a motion β of R extending α_D, i.e., $\beta x = \alpha_D x = \alpha x$ on D. Therefore $\beta^{-1}\alpha$ is an equilong map of R which leaves all points of D fixed. Denote the set of all fixed points of $\beta^{-1}\alpha$ by M. Then M is closed, contains D and $\beta^{-1}\alpha x = x$ or $\alpha x = \beta x$ on M hence α_M is injective. By (5) M contains $T(x, y)$ when x, y lie in M and $T(x, y)$ is unique. It follows from (21) that the interior M_i of M contains with any two points x, y all $T(x, y)$ and that $M = \bar{M}_i$. Since M_i is connected and contains D, moreover α is injective on M_i and D is maximal we conclude $M_i = D$.

Under the hypotheses of (25) we call α *locally finite* if R is the union of the closures of the regions of injectivity of α and if this covering of R is locally finite. There will then be a finite or countable number of regions of injectivity $D_1, D_2, \cdots$. For each D_i there is a motion β_i such that $\beta_i^{-1}\alpha$ leaves D_i pointwise fixed. Therefore studying the properties of α we may *assume that α leaves D_1 pointwise fixed.*

An ellipsoid R in E^2 with three axes of different lengths and with its intrinsic metric admits a finite number of equilong maps.

These are generated by the reflections in the three planes P_i containing two axes and the following maps β_i: If R_i, R'_i are the sets on R bounded by P_i then β_i leaves R_i pointwise fixed and maps R'_i on R_i by reflection in P_i. Obviously there is such a variety of spaces possessing isolated equilong maps that neither is it feasible nor would it be worthwhile to determine all spaces admitting equilong maps.

Clearly, the interesting spaces are those which possess large groups of motions. We are now going to examine such spaces for *proper equilong maps*, that is, length preserving maps which are not motions.

5. Spaces without proper equilong maps. The *one dimensional* cases, although trivial, are basic. The regions of injectivity of a proper locally finite equilong map α of the real axis are intervals or rays whose endpoints form a discrete set. We may assume that D_1 has a right endpoint x_2. Let $D'_2, D'_3, \cdots$ be the regions of injectivity to the right of x_2 in their natural order. Denote the left endpoint of D'_i by x_i and let R_i be the reflection $x' = 2x - x_i$ of the real axis in x_i. Then for $x > x_2$ the map α is given by

$$\alpha x = R_2 R_3 \cdots R_j x \text{ for } x \in D'_j . \tag{26}$$

The procedure is analogous for the D_j preceding D_1 (if any).

If the space is the unitcircle (with length as metric) the construction is similar. Orient the circle. There is a finite number of regions of injectivity for a proper locally finite equilong map α which are arcs and which we call $D_1, \cdots, D_m$ in the order of the orientation. Denote by R_i the reflection in the diameter of the circle passing through the left endpoint of D_i. Then we still have

$$\alpha x = R_2 R_3 \cdots R_j \text{ for } x \in D_j ,$$

but the D_j must satisfy two conditions. Their number m must be even. $m = 2k$, otherwise α would be injective on $D_m \cup D_1$. The right endpoint of D_m must stay fixed. If δ_i is the length of D_i this yields

$$\sum_{i=1}^{k} \delta_{2i-1} = \sum_{i=1}^{k} \delta_{2i} \qquad (=\pi) . \tag{27}$$

We want to establish that certain spaces with at least transitive groups of motions do not possess proper equilong maps.

(28) THEOREM. *A Minkowski space* R (dim $R = n \geqq 2$) *with strictly convex spheres*[5] *admits a locally finite proper equilong map, if, and*

[5] The validity of (28) is not contingent upon the strict convexity of the spheres. The latter is equivalent to II and hence necessary for applying (25).

only if, it possesses the reflection in some hyperplane.

Proof. Assume that R can be reflected in the hyperplane H. Then the reflections in all hyperplanes parallel to H also exist. Choose affine coordinates $x_1, \cdots, x_n$ such that x_i is Minkowski length on the x_i-axis, H is given by $x_n = 0$ and the line $x_1 = 0, \cdots, x_{n-1} = 0$ is normal to H in the Minkowski sense. Then all lines $x_i = \text{const.}$ $(i = 1, \cdots, n-1)$ are normal to all $x_n = \text{const.}$ Let $x_n \rightarrow \beta(x_n)$ be a locally finite equilong map of the x_n-axis on itself. Then

$$(x_1, \cdots, x_{n-1}, x_n) \rightarrow (x_1, \cdots, x_{n-1}, \beta(x_n))$$

defines an equilong map of R. Thus a considerable variety of equilong maps can be derived alone from the reflection in H.

Conversely, assume R possesses a locally finite proper equilong map α. Then at least two regions $D_1, D_2, \cdots$ of injectivity exist and by (25) all these are convex polyhedral regions. D_1, which by agreement is left pointwise fixed by α, has a boundary point p such that for a suitable $\rho > 0$ the sphere $S(p, \rho)$ intersects only $\bar{D}_1$ and a single other $\bar{D}_i$, say $\bar{D}_2$. Then α leaves the disk $S(p, \rho) \cap \bar{D}_1 \cap \bar{D}_2$ fixed and coincides on $S(p, \rho) \cap \bar{D}_2$ with a motion β of R. But β must be either the identity, which is impossible because then D_1 would not be a region of injectivity, or the reflection in the hyperplane carrying $\bar{D}_1 \cap \bar{D}_2$.

Then same can be proved for plane quasihyperbolic geometry (see [1, pp 360, 363, 371, 407]) and also for its higher dimensional analogues. Hyperplanes through arbitrary n points (if $\dim R = n > 2$) do in general not exist and the result must be interpreted to mean that a hyperplane H and the reflection in H exist.

Next we show that the spaces, which after the elementary and elliptic spaces, have the greatest degree of mobility, do not possess proper equilong maps:

(29) THEOREM. *The hermitian elliptic and hyperbolic spaces of (real) dimension greater than* 2, *the quaternian elliptic and hyperbolic spaces of dimension greater than* 4, *and the Cayley elliptic and hyperbolic planes*[6] *do not possess locally finite length preserving maps other than motions.*

None of the spaces in (29) have constant curvature. Therefore, using the result mentioned in the preceding section, it suffices to prove:

---

[6] The Cayley planes have dimension 16. The hermitian spaces of dimension 2 and the quarternion spaces of dimension 4 are elementary, see [**1**].

(30) *Let R be a finitely compact metric space with an intrinsic metric and a pairwise transitive group of motions. If R possesses a proper equilong map, then R has constant curvature.*

Even without the hypothesis that R admits a proper equilong map all spaces in question of dimension less than 4 have constant curvature, so that we may assume that $n = \dim R \geqq 3$. Moreover all spaces satisfy I, II, III.

As in the preceding proof there is a point p on the boundary of D_1 and a sphere $S(p, \rho)$ $(\rho > 0)$ such that $S(p, \rho) \subset \bar{D}_1 \cup \bar{D}_2$. We choose ρ so small that $T(x, y)$ is unique for $x, y \in S(p, \rho)$. Then $S(p, \rho)$ is homeomorphic to E^n. By (25) the set

$$N = \bar{D}_1 \cap \bar{D}_2 \cap S(p, \rho)$$

contains with any two points the segment $T(x, y)$ and separates $S(p, \rho)$ into two sets. In the language of differential geometry N is therefore an $(n-1)$-dimensional totally geodesic set. Let (p, dx_N) be the lineal element normal to N at p. If any other lineal element (q, dy) is given, then pairwise transitivity guarantees the existence of a motion taking (p, dx_N) into (q, dy). This motion takes N into a totally geodesic set through q and normal to dy. It follows from the wellknown theorem of Beltrami, that R has constant curvature.

6. Equilong maps of the elementary spaces. We now study the locally finite equilong maps α of the elementary spaces of dimension $n \geqq 2$ which are not motions. Then there are at least two regions of injectivity. As before we denote these by $D_1, D_2, \cdots$ and assume that α leaves D_1 pointwise fixed. By (25) the D_i and $\bar{D}_i$ are convex sets. We remember that a set in a spherical space S^n is called convex if it contains with two points at least one segment. Such a set either lies in a closed hemisphere or is the entire S^n. Therefore each D_i and D_i is indeed convex and D_i lies in an open hemisphere of S^n. For brevity we write (D) for the set of the D_i.

The r-faces $(0 \leqq r \leqq n-1)$ of (D) are the r-faces of the individual D_i. Of course, we will call the 0-faces, and 1-faces also vertices and edges. If the D_i are known then α is easily reconstructed. A *string* $s = (D'_1, \cdots, D'_r)$, where each D'_j is a D_i, has the property that D'_i and D'_{i+1} $(j = 1, \cdots, r-1)$ have a common $(n-1)$-face. This face is unique if $D'_j \neq D'_{j+1}$. In this case we define R_{j+1} as the reflection of the space in the hyperplane containing $D'_j \cap D'_{j+1}$; or $D'_j = D'_{j+1}$ then R_j is the identity map ε of the space R. We define

$$\beta(s) = R_2 R_3 \cdots R_r$$

and complete this definition by putting $\beta(D_i) = \varepsilon$ for strings consisting of a single D_i.

In terms of strings the map can be described as follows:

Let $s_k = (D'_1, \cdots, D'_r)$ be a string from D_1 to D_k, i.e., $D'_1 = D_1, D'_r = D_k$. Then

$$\alpha x = \beta(s_k)x \text{ for } x \in D_k . \tag{31}$$

This follows from our discussion of the Minkowski case.

(31) *implies that $\beta(s_k)x$ is independent of the string from D_1 to D_k.*

Conversely, if for a given locally finite division $D = (D_1, D_2, \cdots)$ of an elementary space into convex regions the map $\beta(s_k)$ is independent of the string s_k leading from D_1 to D_k, then (31) *defines an equilong map with the D_i as regions of injectivity which leaves D_1 pointwise fixed.*

Let (D) consist of the regions of injectivity of a locally finite equilong map and consider an m-face f_m of (D) $(0 \leqq m \leqq n - 1)$. Take an interior point w of f_m ($w = f_0$ for $m = 0$) and a hypersphere K'_w about w whose radius $\rho > 0$ is so small that the ball $wx \leqq \rho$ intersects no other D_i than those, $D'_1, \cdots, D'_k$, which have f_m as a face. Let the $(n - m)$-flat normal to f_m at w intersect K'_w in the $(n - m - 1)$-sphere K_w ($K_w = K'_w$ if $m = 0$). The equilong map β with (D) as regions of injectivity which leaves D'_1 pointwise fixed induces an equilong map of K_w in itself for which the $D'_j \cap K_w$ are the regions of injectivity.

If (D) is an arbitrary locally finite division of the space into (at least two) convex regions, and its m-faces are again defined as those of the individual D_i, we may define K_w for a given f_m as before and denote by $C(f_m)$ *the condition that the $D'_j \cap K_w$ be the regions of injectivity for an equilong map of K_w.* These conditions are essentially independent of the choice of w and ρ in the sense that for different choices leading to $K^*_{w^*}$ a homothetic transformation will send the $D'_j \cap K_w$ into the $D'_j \cap K^*_{w^*}$.

The conditions $C(f_{n-1})$ are trivial, they are satisfied by any (D) and hence will no longer be mentioned. The $C(f_{n-2})$ are particularly simple. In this case K_w is a circle and we obtain from (27):

(32) *The condition $C(f_{n-2})$ means: If $f^1_{n-1}, \cdots, f^l_{n-1}$ are, in cyclic order, the $(n - 1)$-faces of (D) which have f_{n-2} as face then l is even and, if δ_i is the angle between f^i_{n-1} and f^{i+1}_{n-1} ($f^{l+1}_{n=1} = f^1_n$) then*

$$\delta_1 + \delta_3 + \cdots + \delta_{l-1} = \delta_2 + \delta_4 + \cdots + \delta_l .$$

We have shown:

(33) If $D_1, D_2, \cdots$ are the regions of injectivity of a locally finite equilong map of an n-dimensional ($n > 2$) elementary space, then they satisfy the conditions $C(f_m)$ for all m-faces f_m ($0 \leqq m \leqq n-2$).

The converse of (33) also holds, but it is clear that the $C(f_m)$ are not independent. Our discussion of the general case applied to K_w contains

(34) If f_m is a face of f_{m+k} ($k \geqq 1$) then $C(f_m)$ implies $C(f_{m+k})$.

Thus, if all f_m with $m > 0$ have vertices, then the $C(f_0)$ yield the remaining $C(f_m)$. It is much more surprising that the simple conditions $C(f_{n-2})$ are also sufficient. Although our proof of the converse of (33) will use the $C(f_{n-2})$ only, it is of interest to see directly why the $C(f_m)$ with $m < n-2$ are redundant. Therefore we show:

(35) If the converse of (33) holds, then the conditions $C(f_{n-2})$ are sufficient.

For $n = 2$ the conditions $C(f_0) = C(f_{n-2})$ are the only ones in (34), hence (35) is true. Assume (35) has been proved for $n \leqq N-1$.

Let f_m be an m-face of (D) in E^N ($m \leqq N-2$) and construct a corresponding $(N-m-1)$-sphere K_w ($w \in f_m$) as above. With the previous notations $D_1', \cdots, D_r'$ with f_m as face determine regions $D_i' \cap K_w$ on K_w. An $(m+k)$-face of $(D') = (D_1', \cdots, D_r')$ containing f_m intersects K_w in a $(k-1)$-face f'_{k-1} of $(D_1' \cap K_w, \cdots, D_k' \cap K_w)$. The condition $C(f'_{k-1})$ for this set is equivalent to $C(f_{m+k})$ for (D'), in particular $C(f'_{N-m-3})$ to $C(f_{N-2})$.

By the induction hypothesis applied to the $(N-m-1)$-sphere K_w it follows from the $C(f'_{N-m-3})$ that the $D_i' \cap K_w$ are the regions of injectivity for an equilong map of K_w. Therefore $C(f_m)$ holds and it follows from (34) that the D_i are the regions of injectivity for an equilong map of R. Thus our principal result on elementary spaces is this:

(36) THEOREM. *Let* $(D) = (D_1, D_2, \cdots)$ *be a locally finite division of an elementary space* R ($\dim R = n \geqq 2$) *into convex regions. Necessary and sufficient for the* D_i *to be the regions of injectivity of an equilong map* α *of* R*—which is then determined up to motions by* (31)*—is that every* $(n-2)$*-face* f_{n-2} *of any* D_i *satisfy the condition:*

If $f^1_{n-1}, \cdots, f^l_{n-1}$ *are the* $(n-1)$*-faces of the* D_i *having* f_{n-2} *as face in cyclic order and* δ_i *is the angle between* f^i_{n-1} *and* f^{i+1}_{n-1} ($f^{l+1}_{n-1} = f^1_{n-1}$) *then* l *is even and*

$$\delta_1 + \delta_3 + \cdots + \delta_{l-1} = \delta_2 + \delta_4 + \cdots + \delta_l \,.$$

The proof is the content of the last section.

7. Proof of the main theorem on elementary spaces. If $s = (D_1', \cdots, D_r')$ and $s' = (A_1 = D_r', A_2, \cdots, A_t)$ are strings (always in the given system (D)), then we denote the string $(D_1', \cdots, D_r', A_1, \cdots, A_t)$ by $s \cdot s'$. Our conventions on $\beta(s)$ show that then

$$\beta(s)\beta(s') = \beta(s \cdot s') . \tag{37}$$

With s^{-1} standing for $(D_r', D_{r-1}', \cdots, D_1')$ we have

$$\beta(s)\beta(s^{-1}) = \beta(s \cdot s^{-1}) = \beta(s^{-1})\beta(s) = \varepsilon . \tag{38}$$

The string s is closed if $D_1' = D_r'$. If s is closed then $s_c = (D_j', \cdots, D_r', D_2', \cdots, D_{j-1}', D_j')$ $(1 < j \leqq r)$ is closed and it follows from (37) that

(39) $\beta(s) = \varepsilon$ *implies* $\beta(s_c) = \varepsilon$ *and conversely.*

(40) *If* $s_1 = s_1's$ *and* $s_2 = s^{-1}s_2'$ *then* $\beta(s_1s_2) = \beta(s_1's_2')$.

This is a corollary of (37) and (38). Finally: If $s = s_1s_2s_3$ and $s' = s_1's_2^{-1}s_3'$ are closed strings and $\beta(s) = \beta(s') = \varepsilon$ then (39) yields $\beta(s_3s_1s_2) = \beta(s_2^{-1}s_3's_1') = \varepsilon$, hence

$$\beta(s_3s_1s_3's_1') = \beta(s_1s_3's_1's_3) = \varepsilon . \tag{41}$$

Let $s = (D_1', \cdots, D_r' = D_1')$ be any closed string. A polygonal path π belongs to s if it has the following properties: it begins and ends at a point of D_1'. It is the product $\pi = \pi_1 \cdots \pi_r$ of paths π_i (in the sense of homotopy theory), where π_i lies, except possibly for its endpoints, in D_i'. The endpoint of π_i $(i < r)$ and hence the initial point of π_{i+1} is an interior point of the common $(n-1)$-face of D_i' and D_{i+1}' if $D_i' \neq D_{i+1}'$ and lies in D_i' if $D_i' = D_{i+1}'$.

Conversely, let a closed polygonal path $\pi = \pi_1 \cdots \pi_r$ be given such that it begins and ends at a point of D_1'', each π_i lies, the endpoints possibly excepted, in a D_i'', and if the endpoint of π_i $(i < r)$ does not lie in D_i'' then D_i'' and D_{i+1}'' have exactly one common $(n-1)$-face and the point is an interior point of this face. Then $(D_1'', \cdots, D_r'' = D_1'')$ is a closed string, and the only one, to which π belongs.

Our rules (37) to (41) contain the following fact:

(42) *If* π, π_i *are polygonal paths belonging to closed strings* s *and* s_i $(i = 1, \cdots, m)$ *and if* $\pi \sim \pi_1 \cdots \pi_m$, *in the sense that* π *remains after subpaths of the* π_i *traversed in opposite senses have been cancelled, then* $\beta(s_i) = \varepsilon$ *implies* $\beta(s) = \varepsilon$.

The considerations of the preceding section reduce the proof of

(36) to the following:

(43) *If* (D) *satisfies the conditions* $C(f_{n-2})$ *then* $\beta(s) = \varepsilon$ *for every closed string* s *in* (D).

Let the closed string $s = (D'_1, \cdots, D'_r = D'_1)$ be given. We may assume that $D'_i \neq D'_{i+1}$ $(1 \leqq i \leqq r-1)$. Then points $p_i \in D'_i$ $(p = p_1 = p_r)$ can be chosen to satisfy the following conditions:

(a) The points p, p_i, p_{i+1} $(2 \leqq i \leqq r-2)$ are not collinear (do not lie on a great circle in the spherical case). The segments $T'_i = T(p, p_i)$ $(2 \leqq i \leqq r-1)$ and $T_i = T(p_i, p_{i+1})$ $(1 \leqq i \leqq r, p_{r+1} = p)$ are then unique also in the spherical case.

(b) T'_i or T_i have at most one common point with a given $(n-1)$-face of (D) and do not intersect a face of dimension lower than $n-1$. The path π formed by the segments T_i oriented from p_i towards p_{i+1} then belongs to s.

(c) The 2-simplex S_i $(2 \leqq i \leqq r-2)$ spanned by p, p_i, p_{i+1} does not intersect a face of (D) of dimension less than $n-2$.

There is a finite number (if any) of points $u_1, \cdots, u_t$ in which $\cup S_i$ intersects the $(n-2)$-faces of (D), and the u_j are interior points of the S_i in which they lie.

Let $q_{m_i+1}, q_{m_i+2}, \cdots, q_{m_{i+1}}$ $(m_2 = 0)$ with $q_1 = p_2$ and $q_{m_i} = p_i$ denote points lying in this order on T_i. Let $t_0 = 0 < t_1 < \cdots < t_h = 1$, and denote by $q_{\nu,i}$ the point on the (even in the spherical case unique) segment $T(p, q_i)$ $(i = 1, \cdots, m_{r-2})$ for which $pq_{\nu,i} = t_\nu(pq_i)$.

The points q_i and the numbers t_ν can be chosen in such a manner that the $q_{\nu,i}$ have the following properties: No $T(q_{\nu,i}, q_{\nu+1,i})$ or $T(q_{\nu,i}, q_{\nu,i+1})$ has more than one common point with a given $(n-1)$-face of (D), or intersects a face of (D) of dimension less than $n-1$. Consequently, these segments also avoid the points u_j. Denote by $Q_{\nu,i}$ the (convex) quadrangle with vertices $q_{\nu,i}, q_{\nu+1,i}, q_{\nu+1,i+1}, q_{i+1,\nu}$ and by $\pi_{\nu,i}$ its boundary with the orientation corresponding to this order of the vertices. If the q_i and t_ν are properly chosen then these quadrangles have the following further properties: $Q_{0,i} \subset D'_1$ for all i. For $\nu > 0$ a $Q_{\nu,i}$ lies either in one D'_i, or $Q_{\nu,i}$ has common points with exactly two D'_i which have a common $(n-1)$-face and intersects this $(n-1)$-face in interior points, or, finally, $Q_{\nu,i}$ contains exactly one u_j and lies in the union of D'_i with a common $(n-2)$-face.

Then $\pi_{\nu,i}$ belongs to a closed string $s_{\nu,i}$ and $\pi \sim \prod_{\nu}\prod_{i} \pi_{\nu,i}$. It is clear that $\beta(s_{0,i}) = \varepsilon$ and the $\beta(s_{\nu,i}) = \varepsilon$ for $\nu > 0$ in the first two cases because of our rules (37) to (41). In the last case, if $Q_{\nu,i}$ contains u_j and $u_j \in f^j_{n-2}$ then $C(f^j_{n-2})$ and (37) to (41) guarantee that $\beta(s_{\nu,i}) = \varepsilon$. It now follows from (42) that $\beta(s) = \varepsilon$.

In conclusion we point out that this investigation leads to a variety of questions: A first type concerns general shrinkages of special spaces and is exemplified by (16). A second type inquires into the structure of the regions of injectivity of equilong maps of spaces which do not satsify III. A third deals with the equilong maps of special (e.g. the locally elementary) spaces. While it does not seem worth the effort to determine all equilong maps of the elementary spaces, it should be decided, whether the spaces in (29) possess proper equilong maps which are not locally finite.

In addition there are many topics suggested by length preserving maps, for instance, maps of E^n into itself which preserve, with a suitable definition, the areas of all two-dimensional surfaces. The locally finite equilong maps of E^n have this property for any reasonable area.

References

1. H. Busemann, *The Geometry of Geodesics*, New York 1955.
2. ———, *Geometries in which the planes minimize area*, Annali Mat pur. appl. (IV) **55** (1961), 171-190.
3. D. G. Bourgin, *Modern Algebrai Topology*, New York 1963.
4. H. Freudenthal and W. Hurewicz, *Dehnungen, Verkürzungen, Isometrien* Fund. Math., **26** (1930), 120-122.
5. A. N. Kolmogorov and B. V. Fomin, *Functional Analysis vol I*, Rochester 1957.
6. Yu. G. Rešetnyak, *On a special map of a polyhedron* (Russian), Mat. Sbornik **53** (95) (1961), 39-52.
7. W. Rinow, *Die innere Geometrie der metrischen Räume*, Berlin-Göttingen-Heidelberg 1961.
8. J. Szenthe, *Über lokalisometrische Abbildungen von G-Räumen auf sich*, Annali Mat. pur. appl. (IV) **55** (1961) 37-46.
9. ———, *Über metrische Räume, deren lokalisometrische Abbildungen Isometrien sind*, Acta Math. Acad. Sci. Hungar., **13** (1962), 443-441.
10. J. Tits, *Sur certaines classes d'espaces homogènes de groupes de Lie*, Mem. Acad. royale Belgique, Cl. Sciences, vol in 8°, **29** (1955), fasc. 3.

CHARACTERIZATIONS OF CONVOLUTION SEMIGROUPS OF MEASURES

H. S. COLLINS

A problem of fundamental importance in the study of compact topological semigroups is that of classifying in an *intrinsic* way each of a certain class of such semigroups. Unfortunately, virtually nothing has been done along these lines, even for such geometrically pleasing semigroups as the affine semigroups introduced by the author and H. Cohen in [3]. It is the purpose of this note to rectify this situation, at least for several particular types of compact affine topological semigroups; namely, certain convolution semigroups of real valued regular Borel measures on compact topological semigroups. The author's interest in this problem dates back to the early papers of Peck [13] and Wendel [21], and to some unpublished work of Wendel. Since that time, quite a literature has developed as regards these semigroups (e.g., see the bibliography), but almost without exception these papers merely study the *properties* of the semigroups without making any attempt to abstract sufficiently many of their properties to characterize them.

If S is a compact Hausdorff space and $P(S)$ denotes the set of all nonnegative regular Borel measures on S of variation norm one, it is known that $P(S)$ is a convex set which is compact in the weak-* topology $\left(\text{a net } \{\mu a\} \text{ of measures in } P(S) \text{ converges weak-* to } \mu \in P(S) \text{ if } \int f d\mu a \to \int f d\mu, \text{ for each real continuous function } f \text{ on } S\right)$. In similar fashion, the unit ball $B(S)$ of real-valued regular Borel measures of norm $\leqq 1$ is a compact convex set. When S is endowed with a continuous associative multiplication, each of $P(S)$ and $B(S)$ becomes a compact affine topological semigroup relative to convolution multiplication (see [10]); when such is the case, we denote these semigroups by $\tilde{S}$ and $\approx{S}$ respectively. Note that our use of the symbol $\approx{S}$ differs from that of Glicksberg in [10], where $\approx{S}$, denoted the ball semigroup of *complex* measures.

In § 2, the following three types of images of the sets $P(S)$ and $B(S)$ are determined:

(a) all *extremal* images of $P(S)$; i.e., all continuous affine images under mappings which preserve extreme points,

(b) all one-to-one affine bicontinuous images of $P(S)$, and

(c) all one-to-one affine bicontinuous images of $B(S)$. The common requirements in each of (a), (b), and (c) are that the image K be

Received July 8, 1963. This research was supported by the National Science Foundation, Contract NSF GP-1449.

compact and convex, have a separating family of real continuous affine functions, and have a compact set of extreme points. In (b), the additional requirement is that K be a simplex in the sense of Choquet [2] or Loomis [12]. In (c), one must require the existence of a compact subset T of K and a point z in K such that K is "symmetric relative to z", $T \cup (2z - T)$ is the set of extreme points of K, the closed convex hull K_1 of T is a simplex, and there exists on K a continuous real affine function which vanishes at z and is one on K_1.

In § 3, the imposition of a topological semigroup structure on S (and consequently on $P(S)$ and $B(S)$ via convolution) enables us to use the results of § 2 to characterize

(a) all extremal homomorphic images of $\tilde{S}$,

(b) all one-to-one affine bicontinuous and isomorphic images of $\tilde{S}$,

and (c) all one-to-one affine bicontinuous and isomorphic images of $\tilde{\tilde{S}}$. The only requirement needed in addition to the corresponding ones in § 2 is that the set of extreme points of K in cases (a) and (b) be a topological semigroup, while case (c) requires that the set T be a topological semigroup and the point z be a zero of K.

In each of §§ 2 and 3 additional characterizations of some interest are given. In our use of the Choquet simplex condition we prefer the formulation Loomis gives in [12], and it is a pleasure to record here the author's indebtedness to Professor Loomis for recent conversations during a visit by him as consultant to a Banach Algebra seminar at Louisiana State University.

1. Preliminaries. Throughout this paper the letter K will denote a compact convex subset of some real Hausdorff topological vector space. A mapping f with domain K and range another such set is *affine* if $x, y \in K$ and $0 \leqq a \leqq 1$ imply $f(ax + (1-a)y) = af(x) + (1-a)f(y)$. The symbol $L(K)$ will be used for the set of all continuous real valued affine functions defined on K, and it is clear that $L(K)$ needs not in general distinguish points of K. If, however, the vector space containing K is locally convex, the set $L(K)$ will distinguish points, and thus the assumption of local convexity (which we do not make) would permit a considerable simplification in the statements of the theorems to follow. If $z \in K$, the symbol $L_0(K)$ denotes the subset of $L(K)$ consisting of those functions each of which vanishes at z. It is easy to see that $L(K)$ separates points of K if and only if $L_0(K)$ does. If each of $L(K)$ and $L_0(K)$ is given the supremum norm, then both bocome real Banach spaces and as such have adjoint spaces of real continuous linear functionals, denoted respectively by $L(K)^*$ and $L_0(K)^*$. In each of these spaces we make use of the weak-* topology to embed K. Explicitly, if $x \in K$, denote by $\bar{x}$ [and x'] respectively the element of

$L(K)^*$ [of $L_0(K)^*$] for which $\bar{x}(l) = l(x)$ for all $l \in L(K)$ [$x'(l) = l(x)$ for all $l \in L_0(K)$]. It is obvious that these mappings are one-to-one if and only if $L(K)$ separates points of K, and that each is affine and continuous between K and its image, the latter given the relativized weak-* topology. The embedding $x \to \bar{x}$ was used by Loomis in [12] to formulate and extend Choquet's work [2]. Following Loomis, we say that K is a *simplex* if (i) $L(K)$ separates points of K and (ii) the truncated cone $T_r(K) = \{a\bar{z}: 0 \leqq a \leqq 1, z \in K\}$ determined by K in $L(K)^*$ is a lattice relative to the partial order: if $x, y \in T_r(K)$, then $\bar{x} \leqq \bar{y} \leqq \bar{y}$ means $\bar{y} - \bar{x} \in T_r(K)$. Our only contact with Loomis's work here (aside from borrowing some of his notation) is the use of his Theorem 6 to prove (when K has a compact set of extreme points) that K is a simplex if and only if K is the one-to-one affine bicontinuous image of some $P(S)$. The statement that K is an *affine semigroup* means (see [3]) there exists an associative separately affine multiplication on K; K is an *affine topological semigroup* if the multiplication function is also (doubly) continuous. The semigroups $\tilde{S}$ and $\tilde{\tilde{S}}$ are important examples of such semigroups, as are many semigroups of matrices. Another important class of such semigroups is the class of *group extremal* semigroups (the term is Wendel's), where by definition the compact affine topological semigroup K is *group extremal* if (i) it has an identity element and (ii) the set of elements with inverse coincides with the set of extreme points of K. Peck in [13] proved that each such semigroup has a zero, and Wendel (unpublished) observed that this result follows also from the fact that each such semigroup is the homomorphic image of some $\tilde{S}$, with S a compact group.

2. Affine images of $P(S)$ and $B(S)$. In this section of the paper K (as above) will be a convex compact set and $E(K)$ will denote its set of extreme points (a priori, possibly void). However, if $L(K)$ separates points of K, such is not the case; in fact, the Krein-Milman theorem holds for K. The first theorem gives conditions on K necessary and sufficient that K be the extremal image of some $P(S)$.

THEOREM 2.1. *Suppose that $L(K)$ separates points of K. Then K is the extremal image of some $P(S)$ if and only if $E(K)$ is compact.*

Proof. Suppose first that there exists a compact space S and an extremal mapping F (F is continuous affine onto and preserves extreme points) of $P(S)$ to K. By the Kelley-Arens theorem [1, Lemmas 3.1 and 3.2], the set of point measures on S is the set of extreme points of $P(S)$, hence is compact. Thus $F(E(P(S)))$ is compact and contains $E(K)$ (this inclusion holds always). Since F is extremal, the other inclusion is true also; i.e., $E(K)$ is compact.

Conversely, suppose $E(K)$ is compact, and let $S = E(K)$. By assumption, the embedding $\bar{K}$ of K in $L(K)^*$ is one-to-one bicontinuous affine and onto. We now define a mapping R on $P(S)$ onto $\bar{K}$ which is continuous affine and extremal. For $\mu \in P(S)$, and $l \in L(K)$, let $R_\mu(l) = \int_S l d\mu$. Fix a partition $\{E_i\}_{i=1}^n$ of S by Borel sets, $t_i \in E_i$, $1 \leqq i \leqq n$. Then $\sum_{i=1}^n l(t_i)\mu(E_i) = \sum_{i=1}^n \bar{t}_i(l)\mu(E_i) = \sum_{i=1}^n l(\bar{t}_i)\mu(E_i)$ (regarding l as a linear functional on $L(K)^*) = l(\sum_{i=1}^n \mu(E_i)\bar{t}_i)$. Since $\sum_{i=1}^n \mu(E_i) = 1$, $\mu(E_i) \geqq 0$, and $\bar{t}_i \in \bar{S}$, clearly the sum $\sum_{i=1}^n \mu(E_i)\bar{t}_i \in$ convex hull of $\bar{S}$. Since sums of the form $\sum_{i=1}^n l(t_i)\mu(E_i)$ converge to $\int_S l d\mu$, this implies that $R_\mu \in$ weak-* closed convex hull of $\bar{S} = \bar{K}$. It it clear that R maps the extreme points of $P(S)$ onto those of $\bar{K}$ and hence (since R is obviously continuous and affine) R maps $P(S)$ onto $\bar{K}$. This completes the proof.

The next theorem gives several different sets of necessary and sufficient conditions that K be the one-to-one affine bicontinuous image of a $P(S)$. It perhaps should be remarked that the requirement that K be a simplex can be stated without mentioning explicitly the embedding $\bar{K}$. We now do this, merely referring the reader to Loomis [12, Theorem 6] for the verification. The result will be stated as Lemma 2.1.

LEMMA 2.1. *Suppose that $L(K)$ separates points of K. Then K is a simplex if and only if given $\{a_i x_i\}_{i=1}^m$ and $\{b_j y_j\}_{j=1}^n$, where $\sum_{i=1}^m a_i = \sum_{j=1}^n b_j = 1$, $a_i \geqq 0$, $b_j \geqq 0$, $x_i, y_j \in K$, $1 \leqq i \leqq m$, $1 \leqq j \leqq n$ and $\sum_{i=1}^m a_i x_i = \sum_{j=1}^n b_j y_j$, there exists $\{c_k z_k\}_{k=1}^p$ with $\sum_{k=1}^p c_k = 1$, $c_k \geqq 0$ and $z_k \in K$, $1 \leqq k \leqq p$, for which* (a) $\sum_{k=1}^p c_k z_k = \sum_{i=1}^m a_i x_i = \sum_{j=1}^n b_i y_j$, (b) $\{1, 2, \cdots, p\}$ *can be written as the pairwise disjoint union of sets $\{I_i\}_{i=1}^m$ and $\{J_j\}_{j=1}^n$, with $a_i = \sum_{k \in I_i} c_k$, $a_i x_i = \sum_{k \in I_i} c_k z_k$, $1 \leqq i \leqq m$, and $b_j = \sum_{k \in J_j} c_k$, $b_j y_j = \sum_{k \in J_j} c_k z_k$, $1 \leqq j \leqq n$.*

THEOREM 2.2. *The following conditions are mutually equivalent for K:* (1) *K is the one-to-one affine bicontinuous image of some $P(S)$,* (2) (a) *$E(K)$ is compact and* (b) *K is a simplex,* (3) (a) *$L(K)$ separates points of K,* (b) *$E(K)$ is compact, and* (c) *each continuous real function on $E(K)$ is extendable to be in $L(K)$.*

Proof. (1) $\rightarrow$ (2). Suppose F on $P(S)$ to K is one-to-one affine bicontinuous onto, where S is a compact Hausdorff space. It is easily verified then that $E(K)$ is compact and $L(K)$ separates points of K, for $P(S)$ has these properties. It thus remains to prove that the truncated cone $T_r(K)$ determined by $\bar{K}$ in $L(K)^*$ is a lattice. Let C be the truncated cone determined by $P(S)$ in the vector space of all real regular Borel measures on S. Clearly C itself is a lattice; we will now show that F has an extension G to all of C such that

(i) G is one-to-one and affine on C onto $T_r(K)$,

(ii) $G(0) = 0$, and

(iii) $\mu, \nu \in C$ implies $\mu \leqq \nu$ if and only if $G(\mu) \leqq G(\nu)$. If this can be proved, it will easily follow that $T_r(K)$ is a lattice. Thus, we proceed to the definition of G. If $a\mu \in C$, with $0 \leqq a \leqq 1$, $\mu \in P(S)$, define $G(a\mu) = aF(\mu)$. If $a\mu = b\nu$, with $a\mu, b\nu \in C$ and $f \equiv 1$ on S, then $a = a\int_S f d\mu = \int_S fd(a\mu) = \int_S fd(b\nu) = b\int_S fd\nu = b$; i.e., $a=b$. Thus, if $a = 0$, $a\mu = b\nu = 0$, and $G(a\mu) = aF(\mu) = 0 = bF(\nu) = G(b\nu)$. If $a \neq 0$, then $\mu = \nu$, so $G(a\mu) = aF(\mu) = bF(\nu) = G(b\nu)$; i.e., G is well defined on C into $T_r(K)$. Clearly G is onto $T_r(K)$. To show G is one-to-one, let $G(a\mu) = G(b\nu)$. Then $aF(\mu) = bF(\nu)$; i.e., $a\bar{x} = b\bar{y}$, with $x, y \in K$. If l is identically one on K then $l \in L(K)$ and $a = al(x) = a\bar{x}(l) = b\bar{y}(l) = bl(y) = b$; i.e., $a = b$, and $aF(\mu) = aF(\nu)$. Since $a = 0$ implies $a\mu = 0 = b\nu$, we can assume $F(\mu) = F(\nu)$, whence $\mu = \nu$ and $a\mu = b\nu$. It is obvious that (and this has been proved already) $G(0) = 0$, so it remains only to verify that G extends F (this is clear), that G is affine, and (iii) holds. To prove G is affine, let $a\mu, b\nu \in C$, $0 \leqq c \leqq 1$, and let $d = ca + (1 - c)b$. If $d \neq 0$, then

$$\begin{aligned} G(ca\mu + (1-c)b\nu) &= G\Big[d\Big(\frac{ca}{d}\mu + \frac{(1-c)b\nu}{d}\Big)\Big] \\ &= d\Big[F\Big(\frac{ca}{d}\mu + \frac{(1-c)b\nu}{d}\Big)\Big] = d\Big[\frac{ca}{d}F(\mu) + \frac{(1-c)b}{d}F(\nu)\Big] \\ &= caF(\mu) + (1-c)bF(\nu) = cG(a\mu) + (1-c)G(b\nu)\,. \end{aligned}$$

If $d = 0$, then $\int_S 1d(ca\mu + (1 - c)b\nu) = ca\int_S 1d\mu + (1 - c)b\int_S 1d\nu = ca + (1 - c)b = 0$, hence $ca\mu + (1 - c)b\nu = 0$, and $ca = (1 - c)b = 0$. The desired result easily follows. Now, suppose $\phi, \psi \in C$, with $\phi \leqq \psi$. Then $(1/2)0 + (1/2)\psi = (1/2)\psi = (1/2)\phi + (1/2)(\psi - \phi)$, whence $(1/2)G(\psi) = (1/2)G(0) + (1/2)G(\psi) = (1/2)G(\phi) + (1/2)G(\psi - \phi)$ and this implies $G(\psi) - G(\phi) = G(\psi - \phi) \in T_r(K)$; i.e., $G(\phi) \leqq G(\psi)$. To conclude (1) $\to$ (2), we suppose $G(\phi) \leqq G(\psi)$, with $\phi, \psi \in C$. Then there is $\alpha \in C$ such that $G(\alpha) = G(\psi) - G(\phi)$, so $G(\psi) = G(\alpha) + G(\phi)$. From this, it follows that $G(\psi/2) = (1/2)G(\psi) = G((\alpha + \phi)/2)$ and since G is one-to-one, $(\psi/2) = ((\alpha + \phi)/2)$. Thus $\psi = \alpha + \phi$, $\psi - \phi = \alpha \in C$, and $\phi \leqq \psi$.

(2) $\to$ (1). Let $S = E(K)$ and define R on $P(S)$ into $L(K)^*$ as in the proof of Theorem 1; thus R is an extremal mapping of $P(S)$ onto $\bar{K}$, and there remains only the proof that R is one-to-one. Let $\mu, \nu \in P(S)$ with $R_\mu = R_\nu$. Extend μ and ν to the Borel sets of K by defining them to be zero at Borel sets missing S (call these extensions $\bar{\mu}$ and $\bar{\nu}$), and note then that $R'_{\bar{\mu}}\bar{\mu} = R'_{\bar{\nu}}$, where, for example, $R'_{\bar{\mu}}(l) = \int_K ld\bar{\mu}$, all $l \in L(K)$. The mapping R' is the resultant mapping

used by Loomis in [12], and in Theorem 6 he proved that (since $\bar{K}$ a lattice clearly implies the set of all subelements of $\bar{x} = R'_{\bar{\mu}} = R'_{\bar{\nu}}$ is a lattice) there exists a unique extremal measure whose resultant is $\bar{x}$. Since $S = E(K)$ is known here to be compact, this says [12, p. 517] that $\mu = \nu$. Thus R is one-to-one, and (1) is proved.

(2) $\to$ (3). As was seen in the proof of (2) $\to$ (1), the mapping R on $P(S)$ to $\bar{K}$ is one-to-one bicontinuous and affine onto, where $S = E(K)$. If f is continuous real valued on S, denote by h the restriction to $P(S)$ of the linear functional on the space of all real regular Borel measures determined by $f: h(\mu) = \int_S f d\mu$, all $\mu \in P(S)$. Then $x \in E(K)$ implies (denote by $\underset{\cdot}{x}$ the point measure on S determined by x) $\bar{f}(x) \equiv h(R^{-1}(\bar{x})) = h(\underset{\cdot}{x}) = \int_S f d\underset{\cdot}{x} = f(x)$; i.e., $\bar{f}$ extends f to be in $L(K)$.

(3) $\to$ (1). If $\mu, \nu \in P(S)$ and $R_\mu = R_\nu$, where $S = E(K)$, and f is continuous real valued on S, let $\bar{f}$ be its extension to K to be continuous and affine. Then $\int_S f d\mu = \int_S \bar{f} d\mu = \int_S \bar{f} d\nu = \int_S f d\nu$, i.e., μ and ν are equal as functionals on the space of real continuous functions on S. The Riesz theorem then implies $\mu = \nu$ as measures. This concludes the proof of Theorem 2.

REMARK 2.1. It is easy to verify that in the preceding theorem the condition (c) of part (3) may be replaced by (c′): each continuous real function f on $S = E(K)$ is *uniquely* extendable to $\bar{f} \in L(K)$. It follows then that $f \to \bar{f}$ is an isometric isomorphism of $C(S)$ onto $L(K)$, where $C(S)$ is the space of real continuous functions on S, and each space is given the supremum norm.

We conclude § 2 now with our characterization of all one-to-one affine and bicontinuous images K of real unit balls of measures. The conditions given here (in Theorem 2.3) are quite natural with possibly one exception: the requirement that $L_0(K)$ contain a function which is identically one on T seems somewhat artificial. However, some remarks regarding this condition are made following the proof of the theorem, and these may help place the condition in proper perspective.

THEOREM 2.3. *The following conditions are mutually equivalent for* K:

(1) K *is the one-to-one affine bicontinuous image of some* $B(S)$,

(2) (a) *there exists* $z \in K$ *and compact* $T \subset K$ *such that* $x \in K$ *implies* $2z - x \in K$ *and* $E(K) = T \cup (2z - T)$ (b) *if* K_1 *denotes the closed convex hull of* T, *then* K_1 *is a simplex,* (c) $L(K)$ *separates points of* K *and* $L_0(K)$ *contains a function which is identically one on* K_1,

(3) (a) *part* (a) *of* (2) *holds and* (b) $L(K)$ *separates points of* K *and each continuous function* f *on* T *is extendable to an* $\bar{f} \in L_0(K)$.

Proof. (1) $\rightarrow$ (2). Let F be a one-to-one affine bicontinuous mapping of $B(S)$ onto K, where S is a compact Hausdorff space. Let $z = F(0)$, $T =$ the image under F of all measures determined by the points of S. If $x \in K$, there exists $\mu \in B(S)$ such that $F(\mu) = x$. Then $2z - x = 2F(0) - F(\mu) = F(2 \cdot 0 - \mu) = F(-\mu) \in K$. The Kelley-Arens theorem [1, Lemmas 3.1 and 3.2] says $\underset{\bullet}{S} \cup (-\underset{\bullet}{S})$ is the set of extreme points of $B(S)$, where $\underset{\bullet}{S}$ is the set of point measures. Hence $E(K) = F[\underset{\bullet}{S} \cup (-\underset{\bullet}{S})] = F(\underset{\bullet}{S}) \cup F(-\underset{\bullet}{S}) = T \cup (2z - T)$. Thus (a) of (2) is verified. Using the Kelley-Arens result again ($\underset{\bullet}{S}$ is the set of extreme points of $P(S)$), $K_1 = F(P(S))$ hence Theorem 2.2 implies K_1 is a simplex. Since $B(S)$ has a separating family of continuous real affine functions vanishing at 0 and contains one which is identically one on $P(S)$, part (c) of (2) follows easily. Thus, (2) is proved.

(2) $\rightarrow$ (3). Since $T \subset E(K)$, it is obvious that $T \subset E(K_1)$. On the other hand, the closed convex hull of T is K_1, hence T (being closed) contains $E(K_1)$; i.e., $T = E(K_1)$. Consider the embeddings K', K_1', and $-K_1'$ in $L_0(K)^*$ of (respectively) K, K_1, and $(2z - K_1)$. Since K_1' and $-K_1'$ are compact convex sets whose union contains $T' \cup -T' = E(K')$, it is clear that the convex hull of $K_1' \cup -K_1'$ is compact (and convex) and thus coincides with K'. Now K_1' is a simplex (by Theorem 2.2) with $E(K_1) = T'$, so each continuous real function on T' can be extended to be continuous and affine on K_1'. This fact together with the fact that $x \rightarrow x'$ is one-to-one affine and bicontinuous on K onto K' reduces the problem to proving that each continuous affine f on K_1' extends to a continuous affine function on K' which vanishes at $z' = 0 \in L_0(K)^*$. Fix such an f, and let $ax' + (1-a)(-y') \in K'$, where $0 \leqq a \leqq 1$, $x, y \in K_1$ (note that K' is the union of the line segments $[p, q]$, with $p \in K_1'$, $q \in -K_1'$). Define $\bar{f}[ax' + (1-a)(-y')] = af(x') - (1-a)f(y')$. We show first that $\bar{f}$ is well defined. Let $ax' + (1-a)(-y') = bw' + (1-b)(-t')$, with $0 \leqq a, b \leqq 1$, $x, y, w, t \in K_1$, and let $l_0 \in L_0(K)$ be one on K_1. Then $l_0(ax' + (1-a)(-y')) = al_0(x) + (1-a)l_0(-y) = a + (1-a)(-1)$, since $l_0 \equiv -1$ on $2z - K_1$. Similarly, $l_0(bw' + (1-b)(-t')) = b + (1-b)(-1)$ so $2a - 1 = 2b - 1$, and $a = b$. But then $ax' + (1-a)(-y') = aw' + (1-a)(-t')$, hence $ax' + (1-a)t' = aw' + (1-a)y'$. Since f is affine, $af(x') + (1-a)f(t') = af(w') + (1-a)f(y')$, so $af(x') - (1-a)f(y') = af(w') - (1-a)f(t')$; i.e., $\bar{f}$ is well defined. That $\bar{f}$ extends f follows from $\bar{f}(x') = f(1 \cdot x' + 0 \cdot (-x')) = 1 \cdot f(x') - 0 \cdot f(-x') = f(x')$. To prove $\bar{f}$ is continuous on K', let $\{a_\alpha x_\alpha' + (1-a_\alpha)(-y_\alpha')\}$ be a net in K' converging (weak-*) to $ax' + (1-a)(-y')$. The net $\{(a_\alpha, x_\alpha', -y_\alpha')\}$ in the compact space $[0, 1] \times K_1' \times (-K_1')$ has a subnet, say $\{(a_\beta, x_\beta', -y_\beta')\}$, converging in the product

space to $(b, w, -t)$. It follows that $a_\beta x'_\beta + (1 - a_\beta)(-y'_\beta) \rightarrow bw' + (1 + b)(-t')$, whence $bw' + (1 - b)(-t') = ax' + (1 - a)(-y')$, and $a=b$. Thus, $ax' + (1 - a)(-y') = aw' + (1 - b)(-t')$. But then (as above) $af(x')+(1-a)f(t')=af(w')+(1-a)f(y')$, whence $\bar{f}[a_\beta x'_\beta+(1-a_\beta)(-y'_\beta)]= a_\beta f(x'_\beta) - (1 - a_\beta)f(y'_\beta) \rightarrow af(w') - (1-a)f(t') = af(x') - (1-a)f(y') = \bar{f}[ax' + (1 - a)(-y')]$; i.e., $\bar{f}$ is continuous. Note also that $\bar{f}(0) = \bar{f}[(1/2)x' + (1/2)(-x')] = (1/2)f(x') - (1 - (1/2))f(x') = 0$. Finally, we show $\bar{f}$ is affine. To this end, let $ax' + (1 - a)(-y')$ and $bw' + (1 - b)(-t') \in K'$, $0 \leqq c \leqq 1$. Then $\phi = c[ax' + (1 - a)(-y')] + (1 - c)[bw' + (1 - b)(-t')] = cax' + (1 - c)bw' + c(1 - a)(-y') + (1 - c)(1-b)(-t')$. If $d = ca + (1-c)b$, then $1 - d = c(1-a) + (1-c)(1-b)$. If then $d \neq 0$, $d \neq 1$,

$$\begin{aligned}
\bar{f}(\phi) &= \bar{f}\Big[d\Big\{\frac{ca}{d}x' + \frac{(1-c)b}{d}w'\Big\} \\
&\quad + (1-d)\Big\{\frac{c(1-a)}{1-d}(-y') + \frac{(1-c)(1-b)}{1-d}(-t')\Big\}\Big] \\
&= d\Big[f\Big(\frac{ca}{d}x' + \frac{(1-c)b}{d}w'\Big)\Big] \\
&\quad - (1-d)\Big[f\Big(\frac{c(1-a)}{1-d}y' + \frac{(1-c)(1-b)}{1-d}t'\Big)\Big] \\
&= caf(x') + (1-c)bf(w') - c(1-a)f(y') \\
&\quad - (1-c)(1-b)f(t') = c[af(x') - (1-a)f(y')] \\
&\quad + (1-c)[bf(w') - (1-b)f(t')] \\
&= c\bar{f}[ax' + (1-a)(-y')] + (1-c)\bar{f}[bw' + (1-b)(-t')] .
\end{aligned}$$

Each of the cases $d = 0$ and $d = 1$ is resolved into easily handled sub-cases, and the arguments will be omitted. This completes the proof of (2) $\rightarrow$ (3).

(3) $\rightarrow$ (1). Define R on $B(T)$ into $L_0(K)^*$ as usual: for $\mu \in B(T)$ and $l \in L_0(K)$, let $R_\mu(l) = \int_T l d\mu$. By an argument similar to one used before, R maps $B(T)$ onto the weak-* closed convex symmetric hull in $L_0(K)^*$ of T'. This set is K', and as before R is affine and continuous, so the proof that R is one-to-one is all that remains. Let $\mu, \nu \in B(T)$ and $R_\mu = R_\nu$. Then if f is continuous on T, it has an extension $\bar{f}$ to be in $L_0(K)$. But then $\int_T f d\mu = \int_T \bar{f} d\mu = R_\mu(\bar{f}) = R_\nu(\bar{f}) = \int_T \bar{f} d\nu = \int_T f d\nu$, and $\mu = \nu$ as functionals on the real continuous functions on T. The Riesz theorem completes the proof that $\mu = \nu$, and thus the theorem is concluded.

LEMMA 2.2. *Suppose K is compact convex, with $z \in K$ and $T \subset K$ such that $x \in K$ implies $2z - x \in K$ and T is compact. Let further*

$L_0(K)$ separate points of K, $E(K) = T \cup (2z - T)$, and K_1 be the closed convex hull of T. The following conditions are then mutually equivalent:

(1) *$L_0(K)$ contains l_0 which is one on K_1,*

(2) *$0 \leqq a,\ b \leqq 1,\ x, y, w, t \in K_1$ and $ax' + (1-a)(-y') = bw' + (1-b)(-t')$ imply $a = b$,*

(3) *each $l \in L(K_1)$ can be extended to an $\bar{l} \in L_0(K)$.*

Proof. The implication (1) $\rightarrow$ (3) is part of the proof of (2) $\rightarrow$ (3) of the previous theorem. If (3) holds, then since $L(K_1)$ contains the function which is constantly one on K_1, clearly (1) holds; i.e., (3) $\rightarrow$ (1). The proof that (1) $\rightarrow$ (2) is also in the proof of (2) $\rightarrow$ (3) of Theorem 2.3, so it remains only to show that (2) $\rightarrow$ (1). This proof, however, is also found in (2) $\rightarrow$ (3) of the previous theorem, for all that was needed to extend $f \in L(K_1)$ to $\bar{f} \in L_0(K)$ was condition (2) of the present lemma. In particular, then, the function identically one on K_1 is extendable; i.e., (1) holds.

REMARK 2.2. Given the hypotheses of Lemma 2.2., each of (1) through (3) of that lemma is equivalent to the geometric condition: *Let C be the cone $\{ax' : a \geqq 0,\ x \in K_1\}$ in $L_0(K)^*$ determined by K_1'. Then 0 is not in K_1' and each $\phi \neq 0$ in C is uniquely representable as $\phi = ax'$, for some $a > 0$ and $x \in K_1$.* The proof of this statement is quite easy, as follows. Let $ax' = by'$, with $a, b > 0$, $x, y \in K_1$. Then, by (1) of Lemma 2.2, $a = al_0(x) = l_0(ax') = l_0(by') = bl_0(y) = b$, hence $a = b$ and thus $x' = y'$. Clearly, 0 is not in K_1'. Conversely, if this geometric condition obtains, let $0 \leqq a,\ b \leqq 1,\ x, y, w, t \in K_1$ and $ax' + (1-a)(-y') = bw' + (1-b)(-t')$. Then $ax' + (1-b)t' = bw' + (1-a)y'$. Let $d = a - b + 1$, and note that $b - a + 1 = 2 - d$. Thus if $d \neq 0$, $d \neq 2$, we have that $d[(a/d)x' + ((1-b)/d)t'] = 2 - d[(b/(2-d))w' + ((1-a)/(2-d))y']$, hence (by the condition) $d = 2 - d$. Thus $d = 1$, and $a = b$. Note that if $d = 0$, then $a + 1 = b \leqq 1$ implies $a \leqq 0$, so $a = 0$ and $b = 1$. But then $-y' = w'$, which says $(1/2)(w' + y') = 0 \in K_1'$. If $d = 2$, then $b + 1 = a \leqq 1$ implies $b = 0$ and $a = 1$. Then $x' = -t'$, and again $0 \in K_1'$. This completes the proof.

3. Affine homomorphic and isomorphic images of $\tilde{S}$ and $\tilde{\tilde{S}}$. In this section we are interested in homomorphic and isomorphic (as well as affine) images K of the convolution semigroups $\tilde{S}$ and $\tilde{\tilde{S}}$. The essential difficulties involved in the characterizations we obtain have already been solved in § 2, and the additional requirements are (primarily) that (a) K be a compact affine topological semigroup and (b) $E(K)$ or T be a compact topological semigroup.

The following lemma takes care of most of the additional difficulties encountered when one requires a topological semigroup structure on S and K.

LEMMA 3.1. *Let K be a compact affine topological semigroup, L a norm closed linear subspace of $L(K)$ separating points of K, and $T \subset E(K)$ be a compact sub-semigroup of K. Denote by $x \rightarrow x^0$ the embedding of K into L^*, giving L^* the weak-* topology determined by pointwise convergence on L, and let A* (let B) *respectively denote the closed convex symmetric hull of T^0 (the closed convex hull of T^0). Then:*

(1) *If R on $P(T)$ into L^* is defined (for $\mu \in P(T), l \in L$) by $R_\mu(l) = \int_T l d\mu$, then R is a continuous affine homomorphism of $\tilde{T}$ onto B.*

(2) *If A is contained in K^0 and has 0 as a zero and Q on $B(T)$ into L^* is defined (for $\mu \in B(T)$ and $l \in L$) by $Q_\mu(l) = \int_T l d\mu$, then Q is a continuous affine homomorphism of $\tilde{\tilde{T}}$ onto A. Note that R and Q are the mappings of Theorem* 2.1 *and* 2.3 *respectively, if $L = L(K)$ and $L_0(K)$ respectively.*

Proof. The statements regarding Q and R (except for those involving the homomorphism properties) are proved exactly as in Theorems 2.1 and 2.3. It therefore suffices, for example, to prove that Q is a homomorphism, so let $\mu, \nu \in B(T)$. Suppose first $\mu = \sum_{i=1}^m a_i\mu_i$, $\nu = \sum_{j=1}^n b_j\nu_j$, with $\sum_1^m a_i = 1 = \sum_1^n b_j$ and $a_i, b_j \geqq 0$, μ_i, ν_j extreme points of $B(T)$, all i and j. Then $\mu\nu = \sum_{i,j} a_i b_j \mu_i \nu_j$, so $Q_{\mu\nu} = \sum_{i,j} a_i b_j Q_{\mu_i\nu_j}$. If both μ_i and ν_j are point measures determined respectively by t_i and $s_j \in T$, then $Q_{\mu_i\nu_j} = Q_\alpha$, where α is the point measure determined by $t_i s_j \in T$. But then $Q_\alpha(l) = \int_T l d\alpha = l(t_i s_j) = (t_i s_j)^0(l) = (t_i^0 \cdot s_j^0)(l) = (Q_{\mu_i} \cdot Q_{\nu_j})(l)$; i.e., $Q_{\mu_i\nu_j} = Q_{\mu_i} \cdot Q_{\nu_j}$. Now if $\alpha, \beta \in B(T)$, then $-\alpha \cdot \beta = -(\alpha \cdot \beta) = \alpha \cdot (-\beta)$ and $(-\alpha) \cdot (-\beta) = \alpha \cdot \beta$. Thus if $\mu_i(\nu_j)$ is each the minus of a point measure, say $\mu_i = -\alpha_i$, $\nu_j = -\beta_j$, then $Q_{\mu_i\nu_j} = Q_{\alpha_i\beta_j} = Q_{\alpha_i} \cdot Q_{\beta_j} = x^0 \cdot y^0$, with $x, y \in T$. On the other hand $Q_{\mu_i} \cdot Q_{\nu_j} = (-x^0)\cdot(-y^0) = (2\cdot 0 - x^0)(2\cdot 0 - y^0) = 4\cdot 0 - 2\cdot 0 - 2 \cdot 0 + x^0 \cdot y^0 = x^0 \cdot y^0$; i.e., $Q_{\mu_i\nu_j} = Q_{\mu_i} \cdot Q_{\nu_j}$. If (say) μ_i and α_j are point measures, with $\nu_j = -\alpha_j$, then $Q_{\mu_i\nu_j} = -Q_{\mu_i\alpha_j} = -(Q_{\mu_i} \cdot Q_{\alpha_j}) = -(x^0 \cdot y^0)$, with $x, y \in T$. On the other hand $Q_{\mu_i} \cdot Q_{\nu_j} = x^0 \cdot (-y^0) = x^0 \cdot (2 \cdot 0 - y^0) = 2 \cdot 0 - (x^0 \cdot y^0) = -(x^0 \cdot y^0)$; i.e., in all cases, $Q_{\mu_i\nu_j} = Q_{\mu_i} \cdot Q_{\nu_j}$. Thus, $Q_{\mu\nu} = \sum_{i,j} a_i b_j Q_{\mu_i\nu_j} = (\sum_i a_i Q_{\mu_i}) \cdot (\sum_j b_j Q_{\nu_j}) = Q_\mu \cdot Q_\nu$. Suppose next that $\mu = \sum_{i=1}^n a_i\mu_i$, with μ_i extreme points of $B(T)$. Then $\mu\nu = \sum_{i=1}^n a_i\mu_i\nu$, so $Q_{\mu\nu} = \sum_{i=1}^n a_i Q_{\mu_i\nu}$. Let $\{\nu_a\}$ be a net of convex combinations of extreme points of $B(T)$ converging to ν; then $\mu_i\nu_a \rightarrow_a \mu_i\nu$, so Q continuous implies $Q_{\mu_i} \cdot Q_{\nu_a} = Q_{\mu_i\nu_a} \rightarrow_a Q_{\mu_i\nu}$. Since $Q_{\mu_i} \cdot Q_{\nu_a} \rightarrow_a Q_{\mu_i} \cdot Q_\nu$, it follows that $Q_{\mu_i\nu} = Q_{\mu_i} \cdot Q_\nu$, $1 \leqq i \leqq n$, hence $Q_{\mu\nu} = \sum_{i=1}^n a_i(Q_{\mu_i} \cdot Q_\nu) =$

$(\sum_{i=1}^{n} a_i Q_{\mu_i}) \cdot Q_\nu = Q_\mu \cdot Q_\nu$. Now let μ, ν be arbitrary, $\{\mu_a\}$ a net of convex combinations of extreme points converging to μ. Then, by the preceding, $Q_{\mu_a} \cdot Q_\nu = Q_{\mu_a \nu} \to_a Q_{\mu\nu}$, while $Q_{\mu_a} \cdot Q_\nu \to_a Q_\mu \cdot Q_\nu$. Thus, Q is a homomorphism and the argument for R is similar, though simpler.

THEOREM 3.1. *Suppose K is a compact affine topological semigroup with $L(K)$ separating points of K. Then K is the extremal homomorphic image of the convolution semigroup $\tilde{T}$ of measures over some compact Hausdorff semigroup T if and only if $E(K)$ is a compact semigroup.*

Proof. It is now obvious that the extremal homomorphic image of a $\tilde{T}$ has a compact semigroup of extreme points. For the converse, use Theorem 2.1 and Lemma 3.1, letting the L of Lemma 3.1 be $L(K)$, $T = E(K)$, and letting the mapping be the R of Lemma 3.1, part (1).

COROLLARY 3.1.1. *Suppose K is a compact affine topological semigroup with $L(K)$ separating points of K. Then K is group extremal (i.e., K has an identity and $E(K)$ is a compact group) if and only if K is the affine continuous homomorphic image of some $\tilde{T}$, with T a compact topological group.*

Proof. Suppose first that K has an identity element and $E(K)$ is a compact group. Now by Wendel's theorem [3, Theorem 1] the maximal group T containing the identity is contained in $E(K)$, hence $E(K) = T$. Now the mapping R of Theorem 3.1 and Lemma 3.1 (since T is a group and R is a homomorphism) is extremal, so K is the affine continuous and homomorphic image under R of $\tilde{T}$. If this condition holds, then K is the extremal image of a semigroup of measures over a compact group, hence $E(K)$ is a compact group and K has an identity. This completes the proof.

REMARK 3.1. The group extremal semigroups of the preceding corollary are known to always have a zero [13, 3]. Familiar examples of such semigroups are the closed unit disc of complex numbers, with ordinary complex multiplication, and the interval $[-1, 1]$ of reals, with ordinary multiplication.

In the following theorems S will always be a compact semigroup.

THEOREM 3.2. *The following conditions are mutually equivalent for the compact affine topological semigroup K:*

(1) *K is the one-to-one affine bicontinuous and isomorphic image of some probability semigroup $\tilde{S}$,*

(2) (a) *$E(K)$ is a compact topological semigroup, and* (b) *K is a simplex,*

(3) (a) *$L(K)$ separates points of K,* (b) *$E(K)$ is a compact semigroup, and* (c) *each continuous real function f on $E(K)$ is extendable to $\bar{f} \in L(K)$.*

Proof. If (1) holds, then Theorem 2.2 implies everything claimed in (2) save the statement that $E(K)$ is a semigroup, and this follows because $E(K)$ is the isomorphic image of a semigroup, namely S. Now the implication (2) $\rightarrow$ (3) follows directly from (2) and Theorem 2.2. To conclude the proof of the theorem, we show that (3) $\rightarrow$ (1). Here again we use Lemma 3.1, letting $T = E(K)$, $L = L(K)$, and the mapping be the R of Lemma 3.1. Note this function is the same as that used in (3) $\rightarrow$ (1) of Theorem 2.2. The result of applying Lemma 3.1 and Theorem 2.2 is that K is the one-to-one bicontinuous affine and isomorphic image of $\tilde{T}$ under R.

THEOREM 3.3. *Let K be a compact affine topological semigroup. The following conditions are mutually equivalent.*

(1) *K is the one-to-one affine bicontinuous and isomorphic image of some real unit ball semigroup $\tilde{\tilde{S}}$,*

(2) *the same as* (2) *of Theorem* 2.3 *except for the additional requirements that the z and T of that theorem be a zero for K and a semigroup, respectively,*

(3) *the same as* (3) *of Theorem* 2.3 *except requiring additionally that the z and T of that theorem be a zero for K and a semigroup, respectively.*

Proof. The only conditions which need to be checked (in virtue of Theorem 2.3) are those involving the semigroup structures on the spaces involved. Thus, in (1) $\rightarrow$ (2), the zero of $\tilde{\tilde{S}}$ maps onto z (hence z is a zero for K) and S maps onto T (hence T is a semigroup). The implication (2) $\rightarrow$ (3) follows immediately from Theorem 2.3 and the additional assumptions on z and T. To prove, finally, that (3) $\rightarrow$ (1), note that z maps into $0 \in L_0(K)^*$ under the embedding K into K'; thus, 0 is a zero for K'. Further, T maps onto T' and the closed convex symmetric hull A of T' in $L_0(K)^*$ is K'. In Lemma 3.1, then, we take $L = L_0(K)^*$, and let Q on $B(T)$ onto K' be as in that lemma. Then Lemma 3.1 together with Theorem 2.3 insure that K' is the one-to-one affine bicontinuous and isomorphic image of $\tilde{\tilde{T}}$, so also then is K. This completes the proof.

REMARK 3.2. A simple example illustrating the last two theorems may be constructed in the plane, as follows. Let the K of Theorem

3.3. be all pairs (x, y) of reals such that $|x| + |y| \leqq 1$, and let $T = \{i, j\}$, where $i = (1, 0)$, $j = (0, 1)$. Define $i^2 = i$, $j^2 = j$, $ij = ji = j$. The multiplication (on the entire plane) is defined as follows: $(ai + bj)(ci + dj) = aci + (ad + bc + bd)j$. Then K_1 is $\{(x, y): x, y \geqq 0, x + y = 1\}$, a simplex, and is affinely isomorphic with $[-1, 1]$ with usual multiplication. K itself is, of course, a unit ball semigroup of measures, with $z = (0, 0)$.

Examples of similar nature could be constructed on any finite simplex, of course, the requirements being that multiplications of a suitable nature be defined on the set of vertices. It is clear that exactly n distinct geometric figures exist in n-space on which probability semigroup structures can be defined; namely, the n simplexes each with i vertices, $2 \leqq i \leqq n + 1$. Thus the number of distinct probability semigroups in n-space is $= \sum_{i=2}^{n+1} A(i)$, where $A(i)$ is the number of distinct associative multiplications on a set of i elements (isomorphic and anti-isomorphic semigroups are identified).

REFERENCES

1. R. F. Arens and J. L. Kelley, *Characterizations of the space of continuous functions over a compact Hausdorff space*, Trans. Amer. Math. Soc., **62** (1947), 499-508.

2. G. Choquet, *Unicité des Représentations Intégrales au Moyen en points Extremaus dans les Cônes Convex Reticulés*, Comptes Rendues **243** (1956), 555-557.

3. H. Cohen and H. S. Collins, *Affine semigroups*, Trans. Amer. Math. Soc., **93** (1959), 97-113.

4. H. S. Collins, *Idempotent measures on compact semigroups*, Proc. Amer. Math. Soc., **13** (1962), 442-446.

5. ———, *Primitive idempotents in the semigroup of measures*, Duke Math. J., **27** (1960), 397-400.

6. ———, *The kernel of a semigroup of Measures*, Duke Math. J., **28** (1961), 387-392.

7. ———, *Convergence of convolution iterates of measures*, Duke Math. J., **29** (1962), 259-264.

8. ———, *Remarks on affine semigroups*, Pacific J. Math., **12** (1962), 449-455.

9. H. S. Collins and R. J. Koch, *Regular D-classes in measure semigroups*, Trans. Amer. Math. Soc., **105** (1962), 21-31.

10. I. Glickcberg, *Convolution semigroups of measures*, Pacific J. Math., **9** (1959), 51-67.

11. K. Ito and Y. Kawada, *On the probability distributions on a compact group*, Proc. Phys.-Math. Soc., Japan **22** (1940), 977-998.

12. L. H. Loomis, *Unique direct integral decompositions on convex sets*, Amer. J. Math., **86** (1962), 509-526.

13. J. E. L. Peck, *An Ergodic theorem for a noncommutative semigroup of linear operators*, Proc. Amer. Math. Soc., **2** (1951), 414-421.

14. J. S. Pym, *Idempotent measures on semigroups*, Pacific J. Math., **12** (1962), 685-698.

15. M. Rosenblatt, *Limits of convolution sequences of measures on a compact topological semigroup*, J. Math. and Mechanics **9** (1960), 293-306.

16. M. Rosenblatt and M. Heble, *Idempotent measures on a compact topological semigroup*, Proc. Amer. Math. Soc., **14** (1963), 177-184.

17. S. Schwarz, *On the structure of the semigroup of measures on a finite semigroup*, Czechoslovak Math. J., **7** (1957), 358-373.

18. ———, *Probability measures on noncommutative semigroups, general topology and its relations to modern analysis and algebra*, Czechoslvak Academy of Sciences, Prague, 1962.

19. Karl Stromberg, *Probabilities on a compact group*, Trans. Amer. Math. Soc., **94** (1960), 295–309.

20. A. D. Wallace, *Remarks on affine semigroups*, Bull. Amer. Math. Soc., **66** (1960), 110–112.

21. J. G. Wendel, *Haar measure and the semigroup of measures on a compact group*, Proc. Amer. Math. Soc., **5** (1954), 923–929.

LOUISIANA STATE UNIVERSITY

THE RELATIONSHIP BETWEEN THE RADICAL OF A LATTICE-ORDERED GROUP AND COMPLETE DISTRIBUTIVITY

PAUL CONRAD

1. Introduction. Throughout this note let G be a lattice-ordered group (notation 1-group). G is said to be *representable* if there exists an 1-isomorphism of G onto a subdirect sum of a cardinal sum of totally ordered groups (notation 0-groups). In particular, every abelian 1-group is representable. G is said to be *completely distributive* if for $g_{ij} \in G$

$$\bigwedge_{i \in I} \bigvee_{j \in J} g_{ij} = \bigvee_{f \in J^I} \bigwedge_{i \in I} g_{if(i)}$$

provided the indicated joins and intersections exist.

For each $0 \neq g$ in G let R_g be the subgroup of G that is generated by the set of all 1-ideals of G not containing g. Then R_g is an 1-ideal of G and the radical of G is defined to be

$$R(G) = \cap R_g \qquad (0 \neq g \in G) .$$

In [2] it is shown that if G is a divisible abelian 1-group, then there exists a minimal Hahn-type embedding of G into an 1-group of real valued functions if and only if $R(G) = 0$. Thus it would be useful to identify the class of abelian 1-groups with zero radicals, and to examine the properties of non-abelian 1-groups with zero radicals. In our main theorem we show that a representable 1-group G is completely distributive if and only if $R(G) = 0$. We also show $R(G) = 0$ if and only if G has a regular representation. This settles a question raised by Weinberg [6].

With no restrictions on G we show that $R(G)$ is completely determined by the lattice $\mathscr{L}$ of all 1-ideals of G. In particular, if G is a representable 1-group, then whether or not G is completely distributive depends only on $\mathscr{L}$.

The author would like to express his gratitude to A. H. Clifford who read a rough draft of this note and made valuable suggestions. In particular, the present forms of Lemmas 1 and 2 are due to him.

2. Regular and essential L-ideals. If $g \in G$ and M is an 1-ideal of G that is maximal with respect to $g \notin M$, then M is called a *regular* 1-*ideal* of G. Let M^* be the intersection of all 1-ideals of G that

Received June 6, 1963. This research was supported by a grant from the National Science Foundation.

properly contain M. Then since $g \in M^*$, it follows that M^* is the unique 1-ideal of G that covers M. Let Γ be an index set for the set of all pairs (G^γ, G_γ) of 1-ideals of G such that G_γ is regular and G^γ covers G_γ. Define $\alpha < \beta$ if $G^\alpha \subseteqq G_\beta$. Then Γ is a *po*-set, and we say that $\gamma \in \Gamma$ is a *value* of g if $g \in G^\gamma \backslash G_\gamma$. In particular, the set of all values of g is a trivially ordered subset of Γ. An element $\gamma \in \Gamma$ is called *essential* if there exists an $0 \neq h$ in G such that all the values of h are $\leqq \gamma$. In this case G_γ is called an *essential* 1-*ideal* of G, and if $g \in G^\gamma \backslash G_\gamma$, then we say that γ is an *essential value* of g.

Clearly the set E of all essential elements in Γ is a dual ideal of Γ ($\alpha < \beta \in \Gamma, \alpha \in E \rightarrow \beta \in E$). The following lemma shows that the radical $R(G)$ of G is completely determined by the essential ideals of G.

LEMMA 1. *The radical of G is the intersection of essential* 1-*ideals of G*: $R(G) = \cap G_\gamma (\gamma \in E)$.

Proof. If $g \notin R(G)$, then $g \notin R_h$ for some h in G and by Zorn's lemma there exists an 1-ideal M of G that is maximal with respect to $g \notin M \supseteqq R_h$. Thus $M = G_\gamma$ for some $\gamma \in E$, $g \in G^\gamma \backslash G_\gamma$ and hence g has an essential value. If $x \in \cap G_\gamma$, then x has no essential value and hence $x \in R(G)$. Therefore $\cap G_\gamma \subseteqq R(G)$. If E is the null set, then $G = \cap G_\gamma \supseteqq R(G)$ and if $\gamma \in E$, then there exists $0 \neq h_\gamma \in G$ such that if δ is a value of h_γ, then $\delta \leqq \gamma$ and hence $G_\delta \leqq G_\gamma$. Thus $R_{h_\gamma} \subseteqq G_\gamma$ and so

$$\bigcap_{\gamma \in E} G_\gamma \supseteqq \bigcap_{\gamma \in E} R_{h_\gamma} \supseteqq \bigcap_{0 \neq g \in G} R_g = R(G) .$$

COROLLARY. $R(G) = 0$ *if and only if each nonzero element in G has at least one essential value.*

We next show that $R(G)$ depends only on the lattice $\mathscr{L}$ of all 1-ideals of G. Note that a regular 1-ideal M of G is characterized by the fact that it is meet irreducible in $\mathscr{L}$. That is, if M^* is the intersection of all 1-ideals of G that properly contain M, then M is properly contained in M^*.

LEMMA 2. $\beta \in \Gamma$ *is essential if and only if* $\cap \{G_\gamma : \gamma \in \Gamma$ *and* $\gamma \not\leqq \beta\} \neq 0$.

Proof. Suppose that $0 < h \in \cap \{G_\gamma : \gamma \in \Gamma$ and $\gamma \not\leqq \beta\}$ and let α be a value of h. Then $h \notin G_\alpha$ and so $\alpha \leqq \beta$. Thus all the values of h are $\leqq \beta$, and hence β is essential. Conversely assume that G_β is essential and pick $0 < h \in G$ such that all the values of h are $\leqq \beta$. Then

$h \in \cap\{G_\gamma : \gamma \in \Gamma$ and $\gamma \not\leqq \beta\}$. For if $h \notin G_\gamma$, where $\gamma \not\leqq \beta$, then h must have a value $\alpha \geqq \gamma$ which is impossible.

COROLLARY. *$R(G)$ is an invariant of the lattice $\mathscr{L}$ of all 1-ideals of G.*

LEMMA 3. *For an 1-group G the following are equivalent.*
(1) *G/M is an 0-group for each regular 1-ideal M of G.*
(2) *G is representable.*

Proof. For each $0 \neq g$ in G pick an l-ideal M_g of G that is maximal with respect to not containing g. Then $\cap M_g = 0$, and if (1) is satisfied, then each G/M_g is an 0-group and the mapping of $x \in G$ upon $(\cdots, M_g + x, \cdots)$ is a representation of G. Conversely suppose that G has a representation, then clearly

(3) if $a, b \in G^+$ and $a \wedge b = 0$, then $a \wedge (-x + b + x) = 0$ for all $x \in G$. In fact, Sik [5] established that (2) and (3) are equivalent, but we only need that (2) implies (3). Let M be an 1-ideal of G that is maximal with respect to not containing $0 < a \in G$, and let $A = M + a$. Suppose (by way of contradiction) that G/M is not an 0-group. Then there exist strictly positive elements X and Z in G/M such that $X \wedge Z = M$.

Case I. $X \wedge A = M$. Then $P(A) = \{Y \in G/M : |Y| \wedge A = M\}$ is a convex 1-subgroup of G/M that contains X but not A. If $M < Y \in P(A)$, then $Y = M + y$, where $0 < y \in G$, and $a = a \wedge y + a'$, $y = a \wedge y + y'$, $a' \wedge y' = 0$. Moreover

$$M = A \wedge Y = M + a \wedge M + y = M + a \wedge y\ .$$

Thus $a \wedge y \in M$ and so $Y = M + y'$ and $A = M + a'$. But by (3), $a' \wedge (-g + y' + g) = 0$ for all g in G and hence $A \wedge -(M + g) + Y + (M + g) = M$. Thus $P(A)$ is a nonzero 1-ideal of G/M that does not contain A, and hence there exists an 1-ideal of G that properly contains M but not a, but this contradicts the maximality of M.

Case II. $X \wedge A \neq M$. Then $P(X)$ is an 1-ideal of G/M that contains Z but not A, and once again we contradict the maximality of M. Therefore G/M is an 0-group, and hence (2) implies (1).

COROLLARY. *If G is representable and $R(G) = 0$, then an element g is positive in G if and only $G_\gamma + g$ is positive for all essential values γ of g.*

Proof. If g is positive in G, then $G_\gamma + g$ is positive for all values γ of g, essential or otherwise. If g is not positive, then $g = g \vee 0 +$

$g \wedge 0 = g^+ + g^-$, where $g^- \neq 0$ and $g^+ \wedge -g^- = 0$. By the Corollary to Lemma 1 there exists an essential value γ of g^- and by Lemma 3, G/G_γ is an 0-group, and so $g^+ \in G_\gamma$. Thus γ is also an essential value of g and $G_\gamma + g = G_\gamma + g^-$ is negative.

LEMMA 4. *If* $0 < g \in \vee A_\lambda$, *where the* A_λ *are* 1-*ideals of* G, *then* $g = g_1 \vee \cdots \vee g_n$, *where* $0 \leqq g_i \in \cup A_\lambda$ *for* $i = 1, \cdots, n$.

Proof. This proof is due to T. Lloyd. Clearly $g = a_1 + \cdots + a_n$, where the $a_i \in A_{\lambda i}$ for $i = 1, \cdots, n$. Thus it suffices to show that $g \leqq a_1' \vee \cdots \vee a_n'$, where $a_i' \in A_{\lambda i}$ for $i = 1, \cdots, n$. For then

$$g = ((a_1' \vee 0) \wedge g) \vee \cdots \vee ((a_n' \vee 0) \wedge g)$$
$$= g_1 \vee \cdots \vee g_n$$

where $0 \leqq g_i \in A_{\lambda i}$ for $i = 1, \cdots, n$. If $n = 2$, then

$$a_1 + a_2 \leqq 2a_1 \vee (a_1 + a_2 - a_1 + a_2) = a_1' \vee a_2'$$

because

$$0 \leqq |a_1 - a_2| = (a_1 - a_2) \vee (a_2 - a_1)$$
$$= -a_1 + (2a_1 \vee (a_1 + a_2 - a_1 + a_2)) - a_2 \,.$$

Thus $a_1 + \cdots + a_n \leqq (a_1 + \cdots + a_{n-1})' \vee a_n'$, and since $(a_1 + \cdots + a_{n-1})' \in \vee A_{\lambda i}$ $(i = 1, \cdots, n-1)$, $(a_1 + \cdots + a_{n-1})' = b_1 + \cdots + b_{n-1}$, where $b_i \in A_{\lambda i}$ for $i = 1, \cdots, n-1$. Thus by induction $b_1 + \cdots + b_{n-1} \leqq a_1' \vee \cdots \vee a_{n-1}'$ and hence $g \leqq a_1' \vee \cdots \vee a_n'$.

3. Completely distributive *L*-groups. Let A be a sublattice and and subdirect sum of a cardinal sum B of 0-groups $B_\lambda (\lambda \in \Lambda)$. If for each λ in Λ, the projection ρ_λ of A onto B_λ preserves infinite joins, then A is called a *regular* subgroup of B. An 1-group G is said to have a *regular representation* if it is 1-isomorphic to a regular subgroup of a cardinal sum of 0-groups. It is easy to prove that an 1-group G with a regular representation is completely distributive [6]. Weinberg has also shown ([6] Proposition 1.3) that the natural homomorphism of an 1-group G onto G/J, where J is an 1-ideal of G, preserves infinite joins if and only if J is closed $(\vee j_\lambda \in G, \{j_\lambda : \lambda \in \Lambda\} \subseteqq J \rightarrow \vee j_\lambda \in J)$. Thus it follows that G has a regular representation if and only if there exists a family of closed 1-ideals J_λ of G such that $\cap J_\lambda = 0$ and each G/J_λ is an 0-group.

LEMMA 5. (Weinberg) *An* 1-*group* G *is completely distributive if and only if for each* $0 < g$ *in* G *there exists* $0 < g^*$ *in* G *such that*

$$g = \vee g_\lambda,\ g_\lambda \in G^+ \rightarrow g^* \leqq g_\lambda \text{ for some } \lambda.$$

THEOREM. *For a representable* l-*group* G *the following are equivalent.*

(1) $R(G) = 0$.

(2) *Each essential* l-*ideal of* G *is closed and* $\cap G_\gamma = 0$ $(\gamma \in E)$.

(3) G *has a regular representation.*

(4) G *is completely distributive.*

Proof. By Lemma 3, for each γ in E, G/G_γ is an 0-group, and hence by the preceding discussion (2) implies (3) and (3) implies (4). Suppose that G is completely distributive, and assume (by way of contradiction) that $0 < g \in R(G)$. Then by Lemma 5 there exists $0 < g^* \in G$ such that if $g = \vee g_\alpha$ $(g_\alpha \in G^+)$, then $g^* \leqq g_\alpha$ for some α. Since $g \in R(G)$ it follows that $g \in R_{g^*} = \vee A_\lambda$. where the A_λ are the l-ideals of G not containing g^*. Thus by Lemma 4, $g = g_1 \vee \cdots \vee g_n$, where $0 \leqq g_i \in \cup A_\lambda$. But then $g^* \leqq g_i$ for some i, and hence $g^* \in \cup A_\lambda$ *a* contradiction. Therefore (4) implies (1).

To complete the proof we must show that (1) implies (2). If (1) is satisfied, then by Lemma 1, $\cap G_\gamma = 0$ $(\gamma \in E)$. Let G_δ be an essential l-ideal of G and assume (by way of contradiction) that G_δ is not closed. Then there exists $g \in G^+ \backslash G_\delta$ such that $g = \vee g_j (g_j \in G_\delta^+)$. Since G_δ is essential there exists $0 < h \in G$ such that all the values of h are $\leqq \delta$. We shall show that for some such h, $g - h \geqq g_j$ for all j, and hence $\vee g_j > \vee g_j - h = g - h \geqq \vee g_j$.

Case I. There exists $0 < h \in G$ such that all the values of h are $\leqq \delta$ and $G_\delta + h < G_\delta + g$. Since $g - h \notin G_\delta$ and $g_j \in G_\delta$, $g - h - g_j \neq 0$. By the Corollary to Lemma 3 it suffices to show that $G_\beta + g - h - g_j$ is positive for all values β of $g - h - g_j$ in E. If $h \in G_\beta$, then $G_\beta + g - h - g_j = G_\beta + g - g_j$ is positive. If $h \notin G_\beta$, then there exists a value γ of h such that $\gamma \geqq \beta$. But then $\beta \leqq \gamma \leqq \delta$, and since $g - h - g_j \in G^\beta \backslash G_\delta$, $\beta = \delta$. Therefore $G_\beta + g - h - g_j = G_\delta + g - h$ is positive.

Case II. For each $0 < h \in G$ such that all of the values of h are $\leqq \delta$, $G_\delta + h \geqq G_\delta + g$. If $\delta > \gamma \in E$, then we may choose $0 < k \in G$ such that all of the values of k are $\leqq \gamma < \delta$. But then $G_\delta + g > G_\delta = G_\delta + k$. Therefore δ is minimal in E. If all values of $0 < h$ are $\leqq \delta$, then $G_\delta + h \geqq G_\delta + g$ and so $G_\delta + g \wedge h = G_\delta + g$. If β is a value of $g \wedge h$ in E, then $g \wedge h \in G^\beta \backslash G_\beta$ and hence $h \notin G_\beta$. Thus there exists a value γ of h such that $\beta \leqq \gamma \leqq \delta$ and since δ is minimal in E, $\beta = \delta$. Thus without loss of generality, $0 < h \in G$, δ is the only value of h in E and $G_\delta + h = G_\delta + g$. If $g - h - g_j \neq 0$ and β is a value of $g - h - g_j$ in E then $h \in G_\beta$. Otherwise $\beta = \delta$, but $g - h - g_j \in G_\delta$. Therefore $G_\beta + g - h - g_j = G_\beta + g - g_j$ is positive for all values β

of $g - h - g_j$ in E. This completes the proof of our theorem. In proving that (4) implies (1) we did not use the hypothesis that G is representable. Thus we have

COROLLARY I. *If G is a completely distributive* 1-*group, then* $R(G) = 0$.

From the Corollary to Lemma 2 we have

COROLLARY II. *If G is a representable* 1-*group, then whether or not G is completely distributive depends only on the lattice $\mathscr{L}$ of all* 1-*ideals of G.*

4. Remarks and examples. Let P be the 1-group of all order preserving permutations of the real line (with $fg(x) = f(g(x))$ and f positive if $f(x) \geqq x$ for all x). Let

$A = \{f \in P : f$ induces the identity on $(-\infty, a]$ for some $a\}$, and

$B = \{f \in P : f$ induces the identity on $[a, \infty)$ for some $a\}$.

Let $C = A \cap B$. Then Holland [4] has shown that A, B and C are the only proper 1-ideals of G, and Higman [3] has shown that C is algebraically simple. Therefore 0 is the only essential 1-ideal of C and since $C/0$ is not an 0-group it follows from Lemma 3 that C is not representable. Therefore C satisfies property (2) of the theorem, but not property (3).

(G, B) is the only value of each element in $A \backslash B$ and $(C, 0)$ is the only value of each nonzero element in C. Thus B and 0 are essential 1-ideals of P, and in particular, P satisfies (1). For each $n = 1, 2, \cdots$ let

$$f_n(x) = \begin{cases} 2x & \text{if } x \leqq n \\ \dfrac{x + 3n}{2} & \text{if } n \leqq x \leqq 3n \\ x & \text{if } 3n \leqq x \, . \end{cases}$$

Then $(\vee f_n)(x) = 2x$, and hence the f_n belong to B, but $\vee f_n \notin B$. Therefore P satisfies (1) but not (2).

A simple application of Lemma 5 shows that P is completely distributive (or see [6] Example 3.3). Therefore (4) does not imply (2) or (3). On the other hand for arbitrary 1-groups, (3) $\rightarrow$ (2) $\rightarrow$ (1). The remaining question is whether or not (1) or (2) implies (4) for non-representable 1-groups? Note that if $R(G) = 0$ implies complete distributivity, then every 1-group with no proper 1-ideals is completely distributive, and in particular, every 1-group that is algebraically simple is completely distributive.

If the radical used in this note is replaced by one constructed in exactly the same way, but with 1-ideals replaced by convex 1-subgroups, then if this new radical is zero, the group is completely distributive. Also the new radical is an invariant of the lattice of all convex 1-subgroups of G. The proofs of these statements are analogous to those in this paper using the fact that if C is a regular convex 1-subgroup, then the set of right cosets of C in G is totally ordered by

$$C + x \leqq C + y \text{ if } x \leqq y + c \text{ for some } c \in C.$$

Unfortunately the converse to the above is false. For example, the new radical for P is P itself and yet P is completely distributive.

Let G be an Archimedean 1-group. By Theorem 5.7 in [2], $R(G) = 0$ if and only if G has a basis, and by Theorem 7.3 in [1], G has a basis if and only if G is (isomorphic to) a subdirect sum of a cardinal sum of subgroups R_γ of the reals which contains the finite cardinal sum of the R_γ. Thus we have a new proof for one of the main results in [6].

REFERENCES

1. P. Conrad, *Some structure theorems for lattice-ordered groups*, Trans. Amer. Math. Soc., **99** (1961), 212–240.

2. P. Conrad, J. Harvey and C. Holland, *The Hahn embedding theorem for lattice-ordered groups*, Trans. Amer. Math. Soc., **108** (1963), 143–169.

3. G. Higman, *On infinite simple permutation groups*, Publ. Math. Debrecen 3, (1953–54), 221–226.

4. C. Holland, *The lattice-ordered group of automorphisms of an ordered set*, Michigan Math. J., **10** (1963), 399–408.

5. F. Sik, *Über Summen einfach geordneter Gruppen*, Czech. Math. J., **8** (1958), 22–53.

6. E. C. Weinberg, *Completely distributive lattice-ordered groups*, Pacific J. Math., **12** (1962), 1131–1137.

A SUFFICIENT CONDITION THAT AN ARC IN S^n BE CELLULAR

P. H. DOYLE

An arc A in S^n, the n-sphere, is cellular if $S^n - A$ is topologically E^n, euclidean n-space. A sufficient condition for the cellularity of an arc in E^3 is given in [4] in terms of the property local peripheral unknottedness (L.P.U) [5]. We consider a weaker property and show that an arc in S^n with this property is cellular.

If A is an arc in S^n we say that A is p-shrinkable if A has an end point q and in each open set U containing q in S^n, there is a closed n-cell $C \subset U$ such that q lies in Int C (the interior of C), while BdC (the boundary of C) meets A in exactly one point. We note that A is p-shrinkable is precisely the condition that A be L.P.U. at an endpoint [5]. There is, however, a good geometric reason for using the p-shrinkable terminology here; the letter p denotes pseudo-isotopy.

LEMMA 0. *Let C^n be a closed n-cell and D^n a closed n-cell which lies in int C^n except for a single point q which lies on the boundary of each n-cell. If there is a homeomorphism h of C^n onto a geometric n-simplex such that $h(D^n)$ is also an n-simplex, then there is a pseudo-isotopy ρ_t of C^n onto C^n which is the identity on BdC^n, while $\rho_1(D^n)$, the terminal image of D^n, is the point q.*

The proof of this is omitted since it depends only on the same result when C^n and D^n are simplices.

LEMMA 1. *Let C^n be a closed n-cell and B an arc which lies in int C^n except for an endpoint b of B on BdC^n. Then there is a pseudo-isotopy of C^n onto C^n which is fixed on BdC^n and which carries B to b.*

Proof. Since $B \cap BdC^n = b$ we note that there is in C^n an n-cell D^n which contains B in its interior except for the point b, $D^n - b \subset \text{Int}\, C^n$, and D^n is embedded in C^n as in Lemma 0. Thus Lemma 0 can be applied to shrink B in the manner required by the Lemma.

THEOREM 1. *Let A be an arc in S^n such that for each subarc B of A, B is p-shrinkable. Then every arc in A is cellular.*

Proof. The proof is by contradiction. If A contains a non-cellular

Received January 30, 1963.

subarc there is no loss of generality in assuming this arc is A. Then $S^n - A \neq E^n$. By the characterization theorem of E^n in [1], there is a compact set C in $S^n - A$ and C lies in no open n-cell in $S^n - A$. By the Generalized Schoenflies Theorem [2], this is equivalent to the condition that no bicollared $(n-1)$-sphere in S^n separates C and A.

Let G be the set of all subarcs of A which cannot be separated from C by a bicollared sphere in S^n. We partially order G by set inclusion and select a maximal chain in G. Let B be the intersection of all arcs in this maximal chain. Evidently B cannot be separated from C by a bicollared sphere in S^n. Thus B is an arc and each proper subarc of B can be so separated from C in S^n.

By the hypothesis of the theorem, B is p-shrinkable. So let B be L.P.U. at an endpoint q. Let U be an open set containing q and $U \cap C = \square$. Then there is an n-cell $C^n \subset U$, $C^n \cap B = B^1$, an arc, while $B^1 \cap BdC^n = p$, a point. So by Lemma 1 there is a pseudo-isotopy ρ_t of S^n onto S^n, ρ_t is the identity in $S^n - C^n$, and $\rho_1(B^1) = p$. But $\rho_1(B)$ is a proper subarc of B which cannot be separated from C in S^n by a bicollared sphere. But this is a contradiction. Thus A is cellular as well as each subarc of A.

COROLLARY 1. *Let A be an arc in S^n which is the union of two p-shrinkable arcs, $A_1 \cup A_2$, which meet in a common endpoint p. Then A is cellular if A_1 is* L.P.U.

Proof. Each subarc of A is p-shrinkable.

COROLLARY 2. *Each non-cellular arc A in S^n contains a subarc which is not* L.P.U. *at either of its endpoints.*

Even in S^3 there is a difference between an arc being L.P.U. at each point and having the p-shrinkable property for each subarc. The simplest example is perhaps a mildly wild arc which is not a Wilder arc. [3].

REFERENCES

1. M. Brown, *The monotone union of open n-cells is an open n-cell,* Proc. Amer. Math. Soc., **12** (1961), 812–814.

2. ———, *A proof of the generalized Schoenflies theorem,* Bull. Amer. Math. Soc., **66** (1960), 74–76.

3. R. H. Fox, O. G. Harrold, *The Wilder arcs, Topology of* 3-*manifolds and related topics,* Prentice-Hall, (1962).

4. O. G. Harrold, *The enclosing of simple arcs and curves by polyhedra in* 3-*space,* Duke Math. J., **21** (1959), 615-622.

5. ———, *Combinatorial structures, local unknottedness, and local peripheral unknottedness. Topology of 3-Manifolds and related topics,* Prentice-Hall, (1962).

VIRGINIA POLYTECHNIC INSTITUTE

INTRINSIC EXTENSIONS OF RINGS

CARL FAITH AND YUZO UTUMI

A module M is an *essential extension* of a submodule N in case $K \cap N \neq 0$ for each nonzero submodule K of M. If S is a subring of a ring R, and if ${}_SR$, ${}_SS$ denote the left S-modules naturally defined by the ring operations of R, then R is a *left quotient ring of* S in case ${}_SR$ is an essential extension of ${}_SS$.

We shall discuss the following problem: (1) Characterize the condition that a ring extension R of S is a left quotient ring of S wholly in terms of the relative left ideal structures of R and S.

A ring extension R of S is *left intrinsic over* S in case $K \cap S \neq 0$ for each nonzero left ideal K of R. Evidently each left quotient ring R of S is left intrinsic over S but an obvious example (when R is a field and S a subfield $\neq R$) shows that the converse fails. Nevertheless, we ask: (2) When is the condition R is left intrinsic over S a solution to (1)?

We now specialize S by requiring that:

(i) S possesses a left quotient ring which is a (von Neumann) regular ring, or equivalently (R. E. Johnson [2]) by requiring that the left singular ideal of S vanishes. For such a ring there exists a maximal left quotient ring $\hat{S}$ which is unique up to isomorphism over S, and which is itself a regular ring ([2]). To eliminate the field example we require that:

(ii) $\hat{S}$ possesses no strongly regular ideals $\neq 0$. Under these hypotheses we present the following solution to (1).

A. THEOREM (2.6). *Let S satisfy* (i) *and* (ii). *Then an extension ring R of S is a left quotient ring of S if and only if R is a left intrinsic extension of S such that for each closed left ideal A of S there corresponds a left ideal B of R such that $B \cap S = A$.*

(See § 1 for definitions.)

Regarding (2) we add a rather dubious final hypothesis:

(iii) $\hat{S}$ is right intrinsic over S.

B. THEOREM (3.1). *If S satisfies* (i)–(iii), *then an extension ring R of S is a left quotient ring of S if and only if R is left*

Received May 15, 1963. The first author gratefully acknowledges partial support from the National Science Foundation under grants G-19863 and G-21514.

intrinsic over S. (Then there exists a ring monomorphism of R into $\hat{S}$ which is the identity on S.)

Combining B with a theorem of Goldie [1] we obtain:

C. THEOREM (3.2). *Let* S *be a prime ring which is both left and right noetherian, and assume that* S *is not an integral domain. Let* Q *denote the classical quotient ring of* S ([1]). *Then an extension ring* R *of* S *is left intrinsic over* S *if and only if there exists a ring monomorphism of* R *into* Q *which is the identity on* S.

1. A ring S is *strongly regular* (resp. *regular*) if for any $x \in S$ there exists $y \in S$ such that $x^2y = x$ (resp. $xyx = x$); an ideal I of S is *strongly regular* if I is a strongly regular ring.

Let S be a ring. Then ${}_SM$ will denote that M is a left S-module, and ${}_SS$ denotes the left S-module defined naturally by the ring operations in S. ${}_SM$ is an *essential extension of a submodule* N in case $K \cap N \neq 0$ for each submodule $K \neq 0$ of M. Then, N is said to be an *essential submodule of* M. An *essential left ideal* of S is a left ideal of S which is an essential submodule of ${}_SS$. (Thus a left ideal I of S is essential if and only if S is a left intrinsic extension of I.)

An element $x \in {}_SM$ is *singular* in case the annihilator of x in S is an essential left ideal of S. It is known that the set $Z({}_SM)$ of singular elements of ${}_SM$ is a submodule of ${}_SM$, called the *singular submodule of* ${}_SM$; $Z({}_SS)$ is an ideal of S called the *left singular ideal of* S.

If $Z({}_SS) = 0$, then S is said to be a J_l*-ring*, and $\hat{S}$ denotes its maximal left quotient ring; $\hat{S}$ is a regular ring with identity, and is left self-injective. If R is any left self-injective ring with identity, then it is known (Utumi [4], Lemma 8) that $Z({}_RR)$ coincides with the Jacobson radical $J(R)$ of R, and that the difference $R - J(R)$ is a regular ring.

A left ideal A of a ring S is *closed* if there is no left ideal of S which is a proper essential extension of A.

In case S is a J_l-ring it is known that the set of closed left ideals of S forms a complete complemented modular lattice $L(S)$. If R is a left quotient ring of S, R is also a J_l-ring, and $L(R)$ is isomorphic to $L(S)$ by the correspondence $A(\in L(R)) \rightarrow A \cap S$. Thus, in this case, the following condition is satisfied:

(1.1) *Condition*. Let R be an extension ring of a ring S. For any closed left ideal A of S there is a left ideal B of R such that $B \cap S = A$.

We call an extension ring R of S *left strongly intrinsic* if R is a left intrinsic extension of S, and if Condition 1.1 is fulfilled.

(1.2) LEMMA. *Let R be a left intrinsic extension of a J_l-ring S. Suppose that the maximal left quotient ring of S is right intrinsic over S. Then R is a left strongly intrinsic extension of S.*

Proof. Let A be a closed left ideal of S. By [5, Theorem 2.2] A is an annihilator left ideal of S. Hence $B \cap S = A$ for some annihilator left ideal B of R, as desired.

It is evident that Condition 1.1 is equivalent to the following:

(1.3) If A and B are left ideals of S such that $A \cap B = 0$, then there exists a left ideal C of R such that $C \supset A$ and $C \cap B = 0$.

Let R be a left strongly intrinsic extension of a ring S. Then the following three properties are easily seen:

(1.4) If $A \cap B = 0$ for a left ideal A of R and a left ideal B of S, there is a left ideal C of R such that $A \cap C = 0$.

(1.5) If the sum of left ideals $\{A_i\}$ of S is direct, so is the sum of $\{A_i + RA_i\}$.

(1.6) If A is an essential left ideal of S, then $A + RA$ is an essential left ideal of R. If C is an essential left ideal of R, $C \cap S$ is also an essential left ideal of S.

(1.7) LEMMA. *Let R be a left strongly intrinsic extension of a ring S. Then the following properties are equivalent:*

(i) *S is a J_l-ring;*

(ii) *R is a J_l-ring;*

(iii) *the singular submodule $Z({}_SR)$ of the left S-module R is zero.*

Proof. By (1.6) it is obvious that $Z({}_SR) \subset Z({}_RR) \subset Z({}_SR)$. Hence $Z({}_SR) = Z({}_RR)$. This shows the equivalence of (ii) and (iii). Now $Z({}_SS) = Z({}_SR) \cap S = Z({}_RR) \cap S$. Since $Z({}_RR)$ is an ideal of R, $Z({}_RR) = 0$ if and only if $Z({}_RR) \cap S = 0$. Therefore (i) is equivalent to (ii), as desired.

The following proposition is known:

(1.8) Let S be a J_l-ring. A left ideal A of S is closed if and only if $Bx \subset A$ for $x \in S$ and an essential left ideal B of S implies that $x \in A$.

(1.9) LEMMA. *Let R be a left strongly intrinsic extension of a J_l-ring S. Then the lattice $L(R)$ of closed left ideals of R is isomorphic to the lattice $L(S)$ of closed left ideals of S under contraction $A \to A \cap S$.*

Proof. It is direct from (1.6) and (1.8) that $A \cap S$ is closed for any closed A. Suppose that $A_1 \cap S = A_2 \cap S$ for $A_1, A_2 \in L(R)$. It is known that any intersection of closed left ideals of a J_l-ring is closed. Thus, $(A_3 =)\ A_1 \cap A_2 \in L(R)$. Let B be a left ideal of R such that $B \subset A_1$, $B \cap A_3 = 0$. It follows then that $B \cap S = 0$. Hence $B = 0$, which shows that $A_1 = A_3$, since A_3 is closed. Similarly $A_2 = A_3$, and therefore $A_1 = A_2$. Finally we shall show that the correspondence is onto. Let $C \in L(S)$. By (1.1), $C = D \cap S$ for some left ideal D of R. By Zorn's lemma there exists a maximal left ideal E of R such that $E \cap S = C$. Let F be a left ideal of R which contains E properly. Then $F \cap S \neq C$. Since C is closed, we can find a nonzero left ideal G of S such that $G \subset F \cap S$ and $G \cap C = 0$. By (1.1) there is a left ideal H of R such that $H \cap S$ is an essential extension of G. Then $0 = (H \cap S) \cap C = ((H \cap F) \cap E) \cap S$. Since R is left intrinsic over S, we have that $(H \cap F) \cap E = 0$. This implies that F is not an essential extension of E. Therefore $E \in L(R)$, completing the proof.

2. The following proposition is easily verified:

(2.1) Let M be a left S-module with zero singular submodule. Suppose that a left ideal A of S is an essential extension of a left ideal B of S. Let v and w be left S-homomorphisms of A into M. If $(v - w)B = 0$, then $v = w$.

In fact, $v - w$ induces a homomorphism of A/B into M. By assumption $Z(A/B) = A/B$. Hence $(v - w)A = Z((v - w)A) \subset Z(M) = 0$, as desired.

A left S-module M is called injective if for any left ideal A of S, and for any left S-homomorphism v of A into M there exists an element x such that $v(a) = ax$ for every $a \in A$. A ring S is called left self injective if the left S-module S is injective, and S has a unit element. Any left self injective ring which is semisimple (in the sense of Jacobson) is regular (in the sense of von Neumann). As is known, the maximal left quotient ring of a J_l-ring is semisimple, in the sense of Jacobson, and left self injective.

We denote by $l(P, Q)$ the set of $x \in P$ with $xQ = 0$. Similarly $r(P, Q)$ denotes the right annihilator, in P, of Q.

(2.2) LEMMA. *Let R be a semisimple left self injective ring, and suppose that it is a left strongly intrinsic extension of a ring*

S. Then the left S-module R is an injective module with zero singular submodule.

Proof. Since R is regular, it is a J_l-ring. By Lemma (1.7), $Z({}_SR) = 0$. Let A be a left ideal of S, and v a left S-homomorphism of A into R. Denote the left ideal of S generated by an element x by $(x)_l$. By Zorn's lemma there is a maximal subset $\{x_i\}$ of A such that the sum B of $(x_i)_l$ is direct. Evidently A is an essential extension of B. By (1.5), the sum of $\{Sx_i\}$ is also direct. Now $l(S, x_i) \subset l(S, v(x_i))$, that is, $l(R, x_i) \cap S \subset l(R, v(x)) \cap S$. Since any annihilator left ideal of the J_l-ring R is closed, it follows by Lemma 1.9 that $l(R, x_i) \subset l(R, v(x_i))$. This shows that $x_i \to v(x_i)$ generates a left R-homomorphism w of $\sum_i Rx_i$ into R. By the injectivity of ${}_RR$ there exists an element $a \in R$ such that $v(x_i) = x_i a$ for every i. Thus the homomorphism v and the right multiplication of a coincide on B. Since A is essential over B, it follows by (2.1) that v is given by the right multiplication of a, completing the proof.

(2.3) THEOREM. *Let R be an extension ring of a ring S, and suppose that the left S-module R is an injective module with zero singular submodule. Then S is a J_l-ring. Let T be the maximal one among such submodules of the left S-module R that are essential over the left S-module S. Then T forms a subring of R, and in fact it is the maximal left quotient ring of S.*

Proof. Since $Z({}_SS) = Z({}_SR) \cap S = 0$, S is a J_l-ring. Let E be the endomorphism ring of ${}_SR$, and let $v \in E$. In case ${}_SR$ is essential over Ker v, $Z({}_SR/\text{Ker } v) = R/\text{Ker } v$, and so $Im\, v = Z(Im\, v) \subset Z(R) = 0$, whence $v = 0$. In view of [4, Lemma 8] it follows from this that E is semisimple, and T is uniquely determined. Since ${}_ST$ is essential over ${}_SS$, it is easy to see that the set $D(x)$ of elements y of S such that $yx \in S$ is an essential left ideal of S for each $x \in T$. Now we denote by U the set of $x \in R$ such that $D(x)$ is an essential left ideal of S. Clearly $T \subset U$. It is not difficult to show that U is a subring of R. Since $Z({}_SR) = 0$, ${}_SU$ is essential over ${}_SS$. Hence $U \subset T$, therefore $T = U$. Thus, T forms a subring of R. T is the maximal left quotient ring of S because ${}_ST$ is the maximal essential extension of ${}_SS$.

(2.4) LEMMA. *Let R be a ring with unit, and suppose that it is a left strongly intrinsic extension of a semisimple left self injective ring S. Then every idempotent of R belongs to S.*

Proof. Let $e = e^2 \in R$. Then $Re \in L(R)$, the lattice of closed left ideals of R. Hence $Re \cap S \in L(S)$, the lattice of closed left ideals of

S, by Lemma 1.9. Since every closed left ideal of S is a principal left ideal generated by an idempotent, $S \cap Re = Sf$ for some $f = f^2 \in S$. Evidently $Rf \cap S = Sf = Re \cap S$, and hence $Re = Rf$ by Lemma 1.9. Similarly we can find an idempotent $g \in S$ such that $R(1-e) = Rg$. Since S is regular, $Sf + Sg = Sh$ for some $h = h^2 \in S$. Then $Rh = R$, and $h = 1$. Hence $1 = xf + yg$ for some $x, y \in S$. Also,

$$e - xf = e - (1 - yg) = yg - (1 - e) \in Rg + R(1 - e) = R(1 - e) .$$

Hence $e - xf \in Re \cap R(1-e) = 0$, and therefore $e = xf \in S$, completing the proof.

By virtue of [5, Corollary to Theorem 4], any semisimple left self injective ring R is decomposed into the direct sum of two ideals R_1 and R_2 in such a way that R_1 is strongly regular, and R_2 does not contain any nonzero strongly regular ideals. The decomposition is unique. By [5, Theorem 2], R_2 is generated by idempotents.

(2.5) THEOREM. *Let R be semisimple left self injective ring, and let $R = R_1 \oplus R_2$ be the decomposition into ideals mentioned above. Suppose that R is a left strongly intrinsic extension of a ring S. Then there is a subring T of R_1 with the following properties*:

(i) *T contains every idempotent of R_1,*

(ii) *T is a strongly regular, (left) self injective ring and*

(iii) *$T \oplus R_2$ is the maximal left quotient ring of S.*

Proof. By Lemma 2.2, ${}_SR$ is injective and $Z({}_SR) = 0$. Thus, by Lemma 2.3, R contains as a subring the maximal left quotient ring Q of S. Since R is left intrinsic over S, and $R \supset Q \supset S$, it is evident that R is left intrinsic over Q. Let A be a closed left ideal of Q. Then $A = Qe$, $e = e^2 \in Q$, and hence $A = Re \cap Q$. This shows that R is left strongly intrinsic over Q. Thus, by Lemma 2.4 every idempotent of R belongs to Q. Since R_2 is generated by idempotents, $R_2 \subset Q$, and so $Q = (Q \cap R_1) \oplus R_2$. Set $T = Q \cap R_1$. Since Q is regular, so is its ideal T. Thus, the strong regularity of T follows from the fact that a regular ring is strongly regular if and only if it has no nonzero nilpotent elements, completing the proof.

(2.6) THEOREM. *Let S be a J_l-ring. Suppose that the maximal left quotient ring of S does not contain any nonzero strongly regular ideals. Then any left strongly intrinsic extension of S is a left quotient ring of S.*

Proof. Let R be a left strongly intrinsic extension of S, and denote by Q the maximal left quotient ring of R. By Lemma 1.9,

Q is a left strongly intrinsic extension of S. Let $Q = Q_1 \oplus Q_2$ where Q_1 is a strongly regular ideal of Q, and Q_2 is an ideal of Q which does not contain any nonzero strongly regular ideals. By Theorem 2.5 there is a strongly regular subring T of Q_1 such that $T \oplus Q_2$ is the maximal left quotient ring of S. T contains every idempotent of Q_1, especially the unit element of Q_1. Since $T = 0$ by assumption, it follows that $Q_1 = 0$. Thus, $Q = Q_2$ is the maximal left quotient ring of S. Since $Q \supset R \supset S$, R is a left quotient ring of S, as desired.

3. The following is the main theorem.

(3.1) THEOREM. *Let S be a J_l-ring. Suppose that the maximal left quotient ring of S is right intrinsic, and does not contain any nonzero strongly regular ideals. Then any left intrinsic extension of S is a left quotient ring.*

Proof. By Lemma 1.2 any left intrinsic extension of S is left strongly intrinsic over S, and hence it is a left quotient ring of S by Theorem 2.6, as desired.

Goldie proved in [1; Theorem 13] that if a prime ring S satisfies the maximum conditions for left and right ideals, then S may be imbedded into a simple ring R with minimum condition in such a way that R is the classical quotient ring of S in the following sense:

(i) Every non-zero-divisor of S has the inverse in R;

(ii) every element x of R is of the forms $a^{-1}b$ and cd^{-1} for some a, b, c and $d \in S$. In this case S is a J_l-ring, and R is the maximal left quotient ring of S. Since R is a right quotient ring of S, it is right intrinsic over S. R contains a nonzero strongly regular ideal if and only if R is a division ring, that is, S is an Ore domain. Thus, by Theorem 3.1 we obtain the following.

3.2. THEOREM. *Let S be a prime ring with maximum conditions for left and right ideals, and suppose that it is not an Ore domain. Then an extension ring of S is left (or right) intrinsic over S if and only if it is isomorphic, over S, to a between ring of S and the classical quotient ring of S.*

REFERENCES

1. A. W. Goldie, *The structure of prime rings under ascending chain conditions*, Proc. London Math. Sec., (3) **8** (1958), 589-608.
2. R. E. Johnson, *The extended centralizer of a ring over a module*, Proc. Amer. Math. Soc., **2** (1951), 891-895.
3. Yuzo Utumi, *On quotient rings*, Osaka Math. J., **8** (1956), 1-18.
4. ———, *On a theorem on modular lattices*, Proc. Japan Acad., **35** (1959), 16-21.

5. Yuzo, Utumi, *On continuous regular rings and semisimple self injective rings*, Can. J. Math., **12** (1960), 597-605.
6. ———, *On rings of which any one-sided quotient rings are two-sided*, Proc. Amer. Math. Soc., **14** (1963), 141-147.

THE INSTITUTE FOR ADVANCED STUDY AND RUTGERS UNIVERSITY,
THE UNIVERSITY OF ROCHESTER

AN APPROXIMATE GAUSS MEAN VALUE THEOREM

W. FULKS

1. Introduction. The mean value theorem of Gauss, and its converse, due to Koebe, have long been known to characterize harmonic functions. Since any second order homogeneous elliptic operator L can, by an appropriate linear change of variables, be reduced (at a given point) to the Laplacian, it seems reasonable to expect that solutions of $Lu = 0$ should, when averaged over appropriate small ellipsoids, satisfy an approximate Gauss-type theorem, and one could hope that such a mean value property would characterize the solutions of the equation.

It turns out that this is the case. In fact the operator need not be elliptic, but may be parabolic, or of mixed elliptic and parabolic type. While the methods used here do not permit the weak smoothness conditions on the solutions admitted by Koebe's theorem, the result is stronger than might be expected in that no smoothness, not even measurability, is required of the coefficients of L: they need only be defined.

Since the result applies to parabolic equations, it seems of interest to examine the heat equation, for it can be cast in the required form. This leads to a characterization of its solutions in terms of averages over parabolic arcs.

2. The basic theorem. In the following $D_i = \partial/\partial y_i$, $D_{ij} = \partial^2/\partial y_i \partial y_j$, $u_{,ij} = D_{ij}u$, and ∇_y is the gradient operator with respect to the components of y.

It is convenient to consider equations of the form $Lu = f$, where f need only be defined, and may depend on u and any of its derivatives.

LEMMA. *Let $A = [a_{ij}]$ be an $n \times n$ constant nonnegative definite symmetric matrix, and denote by $B = [b_{ij}]$ the unique nonnegative definite symmetric square root of A. Let u be defined in a neighborhood of a point y in E_n, and be twice differentiable at y. For this y define the quadratic function q of x by*

$$q(x) \equiv (Bx \cdot \nabla_y)^2 u(y) \ .$$

Then the sum of the coefficients of the squared terms of $q(x)$ is $\sum_{i,j} a_{ij} u_{ij}(y)$.

Received July 24, 1963. This paper was prepared at Oregon State University under an NSF grant.

Proof. We have

$$q(x) = (Bx\cdot\nabla)^2 u = \Big(\sum_{i,m} b_{im}x_m D_i\Big)\Big(\sum_{j,k} b_{jk}x_k D_j\Big)u = \sum_{k,m}\Big(\sum_{ij} b_{im}b_{jk}u_{,ij}\Big)x_k x_m .$$

The sum of the coefficients of the squared terms is then

$$\sum_k\Big(\sum_{i\,j} b_{ik}b_{jk}\Big)u_{,ij} = \sum_{i,j}\Big(\sum_k b_{ik}b_{kj}\Big)u_{,ij} = \sum_{i,j} a_{ij}u_{,ij} .$$

THEOREM. *Let $L = \sum_{i,j} a_{ij}(y)D_{ij}$ be a well defined symmetric differential operator with a nonnegative definite matrix $A(y) = [a_{ij}(y)]$ in an open region R in E_n. Let $B(y) = [b_{ij}(y)]$ be the unique nonnegative definite square root of A, and for $y \in R$ and r sufficiently small, define*

$$(1)\qquad u_r(y) \equiv \frac{1}{\Omega_r}\int_{|x|=r} u(y + B(y)x)d\Omega_r$$

where Ω_r is the area of the sphere $\{|x| = r\}$. Let u be a function defined in a neighborhood of a point $y_0 \in R$, which is twice differentiable at y_0. Then for u to be a solution of $Lu = f$ at y_0 it is necessary and sufficient that

$$(2)\qquad u_r(y_0) = u(y_0) + C_n r^2 f(y_0) + o(r^2) \quad as\ r \to 0 ,$$

where C_n is a certain constant depending only on n, in fact it is easily verified that

$$C_n = \frac{n-1}{2n}\,\frac{\Gamma(n/2)}{\Gamma((n+1)/2)} .$$

Proof. Denote the *constant* matrices $A(y_0)$, $B(y_0)$ by A and B respectively. Since u is twice differentiable at y_0 we have

$$(3)\qquad u(y_0 + B(y_0)x) = u(y_0 + Bx) = u(y_0) + (Bx\cdot\nabla_y)u(y)\,|_{y_0} + \frac{1}{2}(Bx\cdot\nabla_y)^2u(y)\,|_{y_0} + o(|Bx|^2) .$$

But $|Bx| \leqq \|B\|\ |x|$. Thus on $\{|x| = r\}$, (3) becomes

$$(4)\qquad u(y_0 + B(y_0)x) = u(y_0) + (Bx\cdot\nabla_y)u(y)\,|_{y_0} + \frac{1}{2}(Bx\cdot\nabla)^2u(y)\,|_{y_0} + o(r^2) .$$

Dividing (4) by Ω_r and integrating over $\{|x| = r\}$ we get

$$(5)\qquad u_r(y_0) = u(y_0) + \frac{1}{2\Omega_r}\int_{|x|=r}(Bx\cdot\nabla_y)^2u(y)\,|_{y_0}\,d\Omega_r + o(r^2) .$$

We next observe

$$\frac{1}{2\Omega_r}\int_{|x|=r} x_i x_j d\Omega_r = C_n r^2 \delta_{ij}$$

where C_n is a constant depending only on n. Thus (5) becomes, by the lemma,

$$u_r(y_0) = u(y_0) + C_n r^2 \sum_{i\,j} a_{ij}(y_0) u_{,ij}(y_0) + o(r^2)\ . \tag{6}$$

But (6) is compatible with (2) if and only if $Lu = f$ at y_0.

3. The heat equation. As an application of the main result let us consider the heat operator $Hu = u_{xx} - u_t$. If we make the change of variables given by $x = \xi$, $t = \tau - (1/2)\xi^2$ and set $u(x, t) = v(\xi, \tau)$ then we see that our operator takes the form $v_{\xi\xi} + 2\xi v_{\xi\tau} + \xi^2 v_{\tau\tau}$. In this case the matrix A is given by

$$A = \begin{pmatrix} 1 & \xi \\ \xi & \xi^2 \end{pmatrix}.$$

To compute B we observe that $A^2 = (1 + \xi^2)A$, so that $B = A/\sqrt{1 + \xi^2}$. Then

$$B\begin{pmatrix} r\cos\theta \\ r\sin\theta \end{pmatrix} = \frac{1}{\sqrt{1+\xi^2}}\begin{pmatrix} r\cos\theta + \xi r\sin\theta \\ \xi r\cos\theta + \xi^2 r\sin\theta \end{pmatrix}. \tag{7}$$

For each ξ, there is an α satisfying $-(\pi/2) \leqq \alpha \leqq (\pi/2)$ for which

$$\frac{\cos\theta + \xi\sin\theta}{\sqrt{1+\xi^2}} = \cos(\theta - \alpha)\ ,$$

so that (7) takes the form

$$B\begin{pmatrix} r\cos\theta \\ r\sin\theta \end{pmatrix} = \begin{pmatrix} r\cos(\theta-\alpha) \\ r\xi\cos(\theta-\alpha) \end{pmatrix}. \tag{8}$$

Then $v_r(\xi_0, \tau_0)$ becomes

$$v_r(\xi_0, \tau_0) = \frac{1}{2\pi}\int_0^{2\pi} v(\xi_0 + r\cos(\theta-\alpha), \tau_0 + r\xi_0\cos(\theta-\alpha))d\theta\ .$$

Replacing $\theta - \alpha$ by θ and using the symmetry of the cosine function this reduces to

$$v_r(\xi_0, \tau_0) = \frac{1}{\pi}\int_0^{\pi} v(\xi_0 + r\cos\theta, \tau_0 + r\xi_0\cos\theta)d\theta\ .$$

By changing back to (x, t) coordinates and defining $x_0 = \xi_0$, $t_0 =$

$\tau_0 - (1/2)\xi_0^2$ and $u_r(x_0, t_0) \equiv v_r(\xi_0, \tau_0)$ we get

$$u_r(x_0, t_0) = \frac{1}{\pi}\int_0^\pi u\left(x_0 + r\cos\theta, \tau_0 + rx_0\cos\theta - \frac{1}{2}(x_0 + r\cos\theta)^2\right)d\theta ,$$
$$= \frac{1}{\pi}\int_0^\pi u\left(x_0 + r\cos\theta, t_0 - \frac{1}{2}r^2\cos^2\theta\right)d\theta ,$$
$$= \frac{1}{\pi}\int_{-r}^r u\left(x_0 + z, t_0 - \frac{1}{2}z^2\right)\frac{dz}{\sqrt{r^2 - z^2}}$$

or finally

$$u_r(x_0, t_0) = \frac{1}{\pi}\int_{-1}^1 u\left(x_0 + rz, t_0 - \frac{1}{2}r^2z^2\right)\frac{dz}{\sqrt{1 - z^2}} \tag{9}$$

which is easily seen to be a weighted average of u over the tip of a parabola with vertex at (x_0, t_0), having the line $t = t_0$ as its axis and opening down.

This gives us the following theorem.

THEOREM. *If u is twice differentiable at a point (x_0, t_0), then a necessary and sufficient condition that $Hu = f$ at (x_0, t_0) is that*

$$u_r(x_0, t_0) = u(x_0, t_0) + C_2r^2f + o(r^2) \quad as\ r \to 0 ,$$

where $u_r(x_0, t_0)$ is given by (9).

To study the heat equation in higher dimensions one can make similar transformations. But it is easier to guess the form the previous theorem would take and verify it directly by the methods which established our basic theorem. The result is given below where Δu is the n-dimensional Laplacian, and Ω is the area of the unit sphere in $n + 1$ dimensions.

THEOREM. *If u is twice differentiable at a point (x_0, t_0) in $n + 1$ dimensions, then a necessary and sufficient condition that $\Delta u - u_t = f$ at (x_0, t_0) is that*

$$u_r(x_0, t_0) = u(x_0, t_0) + C_{n+1}r^2f + o(r^2) \quad as\ r \to 0$$

where

$$u_r(x_0, t_0) = \frac{2}{\Omega}\int_{|z|<1} u\left(x_0 + zr, t_0 - \frac{1}{2n}z^2r^2\right)\frac{dz}{\sqrt{1 - |z|^2}}$$

with $dz = dz_1dz_2 \cdots dz_n$.

UNIVERSITY OF COLORADO

STRONGLY RECURRENT TRANSFORMATIONS

ARSHAG HAJIAN

Let $(X, \mathscr{B}, m)$ be a finite or σ-finite and non-atomic measure space. A set B is said to be measurable if it is a member of $\mathscr{B}$. Two measures on $\mathscr{B}$, finite or σ-finite (one may be finite and the other σ-finite), are said to be equivalent if they have the same null sets. In this paper we consider a one-to-one, nonsingular, measurable transformation ϕ of X onto itself. By a nonsingular transformation ϕ we mean $m(\phi B) = m(\phi^{-1}B) = 0$ for every measurable set B with $m(B) = 0$, and by a measurable transformation ϕ we mean $\phi B \in \mathscr{B}$ and $\phi^{-1}B \in \mathscr{B}$ for every $B \in \mathscr{B}$. We shall say that the transformation ϕ is measure preserving (with respect to a measure μ) or equivalently, μ is an invariant measure (with respect to the transformation ϕ) if $\mu(\phi B) = \mu(\phi^{-1}B) = \mu(B)$ for every measurable set B.

A recurrent transformation is a common notion in ergodic theory. This is a measurable transformation ϕ defined on a finite or σ-finite measure space $(X, \mathscr{B}, m)$ with the following property: if A is a measurable set of positive measure, then for almost all $x \in A$ $\phi^n x$ belongs to A for infinitely many integers n. It is not difficult to see that every measurable transformation which preserves a finite invariant measure μ equivalent to m is recurrent. The converse statement is not in general true; for example an ergodic transformation which preserves an infinite and σ-finite measure is always recurrent yet it does not preserve a finite invariant equivalent measure. In this paper we restrict the notion of a recurrent transformation. We introduce the notion of a strongly recurrent set and define a strongly recurrent transformation. We show that a transformation ϕ is strongly recurrent if and only if there exists a finite invariant measure μ equivalent to m (Theorem 2). This is accomplished by showing the connection between strongly recurrent sets and weakly wandering sets (Theorem 1). Weakly wandering sets were introduced in [1], and the condition that a transformation ϕ does not have any weakly wandering set of positive measure was further strengthened (see condition $(W)^*$ below). It was shown in [1] that this stronger condition was again a necessary and sufficient condition for the existence of a finite invariant measure μ equivalent to m. We show that a similar strengthening for a strongly recurrent transformation is false for a wide class of measure preserving transformations defined on a finite measure space (Theorem 3).

DEFINITION. A measurable set S is said to be *strongly recurrent*

Received July 22, 1963.

(with respect to ϕ) if the set of all integers n such that $m(\phi^n S \cap S) > 0$ is relatively dense, i.e., if there exists a positive integer k such that

$$\max_{0 \leq i \leq k-1} m(\phi^{n-i} S \cap S) > 0 \tag{1}$$

for $n = 0, \pm 1, \pm 2, \cdots$. This condition is obviously equivalent to the following:

$$m\Big(\bigcup_{i=0}^{k-1} \phi^{n-i} S \cap S\Big) > 0 \tag{2}$$

or

$$m\Big(\phi^n S \cap \Big[\bigcup_{i=0}^{k-1} \phi^i S\Big]\Big) > 0 \tag{3}$$

for $n = 0, \pm 1, \pm 2, \cdots$, This last condition means that there exists a finite number of images of S by the powers of ϕ such that any image of S by any power of ϕ has an intersection of positive measure with at least one of them.

The transformation ϕ is said to be strongly recurrent if every set of positive measure is strongly recurrent. We note that the property of a transformation ϕ being strongly recurrent is preserved under equivalent measures.

The following notion was introduced in [1]: A measurable set W is said to be weakly wandering (with respect to ϕ) if there exists a sequence of integers $\{n_k : k = 1, 2, \cdots\}$ such that the sets $\phi^{n_k} W$, $k = 1, 2, \cdots$ are mutually disjoint.

THEOREM 1. *Let* $(X, \mathscr{B}, m)$ *be a finite or* σ*-finite measure space, and let* ϕ *be a one-to-one, nonsingular, measurable transformation of* X *onto itself. Then the following two conditions are equivalent:*

(W) $m(A) > 0$ *implies that there exists at most a finite number of mutually disjoint images of* A *by the powers of* ϕ; *in other words,* A *is not weakly wandering.*

(S) $m(A) > 0$ *implies that* A *is strongly recurrent.*

We first prove a Lemma which is by itself of some interest.

LEMMA 1. *Let* $(X, \mathscr{B}, m)$ *and* ϕ *be as in Theorem* 1, *and let* A *be a measurable set of positive measure such that*

$$\liminf_{n \to \infty} m(\phi^n A) = 0 . \tag{4}$$

Then given ε *with* $0 < \varepsilon < m(A)$, *there exists a measurable subset* A' *of* A *with* $m(A') < \varepsilon$ *such that the set* $S = A - A'$ *is not strongly recurrent.*

Proof. Let A be a measurable set with $m(A) = \alpha > 0$ and $\liminf_{n\to\infty} m(\phi^n A) = 0$. Let ε be a positive number with $0 < \varepsilon < \alpha$. Let

$$\varepsilon_k = \frac{\varepsilon}{k2^k}$$

for $k = 1, 2, \cdots$. Next, for each $k = 1, 2, \cdots$ we choose a positive integer n_k such that

$$m(\phi^{n_k - i}A) < \varepsilon_k$$

for $i = 0, 1, 2, \cdots, k - 1$. This is possible since ϕ is nonsingular and (4) is satisfied by A. Let us put

$$A' = \bigcup_{k=1}^{\infty} \bigcup_{i=0}^{k-1} \phi^{n_k - i}A \cap A .$$

Then

$$m(A') \leqq \sum_{k=1}^{\infty} \sum_{i=0}^{k-1} m(\phi^{n_k - i}A) < \sum_{k=1}^{\infty} k\varepsilon_k = \varepsilon .$$

Let $S = A - A'$, then it is easy to see that

$$\phi^{n_k - i}S \cap S \subset \phi^{n_k - i}A \cap (A - A') = \phi$$

for $i = 0, 1, 2, \cdots, k - 1$ and $k = 1, 2, \cdots$. This shows that S is not strongly recurrent.

Proof of Theorem 1. If a measurable set S of positive measure is not strongly recurrent, then it is possible to find a measurable subset N of S with $m(N) = 0$ such that $S' = S - N$ is weakly wandering. This is easy, since S not strongly recurrent means that for each positive integer n_k there exists another positive integer n_{k+1} such that

$$m\left(\phi^{n_{k+1}}S \cap \bigcup_{i=0}^{n_k} \phi^i S\right) = 0 .$$

In this way we may obtain a sequence of integers $\{n_k : k = 1, 2, \cdots\}$ such that

$$m(\phi^{n_k}S \cap \phi^{n_j}S) = m(S \cap \phi^{n_k - n_j}S) = 0 \text{ for } k \neq j .$$

It follows that $S' = S - N$ is weakly wandering, where

$$N = \bigcup_{k=1}^{\infty} \bigcup_{j=1}^{k-1} \phi^{n_k - n_j}S \cap S$$

and $m(N) = 0$.

Conversely, let W be a weakly wandering set of positive measure.

Since the measure space is σ-finite we can find a sequence of measurable sets $\{A_i : i = 1, 2, \cdots\}$ which are mutually disjoint, such that $0 < m(A_i) < \infty$ for $i = 1, 2, \cdots$ and $X = \bigcup_{i=1}^{\infty} A_i$. We let

$$m'(B) = \sum_{i=1}^{\infty} \frac{m(A_i \cap B)}{2^i m(A_i)} \text{ for } B \in \mathscr{B} .$$

It follows that m' and m are equivalent. Since $\phi^{n_k} W, k = 1, 2, \cdots$ are mutually disjoint and $m'(X) < \infty$ it follows that $\liminf_{n\to\infty} m'(\phi^n W) = 0$. Thus, whether m is finite or σ-finite, the set W satisfies (4) with m replaced by the equivalent and finite measure m'. By applying Lemma 1 we obtain a measurable subset S of W such that $m'(S) > 0$ and S is not strongly recurrent. Since m and m' are equivalent, this proves the theorem.

THEOREM 2. *Let* $(X, \mathscr{B}, m)$ *and* ϕ *be as in Theorem* 1. *Then condition* (S) *is equivalent to the existence of a finite invariant measure* μ *equivalent to* m.

Proof. Theorem 2 is an immediate consequence of Theorem 1 above and Theorem 1 of [1], where it was shown that condition (W) is equivalent to the existence of a finite invariant measure μ equivalent to m.

In [1] it was further shown that the following condition:

(W)* Given $\varepsilon > 0$, there exists a positive integer N such that $m(A) \geqq \varepsilon$ implies that there exists at most N mutually disjoint images of A by the powers of T,

is again a necessary and sufficient condition for the existence of a finite invariant measure μ equivalent to m (see condition $(V)^*$, § 3 of [1]).

Condition $(W)^*$ is in appearance a stronger condition than condition (W). We note that in condition (W)* the positive integer N depends on ε only and not on the measurable set A. However, it turns out that these two conditions are equivalent to each other and are in turn necessary and sufficient conditions for the existence of a finite invariant measure μ equivalent to m (see Theorem 1 of [1]). By analogy, we may attempt to strengthen condition (S) in the following manner:

(S)* Given $\varepsilon > 0$, there exists a positive integer N such that $m(A) \geqq \varepsilon$ implies

$$m\left(\phi^n A \cap \bigcup_{i=0}^{N-1} \phi^i A\right) > 0 \quad \text{for } n = 0, \pm 1, \pm 2, \cdots$$

We show that condition (S)* is not a necessary condition for the existence of a finite invariant measure μ equivalent to m. In fact, we shall show that for any ergodic measure preserving transformation ϕ defined on a finite measure space $(X, \mathscr{B}, \mu)$ condition (S)* is not satisfied.

We say that a transformation ϕ is ergodic if $\phi A = A$ implies $m(A) = 0$ or $m(X - A) = 0$.

LEMMA 2. *Let $(X, \mathscr{B}, \mu,)$ be a finite or σ-finite measure space, and let ϕ be an ergodic measure preserving transformation defined on it. Then given $\varepsilon > 0$ and a positive integer $N > 0$, there exists a measurable set C with $\mu(C) \leqq \varepsilon$ such that*

$$X - C = \bigcup_{i=0}^{N-1} \phi^i E$$

for some measurable set E where $E, \phi E, \cdots, \phi^{N-1}E$ are mutually disjoint.

Proof. Given $\varepsilon > 0$ and an integer $N > 0$, let F be any measurable set with $0 < \mu(F) \leqq \varepsilon/N$. Let

$$F_0 = F$$
$$F_1 = \phi^{-1}F - F_0$$
$$F_2 = \phi^{-2}F - F_0 \cup F_1$$

and in general

$$F_n = \phi^{-n}F - \bigcup_{j=0}^{n-1} F_j \quad \text{for } n = 1, 2, \cdots .$$

It follows that $F_n, n = 0, 1, 2, \cdots$ are mutually disjoint, and furthermore;

$$\phi^k F_n \subset F_{n-k} \qquad \begin{array}{l} \text{for } k = 0, \cdots, n \\ \text{and } n = 0, 1, 2, \cdots . \end{array}$$

We let

$$E_i = F_{iN} = \phi^{-iN}F - \bigcup_{j=0}^{iN-1} F_j = \phi^{-iN}F - \bigcup_{j=0}^{iN-1} \phi^{-j}F$$

then

$$\phi^k E_i \subset F_{iN-k} \quad \text{for } k = 0, 1, \cdots, iN; \text{ and } i = 1, 2, \cdots$$

which implies that the sets

$$(5) \qquad \phi^k E_i \quad \text{for } k = 0, 1, \cdots, iN; \text{ and } j = 1, 2, \cdots$$

are mutually disjoint.

Next we let

$$E = \bigcup_{i=1}^{\infty} E_i$$

and

$$C = X - \bigcup_{k=0}^{N-1} \phi^k E\ .$$

It follows from (5) that $E, \phi E, \cdots, \phi^{N-1}E$ are mutually disjoint, and

$$\mu(C) = \mu\Big(X - \bigcup_{k=0}^{N-1} \phi^k E\Big) \leqq N\mu(F) \leqq \varepsilon\ .$$

THEOREM 3. *Let ϕ be on ergodic measure preserving transformation defined on a finite measure space $(X, \mathscr{B}, \mu)$ with $\mu(X) = 1$. Then condition $(S)^*$ is not satisfied.*

Proof. Let $\varepsilon = 1/(q+1)$ for some positive integer $q > 3$. Let $k > 1$ be an arbitrary positive integer. We show that there exists a measurable set A with $\mu(A) \geqq \varepsilon$ and

$$\mu\Big(\phi^{n_k} A \cap \bigcup_{i=0}^{k-1} \phi^i A\Big) = 0$$

for some integer $n_k > k$. Let us put $N = qk$. Then by Lemma 2 there exists a measurable set E with $E, \phi E, \cdots, \phi^{N-1}E$ mutually disjoint and

$$\mu\Big(X - \bigcup_{k=0}^{N-1} \phi^k E\Big) \leqq \varepsilon = \frac{1}{q+1}\ .$$

Since $\mu(X) = 1$, this implies $1 - N\mu(E) \leqq \varepsilon$ or $\mu(E) \geqq (1-\varepsilon)/N$. Let

$$A = \bigcup_{i=0}^{k-1} \phi^i E\ .$$

Since $k = N/q$ we have

$$\mu(A) = k\mu(E) \geqq \frac{N}{q}\frac{(1-\varepsilon)}{N} = \frac{1 - \dfrac{1}{q+1}}{q} = \frac{1}{q+1} = \varepsilon$$

and

$$\mu\Big(\phi^{n_k} A \cap \bigcup_{i=0}^{k-1} \phi^i A\Big) = \mu\Big(\bigcup_{i=n_k}^{n_k+k-1} \phi^i E \cap \bigcup_{i=0}^{2k-2} \phi^i E\Big) = 0$$

for some n_k where $2k < n_k < (q-1)k = N - k$.

This shows that condition (S)* is not satisfied since ε is fixed, k is arbitrary, and $n_k > k$.

REFERENCE

1. A. Hajian and S. Kakutani, *Weakly wandering sets and invariant measures,* Transactions A. M. S. **110** (1964), 136–151.

CORNELL UNIVERSITY

DOUBLY INVARIANT SUBSPACES, II

MORISUKE HASUMI AND T. P. SRINIVASAN

1. Introduction. Let X be a locally compact Hausdorff space and μ a positive Radon measure on X. Let $\mathscr{H}$ be a separable Hibert space and let $L^p_{\mathscr{H}}$ $(1 \leqq p \leqq +\infty)$ denote the space of $\mathscr{H}$-valued functions on X which are weakly measurable and whose norms are in scalar $L^p(d\mu)$. Call P a *measurable range function* if P is a function on X defined a.e. $(d\mu)$ to the space of orthogonal projections on $\mathscr{H}$ which is weakly measurable. We shall regard two range functions P, P' to be the same if $P(x) = P'(x)$ l.a.e., i.e. $P(x) = P'(x)$ a.e. on every compact subset of X. We shall denote by $\hat{P}$ the operator on $L^p_{\mathscr{H}}$ defined by $(\hat{P}f)(x) = P(x)f(x)$ l.a.e. Let A be a subalgebra of the algebra $C(X)$ of bounded continuous functions on X such that $A \cup \bar{A}$ (where the bar denotes complex conjugation) is weakly* dense in $L^\infty(d\mu)$. Say that a subspace $\mathscr{M}$ of $L^p_{\mathscr{H}}$ is *doubly invariant* if

(i) $\mathscr{M}$ is closed in $L^p_{\mathscr{H}}$ if $1 \leqq p < \infty$ and weakly* closed if $p = \infty$,

(ii) $\mathscr{M}$ is invariant under multiplication by functions in $A \cup \bar{A}$.

We shall refer to the following theorem as Wiener's theorem for $L^p_{\mathscr{H}}$:

THEOREM. *Every doubly invariant subspace $\mathscr{M}$ of $L^p_{\mathscr{H}}$ $(1 \leqq p \leqq \infty)$ is of the form $\hat{P}L^p_{\mathscr{H}}$ for some measurable range function P (and trivially conversely); $\mathscr{M}$ determines P uniquely.*

For compact spaces X, Wiener's theorem was proved in [4] for arbitrary $\mathscr{H}$ for $p = 2$ and for the scalar $\mathscr{H}$ (the space of complex numbers) for arbitrary p. It was pointed out in [4] that the $L^2_{\mathscr{H}}$ theorem is true for locally compact spaces and the proof was outlined considering the real line as an example. It was also mentioned in [4] that the $L^2_{\mathscr{H}}$ theorem is a special case of a known theorem on rings of operators [2; p. 167, Théorème 1]. But the proof in [4] and the proof of the more general theorem in [2] implicitly assume the σ-finiteness of μ or at least of the separability of $L^2_{\mathscr{H}}$ (as opposed to the separability of $\mathscr{H}$). The theorem itself is true without this restriction not only for $p = 2$ but for all p and all (separable) $\mathscr{H}$ (not necessarily the scalar $\mathscr{H}$). Indeed the general $L^p_{\mathscr{H}}$ theorem is true even under the weaker assumption that the restriction of $A \cup \bar{A}$ to every compact subset K of X is L^2-dense in $L^2(d\mu \mid K)$, instead of being weakly* dense in L^∞. In this paper we prove this theorem

Received July 18, 1963. This work was done while both authors held visiting appointments in the University of California, Berkeley, and were respectively sponsored in part by the National Science Foundation, Grants NSF GP-2 and NSF G-18974.

(Theorem 4) in its full generality (with the above weaker assumption). This is done as follows: Using the techniques employed in [5] we first show in § 2 (Theorem 2) that a general class of subalgebras dense in L^2 is weakly* dense, which seems to be of independent interest. This enables us to reduce the L^2-density case to that of weak* density. To overcome the difficulties caused by the (possible) non-separability of $L^2_{\mathscr{H}}$ we extend in § 3 (Theorem 3) a theorem of Dunford-Pettis [1; p. 46, Corollaire 2] to apply to our setup. We finally use the $L^2_{\mathscr{H}}$ theorem for compact X in [4] and the broad techniques in [4] to complete the proof. As pointed out in [4], the $L^p_{\mathscr{H}}$ theorem for $p \neq 2$ is of special interest as it shows that the doubly invariant subspaces of $L^p_{\mathscr{H}}$ admit projections of norm 1 commuting with bounded (scalar) functions; as is well known, a closed linear subspace of a Banach space does not in general have any bounded projection at all. In the final section of the paper we extend a known theorem [2] on operators in $L^p_{\mathscr{H}}$ which commute with multiplication by bounded (scalar) functions (Theorem 5).

2. Weak* density of certain subalgebras of L^∞.

THEOREM 1. *Let (X, m) be a finite measure space. Any subalgebra $\mathscr{A}$ of $L^\infty(dm)$ which is conjugate-closed and dense in $L^2(dm)$ is weakly* dense in $L^\infty(dm)$.*[1]

The following three lemmas will lead to the proof of the theorem.

LEMMA 1. *Let $\mathscr{B}$ be a conjugate-closed subalgebra of $L^\infty(dm)$ which contains constants and is closed in $L^\infty(dm)$. Then $\mathscr{B}$ is closed for absolute values.*

Proof. Let $f \in \mathscr{B}$, $0 \leqq f \leqq 1/2$, say. Then $f^{\frac{1}{2}} = (1-(1-f))^{\frac{1}{2}}$ can be expressed as the sum of a convergent series in $L^\infty(dm)$ whose terms come from $\mathscr{B}$; it follows that $f^{\frac{1}{2}} \in \mathscr{B}$ for all non-negative $f \in \mathscr{B}$. Since $\mathscr{B}$ is conjugate-closed, the lemm follows.

LEMMA 2. *Let (X, m) be a finite measure space and A a subalgebra of $L^\infty(dm)$ such that $A \cup \bar{A}$ is dense in $L^2(dm)$. Then every closed subspace $\mathscr{M}$ of $L^2(dm)$ which is invariant under multiplication by functions in $A \cup \bar{A}$ is of the form $C_S L^2(dm)$ for some measurable subset S of X (where C_S denotes the characteristic function of S).*

Proof. Let $\mathscr{B}$ be the closed subalgebra of $L^\infty(dm)$ generated by $A \cup \bar{A}$ and the constants. Then $\mathscr{M}$ is clearly invariant under multi-

[1] A weaker result was proved in [5].

plication by functions in $\mathscr{B}$. By Lemma 1, $\mathscr{B}$ is closed for absolute values. Let q be the orthogonal projection of the constant function 1 on $\mathscr{M}$. Then $1 - q \perp \mathscr{M}$. Since $\mathscr{M}$ is invariant under multiplication by function in $\mathscr{B}$, it follows that

$$\int fq dm = \int f \,|\, q \,|^2 dm \tag{2.1}$$

for all $f \in \mathscr{B}$. Let Y be any measurable subset of X and let $\{f_n\}$ be a sequence of functions from $\mathscr{B}$ which converges to C_Y in $L^2(dm)$. Since $|f_m - f_n| \in \mathscr{B}$, we have from (2.1)

$$\int |f_m - f_n| \,|\, q \,|^2 dm = \int |f_m - f_n| \, q dm$$

and the last integral is less than $\left(\int |f_m - f_n|^2 dm\right)^{\frac{1}{2}} \times \left(\int |\, q \,|^2 dm\right)^{\frac{1}{2}}$. It follows that $\{f_n \,|\, q\,|^2\}$ is a Cauchy sequence in $L^1(dm)$. Hence $f_n \,|\, q \,|^2 \to C_Y \,|\, q \,|^2$ in $L^1(dm)$; in particular,

$$\int f_n \,|\, q \,|^2 dm \to \int_Y |\, q \,|^2 dm \,. \tag{2.2}$$

Since $f_n \to C_Y$ in $L^2(dm)$, $f_n q \to C_Y q$ in $L^1(dm)$ and thus

$$\int f_n q dm \to \int_Y q dm \,. \tag{2.3}$$

It follows from (2.1)–(2.3) that $\int_Y |\, q \,|^2 dm = \int_Y q dm$ for all measurable subsets Y; hence $|\, q \,|^2 = q$ a.e. Thus $q = C_S$ a.e. for some $S \subset X$.

Because of invariance, $C_S L^2(dm) \subset \mathscr{M}$. If the inclusion were strict, let $g \in \mathscr{M} \ominus C_S L^2(dm)$. Then $g \perp C_S \mathscr{B}$ also $C_{S'} \in \mathscr{M}^{\perp}$ (where $S' = X - S$) and $\mathscr{M}^{\perp}$ is also invariant along with $\mathscr{M}$ under multiplications by functions in $\mathscr{B}$. So $g \perp C_{S'} \mathscr{B}$. It follows that $g \perp \mathscr{B}$ and because of density of $\mathscr{B}$ in $L^2(dm)$, we have $g = 0$ a.e. Thus $\mathscr{M} = C_S L^2(dm)$.

LEMMA 3. *Let (X, m) and A be as in Lemma 2. Then every closed subspace of $L^1(dm)$ which is invariant under multiplication by functions in $A \cup \bar{A}$ is of the form $C_S L^1(dm)$ for some measurable subset S.*

Proof. This follows from Lemma 2 above and Theorem 7 in [4].

Proof of Theorem 1. Let $\mathscr{M} = \left\{ f \in L^1(dm) \colon \int fg dm = 0 \text{ for all } g \in \mathscr{A} \right\}$. Then $\mathscr{M}$ is $\mathscr{A}$-invariant, meaning invariant under multiplication by functions in $\mathscr{A}$ and Lemma 3 applies for $\mathscr{M}$ (with $\mathscr{A}$

replacing A). Thus $\mathscr{M} = C_S L^1(dm)$ for some S, so $\mathscr{M} \cap L^2(dm) = C_S L^2(dm)$. But $\mathscr{M} \cap L^2(dm) = L^2(dm) \ominus \mathscr{A}$. Since $\mathscr{A}$ is dense in $L^2(dm)$ by assumption, it follows that $C_S = 0$ a.e. Therefore $\mathscr{M} = \{0\}$ and the theorem follows.

REMARK. One of the corollaries of Theorem 1 is the "uniqueness" of the Fourier coefficients of any function in $L^1(G)$, for a compact Abelian group G. The characters are dense in $L^2(G)$ so that the subspace $\mathscr{A}$ of their finite linear combinations is weakly* dense in $L^\infty(dm)$ by Theorem 1 and the uniqueness follows.

We now extend Theorem 1 to infinite measure spaces. For convenience we state the result in terms of Radon measures on locally compact spaces. We have

THEOREM 2. *Let X be a locally compact Hausdorff space and μ a positve Radon measure on X. Let $\mathscr{A}$ be a subalgebra of the algebra of bounded continuous functions on X such that*

(i) *$\mathscr{A}$ is conjugate-closed,*

(ii) *$\mathscr{A} \mid K$ is dense in $L^2(d\mu \mid K)$ for every compact subset K of X. Then $\mathscr{A}$ is weakly* dense in $L^\infty(d\mu)$.*

Proof. Let $\mathscr{M} = \left\{f \in L^1(d\mu) \colon \int fg d\mu = 0 \text{ for all } g \in \mathscr{A}\right\}$. If we show that $\mathscr{M} = \{0\}$, the theorem is proved. Now $\mathscr{M}$ is clearly a closed subspace of $L^1(d\mu)$ and is $\mathscr{A}$-invariant. We need the following lemma which will be proved below.

LEMMA 4. *Every closed $\mathscr{A}$-invariant subspace $\mathscr{M}$ of $L^1(d\mu)$ is of the form $C_S L^1(d\mu)$ for some measurable subset S* (where $\mathscr{A}$ is as in Theorem 2).

Assuming Lemma 4, the main theorem follows at once. For, since $\mathscr{M} = C_S L^1(d\mu)$, $\mathscr{A} \subset \mathscr{M}^\perp = C_{S'} L^\infty(d\mu)$. If $\mu(S) > 0$, then S contains a compact subset K of positive measure. Since $\mathscr{A} \subset C_{S'} L^\infty(d\mu)$, $\mathscr{A} \mid K = \{0\}$, contradicting the density of $\mathscr{A} \mid K$ in $L^2(d\mu \mid K)$. Hence $\mu(S) = 0$, so $\mathscr{M} = \{0\}$, completing the proof of the theorem.

Proof of Lemma 4. Let $\mathscr{M}_K = C_K \mathscr{M}$, $\mathscr{A}_K = C_K \mathscr{A}$ and $\mu_K = C_K \mu$. We shall identify $L^p(d\mu \mid K)$, $L^p(d\mu_K)$ and $C_K L^p(d\mu)$ which are clearly mutually isometrically isomorphic. Each $\mathscr{M}_K$ is closed and $\mathscr{A}_K$-invariant in $L^1(d\mu_K)$, so by Lemma 3, $\mathscr{M}_K = C_{S(K)} L^1(d\mu_K)$ for some $S(K) \subset K$. If $K' \supset K$, compact, then

$$\begin{aligned} C_{S(K)}L^1(d\mu) &= C_{S(K)}L^1(d\mu_K) = \mathscr{M}_K = C_K C_{K'}\mathscr{M} \\ &= C_K C_{S(K')}L^1(d\mu_{K'}) = C_{S(K')\cap K}L^1(d\mu_{K'}) \\ &= C_{S(K')\cap K}L^1(d\mu) , \end{aligned}$$

so that $S(K) = S(K') \cap K$ (modulo null sets).

Let $\mathscr{K}$ denote the set of all continuous functions with compact support and let σ be the linear functional on $\mathscr{K}$ defined by

$$\sigma(\varphi) = \int_{S(K)} \varphi d\mu \tag{2.4}$$

for $\varphi \in \mathscr{K}$ where K is any compact subset containing the support of φ. Then σ is well-defined and is continuous in the L^1-norm, so can be uniquely extended to a bounded linear functional on $L^1(d\mu)$, which we again denote by σ. Let σ be realized by the L^∞-function g so that

$$\sigma(f) = \int f g d\mu \tag{2.5}$$

for all $f \in L^1(d\mu)$. From (2.4) and (2.5) it is easy to see that $g \mid K = C_{S(K)}$ a.e. for every compact subset K; so we may assume $g = C_S$ for some measurable S with $S \cap K = S(K)$ (modulo null sets). Now

$$C_K C_S L^1(d\mu) = C_{S\cap K}L^1(d\mu) = C_{S(K)}L^1(d\mu) = \mathscr{M}_K = C_K\mathscr{M}$$

for all compact K. Since for any $f \in L^1(d\mu)$, $C_K f \rightarrow f$ in $L^1(d\mu)$, it follows from the above that $C_S L^1(d\mu) = \mathscr{M}$.

REMARK. The assumption that $\mathscr{A}$ is an algebra is crucial in both Theorems 1 and 2; the conclusion would be false if $\mathscr{A}$ were merely a linear subspace satisfying the rest of the assumptions. The following example shows that, in the locally compact case for instance, a conjugate-closed linear subspace of $L^\infty(d\mu)$ may be weakly* dense on every compact subset but not on the whole space.

Let X be a locally compact space and μ a non-finite Radon measure on X. Let $f \in L^1(d\mu)$ be real and have a support of infinite μ-measure. Then the support is non-compact. Let $\mathscr{A} = \left\{g \in L^\infty(d\mu)\colon \int g f d\mu = 0\right\}$. Then $\mathscr{A}$ is clearly not weakly* dense in $L^\infty(d\mu)$. But if g is any continuous function with compact support which is "orthogonal" to $\mathscr{A}$, then g must be in the linear span of f in $L^1(d\mu)$. It follows from our assumption on f that g is the zero function. Hence $\mathscr{A}$ is weakly* dense on every compact subset.

3. Dunford-Pettis theorem. Let X denote a locally compact Hausdorff space and μ a positive Radon measure on X. Let E be a separable Banach space and $\mathscr{K}_E$ denote the space of continuous functions from X into E with compact support. For $1 \leqq p < \infty$, let $\mathscr{F}_E^p$ be the space of all functions f from X into E with

$$N_p(f) = \left(\int_X^* \|f(x)\|^p \, d\mu(x)\right)^{1/p} < \infty$$

where $\int^*$ denotes the upper integral. $\mathscr{F}_E^p$ is then a locally convex space with respect to the seminorm N_p. Let $\mathscr{L}_E^p$ denote the closure of $\mathscr{K}_E$ in $\mathscr{F}_E^p$ and let $L_E^p = \mathscr{L}_E^p / \mathscr{N}_E^p$ where $\mathscr{N}_E^p$ is the set of all functions $f \in \mathscr{L}_E^p$ with $N_p(f) = 0$. Then L_E^p is a Banach space with the norm induced by N_p in the obvious way.

Denote by $\mathscr{L}_{E^*}^\infty$ the space of all weakly* measurable functions f on X to the dual E^* of E such that $\|f(x)\| \leqq A < \infty$ l.a.e. ($\|f(x)\| \leqq A$ a.e. on every compact subset). For $f \in \mathscr{L}_{E^*}^\infty$ let

$$N_\infty(f) = \sup_K (\text{ess.}\sup_{x \in K} \|f(x)\|)$$

where K ranges over all compact subsets of X. Then N_∞ is a seminorm which makes $\mathscr{L}_{E^*}^\infty$ a locally convex space. Let $L_{E^*}^\infty$ be the quotient of $\mathscr{L}_{E^*}^\infty$ by the space of all functions in $\mathscr{L}_{E^*}^\infty$ which vanish l.a.e. Then $L_{E^*}^\infty$ is a Banach space.

The following theorem is well-known (cf. for instance [1; p. 46, Corollaire 2]):

THEOREM (Dunford-Pettis). *Let F be a separable Banach space. For $f \in L_{F^*}^\infty$ and $g \in L^1(d\mu)$, let*

$$w_f(g) = \int_X g f d\mu\,.$$

Then $w_f(g) \in F^*$ and the mapping $f \to w_f$ induces an isometric isomorphism from $L_{F^*}^\infty$ onto $\mathscr{L}(L^1, F^*)$, the space of bounded linear maps from $L^1(d\mu)$ to F^*.

We need the following variant of the Dunford-Pettis theorem:

THEOREM 3. *Let E, F be separable Banach spaces. For any bounded linear map u of L_E^1 into F^* there exists a function Φ from X into $\mathscr{L}(E, F^*)$ such that*

(i) *$\langle \Phi(x)s, t\rangle$ is measurable for every $s \in E$, $t \in F$,*

(ii) *$N_\infty(\Phi) < \infty$, and*

(iii) *$u(f) = \int_X \Phi(x) f(x) d\mu(x)$ for every $f \in L_E^1$ with $\|u\| = N_\infty(\Phi)$.*

Conversely, any function Φ satisfying (i) *and* (ii) *defines a bounded linear map u satisfying* (iii).

Proof. Only the direct part needs a proof. First we note that $\mathscr{L}(E, F^*)$ can be regarded as the strong dual of the projective tensor product $E \hat{\otimes} F$. Indeed, the strong dual of $E \hat{\otimes} F$ is canonically identified with the space $B(E, F)$ of bounded bilinear forms on $E \times F$ and $\mathscr{L}(E, F^*)$ is canonically isomorphic with $B(E, F)$. Since E, F are separable, so is $E \hat{\otimes} F$ and therefore $\mathscr{L}(E, F^*)$ can be regarded as the strong dual of a separable Banach space.

Let u be a bounded linear map of L^1_E into F^*. Then u induces a bounded bilinear form $\tilde{u}$ on $L^1 \times E$ into F^* by $\tilde{u}(f, s) = u(f \otimes s)$ for $f \in L^1$, $s \in E$. For any fixed $f \in L^1$, $s \to \tilde{u}(f, s)$ is a bounded linear map of E into F^* which we shall denote by u_f. Then u_1: $f \to u_f$ is a bounded linear map from L^1 into $\mathscr{L}(E, F^*)$ with $\|u_1\| = \|u\|$. By the Dunford-Pettis theorem, there exists a function Φ: $X \to \mathscr{L}(E, F^*)$ such that

(i) $\langle \Phi(x)s, t \rangle$ is measurable for each $s \in E$, $t \in F$

(ii) $N_\infty(\Phi) = \|u_1\|$, and

(iii) $u_1(f) = u_f = \int_X f(x)\Phi(x)d\mu(x)$.

Hence

$$u(f \otimes s) = \tilde{u}(f, s) = u_f(s) = \int_X f\Phi s d\mu$$
$$= \int_X \Phi(f \otimes s)d\mu \,.$$

Because of the continuity of u, the theorem follows.

4. **Doubly invariant subspaces.** In this section we prove Wiener's theorem in the general setup. Let as usual X denote a locally compact Hausdorff space, μ a positive Radon measure on X, $\mathscr{H}$ a separable Hilbert space and $\mathscr{K}_{\mathscr{H}}$ the space of continuous functions from X into $\mathscr{H}$ with compact support. Let A be a subalgebra of the algebra of bounded continuous functions on X and $\mathscr{A}$ denote the algebra generated by $A \cup \bar{A}$ and the constants. A subspace $\mathscr{M}$ of $L^p_{\mathscr{H}}$ is clearly invariant under multiplication by functions is $A \cup \bar{A}$ if and only if it is $\mathscr{A}$-invariant. We recall that $\mathscr{M}$ is *doubly invariant* if

(i) $\mathscr{M}$ is closed in $L^p_{\mathscr{H}}$ if $1 \leqq p < \infty$ and weakly* closed if $p = \infty$,

(ii) $\mathscr{M}$ is $\mathscr{A}$-invariant.

Then we have

THEOREM 4. *If $\mathscr{A} \mid K$ is dense in $L^2(d\mu \mid K)$ for every compact subset K, then every doubly invariant subspace $\mathscr{M}$ of $L^p_{\mathscr{H}}$ $(1 \leqq p \leqq \infty)$ is of the form $\hat{P}L^p_{\mathscr{H}}$ for some measurable range function P; $\mathscr{M}$ determines P uniquely.*

Proof. We divide the proof into three parts; in the first and the second we assume $\mu(X) < \infty$ and the proof is an imitation of that of the scalar case in [4]. In the last part we treat the case of arbitrary measure spaces and an indication of the proof in this case was given in the proof of Theorem 2.

(i) $\mu(X) < \infty$, $1 \leqq p \leqq 2$. By Theorem 2, $\mathscr{A}$ is weakly* dense in $L^\infty(d\mu)$ and in this case the theorem has been proved in [4] for $p = 2$. Let $1 \leqq p < 2$ and $\mathscr{N} = \mathscr{M} \cap L^2_{\mathscr{H}}$. Then $\mathscr{N}$ is a doubly invariant subspace of $L^2_{\mathscr{H}}$ and so $\mathscr{N} = \hat{P}L^2_{\mathscr{H}}$ for some measurable range function P. We wish to show that $\mathscr{M} = \hat{P}L^p_{\mathscr{H}}$.

For any $f \in \mathscr{M}$ let $f_1(x) = \|f(x)\|^{1-(p/2)}$ and $f_2(x) = f_1(x)^{-1}f(x)$ (of course $f_2(x) = 0$ if $f_1(x) = 0$). Then $f_1 \in L^s(d\mu)$ where $(1/s) + (1/2) = (1/p)$ and $f_2 \in L^2_{\mathscr{H}}$. Let $\mathscr{N}_2$ be the doubly invariant subspace of $L^2_{\mathscr{H}}$ generated by f_2. Then $\mathscr{N}_2 = \hat{P}_2L^2_{\mathscr{H}}$ for a measurable range function P_2. Here we may assume that $P_2(x) = 0$ for those x for which $f_1(x) = 0$. For any $\varphi \in \mathscr{K}_{\mathscr{H}}$

$$f_1\hat{P}_2\varphi \in f_1\hat{P}_2L^2_{\mathscr{H}} = f_1\mathscr{N}_2 \subset \mathscr{M} .$$

On the other hand, since $s > 2$,

$$f_1\hat{P}_2\varphi \in L^s_{\mathscr{H}} \subset L^2_{\mathscr{H}}$$

as $f_1 \in L^s$, $\hat{P}_2\varphi$ is bounded and $\mu(X) < \infty$. Hence

$$f_1\hat{P}_2\varphi \in \mathscr{M} \cap L^2_{\mathscr{H}} = \mathscr{N} = \hat{P}L^2_{\mathscr{H}} .$$

This means that $\hat{P}\hat{P}_2f_1\varphi = \hat{P}_2f_1\varphi$ for all $\varphi \in \mathscr{K}_{\mathscr{H}}$. So, $P_2(x) \leqq P(x)$ l.a.e. Thus we have $\mathscr{N}_2 = \hat{P}_2L^2_{\mathscr{H}} \subset \hat{P}L^2_{\mathscr{H}}$. Hence

$$f = f_1f_2 \in f_1\mathscr{N}_2 \subset f_1\hat{P}L^2_{\mathscr{H}} \subset \hat{P}L^p_{\mathscr{H}} ;$$

the last inclusion resulting from the fact that $f_1 \in L^s$ where $(1/s) + (1/2) = (1/p)$. This shows that $\mathscr{M} \subset PL^p_{\mathscr{H}}$.

Since $\mathscr{M} \supset \mathscr{N} = \hat{P}L^2_{\mathscr{H}}$, we have $\mathscr{M} \supset \hat{P}\mathscr{K}_{\mathscr{H}}$. But $\mathscr{K}_{\mathscr{H}}$ is dense in $L^p_{\mathscr{H}}$ and $\hat{P}$ is L^p-continuous. So $\mathscr{M} \supset \hat{P}L^p_{\mathscr{H}}$ and we have $\mathscr{M} = PL^p_{\mathscr{H}}$.

(ii) $\mu(X) < \infty$, $2 < p \leqq \infty$. Let $\mathscr{M}' = \{f \in L^q_{\mathscr{H}} : f \perp \mathscr{M}\}$ where $(1/q) + (1/p) = 1$. Then $1 \leqq q < 2$ and $\mathscr{M}'$ is doubly invariant in $L^q_{\mathscr{H}}$. Hence by (i) $\mathscr{M}' = \hat{P}'L^q_{\mathscr{H}}$ for some measurable range function P'. Then it is easy to see that $\mathscr{M} = \hat{P}L^p_{\mathscr{H}}$ where $P(x) = I - P'(x)$, I denoting the identity operator on $\mathscr{H}$.

(iii) $\mu(X)$ not necessarily finite, $1 \leqq p \leqq \infty$. Consider any compact subset K of X. Let $\mathscr{M}_K = C_K\mathscr{M}$, $\mathscr{A}_K = C_K\mathscr{A}$ and $\mu_K = C_K\mu$. We shall identify $L^p_{\mathscr{H}}(d\mu \mid K)$, $L^p_{\mathscr{H}}(d\mu_K)$ and $C_KL^p_{\mathscr{H}}(d\mu)$ which are obviously mutually isometrically isomorphic and denote any of them by $L^p_{\mathscr{H}}(K)$. Now $\mathscr{M}_K$ is a doubly invariant subspace of $L^p_{\mathscr{H}}(d\mu_K)$ (with $\mathscr{A}_K$ replacing $\mathscr{A}$) and $\mathscr{A}_K$ is dense in $L^2(d\mu_K)$. Hence by (i) and (ii) above, $\mathscr{M}_K = \hat{P}_K L^p_{\mathscr{H}}(K)$. We extend P_K to the whole of X by defining $P_K(x) = 0$ outside of K.

For any two compact subsets K_1, K_2 with $K_1 \supset K_2$ we have

$$\begin{aligned}\hat{P}_{K_2} L^p_{\mathscr{H}} &= \hat{P}_{K_2} L^p_{\mathscr{H}}(K_2) = \mathscr{M}_{K_2} = C_{K_2} C_{K_1} \mathscr{M} = C_{K_2} \hat{P}_{K_1} L^p_{\mathscr{H}}(K_1) \\ &= \hat{P}_{K_1} C_{K_2} L^p_{\mathscr{H}}(K_1) = \hat{P}_{K_1} C_{K_2} L^p_{\mathscr{H}} .\end{aligned}$$

Hence $P_{K_2} = P_{K_1} C_{K_2}$ a.e. It follows from this that the map σ: $\mathscr{K}_{\mathscr{H}} \rightarrow \mathscr{H}$ given by

$$\sigma(\varphi) = \int_X P_K(x)\varphi(x)d\mu(x) ,$$

where K is any compact subset containing the support of φ, is well-defined. σ is clearly continuous with respect to the $L^1_{\mathscr{H}}$-norm and so can be uniquely extended to the whole of $L^1_{\mathscr{H}}$ to be continuous. We shall denote the extended map by $\tilde{\sigma}$. By Theorem 3 there exists a weakly measurable bounded operator-valued function Φ: $X \rightarrow \mathscr{L}(\mathscr{H}, \mathscr{H})$ such that

$$\tilde{\sigma}(f) = \int_X \Phi(x) f(x) d\mu(x)$$

for all $f \in L^1$. Then, since $\tilde{\sigma}$ entends σ, it is obvious that

$$\Phi \mid K = P_K \text{ a.e.}$$

for every compact set K; so there exists a measurable range function P such that $\Phi = P$ l.a.e.

We assert that $\mathscr{M} = \hat{P} L^p_{\mathscr{H}}$. This follows from the fact that $C_K \mathscr{M} = C_K \hat{P} L^p_{\mathscr{H}}$ for every compact set K and every $f \in \mathscr{M}$ is the L^p-limit (or the weak* limit if $p = \infty$) of $C_K f$. This completes the proof.

The uniqueness of P (for a given $\mathscr{M}$) follows from the uniqueness established in [4] for finite measure spaces.

5. **Decomposable operators.** Let X, μ, A and $\mathscr{A}$ be as in § 4 and let T be an operator in $L^p_{\mathscr{H}}$ bounded if $1 \leqq p < \infty$ and in addition weakly* continuous if $p = \infty$. Clearly T commutes with multiplication by functions in $A \cup \bar{A}$ if and only if it commutes with functions in $\mathscr{A}$, and any operator T which operates pointwise (l.a.e.), meaning

$$(Tf)(x) = T(x)f(x) \text{ l.a.e.}$$

for an operator-valued function $T(x)$, clearly has this property. We wish to prove the following converse.

THEOREM 5. *If T is a bounded (and weakly* continuous, if*

$p = \infty$) linear map from $L^p_{\mathscr{H}}$ into $L^p_{\mathscr{H}}$ ($1 \leqq p \leqq \infty$) which commutes with multiplication by functions in $\mathscr{A}$, then there exists an operator-valued function $T(x)$ defined a.e. *with $T(x) \in \mathscr{L}(\mathscr{H}, \mathscr{H})$ which is weakly measurable and uniformly bounded such that*

$$(Tf)(x) = T(x)f(x) \text{ a.e. } ((Tf)(x) = T(x)f(x) \text{ l.a.e. } if\ p = \infty)$$

This theorem is usually stated for $L^2_{\mathscr{H}}$ [2; p. 162, Theoreme 1] and as far as we are aware, the existing proofs require $L^2_{\mathscr{H}}$ to be separable. We use the variant of Dunford-Pettis theorem established by us in § 3 to get around the difficulties that may be caused by non-separability (we of course assume that the Hilbert space $\mathscr{H}$ is separable).

Proof of Theorem 5. We first consider the case $1 \leqq p < \infty$, for convenience we assume that T is bounded by 1. Let $f \in L^p_{\mathscr{H}}$. Then

$$\int_X \| (Tf)(x) \|^p d\mu(x) \leqq \int_X \| f(x) \|^p d\mu(x) .$$

Since T commutes with multiplication by functions in $\mathscr{A}$, this yields

$$\int_X | \alpha(x) |^p \| (Tf)(x) \|^p d\mu(x) \leqq \int_X | \alpha(x) |^p \| f(x) \|^p d\mu(x)$$

for all $\alpha \in \mathscr{H}$. From the weak* density of $\mathscr{A}$ in L^∞, it follows that

$$\| (Tf)(x) \| \leqq \| f(x) \| \text{ a.e.}$$

If $L^p_{\mathscr{H}}$ is separable, we can obtain $T(x)$ by an explicit construction. In the general case we argue as follows:

Define a map u: $\mathscr{K}_{\mathscr{H}} \to \mathscr{H}$ by setting

$$u(\varphi) = \int_X (T\varphi)(x) d\mu(x) , \qquad \varphi \in \mathscr{K}_{\mathscr{H}} .$$

Then u is continuous with respect to the $L^1_{\mathscr{H}}$-norm on $\mathscr{K}_{\mathscr{H}}$ because

$$\left\| \int_X (T\varphi)(x) d\mu(x) \right\| \leqq \int_X \| (T\varphi)(x) \| d\mu(x)$$
$$\leqq \int_X \| \varphi(x) \| d\mu(x) .$$

Since $\mathscr{K}_{\mathscr{H}}$ is dense in $L^1_{\mathscr{H}}$, u can be extended by continuity to the whole $L^1_{\mathscr{H}}$ without increasing its norm. We denote the extended map also by u. By Theorem 3 there exists a function $\Phi(x)$ from X into $\mathscr{L}(\mathscr{H}, \mathscr{H})$ such that Φ is weakly measurable, uniformly bounded with $\| \Phi(x) \| \leqq \| u \| \leqq 1$ and

$$u(f) = \int_X \Phi(x) f(x) d\mu(x)$$

for every $f \in L^1_{\mathscr{H}}$. Thus for any $\varphi \in \mathscr{K}_{\mathscr{H}}$

$$\int_X (T\varphi)(x)d\mu(x) = u(\varphi) = \int_X \Phi(x)\varphi(x)d\mu(x) .$$

Since T commutes with multiplication by functions in $\mathscr{A}$ and every $\alpha \in \mathscr{A}$ is continuous, we get

$$\begin{aligned}\int_X \alpha(x)\Phi(x)\varphi(x)d\mu(x) &= \int_X \Phi(x)\alpha(x)\varphi(x)d\mu(x) \\ &= \int_X (T\alpha\varphi)(x)d\mu(x) = \int_X \alpha(x)(T\varphi)(x)d\mu(x) .\end{aligned}$$

By the weak* density of $\mathscr{A}$ in L^∞, this implies

$$(T\varphi)(x) = \Phi(x)\varphi(x) \text{ a.e.}$$

for all $\varphi \in \mathscr{K}_{\mathscr{H}}$. If $\hat{\Phi}$ denotes the operator in $L^p_{\mathscr{H}}$ defined by

$$(\hat{\Phi}f)(x) = \hat{\Phi}(x)f(x) \text{ a.e.,}$$

then we have $T\varphi = \hat{\Phi}\varphi$ for all $\varphi \in \mathscr{K}_{\mathscr{H}}$. Since both T and $\hat{\Phi}$ are bounded in $L^p_{\mathscr{H}}$ and $\mathscr{K}_{\mathscr{H}}$ is dense in $L^p_{\mathscr{H}}$, it follows that $T = \hat{\Phi}$. Now we have only to put $\Phi(x) = T(x)$ in order to get the theorem.

If $p = \infty$ and T is bounded and weakly* continuous, then the transposed map T^* of T maps $L^1_{\mathscr{H}}$ into $L^1_{\mathscr{H}}$. Since T^* commutes with multiplication by functions in $\mathscr{A}$, T^* is expressed by an operator-valued function which is weakly measurable and uniformly bounded. Therefore T is also a uniformly bounded and weakly measurable operator-valued function $T(x)$. In this case, we clearly have

$$(Tf)(x) = T(x)f(x) \text{ l.a.e.}$$

for all $f \in L^\infty_{\mathscr{H}}$.

References

1. N. Bourbaki, *Intégration*, Chap. 6: Intégration vectorielle, Hermann, Paris, 1960.
2. J. Dixmier, *Les algèbres d'opérateurs dans l'espace hilbertien*, Gauthier-Villars, Paris, 1957.
3. N. Dunford and B. Pettis, *Linear operations on summable functions*, Trans. Amer. Math. Soc., **47** (1940), 323-392.
4. T. P. Srinivasan, *Doubly invariant subspaces*, Pacific J. Math., **14** (1964), 691-697.
5. ———, *Simply invariant subspaces and generalized analytic functions* (to appear).

UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA

IBARAKI UNIVERSITY
MITO, JAPAN

PANJAB UNIVERITY
CHANDIGARH, INDIA

FIELDS DEFINED BY POLYNOMIALS

LOWELL A. HINRICHS, IVAN NIVEN AND C. L. VANDEN EYNDEN

1. Introduction. First we consider the following question, where F is any field. For what pairs P and Q of polynomials in two variables with coefficients in F do the definitions

$$\text{(I)} \qquad a \oplus b = P(a, b) , \qquad a \odot b = Q(a, b) ,$$

for all a and b in F yield a field $(F, \oplus, \odot)$? It turns out that the answer is different for infinite fields than for finite fields, as shown in §§ 2 and 3.

Next let R be the field of real numbers. For what quadruples P_1, P_2, Q_1, Q_2 of real polynomials in four variables is $(R \times R, \oplus, \odot)$ a field, when we set

$$\text{(II)} \qquad \begin{aligned} (a, b) \oplus (c, d) &= (P_1(a, b, c, d), P_2(a, b, c, d)) , \\ (a, b) \odot (c, d) &= (Q_1(a, b, c, d), Q_2(a, b, c, d)) , \end{aligned}$$

where (x, y) denotes an ordered pair of real numbers? This question is partially answered in §§ 4 and 5, and in § 6 it is shown that the polynomials may be of arbitrarily high degree. In § 7 it is proved that if definitions (II) do give a field, it must be isomorphic to the field of complex numbers.

2. The one-dimensional case.

THEOREM 1. *Let F be an infinite field. The system $(F, \oplus, \odot)$ in* (I) *is a field if and only if*

$$\text{(1)} \qquad \begin{aligned} P(a, b) &= a \oplus b = a + b + \gamma \\ Q(a, b) &= a \odot b = \gamma\sigma(a + b) + \sigma ab + \gamma^2\sigma - \gamma , \end{aligned}$$

where $\gamma \in F$, $\sigma \in F$ and $\sigma \neq 0$. When these conditions are satisfied the field $(F, \oplus, \odot)$ is isomorphic to F, thus $(F, \oplus, \odot) \cong (F, +, \cdot)$.

Proof. We first assume that $(F, \oplus, \odot)$ is a field and show that the polynomials P and Q have the prescribed form. By associativity we have $P(P(a, b), c) = P(a, P(b, c))$ identically in a, b, c. Now if P is of degree n in a, the degrees of the left and right sides of this identity in a are n^2 and n respectively. Since F is infinite it follows

Received November 12, 1962, and in revised form June 14, 1963. The first author listed was supported by NSF contract G 19859, the other two authors by NSF contract G 19016.

that $n^2 = n$ and hence $n = 1$. We conclude that $P(a, b)$ is linear in a and b, and the same holds for $Q(a, b)$.

Using this linearity and also the commutative properties, we can write

$$a \oplus b = \alpha(a + b) + \beta ab + \lambda \,,$$
$$a \odot b = \rho(a + b) + \sigma ab + \tau \,.$$

Now $\beta = 0$, for if $\beta \neq 0$ we would have

$$(-\alpha/\beta) \oplus b = -\alpha^2/\beta + \lambda \,,$$

and the right member is independent of b.

Suppose first that the additive and multiplicative identities are 0 and 1. Then the equations

$$a \oplus 0 = a \,, \qquad a \odot 0 = 0 \,, \qquad a \odot 1 = a$$

show that $\alpha = 1$ and $\lambda = 0$, that $\rho = 0$ and $\tau = 0$, and that $\sigma = 1$. Thus we have

$$a \oplus b = a + b \,, \qquad a \odot b = a \cdot b \,,$$

so that $\oplus$ and $\odot$ are simply the ordinary operations.

But now suppose that z and u denote the additive and multiplicative identities of the field $(F, \oplus, \odot)$. Then the mapping

$$a \to f(a) = (u - z)a + z$$

gives $f(0) = z$ and $f(1) = u$. Since f is a one-to-one mapping of F onto F, the operations $\oplus'$ and $\odot'$ defined by

$$(2) \qquad \begin{aligned} x \oplus' y &= f^{-1}(f(x) \oplus f(y)) \,, \\ x \odot' y &= f^{-1}(f(x) \odot f(y)) \,, \end{aligned}$$

yield a field $(F, \oplus', \odot')$ which is isomorphic under f to $(F, \oplus, \odot)$. But it is easily checked that $\oplus'$ and $\odot'$ are again polynomial operations in the sense of (I). Furthermore note that

$$x \oplus' 0 = x \odot' 1 = x \,, \qquad x \odot' 0 = 0 \,,$$

and so by the argument of the preceding paragraph we conclude that $\oplus'$ and $\odot'$ are just $+$ and $\cdot$. Now if we substitute $x = f^{-1}(a)$ and $y = f^{-1}(b)$ into equations (2) and apply f to both sides we get

$$(3) \qquad \begin{aligned} a \oplus b &= f(f^{-1}(a) + f^{-1}(b)) = a + b - z \,, \\ a \odot b &= f(f^{-1}(a) \cdot f^{-1}(b)) = (a - z)(b - z)(u - z)^{-1} + z \,. \end{aligned}$$

Writing γ for $-z$ and σ for $(u - z)^{-1}$ we see that equations (3) are the same as (1).

Conversely, given any elements γ and $\sigma \neq 0$ of F we see that the operations defined by equations (1) give a field isomorphic to $(F, +, \cdot)$, because the mapping f^{-1} is an isomorphism:

$$f^{-1}(a \oplus b) = f^{-1}(a) + f^{-1}(b) ,$$
$$f^{-1}(a \odot b) = f^{-1}(a) \cdot f^{-1}(b) .$$

3. **Finite fields.** The restriction of Theorem 1 to infinite fields was necessary because in the proof use was made of the fact that polynomials agreeing on infinite sets must be identical. Now for a finite field F of order $q = p^n$ we see that a system $(F, \oplus, \odot)$ in (I) is a field with

$$P(a, b) = a \oplus b = a^q + b^q , \qquad Q(a, b) = a \odot b = a^q b^q .$$

But these are artificial definitions since $a^q = a$ identically in a in the finite field. However, Theorem 1 fails in a genuine sense for all cases except $q = 2, 3, 4$, as can be seen as follows.

Let g be any permutation on F leaving 0 and 1 invariant. Now g is a polynomial function because we can construct a polynomial to agree with g over the q elements of the field. Similarly the operations $\oplus$ and $\odot$ defined by

$$(4) \qquad \begin{aligned} a \oplus b &= g^{-1}(g(a) + g(b)) , \\ a \odot b &= g^{-1}(g(a) \cdot g(b)) , \end{aligned}$$

are polynomial functions. If Theorem 1 were true for the finite field F then equations (4) would be of the form (1) for some γ and σ. But from (4) we see that $a \oplus 0 = a$ and $a \odot 1 = a$, so that 0 and 1 are the additive and multiplicative identities of $(F, \oplus, \odot)$. Hence in (1) we see that $\gamma = 0$ and $\sigma = 1$. Thus $\oplus$ and $\odot$ would be the ordinary operations and (4) would be

$$a + b = g^{-1}(g(a) + g(b)) ,$$
$$a \cdot b = g^{-1}(g(a) \cdot g(b)) .$$

It follows that g is an automorphism of $(F, \oplus, \odot)$. But there exist exactly n automorphisms of a field with p^n elements [4, § 38], Since there are $(p^n - 2)!$ permutations g of F leaving 0 and 1 invariant, and since $(p^n - 2)! > n$ if $p^n \geqq 5$, it follows that Theorem 1 fails for finite fields of order $q = p^n \geqq 5$.

On the other hand suppose that F is a finite field of order $q = p^n = 2, 3$, or 4. Suppose further that there are polynomials P and Q for which the operations $a \oplus b = P(a, b)$ and $a \odot b = Q(a, b)$ yield a field $(F, \oplus, \odot)$. Using the mapping $f(a) = (u - z)a + z$, we apply f^{-1} as in equations (2). Thus we move from $(F, \oplus, \odot)$ to $(F, \oplus', \odot')$

having 0 and 1 as additive and multiplicative identities. Now simple examination of the addition and multiplication tables for finite fields with 2, 3 or 4 elements shows that the operations $\oplus'$ and $\odot'$ must be the ordinary operations of addition and multiplication. Thus we can get equations (3) and the rest of the proof follows as in Theorem 1. We have proved the following result.

THEOREM 2. *Theorem* 1 *holds for only those finite fields with* 2, 3 *or* 4 *elements.*

4. The complex case: a simplification. The definition (II) allows considerably more latitude for the operations $\oplus$ and $\odot$ than exists in the one-dimensional case, and the problem appears to be correspondingly more difficult. To simplify things we show first that there is no great loss in generality in presuming that the additive ard multiplicative identities of the field $(R \times R, \oplus, \odot)$ are $(0, 0)$ and $(1, 0)$. For let the zero and unity of the field be denoted by (p, q) and (r, s). We define

$$[a, b] = (ar - ap - bs + bq + p, as - aq + br - bp + q) , \tag{5}$$

and note that

$$[0, 0] = (p, q), [1, 0] = (r, s) .$$

The right member of (5) is simply

$$(a, b)(r - p, s - q) + (p, q) ,$$

where the multiplication and addition are as in the field of complex numbers. Since $(p, q) \neq (r, s)$ we see that $(r - p, s - q) \neq (0, 0)$ and so (5) is a one-to-one mapping of $R \times R$ onto $R \times R$. If we extend the multiplications $\oplus$ and $\odot$ to the pairs $[a, b]$ by the use of (5) we see that

$$[0, 0] \oplus [a, b] = [1, 0] \odot [a, b] = [a, b] .$$

Furthermore, $[a, b] = (x, y)$ implies not only that x and y are polynomials in a and b by (5), but also that a and b are polynomials in x and y. Hence any system of pairs (a, b) with $\oplus$ and $\odot$ defined by (II) can be transformed into an isomorphic system of pairs $[a, b]$ with $\oplus$ and $\odot$ defined by (5) and (II). Thus all fields of the required sort can be generated in a simple way as in § 2 from those having $(0, 0)$ and $(1, 0)$ as zero and unit.

5. The complex case with linearity.

THEOREM 3. *Let the operations* $\oplus$ *and* $\odot$ *be defined as in* (II), *and assume that each of* P_1, P_2, Q_1, Q_2 *is linear in each argument* *separately. Then* $(R \times R, \oplus, \odot)$ *is a field with* $(0, 0)$ *and* $(1, 0)$ *as zero and unity if and only if*

$$(a, b) \oplus (c, d) = (a + c, b + d) \quad \text{and}$$
$$(a, b) \odot (c, d) = (ac + \gamma bd, ad + bc + \delta bd)$$

for some $\gamma \in R$ *and* $\delta \in R$ *with* $\delta^2 + 4\gamma < 0$. *When these conditions are satisfied,* $(R \times R, \oplus, \odot)$ *is isomorphic to the field of complex numbers, that is,* $(R \times R, \oplus, \odot) \cong (C, +, \cdot)$.

Proof. First we assume that $(R \times R, \oplus, \odot)$ is a field. By the commutative property $P_1(a, b, c, d)$ is symmetric in a and c and also in b and d; likewise for P_2, Q_1 and Q_2. Thus we can write

$$\begin{aligned} P_1(a, b, c, d) = \alpha_0 &+ \alpha_1(a + c) + \alpha_2(b + d) + \alpha_{12}(ab + cd) \\ &+ \alpha_{13}ac + \alpha_{24}bd + \alpha_{14}(ad + bc) + \alpha_{123}(abc + acd) \\ &+ \alpha_{124}(abd + bcd) + \alpha_{1234}abcd \,. \end{aligned}$$

We represent P_2, Q_1 and Q_2 by similar expressions with the α's replaced by β's, γ's and δ's respectively. From the relation $(a, b) \oplus (0, 0) = (a, b)$ we deduce

$$P_1(a, b, 0, 0) = a \,, \qquad P_2(a, b, 0, 0) = b \,,$$

from which it follows that

$$\alpha_1 = \beta_2 = 1 \quad \text{and} \quad \alpha_0 = \beta_0 = \alpha_2 = \beta_1 = \alpha_{12} = \beta_{12} = 0 \,.$$

Now define (h, k) by the relation $(1, 0) \oplus (1, 0) = (h, k)$. Then the distributive property implies that

$$(a, b) \odot (h, k) = (a, b) \oplus (a, b)$$

and so we obtain

$$\begin{aligned} P_1(a, b, a, b) &= Q_1(a, b, h, k) \\ &= 2a + \alpha_{13}a^2 + \alpha_{24}b^2 + 2\alpha_{14}ab + 2\alpha_{123}a^2b \\ &\quad + 2\alpha_{124}ab^2 + \alpha_{1234}a^2b^2 \,. \end{aligned}$$

But $Q_1(a, b, h, k)$ is linear in a and b, and hence

$$\alpha_{13} = \alpha_{24} = \alpha_{14} = \alpha_{123} = \alpha_{124} = \alpha_{1234} = 0 \,.$$

The relation $P_2(a, b, a, b) = Q_2(a, b, h, k)$ yields an analogous result for the β's, and so we get

$$(a, b) \oplus (c, d) = (a + c, b + d) \,.$$

Next, from the relation $(a, b) \odot (0, 0) = (0, 0)$ we see that

$$Q_1(a, b, 0, 0) = Q_2(a, b, 0, 0) = 0 ,$$

and so

$$\gamma_0 = \gamma_1 = \gamma_2 = \gamma_{12} = \delta_0 = \delta_1 = \delta_2 = \delta_{12} = 0 .$$

From $Q_1(a, b, 1, 0) = a$ and $Q_2(a, b, 1, 0) = b$ we obtain

$$\gamma_{13} = \delta_{14} = 1 , \qquad \delta_{13} = \gamma_{14} = \gamma_{123} = \delta_{123} = 0 .$$

Thus we have

$$Q_1(a, b, c, d) = ac + \gamma_{24}bd + \gamma_{124}(bcd + abd) + \gamma_{1234}abcd ,$$
$$Q_2(a, b, c, d) = ad + bc + \delta_{24}bd + \delta_{124}(bcd + abd) + \delta_{1234}abcd .$$

Also the equations

$$\begin{aligned}(a, b) \odot (1, 1) &= (a, b) \odot (1, 0) \oplus (a, b) \odot (0, 1) \\ &= (a, b) \oplus (a, b) \odot (0, 1)\end{aligned}$$

imply that

$$Q_1(a, b, 1, 1) = a + Q_1(a, b, 0, 1), Q_2(a, b, 1, 1) = b + Q_2(a, b, 0, 1) .$$

This yields

$$\gamma_{124} = \gamma_{1234} = \delta_{124} = \delta_{1234} = 0 ,$$

and so we have, removing subscripts,

$$(a, b) \odot (c, d) = (ac + \gamma bd, ad + bc + \delta bd) .$$

Finally, if $(a, b) \neq (0, 0)$, there must exist real numbers x and y such that $(a, b) \odot (x, y) = (1, 0)$. This gives a pair of linear equations with determinant $a^2 + \delta ab - \gamma b^2$. This must not vanish except for $a = 0$ and $b = 0$, and so we conclude that

$$\delta^2 + 4\gamma < 0 .$$

Conversely, to prove that the operations $\oplus$ and $\odot$ in the statement of the theorem do give a field isomorphic to the field of complex numbers, define α and β by

$$\alpha = \frac{\delta}{2} , \qquad \beta = \frac{\sqrt{-4\gamma - \delta^2}}{2} .$$

Since $\beta \neq 0$ the mapping

$$\phi\colon (a, b) \to (a + \alpha b, \beta b)$$

is one-to-one from C onto itself. As in Theorem 1 we point out that by a not difficult calculation

$$(a, b) \oplus (c, d) = \phi^{-1}(\phi(a, b) + \phi(c, d))$$

and

$$(a, b) \odot (c, d) = \phi^{-1}(\phi(a, b) \cdot \phi(c, d)) .$$

Thus the mapping ϕ is an isomorphism from $(R \times R, \oplus, \odot)$ to $(C, +, \cdot)$.

As a variation on Theorem 3 we prove the following; see [2, p. 251] for a related result.

THEOREM 4. *In Theorem 3 replace the hypothesis that P_1, P_2, Q_1 and Q_2 are linear by the assumption*

$$(a, b) \odot (c, 0) = (ac, bc) \tag{6}$$

for all a, b, c in R. Then the conclusion of Theorem 3 holds.

Proof. If first we assume the definitions of $\oplus$ and $\odot$ as in the equations of Theorem 3, then we have a field, and we note that (6) follows. Conversely, suppose that $(R \times R, \oplus, \odot)$ is a field with the usual zero and unity and such that (6) holds. Then we note that

$$\begin{aligned}(aP_1(x, y, z, w), aP_2(x, y, z, w)) \\ &= (a, 0) \odot (P_1(x, y, z, w), P_2(x, y, z, w)) \\ &= (a, 0) \odot ((x, y) \oplus (z, w)) \\ &= (ax, ay) \oplus (az, aw) \\ &= (P_1(ax, ay, az, aw), P_2(ax, ay, az, aw)) .\end{aligned}$$

Thus P_1 and P_2 are homogeneous and linear.

Turning to the operation $\odot$ we note that

$$\begin{aligned}(aQ_1(x, y, z, w), aQ_2(x, y, z, w)) \\ &= (a, 0) \odot ((x, y) \odot (z, w)) \\ &= (ax, ay) \odot (z, w) \\ &= (Q_1(ax, ay, z, w), Q_2(ax, ay, z, w)) .\end{aligned}$$

Applying the commutative property we get

$$\begin{aligned}(a^2Q_1(x, y, z, w), a^2Q_2(x, y, z, w)) \\ &= (Q_1(ax, ay, az, aw), Q_2(ax, ay, az, aw))\end{aligned}$$

and hence Q_1 and Q_2 are homogeneous of degree 2. Now the relations

$$Q_1(a, b, 0, 0) = Q_2(a, b, 0, 0) = 0$$

show that $Q_1(a, b, c, d)$ and $Q_2(a, b, c, d)$ have no a^2 or b^2 terms. From the commutative property it follows that Q_1 and Q_2 have no c^2 or d^2 terms. Thus Q_1 and Q_2 are linear in each argument separately, as also are P_1 and P_2, and so we can apply Theorem 3 to complete the proof.

6. Linearity not necessary. Here we show that $(R \times R, \oplus, \odot)$ with operations defined by (II) may be a field with the usual zero and unity even though P_1, P_2, Q_1 and Q_2 are not linear in the separate arguments. For let T be any polynomial in one variable with real coefficients and set $S(x) = x(x-1)T(x)$. Define the mapping ϕ by

$$\phi\colon (a, b) \to (a + S(b), b) .$$

Then ϕ is a one-to-one mapping of C onto itself which leaves $(0, 0)$ and $(1, 0)$ invariant. Thus if we define

$$(a, b) \oplus (c, d) = \phi^{-1}(\phi(a, b) + \phi(c, d)),$$
$$(a, b) \odot (c, d) = \phi^{-1}(\phi(a, b), \phi(c, d)),$$

we get $(R \times R, \oplus, \odot)$ isomorphic to $(C, +, \cdot)$, the two field representations having common zero and unity. It is clear that the polynomials P_1, P_2, Q_1 and Q_2 may be given arbitrarily high degrees by the proper choice of T.

7. A general theorem. A question left unanswered in the preceding three sections is whether any field satisfying (II) must be isomorphic to the complex numbers. That the answer is yes is a special case of the following result.

THEOREM 5. *Let f and g be continuous mappings from $R^n \times R^n$ into R^n, and suppose that the binary operations $\oplus$ and $\odot$ defined on R^n by*

$$x \oplus y = f(x, y), x \odot y = g(x, y)$$

make $(R^n, \oplus, \odot)$ a field. Then $n = 1$ or 2 and the field is the real field or the field of complex numbers accordingly.

Proof. Let $\ominus x$ and x^* denote the inverses of x under $\oplus$ and $\odot$ respectively. We will show that the maps

$$x \to \ominus x \quad \text{and} \quad x \to x^*$$

are continuous and thus $(R^n, \oplus, \odot)$ is a topological field. Then the known result that any locally compact connected topological field satisfying the first axiom of countability is either the real or the complex numbers will yield the theorem; cf. [3, p. 173].

Consider the map $T: R^{2n} \to R^{2n}$ defined by $T: (x, y) \to (x, x \oplus y)$, where x and y belong to R^n. It is easily seen that T is continuous, one-to-one and onto. It is claimed that T is a homeomorphism. For suppose that A is an open subset of R^{2n} and $a \in A$. Let K be a compact neighborhood of a contained in A. Then T is a homeomorphism of K onto $T[K]$ and so by Brouwer's theorem [1, p. 100] on the invariance of domains the interior of K maps onto an open set. Thus $T(a)$ is an interior point of $T[A]$; we see that T takes open sets onto open sets.

Now T^{-1} is the mapping $(x, s) \to (x, s \ominus x)$, and so, letting s be the additive identity of $(R^n, \oplus, \odot)$, we see that the map $x \to \ominus x$ is continuous. The verification that $x \to x^*$ is a continuous map runs along the same lines. Thus with the usual topology $(R^n, \oplus, \odot)$ is either the reals or the complexes. Since R^m homeomorphic to R^n implies $m = n$, the theorem follows.

References

1. S. Lefschetz, *Topology*, 2nd edition, Chelsea, New York, 1959.
2. L. J. Paige and J. D. Swift, *Elements of Linear Algebra*, Ginn, Boston, 1961.
3. L. Pontrjagin, *Topological Groups*, Princeton, 1939.
4. B. L. van der Waerden, *Modern Algebra*, vol. 1, Unger, 1949.

UNIVERSITY OF OREGON

DECOMPOSITION OF SETS OF GROUP ELEMENTS

W. B. LAFFER AND H. B. MANN

In this paper small letters will denote group elements or integers. Large letters will denote sets of these. The cardinal of a set S will be denoted by (S).

1. Sets in Abelian Groups. The problem of decomposition of sets of elements of a finite additive Abelian group, G, of order v, is the following. Given a set of group elements, C, when do there exist sets of group elements, A and B, with Min $(A), (B) \geqq 2$ and $C = A + B = \{a + b \mid a \in A, b \in B\}$? If there are such sets, A and B, then we say that A and B are components of C, and that C is decomposable. We are also concerned with the following question, given a set C and a set A, when is A a component of C? The problems of decomposition are stated analogously when C, A, and B are sets of nonnegative integers. The results for sets of group elements are analagous to the results for sets of nonnegative integers. We include the proofs for both cases because although the techniques used in handling additive problems in finite Abelian groups are analogous to the techniques used in handling additive problems for sets of nonnegative integers (see Mann [5], [6], [7]; Dyson [1]; and Kneser [4]), they are not identical.

In Theorems 1–5 all sets shall be sets of elements from a finite Abelian group, G, of order v.

THEOREM 1. *Let C be sets of elements from the finite Abelian group, G. Let $\bar{C} = \{\bar{c}_1, \bar{c}_1, \cdots, \bar{c}_r\}$ be the complement of C in G. Let $D = \{\bar{c}_r - \bar{C}\} = \{\bar{c}_r - \bar{c}_j \mid j = 1, \cdots, r\}$. Then A is a component of C, if and only if, for every $k \notin D$ we have $A + k \not\subset A + D$.*

Proof. Put $B = \bigcap_{i=1}^{r} \{\bar{c}_i - \bar{A}\}$. Then A is a component of C if and only if $A + B = C$.

Suppose for every $k \notin D$ we have $A + k \not\subset A + D$. Then, for every $k \notin D$ there is an $a \in A$ such that $a + k = \bar{a}_i + d_i$ for every $i = 1, \cdots, r$ where $d_i = \bar{c}_r - \bar{c}_i$ and $\bar{a}_i \in \bar{A}$. Hence for every $i = 1, \cdots, r$ we have $\bar{c}_r - k = a - \bar{a}_i + \bar{c}_i = a + \bar{c}_i - \bar{a}_i = a + b$ where $b \in B$. For every $c \in C$ put $k = \bar{c}_r - c$. Hence $c = a + b$ which implies that $A + B = C$. Thus A is a component of C.

Suppose $A + B = C$. If there is a $k \notin D$ such that $A + k \subset A + D$,

Received July 1, 1963. The first author was supported in part by an NSF cooperative fellowship.

then for every $a \in A$ there is an i such that $a + k = a_i + d_i$. Therefore, $\bar{c}_r - k = a + \bar{c}_i - a_i = a + \bar{b}$ where $\bar{b} \in \bar{B} = \bigcup_{i=1}^{r} \{\bar{c}_i - A\}$. Since $k \neq \bar{c}_r - \bar{c}_j$, we must have $k = \bar{c}_r - c$ for some $c \in C$. Hence for every $a \in A$ we have $c - a = \bar{b} \in \bar{B}$. Therefore, $c \notin A + B$. This is a contradiction and hence the theorem is true.

Corollary 1.1. *Let $\bar{C} = \{\bar{c}\}$. Then A is a component of C if and only if A is not a coset of some subgroup of G.*

Corollary 1.2. *If G is cyclic of prime order and $\bar{C} = \{\bar{c}\}$, then A is a component of C if and only if $1 < (A) < (C)$.*

Definition. We say that A is an m component of C if and only if A is a component of C and $(A) = m$.

Corollary 1.3. *If $\bar{C} = \{\bar{c}_1, \cdots, \bar{c}_r\}$, then $\{0, a\}$ is a 2 component of C if and only if $2a = \bar{c}_i - \bar{c}_j$ implies $a = \bar{c}_i - \bar{c}_m$ for some m.*

Proof. Suppose $\{0, a\}$ is a 2 component of C. Let $2a = \bar{c}_i - \bar{c}_j$. By Theorem 1 if $k \notin D$ then $\{k, a + k\} \not\subset \{0, a\} + D$. Put $k = a + d_i$. Then $\{k, a + k\} = \{a + d_i, 2a + d_i\} = \{a + d_i, d_j\} \subset \{0, a\} + D$. If $k \notin D$, then $\{0, a\}$ is not a 2 component of C. Hence $k \in D$. Thus $k = a + d_i = d_m$ which implies $a = \bar{c}_i - \bar{c}_m$.

Suppose that $2a = \bar{c}_i - \bar{c}_j$ implies $a = \bar{c}_i - \bar{c}_m$ for some m. If $\{0, a\}$ is not a 2 component of C, then by Theorem 1 there is a $k \notin D$ such that $\{k, a + k\} \subset \{0, a\} + D$. This implies that $k = a + d_u$ for some $u = 1, \cdots, r$, and $a + k = 2a + d_u = d_s$. Thus $2a = \bar{c}_u - \bar{c}_s$ and by assumption this implies $a = \bar{c}_u - \bar{c}_t$. Therefore, $k = a + d_u = d_t \in D$. This is a contradiction, and hence $\{0, a\}$ is a 2 component of C.

In Corollaries 1.4–1.9 we shall assume that $\bar{C}$ is a difference set with parameters $v = (G)$, $r = (\bar{C})$, and $\lambda =$ the number of representations which each nonzero element of G has in the form $\bar{c}_i - \bar{c}_j$.

Corollary 1.4. *Let $(\bar{C}) = r < v$. If $2a = 0$ for $a \neq 0$, then $\{0, a\}$ is not a 2 component of C.*

Proof. If $\{0, a\}$ is a 2 component of C then $2a = \bar{c}_i - \bar{c}_i$ for every $i = 1, \cdots, r$, and by Corollary 1.3 we have $r = v$.

Corollary 1.5. *If $(\lambda, v) = 1$, then there does not exist an $a \neq 0$ such that $\{0, a\}$ is a 2 component of C.*

Proof. Suppose there is an $a \neq 0$ such that $\{0, a\}$ is a 2 component of C. Because of Corollary 1.4 we may assume $2a \neq 0$. Since

$\bar{C}$ is a difference set we have

(i) $2a = \bar{c}_{i_1} - \bar{c}_{j_1} = \bar{c}_{i_2} - \bar{c}_{j_2} = \cdots = \bar{c}_{i_\lambda} - \bar{c}_{j_\lambda}$.

By Corollary 1.3 we must have

(ii) $a = \bar{c}_{i_1} - \bar{c}_{m_1} = \bar{c}_{i_2} - \bar{c}_{m_2} = \cdots = \bar{c}_{i_\lambda} - \bar{c}_{m_\lambda}$.

Subtracting a from $2a$ we get

(iii) $a = \bar{c}_{m_1} - \bar{c}_{j_1} = \bar{c}_{m_2} - \bar{c}_{j_2} = \cdots = \bar{c}_{m_\lambda} - \bar{c}_{j_\lambda}$.

If there is an index m_s in (iii) that does not appear as a first index i_t in (ii), then a has at least $\lambda + 1$ distinct representations as an element of $\{\bar{C} - \bar{C}\}$. This contradicts the choice of $a \neq 0$. Hence we must have that every first index in (iii) appears as a first index in (ii).

Thus from (ii) we obtain $\lambda a = \sum_{k=1}^{\lambda} \bar{c}_{i_k} - \bar{c}_{m_k} = 0$. Since $(\lambda, v) = 1$, it follows that $a = 0$. This contradicts our choice of a, hence Corollary 1.5 is true.

COROLLARY 1.6. *If G is cyclic of order $n^2 + n + 1$, and if G has no simple difference set of order $n + 1$, then every set of n^2 elements has a 2 component.*

COROLLARY 1.7. *Let G be cyclic of order $n^2 + n + 1$, and let $n \leqq 1600$. If n is not a prime power, then every set of n^2 elements has a 2 component.*

Proof. This follows from Corollary 1.6 and the paper of Evans and Mann [2].

COROLLARY 1.8. *Let $\lambda > 2$. Suppose there is a cyclic subgroup H of G such that $(H) = \lambda$. Suppose $H + g \subset \bar{C}$ for some $g \in G$. Let $a \in H$ such that a has order λ. Then $\{0, a\}$ is a 2 component of C.*

Proof. If $\bar{C}$ is a difference set, then so is $\bar{C} - g$. Hence without loss of generality we may assume that $H \subset \bar{C}$.

If $a \in H$ and a has order λ, then $ja \in H$ for $0 \leqq j \leqq \lambda - 1$. Since $2a \neq 0$, we have that $2a = ja - (j - 2)a$ for $0 \leqq j \leqq \lambda - 1$ are the λ distinct representations of $2a$ as an element of $\{\bar{C} - \bar{C}\}$. Clearly, $a = ja - (j - 1)a$ for $0 \leqq j \leqq \lambda - 1$ are the λ distinct representations of a as an element of $\{\bar{C} - \bar{C}\}$. By Corollary 1.3 $\{0, a\}$ is a 2 component of C.

An example of a case where this situation actually occurs is the following. Let G be the residues modulo 15. Let $\bar{C} = \{0, 1, 2, 4, 5, 8, 10\}$. Here $\lambda = 3$, $a = 5$, and $g = 0$. This and other examples can be found in [3].

An immediate generalization of Corollary 1.8 is the following.

COROLLARY 1.9. *Let $(\lambda, v) = d > 2$. Suppose H is a cyclic sub-*

group of G of order d. Suppose there are exactly λ/d cosets of H contained in $\bar{C}$. If $a \in H$ such that a is of order d, then $\{0, a\}$ is a 2 component of C.

An example of a case where this situation occurs is again from [3]. It is the geometry modulo 63. We have $\bar{C} = \{0, 1, 2, 3, 4, 6, 7, 8, 9, 12, 13, 14, 16, 18, 19, 24, 26, 27, 28, 32, 33, 35, 36, 38, 41, 45, 48, 49, 52, 54, 56\}$. Here $\lambda = 15$, $d = 3$, and $a = 21$. We have $H = \{0, 21, 42\}$; and the 5 cosets are: $\{3, 24, 45\}$; $\{6, 27, 48\}$; $\{7, 28, 49\}$; $\{12, 33, 54\}$; and $\{14, 35, 56\}$.

DEFINITION. We say that C is indecomposable, if there do not exist sets, A and B, such that $\text{Min}\,(A), (B) \geqq 2$ and $A + B = C$.

THEOREM 2. *Let C be such that $(C) = 3$. If the elements of C are not in progression, then C is indecomposable.*

THEOREM 3. *Let $C = \{0, c_1, c_2, \cdots, c_s\}$. Let $s \geqq 3$ and let $C^* = \{0, c_1, c_2, \cdots, c_{s-1}\}$. If $c_s \notin \{C^* - C^*\} \cup \{C^* + C^*\}$, then C is indecomposable.*

Proof. Suppose there are sets, A and B, with $\text{Min}\,(A), (B) \geqq 2$ and such that $A + B = C$. Since $0 \in C$ we must have $0 = a_0 + b_0$ where $a_0 \in A$ and $b_0 \in B$. Let $A' = A - a_0$ and $B' = B - b_0$. Since $b_0 = -a_0$ we must have $A' + B' = A + B = C$. Thus we may assume without loss of generality that $0 \in A \cap B$. Hence $A \cup B \subset C$.

If $c_s \in A + B$, then $c_s = c_i + c_j$. If $i \neq s$ and $j \neq s$, then $c_s \in \{C^* + C^*\}$ contrary to hypothesis. Thus $c_s \in A \cup B$. Suppose $c_s \in A$. Since $(B) \geqq 2$, we have a $c \in B$ such that $c \neq 0$. We must have $c_s + c = c_i \in C$. Hence $c_s = c_i - c \in \{C^* - C^*\}$ which is contrary to hypothesis. Thus C is indecomposable.

The fact that $(C) \geqq 4$ is necessary is illustrated by the following example. Let $C = \{0, 2c, c\}$. Put $C^* = \{0, 2c\}$. We have $\{C^* - C^*\} \cup \{C^* + C^*\} = \{0, 2c, -2c, 4c\}$, and $c \notin \{C^* - C^*\} \cup \{C^* + C^*\}$ for any choice of c such that $3c \neq 0$. Yet $C = \{0, c\} + \{0, c\}$.

COROLLARY 3.1. *Let $s \geqq 3$. If $(G) = v > [3s(s-1)/2] + 1$, then there exists a set $C \subset G$ such that $(C) = s + 1$ and C is indecomposable.*

Proof. Let $C^* = \{0, c_1, c_2, \cdots, c_{s-1}\}$ be any set of $s - 1$ nonzero elements of G and zero.

We have $(\{C^* + C^*\}) \leqq [s(s-1)/2] + s$ and $(\{C^* - C^*) \leqq s(s-1) + 1$. Since $C^* \subset \{C^* + C^*\} \cap \{C^* - C^*\}$, we have $(\{C^* + C^*\} \cap \{C^* - C^*\}) \geqq s$.

Hence $(\{C^* + C^*\} \cup \{C^* - C^*\}) \leqq [3s(s-1)/2] + 1$.

Since $v > [3s(s-1)/2] + 1$, we must have an element $c_s \in G$ such that $c_s \notin \{C^* + C^*\} \cup \{C^* - C^*\}$. By Theorem 3 $C^* \cup \{c_s\}$ is indecomposable.

Theorem 2 and Corollary 3.1 give us the following.

THEOREM 4. *For any positive integer $s \geqq 2$ there always exists an Abelian group G and a subset C of G such that $(C) = s + 1$ and C is indecomposable.*

THEOREM 5. *Let $\{e_1, \cdots, e_u\} = \{\bar{C} - \bar{C}\}$. Let $s(x, y)$ be the number of solutions, of the group equation $xg = y$ where x is an integer. Suppose $m + 1 \leqq v - (\bar{C})$. If $v > \sum_{t=1}^{u} \sum_{x=2}^{m+1} s(x, e_t) = \delta$, then there exist sets A and B such that $0 \in A$, $(A) = m + 1$, A is in progression and $A + B = C$.*

Proof. There are at most δ solutions of the equations $xg = e_t$ where $2 \leqq x \leqq m + 1$ and $1 \leqq t \leqq u$. Hence if $v > \delta$, there exists an element $a \in G$ such that $xa \neq \bar{c}_i - \bar{c}_j$ for all $\bar{c}_i, \bar{c}_j \in \bar{C}$ and $x = 2, \cdots, m + 1$. We distinguish two cases:

I $m \equiv 0(2)$;

II $m \equiv 1(2)$;

Case I. $m \equiv 0(2)$. Put $A = \{0, a, -a, \cdots, ma/2, -ma/2\}$. For $k \notin D$ suppose that $A + k \subset A + D$. Then $k = ha + d_i$. Choose $|h|$ minimal.

If $h > 0$, then $k + [((m + 2)/2) - h]a = [(m + 2)/2]a + d_i = ja + d_t$. From our choice of a, in follows that $j = m/2$. Hence $a + d_i = d_t$, and so $k = (h - 1)a + d_t$, contradicting our choice of h.

If $h < 0$, then $k + [(-(m + 2)/2) - h]a = -[(m + 2)/2]a + d_i = ja + d_t$. From our choice of a it follows that $j = -m/2$. Hence $d_i = a + d_t$, and so $k = (h + 1)a + d_t$ contradicting our choice of h.

Thus by Theorem 1.1 A is a component of C.

Case II. $m \equiv 1(2)$. Put $A = \{0, a, -a, \cdots, [(m - 1)/2]a, -[(m - 1)/2]a, [(m + 1)/2]a\}$. The argument is the same as above replacing $m + 2$ with $m + 3$ and j will be either $(m + 1)/2$ or $-(m + 1)/2$.

COROLLARY 5.1. *Let $v > \delta$. Then there exist an A_i and a B_i for every i such that $2 \leqq i \leqq m + 1$ such that $0 \in A_i$, $(A_i) = i$, A_i is in progression, and $A_i + B_i = C$.*

COROLLARY 5.2. *If v is a prime, then $v > mu$ implies that for every i for which $2 \leqq i \leqq m+1$ there exist an A_i and a B_i such that $0 \in A_i$, $(A_i) = i$, A_i is in progression, and $A_i + B_i = C$.*

We note by an example that a set C may have an $i+1$ component and not have an i component. Let G be the integers modulo 13. Let $\bar{C} = \{1, 2, 4, 10\}$. By Corollary 1.5 C does not have a 2 component, since $\bar{C}$ is a simple difference set. But $\{0, 9, 12\} + \{0, 7, 9, 12\} = \{0, 3, 5, 6, 7, 8, 9, 11, 12\} = C$.

Sets of integers. From now on our sets shall be sets of nonnegative integers. In particular the complement of a set S, which shall be denoted by $\bar{S}$, shall mean the set of all nonnegative integers which are not in S.

DEFINITION. An n section is a set of nonnegative integers which contain all integers greater than n but does not contain n.

In Theorems 6 and 7 we shall assume that the sets A and C are $\bar{c}_r$ sections.

THEOREM 6. *Let $\bar{C} = \{\bar{c}_1 < \bar{c}_2 < \cdots < \bar{c}_r\}$. Let $D = \{d_i = \bar{c}_r - \bar{c}_i, i = 1, \cdots, r\}$. Then A is a component of C if and only if for every $k \notin D$, $k \leqq c_r$ we have $A + k \not\subset A + D$.*

Proof. Put $B = \bigcap_{i=1}^{r} \{\bar{c}_i - \bar{A}\}$. If $A + B_1 = C$, then $B_1 \subset B$ and also $A + B = C$. Hence, A is a component of C if and only if $A + B = C$.

Suppose for every $k \notin D$, $k \leqq \bar{c}_r$ we have $A + k \not\subset A + D$. Then, for every $k \notin D$, $k \leqq \bar{c}_r$ there is an $a \in A$ such that $a + k = \bar{a}_i + d_i$ for every $i = 1, \cdots, r$ where $d_i = \bar{c}_r - \bar{c}_i$ and $\bar{a}_i \in \bar{A}$. Hence, $\bar{c}_r - k = a - \bar{a}_i + \bar{c}_i$ for every $i = 1, \cdots, r$. Hence, $\bar{c}_r - k = a + b$ where $b \in B$. Put $k = \bar{c}_r - c$ where $c \in C$. Then $k \notin D$ and $k \leqq c_r$. Thus $c = a + b$. Hence $A + B = C$, and A is a component of C.

Now suppose A is a component of C. Hence $A + B = C$. If there is a $k \notin D, k \leqq \bar{c}_r$ such that $A + k \subset A + D$, then for every $a \in A$ there is an i such that $a + k = a_i + d_i$. Hence, $\bar{c}_r - k = a + \bar{c}_i - a_i = a + \bar{b}$, where $\bar{b} \in \bar{B}$. Since $k \neq \bar{c}_r - \bar{c}_j$ for any $j = 1, \cdots, r$, we must have $k = \bar{c}_r - c$ for some $c \in C$. Hence for every $a \in A$ we have $c - a = \bar{b} \in \bar{B}$. Therefore, $c \notin A + B$. This is a contradiction, and hence our theorem is true.

DEFINITION. For m a positive integer we say that A is an m set, if and only if, for all $a \in A$ we have $a + m \in A$.

COROLLARY 6.1. *Let A be an n section such that $0 \in A$. Then*

A is a component of $\overline{\{n\}}$, *if and only if, A is not an m set for* $m \leqq n$.

COROLLARY 6.2. *Let A be an m set for some m such that* $m < n$. *Let* $\bar{A} = \{n_1 < n_2 < \cdots < n_r = n\}$. *Then there is an* $n_i < n_r$ *such that*

$$\text{(i)} \qquad \frac{A(n_r)}{n_r + 1} \geqq \frac{A(n_i)}{n_i + 1}$$

where $A(x)$ *denotes the usual counting function of all* $a \in A$ *such that* $a \leqq x$.

Proof. At the end of Lemma 1 on page 911 of [7] it was shown that if the construction defined there fails at a gap $n_s < n_r = n$, then

$$\frac{C(n)}{n+1} \geqq \frac{A(n_s) + B(n_s) - 1}{n_s + 1} + \left(C(d_s - 1) - \frac{C(n)d_s}{n+1}\right)\frac{1}{n_s + 1}\,.$$

Now let $C = A$ and $B = \{0\}$. Then we have

$$\frac{A(n)}{n+1} \geqq \frac{A(n_s)}{n_s + 1} + \left(A(d_s - 1) - \frac{A(n)d_s}{n+1}\right)\frac{1}{n_s + 1}\,.$$

If we assume that $A(n_i)/(n_i + 1) > A(n)/(n + 1)$ for all $i = 1, \cdots, r - 1$ then it follows that for all $a \in A$ we have $A(a)/(a + 1) > A(n)(n + 1)$. Hence the remainder term, $[A(d_s - 1) - \{A(n)d_s/(n + 1)\}]1/(n_s + 1)$ is positive, and we have $A(n)/(n + 1) > A(n_s)/(n_s + 1)$. This contradicts the assumption that $A(n_i)/(n_i + 1) > A(n)/(n + 1)$ for all $i - 1, \cdots, r - 1$.

If the construction does not fail, then all gaps in C are filled except $n = n_r$. Since in our case $C = A$, we would have that A is a component of $\overline{\{n\}}$. But A is an m set for some $m < n$, and by Corollary 6.1 this is impossible.

An example of a case where we have equality in (i) is $A = \{0, 2, 4\}$. We have $A(1)/2 = A(3)/4 = A(5)/6 = 1/2$.

DEFINITION. The $\bar{c}_r$ section C has an m component if and only if there exist sets, A and B such that $A(\bar{c}_r) = m$ and $A + B = C$.

COROLLARY 6.3. *Let* $\bar{C} = \{\bar{c}_1 < \bar{c}_2\}$ *and let* $0 \in C$. *Then C has a* 2 *component, if and only if,* $\bar{C}$ *is not one of the following three sets*: $\{1, 2\}$; $\{2, 4\}$; $\{3, 5\}$.

COROLLARY 6.4. *If* $\bar{c}_r$ *is such that* $\bar{c}_r - \bar{c}_1 > \bar{c}_{r-1}$, *then C has a* 2 *component.*

Proof. One can easily verify that $\{0, \bar{c}_r - \bar{c}_1\}$ is a 2 component by Theorem 6.

THEOREM 7. *For a given r and m there exist at most a finite number of sets with r gaps which do not have an $m+1$ component.*

That is to say if $\bar{C} = \{\bar{c}_1 < \bar{c}_2 < \cdots < \bar{c}_r\}$, then there are sets A and B such that $A(\bar{c}_r) = m+1$, $0 \in A$, and $A + B = C$ except for at most a finite number of sets C. One can even impose the additional condition that A be in progression.

Proof. We first prove a lemma.

LEMMA. *If there exists an $a \in C$ such that*

$$(1) \qquad xa \neq \bar{c}_i - \bar{c}_j \quad \textit{for} \quad 1 \leqq j < i < r \quad \textit{and} \quad x = 1, \cdots, m+1\,;$$

and

$$(2) \qquad \textit{either} \quad \bar{c}_j < a < (m+1)a < \bar{c}_{j+1} \quad \textit{for some } j \textit{ such that} \quad 1 \leqq j \leqq r-1 \quad \textit{or} \quad 0 < a < (m+1)a < \bar{c}_1\,;$$

then $A = \{0, a, 2a, \cdots, ma\}$ is an $m+1$ component of C.

Proof. By Theorem 6 A is a component of C, if and only if, for all $k \notin D$, $k \leqq \bar{c}_r$, we have $A + k \not\subset A + D$. If for some $k \notin D$, $k \leqq \bar{c}_r$ we have $A + k \subset A + D$, then $k = sa + di$ for some s and i such that $0 < s \leqq m$ and $1 \leqq i \leqq r$. Under the assumption that a satisfies (1) and (2), we shall show that $k + (m - s + 1)a \notin A + D$, and hence $A + k \not\subset A + D$ contrary to the assumption that $A + k \subset A + D$. Hence A will be an $m+1$ component of C.

We have $k + (m - s + 1)a = sa + d_i + (m - s + 1)a = (m+1)a + d_i$ for some i. If $k + (m - s + 1)a \in A + D$ then either

$$(3) \qquad (m+1)a + d_i = ta + d_j \quad \text{for some} \quad t = 0, \cdots, m \quad \text{and} \quad j = 1, \cdots, r \quad \text{and} \quad j \neq i\,;$$

or

$$(4) \qquad (m+1)a + d_i > \bar{c}_r > sa + d_i\,.$$

Now (3) implies $(m + 1 - t)a = \bar{c}_i - \bar{c}_j$ contrary to (1), and (4) implies $(m+1)a > \bar{c}_i > sa \geqq a > 0$ contrary to (2). Hence A is an $m+1$ component of C. This completes the proof of the lemma.

Let $\delta = (\{\bar{c}_i - \bar{c}_j \mid 1 \leqq j < i \leqq r\})$. Then $1 \leqq \delta \leqq (r^2 - r)/2$. If there are at least $(m+1)\delta + 1$ choices for a between $\bar{c}_j + 1$ and $[\bar{c}_{j+1}/(m+1)] - 1$ inclusive or between 1 and $[\bar{c}_1/(m+1)] - 1$ inclusive, then we can choose an a so that conditions (1) and (2) of the lemma are satisfied. Thus conditions (1) and (2) of the lemma are satisfied if either

$$(5) \qquad \bar{c}_{j+1} \geqq (m+1)(\delta(m+1) + \bar{c}_j + 2) \quad \text{for} \quad 1 \leqq j \leqq r-1$$

or

$$(6) \qquad \bar{c}_1 \geqq (m+1)(\delta(m+1)+2) .$$

Let $\delta(m+1) + 2 = n$. If (5) and (6) both fail, then we must have $\bar{c} < (m+1)n$

$$\begin{aligned}
&\bar{c}_2 < (m+1)(n+\bar{c}_1) < n((m+1)+(m+1)^2)\\
&\vdots\\
&\bar{c}_i < (m+1)(n+\bar{c}_{i-1}) < n\sum_{q=1}^{i}(m+1)^q\\
&\vdots\\
&\bar{c}_r < (m+1)(n+\bar{c}_{r-1}) < n\sum_{q=1}^{r}(m+1)^q .
\end{aligned}$$

Hence if $c_r \geqq n\sum_{q=1}^{r}(m+1)^q$ then either condition (5) or condition (6) is satisfied and a can be chosen so that conditions (1) and (2) of the lemma are both satisfied.

Thus if $\bar{c}_r \geqq (\delta(m+1)+2)\sum_{q=1}^{r}(m+1)^q$, then there is an $a \in C$ such that $\{0, a, 2a, \cdots, ma\}$ is an $m+1$ component of C. Since for a fixed r, δ is bounded, it follows that the number of sets C, with r gaps which do not have an $m+1$ component is finite.

In Theorems 8–11, we shall make no restriction on the number of gaps that a set C may have.

DEFINITION. A set C is said to be strictly decomposable if there are sets, A and B, such that $\operatorname{Min}(A), (B) \geqq 2$ and $A + B = C$.

DEFINITION. A set C is said to be asymptotically decomposable if there are sets, A and B, such that $\operatorname{Min}(A), (B) \geqq 2$ and $A + B = C^*$ where $(C^* \cap \bar{C}) < \infty$ and $(C \cap \bar{C}^*) < \infty$. We write $A + B \sim C$.

THEOREM 8. *Let C be given. Let $\{n_i\}$ be a monotonically increasing sequence of nonnegative integers. Let $C_k = C \cap [0, k]$. Then a necessary and sufficient condition that C be strictly decomposable is that for each n_i in the sequence there exist a pair of sets, A_{n_i} and B_{n_i}, such that $\{A_{n_i} + B_{n_i}\} \cap [0, n_i] = C_{n_i}$; and there exists a positive integer N such that whenever $n_i \geqq N$ we have* $\operatorname{Min} A_{n_i}(N), B_{n_i}(N) \geqq 2$.

Proof. If C is strictly decomposable, then we have a pair of sets,

A and B, such that Min $(A), (B) \geqq 2$ and $A+B=C$. Put $A_{n_i} = A \cap [0, n_i]$ and $B_{n_i} = B \cap [0, n_i]$. If for some n_i we have $\{A_{n_i} + B_{n_i}\} \cap [0, n_i] \neq C \cap [0, n_i]$, then $\{A + B\} \cap [0, n_i] \neq C \cap [0, n_i]$ which is a contradiction. Since Min $(A), (B) \geqq 2$ we must have a positive integer N such that Min $A(N), B(N) \geqq 2$. This clearly implies that for $n_i \geqq N$ we have Min $A_{n_i}(N), B_{n_i}(N) \geqq 2$. Hence our condition is necessary.

Now let $\{m_j\}_{j=0}^{\infty}$ be any monotonically increasing sequence of non-negative integers. Put $A_{m_0}, i = A_{n_i} \cap [0, m_0]$ and $B_{m_0}, i = B_{n_i} \cap [0, m_0]$. Since there are only a finite number, 2^{2m_0+2}, of choices for each pair of sets, A_{m_0}, i and B_{m_0}, i, there must be at least one pair, A_{m_0} and B_{m_0}, which is repeated for an infinite number of indices i.

Let $\{n_i^{(1)}\}$ be the subsequence of $\{n_i\}$ for which $A_{m_0} = A_{n_i}(1) \cap [0, m_0]$ and $B_{m_0} = B_{n_i}(1) \cap [0, m_0]$. Now $A_{m_0} + B_{m_0} \subset C$, since in the original construction of A_{m_0} and B_{m_0} the n_i may be chosen arbitrarily large. Also we have $\{A_{m_0} + B_{m_0}\} \cap [0, m_0] = C_{m_0}$.

We repeat this process using m_1 and the sequence $\{n_i^{(1)}\}$. Put $A_{m_1}, i = A_{n_i}(1) \cap [0, m_1]$ and $B_{m_1}, i = B_{n_i}(1) \cap [0, m_1]$. Again we must have at least one pair of sets, A_{m_1} and B_{m_1}, that repeats an infinite number of times. This pair, A_{m_1} and B_{m_1}, determines a subsequence $\{n_i^{(2)}\} \subset \{n_i^{(1)}\}$. We must have $A_{m_1} + B_{m_1} \subset C$ and $\{A_{m_1} + B_{m_1}\} \cap [0, m_1] = C_{m_1}$.

Continuing in this way, we have for each m_j a pair of sets, A_{m_j} and B_{m_j}, and a subsequence $\{n_i^{(j+1)}\} \subset \{n_i^{(j)}\}$ such that $A_{m_j} = A_{n_i}(j+1) \cap [0, m_j]$ and $B_{m_j} = B_{n_i}(j+1) \cap [0, m_j]$. For each m_j we also have $A_{m_j} + B_{m_j} \subset C$ and $\{A_{m_j} + B_{m_j}\} \cap [0, m_j] = C_{m_j}$.

Put $A = \bigcup_{j=0}^{\infty} A_{m_j}$ and $B = \bigcup_{j=0}^{\infty} B_{m_j}$. Since in each subsequence, $\{n_i^{(j)}\}$, there exists an $n_i^{(j)}$ such that $n_i^{(j)} \geqq N$, we have that Min $A_{m_j}(N)$, $B_{m_j}(N) \geqq 2$, and hence Min $(A), (B) \geqq 2$. If $A + B \neq C$, then there is a section of C, say C_k, for which the decomposition fails. Let $m_j > k$. Then for the subsequence, $\{n_i^{(j+1)}\}$, we have

$$\{A_{n_i}(j+1) + B_{n_i}(j+1)\} \cap C_{m_j} \neq C_{m_j} = C_{n_i}(j+1) \cap [0, m_j] .$$

This contradicts the original hypothesis. Hence $A + B = C$ is a strict decomposition of C, and our condition is sufficient. This completes the proof of Theorem 8.

If C has a finite number of gaps, C is a section. If C has infinitely many gaps then Theorem 8 shows that the problem of strict decomposability reduces to the problem of decomposability of sections because we can choose for $\{n_i\}$ a sequence of gaps of C.

COROLLARY 8.1. *Let $\{n_i\}$ be an infinite sequence of elements of $\bar{C}$. Then C is strictly decomposable, if and only if, every section $\{C \cap [0, n_i)\} \cup [n_i + 1, \infty)$ is decomposable.*

THEOREM 9. *Let C be given. Let A be such that $\bar{A} \supsetneqq \bar{C}$ and $(A) \geqq 2$. Let $f(\bar{a})$ be the number of representations of $\bar{a}$ in the form $a_i + \bar{c}_j - a_m$. If for every $\bar{a} \in \bar{A}$, such that $\bar{a} \geqq 0$, and such that $\bar{a} \in A + \bar{C} - A$ we have $f(\bar{a}) < A(\bar{a})$, then there exists a set B such that $A + B = C$ is a strict decomposition of* C.

Proof. Put $B = \bigcap_{i=1}^{\infty} \{\bar{c}_i - \bar{A}\}$. Now $B \neq \phi$ since $0 \in B$ and clearly $A + B \subset C$.

Let $\bar{a} \in \bar{A}$ such that $\bar{a} \in C$. If $\bar{a} \notin A + B$, then it must be true that for every $a \in A$ such that $a < \bar{a}$, we must have $\bar{a} - a \in \bar{B} = \bigcup_{i=1}^{\infty} \{\bar{c}_i - A\}$. Hence for every $a \in A$ such that $a < \bar{a}$ we must have that there exists a $\bar{c}_k \in \bar{C}$ and an $a' \in A$ such that $\bar{a} = a + \bar{c}_k - a' \in A + \bar{C} - A$. There exist at least $A(\bar{a})$ such representations of $\bar{a}$, since there are $A(\bar{a})a$'s in A such that $a < \bar{a}$. Hence $f(\bar{a}) \geqq A(\bar{a})$ contrary to hypothesis. Hence there is an $a \in A$ such that $\bar{a} - a \in B$. Hence $A + B = C$. Since $A \neq C$, $A \subset C$, we must have $(B) \geqq 2$, and thus this decomposition of C is strict.

THEOREM 10. *Let C be an infinite set. If C is asymptotically decomposable, then there is an integer k such that for all positive integers m there are infinitely many pairs of elements $c_1, c_2 \in C$ such that $m < c_1 < c_2 < c_1 + k$.*

Proof. Since C is asymptotically decomposable, there is a C^* such that $C^* \sim C$ and C^* is strictly decomposable. And if C^* satisfies the conclusions of Theorem 10, then so does C. Hence without loss of generality we may assume that C is strictly decomposable.

Since C is an infinite set, and since C is strictly decomposable, at least one of the two components is infinite. Suppose without loss of generality that $(B) = \infty$. Let $a_1, a_2 \in A$ where $0 \leqq a_1 < a_2 < k$. Let m be any positive integer. Then there are an infinite number of elements $b \in B$ such that $a_1 + b = c_1 > m$ and $a_2 + b = c_2$. Now $c_2 - c_1 = a_2 - a_1 < k$, and we have $m < c_1 < c_2 < c_1 + k$.

COROLLARY 10.1. *Let C be an infinite set. Let $f(n) = c_n - c_{n-1}$ for $n \geqq 2$ where c_{n-1} and c_n are consecutive elements of C. If there exists an integer m such that for $n \geqq m$, $f(n)$ is increasing, then C is asymptotically decomposable if and only if $f(n)$ is bounded.*

THEOREM 11. *Let $g(y) = \overline{\lim}_{x \to \infty} [C(x + y) - C(x)]$. If $g(y)$ is bounded for all y, then C is not asymptotically decomposable as the sum of two infinite sets.*

Proof. Suppose $C \sim A + B = C^*$ where $(A) = (B) = \infty$. Clearly

$\overline{\lim}_{x\to\infty}[C^*(x+y)-C^*(x)] = \overline{\lim}_{x\to\infty}[C(x+y)-C(x)] = g(y)$. Let $B = \{b_0 < b_1 < b_2 < \cdots\}$. Then $g(y) \geqq \overline{\lim}_{j\to\infty}[C^*(b_j+y)-C^*(b_j)] \geqq A(y)$. Since for all $a \leqq y$ and $a \in A$ we must have $b_j \leqq b_j + a \leqq b_j + y$. Hence for all $b_j \in B$ and y we must have $C^*(b_j+y) - C^*(b_j) \geqq A(y)$. But if $g(y)$ is bounded, then $A(y)$ is bounded and A is not an infinite set.

Let P be the set of all primes. It is easy to show that P is not strictly decomposable.

THEOREM 12. *If $A + B \sim P$, the set of all primes, then $(A) = (B) = \infty$.*

Proof. Suppose $A = \{a_1 < a_2 < \cdots < a_n\}$ and $A + B \sim P$. Then $A - a_1 + B + a_1 \sim P$. Thus we may without loss of generality assume $a_1 = 0$. Let $N = \max\{\bar{p} \in A + B, p \notin A + B\}$. Then whenever $b \in B$ and $b > N$, we must have $a_i + b \in P$ for $i = 1, \cdots, n$, and in particular $b \in P$.

Choose n primes $p_1, \cdots, p_n$ such that $(a_i, p_i) = 1 = (p_i, p_j)$ for $i \neq j$ and $i = 2, \cdots, n$ and $(p_1, a_i) = 1 = (p_1, p_i)$ for $i = 2, \cdots, n$. Consider the solutions to the simultaneous congruences. $x \equiv a_i(p_i)$ for $i = 2, \cdots, n$ and $x \equiv -a_2(p_1)$. The set of solutions forms an arithemetic progression $\{x + k\prod_{i=1}^n p_i\}_{k=0}^\infty$ with $(x, \prod_{i=1}^n p_i) = 1$. By the Dirichelet theorem there exist an infinite number of primes of the form $x + k\prod_{i=1}^n p_i$. Let q be such a prime, and let $q > N + a_n$. Then $q \in A + B$.

If $q = a_i + b$ for some $i = 2, \cdots, n$ and $b \in B$, then $b > N$ and $b \in P$. But $q = a_i + b$ implies that $b \equiv o(p_i)$ which is impossible for sufficiently large q. If $q = b \in B$, then $q + a_2 \in P$. But $q + a_2 \equiv 0(p_1)$ which is also impossible for sufficiently large q.

Hence $(A) = (B) = \infty$.

BIBLIOGRAPHY

1. F. J. Dyson, *A theorem on the densities of sets of integers*, J. London Math. Soc. Vol. XX, **77** (1945), 8-14.

2. T. A. Evans and H. B. Mann, *On simple difference Sets*, Sankhya: *The Indian Journal of Statistics*, Vol. 11, parts 3 and 4, (1951), 357-364.

3. Marshall, Jr. Hall, *A survey of difference sets*, Proc. Amer. Math. Soc., **7** (1956), 975-986.

4. Martin Kneser, *Abschaetzung der asymptotischen Dichte von Summen mengen*, Math. Zeitschrift Bd. **58** (1953), 459-484.

5. H. B. Mann, *A proof of the fundamental theorem on the density of sums of sets of positive integers*, Ann. of Math., **43**, No. 3, July, (1942), 523-527.

6. ———, *On products of sets of group elements*, Canadian J. Math., **4** (1952), 64-66.

7. ———, *A refinement of the fundamental theorem on the density of the sum of two sets of integers*, Pacific J. Math., **10** (1960), 909-915.

WESTERN WASHINGTON STATE COLLEGE
THE OHIO STATE UNIVERSITY

ALGEBRAIC EXTENSIONS OF COMMUTATIVE BANACH ALGEBRAS

JOHN A. LINDBERG, JR.

1. Introduction. Let A denote a commutative normed algebra with multiplicative unit and norm $\|\cdot\|$. In [2], Arens and Hoffman showed that it is possible to norm $A[x]/(\alpha(x))$, where $\alpha(x) = \sum_{i=0}^{n} \alpha_i x^i$ is a monic polynomial over A, in such a way that the canonical mapping of A into $A[x]/(\alpha(x))$ is an isometry as well as an isomorphism; in fact, they give a family of norms on $A[x]/(\alpha())$, all of which are equivalent. Specifically, let t be a positive number which satisfies $t^n \geqq \|\alpha_0\| + \|\alpha_1\| t + \cdots + \|\alpha_{n-1}\| t^{n-1}$. Let $\sum_{i=0}^{n-1} a_i x^i + (\alpha(x))$ be any coset in $A[x]/(\alpha(x))$. As is well known, $\sum_{i=0}^{n-1} a_i x^i$ is the unique representative of this coset of lowest degree. Thus, $\| \sum_{i=0}^{n-1} a_i x^i + (\alpha(x)) \| = \sum_{i=0}^{n-1} \|a_i\| t^i$ is well defined and makes $A[x]/(\alpha(x))$ into a normed algebra. Clearly, $a \rightarrow a + (\alpha(x))$, $a \in A$, is an isometry of A into $A[x]/(\alpha(x))$. (Unless otherwise stated, we assume without loss of generality that $t = 1$.) From the form of the norm we see that $A[x]/(\alpha(x))$ is a Banach algebra under this norm precisely when A is a Banach algebra under $\|\cdot\|$. In the present paper, we deal mainly with the case where A is a Banach algebra. In section nine we deal with, at some length, more general algebras.

In this paper we are mainly interested in the algebraic aspects of the extension $B = A[x]/(\alpha(x))$. However, we also present results which are Banach algebraic in nature. For example in section three we give a complete description of the Šilov boundary of B. Section four is devoted to the study of the inheritance by B of the Banach algebra properties of regularity and self-adjointness. In particular, we show that if A is regular then B is also regular. Self-adjointness is not always inherited as Example 4.3 shows. A sufficient condition (which is satisfied, for example, when the discriminant of $\alpha(x)$ is invertible) is given under which this property is inherited. (This condition states that the set $S(\alpha(x), A)$ of singular points of $\alpha(x)$ is empty. This means that the natural mapping of the carrier space of B onto the carrier space of A is a local homeomorphism with respect to the weak* topologies. See section two for a complete discussion of this concept.)

In section five we once again make use of the condition that $\alpha(x)$ has no singular points. Theorem 5.2 states that if A is semi-simple

Received January 17, 1962, and in revised form October 23, 1963.

This paper is based on portions of the author's Doctoral dissertation which was prepared under the guidance of Professor B. R. Gelbaum, University of Minnesota. The author was supported, in part, by the Office of Ordnance Research, under contract SAR G DA ORD 12, while at Yale University.

and if $S(\alpha(x), A) = \phi$, then B decomposes into the direct sum of a closed subalgebra of the form $A[b]$, with $\alpha(b) = 0$, and the radical of B.

The next section is motivated by a well-known result in classical field theory. If A is a field and $\alpha(x)$ an irreducible polynomial, then any root $b \in B$ of $\alpha(x) = 0$ gives rise to an automorphism $(\sum_{i=0}^{n-1} a_i x^i + (\alpha(x)) \rightarrow \sum_{i=0}^{n-1} a_i b^i)$ of B which leaves invariant each element of A. In the present context this is no longer generally true. However, we are able to give two sets of conditions which assure us of this conclusion. Theorem 6.1 states that if $A[b]$ is dense in B, then $\sum_{i=0}^{n-1} a_i x^i + (\alpha(x)) \rightarrow \sum_{i=0}^{n-1} a_i b^i$ is an automorphism. Theorem 6.2 requires that the discriminant d of $\alpha(x)$ satisfy the condition that $da \in Rad(A)$ imply $a \in Rad(A)$ ($Rad(A)$ denotes the radical of A) and that the Gelfand transform of b satisfy a certain separation property. Also in section six we give conditions under which the automorphisms of B which leave each element of A invariant are periodic. The period is shown to be a factor of $n!$, $n =$ degree of $\alpha(x)$ over A. Examples can be given which show that in the absence of any restrictions some of the automorphisms of B leaving invariant each element of A have infinite order.

In the next two sections we deal exclusively with polynomials over A which have invertible discriminants in A. Section seven is concerned with the problem of extending a ring isomorphism of A_1 onto A_2 to an isomorphism of $A_1[x]/(\alpha_1(x))$ onto $A_2[x]/(\alpha_2(x))$. A necessary and sufficient condition is given under which such an extension exists. The extension is not necessarily unique. Prior to establishing this theorem we characterize those elements $b \in B$ such that $B = A[b]$ (= algebra of polynomials in b with coefficients in A). Attention is given to extending involutions on A to involutions of B.

In section eight we show that repeated extensions are again simple algebraic extensions (algebraically and topologically) of the type under discussion in this paper.

In the last section we give a complete description of the radical of B. The major results are stated for algebras over fields of characteristic zero. The main theorem (9.2) states that if A is semi-simple, then the radical of B is a nilpotent ideal. The degree of nilpotency is also specified. As a corollary, we have that if B is semi-simple, then A is semi-simple and the discriminant of $\alpha(x)$ is not a zero divisor in A, or zero. Applying this to the case of a tractable normed algebra (intersection of the closed maximal ideals is (0)), we show that the radical of B and the intersection of the closed maximal ideals of B coincide.

We now proceed to section two which contains some preliminaries gathered from other sources.

2. Preliminaries.

If A is a Banach algebra (always assumed to be commutative and to possess a multiplicative unit e), then Φ_A (called the carrier space of A, [12]) is to denote the space of (non-trivial) multiplicative linear functionals on A to C (= complex numbers). If $(h, \lambda) \in \Phi_A \times C$, then (h, λ) can be considered as a multiplicative linear functional on $A[x]$, its action on elements $\Sigma a_i x^i \in A[x]$ being defined by $(h, \lambda)\Sigma a_i x^i = \Sigma h(a_i)\lambda^i$. In [2] it is shown that Φ_B, $B = A[x]/(\alpha(x))$ (throughout this paper, B will be used to denote $A[x]/(\alpha(x))$, $\alpha(x)$ monic), is (identifiable with) the set $\{(h, \lambda) \in \Phi_A \times C : (h, \lambda)\alpha(x) = 0\}$. It should be noted that if $(h, \lambda) \in \Phi_B$, then $|\lambda| \leqq 1$ (recall that we are assuming $\sum_{i=0}^{n-1} \|\alpha_i\| \leqq 1$ so that $\|x + (\alpha(x))\| = 1$). The coset $a + (\alpha(x))$ will be denoted by a for $a \in A$ and $x + (\alpha(x))$ will be denoted by $[x]$.

x will be considered as an indeterminant over $\hat{A}$ (= Gelfand representation of A) and C as well as an indeterminant over A. If $\beta(x) = \Sigma\beta_i x^i \in A[x]$, then $\hat{\beta}(x)$ is to denote the polynomial $\Sigma\hat{\beta}_i x^i$ over $\hat{A}$ and $\beta_h(x)$ is to denote the polynomial $\Sigma\hat{\beta}_i(h)x^i$ over C. If $\beta(x) \in A[x]$ and $\beta_h(\lambda) = 0$, $\lambda \in C$, but $\beta_h(x)$ not the zero polynomial, then we denote the multiplicity of λ as a root of $\beta_h(x) = 0$ by $M_\beta(h, \lambda)$. We call M_β the multiplicity function of $\beta(x)$.

We include for the convenience of the reader several results that we will need from other sources.

2.1. π defined by $\pi(h, \lambda) = h$, $(h, \lambda) \in \Phi_B$, is an open continuous mapping of Φ_B onto Φ_A.

2.2. For each $h \in \Phi_A$ there are disjoint neighborhoods $V_1, \cdots, V_m$ in Φ_B of the points in $\pi^{-1}(h) = \{(h, \lambda_1), \cdots, (h, \lambda_m)\}$ such that $\pi(V_1) = \pi(V_i)$, $i = 2, \cdots, m$, and $\pi^{-1}(\pi(V_1)) = \bigcup_{i=1}^m V_i$.

2.3. M_α is locally constant at $(h, \lambda) \in \Phi_B$ if and only if π is a local homeomorphism at (h, λ).

2.4. (Arens and Calderón) If $\beta(x) \in A[x]$ (not necessarily monic) and if $f \in C(\Phi_A)$ such that $\hat{\beta}(f) = 0$ but $\hat{\beta}'(f)$ never vanishes on Φ_A ($\beta'(x)$ is the formal derivative of $\beta(x)$), then a unique element $b \in A$ exists such that $\beta(b) = 0$ and $\hat{b} = f$. (Arens and Calderón did not assert the uniqueness of b. However, it is easily established. Write $\beta(x) = (x - b)Q(x)$, $Q(x) \in A[x]$ and suppose $b' \in A$, $\beta(b') = 0$ and $\hat{b}' = \hat{b}$. Then $(b' - b)Q(b') = 0$. Since $f(h)$ is a simple root of $\beta_h(x) = 0$, $Q(b)\hat{}(h) \neq 0$ for every $h \in \Phi_A$, so that $Q(b)$ is invertible in A. Hence $b = b'$.)

Related to the above is

2.5. If $\alpha(x) \in A[x]$ is a monic polynomial, if $f \in C(\Phi_A)$ such that $\hat{\alpha}(f) = 0$ and if $M_\alpha(\cdot, f(\cdot))$ is locally constant on Φ_A, then $f \in \hat{A}$. (A stronger conclusion similar to the above can not be drawn here.)

2.1, 2.2, 2.3 and 2.5 are proved in [10] while 2.4 is proved in [1].

Let $\alpha(x) \in A[x]$ be monic. If $h \in \Phi_A$ is such that each point of $\pi^{-1}(h)$ possesses a neighborhood on which M_α is constant, or what is equivalent (in view of 2.3), π is a local homeomorphism at each point of $\pi^{-1}(h)$, then we call h an *ordinary point of* $\alpha(x)$. If $h \in \Phi_A$ is not an ordinary point of $\alpha(x)$, we say that it is a *singular point of* $\alpha(x)$ and the set of such points will be denoted by $S(\alpha(x), A)$. It is clear that if $h \in \Phi_A$ is an ordinary point of $\alpha(x)$, then each h' sufficiently close to h is also an ordinary point of $\alpha(x)$ so that $S(\alpha(x), A)$ must be a closed subset of Φ_A. $S(\alpha(x), A)$ is a subset of the set D where $\hat{d}$ vanishes, where d is the discriminant of $\alpha(x)$ (cf. [2] and page 93, [14]). (Note that $\hat{d}(h)$ is the discriminant of $\alpha_h(x)$.) $S(\alpha(x), A)$ can be null even if D is not null. On the other hand, $S(\alpha(x), A)$ can be all of D. Because the cardinality of the sets $\pi^{-1}(h)$ is uniformly bounded by n (= degree of $\alpha(x)$), $S(\alpha(x), A)$ is easily shown to be nowhere dense in Φ_A.

3. **The Šilov Boundary of $A[x]/(\alpha(x))$.** Let A' be a Banach algebra extension of the Banach algebra A, let $\partial A, \partial A'$ denote respectively the Šilov boundaries of A and A', and π the natural mapping of $\Phi_{A'}$ into Φ_A defined by $h = \pi(h') = h' \mid A, h' \in \Phi_A$. Then it is well known that $\pi(\partial A') \supset \partial A$. If A' is the extension $B = A[x]/(\alpha(x))$, then this result can be sharpened; indeed, we have that $\partial B = \pi^{-1}(\partial A)$. In the proof of this assertion, we need (Theorem 5, Appendix IV, [5]): A necessary and sufficient condition that $h_0 \in \partial A$ is that for each neighborhood V in Φ_A of h_0 there is a function $f \in \hat{A}$ whose absolute value $|f|$ attains its maximum (which we may assume is 1) on V and is less than that on $\Phi_A \sim V$.

THEOREM 3.1. $\partial B = \pi^{-1}(\partial A)$.

Proof. We first show that $\pi^{-1}(\partial A) \subset \partial B$. Let $h_0 \in \partial A$, let W_0 be a neighborhood in Φ_B of $(h_0, \lambda_0^{(1)})$, and let $g \in B^\wedge$ such that $g(h_0, \lambda^{(1)}) = 1$ and zero at the other points $(h_0, \lambda_0^{(i)})$ of the fiber $\pi^{-1}(h_0)$. Let $W_1 \subset W_0$ be an open neighborhood in Φ_B of $(h_0, \lambda_0^{(1)})$ such that $|g(h, \lambda)| > 1/2$ if $(h, \lambda) \in W_1$ and W_i an open neighborhood in Φ_B of $(h_0, \lambda_0^{(i)})$, $i \neq 1$, such that $|g(h, \lambda)| < 1/2$ if $(h, \lambda) \in W_i$. Since π is an open mapping, $V_0 = \bigcap_i \pi(W_i)$ is an open neighborhood in Φ_A of h_0. Let $V_i = W_i \bigcap \pi^{-1}(V_0)$. Now, by the theorem quoted above, there is a function $f \in \hat{A}$ such that $\|f\|_\infty = 1$, $|f(h_1)| = 1$, $h_1 \in V_0$, and $|f(h)| < 1$ if $h \in \Phi_A \sim V_0$. Since $\Phi_A \sim V_0$ is closed, it is compact and hence there is a positive integer N so large that

$$|f(h)|^N \leqq \frac{1}{2\|g\|_\infty} \quad \text{for } h \in \Phi_A \sim V_0 .$$

Then, if $h \notin V_0$ and $(h, \lambda) \in \Phi_B$, we have that

$$|(f^N g)(h, \lambda)| = |f(h)|^N |g(h, \lambda)| \leqq \frac{1}{2\|g\|_\infty} \|g\|_\infty = \frac{1}{2},$$

and if $(h, \lambda) \notin V_1$ but $h \in V_0$, we have that

$$|(f^N g)(h, \lambda)| < |g(h, \lambda)| \leqq \frac{1}{2}.$$

But for $(h_1, \lambda) \in V_1$,

$$|(f^N g)(h_1, \lambda)| = |g(h_1, \lambda)| > \frac{1}{2}.$$

Thus, $|f^N g|$ assumes its maximum value on V_1, and hence, on W, and is less than that outside V_1 or outside of W. By the above quoted theorem, $(h_0, \lambda_0^{(1)}) \in \partial B$, and $\pi^{-1}(\partial A) \subset \partial B$. We next show the reverse inclusion.

Let $(h_0, \lambda_0) \in \partial B$, and let V be any neighborhood in Φ_A of h_0. Let W be an open neighborhood in Φ_B of (h_0, λ_0) such that $\pi(W) \subset V$ and no $(h_0, \lambda_0') \neq (h_0, \lambda_0)$ lies in W. Let $g \in \hat{B}$ be a function such that $\|g\|_\infty$ is assumed by $|g|$ on W and $|g| < \|g\|_\infty$ outside of W. As in the above paragraph, we may assume that $|g(h, \lambda)| < 1/2n$ if $(h, \lambda) \in \Phi_B \sim W$. Let f be the function defined by

$$f(h) = \sum_{i=1}^{n} g(h, \lambda_i(h))$$

where the $\lambda_i(h)$ denote all the roots (each distinct root repeated according to its multiplicity) of $\alpha_h(x) = 0$. Then $f \in \hat{A}$. Now, for $h \notin \pi(W)$

$$|f(h)| = \left|\sum_{i=1}^{n} g(h, \lambda_i)\right| < \sum_{i=1}^{n} |g(h, \lambda_i)| < \frac{1}{2}.$$

There exists $(h_1, \lambda_1) \in W$ such that $|g(h_1, \lambda_1)| = \|g\|_\infty$. (Assume that $\lambda_1 = \lambda_1(h_1)$.) Then

$$\begin{aligned} |f(h_1)| &= \left|g(h_1, \lambda_1) + \sum_{i=2}^{n} g(h_1, \lambda_i(h))\right| \\ &> |g(h_1, \lambda_1)| - \sum_{i=2}^{n} |g(h_1, \lambda_i(h))| > 1 - \frac{n-1}{2n} > \frac{1}{2}. \end{aligned}$$

Thus, $h_0 \in \partial A$ and $\pi(\partial B) \subset \partial A$. Using the fact that $\pi^{-1}(\partial A) \subset \partial B$, we have that $\pi^{-1}(\partial A) \supset \partial B$. This completes the proof of the theorem.

4. Inheritance of the properties of regularity and self-adjointness.

The properties of regularity and self-adjointness are possessed by many important and interesting Banach algebras and hence it is of interest to know whether these properties are inherited by the extension B. G. A. Heuer in [6] has shown that if A is regular and self-adjoint and if in addition the discriminant of the monic polynomial $\alpha(x)$ is invertible in A, then B is both regular and self-adjoint. In this section, we show that regularity is always inherited (without the assumption of self-adjointness). As a corollary we show that if $\hat{A}$ is dense in $C(\Phi_A)$, then $\hat{B}$ is dense in $C(\Phi_B)$. (For a discussion of the inheritance by $\hat{B}$ of the sup norm completeness of $\hat{A}$, the reader is referred to [7].) Example 4.3 shows that the self-adjointness of A is not always inherited by B. We finally show that if $S(\alpha(x), A) = \phi$, then self-adjointness is inherited.

THEOREM 4.1. *Let A be a regular Banach algebra and let $\alpha(x)$ be a monic polynomial over A. Then B is regular.*

Proof. It suffices to show that if given $(h_0, \lambda_0) \in \Phi_B$ and a neighborhood W in Φ_B of (h_0, λ_0), then there exists a function $\hat{b} \in \hat{B}$ such that $\hat{b}(h_0, \lambda_0) = 1$ and $\hat{b}(h, \lambda) = 0$ if $(h, \lambda) \in \Phi_B \sim W$. From 2.2, it follows that there is a neighborhood V in Φ_A of h_0 so small that $V \subset \pi(W)$ and $\pi^{-1}(V) = \bigcup_{i=1}^m V_i$ where the V_i are disjoint neighborhoods of the points in $\pi^{-1}(h_0)$ with $W \supset V_1$. We assume (without loss of generality) that the sets $V_1, \cdots, V_m$ are closed. Since A is regular the set V is hull-kernel closed in Φ_A, from which it follows that $\pi^{-1}(V)$ is hull-kernel closed in Φ_B. Now, let I denote the ideal in B of elements whose transforms in $\hat{B}$ vanish on $\pi^{-1}(V)$. Since I is a closed ideal, B/I is a Banach algebra with carrier space (identifiable with) $\pi^{-1}(V)$ (cf. [11]). By [13], there is an idempotent f in B/I such that $\hat{f}(h, \lambda) = 1$ if and only if $(h, \lambda) \in V_1$. But $\hat{f} = \hat{f}_0 \mid \pi^{-1}(V)$ for some $f_0 \in B$. Since A is regular there is an element $a \in A$ such that $\hat{a}(h_0) = 1$ and $\hat{a}$ vanishes outside of V. Then $b = af_0$ is an element of B such that $\hat{b}(h_0, \lambda_0) = 1$ and $\hat{b}(h, \lambda) = 0$ outside of $V_1 \subset W$. This completes the proof of the theorem.

The corollary below extends the following result of Heuer [6]: If $\hat{A}$ is dense in $C(\Phi_A)$ and if for each singular point h, $\pi^{-1}(h)$ consists of exactly one point, then $\hat{B}$ is dense in $C(\Phi_B)$. The proof given below is essentially due to Heuer.

COROLLARY 4.2. *If A is a Banach algebra and if $\hat{A}$ is dense in $C(\Phi_A)$, then $\hat{B}$ is dense in $C(\Phi_B)$.*

Proof. Since $\hat{A}$ is dense in $C(\Phi_A)$, it is easily shown that $\hat{B} = (\hat{A}[x]/(\hat{\alpha}(x)))^\wedge$ is dense in $B_0 = (C(\Phi_A)[x]/(\hat{\alpha}(x)))^\wedge$, with both algebras being viewed as subalgebras of $C(\Phi_B)$. Thus, it suffices to show that B_0 is dense in $C(\Phi_B)$. (It need not be the case that B_0 is all of $C(\Phi_B)$

as Example 4.3 of this section shows.) Let $h \in \Phi_A$ be arbitrarily given. By the theorem, B_0 is regular so that if $V_1, \cdots, V_m$ are disjoint neighborhoods of the points in $\pi^{-1}(h)$, then there exists a function $f \in B_0$ which takes the value i on V_i, $i = 1, 2, \cdots, m$. Let g be a real-valued function in $C(\Phi_A)$ such that $g(h) = 1$ and g vanishes outside of $\bigcap_{i=1}^{m} \pi(V_i)$. Then $(gf)^\wedge$ is a real-valued function in B_0 which separates the points of $\pi^{-1}(h)$. Since $C(\Phi_A)$ is (isomorphic to) a subalgebra of B_0, any two points $(h, \lambda), (h', \lambda') \in \Phi_B$, with $h \neq h'$, can be separated by a real-valued function in B_0. Hence any two points in Φ_B can be separated by a real-valued function in B_0. The conclusion of the corollary now follows from the Stone-Weierstrass Theorem.

We now turn our attention to the question of inheritance of the property of self-adjointness, and first give an example which shows that this property is not always inherited by the extension.

EXAMPLE 4.3. Let $A = C(\Delta)$, $\Delta = \{z \in C\colon |z| \leqq 1\}$ and $\alpha(x) = x^2 - f_0$, $f_0(z) \equiv z$. Then $A[x]/(\alpha(x))$ is not self-adjoint. For if it were, then $([x]^\wedge)^- = \hat{a}_0 + \hat{a}_1[x]^\wedge$ for some choice of $a_0, a_1 \in A$. But this means that $\hat{a}_1(z) = \exp(-\arg z)$, $z \neq 0$. This is a contradiction since $\exp(-\arg z)$ is not extendable to a continuous function on Δ.

THEOREM 4.4. *Let A be a self-adjoint Banach algebra and $\alpha(x) = \sum_{i=0}^{n} \alpha_i x^i$ be a monic polynomial over A. If $S(\alpha(x), A) = \phi$, then $A[x]/(\alpha(x))$ is self-adjoint.*

Proof. Let $f(h, \lambda) \equiv \bar{\lambda}$ for $(h, \lambda) \in \Phi_B$. Then $f \in C(\Phi_B)$ and $\hat{\beta}(f) = 0$, where $\beta(x) = \sum_{i=0}^{n} \beta_i x^i$, $\hat{\beta}_i = (\hat{\alpha}_i)^-$, $i = 0, 1, \cdots, n-1$, and $\beta_n = e$. Since the multiplicity function M_α of $\alpha(x)$ is locally constant on Φ_B, it follows that $M_\beta(\cdot, f(\cdot))$ is locally constant on Φ_B, where M_β is the multiplicity function of $\beta(x)$ when viewed as a polynomial over B. By 2.5, it follows that $f \in \hat{B}$ so that B is self-adjoint since $(\sum \hat{a}_i([x]^\wedge)^i)^- = \sum (\hat{a}_i)^- f^i \in \hat{B}$.

5. On the Wedderburn decomposition of B. In this section we discuss the Wedderburn decomposition of the extension B, that is, the decomposition of B into the direct sum of a closed subalgebra B_0 of B and the radical Rad(B) of B ($B = B_0 \oplus \text{Rad}(B)$). As is well known, such a decomposition in general does not hold for Banach algebras, even in the weaker sense where one does not require that the subalgebra be closed. We will give an example which supports this statement. Badé and Curtis have given an example in [3]. Feldman, in [4], gave an example where the stronger Wedderburn decomposition failed to hold. For this example, the weaker decomposition holds.

The condition that $S(\alpha(x), A) = \phi$ (A semi-simple) is sufficient for such a decomposition of B to hold. When this condition holds, $\alpha(x)$ is forced to factor; precisely, there exist mutually orthogonal idempotents $e_1, \cdots, e_m$, positive integers k_{ij}, and polynomials $\alpha_{ij}(x) \in A[x]$, $j = 1, \cdots, S_i$; $i = 1, \cdots, m$, such that $e_i\alpha_{ij}(x)$ is monic over e_iA, the discriminant of $\prod_{j=1}^{S_i} e_i\alpha_{ij}(x)$ is invertible in e_iA, and

5.1 $$\alpha(x) = \sum_{i=1}^{m} e_i \prod_{j=1}^{S_i} \alpha_{ij}(x)^{k_{ij}} .$$

Furthermore, the radical of $A[x]/(\alpha(x))$ is a principal ideal generated by $\beta([x])$, where $\beta(x) = \sum_{i=1}^{m} e_i \prod_{j=1}^{S_i} \alpha_{ij}(x)$ (cf. Theorem 2.3, [10]).

THEOREM 5.2. *Let A be a semi-simple Banach algebra and $\alpha(x)$ a monic polynomial over A. If $S(\alpha(x), A) = \phi$, then there exists an element $b \in B$ such that $\alpha(b) = 0$, $A[b]$ is closed in B and $B = A[b] \oplus$ Rad(B).*

Proof. To simplify the proof, we first assume that $m = 1$ in the above paragraph. Thus, $\alpha(x)$ is of the form $\prod_{i=1}^{S} \alpha_i(x)^{k_i}$, where each $\alpha_i(x)$ is monic over A and $\beta(x) = \prod_{i=1}^{S} \alpha_i(x)$ has an invertible discriminant in A.

Since $\beta([x]^\wedge) = 0$ and since $\beta(x)$ has an invertible discriminant in A, and hence in B, there exists an element $b \in B$ such that $\beta(b) = 0$ and $\hat{b} = [x]^\wedge$. Thus, $\alpha(b) = 0$ also. Since $\hat{b} = [x]^\wedge$, there is an element $R \in \text{Rad}(B)$ such that $[x] = b + R$ so that $\sum_{i=0}^{n-1} a_i[x]^i = \sum_{i=0}^{n-1} a_ib^i +$ (polynomial in R, with zero constant term) ($n =$ degree of $\alpha(x)$). Hence, B is the sum of $A[b]$ and Rad(B). We next show that the sum is a direct sum. Let t be the degree of $\beta(x)$ over A. Then $\sum_{i=0}^{n-1} a_i'b^i$ can be expressed in the form $\sum_{i=0}^{t-1} a_ib^i$ for some choice of $a_0, \cdots, a_{t-1}$ in A. Suppose now that $\sum_{i=0}^{t-1} a_ib^i \in \text{Rad}(B)$. Then $\sum_{i=0}^{t-1} a_ix^i$ is a multiple of $\beta(x)$ (this follows since the radical of B is a principal ideal generated by $\beta([x])$). Thus, the a_i's must all be 0. Thus, the sum is direct. (Note also that $\sum_{i=0}^{t-1} a_ib^i = 0$ if and only if $a_i = 0$, $i = 0, 1, \cdots, t-1$.)

In order to show that $A[b]$ is closed, we introduce a mapping ϕ of B onto $A[y]/(\beta(y))$ as follows: $\phi(\Sigma a_i[x]^i) = \Sigma a_i[y]^i$. ϕ is well defined and a homomorphism since $\alpha([y]) = 0$. Furthermore, ϕ is continuous since

$$\left|\left|\phi\left(\sum_{i=0}^{n-1} a_i[x]^i\right)\right|\right| = \left|\left|\sum_{i=0}^{n-1} a_i[y]^i\right|\right| \leqq \sum_{i=0}^{n-1} ||a_i|| \, ||[y]||^i$$
$$\leqq K\sum_{i=0}^{n-1} ||a_i|| = K\left|\left|\sum_{i=0}^{n-1} a_i[x]^i\right|\right| ,$$

where $K = \max\{1, ||[y]||, \cdots, ||[y]||^{n-1}\}$. Since Rad$(B)$ is generated by $\beta([x])$, $\phi(\text{Rad}(B)) = 0$. But $[x] - b \in \text{Rad}(B)$ so that $\phi(b) = \phi([x]) = [y]$.

Thus, if $\sum_{i=0}^{t-1} \|\beta_i\| k^i \leqq k^t$, where $\beta(y) = \sum_{i=0}^{t} \beta_i x^i$, then

$$\sum_{i=0}^{t-1} \|a_i\| k^i = \left\| \sum_{i=0}^{t-1} a_i [y]^i \right\| \leqq K \left\| \sum_{i=0}^{t-1} a_i b^i \right\| \leqq K' \sum_{i=0}^{t-1} \|a_i\|$$

where $K' = K \max\{1, \|b\|, \cdots, \|b\|^{t-1}\}$. Since A is complete, the norm on B restricted to $A[b]$ is complete or equivalently, $A[b]$ is closed in B. This completes the proof of the theorem if we assume that $m = 1$.

The general situation follows immediately from what was proved above and the following observations. Let $e_1, e_2, \cdots, e_m$ be the idempotents which appear in the factorization of $\alpha(x)$ which was displayed in 5.1. Then $A = e_1 A \oplus \cdots \oplus e_m A$ and $B = e_1 B \oplus \cdots \oplus e_m B$, the direct sums being topological. Since the natural isomorphism ϕ_i of $e_i B$ onto $B_i = (e_i A)[x]/(e_i \alpha(x))$ is *bi*-continuous and since $\mathrm{Rad}(B_i) = \phi_i(e_i \, \mathrm{Rad}(B))$, it follows from the above that there exists $b_i \in e_i B$ such that $e_i \alpha(b_i) = 0$, $(e_i A)[b_i]$ is closed in $e_i B$ and $e_i B = (e_i A)[b_i] \oplus e_i(\mathrm{Rad}\, B)$. If we set $b = \sum_{i=1}^{m} b_i$, then $\alpha(b) = 0$, $A[b]$ is closed in B and $B = A[b] \oplus \mathrm{Rad}(B)$. This completes the proof of the theorem.

We now present an example that shows if we drop the condition that $S(\alpha(x), A) = \phi$, then the conclusion of Theorem 5.2 is not assured.

EXAMPLE 5.3. Take A to be the algebra of functions f which are continuous on the disc $\Delta = \{z \in C: |z| \leqq 1\}$, analytic in the interior of Δ and $f'(0) = 0$. For $\alpha(x)$, take $(x - f_0)^2 (x + 2f_0)$ where $f_0(z) \equiv z$, $z \in \Delta$. Then $\alpha(x) \in A[x]$ and $S(\alpha(x), A) = \{0\}$. (Φ_A is identifiable with Δ.) Now, there is no subalgebra B_0 of B isomorphic to $B^\frown$. (If $B = B_0 \oplus \mathrm{Rad}(B)$, then $B_0 \cong \hat{B} \cong B/\mathrm{Rad}(B)$.) For if this were the case, then B_0 would coincide with $A[b]$ for some $b \in B$ and b would have to satisfy $f_0^2(b - f_0)(b + 2f_0) = 0$. This is easily shown to be impossible. It follows from Theorem 9.2 that the degree of nilpotency of $\mathrm{Rad}(B)$ is two.

6. Automorphisms and conjugate roots. If $g: A[x]/(\alpha(x)) \to A[x]/(\alpha(x))$ is an automorphism such that $g(a) = a$ for all $a \in A$, then $g([x])$ is obviously a root of $\alpha(x) = 0$ and $A[g([x])] = A[x]/(\alpha(x))$. Conversely, if $\alpha(b) = 0$, $b \in A[x]/(\alpha(x))$, need the homomorphism $g: \Sigma a_i [x]^i \to \Sigma a_i b^i$ be an automorphism of B? The answer is no in general (recall Theorem 5.2). However, there are various conditions (see 6.1 and 6.3) under which such homomorphisms g are automorphisms. In 6.4 we give conditions under which automorphisms of the above type are periodic. We begin with

THEOREM 6.1. *Let $\alpha(x)$ be a monic polynomial over the Banach algebra A. If $b \in B$ such that $A[b]$ is dense in B and $\alpha(b) = 0$, then $g: \sum_{i=0}^{n-1} a_i [x]^i \to \sum_{i=0}^{n-1} a_i b^i$ is an automorphism.*

Proof. What we actually prove is this: if T is a linear transformation of $A^n = A \times \cdots \times A$ onto a dense subset of A^n such that $a \cdot T(a_1, \cdots, a_n) = T(aa_1, \cdots, aa_n)$, then T is one-to-one and onto. For a norm in A^n we take $\|(a_1, \cdots, a_n)\| = \sum_{i=1}^n \|a_i\|$. (Clearly the homomorphism g has these properties; note that as a Banach space $B = A^n$.)

Let $h \in \Phi_A$ and let T_h denote the mapping $T_h: (h(a_1), \cdots, h(a_n)) \to (h(a_1'), \cdots, h(a_n'))$ where $(a_1', \cdots, a_n') = T(a_1, \cdots, a_n)$. Clearly, T_h is a linear transformation of C^n into itself since $C \cong A/h^{-1}(0)$. (For a norm in C^n, we take $|(\lambda_1, \cdots, \lambda_n)| = \sum_{i=1}^n |\lambda_i|$.) Now, $T_h(C^n)$ must be dense in C^n. For if $(\lambda_1, \cdots, \lambda_n), (\mu_1, \cdots, \mu_n) \in C^n$, then there are elements $a_i, b_i \in A$ such that $h(a_i) = \lambda_i$ and $h(b_i) = \mu_i$, $i = 1, \cdots, n$. If $(a_1', \cdots, a_n') = T(a_i, \cdots, a_n)$, then

$$\begin{aligned} &|T_h(\lambda_1, \cdots, \lambda_n) - (\mu_1, \cdots, \mu_n)| \\ &\quad = |(h(a_1'), \cdots, h(a_n')) - (h(b_1), \cdots, h(b_n))| \\ &\quad = \sum_{i=1}^n |h(a_i') - h(b_i)| \leqq \sum_{i=1}^n \|a_i' - b_i\| \\ &\quad = \|T(a_1, \cdots, a_n) - (b_1, \cdots, b_n)\| . \end{aligned}$$

It follows from the above that $T_h(C^n)$ is dense in C^n. But this means that T_h is one-to-one and hence onto.

Now, consider n-linear equations in a_i (considered to be unknowns) represented by

$$(*) \qquad \sum_{i=1}^n a_i T(e_i) = (b_1, \cdots, b_n) ,$$

where e_i is the vector in A^n with e in the ith place and zero elsewhere. If D is the determinant of the matrix of the coefficients of system $(*)$, then $h(D)$ is precisely the determinant of the matrix associated with the linear transformation T_h. Since T_h is onto, $h(D) \neq 0$.

Since $h \in \Phi_A$ in the above argument is quite arbitrary, $h(D) \neq 0$ for all $h \in \Phi_A$ so that D is invertible in A. But this means that $(*)$ has a unique solution $(a_1, a_2, \cdots, a_n) \in A$ for each $(b_1, \cdots, b_n) \in A^n$. Hence T is both one-to-one and onto.

Let $G(B:A)$ denote the group of automorphisms of B which leave invariant each element of A. If $g \in G(B:A)$, let g^* denote the homeomorphism of Φ_B onto itself which satisfies $g(b)^\wedge(h, \lambda) = \hat{b}(g^*(h, \lambda))$ for all $b \in B$ and all $(h, \lambda) \in \Phi_B$ (cf. [11]). $E(\Phi_B : \Phi_A)$ is to denote the group of homeomorphisms ϕ of Φ_B onto itself such that $\pi \circ \phi = \pi$.

LEMMA 6.2. *If* $g \in G(B:A)$, *then* $g^*(h, \lambda) = (h, g([x])^\wedge(h, \lambda))$ *for every* $(h, \lambda) \in \Phi_B$ *and consequently* $g^* \in E(\Phi_B : \Phi_A)$. *Also,* $(g^*)^{n!} =$ *identity homeomorphism* ($n =$ *degree of* $\alpha(x)$).

Proof. By the definition of g^*, we know that for $a \in A$ and

$(h, \lambda) \in \Phi_B$, $\hat{a}(h) = \hat{a}(h, \lambda) = g(a)\hat{\ }(h, \lambda) = \hat{a}(g^*(h, \lambda)) = \hat{a}(h')$, where $(h', \lambda') = g^*(h, \lambda)$. Since $\hat{A}$ is a separating algebra of functions on Φ_A, it follows that $h = h'$. Thus, $g^*(h, \lambda) = (h, \lambda')$ or equivalently, $\pi \circ g^* = \pi$. The last assertion of the lemma follows from the fact that if $\phi \in E(\Phi_B : \Phi_A)$, then $\phi \mid \pi^{-1}(h)$ is a permutation of $\pi^{-1}(h)$ so that $\phi^{n!}$ must be the identity homeomorphism on Φ_B.

THEOREM 6.3. *Let $\alpha(x)$ be a monic polynomial over the Banach algebra A. If the discriminant d of $\alpha(x)$ has the property that $da \in \operatorname{Rad}(A)$ implies that $a \in \operatorname{Rad}(A)$ and if $b \in B, \alpha(b) = 0$ and $\hat{b}$ separates the points of $\pi^{-1}(h)$ for each $h \in \Phi_A$, then $g\colon \sum_{i=0}^{n-1} a_i[x]^i \to \sum_{i=0}^{n-1} a_i b^i$ is an automorphism.*

Proof. Corresponding to the homomorphism $g\colon \sum_{i=0}^{n-1} a_i[x]^i \to \sum_{i=0}^{n-1} a_i b^i$, let ϕ denote the mapping $\phi(h, \lambda) = (h, g([x])\hat{\ }(h, \lambda))$. Since $\hat{b} = g([x])\hat{\ }$ separates the points of each fiber $\pi^{-1}(h)$, ϕ is one-to-one and onto. Hence, $\phi \in E(\Phi_B : \Phi_A)$. For each i, it is easily shown that $\phi^i(h, \lambda) = (h, (g^i([x]))\hat{\ }(h, \lambda))$ for each $(h, \lambda) \in \Phi_B$. Thus, we have that $\phi^{n!}$ is the identity homeomorphism on Φ_B. It now follows that $g^{n!}([x])\hat{\ } = [x]\hat{\ }$ or equivalently, $g^{n!}([x]) - [x] \in \operatorname{Rad}(B)$.

Let $T = g^{n!}$. Then $T([x]) - [x] \in \operatorname{Rad}(B)$. It further follows that for each $i = 0, \cdots, n-1$, $T([x]^i) - [x]^i \in \operatorname{Rad}(B)$. Since $da \in \operatorname{Rad}(A)$ implies that $a \in \operatorname{Rad}(A)$, where d is the discriminant of $\alpha(x)$, $\hat{d}$ is not a zero divisor in $\hat{A}$ and $\operatorname{Rad}(B) = (\operatorname{Rad}(A))[[x]]$ (cf. [2]). Thus, there exist elements $r_{ij} \in \operatorname{Rad}(A)$, $i, j = 0, \cdots, n-1$, such that $T([x]^i) = [x]^i + \sum_{j=0}^{n-1} r_{ij}[x]^j$. When T is viewed as a linear transformation on A^n, the determinant associated with T is invertible in A so that T is one-to-one and onto. But then g must also be one-to-one and onto. This completes the proof of the theorem.

COROLLARY 6.4. *Maintain the hypothesis (on d) of the theorem. If either*

(i) *$\operatorname{Rad}(A)$ is a nilpotent ideal and d is not a zero divisor in A, or*

(ii) *there exists $\mu > 0$ such that $\| dr \| \geqq \mu \| r \|$ for all $r \in \operatorname{Rad}(A)$, obtains, then each $g \in G(B : A)$ is periodic; in fact, if $(g^*)^p$ is the identity homeomorphism, then g^p is the identity automorphism of B.*

Proof. From the theorem we know that $g^p([x]) - [x] = R \in \operatorname{Rad}(B)$ if $(g^*)^p$ is the identity homeomorphism. We will show that if either (i) or (ii) obtains, then $R = 0$ so that $g^p([x]) = [x]$. If case (i) obtains, then $\operatorname{Rad}(B)$ is a nilpotent ideal (by Corollary 9.4). If we write $\alpha(y) = (y - [x])Q(y)$, $Q(y) \in B[y]$, then $R \cdot Q([x] + R) = 0$. Now there are elements $b_i \in B$, $i = 1, \cdots, n-1$, such that $Q([x] + R) = \alpha'([x]) + \sum_{i=1}^{n-1} b_i R^i$

(by direct computation). If $R^m \neq 0$ but $R^{m+1} = 0$, then $R^m \alpha'([x]) = 0$. If we write $\alpha(y)s(y) + \alpha'(y)t(y) = d$, $s(y), t(y) \in A[y]$ (cf. formula 4, page 96, [14]), then $\alpha'([x])t([x]) = d$. Thus, $R^m \cdot d = 0$; hence $R^m = 0$. This is a contradiction so that $R = 0$.

Suppose next that case (ii) obtains. We first show that d is not a zero divisor in A. For if $da = 0$, then we know that $a \in \mathrm{Rad}(A)$. But $0 = \| da \| \geqq \mu \| a \|$ and hence $a = 0$. Now, as in the above, we have that $R \cdot Q([x] + R) = 0$ or $R \cdot \alpha'([x]) = \sum_{i=1}^{n-1} b_i R^{i+1}$ for some choice of b_i, $i = 1, \cdots, n-1$, in B. Thus, $R \cdot \alpha'([x])t([x]) = R \cdot d = t([x]) \sum_{i=1}^{n-1} R^{i+1}$, $t([x])$ as above. If $R \cdot d = 0$, then $R = 0$. Suppose therefore that $R \neq 0$. Then it follows that $R^k \neq 0$ for all k. For if $R^k = 0$, then $R^{k-1}d = 0$ and hence $R^{k-1} = 0$. Now

$$\| (Rd)^k \|^{1/k} \leqq K \| R^k \|^{2/k}$$

where $K = \| t([x]) \cdot \sum_{i=1}^{n-1} b_i R^{i-1} \| \neq 0$. For each integer k, we have that $\| (Rd)^k \|^{1/k} \geqq \mu \| R^k \|^{1/k}$. For if $R^k = \sum_{i=0}^{n-1} r_i^{(k)} [x]^i$, $r_i^{(k)} \in \mathrm{Rad}(A)$ (recall that $\mathrm{Rad}(B) = (\mathrm{Rad}(A))\,[[x]]$, then

$$\| (Rd)^k \|^{1/k} = \left(\sum_{i=0}^{n-1} \| r_i^{(k)} d^k \| \right)^{1/k} \geqq \left(\mu^k \sum_{i=0}^{n-1} \| r_i^{(k)} \| \right)^{1/k} = \mu \| R^k \|^{1/k} .$$

Combining the above inequalities, we have

$$\mu \| R^k \|^{1/k} \leqq K \| R^k \|^{2/k} .$$

Since $R^k \neq 0$ for all k, we have that $\mu \leqq K \| R^k \|^{1/k}$. But $R \in \mathrm{Rad}(B)$ so that $\lim_{k \to \infty} \| R^k \|^{1/k} = 0$. Thus a contradiction and so R must have been zero.

Condition (ii) of the above corollary is satisfied when d is not a topological divisor of zero in A but may still be satisfied if d is a topological divisor of zero in A.

The case where the discriminant d of $\alpha(x)$ is invertible in A deserves special attention. In this case, if $f \in C(\Phi_B)$ and $\hat{\alpha}(f) = 0$, then there exists a $b \in B$ such that $\alpha(b) = 0$ and $\hat{b} = f$ (cf. 2.4 or [1]). Now, if $\phi \in E(\Phi_B : \Phi_A)$, then define $f(h, \lambda) = \mu$ where $(h, \mu) = \phi(h, \lambda)$. It is easily shown that f is a continuous function on Φ_B. Since $\alpha(f) = 0$, there exists a $b \in B$ with the above properties. Since ϕ is one-to-one, $\hat{b}$ $(= f)$ separates the points of $\pi^{-1}(h)$ for each $h \in \Phi_A$. Hence, it follows from Theorem 6.3 that $g: \sum_{i=0}^{n-1} a_i [x]^i \to \sum_{i=0}^{n-1} a_i b^i$ is an automorphism of B. (Note that $g^* = \phi$.) If we write $(*)$ for the mapping $g \to g^*$, $g \in G(B : A)$, then we have

COROLLARY 6.5. *If d is invertible in A, then $(*)$: $G(B : A) \to E(\Phi_B : \Phi_A)$ is one-to-one and once.*

In closing, we remark that if $g \in G(B : A)$, then g is continuous and hence *bi*-continuous.

7. **Extensions of ring isomorphisms.** If A is a Banach algebra with an involution $(*)$, then we ask: when can $(*)$ be extended to an involution on $A[x]/(\alpha(x))$? Or more generally, if $\phi: A_1 \rightarrow A_2$ is a ring isomorphism (need not commute with scalars), A_1 and A_2 Banach algebras, when can ϕ be extended to a ring isomorphism of $A_1[x]/(\alpha_1(x))$ onto $A_2[y]/(\alpha_2(y))$ (degree $\alpha_1(x)$ = degree $\alpha_2(y)$)? Simple examples show that $(*)$ and ϕ can not always be extended. However, under the added assumption that the discriminants of $\alpha_1(x)$ and $\alpha_2(y)$ are invertible in A_1 and A_2, respectively, then there is a necessary and sufficient condition that ϕ exist. The condition is stated in terms of a topological mapping. The case of extending $(^*)$ is less simple. In the proofs of our results on extending $(^*)$ and ϕ, we must consider elements $b \in A[x]/(\alpha(x))$ such that $\hat{b}$ separates the points of the fibers $\pi^{-1}(h)$, $h \in \Phi_A$. We will show that if the discriminant of $\alpha(x)$ is invertible, then such elements generate all of B over A. Before we prove this, we state a lemma which says that repeated extensions are algebraic in the strict sense of the word. The lemma is more general than needed here but will be used in the next section.

LEMMA 7.1. *Let A be a commutative ring (with unit) and let $B_i = B_{i-1}[x_i]/(\alpha_i(x_i))$, $B_0 = A$, $i = 1, 2, \cdots, m$, where $\alpha_i(x_i)$ is monic over B_{i-1} for each i. If $b \in B_m$, then there exists a monic polynomial $\alpha(x)$ over A of degree $n = \prod_{i=1}^{m} n_i$ ($n_i = \deg \alpha_i(x_i)$) such that $\alpha(b) = 0$.*

A proof of this lemma is to be found in [15] (page 255).

THEOREM 7.2. *Let A be a Banach algebra and let $\alpha(x) \in A[x]$ be a monic polynomial with an invertible discriminant in A. Then $b \in B$ has the property that $A[b] = B$ if and only if $\hat{b}$ separates the points of $\pi^{-1}(h)$ for each $h \in \Phi_A$.*

Proof. Suppose that $A[b] = B$. Then there are elements $a_i \in A$ such that $[x] = \Sigma a_i b^i$. If $\hat{b}(h, \lambda) = \hat{b}(h, \lambda')$ where (h, λ) and (h, λ') are points in Φ_B, then $[x]^\wedge(h, \lambda)$ must be equal to $[x]^\wedge(h, \lambda')$ so that $\lambda = \lambda'$ since $[x]^\wedge$ separates points of $\pi^{-1}(h)$. Hence, $\hat{b}$ separates the points of $\pi^{-1}(h)$ for each $h \in \Phi_A$.

Suppose now that $\hat{b}$ separates the points of $\pi^{-1}(h)$ for each $h \in \Phi_A$. By Lemma 7.1, we know that b satisfies a monic polynomial $\beta(x)$ of degree n ($= \deg \alpha(x)$). Since for each $h \in \Phi_A$, $\hat{b}$ takes on n distinct values on $\pi^{-1}(h)$, the discriminant of $\beta(x)$ must be invertible in A. Let B_0 denote the extension $A[y]/(\beta(y))$. Then $\Phi_{B_0} = \{(h, \mu) \in \Phi_A \times C: (h, \mu)\,\beta(y) = 0\}$, and $\theta: (h, \lambda) \rightarrow (h, \hat{b}(h, \lambda))$ is a continuous one-to-one mapping to Φ_B onto Φ_{B_0} and hence a homeomorphism. Therefore, $[x]^\frown \circ \theta^{-1}$ is a function continuous on Φ_{B_0} and $\hat{\alpha}([x]^\frown \circ \theta^{-1}) = 0$. Hence by the Arens-Calderón theorem (see 2.4 or [1]) there is an element $b_0 \in B_0$ such that $\alpha(b_0) = 0$ and $\hat{b}_0 = [x]^\frown \circ \theta^{-1}$. If ϕ denotes the homomorphism

$$\sum_{i=0}^{n-1} a_i[y]^i \rightarrow \sum_{i=0}^{n-1} a_i b^i ,$$

and if

$$b_0 = \sum_{i=0}^{n-1} a_i[y]^i ,$$

then

$$\begin{aligned}\phi(b_0)^\frown(h, \lambda) &= \phi\Big(\sum_{i=0}^{n-1} a_i[y]^i\Big)^\frown(h, \lambda) = \Big(\sum_{i=0}^{n-1} a_i b^i\Big)^\frown(h, \lambda) = \sum_{i=0}^{n-1} \hat{a}_i(h)\,(\hat{b}(h, \lambda))^i \\ &= \sum_{i=0}^{n-1} \hat{a}_i(h)\,([y])^\frown(\theta(h, \lambda)))^i = \hat{b}_0(\theta(h, \lambda)) = [x]^\frown(h, \lambda)\end{aligned}$$

for all $(h, \lambda) \in \Phi_B$. Hence, $\phi(b_0)^\frown = [x]^\frown$ and since $\alpha(\phi(b_0)) = 0$, we have that $\phi(b_0) = [x]$ by 2.4. Thus, ϕ is onto and $A[b] = B$.

COROLLARY 7.3. *Maintain the hypotheses on A and $\alpha(x)$. If $f \in C(\Phi_B)$ $\beta(y) \in B[y]$ such that*
(i) $\hat{\beta}(f) = 0$,
(ii) *f separates the points of $\pi^{-1}(h)$ for each $h \in \Phi_A$, and*
(iii) *$M_\beta((h, \lambda), f(h, \lambda))$ (M_β = multiplicity function of $\beta(y)$) is locally constant on Φ_B, then there exist $b \in B$ such that $A[b] = B$ and $\hat{b} = f$.*

The corollary follows immediately from 2.5 and the theorem.

COROLLARY 7.4. *Maintain the hypotheses on A and $\alpha(x)$. If $\hat{b}$ separates the points of $\pi^{-1}(h)$ for each h and $\beta(y) \in A[y]$ is a monic polynomial (of degree equal to the degree of $\alpha(x)$) satisfied by b, then $\phi\colon \Sigma a_i[y]^i \rightarrow \Sigma a_i b^i$ is an isomorphism of $A[y]/(\beta(y))$ onto $A[x]/(\alpha(x))$.*

Proof. (We use the notation of the theorem.) By the theorem we know that $A[b_0] = A[y]/(\beta(y))$ so that if $\phi(\sum_{i=0}^{n-1} a_i[y]^i) = \phi(\sum_{i=0}^{n-1} a_i' b_0^i) = 0$, then $\sum_{i=0}^{n-1} a_i'[x]^i = 0$. But this means that $a_i' = 0$ for each i and ϕ is an isomorphism.

Note that the above ϕ is continuous and hence *bi*-continuous.

Before we state and prove the next result, we require the following comments. Let $g\colon A_1 \rightarrow A_2$ be a ring isomorphism (onto). Define $g^*\colon \Phi_{A_1} \rightarrow \Phi_{A_2}$ as follows: for $h \in \Phi_{A_1}$, let $g^*(h)$ be the linear functional associated with the maximal ideal $g(h^{-1}(0))$ in A_2. Since g is one-to-one and onto, so is g^* one-to-one and onto. We now prove

LEMMA 7.5. *Let A_1 and A_2 be Banach algebras. If $g: A_1 \to A_2$ is a ring isomorphism (onto), then $g^*: \Phi_{A_1} \to \Phi_{A_2}$ is a homeomorphism (with respect to the weak* topologies on Φ_{A_1} and Φ_{A_2}).*

Proof. We can assume that A_1 and A_2 are semi-simple since g induces an isomorphism of $A_1/\mathrm{Rad}(A_1)$ onto $A_2/\mathrm{Rad}(A_2)$. Now, by a theorem of Kaplansky [9], $A_1 = \sum_{i=1}^{p} \oplus e_i A_1$ where the e_i are mutually orthogonal idempotents in A_1, $e_i A_1 \cong C$ for $i = 3, 4, \cdots, p$, and $g \mid e_1 A_1$ is linear while $g \mid e_2 A_1$ is conjugate linear. Thus, $\Phi_{A_1} = U_{i=1}^{p} \Phi_{e_i A_1}$ and the $\Phi_{e_i A_1}$ are disjoint open subsets of Φ_{A_1}. Since each $\Phi_{e_i A_1}$ consists of exactly one point if $3 \leqq i \leqq p$, $g^* \mid U_{i=3}^{p} \Phi_{e_i A_1}$ is continuous. That $g^* \mid \Phi_{e_1 A_1}$ is continuous follows from a now classical result (cf. Theorem 24B, [11]). To show shat $g^* \mid \Phi_{e_2 A_1}$ is continuous, we take $a \in e_2 A_1$ and let $\lambda = h(a), h \in \Phi_{e_2 A_1}$. Then $(a - \lambda e_2) \in h^{-1}(0)$. Since $g \mid e_2 A_1$ is conjugate linear, $g(a - \lambda e_2) = g(a) - \bar{\lambda} g(e_2) \in g^*(h)^{-1}(0)$, and hence $g(a)\hat{}(g^*(h)) = (\hat{a}(h))^{-}$. From this it follows immediately that $g^* \mid \Phi_{e_2 A_1}$ is a continuous mapping.

THEOREM 7.6. *Let A_1 and A_2 be Banach algebras, $\alpha_1(x_1) \in A_1[x_1]$ and $\alpha_2(x_2) \in A_2[x_2]$ be monic polynomials with invertible discriminants in A_1 and A_2, respectively, and $B_i = A_i[x_i]/(\alpha_i(x_i))$, $i = 1, 2$. If g is a ring isomorphism of A_1 onto A_2, then there exists an isomorphism $\tilde{g}$ of B_1 onto B_2 which extends g if and only if there exists a homeomorphism γ of Φ_{B_1} onto Φ_{B_2} such that $\pi_2 \circ \gamma = g^* \circ \pi_1$, where π_i is the usual mapping of Φ_{B_i} onto Φ_{A_i}. If g_1 and g_2 are any two such extensions of g, then $g_1 \circ g_2^{-1} \in G(B_2 : A_2)$.*

(Note that if γ exists, then $\alpha_1(x_1)$ and $\alpha_2(x_2)$ must have the same degree since for $h \in \Phi_{A_1}$, $\pi_1^{-1}(h)$ and $\pi_2^{-1}(g^*(h))$ have the same number of points.)

Proof. If $\tilde{g}$ extends g, then we take $\gamma = \tilde{g}^*$. By the above lemma, γ is a homeomorphism. γ is onto since $\tilde{g}$ is onto. Now, if M is a maximal ideal in B_1, then

$$g(M \cap A_1) = \tilde{g}(M \cap A_1) = \tilde{g}(M) \cap \tilde{g}(A_1) = \tilde{g}(M) \cap A_2 .$$

But this means that the restriction of $\tilde{g}^*(h, \lambda)$ to A_2 is $g^*(h)$ if $(h, \lambda)^{-1}(0) = M$. Thus, $\pi_2 \circ \tilde{g}^* = g^* \circ \pi_1$.

Suppose, now, that $\gamma: \Phi_{B_1} \to \Phi_{B_2}$ has the prescribed properties. Let $\beta(x_1) = \sum_{i=0}^{n} (g^{-1}(\alpha_{2,i}))\, x_1^i = 0$, where $\alpha_2(x_2) = \sum_{i=0}^{n} \alpha_{2,i} x_2^i$. We will show that there is a function f in $B_1\hat{}$ which separates the points of $\pi_1^{-1}(h)$ for each h in Φ_{A_1} and $\hat{\beta}(f) = 0$. Let $e_1, \cdots, e_p$ be the mutually orthogonal idempotents discussed in the proof of the above lemma. We define f

as follows. If $(h, \lambda) \in \pi_1^{-1}(\Phi_{e_1 A_1})$, let $f(h, \lambda) = [x_2]^\wedge(\gamma(h, \lambda))$ and if $(h, \lambda) \in \pi_1^{-1}(\Phi_{e_2 A_1})$, let $f(h, \lambda) = ([x_2]^\wedge(\gamma(h, \lambda)))^-$. For $h \in U_{i=3}^{p} \Phi_{e_2 A_1}$, let $\mu_1(h), \cdots, \mu_n(h)$ denote the n distinct roots of $\sum_{i=0}^{n} (g^{-1}(\alpha_{2,i}))^\wedge(h) x_1^i = 0$ and let $(h, \lambda_i(h))$ be the n points in $\pi_1^{-1}(h)$. For $(h, \lambda_i(h))$, let $f(h, \lambda_i(h)) = \mu_i(h)$. As defined, f is a continuous function on Φ_{B_1} and satisfies $\hat{\beta}(x_1) = 0$. Since f separates the points of $\pi_1^{-1}(h)$ for each $h \in \Phi_{A_1}$, and since $\beta(x)$ has an invertible discriminant in A_1, the Arens-Calderón theorem tells us that there exists $b \in B_1$ such that $\hat{b} = f$ and $\sum_{i=0}^{n} g^{-1}(\alpha_{2,i}) b^i = 0$. It follows from Corollary 7.4 that $A_1[b] = B_1$ and B_1 is isomorphic to $B_0 = A_1[y]/(\Sigma g^{-1}(\alpha_{2,i}) y^i)$. But B_0 is, of course, isomorphic to $B_2 = A_2[x_2]/(\alpha_2(x_2))$ so that B_1 and B_2 are isomorphic.

Suppose, now, that g_1 and g_2 are any two extensions of g. Then $g_1 \circ g_2^{-1}$ is clearly an automorphism of B_2 onto itself. Since $g_1 = g_2$ on A_1, $g_1 \circ g_2^{-1}$ leaves A_2 invariant elementwise, that is, $g_1 \circ g_2^{-1} \in G(B_2 : A_2)$.

The above theorem has the following interesting consequence if A is the group algebra $L^1(G)$, $G =$ integers. Let $\alpha(x) \in A[x]$ be an irreducible monic polynomial with an invertible discriminant. The irreducibility of $\alpha(x)$ together with the fact that the discriminant is invertible imply that Φ_B is connected (cf. Theorem 2.4, [10]). Then the above theorem implies that $A[x]/(\alpha(x))$ and $A[x]/(x^n - a_0)$ are isomorphic, where $n =$ degree $\alpha(x)$ and $a_0 \in A$ is the unique element such that $\hat{a}_0(z) = z$, $z \in \{\mu \in C : |\mu| = 1\} = \Phi_A$. If $a \in A$, let $\phi(a) = b$ where $\hat{b}(z) = \sum_{i=-\infty}^{\infty} b_i z^{ni}$ and $\hat{a}(z) = \sum_{i=-\infty}^{\infty} b_i z^i$. Then $\tilde{\phi}$: $\sum_{i=0}^{n-1} a_i [x]^i \rightarrow \sum_{i=0}^{n-1} \phi(a_i) a_0^i$ is clearly an isomorphism of $A[x]/(x^n - a)$ onto A so that $A[x]/(\alpha(x))$ is isomorphic to $A = L^1(G)$.

Another interesting consequence is that if $\alpha(x) \in A[x]$ is a monic polynomial with an invertible discriminant, then $A[x]/(\alpha(x))$ is isomorphic to $A[x]/(\alpha(x) + R(x))$ where $R(x) \in (\text{Rad } A)[x]$ and $\deg R(x) < \deg \alpha(x)$.

We now turn our attention to the case where $g: A \rightarrow A$ is a periodic automorphism and, in particular, an involution of a certain type. The following example shows that not every such automorphism is extendable. Let $A = C(\{z \in C : |z + 1| = 1 \text{ or } |z - 1| = 1\}$ and $\alpha(x) = x^2 - f$, $f(z) = z + 1$ if $|z + 1| = 1$ and $f(z) = 1$ if $|z - 1| = 1$. For an involution, we take $f^*(z) = (f(-z))^-$. g has no extension to B since this would imply that there exists a homeomorphism γ of Φ_B onto Φ_B such that $\gamma(z, \lambda) = (-z, [x]^\wedge(\gamma(z, \lambda))$. But it is impossible for such a homeomorphism to exist. Hence, g has no extension.

However, if $g: A \rightarrow A$ is a periodic automorphism which has an extension $\tilde{g}$ to B (we are assuming that $\alpha(x)$ has an invertible discriminant), then $\tilde{g}$ is periodic and its period divides $n!\, p$, $p =$ period of g. For if $g^p =$ identity automorphism, then $\tilde{g}^{*p}(h, \lambda) = (g^{*p}(h), [x]^\wedge(\tilde{g}^{*p}(h, \lambda))) = (h, [x]^\wedge(\tilde{g}^{*p}(h, \lambda))$ so that $\tilde{g}^{*p} \in E(\Phi_B : \Phi_A)$. Hence $(\tilde{g}^{*p})^{n!} =$ identity homomorphism. Thus, $\tilde{g}^*$ is periodic. By Corollary 6.4, $\tilde{g}^{pn!}$ is the identity automorphism. Simple examples show that the period of $\tilde{g}$

may be $p \cdot n!$ We now restrict our attention to the case where g is a symmetric involution, that is, $(a^*)^\wedge(h) = (\hat{a}(h))^-$.

THEOREM 7.6. *Let A be a Banach algebra and $\alpha(x) \in A[x]$ a monic polynomial with an invertible discriminant in A. If $(^*): A \to A$ is a symmetric involution, then there exists a unique symmetric involution $('): B \to B$ which extends $(^*)$. If $('')$ is any involution extending $(^*)$, then $('') = (') \circ g$ for some $g \in G(B:A)$ which is of period two.*

Proof. Let $\alpha^*(x) = \sum_{i=0}^{n} \alpha_i^* x^i$ where $\alpha(x) = \sum_{i=0}^{n} \alpha_i x^i$. Then $\hat{\alpha}^*(f) = 0$ where $f(h, \lambda) = \bar{\lambda}$. By the Arens-Calderón theorem, there is an element $b_0 \in B$ such that $\alpha^*(b_0) = 0$ and $\hat{b}_0 = f$. Let $(')$ denote the mapping defined by $(\sum_{i=0}^{n-1} a_i[x]^i)' = \sum_{i=0}^{n-1} a_i^* b_0^i$. Clearly $(')$ is a homomorphism and $\alpha(b_0') = 0$. But

$$(b_0')^\wedge(h, \lambda) = \left(\sum_{i=0}^{n-1} a_i^* b_0^i\right)^\wedge (h, \lambda) = \left(\sum_{i=0}^{n-1} (\hat{a}_i(h))^- (\bar{\lambda})^i\right)$$
$$= \left(\sum_{i=0}^{n-1} (\hat{a}_i(h))\lambda^i\right)^- = \lambda = [x]^\wedge(h, \lambda)$$

where $b_0 = \sum_{i=0}^{n-1} a_i[x]^i$, and (h, λ) is any point of Φ_B. Thus, $(b_0')^\wedge = [x]^\wedge$, and it follows that $b_0' = [x]$. Thus, $(')$ is an involution. That $(')$ is symmetric follows from the fact that $((\Sigma a_i[x]^i)')^\wedge = \Sigma((\hat{a}_i)^-)f^i, f = ([x]^\wedge)^-$.

If $('')$ is any symmetric involution on B which extends $(^*)$, then $\alpha^*([x]'') = 0$. But $([x]'')^\wedge = \hat{b}_0$ so that $[x]'' = b_0$. Thus $(')$ is a unique symmetric involution extending $(^*)$.

If $('')$ is any involution (not necessarily symmetric), then $(')^{-1} \circ ('') = g$ belongs to $G(B:A)$. To show g is of period two, consider the following. Since the involution defined on $B^\wedge$ by conjugation commutes with every involution, $g^2(b)^\wedge$ is equal to $\hat{b}$ for every $b \in B$; hence, in particular, $g^2([x])^\wedge = [x]^\wedge$. But $\alpha(g^2([x])) = 0$ so that $g^2([x]) = [x]$ and g is of period two.

8. Primitive elements in repeated extensions.

As seen in § 6, there is some analogy between the present study and the classical case of field extensions. We carry this analogy one step further by proving a theorem about the existence of primitive elements in repeated extensions. It will follow from our theorem, that if $\alpha(x)$ is a monic polynomial with an invertible discriminant, then there exists an extension of the form $A[x]/(\beta(x))$ over which $\alpha(x)$ factors into linear factors.

THEOREM 8.1. *Let A be a Banach algebra. If $B_0 = A$ and $B_i = B_{i-1}[x_i]/(\alpha_i(x_i))$, $i = 1, 2, \cdots, m$, where x_i is an indeterminate over B_{i-1} and $\alpha_i(x_i) \in B_{i-1}[x_i]$ is a monic polynomial with an invertible discrimi-*

nant in B_{i-1}, then there exists a monic polynomial $\alpha(x) \in A[x]$ with an invertible discriminant and an element $b \in B_m$ such that $\alpha(b) = 0$ and $A[b] = B_m \cong A[x]/(\alpha(x))$ (algebraically and topologically).

Proof. The proof is by induction. We shall prove the case $m = 2$. Consider $[x_1]^\wedge(h, \lambda) + c[x_2]^\wedge(h, \lambda, \mu) = \lambda + c\mu$, where c is a complex number, and $(h, \lambda, \mu) \in \Phi_{B_2}$. We will show that we can choose $c > 0$ such that $\lambda + c\mu \neq \lambda' + c\mu'$ if $(h, \lambda, \mu) \neq (h, \lambda', \mu')$. If

$$F(h) = \min\{|\lambda - \lambda'| : (h, \lambda), (h, \lambda') \in \Phi_{B_1} \text{ and } \lambda \neq \lambda'\} \text{ for each } h \in \Phi_A ,$$

then F is a continuous function on Φ_A since $\alpha_1(x_1)$ has no singular points in Φ_A. Since Φ_A is compact and since $F(h) > 0$ for each $h \in \Phi_A$, there exists $s > 0$ such that $F(h) > s$ on Φ_A. Choose $c > 0$ so that $s > 2 \cdot c \cdot \|[x_2]^\wedge\|_\infty$. For this choice of c, let $b = [x_1] + c[x_2]$. Now, if $(h, \lambda, \mu) \neq (h, \lambda, \mu')$, then $\hat{b}(h, \lambda, \mu) \neq \hat{b}(h, \lambda, \mu')$ and if $(h, \lambda) \neq (h, \lambda')$, then

$$\begin{aligned} |\hat{b}(h, \lambda, \mu) - \hat{b}(h, \lambda', \mu')| &\geqq |\lambda - \lambda'| - c \cdot |\mu - \mu'| \\ &\geqq s - c \cdot |\mu - \mu'| > s - 2 \cdot c \cdot \|[x_2]^\wedge\|_\infty > 0 . \end{aligned}$$

From this it follows that if $\alpha(x)$ is the monic polynomial (constructed in Lemma 7.1) of degree $n = n_1 n_2$ satisfied by b, then its discriminant is invertible since corresponding to each h, $\alpha_h(x) = 0$ has $n_1 n_2$ distinct roots.

Let $B = A[x]/(\alpha(x))$. Then Φ_B is (identifiable with) $\{(h, \lambda) \in \Phi_A \times C: (h, \lambda)\alpha(x) = 0\}$. Hence $\gamma: (h, \lambda, \mu) \to (h, \hat{b}(h, \lambda, \mu))$ is a homeomorphism of Φ_{B_2} onto Φ_B. Thus, $[x_1]^\wedge \circ \gamma^{-1}$ is continuous on Φ_B and $\hat{\alpha}_1([x_1]^\wedge \circ \gamma^{-1}) = 0$. By the Arens-Calderón theorem, there exists $b_1 \in B$ such that $\hat{b}_1 = [x_1]^\wedge \circ \gamma^{-1}$ and $\alpha_1(b_1) = 0$. Now, if $g: \sum_{i=0}^{n-1} a_i[x]^i \to \sum_{i=0}^{n-1} a_i b_1^i$, then g is a homomorphism of B onto $A[b]$. By an argument in the proof of Theorem 7.2, we have that $g(b_1)^\wedge = [x_1]^\wedge$. But $\alpha_1(b_1) = 0$ so that $\alpha_1(g(b_1)) = 0$ from which it follows that $g(b_1) = [x_1]$ since the discriminant of $\alpha_i(x_1)$ is invertible. Thus, $A[b]$ contains $[x_1]$ and hence $[x_2] \in A[b]$, i.e., $A[b] = B_2$. It remains to show that g is one-to-one and *bi*-continuous. Clearly, $g \mid A[b_1]$ is one-to-one so that there is an element $b_2 \in B$ which satisfies $\sum_{i=0}^{n_2} (g \mid A[b_1])^{-1}(\alpha_i^{(2)}) b_2^i = 0$ and $\hat{b}_2 = [x_2]^\wedge \circ \gamma^{-1}$, where $\alpha_2(x) = \sum_{i=0}^{n_2} \alpha_i^{(2)} x^i$. As before, $\alpha_2(g_2)) = 0$ and $g(b_2)^\wedge = [x_2]^\wedge$ so that $g(b_2) = [x_2]$. Hence, $g \mid A[b_1 + cb_2]$ is a one-to-one mapping. But $(b_1 + cb_2)^\wedge = [x]^\wedge$ so that $A[b_1 + cb_2] = A[x]/(\alpha(x))$. Thus, g is one-to-one. (Note that this means that $b_1 + cb_2 = [x]$.) The continuity of g follows as in Theorem 7.2. The *bi*-continuity follows from the closed graph theorem.

COROLLARY 8.2. *If $\alpha(x) \in A[x]$ is a monic polynomial with an invertible discriminant in A, then there exists an extension of the form $A[x]/(\beta(x))$ over which $\alpha(x)$ factors into linear factors, where*

$\beta(x)$ is a monic polynomial with an invertible discriminant.

In view of the theorem, the proof of the corollary follows from the fact that if $\alpha(x) = (x - b_1), \cdots, (x - b_t)Q(x)$ over $A[x]/(\alpha(x))$, then $Q(x)$ must have an invertible discriminant over $A[x]/(\alpha(x))$.

9. On the radical of B. Let A be a normed algebra and let $K(A)$ denote the intersection of the closed maximal ideals of A. If $K(A) = (0)$, we say that A is tractable. In [2] it is shown that if A is tractable and if the discriminant of $\alpha(x)$ is not a zero divisor in A, or zero, then B is also tractable. It is further shown that if A is tractable and if $\alpha(x) = x^n - a$, then B is tractable if and only if a is not a zero divisor in A, or zero. Actually, these results are true for a wider class of algebras, namely, commutative algebras (with unit) over fields of characteristic zero, with "tractable" replaced by "semi-simple."

In this section, we will show that the converse of the above theorem is also valid; indeed, we formulate our theorems and corollaries in the general context of algebras over fields of characteristic zero. To do so requires no extra effort, except that of characterizing the maximal ideals of B in terms of those of A. It will follow from the general results presented that when A is tractable, then the radical of B and the intersection of the closed maximal ideals of B coincide, a result that is generally not valid for normed algebras. (An example of a semi-simple normed algebra which is not tractable is given at the end of this section.) Thus, until further notice, we assume that A is a commutative algebra (with unit) over a field F of characteristic zero. Let M_A denote the maximal ideal space of A. We first identify M_B in terms of M_A. If m_0 is a maximal ideal in B, then B/m_0 is a field which contains an isomorphic copy of F and hence is also of characteristic zero. Let ϕ denote the canonical homomorphism of B onto B/m_0. Then $\phi(A)$ is a subfield of B/m_0 since the latter is a simple algebraic extension of $\phi(A)$ (cf. page 259, [15]). Thus we see that $m_0 \cap A$ is a maximal ideal of A. On the other hand, if m is a maximal ideal in A, then we can extend m to (at most $n =$ degree of $\alpha(x)$) a maximal ideal of B. We proceed to show this assertion and at the same time give a description of the extensions.

If I is an ideal in A, then let $\beta_I(x)$ denote $\Sigma(\beta_i + I)x^i$ where $\beta(x) = \Sigma\beta_i x^i$.

Let $m \in M_A$ and $\gamma(x)$ denote a monic polynomial over A such that $\gamma_m(x)$ is an irreducible factor of $\alpha_m(x)$. Let $(m, \gamma(x))$ denote the set

$$\left\{\left(\sum_{i=0}^{n-1} a_i[x]^i\right)\gamma([x]) + \sum_{i=0}^{n-1} m_i[x]^i : a_i \in A,\ m_i \in m\right\}.$$

It is clear that $(m, \gamma(x))$ is an ideal in B. If we define θ by

$$\theta\Big(\sum_{i=0}^{n-1} a_i[x]^i\Big) = \sum_{i=0}^{n-1} (a_i + m)\,(x + (\gamma_m(x)))^i$$

then θ is a homomorphism of B onto $(A/m)[x]/(\gamma_m(x))$. Clearly $(m, \gamma(x)) \subseteq \theta^{-1}(0)$. Now if $\sum_{i=0}^{n-1} (a_i + m)\,(x + (\gamma_m(x)))^i = 0$, then

$$\sum_{i=0}^{n-1} (a_i + m)x^i = \gamma_m(x)Q_m(x) \ ,$$

where $Q(x) \in A[x]$ or equivalently,

$$\sum_{i=0}^{n-1} a_i x^i - \gamma(x)Q(x) \in m[x] \ ,$$

Thus, $\sum_{i=0}^{n-1} a_i[x]^i \in (m, \gamma(x))$. Hence, $\theta^{-1}(0) = (m, \gamma(x))$ and so $(m, \gamma(x))$ is a maximal ideal of B.

From the above, it is clear that if $\gamma_1(x) - \gamma_2(x) \in m[x]$, then $(m, \gamma_1(x)) = (m, \gamma_2(x))$. We now show the converse. Suppose $(m, \gamma_1(x)) = (m, \gamma_2(x))$. There exists $p(x) \in A[x]$ and $m(x) \in m[x]$ such that $\gamma_2(x) = \gamma_1(x)p(x) + m(x)$. Now, $\gamma_{2m}(x) = \gamma_{1m}(x)p_m(x)$. Since both $\gamma_{1m}(x)$ and $\gamma_{2m}(x)$ are irreducible, and monic, $p_m(x) = e + m$. The degrees of $\gamma_1(x)$, $\gamma_2(x)$, $\gamma_{1m}(x)$ and $\gamma_{2m}(x)$ are all equal so that $p(x) = e$. Thus, $\gamma_2(x) - \gamma_1(x) \in m[x]$.

So far we have shown that each maximal ideal of A extends to at least one maximal ideal of B. Furthermore, each maximal ideal of B extends a unique maximal ideal of A. We shall now show that each maximal ideal m_0 of B is of the form given above, with $m = m_0 \cap A$. From earlier comments we know that B/m_0 is a simple algebraic extension of the field $\phi(A)$, where $\phi\colon B \to B/m_0$ is the canonical homomorphism. Since $\phi([x])$ is a root of $\alpha_m(x) = 0$, $\phi([x])$ must satisfy one of its irreducible factors, say $\beta_m(x)$. Hence B/m_0 must be isomorphic to $\phi(A)[x]/(\beta_m(x))$. Thus, if $\phi(\sum_{i=0}^{n-1} a_i[x]^i) = 0$, then $\sum_{i=0}^{n-1} (a_i + m)x^i = Q_m(x)\beta_m(x)$. Thus, $m_0 = (m, \beta(x))$.

In summary, we have that M_B may be viewed as the set of ordered pairs $(m, \beta(x))$, $m \in M_A$, $\beta(x)$ monic and $\beta_m(x)$ an irreducible factor of $\alpha_m(x)$. Of course, we identify any two such pairs $(m, \beta(x))$ and $(m', \gamma(x))$ if and only if $m = m'$ and $\beta(x) - \gamma(x) \in m[x]$. As before, we let π denote the (onto) mapping $(m, \beta(x)) \to m$.

In what follows, let $a(m)$ denote the coset $a + m$, $a \in A$, $m \in M_A$.

In order to avoid interrupting the proof of the main theorem, we will next state and prove a lemma about the existence of a common factor of $a\alpha(x)$ and $b\alpha'(x)$ for suitable elements a and b in A. In general, a and b will not be invertible elements (consider the $\alpha(x)$ in Example 5.3). We will need the following result [15]: Let $f(x)$ and $g(x)$ be polynomials over A of respective degrees m and n, let $k = \max(m - n + 1, 0)$ and let a be the leading coefficient of $g(x)$. Then there exist polynomials $Q(x)$ and $R(x)$ over A such that

$$a^k f(x) = Q(x)g(x) + R(x)$$

and $R(x)$ is either of degree less than n or is the zero polynomial.

LEMMA 9.1. *Let A be semi-simple. If the discriminant d of $\alpha(x)$ is a zero divisor in A (say $dc = 0$, $c \neq 0$) or if $d = 0$, then there are nonzero elements a and b in A and polynomials $\gamma(x)$, $\delta(x)$ and $R(x)$ over A such that*

(i) $a\alpha(x) = \gamma(x)R(x)$

(ii) $b\alpha'(x) = \delta(x)R(x)$

(iii) *for $m \in M_A$, $a(m) = 0$ if and only if $b(m) = 0$, and if $c(m) = 0$, then $a(m) = 0$, and*

(iv) *if $\beta_m(x)$ ($m \in M_A$) is a factor of $\alpha_m(x)$ and $\alpha'_m(x)$, then $\beta_m(x)$ is a factor of $R_m(x)$.*

Proof. We first prove the lemma for the case $d = 0$. Let $R_{-1}(x)$ and $R_0(x)$ denote $\alpha(x)$ and $\alpha'(x)$, respectively. In view of the above quoted result, we assume that we have found polynomials $Q_{j+1}(x)$, $R_{j+1}(x)$, $0 \leqq j \leqq i$, over A such that

$$(*) \qquad R_{0\ j}^{2k_j}\, R_{j-1}(x) = R_{0,j-1}R_j(x)Q_{j+1}(x) + R_{j+1}(x)$$

and $R_{0,j}R_{j+1}(x) \neq 0$ for $0 \leqq j \leqq i$, where $R_{0,j}$ denotes the leading coefficient of $R_{0,j-1}R_j(x)$ and $k_i = \max\{\deg(R_{j-1}(x)) - (\deg R_{0,j-1}R_j(x)) + 1, 0\}$. The polynomial $R_{0,j}R_{j+1}(x)$ is never a non-trivial constant polynomial. This follows from the fact that if $m \in M_A$, then $\alpha_m(x)$ and $\alpha'_m(x)$ have at least one irreducible factor in common since $d(m) = 0$ (recall that A/m is a field of characteristic zero). For each m, let $\beta_m(x)$ be one such factor. Thus, it follows that if $R_{0,j}(m) \neq 0$, then $\beta_m(x)$ is a factor of $R_{0,j}(m)\,(R_{j+1})_m(x)$. Thus, if $R_{0,j}R_{j+1}(x)$ were a constant, say c, then $c(m) = 0$ for all $m \in M_A$. Since A is semi-simple, $c = 0$. From this fact and the fact that degree $R_{j+1}(x) <$ degree $R_{0,j}R_{0,j-1}(x)$, we can conclude that there is a first integer, say i_0, such that (*) holds with $j = i_0$ and $R_{0,i_0}R_{i_0+1}(x) = 0$. Since the coefficients of $R_{i_0+1}(x)$ belong to the same maximal ideals that R_{0,i_0} belongs to, we have that $R_{i_0+1}(x)$ is the zero polynomial. Hence

$$R_{0\ i_0}^{2k_{i_0}} R_{i_0-1}(x) = R_{0,i}R_{0,i_0-1}(x)Q_{i_0+1}(x)\ .$$

Let $R(x) = R_{0,i_0}R_{0,i_0-1}R_{i_0}(x)$, $a = \prod_{j=0}^{i_0} R_{0,j}^{2k_j}$ and $b = \prod_{j=1}^{i_0} R_{0\ j}^{2k_j}$. Then a and b are nonzero and belong to the same maximal ideals to which $R_{0,j}$ belongs. Now, by repeated substitutions, we find polynomials $\gamma(x)$ and $\delta(x)$ over A such that $a\alpha(x) = \gamma(x)R(x)$ and $b\alpha'(x) = \delta(x)R(x)$. From the above it is clear that if $\beta_m(x)$ is a factor of $\alpha_m(x)$ and $\alpha'_m(x)$, then it is a factor of $R_m(x)$.

If $dc = 0$ ($d \neq 0$, $c \neq 0$), then let D denote the set of maximal ideals of A to which c doesn't belong, and I denote the intersection of the maximal ideals in D. By the first part of the proof, there are elements $a', b' \in A$ ($a', b' \notin I$) and polynomials $\tilde{\gamma}(x)$, $\tilde{\delta}(x)$ and $\tilde{R}(x)$ over A such that for the cosets $a' + I$ and $b' + I$ and the polynomials $\tilde{\gamma}_I(x)$, $\tilde{\delta}_I(x)$ and $\tilde{R}_I(x)$, the four conditions of the lemma are fulfilled over A/I. It then follows that the same four conditions are fulfilled over A if we take $a = c^2a'$, $b = c^2b'$, $\gamma(x) = c\tilde{\gamma}(x)$, $\delta(x) = c\tilde{\delta}(x)$ and $R(x) = c\tilde{R}(x)$. (Note that a and b are not zero since if so we would have that a' and b' belong to I.) This completes the proof of the lemma.

It is necessary to introduce the following notation at this point. Let A be semi-simple and $\alpha(x)$ a monic polynomial over A. $M_\alpha(m, \beta_m(x))$ is to denote the power to which $\beta_m(x)$ appears in the factorization of $\alpha_m(x)$ into irreducible factors. Let d_k denote the resultant of $\alpha(x)$ and $\alpha^{(k)}(x)$ (= the formal kth derivative of $\alpha(x)$), $1 \leqq k \leqq n-1$ (cf. page 96, [14]) and let $k(\alpha)$ denote the smallest integer k, if it exists, such that d_k is not a zero divisor in A, or zero, and n if all the d_k are zero divisors in A, or zero. From the definition it follows that if $k > k(\alpha)$, then d_k is not a zero divisor in A or zero.

By a nil ideal in A we mean an ideal all of whose elements are nilpotent. If I is an ideal in A for which there exists an integer k such that $a_1 \cdot a_2 \cdot \cdots \cdot a_k = 0$ whenever $a_i \in I$, $i = 1, 2, \cdots, k$, then we say that I is nilpotent (and write $I^k = (0)$) and if k is the smallest such integer, then we call k the degree of nilpotency of I.

THEOREM 9.2. *Suppose that A is semi-simple and that $\alpha(x)$ is a monic polynomial over A for which $k(\alpha) \geqq 2$. Then the radical of B is nontrivial consisting precisely of the nilpotent elements of B. Furthermore, Rad B is nilpotent and its degree of nilpotency is $k(\alpha)$.*

Proof. It is well known that the radical of an algebra contains all the nilpotent elements of the algebra. We show that Rad(B) consists of precisely nilpotent elements by showing the last assertion of the theorem, from which it follows that Rad(B) is nontrivial.

Suppose that $\beta_1([x]), \cdots, \beta_{k(\alpha)}([x]) \in \text{Rad}(B)$ and set $\beta(x) = \prod_{i=1}^{k(\alpha)} \beta_i(x)$. Then there are polynomials $Q(x)$ and $R(x)$ over A such that $\beta(x) = \alpha(x)Q(x) + R(x)$, with degree $R(x) <$ degree $\alpha(x)$. We will show that $\beta([x]) = 0$ by showing that $R(x)$ is the zero polynomial. Suppose first that $m \in M_A$ has the property that $M_\alpha(m, \gamma_m(x)) \leqq k(\alpha)$ for every irreducible factor $\gamma_m(x)$ of $\alpha_m(x)$. Since $\beta_i([x]) \in \text{Rad}(B)$, we know that $\gamma_m(x)$ must divide $(\beta_i)_m(x)$, and hence $\gamma_m(x)^j$, $j = M_\alpha(m, \gamma_m(x))$ divides $\beta_m(x)$. Furthermore $\gamma_m(x)^j$ divides $\alpha_m(x)$ (by definition of j) so that $\gamma_m(x)^j$ also divides $R_m(x)$. But $\gamma_m(x)$ is an arbitrary irreducible factor so it follows that $\alpha_m(x)$ divides $\beta_m(x)$ and consequently also divides $R_m(x)$. Since degree $R_m(x) <$ degree $\alpha_m(x)$, $R_m(x)$ is the zero polynomial over A/m, or equivalently, the coefficients of $R(x)$ lie in m. If there is a $m \in M_A$ such that $M_\alpha(m, \gamma_m(x)) > k(\alpha)$ for some $\gamma(x)$, then $d_{k(\alpha)} \in m$. Thus, the coefficients of $d_{k(\alpha)}R(x)$ lie in every maximal ideal in A and hence are all zero. But $d_{k(\alpha)}$ is neither a zero divisor in A or zero, so that $R(x)$ is the zero polynomial over A. Thus, $\beta(x) = \alpha(x)Q(x)$, or equivalently, $\beta([x]) = \prod_{i=1}^{k(\alpha)} \beta_i([x]) = 0$.

To show that $\mathrm{Rad}(B)^{k(\alpha)-1} \neq (0)$ (recall that $k(\alpha)$ is assumed to be greater than one), it suffices to show that there is an element $f \in \mathrm{Rad}(B)$ such that $f^{k(\alpha)-1} \neq 0$. We will show that $f = a\gamma([x])$ is a suitable choice, where a and $\gamma(x)$ are supplied to us by Lemma 9.1. (We may assume that c in the lemma has the property that $cd_{k(\alpha)-1} = 0, c \neq 0$.) Let us first note that $a\gamma([x]) \neq 0$. For if not, then $a\gamma(x) = Q(x)\alpha(x)$ for some $Q(x) \in A[x]$. But $a\alpha(x) = \gamma(x)R(x)$ so that $a^2\alpha(x) = Q(x)R(x)\alpha(x)$ or $a^2 = Q(x)R(x)$. If $c(m) = 0$, then $a(m) = 0$. If $c(m) \neq 0$, then $d(m) = 0$ so that $\alpha_m(x)$ and $\alpha'_m(x)$ have a common factor which is also a factor of $R_m(x)$ by (iv) of the lemma. Thus, $a(m) = 0$ for all $m \in M_A$ and hence $a = 0$, which is a contradiction. We show next that $a\gamma([x]) \in \mathrm{Rad}(B)$.

Let m be a maximal ideal such that $a(m) \neq 0$ and $\beta_m(x)$ an irreducible factor of $\alpha_m(x)$. If $\beta_m(x)$ is not a factor of $b(m)\alpha'_m(x)$, then $\beta_m(x)$ is not a factor of $R_m(x)$ (cf. lemma). Hence $\beta_m(x)$ must be a factor of $\gamma_m(x)$. If, on the other hand, $\beta_m(x)$ is a factor of $b(m)\alpha'_m(x)$, then $\beta_m(x)$ is a factor of $\alpha'_m(x)$ ($b(m) \neq 0$ since $a(m) \neq 0$). Thus, from the lemma, we can conclude that $\beta_m(x)^k$, $k = M_\alpha(m, \beta_m(x)) - 1$, is also a factor of $\alpha'_m(x)$, hence a factor of $R_m(x)$ since $\beta_m(x)^{k+1}$ is a factor of $\alpha_m(x)$. Thus, $\beta_m(x)$ must be a simple factor of $\gamma_m(x)$. We can now conclude that $a\gamma([x])$ belongs to every maximal ideal of B.

We now show that $(a\gamma([x]))^{k(\alpha)-1} \neq 0$ or equivalently, $a\gamma(x)^{k(\alpha)-1} \neq Q(x)\alpha(x)$ for every $Q(x) \in A[x]$. Since $k(\alpha) \geqq 2$, we know that there is at least one irreducible factor $\beta_m(x)$ of $\alpha_m(x)$ for some $m \in M_A$ such $\beta_m(x)$ is also a factor of $\alpha'_m(x)$ and $\beta_m(x)^{k(\alpha)}$ is a factor of $\alpha_m(x)$ (take any $m \in M_A$ such that $d_{k(\alpha)-1} \in m$). From what we showed above, we have that $\beta_m(x)$ is a simple factor of $\gamma_m(x)$. If $(a\gamma(x))^{k(\alpha)-1} = Q(x)\alpha(x)$ for some $Q(x) \in A[x]$, then $\beta_m(x)^{k(\alpha)}$ would be a factor of $(a(m)\gamma_m(x))^{k(\alpha)-1}$ or else $a(m) = 0$. Since $ad_{k(\alpha)-1} = 0$ (recall our assumption that $cd_{k(\alpha)-1} = 0$), we may assume that $a(m) \neq 0$. Hence a contradiction since $\beta_m(x)$ is only a simple factor of $\gamma_m(x)$. Thus, $(a\gamma([x]))^{k(\alpha)-1} \neq 0$.

COROLLARY 9.3. *If B is semi-simple, then A is semi-simple and the discriminant d of $\alpha(x)$ is not a zero divisor in A, or zero.*

The proof follows immediately from the theorem. To use the theorem, we need to know that A is semi-simple. But this is true since each maximal ideal of A extends to at least one of B. This situation is special. (There are examples of semi-simple algebras with non-semi-simple subalgebras.)

COROLLARY 9.4. *Let A be a commutative algebra with non-trivial radical $R = \mathrm{Rad}(A)$. Then $\mathrm{Rad}(B) = \{b \in B : b^k \in R[[x]]\}$, $k = k(\alpha_R)$. If R is a nil ideal, then so is $\mathrm{Rad}(B)$. If R is nilpotent, say $R^p = (0)$, then so is $\mathrm{Rad}(B)$ and $\mathrm{Rad}(B)^{pk} = (0)$.*

Proof. Since $\mathrm{Rad}(B) \supset R$, it is clear that $\mathrm{Rad}(B) \supseteqq \{b \in B : b^k \in R[[x]]\}$. Now, consider the homomorphism ϕ of B onto $(A/R)[x]/(\alpha_R(x))$ defined by $\phi(\Sigma a_i[x]^i) = \Sigma(a_i + R)x^i + (\alpha_R(x))$. Then $\phi(\mathrm{Rad}(B)) \subseteqq \mathrm{Rad}((A/R)[x]/(\alpha_R(x)))$ (cf. page 10, [8]). The kernel of ϕ is $R[[x]]$. Thus, if $b \in \mathrm{Rad}(B)$, then $\phi(b^k) = (\phi(b))^k = 0$ by the theorem. It follows that $b^k \in \phi^{-1}(0)$ so that $\mathrm{Rad}(B) \subseteqq \{b \in B : b^k \in R[[x]]\}$. Thus equality holds and the first assertion of the corollary is established.

Suppose now that R is a nil ideal. Let $b \in \mathrm{Rad}(B)$. Then by the above, $b^{k(\alpha_R)} \in R[[x]]$. Let $b^{k(\alpha_R)} = \sum_{i=0}^{n-1} b_i[x]^i$, $b_i \in R$. Since A is commutative, the elements $b_0, \cdots, b_{n-1}$ generate a nilpotent ideal in A (cf. page 193, [8]). If p is the degree of nilpotency of this ideal, then $(b^{k(\alpha_R)})^p = 0$. Thus, $\mathrm{Rad}(B)$ is a nil ideal.

The last assertion follows immediately from what we just proved.

If the degree of nilpotency of $\mathrm{Rad}(A)$ is p, it may well be the case that the degree of nilpotency of $\mathrm{Rad}(B)$ is less than $pk(\alpha_R)$. For example, take an algebra for which $p = 2$ and let $\alpha(x) = x^3$. Then $\mathrm{Rad}(B)^4 = \{0\}$. (It is easy to modify this example so that $\alpha(x) = 0$ has no solution in A.) On the other hand, the degree of nilpotency of $\mathrm{Rad}(B)$ may be equal to $k(\alpha_R)p$.

We now turn our attention to the case where A is a commutative normed algebra. For such an algebra, $K(A)$ denotes the intersection of its closed maximal ideals.

THEOREM 9.5. *Let A be a tractable normed algebra. Then $K(B)$ coincides with the radical of B. Hence if B is tractable, then A is tractable and d is not a divisor of zero in A, or zero.*

In order to prove the theorem we only have to establish that the elements of $K(B)$ are nilpotent. To do this, we must know which maximal ideals of B are closed. Of course, each closed maximal ideal of B extends a maximal ideal of A so that Φ_B (= space of closed maximal ideals of B) is a subset of $D = \{(h, \lambda) \in \Phi_A \times C : \alpha_h(\lambda) = 0\}$. Actually, $\Phi_B = D$. To see this, observe that

$$|\alpha_h(\lambda)| \geqq |\lambda|^n - \|\alpha_{n-1}\| \, |\lambda|^{n-1} - \cdots - \|\alpha_1\| \, |\lambda| - \|\alpha_0\| .$$

If $|\lambda| > 1$, then the right hand side is greater than zero so that

$|\alpha_h(\lambda)| > 0$ for all $h \in \Phi_A$. Thus, if $(h, \lambda) \in D$, then $|\lambda| \leqq 1$ and hence (h, λ) defines a continuous multiplicative linear functional (recall that we are assuming that $||\alpha_0|| + ||\alpha_1|| + \cdots + ||\alpha_{n-1}|| \leqq 1$).

Now, using the fact that $\Phi_B = D$, we use the method of proof of the first assertion of Theorem 9.2 to establish that $K(B)$ is nilpotent. Hence $K(B) \subseteqq \mathrm{Rad}(B)$. On the other hand, $\mathrm{Rad}(B) \subseteqq K(B)$ so that $K(B) = \mathrm{Rad}(B)$.

The second assertion now follows from Corollary 9.3.

As we have pointed out earlier, there are normed algebras which are semisimple but not tractable. A simple example illustrating this is as follows: Let A be any normed algebra with no nonzero nil ideals but possessing a nontrivial radical. $A[x]$ is a normed algebra under $||\Sigma a_i x^i|| = \Sigma ||a_i||$. Clearly, $A[x]$ is not tractable. However $A[x]$ is semi-simple (cf. Theorem 4, page 12, [8]).

BIBLIOGRAPHY

1. R. Arens and A. P. Calderón, *Analytic functions of several Banach algebra elements,* Ann. of Math., **6** (1955), 204–216.

2. R. Arens and K. Hoffman, *Algebraic extensions of normed algebras,* Proc. Amer. Math. Soc., **7** (1956), 203–210.

3. W. G. Badé and P. C. Curtis, Jr., *The Wedderburn decomposition of commutative Banach algebras,* Amer. J. Math., **82** (1960), 851–866.

4. C. Feldman, *The Wedderburn principal theorem in Banach algebras,* Proc. Amer. Math. Soc., **2** (1951), 771–777.

5. I. Gelfand, D. Raikov and G. Šilov, *Commutative normed rings* (in Russian), Uspehi Mat. Nauk. (N. S.) 1, No. 2 (12) (1946), 48–146; Amer. Math. Soc. Translations (series 2), **5** (1957), 115–220.

6. G. A. Heuer, *Algebraic extensions of Banach algebras,* Thesis (1958), University of Minnesota.

7. G. A. Heuer and J. A. Lindberg, *Algebraic extensions of continuous function algebras,* Proc. Amer. Math. Soc., **14** (1963), 337–342.

8. N. Jacobsen, *Structure of Rings,* Amer. Math. Soc. Coll. Publ. no. 37, Providence (1956).

9. I. Kaplansky, *Ring isomorphisms of Banach algebras,* Canad. J. Math., **6** (1954), 374–381.

10. J. A. Lindberg, *Factorization of Polynomials over Banach algebras,* to appear in Trans. Amer. Math. Soc.

11. L. H. Loomis, *An Introduction to Abstract Harmonic Analysis,* Van Nostrand, New York (1953).

12. C. E. Rickart, *General Theory of Banach Algebras,* Van Nostrand, New York (1960).

13. G. Šilov, *Decomposition of a commutative normed ring* (in Russian), Mat. Sbornick N.S., 32 (74) (1953), 353–364; Amer. Math. Soc. Translations (series 2), **5** (1955), 37–48.

14. B. L. Van Der Waerden, *Moderne Algebra,* Vol. 1 (3rd Ed.), Springer, Berlin (1950).

15. O. Zariski and P. Samuel, *Commutative Algebra,* Vol. 1, Van Nostrand, New York (1958).

YALE UNIVERSITY AND
SYRACUSE UNIVERSITY

ON THE DIOPHANTINE EQUATION $Cx^2 + D = y^n$

W. LJUNGGREN

1. Introduction. Let C, D and n denote odd positive integers, $D > 1$ and CD without any squared factor > 1. Let $K = Q(\sqrt{-CD})$, where Q is the field of rational numbers. Let further h denote the number of classes of ideals in K and put $D + (-1)^{(D+1)/2} = 2^m \cdot D_1$, $(D_1, 2) = 1$. In two previous papers [4] and [5] I have proved the following three theorems concerning the diophantine equation $Cx^2 + D = y^n$:

I. The diophantine equation

$$(1) \qquad Cx^2 + D = y^n , \qquad n > 1$$

is impossible in rational integers x and y if $h \not\equiv 0 \pmod n$, m is odd and either $CD \equiv 1 \pmod 4$ or $CD \equiv 3 \pmod 8$ with $n \not\equiv 0 \pmod 3$.

II. The diophantine equation

$$(2) \qquad Cx^2 + D = y^q , \qquad q > 3$$

where q denote an odd prime and $CD \not\equiv 7 \pmod 8$, is impossible in rational integers x and y if $h \not\equiv 0 \pmod q$, m is even and $q \not\equiv CD_1 \pmod 8$.

III. If $D \equiv 1 \pmod 4$, $CD \not\equiv 7 \pmod 8$ and m is even, then the equation (2) has only a finite number of solutions in natural numbers x, y and primes q if $CD_1 \equiv 5 \pmod 8$ or if $C = 1$ with $D_1 \equiv 3 \pmod 8$ for given C and D. The possible values of y and an upper limit for the number of primes q may always be determined after a finite number of arithmetical operations.

From the proofs it immediately follows that these theorems also hold good if $CD \equiv 7 \pmod 8$, *provided y is an odd integer.* This gives a far-reaching extension of results obtained by D. J. Lewis in his paper [2]. Putting $C = 1$, $D = 7$ we find, from 1:

The diophantine equation $x^2 + 7 = y^z$, $z > 1$, is impossible in rational integers x, y and z if y is an odd integer.

Equations of the type (1) have also been studied by T. Nagell [6], [8], [9] and B. Stolt [11].

Received July 2, 1963.

2. The equation $Cx^2 + 4D = y^n$, y odd.

THEOREM 1. *Let n be the power of a prime $q > 3$, and suppose that $h \not\equiv 0 \pmod{n}$. Then the diophantine equation*

$$(3) \qquad Cx^2 + 4D = y^n , \qquad n > 1 , \ y \text{ odd}$$

has no solutions in rational integers x, y if $q \not\equiv 3C(-1)^{(C-1)/2} \pmod{8}$. Likewise, if $D \equiv 0 \pmod{q}$, equation (3) has no integral solution.

Proof. We put $n = q^{\alpha}$. The principal ideals

$$[Cx + 2\sqrt{-CD}] \quad \text{and} \quad [Cx - 2\sqrt{-CD}]$$

have the greatest common ideal divisor $[C, \sqrt{-CD}]$, because $[C] = [C, \sqrt{-CD}]^2$, y is an odd integer and $(x, y) = 1$. From (3) it then follows

$$[Cx + 2\sqrt{-CD}] = [C, \sqrt{-CD}] \cdot \mathfrak{i}^{q^{\alpha}} ,$$

where $\mathfrak{i}$ denotes an ideal of the field $Q(\sqrt{-CD})$. Further we get

$$(4) \qquad [Cx + 2\sqrt{-CD}]^2 = [C] \cdot \mathfrak{i}_1^{q^{\alpha}} (\mathfrak{i}_1 = \mathfrak{i}^2) .$$

If the class number h is divisible by q^{β} $(0 \leqq \beta < \alpha)$ and not by $q^{\beta+1}$, there exist two rational integers f and g such that

$$fq^{\alpha} - gh = q^{\beta} .$$

Then by (4) we get the following equivalence

$$\mathfrak{i}_1^{q^{\beta}} \sim \mathfrak{i}_1^{fq^{\alpha}} \sim 1 .$$

Hence we obtain the ideal equation

$$(5) \qquad [Cx + 2\sqrt{-CD}]^2 = [C] \cdot \left[\frac{1}{2}(u + v\sqrt{-CD})\right]^{q^{\alpha-\beta}}$$

where u and v are rational integers, $u \equiv v \pmod{2}$. Since $q > 3$ all the units in the field $Q(\sqrt{-CD})$ are q^{th} powers. Then it follows from (5)

$$(6) \quad (Cx + 2\sqrt{-CD})^2 = C\left(\frac{1}{2}(u_1 + v_1\sqrt{-CD})\right)^q , \qquad u_1 \equiv v_1 \pmod{2} .$$

By means of (6) we derive

$$\frac{1}{2}(u_1 + v_1\sqrt{-CD}) = \left(\frac{1}{2}(a_1\sqrt{C} + b_1\sqrt{-D})\right)^2 , \qquad a_1 \equiv b_1 \pmod{2}$$

Inserting this expression in (6) we get

$$(7) \qquad x\sqrt{C} + 2\sqrt{-D} = \left(\frac{1}{2}(a_2\sqrt{C} + b_2\sqrt{-D})\right)^q , \quad a_2 \equiv b_2 \pmod 2 .$$

Equating the coefficients of $\sqrt{-D}$ we obtain the relation

$$(7') \qquad 2^{q+1} = \sum_{r=0}^{(q-1)/2} \binom{q}{2r+1} a_2^{q-1-2r} b_2^{2r+1} C^{[(q-1)/2]-r}(-D)^r ,$$

whence $b_2 = \pm 2^s$, $0 \leqq s \leqq q+1$.

Equation (7′) gives modulo q

$$b_2^q(-D)^{(q-1)/2} \equiv 2^{q+1} \pmod q ,$$

or

$$b_2\left(\frac{-D}{q}\right) \equiv 4 \pmod q , \qquad \text{i.e.}$$

$$b_2 \equiv \pm 4 \pmod q .$$

For $q > 5$ b_2 and a_2 must be even numbers, so that we have

$$(8) \qquad x\sqrt{C} + 2\sqrt{-D} = (a\sqrt{C} + b\sqrt{-D})^q .$$

If $q = 5$ and $b_2 = \pm 1$ it follows from (7′) that

$$D^2 \pm 8 = 5\left(\frac{1}{2}(Ca^2 - D)\right)^2 ,$$

which is impossible mod 8. Equation (8) is then valid if $q > 3$. Corresponding to (7′) we get

$$(8') \qquad 2 = \sum_{r=0}^{(q-1)/2} \binom{q}{2r+1} (Ca^2)^{[q-1)/2]-r} b^{2r+1}(-D)^r$$

Equation (8′) is impossible if q divides D. If $(D, q) = 1$ it follows from (8′)

$$2 \equiv b^q(-D)^{(q-1)/2} \equiv b\left(\frac{-D}{q}\right) \pmod q ,$$

whence

$$b = 2\left(\frac{-D}{q}\right) .$$

Inserting this expression for b in (8′) we obtain

$$(9) \qquad \left(\frac{-D}{q}\right) = \sum_{r=0}^{(q-1)/2} \binom{q}{2r+1} (Ca^2)^{[(q-1)/2]-r}(-4D)^r .$$

At first we want to prove that (9) *is impossible if* $q \equiv 1 \pmod 4$.

Treating (9) as a congruence mod 4 we find

$$\left(\frac{-D}{q}\right) = 1 .$$

Suppose now that $q-1$ is divisible by 2^{δ}, but not by $2^{\delta+1}$, $\delta \geqq 2$. Equation (9) may be written

$$(10)\quad 1 - q + q(1 - (Ca^2)^{(q-1)/2}) = \sum_{r=1}^{(q-1)/2} \binom{q}{2r+1} (Ca^2)^{[(q-1)/2]-r}(-4D)^r .$$

The general term in the right-hand side in (10) we then prefer to give the following shape

$$(11)\qquad \frac{q(q-1)}{2r(2r+1)} \cdot 2^{2r} \cdot \binom{q-2}{2r-1} (Ca^2)^{[(q-1)/2]-r} \cdot (-D)^r .$$

Here the numerator is divisible by $2^{\delta+2r}$. The denominator is divisible by a power of 2 which is $\leqq 2$. Since for all $r \geqq 1$ $2^{2r} > 2r$, we conclude that the integer (11) is divisible at least by $2^{\delta+1}$. Hence equation (10) is impossible, because $(Ca^2)^{(q-1)/2} - 1$ is divisible at least by $2^{\delta+1}$, while $q-1$ is divisible by 2^{δ} but not by $2^{\delta+1}$.

It remains to consider the case $q \equiv 3 \pmod 4$. From (9) it then follows

$$\left(\frac{-D}{q}\right) \equiv qC \pmod 4 ,$$

whence

$$(12)\qquad \begin{aligned} \left(\frac{-D}{q}\right) &= -1 \text{ for } C \equiv 1 \pmod 4 , \\ \left(\frac{-D}{q}\right) &= +1 \text{ for } C \equiv 3 \pmod 4 . \end{aligned}$$

Treating (9) as a congruence mod 8, we get

$$(13)\qquad \left(\frac{-D}{q}\right) \equiv qC + 4 \pmod 8$$

Combining (12) and (13) we find

$$q \equiv 3C(-1)^{(C-1)/2} \pmod 8$$

which was to be proved.

Remark. Theorem 1 remains true if $q = 3$, provided $CD \not\equiv 3 \pmod 8$: All units in $Q(\sqrt{-CD})$ are still q^{th} powers, such that equation (7) also holds good for $q = 3$. Since $b_2 \equiv \pm 4 \pmod q$, we have in addition to consider the cases $b_2 = \pm 1$ and $b_2 = \pm 2$. If $b_2 = \pm 1$

we deduce from (7) that $D = 3Ca_2^2+16$, which implies $CD \equiv 3 \pmod 8$, a contradiction. If $b_2 = \pm 2$, a_2 must be even. Putting $a_2 = 2a_3$ we find $D = 3Ca_3^2+2$ and $y = 4Ca_3^2+2$. But we assumed y to be an odd integer, and then our assertion is proved.

We now proceed to prove two lemmas.

LEMMA 1. *Putting*

$$S_1 = \sum_{r=0}^{[(n-1)/4]} \binom{n}{4r+1} \quad \text{and} \quad S_2 = \sum_{r=0}^{[(n-3)/4]} \binom{n}{4r+3} \tag{15}$$

we have if $n \equiv 3 \pmod 8$

$$S_1 \equiv 0 \pmod 3, \qquad S_2 \equiv 1 \pmod 3, \tag{16}$$

and if $n \equiv 7 \pmod 8$

$$S_1 \equiv 1 \pmod 3, \qquad S_2 \equiv 0 \pmod 3. \tag{17}$$

Proof. Inserting $x = 1$ and $x = i$ in the identity

$$\frac{1}{2x}((1+x)^n - (1-x)^n) = \binom{n}{1} + \binom{n}{3}x^2 + \binom{n}{5}x^4 + \cdots,$$

we get

$$2^{n-1} = S_1 + S_2,$$

and

$$2^{(n-1)/2}\cdot(-1)^{(n-3)/4} = S_1 - S_2, \qquad n \equiv 3 \pmod 4,$$

from which (16) and (17) easily follow.

LEMMA 2. *Equation* (9) *is impossible for* $q > 3$ *if*

$$D \equiv (-1)^{(C+1)/2} \pmod 3, \tag{18}$$

and besides one of the three following conditions is satisfied:

$$\begin{aligned} &1^\circ \quad C \equiv 0 \pmod 3 \\ &2^\circ \quad C \equiv \pm 1 \pmod 8 \\ &3^\circ \quad C \equiv \pm 3 \pmod 8 \quad \textit{and} \quad C \equiv (-1)^{(C-1)/2} \pmod 3. \end{aligned} \tag{19}$$

Proof. If $a \equiv 0 \pmod 3$ or if $C \equiv 0 \pmod 3$ it follows from (9) and (12) that

$$(-1)^{(C+1)/2} \equiv -(4D)^{(q-1)/2} \equiv -D \pmod 3, \quad \text{because } D^2 \equiv 1 \pmod 3.$$

But this contradicts condition (18).

If $a^2 \equiv 1 \pmod 3$, $C \not\equiv 0 \pmod 3$ we find

$$(-1)^{(\sigma+1)/2} = \binom{q}{1}C - \binom{q}{3}D + \binom{q}{5}C - \binom{q}{7}D + - \cdots,$$

or

$$(-1)^{(\sigma+1)/2} \equiv CS_1 - DS_2 \pmod 3 . \tag{20}$$

The congruence $C \equiv \pm 1 \pmod 8$ may be written $C \equiv (-1)^{(\sigma-1)/2} \pmod 8$. By Theorem 1 we then conclude $q \equiv 3C(-1)^{(\sigma-1)/2} \equiv 3 \pmod 8$. According to Lemma 1 it follows from (20)

$$(-1)^{(\sigma+1)/2} \equiv -D \pmod 3 ,$$

a contradiction.

The congruence $C \equiv \pm 3 \pmod 8$ is equivalent to $C \equiv 3(-1)^{(\sigma+1)/2} \pmod 8$. By means of Theorem 1 we conclude

$$q \equiv 3C(-1)^{(\sigma-1)/2} \equiv 7 \pmod 8 ,$$

and Lemma 1 then gives

$$(-1)^{(\sigma+1)/2} \equiv C \pmod 3 ,$$

which contradicts the second part of the condition 3°.

Our lemma is proved.

THEOREM 2. *Let C, D, n and h be defined as before, $h \not\equiv 0 \pmod n$. If $D \equiv (-1)^{(\sigma+1)/2} \pmod 3$ and if further one of the conditions* (19) *is satisfied, then the diophantine equation*

$$Cx^2 + 4D = y^n , \qquad n > 1,\ y \text{ odd} \tag{21}$$

has no solutions in rational integers x and y, provided $n \not\equiv 0 \pmod 3$ in case $CD \equiv 3 \pmod 8$.

Proof. Suppose that (21) is solvable in integers x, y, where y is odd. There must exist a prime factor q of n with the following property: q^α is a factor of n but not of the class number h. We put $m = q^\alpha$, $n = mr$ and $z = y^r$. Then the equation

$$Cx^2 + 4D = z^m \tag{22}$$

should be solvable in integers x and z. But this is impossible on account of Lemma 2 and the remark to Theorem 1.

EXAMPLE. *The equation $3x^2 + 28 = y^n$, $n \geqq 3$, has no solutions in rational integers x, y with y odd.*

Here is $C = 3$, $D = 7 \equiv 1 \pmod 3$ and $CD \equiv 5 \pmod 8$. Putting $x = 2x_1$, $y = 2y_1$ we get $3x_1^2 + 7 = 2^{n-2}y_1^n$, which implies $n = 3$, because $3x_1^2 + 7 = 2 \pmod 4$. Equation $3x_1^2 + 7 = 2y_1^3$ has at least the solutions $x_1 = \pm 9$, $y_1 = 5$.

3. The equation $x^2 + 4D = y^n$, y **odd.** In this section we restrict ourselves to the simple case $C = 1$. According to Theorem 1 and the remark attached to this it will be sufficient to deal with the case $q \equiv 3 \pmod 8$, $q = 3$ included. Putting

$$\lambda = a + 2\sqrt{-D} \qquad \text{and} \qquad \lambda' = a - 2\sqrt{-D}$$

it follows from (8), with $b = 2(-D/q) = 2(-1)^{(Q+1)/2} = -2$:

$$\frac{\lambda^q - \lambda'^q}{\lambda - \lambda'} = -1 \,. \tag{23}$$

The following identity is easily verified:

$$\frac{\lambda^{(q-1)/2} - \lambda'^{(q-1)/2}}{\lambda - \lambda'} \cdot (\lambda^{(q+1)/2} + \lambda'^{(q+1)/2}) = -(\lambda\lambda')^{(q-1)/2} + \frac{\lambda^q - \lambda'^q}{\lambda - \lambda'} \,. \tag{24}$$

Since $q = 8t + 3$, (24) may be written

$$\frac{\lambda^{4t+1} - \lambda'^{4t+1}}{\lambda - \lambda'} (\lambda^{4t+2} + \lambda'^{4t+2}) = -(a^2 + 4D)^{4t+1} - 1 \,. \tag{25}$$

The second factor on the left-hand side of (25) is divisible by $(\lambda^2 + \lambda'^2)/2 = a^2 - 4D$. Suppose now $a^2 - 4D > 0$. Since $a^2 - 4D \equiv 5 \pmod 8$, this number contains at least one prime factor $p \equiv 7 \pmod 8$ or $p \equiv 5 \pmod 8$. By means of (25) we derive that the Legendre symbol $((-a^2 - 4D)/p) = -1$, which implies $(-2/p) = 1$, i.e. $p = 8t + 1$ or $8t + 3$, contrary to the assumption. We therefore conclude $a^2 - 4D < 0$, or

$$a^2 < 4D \,. \tag{26}$$

These considerations yield the following theorem:

THEOREM 3. *Let $D > 1$ denote an odd positive integer without any squared factor > 1. If the class number of $Q(\sqrt{-D})$ is indivisible by the odd prime q, then the diophantine equation*

$$x^2 + 4D = y^q \,, \qquad y \text{ odd} \tag{27}$$

has no solutions in rational integers if $q \not\equiv 3 \pmod 8$. If $q \equiv 3 \pmod 8$, then (27) has only a finite number of solutions in rational integers x and y and primes q for given D. The possible values of y and an upper limit for the number of primes q may always *be determined after a finite number of arithmetical operations.*

That an upper limit for the number of primes may be determined, follows as a consequence of a theorem due to Th. Skolem [10]. However, in special cases it will be more convenient to use other methods.

Example 1. $x^2 + 28 = y^q$. We have $h = 1$ and must examine the case $q \equiv 3 \pmod 8$. The inequality (26) gives the possibilities:

$a^2 = 1$, $a^2 = 9$ and $a^2 = 25$. The corresponding values of y^q are 29, 37 and 53 respectively.

We make now use of the formula

$$(x + y)^q - x^q - y^q = qxy(x + y)(x^2 + xy + y^2)^r \cdot Q(u, v) ,$$

where $q > 3$ and

$$u = (x^2 + xy + y^2)^3 , \qquad v = (xy(x + y))^2 ,$$
$$r = 2 \quad \text{for } q \equiv 1 \pmod 3$$

and $r = 1$ for $q \equiv 2 \pmod 3$, and $Q(u, v)$ is a polynomial in u and v with integral coefficients [1]. Putting $x = \lambda$, $y = -\lambda'$, we obtain

$$(\lambda - \lambda')^{q-1} - \frac{\lambda^q - \lambda'^q}{\lambda - \lambda'} = -q\lambda\lambda'(\lambda^2 - \lambda\lambda' + \lambda'^2)^r \cdot Q(u, v) ,$$

or

$$(28) \qquad (16D)^{q'} \equiv 1 \pmod{q \cdot (a^2 + 4D) \cdot (a^2 - 12D)} , \qquad q' = \frac{1}{2}(q - 1) .$$

If $a^2 = 1$ we get $112^{q'} \equiv 1 \pmod{29}$, or $2^{q-1} \equiv -1 \pmod{29}$. Since $2^{14} \equiv -1 \pmod{29}$ and $2^s \not\equiv -1 \pmod{14}$ for $0 \leqq s < 14$, we must have $q \equiv 1 \pmod{14}$, which implies $(q/7) = 1$. From (28) we further find $112^{q'} \equiv 1 \pmod q$, i.e.

$$1 = \left(\frac{112}{q}\right) = \left(\frac{7}{q}\right) = -\left(\frac{q}{7}\right) ,$$

a contradiction.

If $a^2 = 9$ we get $112^{q'} \equiv 1 \pmod 5$, or $2^{q'} \equiv 1 \pmod 5$, which is impossible for $q = 8t + 3$.

If $a^2 = 25$ we obtain $112^{q'} \equiv 1 \pmod{53}$, or $6^{q'} \equiv 1 \pmod{53}$. Now 6 belongs to the exponent 26 mod 53, which is impossible since q' is an odd number.

It then remains $q = 3$, where

$$x + 2\sqrt{-7} = (a - 2\sqrt{-7})^3 ,$$

whence $2 = 56 - 6a^2$, i.e. $a^2 = 9$, $x = 225$ and

$$225^2 + 28 = 37^3 .$$

We have then proved:

The diophantine equation $x^2 + 28 = y^z$; $z > 3$ *and odd, has no solutions in integers* x, y *and* z *if* y *is an odd integer. If* $n = 3$ *there are exactly two solutions, namely* $x = \pm 225$ *and* $y = 37$.

This is a comprehensive generalization of a result obtained by D. J. Lewis [2].

Example 2. $x^2 + 12 = y^q$. Here is $h = 1$, and (26) gives $a^2 = 1$ or $a^2 = 9$. The last possibility must be excluded, giving $y \equiv 0 \pmod 3$. If $q > 3$ it follows from (27)

$$48^{q'} \equiv 1 \pmod{13} ,$$

or

$$2^{q-1} \equiv -1 \pmod{13}$$

implying $q \equiv 1 \pmod 6$, or $(q/3) = 1$. But according to (12) $(-3/q) = -1$, or $(q/3) = 1$, a contradiction. It is further known that $x^2 + 12 = y^3$ has no integral solution. This may be shown in the following manner:
1° y odd. We write our equation in the form

$$x^2 + 4 = (y - 2)(y^2 + 2y + 4)$$

Since $(x, 2) = 1$, all prime factors of $x^2 + 4$ must be of the form $4t + 1$. Consequently, $y \equiv 3 \pmod 4$. But this implies that $y^2 + 2y + 4 \equiv 3 \pmod 4$, which clearly is impossible.

2° y even. Then x must be even, and putting $x = 2x_1$, $y = 2y_1$ we get

$$x_1^2 + 3 = 2^{q-2}y_1^q$$

which is impossible modulo 8, because $q \neq 4$.
Then we have proved:

The diophantine equation $x^2 + 12 = y^n$, $n > 1$ *and odd, has no solutions in rational integers* x *and* y.

4. The equation $Cx^2 + DM^2 = y^n$, y **odd,** $(x, y) = 1$. Let M denote any positive integer, such that $(C, M) = 1$. In order to find criteria for the solvability of the equation

$$(29) \qquad Cx^2 + DM^2 = y^n , \quad n > 1, \ y \text{ odd and } (x, y) = 1 ,$$

similar to those obtained in the previous sections, we are again led to deal with an expression of the type

$$(30)\qquad x\sqrt{C}+M\sqrt{-D}=\left(\frac{1}{2}(a_2\sqrt{C}+b_2\sqrt{-D})\right)^q, \qquad a_2\equiv b_2 \pmod 2,$$

q denoting an odd prime. From (29) it follows

$$(31)\qquad 2^q\cdot M=\sum_{r=0}^{(q-1)/2}\binom{q}{2r+1}a_2^{q-1-2r}\cdot b_2^{2r+1}\cdot C^{(q-1)/2}\cdot(-D)^r .$$

It is easily seen that

$$(32)\qquad b_2 \mid M .$$

If $(Db, q)=1$, we find, treating (31) as a congruence

$$2M\equiv\left(\frac{-D}{q}\right)b_2 \pmod q ,$$

from which we conclude

$$(33)\qquad q\mid 2M\pm b_2 .$$

According to (32) and (33) there are only a finite number of possibilities for b_2 and for the primes q if $b_2\neq 2M(-D/q)$. It then only remains to consider the case

$$b_2=2M\left(\frac{-D}{q}\right),$$

where (30) can be written

$$(34)\qquad x\sqrt{C}+\sqrt{-DM^2}=(a\sqrt{C}+b\sqrt{-DM^2})^q ,$$

and

$$b=\left(\frac{-D}{q}\right).$$

But now we can utilize the results obtained for $M=1$.

Example.

$$x^2+63=y^n, \quad y \text{ odd}, \ n>1.$$

If $(x, y)=1$ we solve

$$x+3\sqrt{-7}=\left(\frac{a_2+b_2\sqrt{-7}}{2}\right)^q .$$

Here we have $y=(a_2^2+7b_2^2)/4$, i.e. a_2 and b_2 are even integers because y is odd. This gives

$$x + 3\sqrt{-7} = (a + b\sqrt{-7})^q \tag{35}$$

with $b = \pm 1$ or $b = \pm 3$. It is obvious that $q \neq 7$, such that $3 \equiv b(-7/q) \pmod q$. This implies $b^2 = 1$. For $q = 3$ equation (35) is impossible mod 9. Then we must have $b = 3(-7/q)$. Since y is odd, a must be even, and from (34) we conclude $(-7/q) = 1$ and

$$1 = \binom{q}{1}a^{q-1} - \binom{q}{3}a^{q-3}\cdot 7\cdot 3^2 + \cdots + \binom{q}{q}(-7)^{(q-1)/2}\cdot 3^{q-1} . \tag{36}$$

Since $q \equiv 1 \pmod 3$ and $a^2 \equiv 1 \pmod 3$, it can be shown that (36) is impossible, exactly in the same way as we earlier proved the impossiblity of (10), exchanging only the prime 2 by the prime 3. Our equation is then impossible if $(x, y) = 1$. If $(x, y) = 3$ we get, putting $x = 3x_1$, $y = 3y_1$

$$x_1^2 + 7 \equiv 3^{n-2}y^n \equiv 0 \pmod 3 ,$$

which is impossible. Then we have proved:

The diophantine equation $x^2 + 63 = y^n$ is impossible in integers x, y if y is odd and $n > 1$ is an odd number.

5. Remark on earlier results.

The diophantine equation

$$ax^2 + bx + c = dy^n , \tag{37}$$

where the left-hand side is an irreducible polynomial of the second degree, having integral coefficients and d is an integer $\neq 0$, has only a finite number of solutions in rational integers x, y when $n \geqq 3$. This was first shown by A. Thue and later on by Laundau and Ostrowski. See for instance [7]. However, no general method is known for determining all integral solutions x and y for a given equation of the form (37).

Equation (1) was solved completely by T. Nagell in case y odd, C arbitrary and $D = 1, 2$ or 4 [9]. Nagell has also examined equation 1 when $C = 1$ and D a square-free integer congruent to 1 or 2 modulo 4, but the results obtained are far from being complete [6]. He has further found interesting theorems concerning the equation $x^2 + 8D = y^n$, $(D, 2) = 1$ [8]. The first complete solution of the equation $x^2 + 2 = y^n$ was given by Ljunggren [3]. An upper bound for the number of solutions of (1), in terms of D and n, was derived by Stolt [11]. It must be emphasized that we in this note have deduced *bounds which are independent of n*. For other equations of the type (1) see [9].

If y is odd, but the classnumber h is divisible by n, we have to deal with *irreducible* binary forms of degree $n \geqq 3$. This occurs also if y is even. The problem of representation of rational integers by such forms is not solved. For the determination of an upper bound for the number of solutions of our equations in these cases compare [2], p. 1075.

References

1. L. E. Dickson, *History of the theory of numbers*, Vol. II, N. Y. (1952), p. 748.

2. D. J. Lewis, *Two classes of diophantine equations*, Pacific J. Math., **11** (1961), 1063-1076.

3. W. Ljunggren, *Über einige Arcustangensgleichungen die auf interessante unbestimmte Gleichungen führen*, Arkiv för Mat., Astronomi och Fysik, v. 29 A, N:o 13, Stockholm (1943), 1-11.

4. ———, *On a diophantine equation*, Kong. Norske Videnskabers Selskab Forhandl. v. 18, Nr. 32, Trondheim (1945), 125-128.

5. ———, *New theorems concerning the diophantine equation* $Cx^2 + D = y^n$, Kong. Norske Videnskabers Selskab Forhandl., v. 29, Nr. 1, Trondheim (1956), 1-4.

6. T. Nagell, *Sur l'impossibilité de quelques équations à deux indeterminées*, Norsk Mat. Forenings Skrifter, Ser. I. Nr. 13, Oslo (1923).

7. ———, *L'analyse indeterminée à degré superieure*, Mem. Sci. Math., Fasc 39, Paris 1929.

8. ———, *On the diophantine equation* $x^2 + 8D = y^n$, Arkiv för Mat., v. 3, nr. 6, Stockholm (1955), 103-112.

9. ———, *Contributions to the theory of a category of diophantine equations of the second degree with two unknowns*, Nova Acta Reg. Soc. Upsal. IV Ser. 16, Uppsala 1955, p. 1-38.

10. Th. Skolem, *Ein Verfahren zur Behandlung gewisser exponentialer Gleichungen und diophantischer Gleichungen*, 8de Skand. Mat. Kongr. Förh., Stockholm (1934), 163-188.

11. B. Stolt, *Über einen verallgemeinerten Fermatschen Satz*, Acta aritmetica, **5** (1959), 267-276.

ATOMIC ORTHOCOMPLEMENTED LATTICES

M. DONALD MACLAREN

Introduction. The lattice of all closed subspaces of a separable Hilbert space has the following properties. It is complete, atomic, irreducible, semi-modular, and orthocomplemented. The primary purpose of this paper is to investigate lattices with these properties.

If L is such a lattice, there is a representation theorem for L. The elements in L of finite dimension or finite deficiency form an orthocomplemented modular lattice. It follows that if the dimension of L is high enough, then there is a dual pair of vector spaces U and W such that L is isomorphic to the lattice of W closed subspaces of U. Because L is orthocomplemented the spaces U and W are isomorphic. This isomorphism establishes a "semi-inner product" on U, and L may be described as being the lattice of closed subspaces of a semi-inner product space.

The contents of the paper are as follows. Section 1 contains some definitions and establishes notation. Section 2 is concerned with the completion of an orthocomplemented lattice and § 3 with the center of such a lattice. With the exception of Theorem 3.2 the techniques used in §§ 2 and 3 are standard, and many of the results are widely known. To the best of the author's knowledge, however, the theorems have not previously appeared in print. Therefore we state and prove them in some detail. The representation theorem and other results centering about the semi-modularity condition are proved in § 4. With the other conditions holding for L, semi-modularity is equivalent to certain covering conditions. Because this is not true for arbitrary complete atomic lattices, the results seem to be of some interest. Finally, in § 5, semi-inner product spaces are discussed. A theorem is given relating the existence of a semi-inner product on U to the existence of an orthocomplemented lattice of subspaces of U. This is an easy generalization of a theorem of Birkhoff and von Neumann [4] (Appendix). In two other theorems we investigate the exact relation between the semi-inner product on U and the orthocomplemented lattice L.

1. Definitions and some elementary lemmas. Let S be a partially-ordered set. If a and b are elements of S, we denote the least upper bound or join of a and b by $a \vee b$, provided that the join exists. We denote the greatest lower bound or meet of a and b by ab, provided

Received June 7, 1963. The results of this paper have been drawn from the author's doctoral thesis, Harvard 1962.

that the meet exists. If A is a subset of S which has a least upper bound, we denote the least upper bound by $\bigvee A$. If the elements of A are indexed by a set J, we may also write $A = \bigvee_j a_j$. If A has a greatest lower bound, it is denoted by $\bigwedge A$ or $\bigwedge_j a_j$.

The symbols $\bigcup$ and $\bigcap$ will be used to denote set union and set intersection respectively.

If a and b are elements of a partially ordered set S with $a \leqq b$, we will denote the set of all $x \in S$ such that $a \leqq x \leqq b$ by $[a, b]$.

A partially-ordered set S is said to be *orthocomplemented* if it contains at least element 0 and a greatest element 1, $1 \neq 0$, and if there exists a map $a \to a'$ of S onto itself which satisfies

(1.1) $a \leqq b$ implies $a' \geqq b'$,

(1.2) $a'' = a$,

(1.3) a' is a complement of a, i.e., $aa' = 0$ and $a \vee a' = 1$.

The mapping $a \to a'$ is called an orthocomplementation, and a' is called the orthocomplement of a.

Two elements a and b of S are said to be *orthogonal* if $a \leqq b'$. In this case we write $a \perp b$. The relation of being orthogonal is obviously symmetric.

We will use the following simple lemmas throughout this paper.

LEMMA 1.1. *Let S be an orthocomplemented partially-ordered set, and let $\{a_j\}$ be a subset of S such that $\bigvee_j a_j$ exists. Then $\bigwedge_j a_j'$ exists, and $(\bigvee_j a_j)' = \bigwedge_j a_j'$.*

LEMMA 1.2. *Let A be a subset of an orthocomplemented partially-ordered set S, and suppose that $\bigvee A$ exists. Then if $b \perp a$ for all $a \in A$, $b \perp \bigvee A$.*

An *isomorphism* of a partially-ordered set S onto a partially-ordered set R is a one-to-one mapping θ from S onto R, such that $\theta(x) \leqq \theta(y)$ if and only if $x \leqq y$. An isomorphism preserves any meets and joins which exist. When S and R are orthocomplemented we will say an isomorphism θ is an *ortho-isomorphism* if $\theta(x') = \theta(x)'$ for all x in S.

LEMMA 1.3. *Let S and R be orthocomplemented lattices, and let θ be a one-to-one map of S onto R. Then θ is an ortho-isomorphism if and only if* (1) *$\theta(x') = \theta(x)'$ for all x in S and* (2) *$\theta(xy) = \theta(x)\theta(y)$ for all x and y in S or $\theta(x \vee y) = \theta(x) \vee \theta(y)$ for all x and y in S.*

Let a and b be elements in a partially-ordered set S. a is said to *cover* b if $a > b$, and there does not exist c in S with $a > c > b$. If S has a least element 0, an *atom* is an element of S which covers

0. A lattice S is *atomic* if every element of S is the join of some set of atoms.

2. Completion of an orthocomplemented partially-ordered set. A partially-ordered set S is said to be *complete* if $\bigwedge A$ and $\bigvee A$ exist for all subsets A of S. If S has both a least element and a greatest element, then $\bigvee A$ exists for all subsets A of S if and only if $\bigwedge A$ exists for all subsets A of S. The standard method for embedding a partially-ordered set S in a complete lattice is to use the completion of S by cuts.[1] If S is orthocomplemented, the completion can be constructed in another way by using the orthogonality relation. This is just the construction used in the standard proof that the completion of a Boolean algebra is a Boolean algebra.

In Theorems 2.1 and 2.2 we show that the partial ordering in an orthocomplemented partially-ordered set can be found if one knows only which elements are orthogonal. This fact suggests that we define an abstract notion of an orthogonality relation.

Let S be any set. We will say that the binary relation $\perp$ is an *orthogonality relation* if it has the following properties.

(1) $a \perp b$ implies $b \perp a$.

(2) $a \perp a$ implies $a \perp b$ for all b in S.

(3) ($c \perp a$ if and only if $c \perp b$) implies $a = b$.

THEOREM 2.1. *If S is a set with an orthogonality relation ($\perp$), then a partial ordering ($\leqq$) may be defined on S: $a \leqq b$ if and only if $d \perp b$ implies $d \perp a$.*

Proof. If $a \leqq b$ and $b \leqq a$, then $d \perp a$ if and only if $d \perp b$. Therefore $a = b$, by the definition of an orthogonality relation. If $a \leqq b$ and $b \leqq c$, $a \leqq c$ by the definition of $\leqq$ in S.

THEOREM 2.2. *If S is an orthocomplemented partially-ordered set, then the relation $\perp$, where $a \perp b$ if and only if $a \leqq b'$, is an orthogonality relation. Further the partial ordering induced by this orthogonality relation coincides with the original partial ordering.*

Proof. The relation $\perp$ is symmetric, because $a \leqq b'$ if and only if $b \leqq a'$. If $a \perp a$, $a \leqq a'$. Hence $a \leqq aa' = 0$, and $0 \perp x$ for all x in S. Finally suppose that a and b are elements of S such that $c \perp a$ implies $c \perp b$. Then $a' \perp b$, i.e., $b \leqq a$. Thus the relation $\perp$ induces the original partial ordering in S. Further if a and b are such that $d \perp a$ if and only if $d \perp b$, we have $b \leqq a$ and $a \leqq b$, i.e., $a = b$.

---

[1] See Birkhoff [**3**], Ch. 4, sec. 7.

We will now assume that S is a set with an orthogonality relation $(\perp)$, and that the partial ordering of Theorem 2.1 has been defined on S. S may be an orthocomplemented partially-ordered set, but it does not have to be. For any subset A of S let $A^{\perp}$ be the set of all x in S such that $x \perp a$ for all a in A. Let $A^{-} = A^{\perp\perp}$, and call a subset *closed* if $A = A^{-}$.

We will use the following simple lemma throughout this paper. Its proof uses only familiar arguments.

LEMMA 2.1. *Let S be a set with an orthogonality relation. Let A and B be subsets of S. Then the following relations hold.*

(1) *If* $A \subseteqq B$, $B^{\perp} \subseteqq A^{\perp}$

(2) $A^{\perp} = A^{\perp\perp\perp}$

(3) $A \subseteqq A^{-}$

(4) $A^{--} = A^{-}$

(5) *If* $A \subseteqq B$, $A^{-} \subseteqq B^{-}$

(6) $(A \cup B)^{\perp} = A^{\perp} \cap B^{\perp}$.

An orthogonality relation is just a special type of polarity as defined by Birkhoff.[2] Thus in the following theorem the assertion that the closed subsets of S form a complete orthocomplemented lattice follows from Theorem 9 and Corollary Ch. 4, of [3]. Since the rest of the proof is quite standard, we omit it.

THEOREM 2.3. *Let S be a set with an orthogonality relation and with the partial ordering induced by the orthogonality relation. Then the closed subsets of S, partially ordered by inclusion, form a complete lattice $L(S)$. If $\{A_j\}$ is a family of closed subsets, $\bigwedge_j A_j$, the meet of the A_j in $L(S)$, is just $\bigcap_j A_j$. The mapping $A \rightarrow A^{\perp}$ is an orthocomplementation in $L(S)$. Further there exists a one-to-one mapping of S into $L(S)$ which preserves orthogonality, order, and all existing meets in S. If S is orthocomplemented this map also preserves orthocomplements and all joins existing in S.*

From now on we will always use $L(S)$ to denote the lattice of closed subsets of S. The following theorem justifies our calling $L(S)$ the completion of S.

THEOREM 2.4. *If S is an orthocomplemented partially-ordered set, $L(S)$ is the completion of S by cuts.*

Proof. If A is a subset of S, let $A^{*} = \{x \in S \mid x \geqq a$ for all a in $A\}$ and let $A^{\circ} = \{x \in S \mid x \leqq a$ for all a in $A\}$. The completion by cuts

[2] [3], Ch. 4, sec. 5.

of S is the lattice of all subsets A of S such that $A^{*\circ} = A$[3]. We need only show that $A^{\perp\perp} = A^{*\circ}$ for all subsets A of S. Let $y \in A^*$. Then $y \geqq a$ for all a in A, and hence $y' \leqq a'$ for all a in A. This means y' is in $A^\perp$. Therefore if $x \in A^{\perp\perp}$, $x \perp y'$, i.e., $x \leqq y$. Thus if $x \in A^{\perp\perp}$, $x \leqq y$ all y in A^*, i.e., $x \in A^{*\circ}$. This proves that $A^{\perp\perp} \subseteqq A^{*\circ}$. Now if $y \in A^\perp$, $y' \geqq a$ for all a in A, i.e., $y' \in A^*$. Thus $x \in A^{*\circ}$ implies $x \leqq y'$ all $y \in A^\perp$, i.e., $x \in A^{\perp\perp}$. Therefore $A^{*\circ} \subseteqq A^{\perp\perp}$; $A^{*\circ} = A^{\perp\perp}$.

We will say that a subset I of a partially-ordered set S is *join dense* if every element of S is the join, perhaps infinite, or elements in I. The advantage of using the orthogonality relation to construct the completion of S is that only a join-dense subset of S is actually needed for the construction.

THEOREM 2.5. *Let S be an orthocomplemented partially-ordered set, and let I be a join-dense subset of S. Let $\perp$ be the orthogonality relation in S. Then restricted to I, $\perp$ is an orthogonality relation. Further the partial ordering induced in I by the orthogonality relation $\perp$ coincides with the partial ordering inherited from S. Finally $L(S)$ and $L(I)$ are ortho-isomorphic.*

Proof. (2.1) and (2.2) in the definition of an orthogonality relation are obviously satisfied, because $\perp$ is an orthogonality relation in S. Let a and b be elements of I such that for $c \in I$, $c \perp a$ if and only if $c \perp b$. Let d be an element of S such that $d \perp a$. Since I is join-dense in S, there exist c_j in I such that $d = \bigvee_j c_j$. For each c_j, $c_j \perp a$. Hence $c_j \perp b$. Therefore in S $d \perp b$. Similarly $d \perp b$ implies $d \perp a$. Therefore $a = b$. This proves that (2.3) is satisfied by the relation restricted to I and thus proves that it is an orthogonality relation on I. Now let $\leqq$ be the partial ordering in S, and let $\prec$ be the partial ordering induced in I by the orthogonality relation. If x, y, z are all in I with $x \leqq y$ and $z \perp y$, then $z \perp x$. Thus if x and y are in I with $x \leqq y$ we have $x \prec y$. Now suppose that $x \prec y$. Since I is join-dense $y' = \bigvee_j z_j$ with $z_j \in I$. Clearly $z_j \perp y$ for each j, and hence $z_j \perp x$. Therefore $x \perp \bigvee_j z_j = y'$, i.e. $x \leqq y$. This proves that the two partial orderings $\leqq$ and $\prec$ are the same. To complete the proof of the theorem we must prove that $L(S)$ and $L(I)$ are ortho-isomorphic. We first show that if A is a closed subset of S, $A \cap I$ is closed in I. Note that if B is any set closed in S, and if $x \perp y$ for all $y \in B \cap I$, then $x \perp y$ for all $y \in B$, because I is join dense. In other words, if $B \in L(S)$, $(B \cap I)^\perp = B^\perp$. Now the closure of $A \cap I$ in I is $((A \cap I)^\perp \cap I)^\perp \cap I$. If $A \in L(S)$, we have $((A \cap I)^\perp \cap I)^\perp \cap I = (A^\perp \cap I)^\perp \cap I = A^{\perp\perp} \cap I = A \cap I$, i.e. $A \cap I$ is closed in I. Now for

[3] See [**3**], Ch. 4., Sec. 7.

any subset B of S, let B^- denote the closure of B in S. We next show that if B is closed in I, $B^- \cap I = B$. Suppose that $x \in B^- \cap I$. Then $x \perp y$ for all $y \in B^\perp$. In particular $x \perp y$ for all $y \in B^\perp \cap I$. Therefore $x \in B$, because B is closed in I. Clearly $B \subseteqq B^- \cap I$, so we have $B = B^- \cap I$. Now define a map θ from $L(S)$ to $L(I)$ by $\theta(A) = A \cap I$. We have shown above that if $A \in L(S)$, $A \cap I$ is in $L(I)$. If $B \in L(I)$, $\theta(B^-) = B^- \cap I = B$, so θ is onto. If $A \in L(S)$, $(A \cap I)^- = A$, because I is join dense in S. Therefore θ is one-to-one. Clearly $\theta^{-1}(B) = B^-$, and therefore $\theta(A) \leqq \theta(B)$ if and only if $A \leqq B$. Thus θ is an isomorphism. Finally note that the orthocomplementation in $L(I)$ is $A \to A^\perp \cap I$. But $\theta(A^\perp) = A^\perp \cap I = (A \cap I)^\perp \cap I$. Therefore $\theta(A^\perp) = (\theta(A))^\perp \cap I$. Thus θ preserves orthocomplements.

3. Center of an orthocomplemented lattice. Let $S_j (j \in J)$ be a family of lattices. Let P be the Cartesian product of the S_j, i.e., P is the set of all functions f from J to $\bigcup_j S_j$ such that $f(j) \in S_j$ for all j in J. P has a natural partial ordering: $f \leqq g$ in P if and only if $f(j) \leqq g(j)$ for all j in J. It is easy to verify that this ordering makes P into a lattice. Meets and joins are: $(fg)(j) = f(j)g(j)$ and $(f \vee g)(j) = f(j) \vee g(j)$. P is sometimes called the cardinal product of the S_j, but we will follow von Neumann and call P the direct sum of the S_j. We will write $P = \Sigma \oplus S_j$. We will denote the direct sum of S_1 and S_2 by $S_1 \oplus S_2$. $S_1 \oplus S_2$ may be regarded as the set of all ordered pairs (x_1, x_2) with $x_1 \in S_1$ and $x_2 \in S_2$. The following Theorem is obvious.

THEOREM 3.1. *Let S_j be a family of lattices. Then $P = \Sigma \oplus S_j$ is orthocomplemented if and only if each S_j is orthocomplemented. P is complete if and only if each S_j is complete.*

Now suppose that P is ortho-isomorphic to the direct sum of two orthocomplemented lattices, $P \cong S_1 \oplus S_2$. Let a be the element of P corresponding to $(1, 0)$. Then a' corresponds to $(0, 1)$, S_1 is ortho-isomorphic to $[0, a]$, and S_2 is ortho-isomorphic to $[0, a']$. In this case it will be convenient to write $P = [0, a] \oplus [0, a']$. The *center* of P is the set of all elements a such that $P = [0, a] \oplus [0, a']$. The elements 0 and 1 are always in the center. If the center of P contains only 0 and 1, will say that P is *irreducible*. The next theorem is suggested by a similar result of von Neumann on the center of a continuous geometry.

THEOREM 3.2. *Let P be an orthocomplemented lattice. Then for an element a of P the following three conditions are equivalent.*

(1) $x = xa \vee xa'$ *for all x in P.*

(2) *$(x \vee a)y = xy \vee ay$ for all x and y in P.*
(3) *a is in the center of P.*

Proof. That (1) and (2) hold for central elements is well known. Suppose that a has property (1). Let $x \leqq a$ and $y \leqq a'$. Then $x' \geqq a'$ and $y' \geqq a$, so $x'y'a' = y'a'$ and $x'y'a = x'a$. Thus $x'y' = x'y'a \vee x'y'a' = x'a \vee y'a'$. Taking orthocomplements, we get that for $x \leqq a$ and $y \leqq a'$, $(x \vee y) = (x \vee a')(y \vee a)$. Now define a map θ from P to $[0, a] \oplus [0, a']$ by $\theta(x) = (xa, xa')$. If $\theta(x) = \theta(y)$, $xa = ya$ and $xa' = ya'$. Hence by (1) $x = xa \vee xa' = ya \vee ya' = y$. Thus θ is one-to-one. Let (x, y) be an element of $[0, a] \oplus [0, a']$. Then $x \leqq a$ and $y \leqq a'$. $x' \vee a' = x'a \vee x'a' \vee a' = x'a \vee a'$. Taking orthocomplements, we get $xa = (x \vee a')a$. As was shown above, $(x \vee y) = (x \vee a')(y \vee a)$. Thus $(x \vee y)a = (x \vee a')(y \vee a)a = (x \vee a')a = xa$. Similarly $(x \vee y)a' = y$, i.e., $\theta(x \vee y) = (x, y)$. It is now clear that $\theta^{-1}((x, y)) = x \vee y$, and that $\theta(x) \leqq \theta(y)$ if and only if $x \leqq y$. Thus θ is an isomorphism, and obviously it is an ortho-isomorphism. $P = [0, a] \oplus [0, a']$, i.e., a is in the center. Now if a has property (2), $(a \vee a')x = ax \vee a'x$ for all x in P. Thus a has property (1); a is in the center.

THEOREM 3.3. *If P is a complete, atomic orthocomplemented lattice, the center of P is a complete, atomic Boolean algebra.*

Proof. To prove that the center is a complete Boolean algebra, we need only show that for any subset A of the center $\bigvee A$ is in the center. Let $b = \bigvee A$, and let p be an atom such that $pb = 0$. Then $pa = 0$ for all a in A. Therefore $p' \geqq a$ for all a in A, because $p = pa \vee pa'$ for all a in A. Thus $p' \geqq b$, i.e., $p \leqq b'$. If $pb \neq 0, p \leqq b$, because p is an atom. Thus for every atom $p, p \leqq b$ or $p \leqq b'$. Because P is atomic this means $x = xb \vee xb'$ for all x in P, i.e., b is in the center. To show that the center is atomic, let p be any atom in P, let A be the set of all central elements a such that $p \leqq a$, and let $b = \bigwedge A$. Then b is in the center, and $b \neq 0$. Further b must be an atom of the center, for if not there exists c in the center such that $0 < c < b$. Then $p = pb = pbc \vee pbc'$, so either $p \leqq c$ or $p \leqq c'b$. Thus either $c \in A$ or $c'b \in A$, so either $b \leqq c$ or $b \leqq c'b$. This contradicts the assumption that $0 < c < b$.

LEMMA. *Let L be a complete orthocomplemented lattice, and let a be in the center of L. Then for any family $\{x_j\}$ of elements in L, $a(\bigvee_j x_j) = \bigvee_j (ax_j)$.*

Proof. $a(\bigvee_j x_j) = a(\bigvee_j (x_ja \vee x_ja')) = a(\bigvee_j (x_ja) \vee \bigvee_j (x_ja'))$
$= a(\bigvee_j (x_ja)) \vee a(\bigvee_j (x_ja')) = \bigvee_j (x_ja)$.

THEOREM 3.4. *Let L be a complete, atomic, orthocomplemented lattice. Then L is ortho-isomorphic to the direct sum of irreducible, atomic, orthocomplemented lattices.*

Proof. Let $\{a_j\}$ be the set of all atoms of the center. Let $S_j = [0, a_j]$. Define a mapping θ from L to $\Sigma \oplus S_j$ by $\theta(x)(j) = xa_j$. Let y be in $\Sigma \oplus S_j$, and let $x = \bigvee_j y(j)$. Then $\theta(x)(k) = (\bigvee_j y(j))a_k = \bigvee_j (y(j)a_k) = y(k)$. Thus $\theta(x) = y$; θ is onto. Let p be an atom, and a_j be an atom of the center. Then $p = pa_j$, or $p \perp a_j$. Since $\bigvee_j a_j = 1, p = \bigvee_j (pa_j)$. Since L is atomic it follows that $x = \bigvee_j (xa_j) = \bigvee_j \theta(x)(j)$. Therefore θ is one-to-one. Clearly $\theta(x) \leqq \theta(y)$ if and only if $x \leqq y$, so θ is an isomorphism. Further $\theta(x)'(j) = (xa_j)'a_j = x'a_j = \theta(x')(j)$. Thus $\theta(x') = \theta(x)'$. To complete the proof we need only show that each S_j is irreducible. Suppose that $0 \leqq b \leqq a_j$, and that b is in the center of S_j. Then for $x \leqq a_j, x = xb \vee xb'a_j$. Hence for all x in L,

$$x = xa_jb \vee xa_jb' \vee xa_j' = xb \vee (a_jb' \vee a_j')x = xb \vee xb' .$$

Thus b is in the center of L. Since a_j is an atom of the center, this means that $b = 0$ or $b = a_j$. This proves that $[0, a_j]$ is irreducible.

4. Semi-modular, atomic, orthocomplemented lattices.

Let S be an atomic lattice. Let $x_0 < x_1 < \cdots < x_n$ be a finite chain of elements in S. We will call the integer n the *length* of chain. The chain is a *covering* chain if x_{i+1} covers x_i for $i = 0, 1, \cdots$. We define a function d on the set of ordered pairs (x, y) of elements in S with $x \leqq y$ as follows. If there exists a finite covering chain connecting x and y, $d(x, y)$ is the length of the shortest such covering chain. If no such covering chain exists, $d(x, y) = \infty$.

We will call an element $x \in S$ finite if x is the join of a finite number of atoms. Clearly if $d(0, x)$ is finite, then x is finite. We will let $F(S)$ denote the set of $x \in S$ such that x is finite or x' is finite.

THEOREM 4.1. *Let S be an atomic lattice such that if a and b are finite elements of S which both cover ab, then $a \vee b$ covers a and b. Then the set of all finite elements of S is an ideal. For any finite elements $a \leqq b, d(a, b)$ is finite, and all covering chains connecting a and b have the same length.*

Proof. We first show that if p is an atom, $d(0, a)$ is finite, and $pa = 0$, then $p \vee a$ covers a. We will prove this by induction on $d(0, a)$. If $d(0, a) = 1$, a and p both cover $ap = 0$. Therefore $a \vee p$

covers a and p. Suppose the statement is true for $d(0, a) \leqq n$. Let $d(0, a) = n + 1$, and let p be an atom with $pa = 0$. We need only prove that $p \vee a$ covers a. Because $d(0, a) = n + 1$, there exists b such that a covers b, and $d(0, b) = n$. Now $p \vee b$ covers b, and $(p \vee b)a = b$. Therefore $(p \vee b) \vee a$ covers a, i.e., $p \vee a$ covers a. Next we show that if a is finite $d(0, a)$ is finite. For finite a let $N(a)$ be the smallest number N such that a is the join of N atoms. We will prove this lemma by induction on $N(a)$. Clearly if $N(a) = 1$, $d(0, a) = 1$. Suppose the statement is true for $N(a) \leqq n$, and let a be a finite element with $N(a) = n + 1$. There exists atoms $p_1, \cdots, p_{n+1}$ such that $a = p_1 \vee \cdots \vee p_{n+1}$. Let $b = p_1 \vee \cdots \vee p_n$. Then $b < a$, and $N(b) = n$. Therefore $d(0, b)$ is finite, and $a = p_{n+1} \vee b$ covers b. Therefore $d(0, a) \leqq d(0, b) + 1$, i.e., $d(0, a)$ is finite. Now it follows from the above that $d(a, b)$ is finite if a and b are finite with $a \leqq b$. To complete the proof of the theorem we need only prove the statement, "if $d(a, b) = n$, then all chains connecting a to b have length at most n." That the finite elements form an ideal follows immediately from this. We will prove the statement by induction on n. If $d(a, b) = 1$, b covers a and the statement is clearly true. Suppose the statement is true for $d(a, b) \leqq n$. Let $d(a, b) = n + 1$. Then there exists a covering chain $a < x_1 < \cdots < x_n < b$. Note that $d(x_1, b) = n$. Let $a < y_1 < \cdots < y_m < b$ be any chain connecting a to b. We need only prove that $m \leqq n$. If y_1 does not cover a, there exists an atom p such that $pa = 0$ and $p \leqq y_1$. Then $p \vee a$ covers a. Replacing y_1 by $p \vee a$ we get another chain of length m. Thus we may assume that y_1 covers a. If $y_1 = x_1$, we have $d(y_1, b) = d(x_1, b) = n$. Therefore by the inductive hypothesis $m \leqq n$. If $y_1 \neq x_1$, y_1 and x_1 both cover $y_1 x_1 = a$. Therefore $y_1 \vee x_1$ covers y_1 and x_1. Now let $y_1 \vee x_1 < w_1 < \cdots < w_k = b$ be any chain joining $y_1 \vee x_1$ to b. Then $x_1 < y_1 \vee x_1 < w_1 < \cdots < w_k$ joins x_1 to b. Since $d(x_1, b) = n$, $k + 1 \leqq n$, i.e., $k \leqq n - 1$. It follows that $d(y_1 \vee x_1, b) \leqq n - 1$. Since $y_1 \vee x_1$ covers y_1, this means that $d(y_1, b) \leqq n$. But $m \leqq d(y_1, b)$, so $m \leqq n$.

LEMMA 4.1. *Let S be an atomic orthocomplemented lattice. Then the following covering conditions on S are equivalent.*

(*) If a and b are in $F(S)$, and both cover ab, then $a \vee b$ covers both a and b.

(**) If a and b are in $F(S)$, and $a \vee b$ covers both a and b, then a and b both cover ab.

Proof. Suppose that (*) holds in S, that a and b are in $F(S)$, and that $a \vee b$ covers a and b. Then a' and b' are in $F(S)$, and both cover $a'b'$. Therefore by (*), $a' \vee b'$ covers a' and b'. Hence a and b cover ab. Thus (*) implies (**). A dual argument shows that (**) implies (*).

LEMMA 4.2. *Let S be an atomic orthocomplemented lattice in which the covering condition (*) holds. Then the finite elements of S form an atomic modular lattice.*

Proof. This follows immediately from Theorem 4.1, Lemma 4.1, and Theorem 3, Ch. 5 of Birkhoff [3].

Two elements (b, c) in a lattice are said to form a modular pair if for all $a \leqq c$, $(a \vee b)c = a \vee bc$. A lattice S is semi-modular if the relation of being a modular pair is symmetric in S. Two elements (a, b) form a d-modular pair if for all $c \geqq a$, $(a \vee b)c = a \vee bc$. S is dual semi-modular if the relation of being a d-modular pair is symmetric.

In general semi-modularity is stronger than the covering condition (*). We want to show that with one additional condition (*) implies semi-modularity. Our proof is suggested by the proofs of Theorems III–1 and III–6 of Mackey [5]. We introduce the following notation. If x is in the atomic orthocomplemented lattice S, $\mathscr{A}(x)$ is the set of all atoms p such that $p \leqq x$. $\mathscr{A}(x) + \mathscr{A}(y)$ is the set of all atoms p such that for some $q \in \mathscr{A}(x)$ and $r \in \mathscr{A}(y)$, $p \leqq q \vee r$. If X is a set of atoms, $X^{\perp}$ is the set of all atoms p such that $p \perp q$ for all q in X. It is easy to verify the rules $\mathscr{A}(x') = \mathscr{A}(x)^{\perp}$, $\mathscr{A}(xy) = \mathscr{A}(x) \cap \mathscr{A}(y)$, $(\mathscr{A}(x) + \mathscr{A}(y))^{\perp} = \mathscr{A}(x)^{\perp} \cap \mathscr{A}(y)^{\perp}$.

LEMMA 4.3. *Let S be an atomic orthocomplemented lattice in which the covering condition (*) holds. Assume further that if a and b are atoms in S with $a \neq b$, $a'(a \vee b) \neq 0$. Then if p is an atom in S, $\mathscr{A}(p \vee x) = \mathscr{A}(x) + \mathscr{A}(p)$ for all x in S.*

Proof. We need only show that $\mathscr{A}(x \vee p) \subseteqq \mathscr{A}(x) + \mathscr{A}(p)$. Let p be an atom with $px = 0$. First note that if q and r are atoms with $q \neq r$, then $p'(q \vee r) \neq 0$. This is immediate if $p \leqq q \vee r$. If $p \nleqq q \vee r$, let $c = p \vee q \vee r$. Let $t_1 = p'(p \vee q)$, and $t_2 = p'(p \vee r)$. Then $[0, c]$ is a modular lattice of length 3, $d(0, t_1 \vee t_2) = 2$, and $d(0, q \vee r) = 2$. Hence $(t_1 \vee t_2)(q \vee r) \neq 0$, which means $p'(q \vee r) \neq 0$. Now let s be any atom, y any element such that $y > ys'$, and r any atom in $\mathscr{A}(y)$ but not in $\mathscr{A}(ys')$. If $q \in \mathscr{A}(y)$, and $q \neq r$, then $x = s'(q \vee r)$ is in $\mathscr{A}(ys')$, and $q \leqq r \vee x$. Thus $\mathscr{A}(y) = \mathscr{A}(ys') + \mathscr{A}(r)$. Applying this to x' and p, we have $\mathscr{A}(x') = \mathscr{A}(x'p') + \mathscr{A}(r)$ for some r. Now $\mathscr{A}(x) = \mathscr{A}(x')^{\perp} = \mathscr{A}(x'p')^{\perp} \cap \mathscr{A}(r)^{\perp} = \mathscr{A}((x \vee p)) \cap \mathscr{A}(r') = \mathscr{A}((x \vee p)r')$. But $\mathscr{A}((x \vee p)) = \mathscr{A}((x \vee p)r') + \mathscr{A}(p)$, so $\mathscr{A}(x \vee p) = \mathscr{A}(x) + \mathscr{A}(p)$.

THEOREM 4.2. *Let S be an atomic orthocomplemented lattice satisfying the covering condition* (*). *Assume further that, if a and b are atoms in S, $a \neq b$, then $a'(a \vee b) \neq 0$.*[4] *Then if a is finite (a, x) is a d-modular pair for all x in S.*

Proof. We need only show that for $c \geqq a$, $(a \vee x)c \leqq a \vee xc$. It follows from Lemma 4.3 that $\mathscr{A}(x \vee a) = \mathscr{A}(x) + \mathscr{A}(a)$. Let $a \leqq c$, and let $p \in \mathscr{A}((a \vee x)c)$. Then $p \in \mathscr{A}((a \vee x))$, and hence $p \leqq q \vee r$ where $q \in \mathscr{A}(a)$ and $r \in \mathscr{A}(x)$. If $p = q$ or $p = r$, $p \in \mathscr{A}(xc) + \mathscr{A}(a)$. If p is different from q and r, then $r \leqq p \vee q \leqq c$. Thus $r \in \mathscr{A}(xc)$, which means $p \in \mathscr{A}(a) + \mathscr{A}(xc)$. This proves that $(a \vee x)c \leqq a \vee xc$.

THEOREM 4.3. *Let S satisfy the hypotheses of Theorem* 4.2. *Then $F(S)$ is an atomic, orthocomplemented, modular lattice.*

Proof. We need only show that $F(S)$ is modular. Let a, b, c be in $F(S)$ with $a \leqq c$. If a is finite, $(a \vee b)c = a \vee bc$ by the preceding theorem. Otherwise a' is finite, which means c' is finite. Then we have $[(a \vee b)c]' = (c' \vee b')a' = c' \vee b'a'$. Thus $(a \vee b)c = a \vee bc$.

Let V be a left vector space over a division ring R. Let V^* be the space of all linear functions from V to R. If W is a total subspace of V^*, i.e., $f(x) = 0$ for all $f \in W$ implies $x = 0$, then we say that V, W is a dual pair. If X is a subspace of V, let $X' = \{f \in W: f(x) = 0 \text{ for all } x \in X\}$. Similarly define Y' for Y a subspace of W. Then we say that a subspace X of V is W-closed if $X = X''$.

McLaughlin [6] has given a representation theorem for the completion by cuts of a complemented modular point lattice. Since a complete atomic orthocomplemented lattice S is the completion by cuts of $F(S)$ we can apply the theorem to obtain:

THEOREM 4.4. *Let S be a complete, irreducible, atomic, orthocomplemented lattice in which the covering condition* (*) *holds. Assume further that the $d(0, 1) \geqq 4$, and that if p and q are atoms with $p \neq q$, $p'(p \vee q) \neq 0$. Then there exist a pair of dual vector spaces U, W over a division ring D such that S is isomorphic to the lattice of W-closed subspaces of U.*

COROLLARY. *Let S be a complete, irreducible, atomic, orthocomplemented lattice. Then the following three statements about S are equivalent.*

(1) *S is semi-modular*

(2) *If p is an atom in S and $pa = 0$ then $p \vee a$ covers a.*

[4] Note that this condition holds if L is weakly modular, i.e., if (a, a') is a d-modular pair for all a in L.

(3) *Covering condition* (*) *holds in* S; *and if* p *and* q *are atoms with* $p \neq q$, *then* $p'(p \vee q) \neq 0$.

Proof. It is well known that (1) implies (2) and (2) implies (*). Suppose (2) holds and that p and q are atoms with $p \neq q$. Then $p \vee p'q'$ covers $p'q'$. But $q < p \vee q$, so $p'q' < q'$, which shows that 1 does not cover $p'q'$. Therefore, $p \vee p'q' \neq 1$, i.e. $p'(p \vee q) \neq 0$. This proves that (2) implies (3). Now suppose that (3) holds. If $d(0, 1)$ is finite, S is actually modular (Theorem 4.3). If $d(0, 1)$ is infinite, we can apply the theorem above. Mackey ([5], Theorem III–6) has shown that in such a lattice of closed subspaces the relation of being a d-modular pair is symmetric. Since S is orthocomplemented, this means that the relation of being a modular pair is also symmetric.

THEOREM 4.5. *The completion of a semi-modular atomic orthocomplemented lattice is semi-modular.*

Proof. If S is semi-modular the covering condition (*) holds in $F(S)$. Also if p is an atom and $px = 0$, $p \vee x$ covers x. If p and q are atoms, $p \neq q$, p' covers $p'q'$. Hence $p \vee p'q' < 1$, which gives $p'(p \vee q) \neq 0$. Let $L(S) = \Sigma \oplus R_j$ be the direct sum decomposition of the completion $L(S)$ into irreducible components. Since $F(L(S))$ and $F(S)$ are ortho-isomorphic, the covering condition (*) and the condition $p'(p \vee q) \neq 0$ hold in each R_j. If the dimension of R_j is finite, R_j is actually modular. If the dimension of R_j is infinite, R_j satisfies the hypotheses of Theorem 4.4, so R_j is semi-modular. Thus $L(S)$ is the direct sum of semi-modular lattices; $L(S)$ is semi-modular.

5. Semi-Inner Product Spaces. Let V be a left vector space over a division ring R. A *semi-bilinear functional* B on V is a map $(x, y) \to B(x, y)$ of $V \times V$ into R such that

(1) for all x_1, x_2, y_1, and y_2 in V and α in R, $B(\alpha x_1 + x_2, y_1 + y_2) = \alpha B(x_1, y_1) + \alpha B(x_1, y_2) + B(x_2, y_1) + B(x_2, y_2)$, and

(2) There exists an anti-automorphism θ of R such that for all x and y in V and α in R, $B(x, \alpha y) = B(x, y)\theta(\alpha)$. We will say that a semi-bilinear functional B is a *semi-inner product* if it satisfies the following conditions.

(1) The anti-automorphism θ associated with B is involutory.

(2) $B(x, y) = \theta(B(y, x))$ for all x and y.

(3) $B(x, x) = 0$ implies $x = 0$.

(4) For some x $B(x, x) = 1$.

A left vector space together with a semi-inner product will be called a *semi-inner product space.*

If V is a semi-inner product space, define an orthogonality relation in V by $x \perp y$ if and only if $B(x, y) = 0$. If X is a subset of V define

$X^\perp$ just as in §2. It is easy to verify that $X^\perp$ is always a subspace. The orthocomplemented lattice of all closed subspaces of V will be denoted by $L(V)$.

THEOREM 5.1. *Let V be a left vector space over a division ring R. Then V is a semi-inner product space if and only if there exists a dual space W such that V, W is a dual pair and the lattice, S, of all W-closed subspaces of V is orthocomplemented.*

Proof. Suppose that W exists and that S is orthocomplemented. Let R^* be the ring which is identical with R as an additive group and in which multiplication ($\circ$) is $\alpha \circ \beta = \beta\alpha$. Then W is a left vector space over R^*. If $\lambda \in R^*, f \in W$, $(\lambda f)(x) = f(x)\lambda$. For $x \in V$, let $[x]$ denote the one-dimensional subspace spanned by x. Let $[x]^*$ be the one-dimensional subspace of W spanned by those linear functionals whose nullspaces is $[x]^\perp$. In an obvious way one verifies that $[x] \to [x]^*$ is a one-to-one map of the one-dimensional subspaces of V onto those of W which preserve linear dependence and independence. Hence there exists a semi-linear transformation T from V onto W such that $[x]^* = [T(x)]$. Clearly $[x] \perp [y]$ if and only if $T(y)(x) = 0$. Thus if $x_0 \neq 0$, $T(x_0)(x_0) \neq 0$.

Let φ be the isomorphism from R to R^* associated with T. Then φ may also be regarded as an anti-automorphism of R. Let θ be the inner automorphism of R: $\beta \to (T(x_0)(x_0))\beta(T(x_0)(x_0))^{-1}$. Let $B(x, y) = (T(y)(x))(T(x_0)(x_0))^{-1}$. It is a matter of routine to verify that B is a semi-inner product with anti-automorphism $\sigma = \theta \circ \varphi$. If A is any finite-dimensional subspace of V containing x_0, B defines an orthocomplementation in the lattice of all subspaces of A, and $B(x_0, x_0) = 1$. Hence by proposition 1, page 110 of Baer [1], the anti-automorphism σ associated with B is involutory and $\sigma(B(x, y)) = B(y, x)$, for all x and y in A. Thus B is a semi-inner product. Now suppose that B is a semi-inner product on V. For x in V let f_x be the member of $V^*: f_x(y) = B(y, x)$. It is clear that the set W of all such f_x is a total subspace of V^* and that $L(V)$ is identical with the lattice of W-closed subspaces of V.

Suppose B and B' are two semi-inner products on V which determine the same orthogonality relation. Then there exists α in R such that $B(x, y) = B'(x, y)\alpha$ for all x and y in V[5]. It is quite possible, however, to have two semi-inner products on V, which are not equivalent in this way, but whose associated lattices are ortho-isomorphic. Our last two theorems explore this possibility.

THEOREM 5.2. *Let V_1 be a semi-inner product space over a division*

[5] Baer [**4**], page 105, Proposition 3.

ring R. Let B_1 be the semi-inner product in V_1, let θ be the anti-automorphism associated with B_1, let σ be an automorphism of R, and let τ be an inner automorphism of R. Then there exists a semi-inner product space V_2 over R whose semi-inner product B_2 has anti-automorphism $\varphi = \tau \circ \sigma \circ \theta \circ \sigma^{-1}$ such that $L(V_1)$ and $L(V_2)$ are orthoisomorphic.

Proof. Let x_j ($j \in J$, where J is some indexing set) be a maximal set of nonzero mutually orthogonal vectors in V_1. For $y \in V_1$, let $T(y)$ be the function from J to R such that $T(y)(j) = \sigma(B_1(y, x_j))$. Let V_2 be the set of all such functions $T(y)$. It is clear that V_2 is a left vector space over R, and that T is a semi-linear transformation with automorphism σ from V_1 onto V_2. Further T is one-to-one. For if $T(y) = 0$, then $y \perp x_j$ all $j \in J$; and this means $y = 0$, because $\{x_j\}$ was a maximal orthogonal set. Let the inner automorphism τ be $\tau(\beta) = \alpha^{-1}\beta\alpha$. For f and g in B_2, let $B_2(f, g) = \sigma(B_1(T^{-1}f, T^{-1}g))\alpha$. It is easy to verify that B_2 is a semi-inner product. We include only the proof that $B_2(f, \beta g) = B_2(f, g)\varphi(\beta)$. We have

$$\begin{aligned} B_2(f, \beta g) &= \sigma(B_1(T^{-1}f, T^{-1}(\beta g)))\alpha = \sigma(B_1(T^{-1}f, \sigma^{-1}(\beta)(T^{-1}g)))\alpha \\ &= \sigma(B_1(T^{-1}f, T^{-1}g)\theta(\sigma^{-1}(\beta)))\alpha \\ &= \sigma(B_1(T^{-1}f, T^{-1}g))\sigma(\theta(\sigma^{-1}(\beta)))\alpha \\ &= B_2(f, g)\alpha^{-1}\sigma(\theta(\sigma^{-1}(\beta)))\alpha = B_2(f, g)\varphi(\beta) . \end{aligned}$$

Since $B_2(f, g) = 0$ if and only if $B_1(T^{-1}f, T^{-1}g) = 0$, it is clear that T induces an ortho-isomorphism between the $L(V_1)$ and $L(V_2)$.

THEOREM 5.3. *Let V_1 and V_2 be semi-inner product spaces of dimension greater than two, over division rings R_1 and R_2 respectively, such that $L(V_1)$ and $L(V_2)$ are ortho-isomorphic. Let B_1 and B_2 be the semi-inner products in V_1 and V_2 respectively, and let θ and φ be the associated anti-automorphisms. Then there exists an isomorphism σ from R_1 onto R_2 and a semi-linear transformation T from V_1 to V_2 with isomorphism σ such that T induces the lattice isomorphism. Further there exists an inner automorphism τ of R_2 such that $\varphi = \tau \circ \sigma \circ \theta \circ \sigma^{-1}$.*

Proof. Since $L(V_1)$ and $L(V_2)$ are isomorphic, the lattice of all finite-dimensional subspaces of V_1 is isomorphic to the lattice of all finite dimensional subspaces of V_2. It follows from this that the lattice of all subspaces of V_1 is isomorphic to the lattice of all subspaces of V_2. Therefore the isomorphism σ and the semi-linear transformation T exists. To prove the final assertion, let x be a vector in V_1 such that $B_1(x, x) = 1$. Let y be a nonzero vector in V_1 with $y \perp x$.

Let $x' = T(x)$, and $y' = T(y)$. Since T induces the lattice ortho-isomorphism, $x' \perp y'$. Now for any $\lambda \in R$ with $\lambda \neq 0$,

$$x + \lambda y \perp x - \theta(\lambda^{-1})B_1(y, y)^{-1}y \ .$$

Therefore $T(x + \lambda y) \perp T(x - \theta(\lambda^{-1})B_1(y, y)^{-1}y)$, i.e.,

$$x' + \sigma(\lambda)y' \perp x' - \sigma(\theta(\lambda^{-1}))\sigma(B_1(y, y)^{-1})y' \ .$$

Therefore $B_2(x' + \sigma(\lambda)y', x' - \sigma(\theta(\lambda^{-1}))\sigma(B_1(y, y)^{-1})y') = 0$ for all $\lambda \neq 0$ in R_1. Since $x' \perp y'$, this gives

$$B_2(x', x') + \sigma(\lambda)B_2(y', y')\varphi(\sigma(B_1(y, y)^{-1}))\varphi(\sigma(\theta(\lambda^{-1}))) = 0$$

for all $\lambda \neq 0$ in R_1. Now let $\alpha = B_2(x', x')$. Taking $\lambda = 1$, we get $\alpha - B_2(y', y')\varphi(\sigma(B_1(y, y)^{-1})) = 0$. Thus $\alpha - \sigma(\lambda)\alpha\varphi(\sigma(\theta(\lambda^{-1}))) = 0$ for all $\lambda \neq 0$ in R_1. Let τ be the inner automorphism of R_2: $\tau(\beta) = \alpha^{-1}\beta\alpha$ for all β in R_2. Then taking $\lambda^{-1} = \sigma^{-1}(\beta)$, we get $\beta = (\tau^{-1}\circ\varphi\circ\theta\circ\sigma^{-1})(\beta)$ for all $\beta \neq 0$ in R_2, i.e., $\tau^{-1}\circ\varphi\circ\sigma\circ\theta\circ\sigma^{-1}$ is the identity automorphism of R_2. Since $\sigma\circ\theta\circ\sigma^{-1}$ is an involutory anti-automorphism of R_2, this gives $\varphi = \tau\circ\sigma\circ\theta\circ\sigma^{-1}$.

COROLLARY. *Let n be an integer greater than 2. Then there exist semi-inner product spaces V_1 and V_2 of dimension n over the complex numbers such that $L(V_1)$ and $L(V_2)$ are not ortho-isomorphic.*

Proof. There exists a real closed subfield K of the complex numbers C such that $K(i) = C$, and K is not isomorphic to the real numbers.[6] Let φ be the involutory automorphism of C which has K as its field of fixed points. Let θ be the usual conjugacy automorphism. Let V_1 be an n-dimensional Hilbert space over the complex numbers. Let V_2 be the set of all ordered n-tuples of complex numbers. Let $B((\alpha_1, \cdots, \alpha_n), (\beta_1, \cdots, \beta_n)) = \Sigma\alpha_i\varphi(\beta_i)$. It is easy to verify that B is a semi-inner product with anti-automorphism φ. If $L(V_1)$ and $L(V_2)$ were ortho-isomorphic, we would have $\varphi = \sigma\circ\theta\circ\sigma^{-1}$ for some automorphism σ of C. But this would mean that K was isomorphic to the field of fixed points of θ, i.e., the field of real numbers. This contradiction proves the corollary.

This corollary points up the fact that lattices $L(V_1)$ and $L(V_2)$ may be isomorphic without being ortho-isomorphic.

REFERENCES

1. E. Artin and O. Schreier, *Algebraische Konstruktion reeler Korper*, Abh. Math. Sem. Hamburg, **5** (1926), 83-115.

---

[6] The existence of such a subfield is proved in Artin and Schreier [**1**], Theorem II.

2. R. Baer, *Linear algebra and projective geometry*, New York, 1952.

3. G. Birkhoff, *Lattice Theory*, Amer. Math. Soc. Colloquim Publications, Vol. 25, rev. ed., New York, 1948.

4. G. Birkhoff and J. von Neumann, *The logic of quantum mechanics*, Ann. of Math., **37** (1936), 823-843.

5. G. W. Mackey, *On infinite-dimensional linear spaces*, Trans. Amer. Math. Soc., **57** (1945), 155-207.

6. J. E. McLaughlin, *The normal completion of a complemented modular point lattice*, Proceedings of Symposia in Pure Mathematics, Amer. Math. Soc., Vol. 2, pp. 78-80.

BOEING SCIENTIFIC RESEARCH LABORATORIES

TRANSFORMATIONS OF DOMAINS IN THE PLANE AND APPLICATIONS IN THE THEORY OF FUNCTIONS

MOSHE MARCUS

In this paper we shall consider a family of transformations S_n ($n = 1, 2, \cdots$) operating on open or closed sets in the complex plane z. S_n is defined relatively to a fixed point called the center of transformation, and it transforms an open set into a starlike domain which, for $n > 1$, is also n-fold symmetric with respect to this point. Therefore, for $n > 1$, S_n may be classified as a method of symmetrization. This method of symmetrization was already defined by Szegö [4] for domains which are starlike with respect to the center of transformation.

The definition of S_n will be extended (in the way usually used for symmetrizations) so that S_n will operate also on a certain class of functions and a family of condensers, in the plane. It will be proved that S_n diminishes the capacity of a condenser and this result will be used in order to obtain certain theorems in the theory of functions.

1. Definitions and notations. The transformations S_n are defined as follows.

DEFINITION 1. *Let Ω be an open set in the plane z, which does not contain the point at infinity, and let z_0 be a point of Ω. If $|z - z_0| < \rho$, $(0 < \rho)$, is a circle contained in Ω, we define:*

$$(1) \qquad L_\rho(\varphi) = \int_E \frac{dr}{r} ,$$

where $|z - z_0| = r$ and

$$E = \{z \mid z \in \Omega, |z - z_0| > \rho, \arg(z - z_0) = \varphi\} ;$$

$$(2) \qquad L_\rho^{(n)}(\varphi) = \frac{1}{n} \sum_{k=0}^{n-1} L_\rho\left(\varphi + \frac{2\pi k}{n}\right) ;$$

$$(3) \qquad \begin{cases} R(\varphi) = \rho \exp\{L_\rho(\varphi)\} \\ R^{(n)}(\varphi) = \left[\prod_{k=0}^{n-1} R\left(\varphi + \frac{2\pi k}{n}\right)\right]^{1/n} = \rho \exp\{L_\rho^{(n)}(\varphi)\} . \end{cases}$$

Evidently, $R^{(n)}(\varphi)$ does not depend on ρ.

Received March 28, 1963. This paper represents part of a thesis submitted to the Senate of the Technion, Israel Institute of Technology, in partial fulfillment of the requirements for the degree of Master of Science. The author wishes to thank Professor E. Netanyahu for his guidance and help in the preparation of this work.

Now, the set obtained from Ω by the transformation $S_n = S_n(z_0)$, with center z_0 is defined as follows:

$$S_n\Omega = \{z \mid z - z_0 = re^{i\varphi}, 0 \leqq r < R^{(n)}(\varphi), 0 \leqq \varphi < 2\pi\} . \tag{4}$$

If instead of Ω we have a compact set H, which has an interior point z_0, we define:

$$S_nH = \{z \mid z - z_0 = re^{i\varphi}, 0 \leqq r \leqq R^{(n)}(\varphi), 0 \leqq \varphi < 2\pi\} . \tag{4'}$$

It is easily verified that $S_n\Omega$ is a simply-connected domain and that S_nH is a connected compact set. Both sets are starlike with respect to z_0.

We shall extend the definition of S_n over a family of functions $\mathscr{G}$ which will now be defined. A non-constant real function $g(z)$ belongs to $\mathscr{G}$ if it is continuous over the extended plane z, if it takes its maximum value at infinity and if its minimum is assumed on a set of points, the interior of which is not empty. Let $g(z)$ be a function of $\mathscr{G}$ and let m and M be its minimum and maximum values, respectively. We define the following sets:

$$\begin{cases} G_m = \{z \mid g(z) = m\} , & \\ G_c = \{z \mid g(z) < c\} , & \text{for } m < c \leqq M . \end{cases} \tag{5}$$

G_c (for $m < c < M$) is an open bounded set while G_m is a compact set. Let z_0 be an interior point of G_m and suppose that the circle $|z - z_0| \leqq \rho$, $(0 < \rho)$, is contained in G_m. Denote by $L_\rho(c, \varphi)$, $L_\rho^{(n)}(c, \varphi)$, $R^{(n)}(c, \varphi)$ the functions defined by (1), (2), (3) with G_c replacing Ω. Clearly, for a fixed φ, $L_\rho(c, \varphi)$ is strictly monotonic increasing, for $m \leqq c \leqq M$. We also have:

$$\begin{cases} \lim_{c \to d^-} L_\rho(c, \varphi) = L_\rho(d, \varphi), & \text{for } m < d \leqq M ; \\ \lim_{c \to m} L_\rho(c, \varphi) = L_\rho(m, \rho) . & \end{cases} \tag{6}$$

Let $S_n = S_n(z_0)$. From these properties of $L_\rho(c, \varphi)$, it follows that:

$$S_nG_c \subset S_nG_d , \qquad \text{for } m \leqq c < d \leqq M ; \tag{7}$$

$$S_nG_c = \bigcup_{m \leqq d < c} S_nG_d , \qquad \text{for } m < c \leqq M ; \tag{8}$$

$$S_nG_m = \bigcap_{m < d < M} S_nG_d . \tag{9}$$

Since $\bar{G}_c \subseteqq \bigcap_{c<d<M} G_d$ we also have:

$$S_n\bar{G}_c \subseteqq \bigcap_{c<d<M} S_nG_d , \qquad m \leqq c < M . \tag{10}$$

DEFINITION 2. *Let $g(z) \in \mathscr{G}$. Using the notations introduced* *above, we define the function* $g^{(n)}(z)$ *obtained from* $g(z)$ *by the transformation* $S_n = S_n(z_0)$, *as follows*:

$$(11) \qquad S_n g \equiv g^{(n)}(z) = \begin{cases} \inf\{c \mid z \in S_n G_c\}, & \text{for } z \in S_n G_M, \\ M, & \text{otherwise}. \end{cases}$$

From (8) and (9) we now conclude:

$$(12) \qquad \begin{cases} S_n G_c = \{z \mid g^{(n)}(z) < c\}, & \text{for } m < c \leqq M, \\ S_n G_m = \{z \mid g^{(n)}(z) = m\}. \end{cases}$$

2. A lemma concerning the function $g^{(n)}(z)$.

LEMMA 1. *The function* $g^{(n)}(z)$ *is continuous over the extended plane* z. *If moreover* $g(z)$ *is Lip on every compact subset of* G_M[1] *then* $g^{(n)}(z)$ *is Lip on every compact subset of* $S_n G_M$.

Proof. We begin by proving the continuity of $g^{(n)}(z)$. If $z^* \in S_n G_m$ and $g^{(n)}(z^*) = d > m$ then by (10) and (12), the set $S_n G^*_{d+\varepsilon} - S_n \bar{G}^*_{d-\varepsilon}$ (where $m < d^* - \varepsilon < d^* + \varepsilon < M$) is an open neighbourhood of z^* in which $|g^{(n)}(z) - g^{(n)}(z^*)| \leqq \varepsilon$. If z^* belongs to $S_n G_m$ or z^* belongs to the complement of $S_n G_M$, then the set $S_n G_{m+\varepsilon}(m < m + \varepsilon < M)$, and the complement of $S_n \bar{G}_{M-\varepsilon}(m < M - \varepsilon < M)$ respectively, are open neighbourhoods of z^* in which $|g^{(n)}(z) - g^{(n)}(z^*)| \leqq \varepsilon$.

In order to prove the second assertion of the lemma it is sufficient to show that $g^{(n)}(z)$ is Lip on every set $S_n G_c(m < c < M)$. Without loss of generality we may suppose that $z_0 = 0$ and that $\rho = 1$. (And in this case we shall write $L^{(n)}(c, \varphi)$ instead of $L_1^{(n)}(c, \varphi)$.) We now map the z plane, cut along the positive real axis from zero to infinity, by a branch of $w = \log z$, $(w = u^+ iv)$, onto the strip $0 < v < 2\pi$. (The points of the positive real axis will be mapped both on $v = 0$ and $v = 2\pi$). We denote by H_c and H_c^n the images of G_c and $S_n G_c$ by this mapping, and we put $h(w) = g(e^w)$ and $h^{(n)}(w) = g^{(n)}(e^w)$.

Let c be a fixed number in the open interval (m, M). Since $g(z)$ is Lip on G_c, the function $h(w)$ is Lip on H_c, and if it is shown that $h^{(n)}(w)$ is Lip on H_c^n, the required result follows.

Since $h(w)$ is Lip on H_c, there exists a number $p > 0$ such that: $|h(w_1) - h(w_2)| \leqq p|w_1 - w_2|$, for any $w_1, w_2 \in H_c$.

We need the following assertion:
If δ is a positive number and v_1, v_2, c_1, c_2 are real numbers such that:

$$(13) \qquad |v_1 - v_2| < \delta,\ m < c_1 < c_2 - p\delta < c - p\delta,$$

[1] A function $g(z)$ is Lip on a set E if there exists a constant p, such that for any two points $z_1, z_2 \in E$, we have $|g(z_1) - g(z_2)| \leqq p|z_1 - z_2|$.

then

$$L^{(n)}(c_2, v_2) \geqq L^{(n)}(c_1, v_1) + [\delta^2 - (v_1 - v_2)^2]^{1/2} . \tag{14}$$

Because of the definition of $L^{(n)}(c, v)$, it is enough to prove (14) for $n = 1$. Without loss of generality we may suppose that $0 \leqq v_k < 2\pi$, $(k = 1, 2)$.

Denote by J_k the intersection of the half line $Im w = v_k$, $Re\, w \geqq 0$, with the set H_{c_k}, for $k = 1, 2$. The Lebesgue measure of J_k is $L(c_k, v_k)$. Using (5) and (13) the following is easily verified:

Let $w_1 \in J_1$. If $w_2 = u_2 + iv_2$, $u_2 \geqq 0$ and $|w_1 - w_2| \leqq \delta$, then $w_2 \in J_2$. From this and the fact that J_1 is bounded on the right, (14) follows for $n = 1$.

It will now be shown that

$$|h^{(n)}(w') - h^{(n)}(w'')| \leqq p\,|w' - w''| , \quad \text{for any } w', w'' \in H_c^n .$$

Suppose that there are two points w_1, w_2 in H_c^n for which this inequality does not hold, and let δ be a number such that:

$$|h^{(n)}(w_1) - h^{(n)}(w_2)| > p\delta > p\,|w_1 - w_2| . \tag{15}$$

Let $h^{(n)}(w_1) < h^{(n)}(w_2)$. Then we can find numbers c_1, c_2 such that:

$$m \leqq h^{(n)}(w_1) < c_1 < c_2 - p\delta < h^{(n)}(w_2) - p\delta < c - p\delta . \tag{16}$$

Now the numbers c_1, c_2, $v_1 = Im w_1$, $v_2 = Im w_2$ satisfy (13), and therefore inequality (14) holds. Since, for $m < c < M$,

$$H_c^n = \{w \mid 0 \leqq Im w \leqq 2\pi, h^{(n)}(w) < c\} = \{w \mid 0 \leqq v \leqq 2\pi, u < L^{(n)}(c, v)\} ,$$

it follows (by (16)) that $w_1 \in H_{c_1}^n$ and $w_2 \notin H_{c_2}^n$; hence $u_1 = Re\, w_1 < L^{(n)}(c_1, v_1)$ and $u_2 = Re\, w_2 \geqq L^{(n)}(c_2, v_2)$. These inequalities together with (14) yield $|w_1 - w_2| > \delta$, which is in contradiction to (15). This completes the proof of the lemma.

REMARK. The following is a consequence of the second part of the lemma: If $g(z)$ is Lip on every compact subset of $G_M - G_m$, then $g^{(n)}(z)$ is Lip on every compact subset of $S_n G_M - S_n G_m$.

3. **On a class of functions** (C, z_0). Let $C = (D, E_0, E_1)$ be a condenser in the complex plane z, i.e. a system consisting of a domain D and two disjoint closed sets E_0 and E_1, such that D does not contain the point at infinity, E_0 is bounded, E_1 is unbounded and the union of E_0 and E_1 is equal to the complement of D.

Suppose that E_0 contains an interior point z_0, let $z - z_0 = re^{i\varphi}$ and denote by S_φ the ray $\arg(z - z_0) = \varphi$. Then a subclass (C, z_0) of $\mathscr{G}$ is defined as follows.

A real function $g(z)$, continuous over the extended plane z, belongs to (C, z_0) if:

(i) $g(z)$ possesses continuous first partial derivatives, in D.

(ii) $g(z) \equiv 0$ in E_0, $g(z) \equiv 1$ in E_1 and $0 < g(z) < 1$ in D.

(iii) The set of points on the ray S_φ, at which $g(z)$ assumes a given value $c\,(0 < c < 1)$, is finite.

(iv) Any compact set of points on S_φ, which is contained in D, contains only a finite number of points (possibly zero) at which $\partial g(r, \varphi)/\partial r = 0$.

Suppose that the Dirichlet problem of the equation $\Delta u = 0$, with continuous boundary values, always has a solution in D. Then there exists a real function $\omega(z)$, continuous over the extended plane z, which is harmonic in D, vanishes on E_0 and assumes the value 1 on E_1. This function is the potential functions of C. Evidently, it belongs to (C, z_0).

Let $g(z) \in (C, z_0)$. Using property (iii) we find that (6) may be replaced by

$$\lim_{c \to c_0} L_\rho(c, \varphi) = L_\rho(c_0, \varphi), \qquad \text{for } 0 \leqq c_0 \leqq 1 . \tag{17}$$

Therefore in this case, the function $g^{(n)}(z) \equiv S_n(z_0)g$ may be defined in the following way:

$$g^{(n)}(z) = g^{(n)}(r, \varphi) = \begin{cases} 0, & \text{for } r \leqq R^{(n)}(0, \varphi), \\ c, & \text{for } r = R^{(n)}(c, \varphi),\ 0 < c < 1 , \\ 1, & \text{for } r \geqq R^{(n)}(1, \varphi) . \end{cases} \tag{18}$$

Since, for a fixed φ, $g^{(n)}(r, \varphi)$ is a strictly monotonic increasing function of r in the interval $R^{(n)}(0, \varphi) < r < R^{(n)}(1, \varphi)$ and since $g^{(n)}(r, \varphi)$ is continuous over the entire plane, it follows that $R^{(n)}(c, \varphi)$ is continuous in both variables for $0 < c < 1$, $0 \leqq \varphi < 2\pi$.

The following definition extends the transformation S_n over a family of condensers $\{C\}$.

DEFINITION 3. *Let $C = (D, E_0, E_1)$ be a condenser in the complex plane z, such that E_0 contains an interior point z_0. Put $G_1 = D \cup E_0$ and suppose that S_nG_1 (with $S_n = S_n(z_0)$) does not contain the entire open plane. Then, the condenser $C^{(n)}$ obtained from C by the transformation $S_n = S_n(z_0)$ is defined as follows:*

$$C^{(n)} = (D^{(n)}, E_0^{(n)}, E_1^{(n)}) ,$$

where $D^{(n)} = S_nG_1 - S_nE_0$, $E_0^{(n)} = S_nE_0$ and $E_1^{(n)} =$ the complement of S_nG_1.

4. A theorem concerning the Dirichlet integral of functions belonging to (C, z_0).

THEOREM 1. *Let $C = (D, E_0, E_1)$ be a condenser in the complex plane z, such that E_0 contains an interior point z_0. Suppose that $g(z)$ belongs to (C, z_0) and that its Dirichlet integral over D is finite. If $S_n = S_n(z_0)$, $(n = 1, 2, 3, \cdots)$, $g^{(n)}(z) = S_n g$, and $D^{(n)}$ is the domain mentioned in Definition 3, then:*

$$\iint_{D^{(n)}} (\nabla g^{(n)})^2 dxdy \leqq \iint_D (\nabla g)^2 dxdy \,. \tag{19}$$

REMARK. This theorem was proved by Szegö [4], for $n = 2, 3, \cdots$, in the special case where, D is a doubly-connected domain bounded by two smooth curves which are starlike with respect to z_0; E_0 and E_1 are connected sets; and the function $g(z)$ is the potential function of the condenser C.

Proof. By property (i) of $g(z)$ and by the remark at the end of Lemma 1 it follows that $g^{(n)}(z)$ is Lip on every compact subset of $D^{(n)}$. Therefore the first partial derivatives of $g^{(n)}(x, y)$ exist almost everywhere in $D^{(n)}$ and are bounded in every compact subset of $D^{(n)}$.

Without loss of generality we may suppose that $z_0 = 0$ and that the circle $|z| \leqq \rho = 1$ is contained in E_0. Again we shall write $L^{(n)}(c, \varphi)$ instead of $L_\rho^{(n)}(c, \varphi)$. We also introduce the following notations:

$$D(a, b) = \{z \mid a < g(z) < b\} \,,$$
$$D^{(n)}(a, b) = \{z \mid a < g^{(n)}(z) < b\} \,, \qquad \text{for } 0 < a < b < 1 \,.$$

The sets $D(a, b)$ and $D^{(n)}(a, b)$ will be mapped by $w = \log z$ $(0 \leqq Im w < 2\pi)$ on two sets which we denote by $H(a, b)$ and $H^{(n)}(a, b)$, respectively. Finally we define: $h(w) = g(e^w)$, $h^{(n)}(w) = g^{(n)}(e^w)$ and

$$\gamma_c = \{w \mid 0 < Im\, w < 2\pi, h(w) = c\} \,, \qquad \text{for } 0 < c < 1 \,.$$

The proof of the theorem rests on the following inequality:

$$\iint_{H^{(n)}(a,b)} [1 + \varepsilon^2(\nabla h^{(n)})^2]^{1/2} dudv \leqq \iint_{H(a,b)} [1 + \varepsilon^2(\nabla h)^2]^{1/2} dudv \,, \tag{20}$$

where $w = u + iv$, $0 < a < b < 1$ and $\varepsilon > 0$.

Inequality (19) is derived from (20) by a standard argument which we shall briefly describe.

The closures of the sets $D(a, b)$ and $D^{(n)}(a, b)$ are compact sets contained in D and $D^{(n)}$, respectively. Therefore the first partial derivatives of $h(u, v)$ $(h^{(n)}(u, v))$ are bounded in $H(a, b)$ $(H^{(n)}(a, b))$. It is evident from the definitions that the area of $H(a, b)$ equals that of $H^{(n)}(a, b)$. Taking into account these facts and using the binomial expansion of the integrands in (20), (for ε small enough), we obtain:

$$\frac{\varepsilon^2}{2}\iint_{H^{(n)}(a,b)}(\nabla h^{(n)})^2 dudv + O(\varepsilon^4) \leqq \frac{\varepsilon^2}{2}\iint_{H(a,b)}(\nabla h)^2 dudv + O(\varepsilon^4) .$$

Dividing by ε^2 and letting ε tend to zero we find that

$$\iint_{H^{(n)}(a,b)}(\nabla h^{(n)})^2 dudv \leqq \iint_{H(a,b)}(\nabla h)^2 dudv .$$

Since the Dirichlet integral is invariant under a simple conformal mapping, it follows that

$$\iint_{D^{(n)}(a,b)}(\nabla g^{(n)})^2 dxdy \leqq \iint_{D(a,b)}(\nabla g)^2 dxdy .$$

Hence, letting a tend to zero and b tend to one, we obtain the required inequality.

In the proof of (20) we may suppose that $\varepsilon = 1$.

The first step is the following assertion. Suppose that $w^* = u^* + iv^* \in H^{(n)}(a, b)$ and $0 < v^* < (2\pi/n)$. Put $h^{(n)}(u^*, v^*) = c^*$. If $\partial h/\partial u \neq 0$ at all the points of intersection of the set γ_{c^*} and the lines $Im\, w = v^* + (2\pi m/n)$ $(m = 0, \cdots, n-1)$, then there exists a neighbourhood of w^* in which $h^{(n)}(u, v) \in C^1$.

In order to prove this assertion we shall show first that $L(c, v) \in C^1$ in a neighbourhood of (c^*, v^*). By property (iii) the set γ_{c^*} intersects the line $Im\, w = v^*$ in a finite number of points, which we denote by $w_1, \cdots, w_p$, where $Re\, w_1 < Re\, w_2 < \cdots < Re\, w_p$. By hypothesis, $\partial h/\partial u \neq 0$ at these points. Let q be a positive number such that the circles $K_j : |w - w_j| \leqq q$, $(j = 1, \cdots, p)$, are contained in $H(a, b)$ and $\partial h/\partial u \neq 0$ in them. Then the following is easily verified:

There exists a rectangle

$$P = \{(c, v) \mid |c - c^*| \leqq \delta, |v - v^*| \leqq \delta\} ,$$

(where $a < c^* - \delta < c^* + \delta < b$, $0 < v^* - \delta < v^* + \delta < (2\pi/n)$), such that:

(a) If $(c, v) \in P$ then γ_c intersects the line $Im\, w = v$ in exactly p points, one point in each circles K_j.

(b) The set $H(c^* - \delta, c^* + \delta)$ intersects the strip $v^* - \delta < Im w < v^* + \delta$ in exactly p domains Q_j, where $Q_j \subset K_j$, $(j = 1, \cdots, p)$.

Solving $c = h(u, v)$ for u in Q_j we obtain a function $u = u_j(c, v)$. This function belongs to C^1 in the rectangle P where

$$(21) \qquad \frac{\partial u_j}{\partial c} = \left(\frac{\partial h}{\partial u}\right)^{-1}, \qquad \frac{\partial u_j}{\partial v} = -\left(\frac{\partial h}{\partial v}\right) \times \left(\frac{\partial h}{\partial u}\right)^{-1} .$$

Since by definition:

$$L(c, v) = \sum_{j=1}^{p} (-1)^{j+1} \times u_j(c, v) \tag{22}$$

it follows that $L(c, v) \in C^1[P]$. We observe that in Q_j we have $\partial h/\partial u = (-1)^{j+1} \times |\partial h/\partial u|$ so that

$$\frac{\partial L}{\partial c} = \sum_{j=1}^{p} \left| \frac{\partial u_j}{\partial c} \right| , \qquad \text{in } P . \tag{23}$$

Evidently, similar results hold for any of the points $c = c^*$, $v = v^* + (2\pi m/n)$, for $m = 0, \cdots, n-1$. Therefore it is possible to find a positive number $\eta (\eta \leqq \delta)$ such that $L^{(n)}(c, v) \in C^1$ and $(\partial L^{(n)}/\partial c) > 0$ in the rectangle $|c - c^*| < \eta$, $|v - v^*| < \eta$. By (18), for any fixed v, $c = h^{(n)}(u, v)$ is the inverse function of $u = L^{(n)}(c, v)$ in the interval $0 < c < 1$. Hence it follows that in a certain neighbourhood of (u^*, v^*), $h^{(n)}(u, v) \in C^1$ and

$$\frac{\partial h^{(n)}}{\partial u} = \left(\frac{\partial L^{(n)}}{\partial c}\right)^{-1}, \qquad \frac{\partial h^{(n)}}{\partial v} = -\left(\frac{\partial L^{(n)}}{\partial v}\right) \times \left(\frac{\partial L^{(n)}}{\partial c}\right)^{-1} . \tag{24}$$

Denote by $A(v)$ and $A_n(v)$ the intersections of the line $Im\, w = v$ with the sets $H(a, b)$ and $H^{(n)}(a, b)$ respectively. Let $w \in A(v)$ and $h(w) = c$, $(0 < v < 2\pi)$. If at one of the points of intersection of γ_c with the line $Im\, w = v$, $\partial h/\partial u$ vanishes then we shall say that w is a critical point of $A(v)$. Let $w \in A_n(v)$ and $h^{(n)}(w) = c$. If the intersection of γ_c with one of the sets $A(v + 2\pi m/n)$, $(m = 0, \cdots, n-1)$, contains a crititical point of that set, we shall say that w is a critical point of $A_n(v)$. By properties (iii) and (iv) the set of critical points of $A(v)$ is finite, and consequently, the set of critical points of $A_n(v)$ is finite.

We shall prove now that

$$\int_{A_1(v)} [1 + (\nabla h^{(1)})^2]^{1/2} du \leqq \int_{A(v)} [1 + (\nabla h)^2]^{1/2} du , \tag{25}$$

for $0 < v < 2\pi$. Inequality (20) for $n = 1$, follows from (25).

Let v_0 be a fixed point in the interval $(0, 2\pi)$ and let $\{c_1, \cdots, c_{k-1}\}$ be the set of values (possibly void) taken by $h(w)$ at the critical points of $A(v_0)$. We assume that these values are ordered as follows:

$$a = c_0 < c_1 < \cdots < c_{k-1} < c_k = b .$$

Denote by B_l that subset of $A(v_0)$ which consists of open segments, free from critical points, such that at the endpoints of each segment $h(w)$ assumes the values c_l and c_{l+1}. Evidently, for any l $(l = 0, \cdots, k-1)$ the set B_l is not void and $A(v_0) = \bigcup_{l=0}^{k-1} B_l$.

Now let m be a fixed integer, $0 \leqq m \leqq k-1$, and denote by $\alpha_1, \cdots, \alpha_p$

the segments contained in B_m, which were described above. We shall assume that α_j is at the left of α_{j+1}, $(j=1, \cdots, p-1)$. In some neighbourhood of α_j it is possible to solve $c = h(u, v)$ for u and thereby obtain a function $u = u_j(c, v)$. By (21) we obtain:

$$(26) \qquad \int_{\alpha_j} [1 + (\nabla h(u, v_0))^2]^{1/2} du = \int_{c_m}^{c_{m+1}} [1 + (\nabla u_j(c, v_0))^2]^{1/2} dc \ ,$$

for $j = 1, \cdots, p$.

Denote: $u'_j = L(c_j, v_0)$ and $w'_j = u'_j + iv_0$, $(j = 0, \cdots, k)$. Then w'_0 and w'_k are the endpoints of $A_1(v_0)$ while $w'_1, \cdots, w'_{k-1}$ are the critical points of $A_1(v_0)$. Denote by B'_m the open segment with endpoints w'_m, w'_{m+1}. By (22) and (24) (with $n = 1$) we get:

$$(27) \qquad \begin{aligned} \int_{B'_m} [1 + (\nabla h^{(1)}(u, v_0))^2]^{1/2} du &= \int_{c_m}^{c_{m+1}} [1 + (\nabla L(c, v_0))^2]^{1/2} dc \\ &= \int_{c_m}^{c_{m+1}} \left\{1 + \left[\nabla \sum_{j=1}^{p} (-1)^{j+1} u_j(c, v_0)\right]^2\right\}^{1/2} dc \ . \end{aligned}$$

By (26), (27) and the well known inequality

$$(28) \qquad \left\{\left(\sum_{j=1}^{p} x_j\right)^2 + \left(\sum_{j=1}^{p} y_j\right)^2 + \left(\sum_{j=1}^{p} t_j\right)^2\right\}^{1/2} \leqq \sum_{j=1}^{p} (x_j^2 + y_j^2 + t_j^2)^{1/2} \ ,$$

$(x_j, y_j, t_j$ being real numbers) we finally obtain:

$$(29) \qquad \begin{aligned} \int_{B'_m} [1 + (\nabla h^{(1)}(u, v_0))^2]^{1/2} du &\leqq \int_{B_m} [1 + (\nabla h(u, v_0))^2]^{1/2} du \\ &= \sum_{j=1}^{p} \int_{\alpha_j} [1 + (\nabla h(u, v_0))^2]^{1/2} du \ . \end{aligned}$$

Since (29) holds for any m, $(m = 0, \cdots, k-1)$ inequality (25) follows.

It remains to prove inequality (20) for $n = 2, 3, \cdots$. Since this inequality is proved for $n = 1$, it is enough to show that

$$(30) \quad n \times \int_{A_n(v_0)} [1 + (\nabla h^{(n)}(u, v_0))^2]^{1/2} du \leqq \sum_{j=0}^{n-1} \int_{A_1(v_j)} [1 + (\nabla h^{(1)}(u, v_j))^2]^{1/2} du \ ,$$

where $0 < v_0 < (2\pi/n)$ and $v_j = v_0 + (2\pi j/n)$.

Let $\{c_1^*, \cdots, c_{r-1}^*\}$ be the set of values (possibly void) assumed by $h^{(n)}(w)$ at the critical points of $A_n(v_0)$, these values being ordered as follows:

$$a = c_0^* < c_1^* < \cdots < c_{r-1}^* < c_r^* = b \ .$$

Put $u_m^* = L^{(n)}(c_m^*, v_0)$ and $u_{m,j}^* = L(c_m^*, v_j)$. By (24) we get:

$$\int_{u_m^*}^{u_{m+1}^*}[1+(\nabla h^{(n)}(u, v_0))^2]^{1/2}du = \int_{c_m^*}^{c_{m+1}^*}[1+(\nabla L^{(n)}(c, v_0))^2]^{1/2}dc$$

$$(31) \qquad = \frac{1}{n}\int_{c_m^*}^{c_{m+1}^*}\left[n^2+\left(\sum_{j=0}^{n-1}\nabla L(c, v_j)\right)^2\right]^{1/2}dc\ ;$$

$$\int_{u_{m,j}^*}^{u_{m+1}^{*}{}^{j}}[1+(\nabla h^{(1)}(u, v_j))^2]^{1/2}du = \int_{c_m^*}^{c_{m+1}^*}[1+(\nabla L(c, v_j))^2]^{1/2}dc\ ,$$

for $m = 0, \cdots, r-1$ and $j = 0, \cdots, n-1$. From (31) and (28), inequality (30) follows. This completes the proof of the theorem.

5. The transformation S_n diminishes the capacity of a condenser. Let $C = (D, E_0, E_1)$ be a condenser in the complex plane z, satisfying the conditions of Definition 3. It will be assumed that the Dirichlet problem for $\nabla u = 0$, with continuous boundary values, always has a solution in D. (Sufficient conditions for the validity of this assumption are given, for example, in Hayman [2], Th. 4.2, pp. 63–64. Following Hayman's terminology we shall say that a domain is *admissible* if it satisfies these conditions.) The *capacity* of the condenser C is defined as the Dirichlet integral over D, of the potential function $\omega(z)$ of C, (see § 3).

Let $C^{(n)} = S_nC = (D^{(n)}, E_0^{(n)}, E_1^{(n)})$, (where $S_n = S_n(z_0)$). The domain $D^{(n)}$ is admissible so that the capacity of $C^{(n)}$ is defined. We now prove the following:

THEOREM 2. *Let C and $C^{(n)}$ be the condensers mentioned above and denote their capacities by I and I_n respectively. Then we have $I_n \leqq I$.*

Proof. Let $\omega^{(n)}(z) = S_n\omega(z)$, $(S_n = S_n(z_0))$. Since $\omega(z) \in (C, z_0)$, by Theorem 1 we have

$$(32) \qquad \int_{D^{(n)}}(\nabla\omega^{(n)})^2dxdy \leqq \int_D(\nabla\omega)^2dxdy = I\,.$$

The function $\omega^{(n)}(z)$ is continuous over the extended plane z and Lip in every compact subset of $D^{(n)}$; it vanishes on E_0 and assumes the value 1 on E_1. Hence, by the Dirichlet minimum principle (see, Hayman [2], Th. 4.3, pp. 65–67) we have

$$(33) \qquad I_n \leqq \int_{D^{(n)}}(\nabla\omega^{(n)})^2dxdy\ .$$

The required result follows from (32) and (33).

We shall apply Theorem 2 in order to obtain a result about the inner radius. Let D be a domain in the complex plane z, z_0 a point of D, and $r(D, z_0)$ the inner radius of D at z_0. (We refer here to the definition given, for example, in Hayman [2] pp. 78–80, where the inner radius is defined without any assumptions on D.) The domain D can be approximated from within by a series of bounded analytic domains $\{D_n\}$, which contain the point z_0, such that $\lim_{n\to\infty} r(D_n, z_0) = r(D, z_0)$. (An analytic domain is a domain bounded by a finite number of disjoint, simple closed, analytic curves.) By a well known method of Pólya and Szegö (see Pólya-Szegö [3] pp. 44–45; also Hayman [2] pp. 81–84) the following theorem is obtained as a consequence of Theorem 2.

THEOREM 3. *Let D be a domain in the complex plane z and let $z_0 \in D$. If $S_n = S_n(z_0)$, then*

$$r(D, z_0) \leqq r(S_n D, z_0) \,. \tag{34}$$

6. Applications in the theory of functions. In this section we denote by $w = f(z)$ a function which is regular in $|z| < 1$ and by D the domain of all values w assumed by this function at least once in $|z| < 1$. It is known that

$$|f'(0)| \leqq r(D, f(0)) \,, \tag{35}$$

equality holding if and only if $f(z)$ is a (1,1) mapping, (see Hayman [2], Th. 4.5, p. 80).

As a consequence of Theorem 3 we obtain the following:

THEOREM 4. *Let $S_n = S_n(f(0))$ and suppose that $S_n D$ does not contain the entire open plane. Let $w = F(z)$ be a (1,1) conformal mapping of $|z| < 1$ onto $S_n D$, such that $F(0) = f(0)$. Then we have $|f'(0)| \leqq |F'(0)|$.*

Proof. By (35) we get: $|f'(0)| \leqq r(D, f(0))$ and $|F'(0)| = r(S_n D, F(0))$. From these relations together with (34), the required inequality follows.

The following results are based on Theorem 4.

THEOREM 5. *Let $f(z) = a_1 z + a_2 z^2 + \cdots$. Define $R^{(n)}(\varphi)$ as in Definition 1, for the domain D and the point $w = 0$. Then,*

$$|a_1| \leqq \sqrt[n]{4}\, R^{(n)}(\varphi) \,, \qquad (0 \leqq \varphi < 2\pi) \tag{36}$$

and equality holds for the function

$$w = \psi_n(z) = te^{i(\varphi+\theta)} z/(1 + e^{in\theta} z^n)^{2/n} \,, \qquad (t \text{ and } \theta \text{ real numbers}) \,.$$

Proof. Let φ_0 be a fixed real number and suppose that $R^{(n)}(\varphi_0) = d < \infty$. Denote by D_0 the domain containing the entire w plane, with the exception of n rays: $\arg w = \varphi_0 + (2\pi k/n)$, $d \leqq |w|$, $(k = 0, \cdots, n-1)$. The domain $S_nD(S_n = S_n(0))$ is contained in D_0. The function $w = \sqrt[n]{4}\, de^{i\varphi_0} f_n(z)$ where

$$f_n(z) = z/(1 + z^n)^{2/n} , \tag{37}$$

maps $|z| < 1$ conformally, (1,1) onto D_0. Therefore, by the principle of subordination and Theorem 4 it follows that $|a_1| \leqq \sqrt[n]{4}\, d$, and inequality (36) is proved. The assertion concerning the function $w = \psi_n(z)$ is evident.

The following theorem may be proved by the same method.

THEOREM 6. *Let $f(z) = a_1z + a_2z^2 + \cdots$. Suppose that $R^{(n)}(\varphi) \leqq M < \infty$ for $0 \leqq \varphi < 2\pi$ and that $R^{(n)}(\varphi_0) = \beta M\,(0 < \beta \leqq 1)$. Then*

$$|a_1| \leqq \beta M \cdot \sqrt[n]{4}/(1 + \beta^n)^{2/n} , \tag{38}$$

and equality holds for the function

$$w = \phi_n(z) = Me^{i\varphi_0} f_n^{-1}[q f_n(e^{i\theta} z)] ,$$

where $f_n(z)$ is defined by (37), $0 \leqq \theta < 2\pi$ *and* $q = \sqrt[n]{4}\,\beta/(1 + \beta^n)^{2/n}$.

We now prove

THEOREM 7. *Let $f(z) = a_1z + a_2z^2 + \cdots$ and define*:

$$R_0 = \exp\left[\frac{1}{2\pi}\int_0^{2\pi} \log R(\varphi)d\varphi\right] = \exp\left[\frac{1}{2\pi}\int_0^{2\pi} \log R^{(n)}(\varphi)d\varphi\right] . \tag{39}$$

Then $|a_1| \leqq R_0$, and equality holds for $w = a_1z$.[2]

Proof. First suppose that $w = f(z)$ is regular in $|z| \leqq 1$ and that $f'(z) \neq 0$ on $|z| = 1$. Then $R(\varphi)$ is a continuous function of φ, and we have

$$\lim_{n\to\infty} R^{(n)}(\varphi) = \lim_{n\to\infty} \exp\left[\frac{1}{n}\sum_{k=0}^{n-1} \log R\left(\varphi + \frac{2\pi k}{n}\right)\right] = R_0 , \tag{40}$$

for any real φ. Therefore, if a positive ε is given and n is sufficiently large, the domain S_nD (where $S_n = S_n(0)$) is contained in the circle $|z| < R_0 + \varepsilon$. Hence, by Theorem 4 and the principle of subordi-

[2] The author obtained this result in a weaker form, with $\bar{r}_n = \frac{1}{2\pi}\int_0^{2\pi} R^{(n)}(\varphi)d\varphi$ instead of R_0. (By the geometric-arithmetic mean theorem $R_0 \leqq \bar{r}_n$ for every n). The stronger form written above was suggested by the referee, to whom our thanks are due.

nation, we get $|a_1| \leqq R_0 + \varepsilon$. In order to prove the theorem in the general case, we approximate the function $w = f(z)$ by functions $w = f(\rho z)$, with $0 < \rho < 1$.

Let Ω be an open set in the plane z and let $z_0 \in \Omega$. Denote by $m(\varphi)$ the linear (Lebesgue) measure of the set $E(\varphi) = \{z \mid \arg(z - z_0) = \varphi, z \in \Omega\}$, and define

$$m^{(n)}(\varphi) = \frac{1}{n} \sum_{k=0}^{n-1} m\left(\varphi + \frac{2\pi k}{n}\right). \tag{41}$$

We shall show that Theorems 5, 6, 7, remain true if $R(\varphi)$ is replaced by $m(\varphi)$, and $R^{(n)}(\varphi)$ by $m^{(n)}(\varphi)$. This is a consequence of the following inequalities:

$$R(\varphi) \leqq m(\varphi), \tag{42}$$

$$R^{(n)}(\varphi) \leqq m^{(n)}(\varphi), \qquad \text{for } 0 \leqq \varphi < 2\pi. \tag{42$'$}$$

If $R(\varphi)$ is finite, equality holds in (42) if and only if the set $E(\varphi)$ is contained in a segment E^* such that $E^* - E(\varphi)$ is a set of measure zero. (We shall refer to this condition as the *MR* condition.) Inequality (42′) follows from (42) by the geometric-arithmetic mean theorem. Hence, if $R^{(n)}(\varphi)$ is finite, equality holds in (42′) if and only if

$$R(\varphi) = R\left(\varphi + \frac{2\pi k}{n}\right) = m(\varphi) = m\left(\varphi + \frac{2\pi k}{n}\right), \ (k = 1, \cdots, n-1).$$

From this it follows that when we replace $R(\varphi)$ by $m(\varphi)$ and $R^{(n)}(\varphi)$ by $m^{(n)}(\varphi)$, the functions mentioned at the end of Theorems 5, 6, 7, are in each case, the *only* functions for which equality holds.

In order to prove (42) we may suppose that $m(\varphi)$ is finite. In this case, for any $\varepsilon > 0$ we can find a subset F of $E(\varphi)$, consisting of a finite number of segments, such that the linear measure of $E(\varphi) - F$ is smaller than ε. Therefore it is enough to prove (42) in the case that $E(\varphi)$ consists of a finite number of segments. Suppose that these segments are not adjacent. Then, by shifting them toward z_0 (so that they do not overlap), we increase $R(\varphi)$, while $m(\varphi)$ is invariant. But if the segments are adjacent we have $R(\varphi) = m(\varphi)$. Therefore (42) is proved.

Evidently, the *MR* condition for $E(\varphi)$ is sufficient in order that $R(\varphi) = m(\varphi)$. Suppose now that $R(\varphi)$ is finite and that $E(\varphi)$ does not satisfy the *MR* condition. Then it is possible to find a subset F_1 of $E(\varphi)$ and a subset F_2 of the complement of $E(\varphi)$ on the ray $\arg(z - z_0) = \varphi$, such that the two subsets have equal, positive measures and F_2 separates F_1 from z_0. Replacing F_1 by F_2 we increase $R(\varphi)$, but not $m(\varphi)$. Therefore we must have $R(\varphi) < m(\varphi)$.

References

1. G. M. Golusin, *Interior Problems of the Theory of Schlicht Functions*, translated from the Russian article Uspekhi Matem. Nauk, 6, pp. 26-89 (1939) by T. C. Doyle, A. C. Schaeffer and D. C. Spencer.

2. W. K. Hayman, *Multivalent Functions*, Cambridge University Press, (1958).

3. G. Pólya and G. Szegö, *Isoperimetric Inequalities in Mathematical Physics*, Annals of Mathematical Studies. Nr. 27, Princeton, (1951).

4. G. Szegö, *On a certain kind of symmetrization and its applications* Ann. Mat. Pura Applicata, Ser. 4, Vol. 40, pp. (1955), 113-119.

B^* ALGEBRA UNIT BALL EXTREMAL POINTS

PHILIP MILES

Results of Kadison [3] and Jacobson [2] are combined to show that the points described by the title are unitaries, left shifts, right shifts, or sums of these. The extremality property is preserved by homomorphisms; conversely, when range and domain are AW^* algebras, every extremal point of the range has an extremal point in its pre-image. Exact formulations of these results and of a few simple consequences are given in section one; proofs follow in section two.

In what follows, A will be a self-adjoint subalgebra of some B^* algebra; "x is extremal (A)" will mean that x is an extremal point of the unit ball of A with respect to the B^* norm indicated by the context; "weak topology" will mean the weak operator topology with respect to the representation of A by bounded operators on a Hilbert space which is indicated by the context.

1. Theorems.

Our starting point is a formula due to Kadison ([3], Theorem 1). In a mildly generalized form, his result is:

THEOREM 1. *Let A be a self-adjoint subalgebra of some B^* algebra B. Then x is extremal (A) if and only if*

$$(1 - x^*x)A(1 - xx^*) = \{0\} .$$

Here "1" stands for the identity of A if there is one; otherwise the meaning of the equation is to be found by performing the indicated multiplications for each $y \in A$. It turns out (Theorem 2) that the existence of any element extremal (A) implies that A has an identity.[1]

An obvious consequence of this formula is the perseverance of extremality. Calling "reasonable" any linear topology making involution continuous, and multiplication continuous in each variable separately, we have:

COROLLARY (i) *If $\bar{A}$ is the closure of A in B with respect to a reasonable topology, and if x is in A, then x is extremal (A) if and only if x is extremal $(\bar{A})$.*

(ii) *If ϕ is a *-homomorphism of A into a B^* algebra B_1, then x extremal (A) implies that ϕx is extremal (ϕA).*

Using the methods of [2], one can draw substantial information about the form of an individual extremal element from Theorem 1.

Received February 14, 1963. Supported by NSF Grants G-14362 and G 19050.

[1] This has already been proved by Sakai [5, p. 1.3]

THEOREM 2. *Let A be a self-adjoint subalgebra of the algebra $\mathscr{B}(H)$ of all bounded operators on the Hilbert space H. Let x be extremal (A). Then*

(i) *A has an identity*[1], *which we now take to be the identity operator on H—thus possibly changing the meaning of H.*

(ii) *x satisfies one of the following*

(a) *x is unitary*

(b) *x is semi-unitary—i.e., exactly one of xx^*, x^*x is the identity*

(c) *There is a projection p such that px and $(1-p)x$ are semi-unitary* on pH and $(1-p)H$ respectively. *Further, p can be taken from the center of the weak closure of A or, if A is AW^*, from the center of A.*

(iii) *If x is a semi-unitary with $xx^* = 1$, then*

(a) *$H = \sum_0^\infty \oplus H_i$ where x is an isometry of H_0 onto H_0 and of H_{i+1} onto H_i $(i \geqq 1)$, and maps H_1 onto zero.*

(b) *Let X be the left shift on unilateral l_2. The map taking a polynomial in x and x^* into the same polynomial in X and X^* induces a *-isomorphism from the uniformly closed subalgebra of $\mathscr{B}(H)$ generated by x and x^* onto the uniformly closed subalgebra of $\mathscr{B}(l_2)$ generated by X and X^*.*

(c) *The weakly closed subalgebra W of $\mathscr{B}(H)$ generated by x and x^* is naturally *-isomorphic to $\mathscr{B}(l_2) \oplus Z$, where Z is the weakly closed subalgebra of $\mathscr{B}(H)$ generated by x and x^* restricted to H_0. $Z + 1$ is the center of W.*

Clearly there is a restatement of (iii) applying to semi-unitary operators with $x^*x = 1$; in it X is the right shift on l_2, and x maps H_i onto H_{i+1} for all $i \geqq 1$. It is also clear that (iii) can be applied separately to the components px and $(1-p)x$ of an element satisfying (iic). For example, the uniformly closed algebra generated by such an element is *-isomorphic with the uniformly closed subalgebra of $\mathscr{B}(l_2 \oplus l_2)$ generated by $u_1 + u_2^*$, where u_1 is the left shift on the first l_2, zero on the second, and u_2 is zero on the first l_2, the left shift on the second.

Part (iii) gives us three ways of looking at a semi-unitary element. With regard to the uniformly closed, self-adjoint subalgebras they generate, all semi-unitaries are the same. From the standpoint of weakly closed algebras, semi-unitaries differ only in their unitary parts. Viewed spatially—i.e., as representing a similarity class—a semi-unitary is determined by its unitary part and the dimension of its (or its adjoint's) null space. In the light of (iiia), Putnam's result that similar normal operators are unitarily equivalent is easily seen to imply that similar semi-unitary operators are unitarily equivalent. Another method of classifying extremal elements is considered in [3].

Calling a projection p infinite (A) if there is a partial isometry u in A with $uu^* = p$, $u^*u \neq p$, $u^*u \leqq p$, we see that each projection p which is infinite (A) gives rise to at least one semi-unitary, viz. $u + (1 - p)$. Conversely, the existence of a semi-unitary implies the existence of a projection infinite (A), viz 1. Elements of type (iic) are similarly related to projections p in the weak closure of A having the property that p and $1 - p$ are both infinite (weak closure A).

Clearly the study of extremal points will be most rewarding when they exist in substantial number. We have seen that when an identity is lacking, there are no extremal elements. It is well known that if A is a B^* algebra with identity, there are enough unitaries so that every element of A is a linear combination of four of them. But much more can be asked—namely, that the unit ball be the (somehow) closed convex hull of its extremal points. This fails to happen for the general B^*-algebra. An exercise in Bourbaki shows that if A is a weakly closed subalgebra of $\mathscr{B}(H)$, then A is the weakly closed convex hull of its extremal points; the proof may be written "Alaoglu: Krein-Milman."

The weakly-closed, or similar, situation has another useful feature; restating an argument of Calkin ([1], proofs of Theorems 2.4 and 2.5) we obtain:

THEOREM 3. *If A is an AW^* algebra, ϕ a*-homomorphism of A into a B^* algebra, and y a point extremal (ϕA), then there exists an x extremal (A) with $\phi x = y$.*

As an application of this theorem, we consider how the type (in the sense of [4]) of an AW^* algebra determines the type of an AW^* homomorphic image.

THEOREM 4. *Let A, B be AW^* algebras, with B the image of A under some non-trivial *-homomorphism. Then*

(i) *A of type I_n implies B of type I_n*
(ii) *A of type II_1 implies B of type II_1*
(iii) *A of type II_∞ implies B of type II_∞ or III*
(iv) *A of type III implies B of type III*
(v) *A of type I_∞ implies B of type I_∞, II_∞, or III.*

It is likely that another attack would produce a substantially improved theorem in this direction.

2. Proofs.

Proof of Theorem 1. The proof of [3] may be modified to apply in the case where A is not closed, nor known to have identity. In fact, let x be extremal (A). Letting $h = x^*x$, we observe that $0 \leqq \sigma(h) \leqq 1$, where $\sigma(h)$ is the spectrum of h in the closure of A. Let C be the intersection with A of the uniformly closed subalgebra generated by h. Then C is isometrically*-isomorphic with an algebra of continuous, complex valued functions on $\sigma(h)$. Further, C contains $s = h(1 - h)^2$.

We desire to show the inequality

$$|| (1 \pm s)h(1 \pm s) || \leqq 1 .$$

In view of the identification of C with a function algebra, this reduces to showing that for real t between zero and one,

$$0 \leqq t[1 \pm t(1 - t)^2]^2 \leqq 1 .$$

This is obvious when the ambiguous sign is minus; when it is plus, the expression in t may be expanded as a convex combination of points obviously in $[0, 1]$.

We thus have $|| x(1 \pm s) || \leqq 1$. Writing $x = (1/2)[(x + xs) + (x - xs)]$ and using the extremality of x, we have $xs = 0$ and so $sx^*xs = 0$— i.e., $h^4(1 - h)^4 = 0$. Again viewing C as a function algebra, we conclude from the last equation that the function h assumes only the values zero and one, so h is a projection and x a partial isometry.

Thus if $y \in (1 - x^*x)A(1 - xx^*)$, then $y = (1 - x^*x)y(1 - xx^*)$ and so $xy = yx = 0$. It follows that $|| x \pm y^* ||^2 = || x^*x + yy^* || = \max (|| x^*x ||, || yy^* ||)$. Assuming that $|| y || \leqq 1$, we have $|| x \pm y^* || = 1$ and so, by the extremality of x, $y^* = 0$.

The converse, that an x giving $(1 - x^*x)A(1 - xx^*) = 0$ is extremal (A), is proved in [3]; it also follows from Theorem 2, which is based entirely on the equation $(1 - x^*x)A(1 - xx^*) = 0$.

Proof of Theorem 2. Suppose x satisfies

(1) $$(1 - x^*x)A(1 - xx^*) = 0 ;$$

then, since A is self-adjoint, x also satisfies

(2) $$(1 - xx^*)A(1 - x^*x) = 0 .$$

For each y in A we have, by (1),

$$(1 - x^*x)(1 - xx^*)yy^*(1 - xx^*)(1 - x^*x) = 0 ,$$

and so $(1 - x^*x)(1 - xx^*)y = 0$. Performing the indicated multiplications, we obtain

$$(x^*x + xx^* - x^*x^2x^*)y = y .$$

The same argument may be made with x permuted with x^* and y permuted with y^*; the result is that $x^*x + xx^* - x^*x^2x^*$ is also a right identity for A. As previously agreed, we consider this element to be the identity operator on H, and denote it by "1".

We must now show that (1) implies part (ii) of the theorem. Observe first that (1) implies

$$0 = (1 - x^*x)x^*(1 - xx^*)x = (x^* - x^*xx^*)(x - xx^*x) ,$$

and so, that x is a partial isometry.

We next show:

$$(3) \qquad \begin{aligned} x^k x^{*m}(1 - x^*x) &= x^{*m-k}(1 - x^*x) \\ x^{*k}x^m(1 - xx^*) &= x^{m-k}(1 - xx^*) \end{aligned} \qquad 0 \leqq k \leqq m .$$

The first line of (3) may be rewritten as

$$(1 - x^k x^{*k})x^{*n}(1 - x^*x) = 0 \qquad k \geqq 0,\ n \geqq 0$$

and this equation established by induction on k. It clearly holds for $k=0$ and, by (2) for $k=1$. Writing $1-x^{k+1}x^{*k+1}$ as $(1 - x^k x^{*k}) + x^k(1 - xx^*)x^{*k}$, we see that the induction hypothesis reduces the previous equation to:

$$x^k(1 - xx^*)x^{*k+n}(1 - x^*x) = 0 ;$$

but this is already true by (2). The second line of (3) is proved in the same way, using (1) in place of (2).

That x is a partial isometry, together with (3), gives

$$(4) \qquad x^k x^{*m}(1 - x^*x) = x^{*k}x^m(1 - xx^*) = 0 \qquad 0 \leqq m \leqq k .$$

We can now copy the argument of [2]. Define e_i, f_i by

$$\begin{aligned} e_i &= x^{*i-1}x^{i-1} - x^{*i}x^i \\ f_i &= x^{i-1}x^{*i-1} - x^i x^{*i} \end{aligned} \qquad i \geqq 1 .$$

It follows that, for all $i, j \geqq 1$,

$$e_i e_j = \delta_{ij} e_j , \qquad f_i f_j = \delta_{ij} f_j$$
$$e_i = 0 \text{ if and only if } e_1 = 0 , \qquad f_i = 0 \text{ if and only if } f_1 = 0$$
$$e_i f_j = 0 .$$

These relations are immediate consequences of (2), (3), (4), and the fact that x is a partial isometry. We suppose for a moment that A is an AW^* algebra. Let

$$q = \vee_1^\infty e_i , \qquad r = \vee_1^\infty f_i ,$$

the supremum being that given by the AW^* character of A. Then for any y in A, $ye_i = 0$ for all i implies $yq = 0$, and similarly for yf_i and yp. For any i and j,

$$f_iAe_j = x^{i-1}(1 - xx^*)x^{*i-1}Ax^{*j-1}(1 - x^*x)x^{j-1} = 0$$

by (2). From what we have just said, this implies $rAq = 0$.

Now consider the left annihilator of Aq. Since A is AW^*, the annihilator can be written as Ap for some projection p in A. It is easily shown that, since Aq is a left ideal, p is central in A. We thus have

$$(1 - p)q = q(1 - p) = q$$

and, since $rAq = 0$, r is in Ap—i.e.,

$$pr = rp = r .$$

By definition, $rf_1 = f_1r = f_1$; thus, $pf_1 = p(rf_1) = (pr)f_1 = f_1$, and so,

$$p(1 - xx^*) = 1 - xx^* = (1 - xx^*)p .$$

Rearranging terms,

$$xx^*(1 - p) = (1 - p)xx^* = 1 - p .$$

On the other hand, $e_1 = qe_1$, so $(1 - p)e_1 = (1 - p)qe_1 = e_1$, and

$$(1 - p)x^*x = (1 - p) - e_1 .$$

Thus $(1 - p)x$ is semi-unitary in $(1 - p)A$, and unitary just in case $e_1 = 0$.

In the same way, we show

$$x^*xp = px^*x = p$$
$$p - xx^*p = f_1 ,$$

so px is semi-unitary in pA, unitary just in case $f_1 = 0$.

This proves (ii) when A is AW^*; the statements about the case where A is a self-adjoint subalgebra of $\mathscr{B}(H)$ follow on observing that the weak closure of A is AW^*.

To obtain (iii), observe that if x is semi-unitary with $xx^* = 1$, then x satisfies (1), and we may define elements e_i in terms of x as above (the f_i are of course all zero in this case). Define the spaces H_i of (a) by

$$H_i = \begin{cases} (1 - q)H & i = 0 \\ e_iH & i > 0 . \end{cases}$$

We observe from (3) that $xe_1 = 0$, $xe_i = e_{i-1}x$ for $i > 1$—so $x^*xe_i = e_i$ for $i > 1$. Consequently for any $\xi, \eta \in H$, $i > 1$

$$(xe_i\xi, xe_i\eta) = (x^*xe_i\xi, e_i\eta) = (e_i\xi, e_i\eta) .$$

Since $H = xx^*H \subseteqq xH$, $x(e_iH) = e_{i-1}(xH) = e_{i-1}H$.

Further,

$$(x(1-q)\xi, x(1-q)\eta) = ((1-e_1)(1-q)\xi, (1-q)\eta) = ((1-q)\xi, (1-q)\eta)$$

and

$$(x(1-q)\xi, e_i\eta) = (x^*(1-q)\xi, e_i\eta) = 0 ,$$

so x and x^* take H_0 into itself—and so, since x^* is never zero, x takes H_0 onto itself. Part (a) is now established.

To show (*b*), identify l_2 with a subspace of H by picking some ξ in H_1 with $\|\xi\| = 1$ and identifying the sequence $\{\eta_i\}$ in l_2 with $\Sigma\eta_i x^{*i}\xi$. The restriction map is now clearly a*-isomorphism, from the algebra of polynomials in x and x^* onto the algebra of polynomials in X and X^*; it remains only to prove that this map is an isometry. We know from [7] §2, that certain algebras with involution have a unique B^* norm: a sufficient condition is that the algebra have a faithful *-representation on some Hilbert space, and that for each z in the algebra there is a real k such that $f(z^*z) \leqq kf(1)$ for each functional f on the algebra which is positive on all y^*y.

The algebra of polynomials in X and X^* has been defined as being represented on l_2, and so satisfies the first part of this condition. Further, $X^{*k}X^k - X^{*k+1}X^{k+1}$ is a projection for each $k \geqq 0$. Thus if f is any positive functional,

$$f(1) \geqq f(X^*X) \geqq f(X^{*2}X^2) \geqq \cdots .$$

It is readily shown that any Y^*Y in the polynomial algebra can be written as $\Sigma a_k X^{*k}X^k$—so

$$f(Y^*Y) = \Sigma a_k f(X^{*k}X^k) \leqq (\Sigma |a_k|)f(1)$$

for any positive functional f, and the second part of the condition is also satisfied. Thus there is only one B^* norm on the algebra generated by X and X^*, and the norm this algebra inherits from $\mathscr{B}(l_2)$ is the same it gets from $\mathscr{B}(H)$ via the restriction map.

The isomorphism between the polynomial algebras can be obtained without considering x to be represented on any space; this is done in [2]. The isomorphism can be shown isometric by showing that a polynomial in X and X^* has the same norm as the same polynomial in U and U^*, U being the (unitary) left shift on bi-lateral l_2.

To prove (*c*), let $\{\xi^1_\alpha\}$ be an ortho-normal basis for H_1, and let $\xi^k_\alpha = x^{*k}\xi^1_\alpha$; then $\{\xi^k_\alpha\}$ is an ortho-normal basis for H_k. Let p_α be the projection on the closed linear span of $\{\xi^k_\alpha: k = 1, 2, 3, \cdots\}$, and $\tau_{\alpha,\beta}$ the isometry of $p_\alpha H$ onto $p_\beta H$ which, for each k takes ξ^k_α onto ξ^k_β. Observe that p_α and $\tau_{\alpha,\beta}$ commute with x and x^*. Consequently, if $w \in W$, then w commutes with all $\tau_{\alpha,\beta}$. It follows that there exist scalars $\nu_{i,j}$ such that, for any α,

$$w\xi^i_\alpha = \Sigma_j \nu_{ij}\xi^j_\alpha \,. \tag{5}$$

Suppose z commutes with x and x^*. The equations

$$(z\xi^i_\alpha, \xi^j_\beta) = (zx^{*i-1}\xi^1_\alpha, x^{*j-1}\xi^1_\beta) = (zx^{j-1}x^{*i-1}\xi^1_\alpha, \xi^1_\beta) = (z\xi^1_\alpha, x^{i-1}x^{*j-1}\xi^1_\beta)$$

show that

$$(z\xi^i_\alpha, \xi^j_\beta) = \begin{cases} 0 & i \neq j \\ (z\xi^1_\alpha, \xi^1_\beta) & i = j \,. \end{cases}$$

In other words, there exist scalars $\lambda_{\alpha,\beta}$ such that for each i,

$$z\xi^i_\alpha = \Sigma_\beta \lambda_{\alpha,\beta}\xi^i_\beta \,. \tag{6}$$

Further, if p_0 is the projection on H_0, then z commutes with p_0.

Now, given any w commuting with p_0 and satisfying (5), it follows from (6) and the fact that the ξ^i_α span $H_0^\perp$ that $(1 - p_0)w$ commutes with every z which commutes with x and x^*. Since every element of W commutes with p_0, we have $(1 - p_0)W$ isomorphic to $\mathscr{B}(l_2)$ under the correspondence obtained naturally via (5). Clearly $p_0 W$ is isomorphic to Z, and the proof of (*c*) complete.

Proof of Theorem 3. The first step is to show that, under the conditions of the hypothesis, the pre-image of a partial isometry contains a partial isometry. The proof follows an argument of Calkin [1, Theorems 2.4 and 2.5].

Let y be a partial isometry of B, and let v be any element of $\phi^{-1}(y)$. Since A is AW^*, v has a polar decomposition in A—i.e., there are elements u and h in A such that

$$u \text{ is a partial isometry}$$
$$h = (v^*v)^{1/2} = u^*uh = hu^*u$$
$$v = uh$$

(see [6], Lemma 2.1).

Since $\phi(h^2)$ is a projection, zero and one are in the spectrum of $\phi(h^2)$. Since ϕ is a homomorphism, the spectrum of $\phi(h^2)$ is contained in the spectrum of h^2. Since $h \geqq 0$, this implies that zero and one are in the spectrum of h. Let p_λ be the resolution of the identity for h given by the spectral theorem, let $0 < \alpha < 1$, and let $q = 1 - p_\alpha$. Then $0 \neq q \neq 1$.

Let C be some maximal commutative, self-adjoint subalgebra of A containing 1 and h—and so q as well. Let $x \to \hat{x}$ be the Gelfand representation of C on $C(\Omega)$, the algebra of all continuous, complex-valued functions on Ω, the (compact, Hausdorff) maximal ideal space of C. By definition,

$$\hat{q}(\omega) = 0 \text{ implies } (1 - \hat{h})(\omega) \geqq 1 - \alpha$$
$$\hat{q}(\omega) = 1 \text{ implies } (1 - \hat{h})(\omega) \leqq 1 - \alpha \, .$$

We now assert that there exist self-adjoint r, s, and t in C such that

$$hrq = q \, , \qquad h(1 + h)sq = q$$
$$(1 - h^2)t(1 - q) = 1 - q \, .$$

Since the Gelfand representation is *a* *-isomorphism, this is equivalent to asserting that there are real valued functions $\hat{r}$, $\hat{s}$, and $\hat{t}$ in $C(\Omega)$ such that

$$\hat{r}(\omega) = \hat{h}^{-1}(\omega) \text{ when } \hat{q}(\omega) \neq 0 \, ,$$
$$\hat{s}(\omega) = [\hat{h}(1 + \hat{h})]^{-1} \text{ when } \hat{q}(\omega) \neq 0 \, ,$$
$$\hat{t}(\omega) = [1 - \hat{h}^2]^{-1}(\omega) \text{ when } (1 - \hat{q})(\omega) \neq 0 \, .$$

But $\hat{h}^{-1}$ and $[\hat{h}(1 + \hat{h})]^{-1}$ are bounded and continuous on the closed set $\{\omega : \hat{h}(\omega) \geqq \alpha\}$, which contains the set $(\omega : \hat{q}(\omega) \neq 0\}$, and $[1 - \hat{h}^2]^{-1}$ is bounded and continuous on $\{\omega : (1 - \hat{h})(\omega) \geqq 1 - \alpha\}$, which contains the set $\{\omega : (1 - \hat{q})(\omega) \neq 0\}$. The existence of $\hat{r}$, $\hat{s}$, and $\hat{t}$ is therefore guaranteed by the Tietze extension theorem.

$$qu^*uq = qrhu^*uhrq = qrh^2rq = q \, ,$$

so uq is a partial isometry. Since ϕv is a partial isometry, $uh - uh^3$ is in kernel ϕ. Therefore

$$uh(1 - q) = u(1 - h)h(1 + h)t(1 - q)$$
$$= u(h - h^3)t(1 - q) \, ,$$

which is in kernel ϕ. Also,

$$u(1 - h)q = u(1 - h)h(1 + h)sq$$
$$= u(h - h^3)sq$$

which is in kernel ϕ. Therefore $uh - uq = uh(1 - q) - u(1 - h)q$ is in kernel ϕ, which gives the desired result.

We can now show that if y is extremal (B), then $\phi^{-1}(y)$ contains an element extremal (A). For let v be any partial isometry in $\phi^{-1}(y)$. The fundamental comparability theorem for AW^* algebras—e.g., [4], Theorem 5.6—says that there exist central projections e_1, e_2 in A such that

$$e_1(1 - v^*v) \leqq e_1(1 - vv^*)$$
$$e_2(1 - v^*v) \geqq e_2(1 - vv^*)$$
$$e_1 + e_2 = 1 .$$

In other words, there are partial isometries w and z in A such that

$$w^*w = e_1(1 - v^*v) , \qquad ww^* \leqq e_1(1 - vv^*)$$
$$z^*z = e_2(1 - vv^*) , \qquad zz^* \leqq e_2(1 - v^*v) .$$

The first equation implies $0 = vw^*w$, and so $vw^* = wv^* = 0$; it also gives $w^*w(1 - e_1) = 0$, and so $we_1 = e_1w = w$. The first inequality gives $0 = e_1v^*(1 - vv^*)ww^* = v^*ww^*$, and so $v^*w = w^*v = 0$. Similarly, $vz = zv = 0$ and $e_2z = ze_2 = z$. Define u_1 and u_2 by

$$u_1 = e_1v + w , \qquad u_2 = e_2v + z^* .$$

We have at once from the preceding equations that

$$u_1{}^*u_1 = e_1 , \qquad u_1e_1 = e_1u_1 = e_1$$
$$u_2u_2{}^* = e_2 , \qquad u_2e_2 = e_2u_2 = e_2 .$$

Since $\phi(v)$ is extremal $(\phi(A))$, $(1 - v^*v)A(1 - vv^*)$ is contained in kernel ϕ; in particular, $(1 - v^*v)w(1 - vv^*)$ is in kernel ϕ, but $(1 - v^*v)w(1 - vv^*) = w = u_1 - e_1v$. Thus $u_1 - e_1v$ is in kernel ϕ. Similarly $u_2 - e_2v$ is in kernel ϕ. Consequently $(u_1 + u_2) - v$ is in kernel ϕ. We have already seen that $u_1 + u_2$ is an extremal point of A.

Proof of Theorem 4. All algebras mentioned are assumed to be AW^*. The terminology is taken from [4].

The case where A is of type I_n follows at once from the definitions—i.e., A is of type I_n if and only if it has matrix units e_{ij}, $1 \leqq i, j \leqq n$, with all e_{ii} being Abelian projections. But the properties of being a set of matrix units, or an Abelian projection are both preserved by homomorphisms.

Note that if p is a projection infinite (A), and φ a * homomorphism of A into a B^* algebra, then $\varphi(p)$ is either zero or infinite $\varphi(A)$. For if $p = p_1 + p_2$ with $p \backsim p_1 \sim p_2$ and $\varphi(p) \neq 0$, we have one of $\varphi(p_1), \varphi(p_2) \neq 0$; say $\varphi(p_1)$. Then $\varphi(p_2) < \varphi(p)$, $\varphi(p_2) \sim \varphi(p)$, and so

$\varphi(p)$ is infinite. (We thank the referee for supplying this argument to replace one which was somewhat grandiose.) In consequence, the homomorphic image of an algebra of infinite type is again of infinite type or else zero. An easy consequence of the first part of the proof of Theorem 3 is that each projection in the image algebra comes from a projection in the pre-image. This, with the previous remark, shows that the image of an algebra of Type III is again of Type III.

Conversely, the image of an algebra of finite type is again of finite type; for an AW^* algebra is of finite type if and only if all its extremal elements are unitary. By Theorem 3, the latter property must be inherited by any homomorphic image.

LEMMA. *If q is an abelian projection of B, there is an abelian projection p of A with $\phi p = q$.*

Proof. As we have noted, the proof of Theorem 3 can be used to find a projection p_0 of A with $\phi p_0 = q$. Consider the AW^* algebra $p_0 A p_0$. We know from [4], that any AW^* algebra can be written as a direct sum of two ideals, the first of which is a 2×2 matrix algebra, and the second of which is commutative. Thus we have

$$p_0 A p_0 = A_1 \oplus A_2$$

A_1 a 2×2 matrix algebra, A_2 commutative. Thus $p_0 = p_1 + p_2$, $p_i \in A_i$. Since $p_2 A p_2$ is contained in A_2, it is commutative—i.e., p_2 is an abelian projection.

We observe that a homomorphism of a $n \times n$ matrix algebra $(n \geqq 2)$ into a commutative ring must be zero; for if e_{ij} are matrix units, $\phi e_{ii} = \phi e_{ii} e_{ij} e_{ji} = \phi e_{ij} \phi e_{ii} \phi e_{ji} = 0$. Since ϕ is a homomorphism from $p_0 A p_0$ to qBq, this means that $\phi p_0 = \phi p_2$, as desired.

With these observations, the implications of Theorem 4 may be read on at once.

BIBLIOGRAPHY

1. J. Calkin, *Two-sided ideals and congruences in the ring of bounded operators in Hilbert space*, Annals of Math., (2) **42** (1941), 839–873.

2. N. Jacobson, *Some remarks on one-sided inverses*, Proceedings of the A.M.S., **1** (1950), 352–55.

3. R. V. Kadison, *Isometries of operator algebras*, Annals of Math., (2) **54** (1951), 325–38.

4. I. Kaplansky, *Projections in Banach algebras*, Annals of Math., (2) **53** (1951), 235–49.

5. S. Sakai, *The theory of W^* algebras*, Mimeographed lecture notes, Yale University, 1962.

6. T. Yen, *Trace on finite AW^* algebras*, Duke Math. Journal, **22** (1955), 207–22.

7. P. Miles, *Order isomorphisms of $B*$ algebras*, Transactions of the A.M.S., **107** (1962), 217–236.

UNIVERSITY OF WISCONSIN

ON THE DIFFERENCE AND SUM OF A BASIC SET OF POLYNOMIALS

W. F. NEWNS

1. Introduction. For any basic set (p_n) of polynomials, the differenced set (u_n) and the sum (v_n) have been defined and studied by Mikhail & Nassif [1, 2], who obtained the best possible bound for the orders of (u_n) and (v_n) when (p_n) has a given order ω. Their method was to estimate directly the expressions for the orders of (u_n) and (v_n).

The object of the present note is to indicate how these results can be obtained by an alternative line of reasoning which the author believes may throw more light on them. He observes also that either approach can be used to go a little further and determine not only the order but the type of the sets. In fact:

THEOREM 1. *If* (p_n) *is of increase* (ω, γ), *then* (u_n) *has increase at most* $\max\{(\omega, \gamma), (1, 1/2\pi)\}$.

THEOREM 2. *Let* (p_n) *have increase* (ω, γ). *Then*

(i) *If* $\limsup D_n/n = \alpha < \infty$, (v_n) *has increase at most* $(\omega + \alpha, \infty)$,

(ii) *If* $D_n^{D_n/n} = 0(n^\alpha)$ *and* $\gamma < \infty$ (*so that* $\omega > 0$), *the increase of* (v_n) *is at most* $(\omega + \alpha, 0)$.

Case (ii) *applies in particular* (*with* $\alpha = 1$) *to simple sets.*

2. Spaces of integral functions. Let f be an integral function, ρ its order. If $0 < \rho < \infty$, the rate of increase of f is (ρ, σ) where σ is the type of f. If $\rho = 0$ we put $\sigma = \infty$, and if $\rho = \infty$ we put $\sigma = 0$ and again define the rate of increase of f as (ρ, σ). We use lexicographic order, so that $(\rho_1, \sigma_1) \leqq (\rho_2, \sigma_2)$ means that either $\rho_1 < \rho_2$ or $\rho_1 = \rho_2$ and $\sigma_1 \leqq \sigma_2$.

The set $I(\rho, \sigma)$ of all integral functions of increase not exceeding (ρ, σ) is a vector space under the usual operations. The space $I(\infty, 0)$ of all integral functions is an $\mathscr{F}$-space under the topology of uniform convergence on compact sets (the compact-open topology). If $\rho < \infty$, $I(\rho, \sigma)$ is an $\mathscr{F}$-space under a (unique) topology $\mathscr{T}(\rho, \sigma)$ finer than that induced on it by the topology $\mathscr{T}(\infty, 0)$ of $I(\infty, 0)$, (c.f. [3] § 5, p. 438). These may be defined as follows. Put

$$|f|_{\infty, r} = \sum |a_k| r^k$$
$$|f|_{\rho, r} = \sum (k/e\rho)^{k/\rho} |a_k| r^k$$

Received July 1, 1963.

for $f(z) = \sum a_n z^n$. Then $\mathscr{T}(\rho, 0)$ is defined by the semi-norms $|f|_{\rho,r}$ for all finite r, $\mathscr{T}(\rho, \infty)$ by $|f|_{\rho_1,r}$ for $\rho_1 > \rho$ and all finite r, and $\mathscr{T}(\rho, \sigma)$ for $0 < \sigma < \infty$ by $|f|_{\rho,r}$ for $r < \sigma^{-1/\rho}$.

We denote by $I_0(\rho, \sigma)$ the set of those functions of $I(\rho, \sigma)$ which vanish at the origin: $I_0(\rho, \sigma) = \{g \in I(\rho, \sigma): g(0) = 0\}$. $I_0(\rho, \sigma)$, being a closed subspace of $I(\rho, \sigma)$, is an $\mathscr{F}$-space under the induced structure.

3. Rate of increase of a basic set. As with functions, we define the rate of increase of a basic set to be the pair (ω, γ) where ω is the order, γ the type if $0 < \omega < \infty$ and similar conventions where $\omega = 0$ or ∞. We again use lexicographic order and recall the following result [6]:

THEOREM 3. *A basic set* (p_n) *is of increase not exceeding* (ω, γ) *if and only if it is effective for* $I(\rho, \sigma)$ *in* $\mathscr{T}(\infty, 0)$ *for all* $(\rho, \sigma) < (1/\omega, 1/\gamma)$.

4. The difference operator. For any integral function g we put $\Delta g = f$, where

$$f(z) = g(z + 1) - g(z) .$$

THEOREM 4. *The difference operator* Δ *is a continuous linear mapping of* $I(\rho, \sigma)$ *onto itself.*

A proof that Δ is a linear mapping of $I(\rho, \sigma)$ onto itself will be found in [5] (pp. 21–24) and [4][1] (Theorem I). Continuity of Δ for the compact-open topology (induced by $\mathscr{T}(\infty, 0)$) is easily checked. Continuity for $\mathscr{T}(\rho, \sigma)$ now follows from the closed graph theorem.

Clearly Δ is not a bijection: its kernel contains not only constants but any function of period 1. Since the only functions of period 1 and increase less than $(1, 2\pi)$ are constants, we have:

THEOREM 5. *If* $(\rho, \sigma) < (1, 2\pi)$, *then* Δ *is an isomorphism between the* $\mathscr{F}$*-spaces* $I_0(\rho, \sigma)$ *and* $I(\rho, \sigma)$.

Under the hypotheses of Theorem 5, Δ: $I_0(\rho, \sigma) \to I(\rho, \sigma)$ has a continuous inverse $\mathscr{S}$: $I(\rho, \sigma) \to I_0(\rho, \sigma)$. If $f = \Delta g$ we have $g = \mathscr{S} f$ and call g the *sum* of f.

5. The differenced set. In defining the differenced set (u_n) of a given basic set (p_n), there is no loss of generality in taking $p_0(z) = 1$.

[1] For this reference, which he had failed to trace, the author is indebted to Dr. J. M. Whittaker.

Then

$$u_n = \Delta p_{n+1}$$

and the set (u_n) is basic with respect to the representation

$$z^n = \sum_0^\infty \Pi_{k+1}(\phi_{n+1}) u_k(z) ,$$

where

$$\phi_{n+1}(z) = \mathscr{S} z^n .$$

To prove Theorem 1, let the increase of (p_n) be (ω, γ). If ω is infinite there is nothing to prove, so we suppose $\omega < \infty$. Let $(\rho, \sigma) < \min\{(1/\omega, 1/\gamma), (1, 2\pi)\}$, $\sigma < \infty$ and let $f \in I(\rho, \sigma)$. Then $g = \mathscr{S} f \in I(\rho, \sigma)$ and (Theorem 3)

$$g = \sum_0^\infty \Pi_k(g) p_k \qquad (\mathscr{T}(\infty, 0)) .$$

Since Δ is continuous in $\mathscr{T}(\infty, 0)$,

$$f = \Delta g = \sum_0^\infty \Pi_k(g) \Delta p_k = \sum_1^\infty \Pi_k(g) u_{k-1} = \sum_0^\infty \Pi_{k+1}(g) u_k ,$$

showing that f is represented in $\mathscr{T}(\infty, 0)$ by a series of the required form. To prove that this is the basic series of f, it is obvious that $f \to \Pi_{k+1}(g)$ is continuous (being composed of the continuous functions $\mathscr{S}$ and Π_{k+1}) and hence the series is basic under the inverse matrix

$$\Pi_{k+1}(\mathscr{S} z^n) = \Pi_{k+1}(\phi_{n+1}) .$$

Theorem 1 now follows from Theorem 3.

REMARK. Nothing in this argument depends on the $p_n(z)$ being polynomials. They may be integral functions of any order.

6. **The sum of a basic set.** Given a basic set (p_n) of polynomials,[2] the sum (v_n) is defined by

$$v_n = \mathscr{S} p_{n-1} \quad (n > 0) , \qquad v_0 = 1.$$

This set is basic with respect to the representation

$$z^n = \sum_1^\infty \Pi_{k-1}(\vartheta_{n-1}) v_k \qquad (n > 0) ,$$

where $\vartheta_{n-1}(z) = \Delta z^n$.

[2] In the definition of (v_n) we could allow the (p_n) to be integral functions of increase $<(1, 2\pi)$. However, Theorem 2 applies only to sets of polynomials.

Proceeding heuristically, let f be given (with $f(0) = 0$) and put $g = \Delta f$. Then

$$g = \sum_0^\infty \Pi_k(g) p_k \tag{1}$$

and we obtain

$$f = \mathscr{S} g = \sum_0^\infty \Pi_k(g) \mathscr{S} p_k = \sum_0^\infty \Pi_k(g) v_{k+1} \tag{2}$$

a series with continuous coefficients (composed of Δ and Π_{k-1}) which is therefore basic under

$$\Pi_{k-1}(\Delta z^n) = \Pi_{k-1}(\vartheta_{n-1}) .$$

This argument is valid for all $f \in I_0(\rho_0, \sigma_0)$ only if (ρ_0, σ_0) satisfies certain requirements. For equation (1) to hold in (say) $\mathscr{T}(\infty, 0)$ we need $(\rho_0, \sigma_0) < (1/\omega, 1/\gamma)$. For $\mathscr{S}$ to be well-defined, we need $(\rho_0, \sigma_0) < (1, 2\pi)$. But to apply $\mathscr{S}$ to (1) to obtain (2), we need (1) to hold in a topology $\mathscr{T}(\rho_1, \sigma_1)$ in which $\mathscr{S}$ is continuous, i.e. one for which $(\rho_1, \sigma_1) < (1, 2\pi)$. The problem arises as to which (ρ_0, σ_0) will satisfy these requirements and the answer is given by:

THEOREM 6. *Suppose that* (p_n) *is effective for* $I(\rho, \sigma)$ *in* $\mathscr{T}(\infty, 0)$, $(0 < \rho < \infty,\ 0 < \sigma < \infty)$, *and that* $D_n^{D_n/n} = 0(n^\beta)$. *Given* $\rho_1 (0 < \rho_1 < \infty)$ *put* $(1/\rho_0) = (1/\rho) + (\beta/\rho_1)$. *Then* (p_n) *is effective for* $I(\rho_0, \sigma_0)$ *in* $\mathscr{T}(\rho_1, 0)$ *for all finite* σ_0.

We first complete the proof of Theorem 2. For case (i), let $\rho_0 < (1/\omega + \alpha)$ and choose $\beta > \alpha$ such that $\rho_0 < (1/\omega + \beta)$. Put $(1/\rho_0) = (1/\rho) + \beta$ so that $\rho < (1/\omega)$. The hypotheses of Theorem 6 hold with $\rho_1 = 1$ and so the heuristic argument above holds for (ρ_0, σ_0) for any finite σ_0. This being true for any $\rho_0 < (1/\omega + \alpha)$, case (i) follows from Theorem 3.

For case (ii), we put $\rho = (1/\omega)$, $\beta = \alpha$ and choose $\sigma < (1/\gamma)$. We conclude similarly that (v_n) is effective for $I(\rho_0, \sigma_0)$ in $\mathscr{T}(\infty, 0)$ when $(1/\rho_0) = \omega + \alpha$ and σ_0 is finite. By Theorem 3, this is equivalent to the stated result.

We now prove Theorem 6. Put

$$\gamma = \begin{cases} \sup \{\beta - D_n/n\} & (e\rho_1 \geqq 1) \\ \inf \{\beta - D_n/n\} & (e\rho_1 < 1) . \end{cases}$$

Since $D_n \geqq n$ and $\limsup D_n/n \leqq \beta, \gamma$ is finite. Also we are dealing with a Cannon set so that effectiveness is equivalent to absolute effectiveness. Let $0 < \sigma_0 < \infty$. We have to prove ([3], §§ 7, 8): given

$r_1 < \infty$, there exist M and $r_0 < \sigma_0^{-1/\rho_0}$ such that

$$\sum_l |\pi_{nl}| \sum_k \left(\frac{k}{e\rho_1}\right)^{k/\rho_1} |p_{lk}|\, r_1^k \leqq M\left(\frac{n}{e\rho_0}\right)^{n/\rho_0} r_0^n .$$

Put $s = \rho_0^{1/\rho_0}\rho^{-1/\rho}\rho_1^{-\beta/\rho_1}c^{1/\rho_1}(e\rho_1)^{\ /\rho_1}\sigma^{-1/\rho}\sigma_0^{1/\rho_0}$ where c is chosen large enough for $s \geqq 1$ and $D_n^{D_n/n} \leqq cn^\beta$. The left-hand member of the inequality to be proved may be written

$$\sum_l |\pi_{nl}| \sum_k \left(\frac{k}{e\rho_1}\right)^{k/\rho_1} s^{-k} |p_{lk}|\, (r_1 s)^k .$$

The largest value of k appearing in this is D_n. Since the sequence $(k/e\rho_1)^{k/\rho_1}s^{-k}$ increases to ∞ from some point on, we have $(k/e\rho_1)^{k/\rho_1}s^{-k} \leqq A(D_n/e\rho_1)^{D_n/\rho_1}s^{-D_n}$ for some A and all n. Also

$$\sum_l |\pi_{nl}| \sum_k |p_{lk}|\,(r_1 s)^k \leqq B \sum_l |\pi_{nl}|\, M_l(R) = B\omega_n(R)$$

for $R > r_1 s$, and since (p_n) is effective for $I(\rho, \sigma)$ in $\mathscr{T}(\infty, 0)$, there exist C and $r < \sigma^{-1/\rho}$ such that

$$\omega_n(R) \leqq C\left(\frac{n}{e\rho}\right)^{n/\rho} r^n .$$

Finally, since $D_n \geqq n$ and $s \geqq 1$ we have $s^{-D_n} \leqq s^{-n}$. Thus the left-hand member of the inequality to be proved does not exceed

$$A\left(\frac{D_n}{e\rho_1}\right)^{D_n/\rho_1} s^{-n} BC\left(\frac{n}{e\rho}\right)^{n/\rho} r^n$$
$$\leqq ABCc^{n/\rho_1}\left(\frac{n}{e\rho_1}\right)^{\beta n/\rho_1}(e\rho_1)^{(\beta n - D_n)/\rho_1}s^{-n}\left(\frac{n}{e\rho}\right)^{n/\rho} r^n \leqq M\left(\frac{n}{e\rho_0}\right)^{n/\rho_0} r_0^n ,$$

where $r_0 = \rho_0^{1/\rho_0}c^{1/\rho_1}\rho_1^{-\beta/\rho_1}(e\rho_1)^{\ /\rho_1}s^{-1}\rho^{-1/\rho}r < \sigma_0^{-1/\rho_0}$, as required.

7. **Examples.** Let (ν_n) be a sequence of even nonnegative integers, (γ_n) a sequence of real numbers and ω a nonnegative real number. Consider the set

$$p_{2n}(z) = z^{2n} ,$$
$$p_{2n+1}(z) = z^{2n+1} + ((2n+1)!)^\omega \gamma_n^{2n+1} z^{\nu_n} .$$

EXAMPLE (i). $\nu_n = 2n$, $\gamma_n = \log(2n+1)$. It will be found that (p_n) has increase (ω, ∞) and (v_n) has increase $(\omega + 1, \infty)$.

EXAMPLE (ii). $\nu_n = 2n$, $\gamma_n = (1/\log(2n+1))$ $(n > 0)$, $\omega > 0$. Here (p_n) is of increase $(\omega, 0)$ and (v_n) of increase $(\omega + 1, 0)$.

EXAMPLE (iii). Choose ν_n so that $\nu_n/2n \to \alpha \geqq 1$, but $((\nu_n!)^{1/2n}/(2n)^\alpha) \to$

∞. Put $\gamma_n = \sqrt{}((2n)^\alpha/(\nu_n!)^{1/2n})$, $\omega > 0$. Here $\limsup D_n/n = \alpha$ and (p_n) is of increase $(\omega, 0)$, but (v_n) is of increase $(\omega + \alpha, \infty)$.

REFERENCES

1. N. N. Mikhail and M. Nassif, Assiut Univ. Bull. Sci. Tech.
2. ———, Pacific J. Math., **11** (1961), 1099-1107.
3. W. F. Newns, Phil. Trans. Roy. Soc. London, A 245, (1953), 429-468.
4. Sheila Scott, Proc. Camb. Phil. Soc. XXXI (1935) 543-554.
5. J. M. Whittaker, *Interpolatory function theory*, Cambridge University Press, (1935).
6. ———, *Sur les Séries de Base de Polynomes Quelconques*, Paris: Gauthier-Villars, (1949).

THE UNIVERSITY, LIVERPOOL

RINGS ALL OF WHOSE FINITELY GENERATED MODULES ARE INJECTIVE

B. L. OSOFSKY

The main purpose of this paper is to prove that a ring all of whose finitely generated modules are injective must be semi-simple Artin.[1]

We begin with the following information about the class of rings under consideration:

LEMMA 1. *Let R be a ring with identity, and assume each cyclic right R-module is injective. Then R is regular in the sense of von Neumann and R is right self injective.*

Proof. For any ring R with identity, it is easy to see that a right ideal I of R is generated by an idempotent if and only if I is a direct summand of the right R-module R_R. If I is an injective right ideal of R, then I is a direct summand of R_R, and therefore is generated by an idempotent. Thus if every cyclic right R-module is injective, each principal right ideal aR generated by $a \in R$ is generated by an idempotent, that is $aR = eR$ for some $e = e^2 \in R$. Then there exist $x, y \in R$ such that $e = ax$, and $a = ey$. It follows that $ea = e(ey) = ey = a$ and $a = ea = axa$. Thus R is a regular ring, and since R_R is generated by the identity, R_R is injective.

Let M_R denote a right module over a ring R. If P, N are submodules of M, let $P \,'\!\supseteq N$ signify that P is an essential extension of N. (See Eckmann and Schopf [2].) Then N is an essential submodule of P.

For each $x \in M$, let $x^R = \{r \in R \mid xr = 0\}$. The singular submodule $Z(M)$ is defined by:

$$Z(M) = \{x \in M \mid R_R \,'\!\supseteq x^R\}$$

$Z(R_R)$ is actually a two sided ideal of R.

If $e = e^2 \in R$, and if $x \in e^R \cap eR$, then $x = ex = 0$ and so $e^R \cap eR = 0$. Thus $Z(R_R)$ contains no idempotents $\neq 0$. In particular, if R is a

Received July 18, 1963. This result forms part of the author's doctoral dissertation written at Rutgers University under the direction of Professor Carl Faith. The author gratefully acknowledges partial support from the National Science Foundation under grant G-19863.

[1] After the author obtained this characterization of rings whose cyclic modules are injective, a translation of a recent paper [5] by L. A. Skornjakov was published and brought to the author's attention. Although Skornjakov states the major portion of the author's main theorem, his proof is incorrect. In the proof of his Lemma 9, which is crucial to his proof of the theorem, Skornjakov assumes that the injective hull of a submodule in an injective module must be a unique submodule, whereas in general it is unique only up to isomorphism.

Added in proof April 13, 1964. This lemma is actually false. See the author's dissertation for a counter-example.

regular ring, $Z(R_R) = 0$. (Cf. R. E. Johnson [4]).

We need the following important (and known [3]) lemma:

LEMMA 2. *Let M_R be a module such that $Z(M) = 0$. Then each submodule N of M has a unique maximal essential extension N^* in M.*

Proof. Let $\{M_i \mid i \in I\}$ be the set of all submodules of M such that $M_i\ '\supseteq N$. Set $N^* = \sum_{i \in I} M_i$. Then if $N^*\ '\supseteq N$, N^* must be the unique maximal essential extension of N in M since it contains every essential extension of N in M.

For each $y \in M$, let

$$(N : y) = \{r \in R \mid yr \in N\} .$$

If $M \supseteq Q\ '\supseteq N$, then $R_R\ '\supseteq (N : y)$ for all $y \in Q$. (This follows, since any non-zero right ideal I of R which satisfies $I \cap (N : y) = 0$, also satisfies $yI \neq 0$ and $yI \cap N = 0$, a contradiction.)

Now let $0 \neq x = x_{i_1} + \cdots + x_{i_n}$, $0 \neq x_{i_j} \in M_{i_j}$, $j = 1, \cdots, n$, be any element of N^*. Then

$$(N : x) \supseteq \bigcap_{j=1}^{n} (N : x_{i_j}) .$$

Now $M_{i_j}\ '\supseteq N$, so $R_R\ '\supseteq (N : x_{i_j})$, and therefore $R_R\ '\supseteq \bigcap_{j=1}^{n}(N : x_{i_j})$, hence $R_R\ '\supseteq (N : x)$. Since $Z(M) = 0$, $x(N : x) \neq 0$, and so $x(N : x)$ is a nonzero submodule of $xR \cap N$. This proves $N^*\ '\supseteq N$ as asserted.

We next consider certain properties of idempotents in a right self injective regular ring. Let $N \sim A$ denote the set theoretic complement of a set A in a set N.

LEMMA 3. *Let $\{e_n \mid n \in N\}$ be a set of orthogonal idempotents in a right self injective regular ring. Then for every subset A of N, there exists an idempotent $E_A \in R$ such that*

$$E_A e_n = e_n \qquad \textit{for all } n \in A$$
$$e_{n'} E_A = E_A e_{n'} = 0 \qquad \textit{for all } n' \in N \sim A$$
$$E_A + E_{N \sim A} = E_N .$$

Proof. Since R is regular, $Z(R_R) = 0$. Then, by Lemma 2, each right ideal I of R has a unique maximal essential extension I^* in R. Since R_R is injective, by [2] I has an injective hull in R_R which is a maximal essential extension of I in R_R. Thus each right ideal I has a unique injective hull I^* in R_R. Then as remarked in the proof of Lemma 1, there exists $e = e^2 \in R$ such that $I^* = eR$.

Hence for any subset A of N, there exists an idempotent $e_A \in R$ such that

$$e_A R = \Big(\sum_{n\in A} e_n R\Big)^* .$$

Since $\{e_n \mid n \in N\}$ are orthogonal,

$$\Big(\sum_{n\in A} e_n R\Big) \cap \Big(\sum_{n'\in N\sim A} e_{n'} R\Big) = 0 .$$

Then $e_A R \cap e_{N\sim A} R = 0$ (for $x \neq 0 \in e_A R \cap e_{N\sim A} R$ implies $xR \cap \sum_{n\in A} e_n R \cap \sum_{n'\in N\sim A} e_{n'} R \neq 0$, a contradiction.) Thus the sum $e_A A + e_{N\sim A} R$ is direct, and since each summand is injective, $e_A R \oplus e_{N\sim A} R$ is injective. Since injective hulls of right ideals of R are unique,

$$\Big(\sum_{n\in N} e_n R\Big)^* \supseteqq \Big(\sum_{n\in A} e_n R\Big)^* ,$$

$$\Big(\sum_{n\in N} e_n R\Big)^* \supseteqq \Big(\sum_{n'\in N\sim A} e_{n'} R\Big)^*$$

so $(\sum_{n\in N} e_n R)^* \supseteqq e_A R \oplus e_{N\sim A} R \supseteqq \sum_{n\in N} e_n R$. Then it follows $(\sum_{n\in N} e_n R)^* = e_A R \oplus e_{N\sim A} R$. Set $E_N = e_N$, where $e_N R = (\sum_{n\in N} e_n R)^*$. Let E_A (respectively $E_{N\sim A}$) be the projection of the idempotent E_N on $e_A R$ (respectively $e_{N\sim A} R$). We note that $E_A + E_{N\sim A} = E_N$ by definition, and E_A is simply the projection of the identity element of R on $e_A R$ with respect to the direct decomposition $R = (1 - E_N)R \oplus e_A R \oplus e_{N\sim A} R$. Thus E_A and $E_{N\sim A}$ are orthogonal. Furthermore

$$E_A e_n = e_n \qquad \forall n \in A$$

$$E_A e_{n'} = E_A E_{N\sim A} e_{n'} = 0 \qquad \forall n' \in N \sim A .$$

Since $e_A R' \supseteqq \sum_{n\in A} e_n R$, $(\sum_{n\in A} e_n R : E_A)$ is an essential right ideal of R. But

$$e_{n'} E_A\Big(\sum_{n\in A} e_n R : E_A\Big) \subseteqq e_{n'}\Big(\sum_{n\in A} e_n R\Big) = 0 \qquad \forall n' \in N \sim A .$$

Thus $e_{n'} E_A \in Z(R_R) = 0$, and we conclude $e_{n'} E_A = 0 \; \forall n' \in N \sim A$.

LEMMA 4. *Let R be a right self injective regular ring which contains an infinite set of orthogonal idempotents $\{e_n \mid n \in N\}$. Let $I = \sum \oplus e_n R$. For $A \subseteqq N$, let E_A be defined as in Lemma 3. Then a set $S_{\mathfrak{A}} = \{E_A \mid A \in \mathfrak{A}\}$, where each A is infinite, is independent modulo I, that is, $\sum_{A\in\mathfrak{A}} (E_A R + I)$ is direct in $R - I$, if and only if for any finite set $\{A_i \mid i = 1, \cdots, n\} \subseteqq \mathfrak{A}$, $A_i \cap \bigcup_{j\neq i} A_j$ is a finite subset of N, $1 \leqq i \leqq n$.*

Proof. Assume $S_{\mathfrak{A}}$ is independent modulo I, and let $\{A_i \mid i = 1, \cdots, n\} \subseteqq \mathfrak{A}$. Set $C = C_{ij} = A_i \cap A_j$. For all i and $j \neq i$, $E_{A_i} R \supseteqq \sum_{n\in C} e_n R$ and $E_{A_j} R \supseteqq \sum_{n\in C} e_n R$. Thus $E_{A_i} R \supseteqq (\sum_{n\in C} e_n R)^* = E_C R$ and

$E_{A_j}R \supseteqq (\sum_{n\in C} e_nR)^* = E_CR$. Since $0 = E_C - E_C = E_{A_i}E_C - E_{A_j}E_C$, $E_C \in I$. Then for all but a finite number of $n \in C$, $e_nE_C = 0$. Since this implies $e_n = e_nE_Ce_n = 0$ which is true for no n, C must be finite. Then $A_i \cap \bigcup_{j\neq i} A_j$ is a finite union of finite sets, and thus finite.

Now assume $A_i \cap \bigcup_{j\neq i} A_j$ is finite, and let $\sum_{j=1}^n E_{A_j}r_j \in I$. If $m \notin A_i$, $e_mE_{A_i}r_i = 0$ by Lemma 3. If $m \in A_i \sim \bigcup_{j\neq i} A_j$, $e_m \sum_{j=1}^n E_{A_j}r_j = e_mE_{A_i}r_i$. Since $\sum_{j=1}^n E_{A_j}r_j \in I$, there are at most a finite number of $m \in A_i \sim \bigcup_{j\neq i} A_j$ such that $e_mE_{A_i}r_i \neq 0$. Since $A_i \cap \bigcup_{j\neq i} A_j$ is also finite, the set

$$B = \{m \in N \mid e_mE_{A_i}r_i \neq 0\}$$

must be finite.

Now for all $n' \in N \sim B$,

$$\begin{aligned} 0 = e_{n'}E_{A_i}r_i &= e_{n'}([1 - E_N] + E_{N\sim B} + E_B)E_{A_i}r_i \\ &= e_{n'}E_{N\sim B}E_{A_i}r_i \,. \end{aligned}$$

Assume $E_{N\sim B}E_{A_i}r_i \neq 0$. Then, since $E_{N\sim B}R\,' \supseteqq \sum_{n'\in N\sim B} e_{n'}R$, there is an $s \in R$ such that $E_{N\sim B}E_{A_i}r_is \neq 0 \in \sum_{n'\in N\sim B} e_{n'}R$, so for some $n' \in N \sim B$, $e_{n'}E_{N\sim B}E_{A_i}r_is \neq 0$, a contradiction.

Then

$$E_{A_i}r_i = ([1 - E_N] + E_{N\sim B} + E_B)E_{A_i}r_i = E_BE_{A_i}r_i \,.$$

Since a finite direct sum of injective modules is injective, $\sum_{i\in B} e_iR = (\sum_{i\in B} e_iR)^* = E_BR$ and $E_BR \subseteqq I$. It follows that $E_{A_i}r_i \in I$.

LEMMA 5. *Let R be a right self injective regular ring which contains an infinite set of orthogonal idempotents $\{e_n \mid n \in N\}$. If $I = \sum_{n\in N} e_nR$, then $R - I$ is not an injective R-module.*

Proof. Let $\{A_i \mid i = 1, 2, \cdots\}$ be a countable family of subsets $A_i \subseteqq N$ such that $\{E_{A_i} \mid i = 1, 2, \cdots\}$ are independent in $R - I$. For example, the A_i may be disjoint countable subsets of N.

Let $\mathscr{P}$ denote the family of sets $S_{\mathfrak{A}} = \{E_{B_\alpha} \mid B_\alpha \subseteqq N, \alpha \in \mathfrak{A}\}$ where $\mathfrak{A}$ is some index set, such that $S_{\mathfrak{A}} \supseteqq \{E_{A_i} \mid i = 1, 2, \cdots\}$ and $S_{\mathfrak{A}}$ is independent modulo I. Partially order $\mathscr{P}$ by inclusion. Since independence modulo I depends only on finite sets of idempotents, $\mathscr{P}$ is inductive. By Zorn's lemma, select a maximal element $S \in \mathscr{P}$.

Let $J = \sum_{E_B\in S} E_BR$. Define $\varphi: J \to R - I$ by

$$\begin{aligned} \varphi(E_{A_i}) &= E_{A_i} + I \qquad \forall i = 1, 2, \cdots \\ \varphi(E_B) &= 0 + I \qquad \forall E_B \in S \sim \{E_{A_i}\} \\ \varphi\Big(\sum_{k=1}^n E_{B_k}r_k\Big) &= \sum_{k=1}^n \varphi(E_{B_k})r_k \qquad E_{B_k} \in S,\ r_k \in R \,. \end{aligned}$$

$\sum_{k=1}^n E_{B_k}r_k = 0$ implies $E_{B_k}r_k \in I$ since the idempotents of S are independent modulo I. Hence $\varphi(E_{B_k})r_k = 0 + I$, and $\varphi(\sum_{k=1}^{n} E_{B_k} r_k) = 0 + I$. Thus φ is a map which is clearly an R homomorphism.

Assume φ is induced by left multiplication by $m + I$ in $R - I$, $m \in R$. Then

$$(1) \qquad mE_{A_i} - E_{A_i} \in I \qquad \forall i = 1, 2, \cdots$$

and

$$(2) \qquad mE_B \in I \qquad \forall E_B \in S \sim \{E_{A_i}\}.$$

From (1) we conclude that for all but a finite number of $n \in A_i$, $e_n(mE_{A_i} - E_{A_i}) = 0$ and $e_n m E_{A_i} e_n = e_n E_{A_i} e_n$. Thus $e_n m e_n = e_n$ by Lemma 3.

From (2) we conclude that for all but a finite number of $n' \in B$, $e_{n'} m E_B = 0$, and $e_{n'} m E_B e_{n'} = e_{n'} m e_{n'} = 0$.

Let $j_1 \in A_1$, $e_{j_1} m e_{j_1} = e_{j_1}$. Select $j_{n+1} \in A_{n+1}$ such that

$$e_{j_{n+1}} m e_{j_{n+1}} = e_{j_{n+1}}$$

and

$$j_{n+1} \notin A_k \qquad \text{for all } k < n + 1 .$$

This is possible since $\{j \in A_{n+1} \mid e_j m e_j = e_j\}$ is infinite and Lemma 4 implies $A_{n+1} \cap \bigcup_{k=1}^{n} A_k$ is finite.

Since S is maximal in $\mathscr{P}$, by Lemma 4 $\{j_n \mid n = 1, 2, \cdots\}$ thus defined must have an infinite number of elements in common with some $B \subseteqq N$ such that $E_B \in S$. $B \neq A_i$ $i = 1, 2, \cdots$ since $j_n \notin A_i$ for all $n > i$. Therefore $\varphi(E_B) = 0$, and $e_{j'} = e_{j'} m e_{j'} = 0$ for all but a finite number of $j' \in B \cap \{j_n \mid n = 1, 2, \cdots\}$. This contradicts the assumption that $B \cap \{j_n \mid n = 1, 2, \cdots\}$ is an infinite set. Thus φ is not induced by left multiplication by $m + I$ in $R - I$. Hence $R - I$ is not injective. (See [1], p. 8.)

THEOREM. *Let R be a ring with* 1. *Then the following conditions are equivalent*:

(a) *R is semi-simple Artin.*

(b) *Every finitely generated right R-module is injective.*

(c) *Every cyclic right R-module is injective.*

Proof. (*a*) $\Rightarrow$ (*b*). By ([1], p. 11, Theorem 4.2), every right module over a semi-simple Artin ring R is injective, and so every finitely generated right R-module is injective.

(b) $\Rightarrow$ (c). Since every cyclic R-module is finitely generated by one element, (c) is a special case of (b).

(c) $\Rightarrow$ (a). If every cyclic R-module is injective, by Lemma 1, R

is right self injective and regular. By Lemma 5, R cannot contain an infinite set of orthogonal idempotents. It is well known that this condition in any regular ring R implies that R satisfies the minimum condition and hence is semi-simple Artin.

COROLLARY. *Let R be a right self injective, hereditary ring with identity. Then R is semi-simple Artin.*

Proof. R hereditary is equivalent to the condition that every quotient of an injective R-module is injective. (See [1], p. 14.) Since every cyclic module is isomorphic to a quotient of the injective module R_R, every cyclic R-module is injective. Therefore by the theorem R is semi-simple Artin.

REFERENCES

1. H. Cartan and S. Eilenberg, *Homological Algebra*, Princeton University Press, 1956.
2. B. Eckmann and A. Schopf, *Über injektive Moduln*, Archiv der Mathematik, **4** (1953), 75-78.
3. Carl Faith, *Theory of Rings*, Lecture Notes (dittoed), Rutgers University, 1963.
4. R. E. Johnson, *The extended centralizer of a ring over a module*, Proc. Amer. Math. Soc., **2** (1951), 891-895.
5. L. A. Skornjakov, *Rings with injective cyclic modules*, Soviet Mathematics, **4** (1963), 36-38.

RUTGERS, THE STATE UNIVERSITY

TOEPLITZ MATRICES AND INVERTIBILITY OF HANKEL MATRICES

C. R. PUTNAM

1. Introduction. Let $\{c_n\}$, for $n = 0, \pm 1, \pm 2, \cdots$, be a sequence of real numbers satisfying $c_0 = 0, c_{-n} = c_n$ and $0 < \sum_{n=1}^{\infty} c_n^2 < \infty$, and let $f(\theta)$ $(\neq 0)$ be the even function of class $L^2(-\pi, \pi)$ defined by

$$(1) \qquad f(\theta) \sim \sum_{n=-\infty}^{\infty} c_n e^{in\theta} = 2 \sum_{n=1}^{\infty} c_n \cos n\theta \ .$$

Define the Toeplitz matrix T and the Hankel matrices H and K by

$$(2) \quad T = (c_{i-j}),\ H = (c_{i+j-1}) \text{ and } K = (c_{i+j}), \quad \text{where } i, j = 1, 2, \cdots .$$

Then

$$(3) \qquad T = F + K\ , \quad \text{where } F = \int_0^{\pi} f(\theta) dE_0(\theta)\ ,$$

and $\{E_0(\theta)\}$ is the resolution of the identity of the matrix belonging to the quadratic form $2\sum_{n=1}^{\infty} x_n x_{n+1}$. (See [12], p. 837.)

A self-adjoint operator A on a Hilbert space, with the spectral resolution $A = \int \lambda dE(\lambda)$, will be called absolutely continuous if $\| E(\lambda)x \|^2$ is an absolutely continuous function of λ for every element x of the Hilbert space. If the function $f(\theta)$ of (1) is (essentially) bounded then T must be bounded (Toeplitz). Since F must also be bounded, so also are H and K. It was shown in [12], p. 840, using methods involving commutators of operators, that if the function $g(\theta)$ defined by

$$(4) \qquad g(\theta) \sim \sum_{n=1}^{\infty} c_n e^{in\theta}$$

is bounded (hence $f(\theta)$ is also bounded) then T must be absolutely continuous if either

(5) 0 is not in the point spectrum of H (that is, H^{-1} exists) ,

or

(6) F is absolutely continuous .

Rosenblum [17] has shown, using results of Aronszajn and Donoghue [1], that in fact T is *always* (with no restrictions) absolutely continuous.

Received February 8, 1963. This research was supported by the National Science Foundation research grant NSF-G18915.

In addition, it was shown in Putnam [12], using a theorem of Rosenblum [16], and generalized by Rosenblum in [17] using results of Kato [7], that if $\sum_{n=1}^{\infty} n \mid c_{n+1} \mid < \infty$ or, equivalently, if

$$(7) \qquad \sum_{n=1}^{\infty} n \mid c_n \mid < \infty \, ,$$

and if (6) holds, then T and F are unitarily equivalent, so that

$$(8) \qquad T = UFU^* \, , \quad U \text{ unitary} \, .$$

The absolute continuity of F is equivalent to the requirement that

$$(9) \qquad \text{meas}\,\{\theta : f(\theta)\, \varepsilon Z\} = 0 \quad \text{whenever meas } Z = 0 \, .$$

In the present paper a sufficient condition, involving the *negation* of (5), for (6), that is, for the validity of (9), will be obtained.

Before stating the theorem it will be convenient to define the operators $F_k(k = 0, 1, 2, \cdots)$ by

$$(10) \qquad F_k = \int_0^{\pi} f_k(\theta) dE_0(\theta) \, , \quad \text{where } f_k(\theta) \sim \sum_{n=1}^{\infty} c_n n^{-k} \cos n\theta \, .$$

(In particular, $F_0 = F$.)

There will be proved the following

THEOREM 1, *Suppose that*

$$(11) \qquad 0 \textit{ is in the point spectrum of } H \, .$$

Then,

(a) *the point spectrum of* F *is empty, and*

(b) *each of the operators* $F_2, F_3, \cdots$ *is absolutely continuous.*

(c) *If, in addition to* (11), *it is assumed that* $\sum_{n=1}^{\infty} \mid c_n \mid < \infty$, *then* F_1 *is absolutely continuous.*

(d) *If, in addition to* (11), *relation* (7) *is assumed, then* (6) *holds.*

From part (d) of the theorem and the results mentioned earlier there follows the

COROLLARY. *Relations* (7) *and* (11) *imply* (8).

It will remain undecided whether (11) alone, without the additional assumption (7), is sufficient to imply not only the assertion of (a) but also (6). It is interesting to observe though that, if the implication (11) $\rightarrow$ (6) is valid, then either (5) or (6) must hold, and, at least if $g(\theta)$ is bounded, the absolute continuity of T (cf. [17]) can be deduced from the commutator methods of [12] (cf. also [11]) as noted above.

It is to be noted that the function $f(\theta)$ determines explicitly the operator F and its spectrum. On the other hand, the structure of T as determined by $f(\theta)$ is not so clear. It is known however that the spectrum of T, in case T is self-adjoint, is the interval $[m, M]$, where m and M denote the essential lower and upper bounds of $f(\theta)$ (Hartman and Wintner [6], pp. 868, 878). Although necessary and sufficient conditions involving $f(\theta)$, or rather $g(\theta)$, for the boundedness of H (Nehari [10]) and the complete continuity of H (Hartman [4]) are known, apparently no similar results are known relating the spectrum of H to the function $f(\theta)$. Concerning the spectrum of H in certain specific cases, see, e.g., Hartman and Wintner [6], p. 366, Magnus [8].

2. Proof of (a) of Theorem 1. Let $\{x_n\}$ and $\{d_n\}$, for $n = 1, 2, \cdots$, be two sequences of complex numbers satisfying $\sum_{n=1}^{\infty} | x_n |^2 < \infty$ and $\sum_{n=1}^{\infty} | d_n |^2 < \infty$, let $x(\theta) \sim \sum_{n=1}^{\infty} x_n e^{in\theta}$ and $h(\theta) \sim \sum_{n=1}^{\infty} d_n e^{in\theta}$. Then it is easily verified that

$$(2\pi)^{-1}\int_{-\pi}^{\pi} x(\theta)(g^*(\theta) + h(\theta))e^{ij\theta}d\theta = \sum_{n=1}^{\infty} c_{n+j}x_n \tag{12}$$

holds for $j = 0, 1, 2, \cdots$, where the asterisk denotes complex conjugation. If $d_n = c_n$ then $g^*(\theta) + h(\theta) = f(\theta)$ and so 0 is in the point spectrum of H if and only if

$$\int_{-\pi}^{\pi} x(\theta)f(\theta)e^{ij\theta}d\theta = 0 , \quad \text{where } j = 0, 1, 2, \cdots , \tag{13}$$

holds for some $x(\theta) \not\equiv 0$ as defined above. Relation (13) implies that the function $x(\theta)f(\theta)$, of class $L(-\pi, \pi)$, has a Fourier series of the form

$$x(\theta)f(\theta) \sim \sum_{n=0}^{\infty} a_n e^{in\theta} . \tag{14}$$

For a fixed constant $p, 0 < p < \infty$, consider the class H_p (after Hardy; see, e.g., Zygmund [19], p. 158) of functions $A(z)$ analytic in the disk $|z| < 1$ and for which $\int_{-\pi}^{\pi} | A(re^{i\theta}) |^p d\theta$ remains bounded for $0 \leqq r < 1$. If $p \geqq 1$, the class L^{p+} of functions $B(\theta) \varepsilon L^p(-\pi, \pi)$ with Fourier series of the form

$$B(\theta) \sim \sum_{n=0}^{\infty} b_n e^{in\theta} \quad \left(b_n = (2\pi)^{-1}\int_{-\pi}^{\pi} B(\theta)e^{-in\theta}d\theta\right) , \tag{15}$$

coincides with the class of boundary functions $B(\theta) = A(e^{i\theta})$; see Rogosinski and Shapiro [15], p. 293. Furthermore, it is known that if $p > 0$ and if $A(z)$ is of class H_p and if $A(z) \not\equiv$ const., then $A(e^{i\theta}) = a$, for an arbitrary constant a, can hold at most on a set of measure zero. For $p = 2$, this result is due to F. and M. Riesz ([14]); for $p \neq 2$, see F. Riesz [13].

Returning to (14), since $x(\theta)f(\theta)\varepsilon L^{1+}$, it follows that $f(\theta) \neq 0$ almost everywhere. A similar argument with $x(\theta)f(\theta)$ replaced by $x(\theta)(f(\theta) - a)$, for any constant a, shows that $f(\theta) \neq a$ almost everywhere, that is,

$$\text{(16)} \qquad \text{meas } \{\theta : f(\theta) = a\} = 0 .$$

But (16) holds if and only if the operator F has no point spectrum and the proof (a) is complete.

3. Proof of (b) of Theorem 1. In order to show that F_2 is absolutely continuous, it must be shown that the set $S_2 = \{\theta : f_2(\theta) \in Z\}$ is a zero set whenever Z is a zero set. Since $\sum_{n=1}^{\infty} | c_n n^{-1} | < \infty$, $f_2'(\theta)$ is continuous and the set $\{\theta : f_2'(\theta) \neq 0\}$ is open. If its canonical decomposition is the finite or infinite union of open intervals I_n $(n = 1, 2, \cdots)$, then $f_2(\theta)$ is strictly monotone on each I_n. Also, on I_n, both f_2 and its inverse g_n are absolutely continuous. Since $I_n \cap S_2$ is the image under g_n of a subset of Z, it follows (cf., e.g., Natanson [9], p. 249) that

$$\text{(17)} \qquad I_n \cap S_2 \text{ has measure } 0 .$$

If it is shown that $f_2'(\theta) \neq 0$ almost everywhere, it will follow from (17) that meas $S_2 = 0$, as was to be proved.

In order to prove that $f_2'(\theta) \neq 0$ almost everywhere, note that $f_2'(\theta)$ is absolutely continuous and that $f_2''(\theta) = (-1/2)f(\theta)$ almost everywhere. Hence, if $f_2'(\theta) = 0$ on a set of positive measure, then also $f(\theta) = 0$ on a set of positive measure, a contradiction. Hence F_2 is absolutely continuous.

Next, it will be shown that F_3 is absolutely continuous. In the definition of $h(\theta)$, choose $d_n = -c_n$, so that in (12), $k(\theta) = g^*(\theta) + h(\theta) = 2i\sum_{n=1}^{\infty} c_n \sin n\theta$. The argument of § 2 shows that $x(\theta)k(\theta)$ is of class L^{1+} and hence $k(\theta) \neq 0$ almost everywhere. Since $f_3'(\theta)$ is continuous, and since $f_3'''(\theta) = (1/2i)k(\theta)$, an argument similar to that used above shows that F_3 is absolutely continuous.

In like manner, it follows that $F_4, F_5, \cdots$ are absolutely continuous and the proof of (b) is complete.

4. Proof of (c) of Theorem 1. In order to prove the absolute continuity of F_1, it must be shown that the set $S_1 = \{\theta : f_1(\theta) \in Z\}$ is a zero set whenever Z is a zero set. The hypothesis of (c) implies that $f_1'(\theta) = (-1/2i)k(\theta)$ is continuous. Since $k(\theta) \neq 0$ almost everywhere, a relation similar to (17) implies that meas $S_1 = 0$, and the proof of (c) is complete.

5. **Proof of (d) of Theorem 1.** Since (7) implies that $f'(\theta)$ is continuous, then $x^2(\theta)f'(\theta)$ is of class $L(-\pi, \pi)$. It will be shown that $x^2(\theta)f'(\theta)$ is also of class L^{1+}, so that

$$x^2(\theta)f'(\theta) \sim \sum_{n=0}^{\infty} b_n e^{in\theta}, \tag{18}$$

and hence (cf. the above reference to [15]) the F. and M. Riesz theorem can be applied to yield $f'(\theta) \neq 0$ almost everywhere. Once this has been shown, the absolute continuity of F follows by an argument similar to that used above.

There remains then to prove (18). Since $f(\theta)$ is now bounded, it follows from the definition of $x(\theta)$ and (14) that both $x(\theta)$ and $x(\theta)f(\theta)$ belong to L^{2+}. Let $u(z)$ and $v(z)$ denote the functions analytic in $|z| < 1$ and possessing the respective boundary functions $x(\theta)$ and $x(\theta)f(\theta)$. Let $U(\theta) = u(e^{i\theta})$ and $V(\theta) = v(e^{i\theta})$, so that $x^2(\theta)f'(\theta) = U^2(\theta)(V(\theta)/U(\theta))'$.

A heuristic argument leading to (18) is the following. Let U' and V' be defined by the formal trigonometrical series obtained by term by term differentiation of the corresponding series for U and V, and suppose that $U^2(V/U)' = UV' - U'V$ is meaningful. Since the trigonometrical series for U, V, U' and V' are of the type $\sum_{n=0}^{\infty} f_n e^{in\theta}$ then so also are the products UV' and $U'V$ as well as their difference.

A rigorous proof of (18) can be given as follows. Let the Fourier series of $U(\theta)$ and $V(\theta)$ be given by

$$U(\theta) \sim \sum_{n=0}^{\infty} a_n e^{in\theta}, \quad V(\theta) \sim \sum_{n=0}^{\infty} b_n e^{in\theta}. \tag{19}$$

Since $V(\theta) = U(\theta)f(\theta)$, where $U(\theta)$ and $f(\theta)$ each belongs to class $L^2(-\pi, \pi)$, then $\sum_{n=0}^{\infty} a_k c_{n-k} = b_n$ for $n = 0, 1, 2, \cdots$, and

$$\sum_{k=0}^{\infty} a_k c_{n-k} = 0 \text{ for } n = -1, -2, \cdots; \tag{20}$$

cf. Zygmund [19], p. 90. Note that the convergence of the series defining the b_n is assured by the Schwarz inequality. Similarly, the Fourier series of $U^2(\theta)$ is given by

$$U^2(\theta) \sim \sum_{n=0}^{\infty} A_n e^{in\theta}, \quad A_n = \sum_{k=0}^{n} a_{n-k}a_k. \tag{21}$$

Since, by (7),

(22) $$f'(\theta) \sim \sum_{n=-\infty}^{\infty} inc_n e^{in\theta} ,$$

and, since $x^2(\theta) = U^2(\theta)$ is of class $L(-\pi, \pi)$ and $f'(\theta)$ is bounded, the Fourier series of $x^2(\theta)f'(\theta)$ is given by

(23) $$x^2(\theta)f'(\theta) \sim \sum_{n=-\infty}^{\infty} B_n e^{in\theta} , \quad B_n = i \sum_{k=-\infty}^{\infty} A_{n-k} k c_k ;$$

cf. Zygmund [19], p. 90.

Since $U^2(\theta) \varepsilon L(-\pi, \pi)$ then, by the Riemann-Lebesgue lemma, $A_n \to 0$ as $n \to \infty$, and the absolute convergence of each of the series defining the B_n is assured by (7). Also the same assertion holds for the series corresponding to the above B_n but where $U(\theta)$ is replaced by the function with the Fourier series $\sum_{n=0}^{\infty} | a_n | e^{in\theta}$. Since $B_n = i\sum_{m=0}^{\infty} A_m(n - m)c_{n-m}$, this implies that each of the iterated series

(24) $$B_n = i \sum_{m=0}^{\infty} \sum_{k=0}^{\infty} a_{m-k} a_k (n - m) c_{n-m}$$

is absolutely convergent. Consequently, an interchange of the order of summation leads to

(25) $$B_n = i \sum_{k=0}^{\infty} a_k \Big[(n - k) \sum_{p=0}^{\infty} a_p c_{n-k-p} - \sum_{p=0}^{\infty} p a_p c_{n-k-p} \Big] .$$

On reversing the order of summation in the second iterated sum, it follows from (20) that $B_n = 0$ for $n = 0, -1, -2, \cdots$, so that (18) follows from (23). This completes the proof of Theorem 1.

6. Some dual results. A theorem similar to Theorem 1 but with the cosines replaced by sines is valid. In particular, whereas (a) of Theorem 1 states that (11) implies (16) while (d) states that (11) and (7) imply (9), the duals of these assertions become the following

THEOREM 2. *Let $j(\theta)$ be defined by*

(26) $$j(\theta) \sim 2 \sum_{n=1}^{\infty} c_n \sin n\theta ,$$

and suppose that (11) *holds. Then, for every constant α,*

(27) $$\text{meas } \{\theta : j(\theta) = \alpha\} = 0 .$$

If, in addition to (11), *relation* (7) *is assumed, then*

(28) $$\text{meas } \{\theta : i(\theta) \varepsilon Z\} = 0 \text{ whenever meas } Z = 0 .$$

The proof follows from the observation that the function $k(\theta) = ij(\theta)$ considered in the beginning of § 3 plays a role similar to that of $f(\theta)$.

7. Remarks. If $A(z) \varepsilon H_p$, then $B(\theta) = A(e^{i\theta})$ satisfies, for every constant α, not only

$$\text{meas } \{\theta : B(\theta) = \alpha\} = 0, \text{ unless } B(\theta) \equiv \alpha \,, \tag{29}$$

but even

$$\int_{-\pi}^{\pi} |\log | B(\theta) - \alpha || \, d\theta < \infty \,. \tag{30}$$

This result was proved by Szego [18] for $p = 2$. Its validity for arbitrary $p > 0$ was pointed out by F. Riesz ([13], pp. 91–92) to be a consequence of his factorization theorem for functions of class H_p. Thus, for every constant α, relations (16) and (27), and even

$$\int_{-\pi}^{\pi} |\log | f(\theta) - \alpha || \, d\theta < \infty \text{ and } \int_{-\pi}^{\pi} |\log | j(\theta) - \alpha || \, d\theta < \infty \,, \tag{31}$$

are seen to be necessary conditions in order that 0 be in the point spectrum of H, or, what is the same thing, in order that the translated sequences $(c_1, c_2, \cdots)$, $(c_2, c_3, \cdots)$, $\cdots$ fail to form a fundamental set for the Hilbert space l^2 of vectors $x = (x_1, x_2, \cdots)$ with $\sum_{n=1}^{\infty} | x_n |^2 < \infty$. (In connection with this latter form of (11), it is interesting to compare the present situation relating to the completeness of shifted sequences, with a similar, but different one considered in the papers of Beurling [2] and Halmos [3].) That the condition (31) is not sufficient for 0 to be in the point spectrum of H can be seen for the case $c_n = 1/n \, (n = 1, 2, \cdots)$. Then $f(\theta)$ of (1) becomes $-2 \log (2 | \sin (\theta/2) |$ and $j(\theta)$ of (26) becomes the odd function on $(-\pi, \pi)$ defined on $(0, \pi)$ by $j(\theta) = \pi - \theta$, and so (31) holds for every constant α. However, 0 is not in the point spectrum of $H = ((i + j - 1)^{-1})$; in fact, the spectrum of H is known to be purely continuous (Magnus [8]).

Since (7) holds if, say, $f''(\theta)$ is continuous, it follows from the Theorems 1 and 2 that for such functions f, in order that (11) hold, not only (16) and (27), but even the more restrictive conditions (9) and (28) must be satisfied. It is to be noted that even if, say, $f''(\theta)$ is continuous, (16) does not imply (9). In order to see this, let C denote a closed, nowhere dense (Cantor) set of *positive* measure on $[0, \pi]$, and define a function $q(\theta)$ so as to have a continuous derivative on $[0, \pi]$ and satisfy $q(\theta) = 0$ on C and $q(\theta) > 0$ on $[0, \pi] - C$.

Then $q(0) = q'(0) = 0$ and $f(\theta) = \int_0^{\theta} q(u) du$ is a strictly increasing function on $[0, \pi]$; hence, if $f(-\theta) = f(\theta)$ for $0 \leqq \theta \leqq \pi, f(\theta)$ is of the form (1), has a continuous second derivative, and satisfies (16). If T denotes the image under f of the set C, then T is measurable and meas $T = \int_c |df| = \int_c q(\theta)d\theta = 0$, so that (9) fails to hold with $T = Z$.

REFERENCES

1. N. Aronszajn and W. F. Donoghue, Jr., *On exponential representations of analytic functions in the upper half-plane with positive real part*, Journal d'Analyse Mathmatiques, **5** (1956-57), 321-388.

2. A. Beurling, *On two problems concerning linear transformations in Hilbert space*, Acta Mathematica, **81** (1949), 239-255.

3. P. R. Halmos, *Shifts on Hilbert spaces*, Journal für die reine und angewandte Mathematik, **208** (1961), 102-112.

4. P. Hartman, *On completely continuous Hankel matrices*, Proc. Amer. Math. Soc., **9** (1958), 862-866.

5. P. Hartman and A. Wintner, *On the spectra of Toeplitz's matrices*, Amer. J. Math., **72** (1950), 359-366.

6. ———, *The spectra of Toeplitz's matrices*, Amer. J. Math., **76** (1954), 867-882.

7. T. Kato, *Perturbation of continuous spectra by trace class operators*, Proc. Japan Acad., **33** (1957), 260-264.

8. W. Magnus, *On the spectrum of Hilbert's matrix*, Amer. J. Math., **72** (1950), 699-704.

9. I. P. Natanson, *Theory of functions of a real variable*, Ungar Pub. Co. (1955).

10. Z. Nehari, *On bounded bilinear forms*, Annals of Math., **65** (1957), 153-162.

11. C. R. Putnam, *Commutators and absolutely continuous operators*, Trans. Amer. Math. Soc., **87** (1958), 513-525.

12. ———, *On Toeplitz matrices, absolute continuity, and unitary equivalence*, Pacific J. Math., **9** (1959), 837-846.

13. F. Riesz, *Über die Randwerte einer analytischen Funktion*, Math. Zeitschrift, **18** (1923), 87-95.

14. ———, *Gesammelte Arbeiten*, Budapest, 1960, 537-554.

15. W. W. Rogosinski and H. S. Shapiro, *On certain extremum problems for analytic functions*, Acta Mathematica, **90** (1953), 287-318.

16. M. Rosenblum, *Perturbation of the continuous spectrum and unitary equivalence*, Pacific J. Math., **7** (1957), 997-1010.

17. ———, *The absolute continuity of Toeplitz's matrices*, Pacific Jour. Math., **10** (1960), 987-996.

18. G. Szego, *Über die Randwerte einer analytischen Funktion*, Mathematische Annalen, **84** (1921), 232-244.

19. A. Zygmund, *Trigonometrical Series*, Warsaw (1935).

PURDUE UNIVERSITY

WEAKLY COMPACT OPERATORS ON OPERATOR ALGEBRAS

SHÔICHIRÔ SAKAI

Let K be a compact space and $C(K)$ be the commutative B^*-algebra of all complex valued continuous functions on K, then Grothendieck [3] (also we can see other proofs in [2]) proved the following remarkable properties:

(I) An arbitrary bounded operator of $C(K)$ into a weakly sequentially complete Banach space is weakly compact.

(II) If T is a weakly compact operator of $C(K)$ into a Banach space, then T maps weakly fundamental sequences into strongly convergent sequences.

On the other hand, let M be a W^*-algebra and M_* be the associated space of M (namely, the dual of M_* is M (cf. [8])) then the author [7] noticed that the Banach space M_* is weakly sequentially complete. Therefore, the above Grothendieck's theorems are applicable in the theory of operator algebras.

In this note, we shall show some applications, and state some related problems.

PROPOSITION 1. *Let A be a B^*-algebra, E an abstract L-space, T be a bounded operator of A into E, then T is weakly compact.*

Proof. Let T^* be the dual of T, then T^* is a bounded operator on the dual E^* of E to the dual A^* of A; E^* is a Banach space of type $C(K)$ (cf. [5]) and the second dual A^{**} of A is a W^*-algebra (cf. [9]), so that A^* is the associated space of a W^*-algebra; hence A^* is weakly sequentially complete; therefore T^* is weakly compact, so that by the well-known theorem, T is weakly compact. This completes the proof.

Now we shall show some applications.

1. Let G be a locally compact group, $L^1(G)$ be the Banach space of all complex valued integrable functions on G with respect to a left, invariant Haar measure μ and $L^2(G)$ be the Banach space of all complex valued square integrable functions on G with respect to μ. Under the convolutions (denoted by "$*$"), $L^1(G)$ is a Banach algebra.

On the other hand, for $f \in L^1(G)$ and $g \in L^2(G)$, put $L_f g = f * g$,

Received March 18, 1963. This research was partially supported by the National Science Foundation Grant 19041.

then L_f is a bounded operator on $L^2(G)$; we shall denote the uniform norm of L_f by $\|L_f\|$.

Now, let T be an operator on $L^1(G)$. T is said to be spectrally continuous, if it satisfies $\|Th\|_1 \leqq r\|L_h\|$ for all $h \in L^1(G)$, where $\|\cdot\|_1$ is the L^1-norm and r is a fixed number.

Using the generalized Planchrel's theorem and the structure theorem of connected locally compact groups, Helgason [4] proved the following: Let G be a separable unimodular locally compact, non-compact, connected group, then a spectrally continuous operator on $L^1(G)$ commuting with all right translations is identically 0.

In his review for the Helgason's paper, Mautner [6] asked whether these restrictions on the group G can be dropped.

Now we shall show

THEOREM 1. *Let G be a locally compact, non-compact group, then a spectrally continuous operator T on $L^1(G)$ commuting with all right translations is identically zero.*

Proof. Let $R(G)$ be the uniform closure of the set $\{L_f \mid f \in L^1(G)\}$ in the B^*-algebra B of all bounded operators on $L^2(G)$, then $R(G)$ is a B^*-algebra; since T is spectrally continuous, it can be uniquely extended to a bounded operator $\widetilde{T}$ of $R(G)$ into $L^1(G)$; by Proposition 1, $\widetilde{T}$ is weakly compact; let S be the unit sphere of $R(G)$; since $\|L_f\| \leqq \|f\|_1$ for $f \in L^1(G)$, $L^1(G) \cap S$ contains the unit sphere of $L^1(G)$; therefore the set $\{Th \mid h \in L^1(G),\ \|h\|_1 \leqq 1\}$ is relatively weakly compact in $L^1(G)$; this implies that T is weakly compact as an operator on $L^1(G)$.

Since T commutes with all right translations, by the theorem of Wendel [10], there is a bounded Radon measure ν such that $Th = \nu * h$ for $h \in L^1(G)$; let f be an element of $L^1(G)$, then the mapping $h \to (f*\nu)^* * (f*\nu) * h$ on $L^1(G)$ is weakly compact, where $(f*\nu)^*(x) = \rho(x)\overline{f*\nu(x^{-1})}$, and $d\mu(x^{-1}) = \rho(x)d\mu(x)$ for $x \in G$; hence the mapping $h \to \{(f*\nu)^* * (f*\nu)\} * \{(f*\nu)^* * (f*\nu)\} * h$ is strongly compact (cf. Cor 3.7 in [2]).

Put $g = \{(f*\nu)^* * (f*\nu)\} * \{(f*\nu)^* * (f*\nu)\}$, then g belongs to $L^1(G)$. Let S_1 be the unit sphere of $L^1(G)$, then $g * S_1$ is relatively strongly compact in $L^1(G)$, so that the set $\{(g*f)^* = f^* * g^* \mid f \in S_1\}$ is also so; hence $S_1 * g^*$ is relatively strongly compact; let $\{v_\alpha\}_{\alpha \in \Pi}$ be a fundamental family of compact neighborhoods at a point s of G and let $\{f_\alpha\}_{\alpha \in \Pi}$ be a family of continuous positive functions on G such that the support of f_α is contained in v_α and $\int_G f_\alpha(x)dx = 1$, then the directed set $\{f_\alpha * g^*\}$ converges to sg^* in the L^1-norm, where $sg^*(x) = g^*(s^{-1}x)$; therefore the set $\{sg^* \mid s \in G\}$ is relatively strongly compact.

Now suppose that $\|g^*\|_1 \neq 0$, then it is enough to assume that

$\| g^* \|_1 = 1$. There is a finite set $\{s_1 g^*, s_2 g^*, \cdots, s_n g^*\}$ where $s_i \in G$ $(i = 1, 2, \cdots, n)$ such that $\inf_{1 \leqq i \leqq n} \| sg^* - s_i g^* \|_1 < 1/2$ for all $s \in G$.

On the other hand, let C be a compact subset of G such that

$$\int_{G-C} | g^*(x) | d\mu(x) < \frac{1}{10} \quad \text{and} \quad \int_{G-C} | g^*(s_i^{-1}x) | d\mu(x) < \frac{1}{10}$$

for $i = 1, 2, \cdots, n$, and s be an element of G such that $s \notin CC^{-1}$, then $s^{-1}C \cap C = (\phi)$; therefore

$$\begin{aligned} &\| sg^* - s_i g^* \|_1 \\ &= \int_C | (sg^* - s_i g)(x) | d\mu(x) + \int_{G-C} | (sg^* - s_i g^*)(x) | d\mu(x) \\ &\geqq \int_C | (s_i g^*)(x) | d\mu(x) - \int_C | (sg^*)(x) | d\mu(x) \\ &\quad + \int_{G-C} | (sg^*)(x) | d\mu(x) - \int_{G-C} | (s_i g^*)(x) | d\mu(x) \\ &\geqq \left(1 - \frac{1}{10}\right) - \int_{s^{-1}C} | g^*(x) | d\mu(x) + \left(1 - \frac{1}{10}\right) - \frac{1}{10} \\ &\geqq \left(1 - \frac{1}{10}\right) - \frac{1}{10} + \left(1 - \frac{1}{10}\right) - \frac{1}{10} = \frac{8}{5} \qquad \text{for all } i\,. \end{aligned}$$

This is a contradiction; hence $g^* = 0$, so that $g = f * \nu = 0$; since f is an arbitrary element of $L^1(G)$, $\nu = 0$, so that $T = 0$. This completes the proof.

2. At first we shall show

PROPOSITION 2. Let A be a weakly sequentially complete B^*-algebra, then A is finite dimensional.

Proof. It is enough to assume that A has unit. Let C be a maximal abelian *-subalgebra of A, then C is a Banach space of type $C(K)$ and weakly sequentially complete; by the Grothendieck's theorem, the identity mapping T on C is weakly compact, so that $T^2 = T$ is strongly compact on C (cf. Cor 3.7 in [2]); hence C is finite-dimensional. Therefore there is a finite family of mutually orthogonal projections $(e_1, e_2, \cdots, e_n)$ by which C is linearly spanned; by the maximality of C, $e_i A e_i$ $(i = 1, 2, \cdots, n)$ is one-dimensional.

For any $x, y \in A$, there is a complex number $\lambda_i(x, y)$ such that $e_i y^* x e_i = \lambda_i(x, y) e_i$; clearly $\lambda_i(x, x) \geqq 0$, and if $\lambda_i(x, x) = 0$, $xe_i = 0$; moreover $\| xe_i \| = \| e_i x^* x e_i \|^{1/2} = \lambda_i(x, x)^{1/2}$; therefore a Banach subspace Ae_i of A is a hilbert space; since $A = \sum_{i=1}^n Ae_i$, A is reflexive, so that A is a reflexive W^*-algebra; since all irreducible *-representations of

A are σ-continuous, A is of type I; since the center of A is finite-dimensional, A is a direct sum of a finite family of type I-factors $(A_1, A_2, \cdots, A_m)$; since A_j can be considered the algebra of all bounded operators on a hilbert space $\mathfrak{h}_j$ for $j = 1, 2, \cdots, m$, the reflexivity of A_j implies the finite-dimensionality of $\mathfrak{h}_j$ and so the finite-dimensionality of A_j; hence A is finite-dimensional. This completes the proof.

COROLLARY 1. *Let A be an infinite dimensional B^*-algebra and E be a Banach space of type L^p $(1 \leqq p < +\infty)$ or the associated space of a W^*-algebra, then the Banach space A is not topologically isomorphic to E.*

Since B^*-algebras are Banach spaces which have many analogous properties with $C(K)$; therefore it is very natural to ask whether the theorems of Grothendieck are positive in B^*-algebras.

We have no solution for the property (I); here we shall show that the property (II) is negative, and show an application.

A negative example. Let $B(\mathfrak{h})$ be the B^*-algebra of all bounded operators on an infinite dimensional hilbert space $\mathfrak{h}$, and e be an one-dimensional projection on $\mathfrak{h}$, then the Banach subspace $B(\mathfrak{h})e$ of $B(\mathfrak{h})$ is isometric to $\mathfrak{h}$ [cf. [7]]; therefore the mapping $x \xrightarrow{T} xe$ of $B(\mathfrak{h})$ into $B(\mathfrak{h})e$ is weakly compact; the unit sphere S of $B(\mathfrak{h})e$ is weakly compact in $B(\mathfrak{h})$; therefore if $B(\mathfrak{h})$ satisfies the property (II), $TS = S$ is strongly compact, so that $B(\mathfrak{h})e$ is finite-dimensional, a contradiction.

Concerning the property (I), we can notice that many operators satisfy the property (I).

For instance, let A be a B^*-algebra, A^* the dual of A. For x, $a \in A$ and $f \in A^*$, put $(Laf)(x) = f(ax)$ and $(Raf)(x) = f(xa)$; we can consider bounded operators $a \xrightarrow{T} Raf$, $a \xrightarrow{S} Laf$ of A into A^*, then T and S are weakly compact (cf. [8]).

Finally we shall show an application.

THEOREM 2. *Let A be a B^*-algebra having an infinite dimensional irreducible $*$-representation, and E be a Banach space of type L^p $(1 \leqq p \leqq +\infty)$ or type $C(\Omega)$, where Ω is a locally compact space and $C(\Omega)$ is the Banach space of all continuous functions vanishing at infinity, or the associated space of a W^*-algebra, then the Banach space A is not topologically isomorphic to E.*

Proof. It is enough to show that A is not topologically isomorphic to $C(\Omega)$. Suppose that A is topologically isomorphic to $C(\Omega)$, then there is an isomorphism T of A onto $C(\Omega)$. Take the second dual T^{**} of T, the T^{**} gives an isomorphism of A^{**} onto $C(\Omega)^{**}$; A^{**} is a W^*-algebra and $C(\Omega)^{**}$ is a Banach space of type $C(K)$; since a $*$-representation of A can be uniquely extended to a σ-continuous $*$-representation of A^{**} (cf. [8]), A^{**} has an infinite dimensional irreducible W^*-representation; hence there is a central projection z of A^{**} such that $A^{**}z$ is a factor of type I_∞; from the above negative example, $A^{**}z$ has not the property (II); on the other hand, since $A^{**} = A^{**}z \oplus A^{**}(1-z)$, where 1 is the unit of A^{**}, $C(\Omega)^{**} = T^{**}(A^{**}z) + T^{**}(A^{**}(1-z))$; since $T^{**}(A^{**}z)$ has the closed complement subspace in $C(\Omega)^{**}$, $T^{**}(A^{**}z)$ has the property (II), so that $A^{**}z$ has the property (II). This is a contradiction, and completes the proof.

COROLLARY 2. *Let F be the associated space of a W^*-algebra without a type I_n part ($n < +\infty$), and E be a Banach space of type L^p or $C(\Omega)$ then F is not topologically isomorphic to E.*

Proof. Suppose that F is topologically isomorphic to E, then F^* is topologically isomorphic to E^*. This is a contradiction.

REMARK. Theorem 2 and Corollary 2 imply that the above mentioned B^*-algebras or associated spaces (for instance, the B^*-algebra $\mathscr{C}$ of all compact operators on an infinite dimensional hilbert space, the B^*-algebra $B(\mathfrak{h})$ of all bounded operators on an infinite dimensional hilbert space $\mathfrak{h}$, the B^*-algebra $R(G)$ corresponding to all non-almost periodic locally compact groups, and all W^*-factor with an exception of type $I_n (n < +\infty)$ and their associated spaces) are not topologically contained in the classes of the so-called classical Banach spaces $((M), (m), (C), (c), (C^{(p)})_{p\geqq 1}, (L^p)_{1\leqq p\leqq +\infty}, (l^p)_{1\leqq p\leqq +\infty})$ mentioned by Banach (cf. [1]); therefore it is very meaningful to examine whether many unsolved problems concerning Banach spaces are positive in these examples.

REFERENCES

1. S. Banach, *Théorie des opérations linéaires*, Varsovie 1932.

2. R. Bartle, N. Dunford and J. Schwartz, *Weak compactness and vector measures*, Can. J. Math., **7** (1955), 289–305.

3. A. Grothendieck, *Sur les applications linéaries compactes d'espace du type $C(K)$*, Can. J. Math., **5** (1953), 129–173.

4. S. Helgason, *Topologies of group algebras and a theorem of Littlewood*, Trans. Amer. Math. Soc., **86** (1957), 269–283.

5. S. Kakutani, *Concrete representation of abstract (M)-Spaces*, Ann. Math., (2) **42** (1941), 994–1024.

6. I. Mautner, *Mathematical Review*, **20** (1959), 319.

7. S. Sakai, *Topological properties of W^*-algebras*, Proc. Jap. Acad., **33** (1957), 439–444.

8. S. Sakai, *The theory of W^*-algebras*, Lecture note, Yale University, 1962.
9. Z. Takeda, *Conjugate spaces of operator algebras*, Proc. Jap. Acad., **30** (1954), 90-95.
10. J. Wendel, *Left centralizers and isomorphisms of group algebras*, Pacific J. Math., **2** (1952), 251-261.

YALE UNIVERSITY AND
WASEDA UNIVERSITY

NILPOTENCY AND SPECTRAL OPERATORS

JAMES E. SIMPSON

1. Introduction. The present paper is concerned with conditions under which the quasi-nilpotent part of a spectral operator is actually nilpotent of some order k. As might be expected, the case of a spectral operator on a Hilbert space has been settled longest. (See [4].) The case of a Banach space has been treated quite thoroughly by C. A. McCarthy [7] who showed that with a certain rate of growth condition on Q, the nilpotent part of the spectral operator $T = S + Q$, satisfies $Q^{m+2} = 0$, where the m is a positive integer involved in the rate of growth condition. He also discusses more special cases in which $Q^{m+1} = 0$ and provides examples to show that these exponents are the lowest possible in each case. The question of extending these results to general locally convex spaces could not even be formulated until a theory of spectral operators in these spaces had been devised. The work of C. Ionescu Tulcea [5] having laid the foundations in this area, we may now attempt to solve the problem of generalizing McCarthy's results. It is shown below that his theorems, and indeed some part of the proofs, may be carried over to the locally convex case, with a suitable reformulation of some of the conditions and reworking of some of the supporting theory.

The basic assumptions are as follows. E denotes a locally convex linear topological space over the field, C, of complex numbers. Moreover, E is assumed to be separated, barrelled and quasi-complete. The strong dual of E is denoted by E'. The space of continuous linear mappings of E into itself is $\mathscr{L}(E, E)$, which we shall always assume to be given the topology $\mathscr{T}_b$ of uniform convergence on the bounded subsets of E. We denote the adjoint of T by ${}^tT$, for each $T \in \mathscr{L}(E, E)$. The resolvent set of T, res T, is a certain subset of $\hat{C}$, the one-point compactification of C. Specifically, $\lambda \in$ res T provided there is a neighborhood V_λ of λ in $\hat{C}$ and a function R_T with domain $V_\lambda \cap C$ and range in $\mathscr{L}(E, E)$ such that

(a) the set $\{R_T(z)x: z \in V_\lambda \cap C\}$ is a bounded subset of E for each $x \in E$, and

(b) $R_T(z)(zI - T) = (zI - T)R_T(z) = I$ for all $z \in V_\lambda \cap C$. The complement of res T, in $\hat{C}$, is the spectrum of T, denoted sp T. If $\infty \notin$ sp T, then sp T is compact in C and we say T is regular. We

Received June 6, 1963. Most of the work appearing in this paper was done while the author was a National Science Foundation Graduate Fellow and appears in somewhat modified form in his dissertation, submitted to Yale University while he was a candidate for the doctoral degree.

denote by $B^\infty(C)$ the algebra of bounded complex-valued Baire functions on C, with norm $\| f \|_u = \sup_{z \in C} |f(z)|$, and by $S(C)$ the Baire sets, i.e., those subsets, A, of C whose characteristic function, φ_A, is in $B^\infty(C)$. We denote by M^1 the set of bounded Radon measures on C, with norm $\| \mu \| = \sup \left\{ \left| \int_C f \, d\mu \right| : f \in B^\infty(C), \| f \|_u = 1 \right\}$. By sup μ we mean the support of the measure μ. Further information about locally convex spaces and Radon measures may be found in the well-known Bourbaki books [1] and [2]. Concerning the resolvent and spectrum of T, see L. Waelbroeck [10] and [11].

Turning now to spectral operators, we review some of the definitions and theorems to be found in the above mentioned work of Ionescu Tulcea [5]. See also F. Maeda [6] and H. Schaeffer [8]. The latter paper contains a monumental quantity of information about spectral measures and extensions and proofs of many of the observations listed in this paragraph. With E as above, let $\mathscr{F} = \{\mu_{x,x'} : x \in E, x' \in E'\}$ be a set of bounded Radon measures on C, indexed as indicated by $E \times E'$. We say that $\mathscr{F}$ is a spectral family of measures if there is a continuous algebraic representation of $B^\infty(C)$ in $\mathscr{L}(E, E)$, denoted by $f \to U_{\mathscr{F}, f}$ (or U_f, if no confusion will ensue) such that $U_l = I$, and $\langle U_f x, x' \rangle = \int_C f d\mu_{x,x'}$ for all $x \in E$, $x' \in E'$, and $f \in B^\infty(C)$. The function l is defined by $l(z) = 1$ for all $z \in C$. By sup $\mathscr{F}$ we mean $\overline{\bigcup_{x \in E, x' \in E'} \sup \mu_{x,x'}}$. We say that $T \in \mathscr{L}(E, E)$ commutes with a spectral family $\mathscr{F}$, provided $TU_f = U_f T$ for all $f \in B^\infty(C)$. Denote by $P_{\mathscr{F}}$ the $\mathscr{L}(E, E)$-valued function defined on $S(C)$ by $P_{\mathscr{F}}(\sigma) = U_{\varphi_\sigma}$. Then $P_{\mathscr{F}}$ has the following properties:

(i) $P_{\mathscr{F}}(\phi) = 0$.

(ii) $P_{\mathscr{F}}(\sigma \cap \delta) = P_{\mathscr{F}}(\sigma) P_{\mathscr{F}}(\delta)$ for all $\sigma, \delta \in S(C)$.

(iii) The set function m_x, defined on $S(C)$ with values in E by $m_x(\sigma) = P_{\mathscr{F}}(\sigma)x$, is countably additive for each $x \in E$.

(iv) $P_{\mathscr{F}}(C) = I$.

We shall call $P_{\mathscr{F}}$ the spectral measure associated with $\mathscr{F}$. It is quite common to write $\int_C f \, dP_{\mathscr{F}}$ for U_f. For each $\sigma \in S(C)$, let $E_\sigma = P_{\mathscr{F}}(\sigma)E$. If $T \in \mathscr{L}(E, E)$ commutes with $\mathscr{F}$, then $T_\sigma : E_\sigma \to E_\sigma$ may be defined by $T_\sigma x = Tx$ for all $x \in E_\sigma$. An element T of $\mathscr{L}(E, E)$ is said to be a spectral operator if there is a spectral family $\mathscr{F} (= \mathscr{F}_T$, if necessary) on C such that T commutes with $\mathscr{F}$ and sp $T_\sigma \subset \sigma$ for every compact subset σ of C. An element Q of $\mathscr{L}(E, E)$ is said to be quasi-nilpotent if $\text{limit}_n |\langle Q^n x, x' \rangle|^{1/n} = 0$ for all $x \in E$, $x' \in E'$. An element S of $\mathscr{L}(E, E)$ is said to be scalar relative to a spectral family $\mathscr{F}$ if the function $f : f(z) = z$ is $\mu_{x,x'}$-measurable for all $\mu_{x,x'} \in \mathscr{F}$ and $\langle Sx, x' \rangle = \int_C f d\mu_{x,x'}$ for all $x \in E$, $x' \in E'$. Finally we mention condition $P\mathscr{C}$), which is described in [5] and [6]. The central decomposition theorem of spectral theory then is that if T is a spectral operator whose associated spectral family $\mathscr{F}_T$ satisfies condition $P\mathscr{C}$), there is a unique decomposition $T = S + Q$ where S is scalar relative to $\mathscr{F}_T$ and Q is quasi-nilpotent.

One last tool will be needed below, namely an "operational calculus" for regular operators. Suppose T is regular. Let $\mathscr{G}(T)$ be the class of all complex-valued functions f, analytic on an open set $D(f)$ which contains sp T. Let D be any Cauchy domain satisfying sp $T \subset D \subset \bar{D} \subset D(f)$. Then for each $f \in \mathscr{G}(T)$, define $f(T) = 1/2\pi i \int_\Gamma f(z) R_T(z) dz$, where Γ is the boundary of D. We then have the following theorem.

THEOREM 1. *For any $f \in \mathscr{G}(T)$, $f(T)$ is a well-defined element of $\mathscr{L}(E, E)$ independent of the choice of D* (provided it satisfies the above conditions). *Moreover, if f and g are both in $\mathscr{G}(T)$, then $(f+g)(T) = f(T) + g(T)$ and $(fg)(T) = f(T)g(T)$. If Γ is a circle of sufficiently large radius to contain* sp T *in its interior, then $T = 1/2\pi i \int_\Gamma z R_T(z) dz$ and $I = 1/2\pi i \int_\Gamma R_T(z) dz$. Finally,*

$$\left\langle \int_\Gamma f(z) R_T(z) dz\, x, x' \right\rangle = \int_\Gamma \langle f(z) R_T(z) x, x' \rangle dz$$

for every $x \in E$, $x' \in E'$, and $f \in \mathscr{G}(T)$, $\Gamma \subset D(f)$.

The definition of Cauchy domain is to be found in Taylor's paper [9], which also contains a proof of a theorem very similar to the above which may easily be adjusted to fit the present situation. The theorem might also be considered a special case of some of the work of H. Cartan [3]. With this background we are prepared to discuss conditions under which $Q^k = 0$ for some positive integer k.

2. The general case. Let E be a separated, locally convex space which is barrelled and quasi-complete. Let T be a spectral operator on E whose spectral family $\mathscr{F}$ satisfies condition $P\mathscr{C}$) so that we may write $T = S + Q$. We now state our rate of growth condition:

DEFINITION. With E and T as described, we say that R_T satisfies an mth-order rate of growth condition (m being a positive integer) if the set $\{d(z, \sigma)^m R_{T_\sigma}(z) P_{\mathscr{F}}(\sigma) : z \notin \sigma, \sigma \text{ compact}\}$ is bounded in $(\mathscr{L}(E, E), \mathscr{T}_b)$. Here $d(z, \sigma)$ is the distance from z to σ.

For the rest of this section we assume that R_T satisfies an mth-order rate of growth of condition.

LEMMA 1. *Let n be a fixed positive integer, σ a compact subset of C, V a neighborhood of zero in $\mathscr{L}(E, E)$. Then there is a finite partition of σ by Borel sets $\{\sigma_j: j = 1, 2, \cdots, k\}$ with the property that if $\{\lambda_j: j = 1, 2, \cdots, k\}$ is any choice of k complex numbers with each $\lambda_j \in \sigma_j$, then*

$$Q^n P_{\mathscr{G}}(\sigma) - \sum_j (T - \lambda_j)^n P_{\mathscr{G}}(\sigma_j) \in V . \tag{1}$$

Proof. Denoting by f_i the function $f_i(z) = z^i \varphi_\sigma(z)$, for $i = 1, 2, \cdots, n$, we observe that for every $\varepsilon > 0$ there is a partition $\{\sigma_j: j = 1, 2, \cdots, k\}$ of σ such that $\| \sum_j \lambda_j^i \varphi_{\sigma_j} - f_i \|_u < \varepsilon$ for all $i = 1, 2, \cdots, n$, where the λ_j are arbitrary in σ_j. Next, since the mapping $f \to U_f$ is continuous, it follows that $S^i P_{\mathscr{G}}(\sigma) = U_{f_i}$ may be approximated in the topology of $\mathscr{L}(E, E)$ by operators of the form $\sum_j \lambda_j^i P_{\mathscr{G}}(\sigma_j)$, uniformly for $i = 1, 2, \cdots, n$, and for $\lambda_j \in \sigma_j$. The theorem may now be proved by considering

$$Q^n P_{\mathscr{G}}(\sigma) = (T - S)^n P_{\mathscr{G}}(\sigma) = \sum_{i=0}^n (-1)^i \binom{n}{i} T^{n-i} S^i P_{\mathscr{G}}(\sigma) .$$

Given V, choose U, an equilibrated neighborhood of zero in $\mathscr{L}(E, E)$, such that $\sum_{i=0}^n \binom{n}{i} U \subset V$. For each $i = 0, 1, \cdots, n$, choose a neighborhood W_i of zero in $\mathscr{L}(E, E)$ such that $A \in W_i$ implies $T^{n-i} A \in U$. Finally, choose $\{\sigma_j\}$ so that $\sum_j \lambda_j^i P_{\mathscr{G}}(\sigma_j) - S^i P_{\mathscr{G}}(\sigma) \in \cap\, i W_i$. Then,

$$\sum_i (-1)^i \binom{n}{i} \sum_j T^{n-i} \lambda_j^i P_{\mathscr{G}}(\sigma_j) - \sum_i (-1)^i \binom{n}{i} T^{n-i} S^i P_{\mathscr{G}}(\sigma) \in V .$$

But the first term in this last expression is just $\sum_j (T - \lambda_j)^n P_{\mathscr{G}}(\sigma_j)$ so that (1) is proved.

LEMMA 2. *For every bounded subset B of E, every equicontinuous subset B' of E', and every positive integer n, there is a positive real number $M = M(B, B', n)$ such that*

$$| \langle Q^n P_{\mathscr{G}}(\sigma) x, x' \rangle | \leqq M \varepsilon^{n-m+1} \tag{2}$$

for all $x \in B$, $x' \in B'$, provided $0 < \varepsilon \leqq 1$ and σ is a Borel set of diameter $\leqq \varepsilon$.

Proof. If σ is empty there is no problem. Next consider the case where σ is a nonvoid compact subset of C, and fix $\eta > 0$. By Lemma 1 there is a partition $\{\sigma_j\}$ of σ such that

$$| \langle Q^n P_{\mathscr{G}}(\sigma) x, x' \rangle - \langle \sum_j (T - \lambda_j)^n P_{\mathscr{G}}(\sigma_j) x, x' \rangle | < \eta \tag{3}$$

for all $x \in B, x' \in B'$. Using any point in σ as center, construct a circle, Γ, of radius 2ε. Then $|z-\lambda| \geqq \varepsilon$ for all $z \in \Gamma$, $\lambda \in \sigma$. Then, for all $x \in B$, $x' \in B'$,

$$\begin{aligned}
&|\langle \sum_j (T-\lambda_j)^n P_{\mathscr{F}}(\sigma_j)x, x'\rangle| \\
&= \left|\left\langle \sum_j \frac{1}{2\pi i}\int_\Gamma (z-\lambda_j)^n R_{T_\sigma}(z)dz P_{\mathscr{F}}(\sigma_j)x, x'\right\rangle\right| \\
&= \frac{1}{2\pi}\left|\int_\Gamma \langle \sum_j (z-\lambda_j)^n R_{T_\sigma}(z)P_{\mathscr{F}}(\sigma_j)x, x'\rangle\, dz\right| \\
&\leqq \frac{1}{2\pi}L(\Gamma)\sup_{z\in\Gamma} |\langle \sum_j (z-\lambda_j)^n P_{\mathscr{F}}(\sigma_j)R_{T_\sigma}(z)P_{\mathscr{F}}(\sigma)x, x'\rangle| \\
&= 2\varepsilon \sup_{z\in\Gamma}\left|\left\langle \int_\sigma \sum_j (z-\lambda_j)^n \varphi_{\sigma_j}(\lambda)dP_{\mathscr{F}}(\lambda)R_{T_\sigma}(z)P_{\mathscr{F}}(\sigma)x, x'\right\rangle\right| .
\end{aligned}$$

Let $g_z(\lambda) = \sum_j (z-\lambda_j)^n \varphi\sigma_j(\lambda)$ be the integrand function in the spectral integral in this last expression. Then the above computation implies

$$(4) \qquad |\langle \sum_j (T-\lambda_j)^n P_{\mathscr{F}}(\sigma_j)x, x'\rangle| \leqq 2\varepsilon \sup_{z\in\Gamma} |\langle U_{g_z} R_{T_\sigma}(z)P_{\mathscr{F}}(\sigma)x, x'\rangle| .$$

Since $\sup_{z\in\Gamma} \| g_z \|_u \leqq (3\varepsilon)^n$ and the set $\{U_f : \| f \|_u \leqq 1\}$ is equicontinuous, corresponding to B' we may find a neighborhood W of zero in E such that

$$(5) \qquad |\langle U_{g_z} y, x'\rangle| \leqq (3\varepsilon)^n$$

for all $y \in W$, and $x' \in B'$. It is apparent that W is independent of ε, of σ, of the choice of $\{\sigma_j\}$, and of $z \in \Gamma$. But B and W determine a neighborhood of zero in $\mathscr{L}(E, E)$. Consequently, from the rate of growth condition, there is $\alpha > 0$ such that

$$(6) \qquad d(z, \sigma)^m \alpha^{-1} R_{T_\sigma}(z)P_{\mathscr{F}}(\sigma)x \in W$$

for all $x \in B$, α being dependent only on B and B' (by way of W).

Substituting (6) in (5) gives, for all $z \in \Gamma$,

$$(7) \qquad |\langle U_{g_z} R_{T_\sigma}(z)P_{\mathscr{F}}(\sigma)x, x'\rangle| \leqq \alpha(3\varepsilon)^n d(z, \sigma)^{-m} \leqq \alpha 3^m \varepsilon^{n-m} .$$

Letting $M = 2\cdot 3^n\alpha$ and substituting (7) in (4) and (3) we have

$$\begin{aligned}
|\langle Q^n P_{\mathscr{F}}(\sigma)x, x'\rangle| &\leqq |\langle \sum_j (T-\lambda_j)^n P_{\mathscr{F}}(\sigma_j)x, x'\rangle| + \eta \\
&\leqq M\varepsilon^{n-m+1} + \eta
\end{aligned}$$

for all $x \in B$, $x' \in B'$, with M dependent only on B, B', and n. Since η is arbitrary, the theorem is proved under the additional assumption that σ be compact. However, the general case may be readily deduced from this one.

THEOREM 2. *Let σ be a Borel set in C whose Hausdorff p-mea-*

sure is zero for some p. Then, for all $k \geqq p + m - 1$, $Q^k P_{\mathscr{F}}(\sigma) = 0$.

Proof. The hypothesis asserts that for every $\varepsilon > 0$ there is a partition $\{\sigma_j\}$ of σ by finitely or countably many Borel sets of diameters ε_j respectivly such that $\varepsilon_j \leqq 1$ for all j and $\sum_j \varepsilon_j^p < \varepsilon$. Let $k = p + m - 1$. Then

$$|\langle Q^k P_{\mathscr{F}}(\sigma)x, x'\rangle| \leqq \sum_j |\langle Q^k P_{\mathscr{F}}(\sigma_j)x, x'\rangle| \leqq \sum_j M\varepsilon_i^{k+1-m} \leqq M\varepsilon ,$$

where M depends on x, x', and k. Since ε, x, and x' are arbitrary, we are done.

THEOREM 3. $Q^{m+2} = 0$.

Proof. Let $p = 3$, $\sigma = \sup \mathscr{F}$, and $k = 3 + m - 1 = m + 2$ in the previous theorem.

3. Variations on the theme. In the case where E is a Banach space, McCarthy has pointed out a number of variations on Theorem 3. The simplest of these, equally valid in our locally convex setting, flow directly from Theorem 2 when the p-measure of $\operatorname{sp} T$ or of $\sup \mathscr{F}$ is 0 for $p = 1$ or 2. An entirely different type of variation (also considered in [7]) may be discovered by observing that certain well-known Banach spaces may be embedded in E or E'. As in part 2 we assume that $T = S + Q$ is a spectral operator whose resolvent satisfies an mth-order rate of growth condition.

THEOREM 4. *For every* $x \in E$, *and* $x' \in E'$, *the measure* $\mu_{Q^{m+1}x,x'}$ *has base Lebesgue planar measure,* λ. *In fact, for every bounded set* B *in* E, *and equicontinuous set* B' *in* E', *there is* $N = N(B, B')$ *such that for all Borel sets* σ, *all* $x \in B$, *and all* $x' \in B'$, *we have*

$$\mu_{Q^{m+1}x,x'}(\sigma) \leqq N\lambda(\sigma) .$$

Proof. Actually we prove this for μ_2, Hausdorff 2-measure in the plane, but this is equivalent to proving it for λ. Let σ be a Borel set in the plane of finite μ_2 measure, and $\varepsilon > 0$. Partition σ by Borel sets $\{\sigma_i\}$ such that the diameter of σ_i is less than ε_i, with $0 < \varepsilon_i \leqq 1$, and $\sum_i \varepsilon_i^2 < \mu_2(\sigma) + \varepsilon$. Letting $n = m + 1$ in Lemma 2, we find that for each B, B' as described in the Theorem, there is $M > 0$ such that $|\langle Q^{m+1}P_{\mathscr{F}}(\sigma_i)x, x'\rangle| \leqq M\varepsilon_i^2$ for all $x \in B$, $x' \in B'$, and i. Consequently, $|\langle Q^{m+1}P_{\mathscr{F}}(\sigma)x, x'\rangle| \leqq \sum_i M\varepsilon_i^2 \leqq M(\mu_2(\sigma) + \varepsilon)$ for all $x \in B$, and $x' \in B'$. Since ε is arbitrary, we are done.

We now denote by $g_{x,x'}$ the Radon-Nikodym derivative of $\mu_{Q^{m+1}x,x'}$ with respect to λ. If f and g are two elements of $B^\infty(C)$ which agree

λ-almost everywhere, then for all $x \in E$, $x' \in E'$, we have

$$|\langle U_{f-g}Q^{m+1}x, x'\rangle| = \left|\int_C (f-g)g_{x,x'}d\lambda\right| = 0 .$$

Hence $U_f Q^{m+1} = U_g Q^{m+1}$. Moreover, if $\{f_n\}$ is a sequence of simple functions which converges to f in $L_1(C)$, then $U_{f_n}Q^{m+1}$ is a sequence which converges in $\mathscr{L}(E, E)$ to an operator which we denote by $U_f Q^{m+1}$. In this case $\langle U_f Q^{m+1}x, x'\rangle = \int_C f d\mu_{Q^{m+1}x,x'}$.

We have already seen that $Q^{m+2} = 0$. If we now assume $Q^{m+1} \neq 0$, by letting $\sigma = \sup \mathscr{F}$, we see that for some x, x', $\mu_{Q^{m+1}x,x'}(\sigma) \neq 0$. Consequently there is a set τ, compact if necessary, with nonvoid interior and a number $a > 0$, such that $\lambda(\tau) \neq 0$ and $|g_{x,x'}(z)| > a$ for all $z \in \tau$.

These two constructions are the basis for the embedding procedures mentioned above. The basic idea is to assume E has some property which is inherited by all of its closed linear subspaces. Then, if $Q^{m+1} \neq 0$, we may embed a suitable space (perhaps an L_p-space) in E or E' which does not have the property, thus obtaining a contradiction. In a written communication, McCarthy has suggested that $C(\tau)$ would be better than the $L_\infty(\tau)$ used in his paper for the case where E is assumed weakly complete. As an indication of some of the details involved in such a construction, we prove here the following:

THEOREM 5. *If E is semi-reflexive, then $Q^{m+1} = 0$.*

Proof. As indicated, we assume $Q^{m+1} \neq 0$ and construct τ and a corresponding to some x_0, x_0'. Define $\Phi: L_1(\tau) \to E$ by the formula

$$\Phi(f) = U_{f\varphi_\tau}Q^{m+1}x_0 .$$

Fix $\varepsilon > 0$ and let B' be an arbitrary equicontinuous subset of E'. Then for any $x' \in B'$,

$$\begin{aligned} |\langle \Phi(f), x'\rangle| &= |\langle U_{f\varphi_\tau}Q^{m+1}x_0, x'\rangle| \\ &= \left|\int_\tau f d\mu_{Q^{m+1}x_0,x'}\right| \leqq N \| f \|_{L_1(\tau)} , \end{aligned}$$

where N is chosen according to Theorem 4 with $B = \{x_0\}$. Thus $|\langle \Phi(f), x'\rangle| < \varepsilon$ for all $x' \in B'$ whenever $\| f \|_{L_1} < \varepsilon/N$. Since B' is arbitrary, Φ is continuous. To see that Φ is one-to-one, first define f^* for each f in $L_1^0(\tau)$ by

$$f^*(z) = \begin{cases} \overline{f(z)}/|f(z)| & \text{when } f(z) \neq 0 , \\ 0 & \text{when } f(z) = 0 . \end{cases}$$

Then f^* is in $B^\infty(C)$ so that we may define $y' = {}^t\left(\int_\tau f^*/g_{x_0,x_0'}dP_{\mathscr{G}}\right)x_0'$. Then $\langle\Phi(f), y'\rangle = \int_\tau ff^*d\lambda = \int_\tau |f|\,d\lambda = \|f\|_{L_1(\tau)}$. Thus the kernel of Φ is zero. To show that Φ^{-1} is continuous on the range of Φ it is sufficient to find, for each $\varepsilon > 0$, an equicontinuous set B' in E' such that $\|f\|_{L_1} < \varepsilon$ whenever $|\langle\Phi(f), x'\rangle| \leqq \varepsilon$ for all $x' \in B'$. The set

$$B = \left\{\int_\tau f^*/g_{x_0,x_0'}dP_{\mathscr{G}} : f \in L_1^0(\tau)\right\}$$

is an equicontinuous subset of $\mathscr{L}(E, E)$. Consequently the set $B' = \{{}^tAx_0' : A \in B\}$ is equicontinuous in E'. Finally,

$$\begin{aligned}\|f\|_{L_1(\tau)} &= \left|\int_\tau |f|/g_{x_0,x_0'}d\mu_{Q^{m+1}x_0,x_0'}\right| \\ &= \left|\left\langle U_{f\varphi_\tau}Q^{m+1}x_0, {}^t\left(\int_\tau f^*/g_{x_0,x_0'}dP_{\mathscr{G}}\right)x_0'\right\rangle\right| \\ &\leqq \sup_{A\in B}|\langle\Phi(f), {}^tAx_0\rangle| = \sup_{x'\in B'}|\langle\Phi(f), x'\rangle|\,.\end{aligned}$$

Consequently, $\Phi(L_1(\tau))$ is a closed linear subspace of E which is isomorphic in both algebraic and topological sense to $L_1(\tau)$. This contradicts the assumption that E is semi-reflexive, hence the Theorem is proved.

Bibliography

1. N. Bourbaki, *Espaces vectoriels topologiques, Elements de Mathematique*, Paris, (1953–55).
2. ———, *Intégration. Elements de Mathematique*, Paris (1952–60).
3. H. Cartan, *Theorie spectrale des C-algébres commutatives*, Seminaire Bourbaki, Exposé **125** (1956).
4. N. Dunford, *Spectral operators*, Pacific J. Math., **4** (1954), 321–354.
5. C. Ionescu Tulcea, *Spectral operators on locally convex spaces*, Bull. Amer. Math. Soc., **67** (1961), 125–128.
6. F. Maeda, *A characterization of spectral operators on locally convex spaces*, Math. Annalen, **143** (1961), 59–74.
7. C. A. McCarthy, *The nilpotent part of a spectral operator*, Pacific J. Math., **9** (1959), 1223–1231.
8. H. Schaeffer, *Spectral measures in locally convex algebras*, Acta Math., **107** (1962), 125–173.
9. A. E. Taylor, *Spectral theory of closed distributive operators*, Acta Math., **84** (1950), 189–224.
10. L. Waelbroeck, *Le calcul symbolique dans les algébres commutatives*, J. Math. Pures Appl., 9e Série, **33** (1954), 147–186.
11. ———, *Locally convex algebras: Spectral Theory*, Seminar on complex analysis, Institute for Advanced Study, 1958, unpub.

MARQUETTE UNIVERSITY

ON THE ELEMENTARY RENEWAL THEOREM FOR NON-IDENTICALLY DISTRIBUTED VARIABLES

WALTER L. SMITH

1. Introduction. Let $\{X_n\}$ be a sequence of independent, identically distributed random variables with $0 < EX_n < \infty$; write $S_n = X_1 + X_2 + \cdots + X_n$; let N_x be the number of partial sums $S_n \leq x$; write $H(x) = EN_x$. The Elementary Renewal Theorem states that under certain conditions $H(x)/x \rightarrow \{EX_n\}^{-1}$ as $x \rightarrow \infty$.

Kawata (1956) has proved a result which, as we shall see below, is equivalent to a generalization of the Elementary Renewal Theorem to the case in which the $\{X_n\}$ are non-identically distributed. Unfortunately, he found it necessary to impose quite heavy restrictions upon the distribution functions involved. In this note we shall also be concerned with the proof of the Elementary Renewal Theorem for non-identically distributed random variables, but under substantially weaker conditions than Kawata's. This renewal theorem, essentially, provides an asymptotic estimate to the sum $\sum_{n=1}^{\infty} P\{S_n \leqq x\}$; actually, we shall discuss in this paper the asymptotic behavior of more general sums $\sum_{n=1}^{\infty} a_n P\{S_n \leqq x\}$, for certain general classes of positive coefficient-sequences $\{a_n\}$. Such more general sums have also been considered by Hatori (1959), (1960), who followed Kawata's general line of attack, however, and was consequently led to assume unduly restrictive conditions.

It is well if we point out that there is another line of inquiry which could be pursued in the present context, one with which the present investigation must not be confused. Instead of considering N_x, one could define a random variable M_x as the least m for which $S_m > x$, and then study the asymptotic behavior of EM_x/x. The latter problem (also for non-identically distributed $\{X_n\}$) has been tackled in recent work announced by Robbins and Chow (1962)*. However, as might be expected, the problem we consider and the problem considered by Robbins and Chow differ in important respects, in general. Indeed, a reference to Theorem A, which we quote below, will show that one can construct a sequence of independent and *identically distributed* random variables with a finite first moment, for which EM_x is finite but EN_x is infinite. Evidently conditions which are

Received June 7, 1962, and in revised form March 7, 1963.

This research was supported by the Office of Naval Research under contract No. Ncnr-855 (09) for research in probability and statistics at the University of North Carolina, Chapel Hill, N. C. Reproduction in whole or in part is permitted for any purpose of the United States Government.

* *Footnote added in proof*:—The details of this work have now appeared in Ann. Math. Statist., **34** (1963), 390–395.

adequate for a study of M_x may prove inadequate for a similar study of N_x. However, when all the $\{X_n\}$ are nonnegative we have an exceptional case; for then $M_x = N_x + 1$, and the distinction between the two lines of inquiry disappears. Our main result, Theorem 1 announced below, is more general than the one announced by Robbins and Chow, for the case of nonnegative random variables.

Let us write $F_n(x) \equiv P\{X_n \leqq x\}$ and $G_n(x) \equiv P\{S_n \leqq x\}$; we shall also need the unit function $U(x) \equiv P\{0 \leqq x\}$.

The function $L(x)$, defined for all sufficiently large x, is said to be a function of slow growth if, for every $c > 0$

$$\frac{L(cx)}{L(x)} \to 1\,, \qquad \text{as } x \to \infty\,. \tag{1.1}$$

It follows from the work of Karamata (1930), that a nonnegative function of slow growth can always be represented thus:

$$L(x) = \frac{a(x)}{x} \exp\left\{\int_1^x \frac{a(u)}{u}\, du\right\}, \tag{1.2}$$

where $a(x)$ is a function which tends to unity as x tends to infinity. An easy consequence of this representation (1.2) is that the convergence (1.1) takes place uniformly with respect to c in any interval not containing the origin.

As a final preliminary we must say a word about the non-negative coefficient-sequences $\{a_n\}$ which we consider. For such a sequence we shall suppose there exist numbers $\alpha > 0$, $\gamma \geqq 0$, and some nonnegative function of slow growth $L(x)$, such that

$$\sum_{n=1}^{\infty} a_n x^n \sim \frac{\alpha}{(1-x)^{\gamma}} L\left(\frac{1}{1-x}\right), \qquad \text{as } x \to 1 - 0\,. \tag{1.3}$$

By an appeal to a Tauberian theorem due to Karamata (Hardy, 1949, p. 166) it is possible to deduce from (1.3) that

$$\sum_{n=1}^{N} a_n \sim \frac{\alpha N^{\gamma} L(N)}{\Gamma(1+\gamma)}\,, \qquad \text{as } N \to \infty\,, \tag{1.4}$$

although we shall omit details of this deduction. Conversely, if one starts from (1.4) then an appropriate Abelian theorem will show that (1.3) follows. Thus (1.3) and (1.4) are equivalent assumptions on the nonnegative $\{a_n\}$. We also note, as an easy deduction from (1.4), that

$$a_n = o(n^{\gamma} L(n))\,, \qquad \text{as } n \to \infty\,. \tag{1.5}$$

In connection with these sequences $\{a_n\}$ we need to define an index:

DEFINITION. *An index k of the sequence $\{a_n\}$ is any number k such that $a_n = O(n^k)$.*

If we write k^* for the greatest lower bound of the indexes of the sequence $\{a_n\}$ then k^* may or may not be an index itself. From (1.5) it is clear that $k^* \leqq \gamma$. On the other hand, we can infer from (1.4) that $k^* \geqq \gamma - 1$ and that $\gamma - 1$ can only be an index if $L(n)$ is a bounded function. We have, therefore,

LEMMA 1. *If k^* is the greatest lower bound of the indexes of $\{a_n\}$ then $\gamma - 1 \leqq k^* \leqq \gamma$; the number $\gamma - 1$ cannot be an index unless $L(n)$ is a bounded function.*

The main result of this paper can now be stated.

THEOREM 1. *Suppose the following conditions hold.*

(T1) *$\{X_n\}$ is a sequence of independent random variables with distribution functions $\{F_n(x)\}$ and finite expectations $\mu_n = EX_n$, such that*

$$\frac{\mu_1 + \mu_2 + \cdots + \mu_n}{n} \to \mu , \qquad \text{as } n \to \infty , \tag{1.6}$$

where μ finite and strictly positive.

(T2) *For every $\varepsilon > 0$*

$$\int_{n\varepsilon}^{\infty} \frac{1}{n} \sum_{r=1}^{n} \{1 - F_r(x)\}dx \to 0 , \qquad \text{as } n \to \infty . \tag{1.7}$$

(T3) *For some $\alpha > 0$, $\gamma \geqq 0$, and some nonnegative function of slow growth $L(x)$, the sequence of nonnegative constants $\{a_n\}$ satisfies either of the equivalent asymptotic relations* (1.3) *or* (1.4).

(T4) *$\sum_{n=1}^{\infty} a_n$ diverges.*

Then, if we write $G_n(x) = P\{X_1 + X_2 + \cdots + X_n \leqq x\}$, in order that

$$\sum_{n=1}^{\infty} a_n G_n(x) \sim \frac{\alpha L(x)}{\Gamma(1+\gamma)} \left(\frac{x}{\mu}\right)^{\gamma} , \qquad \text{as } x \to \infty , \tag{1.8}$$

it is sufficient that one of the following two sets of conditions, (T5) *or* (T6), *hold.*

(T5) *The X_n are nonnegative, in which case it will be proved that there necessarily exists some unbounded non-decreasing function $l(n)$ such that*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{r=1}^{n} \int_{0}^{r/l(r)} \{1 - F_r(x)\}dx = \mu . \tag{1.9}$$

(T6) (a) *If k is an index of $\{a_n\}$, then there is a distribution function $K(x)$ of a negative-valued random variable with a finite moment of order $(k+2)$, such that $K(x) \geqq F_n(x)$ for all n and all x*; (b) *If $-\kappa$ is the first moment of $K(x)$, then for some $v > \kappa$ and every $\varepsilon > 0$,*

$$\liminf_{n\to\infty} \int_{-\infty}^{\varepsilon n/\log n} \frac{1}{n} \sum_{r=1}^{n} \{U(x) - F_r(x)\} dx > 2\sqrt{(k+1)v\varepsilon}\,. \tag{1.10}$$

In condition (T4) above we have required that $\sum_{n=1}^{\infty} a_n$ shall diverge; it will be appreciated that this assumption is made only to avoid triviality. A consequence of the divergence of $\sum_{n=1}^{\infty} a_n$ is that the index $k \geqq -1$; therefore the distribution function $K(x)$ which appears in condition (T6)(a) will always have a finite mean; this justifies the introduction of $-\kappa$ in condition (T6)(b).

The special case of Theorem 1 in which the $\{X_n\}$ are nonnegative can be given the following form.

THEOREM 2. *Suppose that* (i) $\{X_n\}$ *is a sequence of nonnegative random variables such that* (T1) *and* (T2) *of Theorem* 1 *hold*; (ii) *$A(n)$ is a non-decreasing function of n for which constants $\alpha > 0$, $\gamma \geqq 0$, and a nonnegative function of slow growth $L(n)$, can be found such that*

$$A(n) \sim \alpha n^{\gamma} L(n)\,, \qquad \text{as } n \to \infty\,. \tag{1.11}$$

Then it follows that

$$EA(N_x) \sim \alpha\left(\frac{x}{\mu}\right)^{\gamma} L(x)\,, \qquad x \to \infty\,. \tag{1.12}$$

We note that by letting $A(n) = n$ in Theorem 2 we obtain a version of the Elementary Renewal Theorem for independent, non-identically distributed, nonnegative random variables. Alternatively, by taking $a_n = 1$ for all n in Theorem 1, we obtain the following version for the case when the random variables may assume negative values.

THEOREM 3. *If conditions* (T1) *and* (T2) *of Theorem* 1 *hold; and if both parts of* (T6) *hold for $k = 0$, then*

$$\frac{H(x)}{x} \to \frac{1}{\mu}\,, \qquad \text{as } x \to \infty\,, \tag{1.13}$$

where $H(x) = \sum_{n=1}^{\infty} G_n(x)$ is the expected number of partial sums $S_n \leqq x$.

From (1.13) we can infer that, for any fixed $h >$

$$\frac{1}{t}\int_t^{t+h} H(x)dx \to \frac{h}{\mu}, \qquad \text{as } t \to \infty \tag{1.14}$$

Therefore

$$\frac{1}{t}\int_0^t \{H(x+h) - H(x)\}dx \to \frac{h}{\mu}, \qquad \text{as } t \to \infty$$

or, in other words,

$$\lim_{t\to\infty} \frac{1}{t}\int_0^t \sum_{n=1}^{\infty} P\{x < S_n \leqq x + h\}dx = \frac{h}{\mu}. \tag{1.15}$$

This last limit (1.15) is the form taken by Kawata's result (1956); we see that it is implied by the simpler statement (1.13). On the other hand it is not difficult to deduce (1.13) from (1.15), so that (1.15) seems an unduly complicated form for the result. For (1.15) is equivalent to (1.14); and from (1.14) and the monotone character of $H(x)$ we can infer that lim sup $H(t)/t \leqq \mu^{-1}$ and lim inf $H(t+h)/t \geqq \mu^{-1}$.

We close this introduction with some remarks about the conditions of Theorem 1. The easiest proofs of the Elementary Renewal Theorem for the case of *identically* distributed random variables make use of the weak law of large numbers to show that $S_n/n \to \mu$ in probability, as $n \to \infty$. The present investigation will also depend on establishing such weak convergence of S_n/n, and conditions (T1) and (T2), aided by (T6)(a) when the random variables can take negative values, are concerned with this task.

To understand the *raison d'être* of condition (T6)(b) it is necessary to inquire a little into our mode of proof. We shall, as just noted, begin by establishing that $S_n/n \to \mu$ in probability. If only S_n would not fluctuate too violently about its expected value $n\mu$ our theorems would then be an easy consequence of this weak law of large numbers. Unfortunately, considerable deviation of S_n from $n\mu$ is possible; the main obstacle we have to overcome is presented by sequences $\{S_n\}$ which tend to decrease steadily over long stretches of n and then indulge in a rare, but very large, increase in value. This kind of awkward behavior is exemplified by sequences $\{F_n(x)\}$ which assign nearly all the probability to the negative values of x and reserve only a very small probability for positive values of x, necessarily located at very high positive values in order to make the expectations come right. Condition (T6)(b) is concerned with controlling this kind of awkwardness.

Condition (T6)(a), which is unnecessary when the random variables are nonnegative, is introduced to ensure the finiteness of the quantities with which we deal; it will be understood better in relation to the following theorem, which we shall use later in this paper but prove elsewhere (Smith (1964)).

THEOREM A.* *If* $\{X_n\}$ *is a sequence of independent and identically distributed random variables with* $0 < EX_n < \infty$, *and if* $k \geqq 1$, *then a necessary and sufficient condition for the convergence of the series*

$$(1.16) \qquad \sum_{n=1}^{\infty} n^k P\{X_1 + X_2 + \cdots + X_n \leqq x\}, \ -\infty < x < +\infty \ ,$$

is that $E\{|\min(0, X_n)|^{k+2}\} < \infty$. *Furthermore, when this condition is met, if* X_0 *is any other random variable, independent of the* $\{X_n\}$, *such that* $E\{|\min(0, X_0)|^{k+2}\} < \infty$, *then*

$$(1.17) \qquad \sum_{n=1}^{\infty} n^k P\{X_0 + X_1 + \cdots + X_n \leqq x\}, \ -\infty < x < +\infty \ ,$$

is also convergent.

Thus we see that (T6)(a) must be satisfied when the $\{X_n\}$ are identically distributed, and it therefore seems reasonable to require the satisfaction of some condition like (T6)(a) even in the general case, if we are to ensure the finiteness of the left-hand side of (1.8).

2. Some preliminary lemmas. We begin by showing that the conditions of Theorem 1 are sufficient to ensure that $S_n/n \to \mu$ in probability as $n \to \infty$. This could be done by appeal to classical results; however, it is not difficult to proceed from first principles, and our proof conveniently introduces an argument of a sort which we shall use several times in the course of this paper.

LEMMA 2. *If conditions* (T1) *and* (T2) *hold, then a sufficient condition for ensuring that* $S_n/n \to \mu$ *in probability as* $n \to \infty$ *is either*: (a) *the random variables* $\{X_n\}$ *are nonnegative*; *or* (b) *condition* (T6)(a) *holds with* $k = -1$.

Proof. Consider first the case of a sequence of independent, nonnegative, random variables $\{X_n\}$ whose distribution functions satisfy (T2) and whose mean values satisfy the condition that $\mu_1 + \mu_2 + \cdots + \mu_n = O(n)$, a condition less restrictive than (1.6). Write $m_n = n^{-1}(\mu_1 + \mu_2 + \cdots + \mu_n)$

* Footnote: This theorem is also a fairly easy deduction from the recently published results of M. L. Katz (1963).

and let η be a small strictly positive number. Then, in virtue of a familiar inequality, for every fixed $t \geqq 0$ we have

$$P\{S_n \leqq n(m_n - \eta)\} \leqq e^{n(m_n-\eta)t} E\{e^{-tS_n}\} .$$

Furthermore, if we make use of the Laplace-Stieltjes transforms

$$\Phi_j(t) = \int_{0-}^{\infty} e^{-tx} dF_j(x) ,$$

we may rewrite this last inequality thus:

$$P\{S_n \leqq n(m_n - \eta)\} \leqq e^{n(m_n-\eta)t} \prod_{j=1}^{n} \Phi_j(t) .$$

But, as may easily be verified,

$$\Phi_j(t) \leqq e^{\Phi_j(t)-1} ;$$

and so we have that

$$P\{S_n \leqq n(m_n - \eta)\} \leqq e^{W_n(t)} , \tag{2.1}$$

where, after some integrations by parts, we see that

$$W_n(t) = n(m_n - \eta)t - t \sum_{j=1}^{n} \int_0^{\infty} e^{-tx}\{1 - F_j(x)\}dx . \tag{2.2}$$

Choose a small $\varepsilon > 0$ and set

$$\delta_n(\varepsilon) = \int_{n\varepsilon}^{\infty} \frac{1}{n} \sum_{j=1}^{n} \{1 - F_j(x)\}dx ;$$

by (T2), $\delta_n(\varepsilon) \to 0$ as $n \to \infty$. From (2.2) we then deduce that

$$W_n(t) \leqq n(m_n - \eta)t - nte^{-nt\varepsilon}[\delta_n(0) - \delta_n(\varepsilon)] . \tag{2.3}$$

If we observe that $\delta_n(0) = m_n$ and put $t = (n\sqrt{\varepsilon})^{-1}$ in (2.3), we find that

$$\begin{aligned} W_n\left(\frac{1}{n\sqrt{\varepsilon}}\right) &\leqq \frac{m_n - \eta}{\sqrt{\varepsilon}} - \frac{\exp(-\sqrt{\varepsilon})}{\sqrt{\varepsilon}}[m_n - \delta_n(\varepsilon)] \\ &= \frac{1}{\sqrt{\varepsilon}}\{m_n[1 - \exp(-\sqrt{\varepsilon})] + \exp(-\sqrt{\varepsilon})\delta_n(\varepsilon) - \eta\} . \end{aligned} \tag{2.4}$$

Recall that m_n is positive and bounded; thus $m_n[1 - \exp(-\sqrt{\varepsilon})]$ can be made arbitrarily small for all n, by choosing ε sufficiently small. Hence the expression in braces in the last inequality can be made $< -\eta/2$, for all sufficiently large n, by choice of ε. Therefore, from (2.1), we see that

$$P\{S_n \leqq n(m_n - \eta)\} \leqq \exp\left\{-\frac{\eta}{2\sqrt{\varepsilon}}\right\},$$

for all sufficiently large n. Since ε can be chosen arbitrarily small we are led to the conclusion that

$$P\{S_n \leqq n(m_n - \eta)\} \to 0 \quad \text{as } n \to \infty \tag{2.5}$$

for every $\eta > 0$.

Let $\rho > 0$ be arbitrarily small. Since $ES_n = nm_n$ we have that

$$\begin{aligned} nm_n \geqq n(m_n + \rho)P\{S_n \geqq n(m_n + \rho)\} \\ + n(m_n - \eta)P\{n(m_n - \eta) < S_n < n(m_n + \rho)\}, \end{aligned}$$

and from this inequality it follows that

$$(\rho + \eta)P\{S_n \geqq n(m_n + \rho)\} < \eta + m_n P\{S_n \leqq n(m_n - \eta)\}. \tag{2.6}$$

If we choose η arbitrarily small, and observe once again that m_n is bounded, we can infer from (2.5) and (2.6) that

$$P\{S_n \geqq n(m_n + \rho)\} \to 0 \quad \text{as } n \to \infty \tag{2.7}$$

for every $\rho > 0$. The coupling of (2.5) and (2.7) produce the desired conclusion that for every $\eta > 0$

$$P\left\{\left|\frac{S_n}{n} - m_n\right| \geqq \eta\right\} \to 0 \quad \text{as } n \to \infty. \tag{2.8}$$

We remark that (2.8) proves the lemma for the case when the $\{X_n\}$ are nonnegative (see the Corollary 2 quoted on page 141 of Gnedenko and Kolmogorov (1954)). We turn now to the general case, and begin by defining

$$\begin{aligned} X_n^+ &= X_n, \quad &&\text{if } X_n \geqq 0, \\ &= 0, &&\text{otherwise}, \end{aligned}$$

and $X_n^- = X_n^+ - X_n$. Thus both X_n^+ and X_n^- are nonnegative random variables and we shall write $\mu_n^+ = EX_n^+$, $\mu_n^- = EX_n^-$, $m_n^+ = n^{-1}(\mu_1^+ + \mu_2^+ + \cdots + \mu_n^+)$, $m_n^- = n^{-1}(\mu_1^- + \mu_2^- + \cdots + \mu_n^-)$, and, of course, $m_n = n^{-1}(\mu_1 + \mu_2 + \cdots + \mu_n)$.

Since (T6)(a) holds for $k = -1$, the mean of $K(x)$ is finite; if we call this mean $-\kappa$ then it also follows from (T6)(a) that $\mu_n^- \leqq \kappa$ for all n, and hence $m_n^- \leqq \kappa$. Moreover, the fact that $K(x) \geqq F_n(x)$, for all n and all x, ensures that (1.7) will hold when the $\{F_n(x)\}$ in that condition are replaced by the corresponding distribution functions of the variables $\{X_n^-\}$. We can now infer from the result established for nonnegative random variables at the start of this proof that, if we write $S_n^- = X_1^- + X_2^- + \cdots + X_n^-$,

$$P\left\{\left|\frac{S_n^-}{n} - m_n^-\right| \geqq \eta\right\} \to 0 \qquad \text{as } n \to \infty , \tag{2.9}$$

for every $\eta > 0$.

Let us turn now to a consideration of $S_n^+ = X_1^+ + X_2^+ + \cdots + X_n^+$. We note first that, since $\mu_n^+ - \mu_n^- = \mu_n$, we have $m_n^+ = m_n^- + m_n$. But we have just shown that m_n^- is bounded by κ; and we may now suppose $m_n \to \mu$ as $n \to \infty$, by (1.6). Thus m_n^+ is a bounded function of the integer n, and we can appeal to our preliminary result to deduce that

$$P\left\{\left|\frac{S_n^+}{n} - m_n^+\right| \geqq \eta\right\} \to 0 \qquad n \to \infty , \tag{2.10}$$

for every $\eta > 0$. The lemma follows from (2.9) and (2.10).

LEMMA 3. *Under the conditions of Lemma* 2,

$$\int_\mu^\infty \{1 - G_n(nx)\}dx \to 0 \qquad as\ x \to \infty . \tag{2.11}$$

Proof. It is easy to verify that

$$\begin{aligned} m_n &= \int_{-\infty}^{+\infty} \{U(x) - G_n(nx)\}dx , \\ &= \int_\mu^\infty \{1 - G_n(nx)\}dx + \int_0^\mu \{1 - G_n(nx)\}dx \\ &\quad - \int_{-\kappa}^0 G_n(nx)dx - \int_{-\infty}^{-\kappa} G_n(nx)dx , \\ &= A_n + B_n - C_n - D_n , \qquad \text{say} . \end{aligned} \tag{2.12}$$

By Lemma 2, $G_n(nx) \to 0$ as $n \to \infty$, for all $x < \mu$. Thus $B_n \to \mu$ and $C_n \to 0$ as $n \to \infty$, by bounded convergence. But $m_n \to \mu$, by (1.6). Thus we can see from (2.12) that in order to establish the required result, $A_n \to 0$, we need only prove that $D_n \to 0$. In the case when the $\{X_n\}$ are nonnegative there is, of course, no need for further argument.

Write $K_n(x)$ for the familiar n-fold Stieltjes convolution of $K(x)$ with itself. Then plainly, since $K(x) \geqq F_n(x)$ for all n and all x, $K_n(x) \geqq G_n(x)$ for all n and all x. Thus it will be enough if we can prove that

$$\int_{-\infty}^{-\kappa} K_n(nx)dx \to 0 \qquad \text{as } n \to \infty . \tag{2.13}$$

However,

$$\int_{-\infty}^{0} K_n(nx)dx = \kappa\,, \qquad \text{for all } n\,,$$

so that (2.13) would follow if we proved that

$$\int_{-\kappa}^{0} K_n(nx)dx \to \kappa \qquad \text{as } n \to \infty\,. \tag{2.14}$$

To prove (2.14) we need only remark that, by the weak law of large numbers for identically distributed random variables, $K_n(nx) \to 1$ as $n \to \infty$ for all $x > -\kappa$; thus (2.14) follows from the theorem on bounded convergence. This proves the lemma.

LEMMA 4. *If the nonnegative constants* $\{a_n\}$ *satisfy* (1.3) *then, as* $s \to 0+$,

$$\sum_{n=1}^{\infty} a_n e^{-\mu s n} \sim \frac{\alpha}{\mu^\gamma s^\gamma} L\left(\frac{1}{s}\right).$$

Proof. Plainly, $e^{-\mu s} \to 1 - 0$ as $s \to 0+$. Therefore, as $s \to 0+$,

$$\sum_{n=1}^{\infty} a_n e^{-\mu s n} \sim \frac{\alpha}{(1-e^{-\mu s})^\gamma} L\left(\frac{1}{1-e^{-\mu s}}\right)$$
$$\sim \frac{\alpha}{\mu^\gamma s^\gamma} L\left(\frac{1}{1-e^{-\mu s}}\right).$$

But, as we have remarked in our introduction, $L(rx)/L(x) \to 1$ as $x \to \infty$ uniformly for r in any interval not containing 0. Thus it transpires that

$$L\left(\frac{1}{1-e^{-\mu s}}\right) \sim L\left(\frac{1}{s}\right)$$

as $s \to 0+$, since $1 - e^{-\mu s} \sim \mu s$ for small s. Thus the lemma is proved.

LEMMA 5. *Under the same conditions as Lemma* 4, *as* $s \to 0+$,

$$\sum_{n=1}^{\infty} n a_n e^{-\mu s n} \sim \frac{\gamma \alpha}{(\mu s)^{\gamma+1}} L\left(\frac{1}{s}\right).$$

Proof. Choose η, $0 < \eta < 1$. Then

$$\mu n e^{-\mu n s} < \frac{e^{-\mu s n \eta} - e^{-\mu s n}}{(1-\eta)s}\,.$$

Thus

$$\mu \cdot \sum_{n=1}^{\infty} n a_n e^{-\mu n s} < \frac{\sum_{n=1}^{\infty} a_n e^{-\mu s n \eta} - \sum_{n=1}^{\infty} a_n e^{-\mu s n}}{(1-\eta)s}$$

and so

$$\frac{\mu s^{\gamma+1}}{L\left(\frac{1}{s}\right)} \sum_{n=1}^{\infty} n a_n e^{-\mu s n} < \frac{s^{\gamma}}{(1-\eta) L\left(\frac{1}{s}\right)} \left\{ \sum_{n=1}^{\infty} a_n e^{-\mu s n \eta} - \sum_{n=1}^{\infty} a_n e^{-\mu s n} \right\} .$$

It follows therefore, from Lemma 4, that

$$\limsup_{s \to 0+} \frac{\mu s^{\gamma+1}}{L\left(\frac{1}{s}\right)} \sum_{n=1}^{\infty} n a_n e^{-\mu s n} \leqq \frac{\alpha}{\mu^{\gamma}} \left\{ \frac{\eta^{-\gamma} - 1}{1 - \eta} \right\} . \tag{2.15}$$

If we let $\eta \to 1 - 0$ in (2.15) we obtain

$$\limsup_{s \to 0+} \frac{\mu s^{\gamma+1}}{L\left(\frac{1}{s}\right)} \sum_{n=1}^{\infty} n a_n e^{-\mu s n} \leqq \frac{\alpha \gamma}{\mu^{\gamma}} . \tag{2.16}$$

Similarly, by taking $\eta > 1$ and using the fact that

$$\mu n e^{-\mu n s} > \frac{e^{-\mu s n} - e^{-\mu s n \eta}}{\eta - 1}$$

we can show

$$\liminf_{s \to 0+} \frac{\mu s^{\gamma+1}}{L\left(\frac{1}{s}\right)} \sum_{n=1}^{\infty} n a_n e^{-\mu s n} \geqq \frac{\alpha \gamma}{\mu^{\gamma}} . \tag{2.17}$$

The lemma follows from (2.16) and (2.17).

3. **Proof of Theorem 1.** We shall write β for an upper bound to the numbers $\{a_n/n^k\}$, where k is the index of the nonnegative coefficient sequence $\{a_n\}$; we shall also write $\eta > 0$ for an arbitrary small number; it is supposed that $\eta < \mu$.

Consider, to begin with,

$$\begin{aligned} K_n &= \int_{n\eta}^{n\mu} e^{-sx} G_n(x) dx , \\ &= n \int_{\eta}^{\mu} e^{-nsx} G_n(nx) dx . \end{aligned} \tag{3.1}$$

Evidently,

$$0 \leqq K_n \leqq n e^{-n\eta s} \int_{\eta}^{\mu} G_n(nx) dx .$$

But $G_n(nx) \to 0$ as $n \to \infty$, for all $x < \mu$, by Lemma 2. Hence we can appeal to bounded convergence and write

$$K_n = ne^{-n\eta s}\delta'_n \,, \tag{3.2}$$

where $\delta'_n \to 0$ as $n \to \infty$, uniformly in $s \geqq 0$.

Next consider

$$\begin{aligned} L_n &= \int_{n\mu}^{\infty} e^{-sx}\{1 - G_n(x)\}dx \\ &= n\int_{\mu}^{\infty} e^{-nsx}\{1 - G_n(nx\}dx \,. \end{aligned} \tag{3.3}$$

In view of Lemma 3 and the assumption that $\eta < \mu$ we may thus conclude that

$$L_n = ne^{-n\eta s}\delta''_n \,, \tag{3.4}$$

where $\delta''_n \to 0$ as $n \to \infty$, uniformly in $s \geqq 0$.

Thus, if we write $\delta_n = \delta'_n - \delta''_n$,

$$\sum_{n=1}^{\infty} a_n(L_n - K_n) = \sum_{n=1}^{\infty} na_n\delta_n e^{-n\eta s} \,. \tag{3.5}$$

Given an arbitrary $\varepsilon > 0$, we can find $n_0(\varepsilon)$ such that $|\delta_n| < \varepsilon$ for all $n > n_0$. Moreover we can assume that $s^{-\gamma}L(s^{-1}) \to \infty$ as $s \to 0+$, since we suppose $\sum a_n$ to be divergent. Thus

$$\left|\sum_{n=1}^{\infty} na_n\delta_n e^{-n\eta s}\right| < \sum_{n=1}^{n_0} na_n \left|\delta_n\right| e^{-n\eta s} + \varepsilon \sum_{n=1}^{\infty} na_n e^{-n\eta s} \,. \tag{3.6}$$

Therefore, by Lemma 5,

$$\limsup_{s\to 0+} \frac{(\mu s)^{\gamma+1}}{L\left(\frac{1}{s}\right)} \left|\sum_{n=1}^{\infty} na_n\delta_n e^{-n\eta s}\right| \leqq \varepsilon\gamma\alpha \,. \tag{3.7}$$

But ε is arbitrary, and we can therefore deduce from (3.7) and (3.5) that, as $s \to 0+$,

$$\frac{(\mu s)^{\gamma+1}}{L\left(\frac{1}{s}\right)} \sum_{n=1}^{\infty} a_n(L_n - K_n) \to 0 \,. \tag{3.8}$$

Now consider the function

$$H_\eta(x) = \sum_{n=1}^{\infty} a_n G_n(x) U(x - n\eta) \,. \tag{3.9}$$

Evidently $H_\eta(x)$ is non-decreasing, since each term in the summation is non-decreasing. We also note that

$$\begin{aligned} H_\eta(x) &= \sum_{n=1}^{\infty} a_n U(x - n\mu) \\ &\quad - \sum_{n=1}^{\infty} a_n\{U(x - n\mu) - G_n(x)\} U(x - n\eta) \,. \end{aligned} \tag{3.10}$$

Let us denote the Laplace transform of a function $A(x)$, say, thus:-

$$A^0(s) = \int_0^\infty e^{-sx} A(x)dx .$$

Then, from (3.10), we have

$$H_\eta^0(s) = \frac{1}{s} \sum_{n=1}^\infty a_n e^{-n\mu s} + \sum_{n=1}^\infty a_n (L_n - K_n) ; \tag{3.11}$$

the term-by-term integration being justified by monotone convergence.

From (3.11), (3.8), and Lemma 4, it now appears that

$$\frac{(\mu s)^{\gamma+1}}{L\left(\frac{1}{s}\right)} H_\eta^0(s) \to \alpha\mu , \qquad \text{as } s \to 0+ . \tag{3.12}$$

An appeal to Doetsch (1950, p. 511) then allows the inference

$$\frac{\mu^\gamma \Gamma(1+\gamma)}{t^\gamma L(t)} H_\eta(t) \to \alpha \qquad \text{as } t \to \infty . \tag{3.13}$$

But, by (3.9),

$$\sum_{n=1}^\infty a_n G_n(x) = H_\eta(x) + \Psi_\eta(x) , \qquad \text{say} , \tag{3.14}$$

where

$$\Psi_\eta(x) = \sum_{n=1}^\infty a_n G_n(x)\{1 - U(x - n\eta)\} . \tag{3.15}$$

If we were to prove that

$$\lim_{\eta\to 0} \limsup_{x\to\infty} \frac{\Psi_\eta(x)}{x^\gamma L(x)} = 0 \tag{3.16}$$

then the theorem would follow from (3.14) and (3.13). The proof of (3.16) under fairly weak hypotheses is quite involved, however, and we therefore present it in the following two separate sections.

4. Completion of proof under (T6). If $\{X_n\}$ is the renewal sequence under study let us write $X_n^- = -X_n$ when $X_n < 0$, $X_n^- = 0$ when $X_n \geqq 0$. When (T6) holds we can introduce the distribution function $K(x)$ which, as has already been explained in § 1, may be assumed to have a finite first moment $-\kappa$. Therefore, if we write $\nu_n = EX_n^-$, we have $0 \leqq \nu_n \leqq \kappa$ for all n.

Let us also write $X_n^+ = X_n + X_n^-$, $S_n^+ = X_1^+ + X_2^+ + \cdots + X_n^+$, and $S_n^- = X_1^- + X_2^- + \cdots + X_n^-$.

LEMMA 6. *When* (T6) *holds we can find* $\eta > 0$, $\delta > 0$, *such that*

$$P\{S_n^+ \leqq \nu_1 + \nu_2 + \cdots + \nu_n + n\eta\} = O\left(\frac{1}{n^{k+\delta+1}}\right), \tag{4.1}$$

where k *is the index of the coefficient sequence* $\{a_n\}$.

Proof. Let us write $\bar{\nu}_n = n^{-1}(\nu_1 + \nu_2 + \cdots + \nu_n)$. Then, for any $t \geqq 0$, $\eta > 0$, it is plain that

$$\begin{aligned} P\{S_n^+ \leqq n(\bar{\nu}_n + \eta)\} &\leqq e^{n(\eta+\bar{\nu}_n)t} E e^{-tS_n^+} \\ &= e^{n(\eta+\bar{\nu}_n)t} \prod_{j=1}^{n} \Phi_j^+(t) , \end{aligned} \tag{4.2}$$

where

$$\Phi_j^+(t) = F_j(0+) + \int_{0+}^{\infty} e^{-tx} dF_j(x) .$$

If we now use the familiar inequality already employed in § 2 we can deduce from (4.2) that

$$P\{S_n^+ \leqq n(\bar{\nu}_n + \eta)\} \leqq e^{W_n(t)} , \tag{4.3}$$

where, after some integrations by parts, we now have

$$W_n(t) = n(\eta + \bar{\nu}_n)t - t\sum_{j=1}^{n}\int_0^{\infty} e^{-tx}\{1 - F_j(x)\}dx . \tag{4.4}$$

In this section we are assuming (1.10) to hold, and so, given any fixed $\varepsilon > 0$, we have for all sufficiently large n that

$$\sum_{j=1}^{n}\int_{-\infty}^{\varepsilon n/\log n}\{U(x) - F_j(x)\}dx > 2n\sqrt{(k+1)v\varepsilon} ,$$

where $v > \kappa$ is independent of ε.

We can rewrite this last inequality as follows.

$$\sum_{j=1}^{n}\int_{0}^{\varepsilon n/\log n}\{1 - F_j(x)\}dx > n\bar{\nu}_n + 2n\sqrt{(k+1)v\varepsilon} \tag{4.5}$$

Thus, from (4.4), it is plain that

$$W_n(t) \leqq n(\eta + \bar{\nu}_n)t - nte^{-t\varepsilon n/\log n}(\bar{\nu}_n + 2\sqrt{(k+1)v\varepsilon}) .$$

If, in the latter inequality, we make the substitutions

$$t = t_n = \frac{\lambda \log n}{\varepsilon n} ,$$

$$\lambda = \sqrt{\frac{\varepsilon(k+1)}{\kappa}} ,$$

and if we note incidentally that $\kappa \geqq \bar{\nu}_n$ for all n, and $1 - e^{-x} < x$ for all $x > 0$, then we find that

$$\frac{W_n(t_n)}{\log n} < \frac{\lambda\eta}{\varepsilon} + (k+1) - 2e^{-\lambda}(k+1)\sqrt{\frac{v}{\kappa}} . \tag{4.6}$$

By taking ε sufficiently small we can make $e^{-\lambda}$ arbitrarily near unity and thus make

$$(k+1) - 2e^{-\lambda}(k+1)\sqrt{\frac{v}{\kappa}} < -(k+2\delta+1)$$

for some small $\delta > 0$ (recall that $v > \kappa$). Next choose η so small that $\lambda\eta < \varepsilon\delta$ and it follows from (4.6) that

$$W_n(t_n) < -(k+\delta+1)\log n\,. \tag{4.7}$$

Lemma 6 follows from (4.7) and (4.3).

In what follows we denote the familiar Stieltjes convolution of two distribution functions, say $A(x)$ and $B(x)$, by $A*B(x)$. We denote $A*A(x)$ by $A^{*2}(x)$, and, generally, $A*A^{*n}(x)$ by $A^{*(n+1)}(x)$, for $n = 1, 2, 3, \cdots$.

LEMMA 7. *When* (T6) *holds*

$$\sum_{n=1}^{\infty} n^k P\{S_n^- > n(\bar{\nu}_n + \delta)\}$$

is convergent for every $\delta > 0$.

Proof. Define $Z_n = \delta/2 + \nu_n - X_n^-$ and write $L_n(x) = P\{Z_n \leqq x\}$. If we recall that $\nu_n \leqq \kappa$ for all n, then we easily see that

$$U\left(x - \frac{1}{2}\delta - \kappa\right) \leqq L_n(x) \leqq K(x)\,.$$

This proves that Z_n is a *stochastically stable sequence* as defined by Smith (1962), whose Theorem 7 allows us to draw the following conclusion.

For every integer p *there is a distribution function* $K_p(x)$ *such that*

$$P\{Z_n + Z_{n+1} + \cdots + Z_{n+p-1} \leqq px\} \leqq K_p(x) \tag{4.8}$$

for all n *and all* x, *where*

$$I_p \equiv \int_{-\infty}^{0} K_p(x)dx \tag{4.9}$$

is finite for all p, *and* $I_p \to 0$ *as* $p \to \infty$.

Thus we can find $p_0(\delta/2)$ such that $I_{p_0} < \delta/2$. If Y is a random variable with distribution function K_{p_0} then it follows from (4.9) that $EY > -\delta/2$. Moreover, it is clear that $E\{|\min(0, Y)|^{k+2}\} < \infty$, since we can certainly suppose $K_p(x) \leqq K^{*p}(px)$.

Write $M(x)$ for the supremum of $P\{Z_1 + Z_2 + \cdots + Z_r \leqq p_0 x\}$ for

$r = 1, 2, \cdots, p_0 - 1$. Then if Y_0 is a random variable with distribution function $M(x)$ it is also apparent that $E\{|\min(0, Y_0)|^{k+2}\} < \infty$.

Now choose and fix $r = 0$, or 1, or 2, $\cdots$, or $p_0 - 1$. It follows from what we have established so far that

$$P\left\{\sum_{j=1}^{np_0+r} Z_j \leqq p_0 x\right\} \leqq M{*}K_{p_0}^{*n}(x)\,. \tag{4.10}$$

Let $Y_1, Y_2, \cdots$ be a sequence of independent random variables, identically distributed, with distribution function $K_{p_0}(x)$; let $Y_1, Y_2, \cdots$ be independent of Y_0. Then $E(Y_j + \delta/2) > 0$, for $j = 1, 2, 3, \cdots$ and $E\{|\min(0, Y_j + 1/2\delta)|^{k+2}\} < \infty$ for $j = 0, 1, 2, \cdots$. Thus it follows from Theorem A, quoted in § 1, that

$$\sum_{n=1}^{\infty} n^k P\left\{Y_0 + \sum_{j=1}^{n}(Y_j + \delta/2) \leqq 0\right\} < \infty\,,$$

that is,

$$\sum_{j=1}^{\infty} n^k M{*}K_{p_0}^{*n}\left(-\frac{1}{2}n\delta\right) < \infty\,. \tag{4.11}$$

From (4.10) and (4.11) we conclude that

$$\sum_{n=1}^{\infty} n^k P\left\{(np_0 + r)\left(\frac{1}{2}\delta + \bar{\nu}_{np_0+r}\right) \leqq S_{np_0+r}^- - \frac{1}{2}np_0\delta\right\} < \infty\,,$$

whence,

$$\sum_{n=1}^{\infty} n^k P\{S_{np_0+r}^- \geqq (np_0 + r)(\delta + \bar{\nu}_{np_0+r})\Big\} < \infty\,. \tag{4.12}$$

The lemma follows from (4.12) by letting $r = 0, 1, 2, \cdots, p_0 - 1$ in turn.

LEMMA 8. *When* (T6) *holds we can find* $\eta > 0$ *such that*

$$\sum_{n=1}^{\infty} n^k G_n(n\eta) < \infty\,.$$

Proof. We observe that

$$\begin{aligned} P\{S_n \leqq n\eta\} &= P\{S_n^+ \leqq n\eta + S_n^-\} \\ &\leqq P\{S_n^+ \leqq n\eta + S_n^-,\ S_n^- < n(\delta + \bar{\nu}_n)\} \\ &\quad + P\{S_n^- \geqq n(\delta + \bar{\nu}_n)\} \end{aligned}$$

for every $\delta > 0$. Hence

$$P\{S_n \leqq n\eta\} \leqq P\{S_n^+ \leqq n(\eta + \delta + \bar{\nu}_n)\} + P\{S_n^- \geqq n(\delta + \bar{\nu}_n)\}\,.$$

The lemma now follows from Lemmas 6 and 7 if we make $\eta + \delta$ sufficiently small.

The proof of (3.16) is now straightforward. We see from (3.15) that

$$\Psi_\eta(x) \leqq \sum_{n=1}^{\infty} a_n G_n(n\eta) ,$$

for all x. Therefore, since k is the *index* of the $\{a_n\}$ sequence, it follows from Lemma 8 that $\Psi_\eta(x)$ is bounded above. Since $x^\gamma L(x) \to \infty$ as $x \to \infty$, the truth of (3.16) is established.

5. Completion of proof under (T5). We begin by showing that, once we have proved (1.9), we can assume, with no loss of generality, certain convenient properties for the function $l(n)$. All that actually matters are the values taken by $l(n)$ for integer values of n; but we may clearly assume $l(x)$ to be a continuous function defined for all $x \geqq 1$. More to the point, we observe that if (1.9) holds for the function $l(x)$ then it also holds for any function $l_1(x) \leqq l(x)$. In this connection we prove the following;

LEMMA 9. *If, for $x \geqq 1$, $l(x)$ is an unbounded, continuous, and non-decreasing function of x then we can find another such function $l_1(x) \leqq l(x)$, defined on the same domain, with the additional property that $l_1(x)/x$ is non-increasing for all sufficiently large x, and tends to zero as $x \to \infty$.*

Proof. For $x \geqq 1$ define

$$l_1(x) = \log x + \inf_{1 \leqq y \leqq x} \{l(y) - \log (y)\} . \tag{5.1}$$

We shall show that $l_1(x)$ has all the requisite properties.

To begin with, since $l(y) - \log y$ is a continuous function in $[1, x]$ it attains its lower bound; we shall write $y(x) \leqq x$ for the largest y-value at which this lower bound is attained. Then

$$l_1(x) = \log x + l(y(x)) - \log y(x) . \tag{5.2}$$

Evidently $y(x)$ is a non-decreasing function of x. If $y(x) \to \infty$ as $x \to \infty$ then the fact that $l(x)$ is unbounded shows, in (5.2), that $l_1(x)$ is also unbounded; if $y(x)$ tends to a finite limit as $x \to \infty$, then the fact that $\log x$ is unbounded shows, also in (5.2), that $l_1(x)$ is unbounded. Incidentally, it is an easy deduction from (5.1) that $l_1(x) \leqq l(x)$.

Next choose an arbitrary value of x, x_1 say. Our argument will be given in two cases.

Case $y(x_1) < x_1$. The continuity of $l(y) - \log y$ ensures the existence of some open interval G, containing x_1, within which $y(x) = y(x_1)$. Hence, in G,

$$l_1(x) = \log x - \log x_1 + l_1(x_1) ;$$

from this equation it is clear that $l_1(x)$ is increasing in G and, by simple differentiation, $l_1(x)/x$ is decreasing in G.

Case $y(x_1) = x_1$. In this case, for any $h > 0$, $y(x_1 + h) \geqq x_1$; thus $l(y(x_1 + h)) \geqq l(x_1) = l_1(x_1)$. Hence, by (5.2),

$$l_1(x_1 + h) \geqq \log (x_1 + h) + l_1(x_1) - \log (y(x_1 + h)) . \tag{5.3}$$

Since $x_1 + h \geqq y(x_1 + h)$, it follows from (5.3) that $l_1(x_1 + h) \geqq l_1(x_1)$, i.e. $l_1(x)$ is increasing at x_1. But, from (5.1),

$$l_1(x_1 + h) - \log (x_1 + h) \leqq l_1(x_1) - \log x_1$$

from which we can infer that

$$\frac{l_1(x_1 + h)}{x_1 + h} - \frac{l_1(x_1)}{x_1} \leqq - \frac{h l_1(x_1)}{x_1(x_1 + h)} + \frac{h}{x_1(x_1 + h)} . \tag{5.4}$$

The right hand side of (5.4) is negative for all sufficiently large x_1, because $l_1(x)$ is unbounded and non-decreasing. Thus $l_1(x)/x$ is non-decreasing at x_1.

Finally, we remark that

$$l_1(x) \leqq \log x + l(1) ,$$

from which it is obvious that $l_1(x)/x \to 0$ as $x \to \infty$.

Let us next establish that conditions (T1) and (T2) of Theorem 1 do indeed imply the existence of some unbounded increasing function $l(n)$ for which (1.9) holds. We see that since (1.7) is true for every $\varepsilon > 0$, there must be an unbounded increasing function $w(n)$, say, such that

$$\frac{1}{n} \sum_{r=1}^{\infty} \int_{n/w(n)}^{\infty} \{1 - F_r(x)\} dx \to 0 \qquad \text{as } n \to \infty .$$

Therefore, because of (1.6), it follows that

$$\frac{1}{n} \sum_{r=1}^{n} \int_{0}^{n/w(n)} \{1 - F_r(x)\} dx \to \mu \qquad \text{as } n \to \infty .$$

Let $l(n)$ be some other unbounded increasing function; we leave this function somewhat arbitrary for the moment except for the supposition that it increases very much more slowly than $w(n)$. Let $t(n)$ be such that $r/l(r) \geqq n/w(n)$ for all $r > t(n)$. Then we can infer that, for any positive ε, and all sufficiently large n,

$$(\mu - \varepsilon) < \frac{1}{n} \sum_{r=1}^{t(n)} \int_0^{n/w(n)} \{1 - F_r(x)\} dx + \frac{1}{n} \sum_{r=t(n)+1}^{n} \int_0^{r/l(r)} \{1 - F_r(x)\} dx$$

$$= T_1(n) + T_2(n) , \qquad \text{say} .$$

But

$$\frac{1}{n} \sum_{r=1}^{t(n)} \int_0^{n/w(n)} \{1 - F_r(x)\} dx \leqq \frac{1}{n} (\mu_1 + \mu_2 + \cdots + \mu_{t(n)}) ,$$

and so we could conclude from (1.6) that $T_1(n) \to 0$ as $n \to \infty$ if only we could be sure that $t(n)/n \to 0$. It would than follow that $T_2(n) > (\mu - 2\varepsilon)$ for all large n; the desired conclusion (1.9) would then be proved, in view of the arbitrariness of ε.

For large n, since $w(n)$ is increasing,

$$w(n) \geqq w\left(n \frac{\log w(n)}{w(n)}\right)$$

and so, if we put $s(n) = n(\log w(n))/w(n)$,

$$\frac{s(n)}{\log w(s(n))} \geqq \frac{n}{w(n)} .$$

It is clear from Lemma 9 that we may assume $n/w(n)$ to be increasing for all large n; from this it is easily seen that $n/(\log w(n))$ is also increasing for all large n. Therefore, if we let $l(n) = \log w(n)$ we have that $r/l(r) \geqq n/w(n)$ for all $r \geqq s(n)$. Hence $t(n) \leqq s(n)$, and it is plain that $s(n)/n \to 0$ as $n \to \infty$; thus we have a function $l(n)$ which exhibits the desired behavior, and (1.9) is proved.

We now turn to the proof of (3.16) under the condition (T5) that the $\{X_n\}$ are nonnegative; for this proof we may, by the immediately preceding discussion, assume that $x/l(x)$ is unbounded and non-decreasing. We can then define $r^* = r^*(x)$ as the greatest integer such that $r^*/l(r^*) \leqq x$. We also write $s^* = s^*(x) = r^*((1 + e)x)$; thus $s^*(x)$ is the greatest positive integer such that $s^*/l(s^*) \leqq (1 + e)x$.

Choose a large positive C and consider the following three cases, in all of which x is assumed to be large.

(i) $cx \leqq n \leqq r^*(x)$. By considerations similar to those in the proof of Lemma 2 we have

$$G_n(x) \leqq e^{W_n(t)} , \tag{5.5}$$

where

$$W_n(t) = tx - t \sum_{j=1}^{n} \int_0^{\infty} e^{-tu} \{1 - F_j(u)\} du . \tag{5.6}$$

If we substitute $t = 1/(2x)$ and truncate the integrals at x in (5.6) then we find

$$W_n(1/(2x)) \leqq \frac{1}{2} - \frac{e^{-1/2}}{2x} \sum_{j=1}^{n} \int_0^x \{1 - F_j(u)\} du \,. \tag{5.7}$$

Since (T5) is assumed to hold we can, by (1.9), find $\delta > 0$ such that

$$\sum_{j=1}^{n} \int_0^{j/l(j)} \{1 - F_j(u)\} du > n\delta \tag{5.8}$$

for all sufficiently large n. But $n \leqq r^*$ so that (since $x/l(x)$ is non-decreasing) $j/l(j) \leqq x$ for $j = 1, 2, \cdots, n$. (Actually we have only shown that $x/l(x)$ is non-decreasing for all sufficiently large x; but this is adequate for our purpose if we note that $x/l(x) \to \infty$ as $x \to \infty$ and assume n large). Hence we can infer from (5.8) that

$$\sum_{j=1}^{n} \int_0^x \{1 - F_j(u)\} du > n\delta \tag{5.9}$$

and so, from (5.7), that

$$W_n(1/(2x)) \leqq \frac{1}{2} - \frac{e^{-1/2} n\delta}{2x} \,.$$

If we use this last inequality in (5.5) we deduce that

$$G_n(x) \leqq \exp\left\{\frac{1}{2} - \frac{e^{-1/2} n\delta}{2x}\right\}, \qquad n \leqq r^* \,. \tag{5.10}$$

(ii) $r^*(x) < n \leqq s^*(x)$. For this case we modify the kind of inequality we have been using on $G_n(x)$. Plainly

$$\begin{aligned} P\{S_n \leqq x\} \leqq P\{X_j \leqq x \text{ for all } j\} &= \prod_{j=1}^{n} F_j(x) \\ &\leqq \exp\left\{- \sum_{j=1}^{n} [1 - F_j(x)]\right\} . \end{aligned}$$

By forming the geometric mean of this last inequality and (5.5) we obtain a new inequality:

$$G_n(x) \leqq e^{R_n(t)} \,, \tag{5.11}$$

where

$$\begin{aligned} R_n(t) = \frac{1}{2} tx - \frac{1}{2} t \sum_{j=1}^{n} \int_0^\infty e^{-tu} \{1 - F_j(u)\} du \\ - \frac{1}{2} \sum_{j=1}^{n} \{1 - F_j(x)\} \,. \end{aligned} \tag{5.12}$$

If we truncate the integrals in (5.12) at x and substitute $t = 1/x$ then we can infer

(5.13)
$$R_n(1/x) \leqq \frac{1}{2} - \frac{e^{-1}}{2x}\sum_{j=1}^{n}\int_0^x \{1 - F_j(u)\}du - \frac{1}{2}\sum_{j=r^*+1}^{n}\{1 - F_j(x)\} .$$

At this point it is convenient to write

$$\lambda_j = \int_0^{j/l(j)} \{1 - F_j(u)\}du ;$$

then (5.9) can be rewritten

(5.14) $\quad (\lambda_1 + \lambda_2 + \cdots + \lambda_n) > n\delta , \quad n$ sufficiently large.

We shall also write

$$\alpha_j(x) = \int_0^x \{1 - F_j(u)\}du .$$

Therefore, for $j > r^*$ (and, consequently, $j/l(j) > x$) we have

$$\int_x^{j/l(j)} \{1 - F_j(u)\}du = \lambda_j - \alpha_j(x) .$$

A consequence of the last equation is that

(5.15)
$$\left(\frac{j}{l(j)} - x\right)(1 - F_j(x)) \geqq \lambda_j - \alpha_j(x) .$$

However, if $j \leqq s^*(x)$, then $j/l(j) \leqq (1 + e)x$, and from this inequality it follows that

$$\left(\frac{j}{l(j)} - x\right) \leqq ex .$$

Thus we have, from (5.15), that

$$1 - F_j(x) \geqq \frac{\lambda_j - \alpha_j(x)}{ex}, \; r^* < j \leqq s^* .$$

Using the last inequality we can infer from (5.13) that

$$\begin{aligned} R_n(1/x) &< \frac{1}{2} - \frac{1}{2ex}\sum_{1}^{r^*}\lambda_j - \frac{1}{2ex}\sum_{r^*+1}^{n}\alpha_j(x) \\ &\quad - \frac{1}{2ex}\sum_{r^*+1}^{n}(\lambda_j - \alpha_j(x)) \\ &= \frac{1}{2} - \frac{1}{2ex}\sum_{1}^{n}\lambda_j \\ &< \frac{1}{2} - \frac{n\delta}{2ex} , \end{aligned}$$

by (5.8).

Therefore, from (5.11), we discover that

$$G_n(x) < \exp\left\{\frac{1}{2} - \frac{n\delta}{2ex}\right\}, \qquad r^* < n \leqq s^* . \tag{5.16}$$

From (5.10) and (5.16) we can conclude that

$$\sum_{n=Cx}^{s^*(x)} a_n G_n(x) < e^{1/2} \sum_{n=Cx}^{\infty} a_n \exp\left\{-\frac{n\delta}{2ex}\right\} . \tag{5.17}$$

We quote here a theorem which can be immediately deduced from some results of Karamata (Hardy (1949), pp. 166–169, especially Theorems 110, 111).

THEOREM B. *Suppose that $\alpha(t)$ is a non-decreasing function of t and that $I(y) = \int_0^\infty e^{-yt} d\alpha(t)$ is convergent for $y > 0$, and that*

$$I(y) \sim \frac{L(y^{-1})}{y^\gamma}, \qquad as \quad y \to 0+ , \tag{5.18}$$

where $\gamma \geqq 0$ and $L(x)$ is a function of slow growth. Then, if $g(x)$ is a continuous function of bounded variation in $(0, 1)$, as $y \to 0+$ we have

(a) *in case $\gamma > 0$;*

$$\frac{y^\gamma}{L(y^{-1})} \int_0^\infty e^{-yt} g(e^{-yt}) d\alpha(t) \sim \frac{1}{\Gamma(\gamma)} \int_0^\infty e^{-t} t^{\gamma-1} g(e^{-t}) dt ,$$

(b) *in case $\gamma = 0$:*

$$\frac{1}{L(y^{-1})} \int_0^\infty e^{-yt} g(e^{-yt}) d\alpha(t) \sim g(1) .$$

Let us put, in this theorem,

$$\alpha(t) = \sum_{n \leqq t} a_n .$$

Then Lemma 4 shows that a relation like (5.18) holds.

Define

$$\begin{aligned} g(x) &= 1 , & 0 \leqq x \leqq a , \\ &= 2 - \frac{x}{a} , & a \leqq x \leqq 2a , \\ &= 0 , & 2a \leqq x \leqq 1 . \end{aligned}$$

Clearly this $g(x)$ satisfies the conditions of Theorem B. Thus, when $\gamma > 0$ we can deduce that

$$\limsup_{y\to 0+} \frac{y^\gamma}{L(y^{-1})} \sum_{n \geqq (1/y)\log(1/a)} a_n e^{-ny}$$
$$\leqq \frac{\alpha}{\mu^\gamma \Gamma(\gamma)} \int_{\log(1/2a)}^{\infty} e^{-t} t^{\gamma-1} dt \ .$$

In this last result, substitute $y = \delta/(2ex)$ and $\log(1/a) = C\delta/(2e)$. On being given any prescribed $\varepsilon > 0$ we can choose C sufficiently large for us to deduce, via (5.16), that

$$\limsup_{x\to\infty} \frac{1}{x^\gamma L(x)} \sum_{n=Cx}^{s^*(x)} a_n G_n(x) < \varepsilon \ . \tag{5.19}$$

When $\gamma = 0$ a similar result to (5.19) can also be proved by appeal to Theorem B.

(iii) $s^*(x) < n$. For this range of values for n we always have $n/l(n) > x$. If we define R_n and λ_n as for case (ii) then by arguments similar to the ones employed in that case we find

$$R_n(x^{-1}) < \frac{1}{2} - \frac{1}{2ex} \sum_1^{s^*} \lambda_j - \frac{1}{2ex} \sum_{s^*+1}^{n} \alpha_j(x)$$
$$- \frac{1}{2} \sum_{s^*+1}^{n} \left\{ \frac{\lambda_j - \alpha_j(x)}{\dfrac{j}{l(j)} - x} \right\} .$$

However, when $j > s^*$, $j/l(j) > (1+e)x$ and one can infer that

$$\left(\frac{j}{l(j)} - x\right) > ex \tag{5.20}$$

and deduce therefore that

$$R_n(x^{-1}) < \frac{1}{2} - \frac{1}{2ex} \sum_1^{s^*} \lambda_j - \frac{1}{2} \sum_{s^*+1}^{n} \left\{ \frac{\lambda_j}{\dfrac{j}{l(j)} - x} \right\} . \tag{5.21}$$

Write

$$T_n = \sum_{s^*+1}^{n} \left\{ \frac{\lambda_j}{\dfrac{j}{l(j)} - x} \right\}$$

and

$$\Lambda_j = \lambda_1 + \lambda_2 + \cdots + \lambda_j \ .$$

Then

$$T_n = \frac{\Lambda_n}{\dfrac{n}{l(n)} - x} - \frac{\Lambda_{s^*}}{\dfrac{s^*+1}{l(s^*+1)} - x} + \sum_{s^*+1}^{n-1} \Lambda_j \left\{ \frac{1}{\dfrac{j}{l(j)} - x} - \frac{1}{\dfrac{j+1}{l(j+1)} - x} \right\} .$$

We may assume s^* to be large, so that, by (5.14), $\Lambda_j > j\delta$ for all j in the range of consideration. Thus

$$\begin{aligned} T_n &> \frac{n\delta}{\dfrac{n}{l(n)} - x} - \frac{\Lambda_{s^*}}{\dfrac{s^*+1}{l(s^*+1)} - x} + \delta \sum_{s^*+1}^{n-1} j \left\{ \frac{1}{\dfrac{j}{l(j)} - x} - \frac{1}{\dfrac{j+1}{l(j+1)} - x} \right\} \\ &= \frac{(\delta s^* - \Lambda_{s^*})}{\dfrac{s^*+1}{l(s^*+1)} - x} + \delta \sum_{s^*+1}^{n} \left\{ \frac{1}{\dfrac{j}{l(j)} - x} \right\} . \end{aligned}$$

If we use this last inequality in (5.21), and also make use of (5.20), then we find

$$\begin{aligned} 2R_n(x^{-1}) &< 1 - \frac{s^*\delta}{\dfrac{s^*+1}{l(s^*+1)} - x} - \delta \sum_{s^*+1}^{n} \left\{ \frac{1}{\dfrac{j}{l(j)} - x} \right\} \\ &< 1 - \frac{s^*\delta}{\dfrac{s^*+1}{l(s^*+1)} - x} - \delta \sum_{s^*+1}^{n} \frac{l(j)}{j} \\ &< 1 - \frac{s^*\delta}{\dfrac{s^*+1}{l(s^*+1)} - x} - \delta l(s^*+1) \log\left(\frac{n+1}{s^*+1}\right) . \end{aligned}$$

We may therefore conclude, from (5.11), that

$$G_n(x) < \left(\frac{s^*+1}{n+1}\right)^{(1/2)\delta l(s^*+1)} e^{\psi(x)},\ n > s^* , \tag{5.22}$$

where

$$\psi(x) = \frac{1}{2} - \frac{s^*\delta}{2\left(\dfrac{s^*+1}{l(s^*+1)} - x\right)} .$$

At this point in our argument we need information about the order of magnitude of

$$U_N = \sum_N^\infty \frac{a_n}{n^\rho},$$

for $\rho > \gamma + 2$. To this end, define $\alpha(t) = \Sigma_{n \le t} a_n$, as before. Then there is some constant c such that $\alpha(t) < ct^\gamma L(t)$ for all large t, by (1.4).

Evidently

$$\begin{aligned} U_N &= \int_{N-0}^\infty \frac{1}{\mu^\rho} d\alpha(u) . \\ &= \left[\frac{1}{u^\rho}\alpha(u)\right]_{N-0}^\infty + \rho\int_N^\infty \frac{\alpha(u)}{u^{\rho+1}} du \\ &< c\rho\int_N^\infty \frac{L(u)}{u^{\rho-\gamma+1}} du , \end{aligned}$$

if we ignore a negative form. Hence, by an obvious substitution in the integral, we have

$$U_N < \frac{c\rho L(N)}{N^{\rho-\gamma}} \int_1^\infty \frac{L(Nv)}{L(N)} \frac{dv}{v^{\rho-\gamma+1}} . \tag{5.23}$$

But, from (1.2),

$$\frac{L(Nv)}{L(N)} = \frac{a(Nv)}{a(N)} \cdot \frac{1}{v} \cdot \exp\left\{\int_N^{Nv} \frac{a(u)}{u} du\right\},$$

where $a(x) \to 1$, as $x \to \infty$. Therefore, given an arbitrarily small $\varepsilon > 0$, we can choose N so large that for all $v \geqq 1$

$$\begin{aligned} \frac{L(Nv)}{L(N)} &< \frac{(1+\varepsilon)}{v} \exp\left\{(1+\varepsilon)\int_N^{Nv} \frac{du}{u}\right\} \\ &= (1+\varepsilon)v^\varepsilon . \end{aligned}$$

Hence we can appeal to dominated convergence to infer that

$$\int_1^\infty \frac{L(Nv)}{L(N)} \frac{dv}{v^{\rho-\gamma+1}} \to \int_1^\infty \frac{dv}{v^{\rho-\gamma+1}}$$

as $N \to \infty$. It then transpires, from (5.23), that

$$U_N = O\left(\frac{L(N)}{N^{\rho-\gamma}}\right) . \tag{5.24}$$

For all sufficiently large x we shall have $(1/2)\delta l(s^* + 1) > \gamma + 2$, and hence may infer from (5.22) and (5.24) that

$$\sum_{n=s^*+1}^\infty a_n G_n(x) = O((s^* + 1)^\gamma L(s^* + 1)e^{\psi(x)})$$

and hence that

$$(5.25)\qquad \frac{1}{x^{\gamma}L(x)}\sum_{n=s^*+1}^{\infty} a_nG_n(x) = O\Big(\Big(\frac{s^*+1}{x}\Big)^{\gamma}\frac{L(s^*+1)}{L(x)}e^{\psi(x)}\Big) .$$

Since

$$\begin{aligned}\frac{s^*+1}{l(s^*+1)} &= \Big(\frac{s^*+1}{s^*}\Big)\Big(\frac{l(s^*)}{l(s^*+1)}\Big)\Big(\frac{s^*}{l(s^*)}\Big)\\ &\leqq \Big(\frac{s^*+1}{s^*}\Big)(1+e)x ,\end{aligned}$$

it is clear that for all large x

$$\frac{s^*+1}{l(s^*+1)} < (1+2e)x .$$

Therefore, from the definition of $\psi(x)$ which follows (5.22) we infer that

$$(5.26)\qquad \psi(x) < \frac{1}{2} - \frac{s^*\delta}{4ex} .$$

In addition, we can deduce from (1.2) that

$$\frac{L(s^*+1)}{L(x)} = \frac{a(s^*+1)}{a(x)}\Big(\frac{x}{s^*+1}\Big)\exp\Big\{\int_x^{s^*+1}\frac{a(u)}{u}du\Big\}$$

and so,

$$(5.27)\qquad \frac{L(s^*+1)}{L(x)} = O\Big(\Big(\frac{s^*+1}{x}\Big)^{\varepsilon}\Big)$$

for any $\varepsilon > 0$.

If we combine (5.26) and (5.27) with (5.25) we discover that

$$(5.28)\qquad \frac{1}{x^{\gamma}L(x)}\sum_{n=s^*+1}^{\infty} a_nG_n(x) = O\Big(\Big(\frac{s^*+1}{x}\Big)^{\gamma+\varepsilon}\exp\Big(-\frac{s^*\delta}{4ex}\Big)\Big) .$$

But $(s^*+1) > (1+e)xl(s^*+1)$, so that $(s^*+1)/x \to \infty$ as $x \to \infty$. We can therefore deduce from (5.28) that

$$\frac{1}{x^{\gamma}L(x)}\sum_{n=s^*+1}^{\infty} a_nG_n(x) \to 0 , \quad \text{as} \quad x \to \infty .$$

On combining this last result with (5.19) we find that given any $\varepsilon > 0$ we can choose a sufficiently large $C > 0$ so that

$$\limsup_{x\to\infty} \frac{1}{x^{\gamma}L(x)}\sum_{n=Cx}^{\infty} a_nG_n(x) < \varepsilon .$$

This result establishes (3.16) and completes our proof, since

$$\Psi_{\eta}(x) = \sum_{n>(x/\eta)}^{\infty} a_nG_n(x) .$$

References

1. G. Doetsch, *Handbuch der Laplace-Transformation*, Vol. I, (1950) Verlag Birkhäuser, Basel.

2. B. V. Gnedenko and A. N. Kolmogorov, *Limit distributions for sums of independent random variables* (Translated from the Russian by K. L. Chung), (1954), Cambridge, Mass., Addison-Wesley.

3. G. H. Hardy, *Divergent Series*, (1949), Oxford University Press.

4. H. Hatori, *Some theorems in an extended renewal theory, I*, Kōdai Math. Sem. Rep., **11** (1959), 139-146.

5. ———, *A note on a renewal theorem*, Kōdai Math. Sem. Rep., **12** (1960), 28-37.

6. J. Karamata, *Sur un mode de croissance régulière des fonctions*, Mathematics (Cluj), **4** (1930), 38-53.

7. M. L. Katz, *The probability in the tail of a distribution*, Ann. Math. Statist., **34** (1963), 312-318.

8. T. Kawata, *A renewal theorem*, J. Math. Soc. Japan, **8** (1956), 118-126.

9. H. Robbins and Y. S. Chow, *Abstracts of short communications*, (1962), International Congress of Mathematicians, Stockholm.

10. W. L. Smith, *On some general renewal theorems for nonidentically distributed variables*, Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Vol. II, (1962), 467-514, Berkeley, University of California Press.

11. ———, *On functions of characteristic functions and their application to some renewal-theoretic random walk problems*, (1964), to appear.

DOUBLY INVARIANT SUBSPACES

T. P. SRINIVASAN

1. Our theme is a theorem on doubly invariant subspaces attributed to Wiener in the folk lore; our discussion was inspired by that of Helson-Lowdenslager [2] on simply invariant subspaces and a course of lectures by Professor Helson on this subject. Let $\mathscr{M}$ denote a closed subspace of L^2 of the circle $|z| = 1$, which we shall denote as $L^2(e^{ix})$. Let λ denote the function on $|z| = 1$ defined by $\lambda(e^{ix}) = e^{ix}$. Say that $\mathscr{M}$ is *doubly invariant* if $f \in \mathscr{M} \Rightarrow \lambda f$, $\lambda^{-1}f \in \mathscr{M}$. An example of such a subspace is the set of all $f \in L^2(e^{ix})$ which vanish on a fixed measurable subset E. Wiener's theorem asserts that every doubly invariant $\mathscr{M}$ is of this form. A similar result holds for L^2 of the real line too (which we shall denote as $L^2(dt)$). In this case a doubly invariant subspace is any closed subspace $\mathscr{M}$ of $L^2(dt)$ such that $f \in \mathscr{M} \Rightarrow e^{iut}f \in \mathscr{M}$ for all real u, and every such subspace consists precisely of all functions in $L^2(dt)$ which vanish on a fixed measurable subset E of the line. In either case—the circle or the line—$\mathscr{M}$ determines E uniquely. We shall refer to either of these cases as the scalar case.

Wiener's theorem extends to L^2 spaces of vector valued functions on the circle or the line. Let $\mathscr{H}$ be any separable Hilbert space and $L^2_{\mathscr{H}}$ denote the set of all functions on $|z| = 1$ with values in $\mathscr{H}$ which are weakly measurable and whose norms are square integrable. $L^2_{\mathscr{H}}$ is a Hilbert space for the inner product

$$(f, g) = \int_{-\pi}^{\pi} (f(e^{ix}), g(e^{ix}))d\sigma$$

where the inner product on the right is the one in $\mathscr{H}$ and $d\sigma = (1/2\pi)dx$. The doubly invariant subspaces of $L^2_{\mathscr{H}}$ are defined exactly as before. An example of such a subspace in this case can be given as follows:

Let $\mathscr{J}$ be a *range function* meaning a function on $|z| = 1$ to the family of closed subspaces of $\mathscr{H}$, defined a.e. Two range functions which agree a.e. are regarded as the same function. Let $P(e^{ix})$ be the self adjoint projection on $\mathscr{J}(e^{ix})$. Say that that $\mathscr{J}$ is "measurable" if P is weakly measurable. Given $\mathscr{J}$ measurable, let $\mathscr{M}_{\mathscr{J}}$ be the set of all functions $f \in L^2_{\mathscr{H}}$ for which $f(e^{ix}) \in \mathscr{J}(e^{ix})$ a.e. Then $\mathscr{M}_{\mathscr{J}}$ is a doubly invariant subspace of $L^2_{\mathscr{H}}$. The version of Wiener's theorem in this case will be that every doubly invariant subspace of $L^2_{\mathscr{H}}$ is obtained as above from a measurable range function $\mathscr{J}$ and

Received May 10, 1963. This work was done while I held a visiting appointment in the University of California, Berkeley.

$\mathscr{M}_{\mathscr{J}}$ determines $\mathscr{J}$ uniquely. The scalar case corresponds to one dimensional $\mathscr{H}$ in which case $\mathscr{J}(e^{ix})$ can have only one of two values, either $\{0\}$ or the whole space, so that specifying $\mathscr{J}(e^{ix})$ is merely prescribing the set on which all functions in $\mathscr{M}_{\mathscr{J}}$ vanish. Thus the above indeed generalizes the scalar case for the circle. The generalization of the line case to the vector context is now obvious.

In both the scalar and vector cases, the circle or the line and the associated Lebesgue measure are inessential. Let X be any locally compact space and m a regular Borel measure on X and let P be any subspace of $L_\infty(dm)$ which is weak* dense. Say that a closed subspace $\mathscr{M}$ of $L^2(dm)$ or $L^2_{\mathscr{H}}(dm)$ is *doubly invariant* if it is invariant for multiplication by functions in P. Then the doubly invariant subspaces of $L^2(dm)$ or $L^2_{\mathscr{H}}(dm)$ have precisely the same structure as in the circle or the line case. The circle corresponds to the situation '$m(X) < \infty$' and the line to '$m(X) = \infty$'; the subspace P corresponds in either case to the set of all trignometric polynomials.

In this paper we first give a proof of Wiener's theorem for the scalar circle case and show that essentially the same proof applies to the line case too. We then generalize our proof to yield the vector case. Our proof for the (scalar and vector) circle case applies word for word (with obvious changes) to the context of finite regular measure spaces mentioned above; our proof of the line case could be adapted to the context of infinite measure spaces. By modifying our proof for the vector case we obtain a theorem (Theorem 5) on range functions of constant dimension which incidentally gives a characterization of range functions associated with *simply invariant* subspaces with no *remote past* (Theorem 6). Finally we show that in the scalar case the $L^2(dm)$ theorem implies a corresponding $L^p(dm)$ theorem (Theorem 7), $1 \leqq p \leqq \infty$[1].

The Wiener L^2 theorem is known. In the scalar case, direct proofs are also known; our proof seems to be simpler. In the vector case our version of the theorem was suggested by Professors Helson and Lowdenslager; we have not seen in the literature a direct proof of the theorem in this case. It could be derived as a corollary from the following general theorem in the theory of 'rings of operators':

Any bounded operator $T: L^2_{\mathscr{H}} \to L^2_{\mathscr{H}}$ which commutes with multiplication by bounded scalar functions is multiplication by a bounded operator valued function. [cf: 1, p. 167, Theorem 1; 3, p. 301, Lemma 1] The proof this way would be more involved. Our L^p theorem and Theorem 6, we believe, are new.

We may point out in passing that the general theorem on multiplication operators quoted above can itself derived from Wiener's

[1] The $L^p(dm)$ theorem for $p \neq 2$ is of interest as it exhibits a class of subspaces of $L^p(dm)$ which admit bounded projections.

theorem by an application of the spectral theorem for self adjoint operators. We omit the proof of this.

We have benefitted considerably by our discussion with Professor Helson in the course of preparation of this paper and our thanks are due to him.

2. THEOREM 1. *Let* $\mathscr{M}$ *be a doubly invariant subspace of* $L^2(e^{ix})$. *Then* $\mathscr{M} = C_E L^2(e^{ix})$ *for some measurable sebset* E (*where* C_E *denotes the characteristic function of* E).

Proof. Let $\mathscr{M}^\perp$ be the orthogonal complement of $\mathscr{M}$ in $L^2(e^{ix})$ and let q be the orthogonal projection on $\mathscr{M}$ of the constant function 1. Then $1 - q \in \mathscr{M}^\perp$, and because of double invariance of $\mathscr{M}$ and hence of $\mathscr{M}^\perp$, $\lambda^n(1 - q) \in \mathscr{M}^\perp$ for all n. So $\int (q - |q|^2) e^{-nix} d\sigma = 0$ for all n so that $|q|^2 = q$ a.e. Hence $q = C_E$ for some measurable subset E.

Trivially $qL^2(e^{ix}) \subset \mathscr{M}$. This inclusion is in fact an equality. For if $g \in \mathscr{M} \ominus qL^2(e^{ix})$ then $g \perp \lambda^n q$ for all n, also $g \perp \lambda^n(1 - q)$ (which lies in $\mathscr{M}^\perp$), so $g \perp \lambda^n$ for all n and hence $g = 0$ a.e. Thus $\mathscr{M} = qL^2(e^{ix}) = C_E L^2(e^{ix})$. We pass now to the line case:

THEOREM 2. *Let* $\mathscr{M}$ *be a doubly invariant subspace of* $L^2(dt)$, $-\infty < t < \infty$. *Then* $\mathscr{M} = C_E L^2(dt)$ *for some measurable subset* E *of the line*.

Proof. Let $\tilde{L}^2 = (1 - it)L^2(dt)$ and $\tilde{\mathscr{M}} = (1 - it)\mathscr{M}$. $\tilde{L}^2$ is a Hilbert space for the inner product

$$(f, g) = \int_{-\infty}^{\infty} f\bar{g} \frac{1}{1 + t^2} dt$$

and $\tilde{\mathscr{M}}$ is a closed subspace of $\tilde{L}^2$ invariant under multiplication by all e^{iut}. Let $\tilde{\mathscr{M}}^\perp$ be the orthogonal complement of $\tilde{\mathscr{M}}$ in $\tilde{L}^2$ and let q be the projection of the constant function 1 (which belongs to $\tilde{L}^2$) on $\tilde{\mathscr{M}}$. Now the arguments are the same as in the circle case:

$(1 - q)e^{iut} \in \tilde{\mathscr{M}}^\perp$ for all u and hence $q \perp (1 - q)e^{iut}$ for all u. That is

$$\int_{-\infty}^{\infty} (q - |q|^2) \frac{1}{1 + t^2} e^{-iut} dt = 0 \qquad \text{for all } u.$$

Hence $(q - |q|^2)(1/(1 + t^2)) = 0$ a.e. Thus $|q|^2 = q$ a.e. and $q = C_E$ for some E. Then as in the circle case, $\tilde{\mathscr{M}} = q\tilde{L}^2 = C_E \tilde{L}^2$, i.e. $(1 - it)\mathscr{M} = (1 - it)C_E L^2$. Hence $\mathscr{M} = C_E L^2$. The uniqueness of E is trivial in both the cases.

3.1. We deal with the vector case for the circle. Let $\mathscr{H}$ be a separable Hilbert-space and $L^2_{\mathscr{H}}$ be defined as in § 1. Then we have

THEOREM 3. *For every doubly invariant subspace $\mathscr{M}$ of $L^2_{\mathscr{H}}$ there exists a unique measurable range function $\mathscr{J}$ such that $\mathscr{M} = \mathscr{M}_{\mathscr{J}}$.*

Proof. Let $\{e_k\}$ $k = 1, 2, \cdots$ be an orthonormal basis for $\mathscr{H}$ and let q_k be the projection of the constant function e_k on $\mathscr{M}$. Then $q_k \in L^2_{\mathscr{H}}$ and of course is measurable. Each q_k is defined a.e. on the circle and hence also all q_k's together. Let $\mathscr{J}(e^{ix})$ be the closed subspace of $\mathscr{H}$ spanned by $\{q_k(e^{ix})\}_k$. Then $\mathscr{J}(e^{ix})$ is defined a.e. We shall show that

(a) $\mathscr{J}$ is measurable
(b) $\mathscr{M} = \mathscr{M}_{\mathscr{J}}$

Proof of (a). Let $P(e^{ix})$ be the orthogonal projection on $\mathscr{J}(e^{ix})$. We have only to show that $P(e^{ix})e_k$ is measurable for all k. We shall actually show that $P(e^{ix})e_k = q_k(e^{ix})$ a.e. Let $\mathscr{M}^{\perp} = L^2_{\mathscr{H}} \ominus \mathscr{M}$. Now $q_k \in \mathscr{M}$ and $e_k - q_k \in \mathscr{M}^{\perp}$. Because of double invariance then, $\lambda^n q_r \in \mathscr{M}$ for all n, and is $\perp e_k - q_k$ for all k. Thus $\int (e_k - q_k(e^{ix}), q_r(e^{ix}))e^{-nix}d\sigma = 0$ for all n and hence $e_k - q_k(e^{ix}) \perp q_r(e^{ix})$ a.e. for every r so that $e_k - q_k(e^{ix}) \perp q_r(e^{ix})$ for all r, a.e. This means $e_k - q_k(e^{ix}) \perp \mathscr{J}(e^{ix})$ a.e. Since $q_k(e^{ix}) \in \mathscr{J}(e^{ix})$ it follows that $P(e^{ix})e_k = q_k(e^{ix})$ a.e.

Proof of (b). Let $\mathscr{N}$ be the closed span of $\{\lambda^n q_k\}$ in $L^2_{\mathscr{H}}$, $k \geqq 1$, $n = 0, \pm 1, \pm 2, \cdots$. Then $\mathscr{N}$ is doubly invariant and $\mathscr{N} \subset \mathscr{M}$. If $\mathscr{N} \neq \mathscr{M}$ let $g \in \mathscr{M} \ominus \mathscr{N}$. Then, using the invariance, we have

(i) $g \perp \lambda^n q_k$ for all k, n
(ii) $\lambda^n g \perp e_k - q_k$ for all k, n.

It follows as in the proof of (a) that

(i) $g(e^{ix}) \perp q_k(e^{ix})$ a.e.
(ii) $g(e^{ix}) \perp e_k - q_k(e^{ix})$ a.e.

Hence $g(e^{ix}) \perp e_k$ a.e. for every k so that $g(e^{ik}) \perp e_k$ for all k, a.e. Hence $g(e^{ix}) = 0$ a.e. This shows that $\mathscr{M} = \mathscr{N}$.

If $f \in \mathscr{N}$ then $f(e^{ix}) \in \mathscr{J}(e^{ix})$ a.e. Hence $\mathscr{M} \subset \mathscr{M}_j$. Let now $g \in \mathscr{M}_{\mathscr{J}} \ominus \mathscr{M}$. Then $g \perp \lambda^n q_k$ for all k, n, so $g(e^{ix}) \perp q_k(e^{ix})$ a.e. for every k and hence $g(e^{ix}) \perp \mathscr{J}(e^{ix})$ a.e. But $g(e^{ix}) \in \mathscr{J}(e^{ix})$ a.e. as $g \in \mathscr{M}_{\mathscr{J}}$. Hence $g = 0$. Thus $\mathscr{M} = \mathscr{M}_{\mathscr{J}}$.

Only the uniqueness part of the theorem remains to be proved. This we prove independently as a lemma.

LEMMA. *If $\mathscr{J}$ and $\mathscr{K}$ are measurable range functions and $\mathscr{M}_{\mathscr{J}} = \mathscr{M}_{\mathscr{K}}$ then $\mathscr{J} = \mathscr{K}$ a.e.*

Proof. Let as before $P(e^{ix})$ be the orthogonal projection on $\mathscr{J}(e^{ix})$ and let $q_k(e^{ix}) = P(e^{ix})e_k$, $k = 1, 2 \cdots$ where $\{e_k\}$ is an o.n. basis for $\mathscr{J}$. q_k is measurable as $\mathscr{J}$ is and $\| q_k(e^{ix}) \|^2 \leqq \| e_k \|^2 = 1$ so that $q_k \in L^2_{\mathscr{H}}$. Also $\{q_k(e^{ix})\}_k$ generate $\mathscr{J}(e^{ix})$ as $\{e_k\}$ generate $\mathscr{H}$. Now $q_k \in \mathscr{M}_{\mathscr{J}} = \mathscr{M}_{\mathscr{K}}$ so that $q_k(e^{ix}) \in \mathscr{K}(e^{ix})$ a.e. for all k. It follows that $\mathscr{J}(e^{ix}) \subset \mathscr{K}(e^{ix})$ a.e. Interchanging $\mathscr{J}$ and $\mathscr{K}$ we conclude that $\mathscr{J} = \mathscr{K}$ a.e.

3.2. The functions $\{q_k\}$ defined in § 3.1 provide a measurable basis pointwise a.e. for $\mathscr{J}$. We shall show that we can secure the $\{q_k\}$ to be orthogonal a.e. The usual orthogonalization process can be applied at every point but the measurability of the resulting functions needs to be proved. This can be avoided by a slight modification of our construction of the q_k's which while preserving their other properties also ensures their pointwise orthogonality. The modification is the following:

Let q_1 be the orthogonal projection of e_1 on $\mathscr{M}$ and let $\mathscr{N}_1$ be the closed span of $\{\lambda^n q_1\}_n$. Then $\mathscr{N}_1$ is doubly invariant and so is $\mathscr{M}_1 = \mathscr{M} \ominus \mathscr{N}_1$. Let now q_2 be the projection of e_2 on $\mathscr{M}_1$ and let $\mathscr{N}_2 \subset \mathscr{M}_1$ be the closed span of $\{\lambda^n q_2\}$. Having obtained $q_1, q_2, \cdots, q_{k-1}$ and $\mathscr{N}_1, \mathscr{N}_2, \cdots \mathscr{N}_{k-1}$ as above, define q_k as the projection of e_k on $\mathscr{M} \ominus \sum_{i=1}^{k-1} \mathscr{N}_i$. The q_k's are easily seen to be mutually orthogonal a.e. If $\mathscr{J}(e^{ix})$ is defined to be the closed span of $\{q_k(e^{ix})\}_k$, the arguments in §3.1 which trivial modifications will show that $\mathscr{M} = \mathscr{M}_{\mathscr{J}}$. We have thus proved

THEOREM 4. *Corresponding to every measurable range function there exist functions $q_k \in L^2_{\mathscr{H}}$, $k = 1, 2, \cdots$ such that the $q_k(e^{ix})$'s are mutually orthogonal and span $\mathscr{J}(e^{ix})$ a.e.*

The question that arises next is: when does $\mathscr{J}(e^{ix})$ have a measurable o.n. basis a.e.? If $\{q_k(e^{ix})\}_k$ is an o.n. basis a.e. for $\mathscr{J}(e^{ix})$ then the dimension of $\mathscr{J}(e^{ix})$ is a constant a.e., being equal to the cardinality of the indexing k's (finite or not). Conversely also we have

THEOREM 5. *If $\mathscr{J}$ is a measurable range function of constant dimension a.e., there exist functions q'_k, $k = 1, 2, \cdots$ in $L^2_{\mathscr{H}}$ such that $\{q'_k(e^{ix})\}$ is an o.n. basis for $\mathscr{J}(e^{ix})$* a.e.

Proof. By our construction in the proof of Theorem 4 we can assume that there exist $q_k \in L_{\mathscr{H}}$, $k = 1, 2, \cdots$ such that $\| q_k(e^{ix}) \| = 1$ or 0 a.e. and $\{q_k(e^{ix})\}_k$ is orthogonal and generates $\mathscr{J}(e^{ix})$. For a given x let $q'_1(e^{ix}) = q_{i_1}(e^{ix})$ where i_1 is the smallest index such that $q_{i_1}(e^{ix}) \neq 0$; having obtained $q'_1(e^{ix}), \cdots, q'_{n-1}(e^{ix})$, let $q'_n(e^{ix}) = q_{i_n}(e^{ix})$ where i_n is the smallest index $\geqq i_{n-1}+1$ such that $q_{i_n}(e^{ix}) \neq 0$.[1] If dimension $\mathscr{J}(e^{ix}) = \infty$ a.e., this construction defines q'_n for every n; if dimension $\mathscr{J}(e^{ix}) = N < \infty$ a.e., the construction proceeds exactly N steps and defines $q'_1, q'_2, \cdots, q'_N$ a.e. The verification that the q'_k's satisfy the requirements of the theorem is not hard.

The above theorem has an interesting corollary. Say that a closed subspace $\mathscr{M} \subset L^2_{\mathscr{H}}$ is "simply invariant" if $\lambda^n \mathscr{M} \subset \mathscr{M}$ for all $n \geqq 0$ but not for all $n < 0$. The range function $\mathscr{J}$ associated with the smallest doubly invariant subspace containing $\mathscr{M}$, we shall call the "range function of $\mathscr{M}$". The subspace $\mathscr{M}_\infty = \bigcap_{n \geqq 0} \lambda^n \mathscr{M}$, we shall call the "remote past" of $\mathscr{M}$. If $\mathscr{M}_\infty = \{0\}$ (when $\mathscr{M}$ is said to be without remote past) it can be shown from the $L^2_{\mathscr{H}}$ version of a theorem of Lax [3, p. 300] that the associated range function is of constant dimension a.e. (meaning finite and equal or infinite a.e.). Conversely, if $\mathscr{J}$ is any measurable range function of constant dimension, by Theorem 5 it has a pointwise o.n. basis $\{q'_k(e^{ix})\}_k$, $q'_k \in L^2_{\mathscr{H}}$. Then $\{\lambda^n q'_k\}_{k,n}$ is an o.n. set in $L^2_{\mathscr{H}}$. If $\mathscr{N}_m$ is the closed span of $\{\lambda^m q'_k\}_k$, the $\mathscr{N}_m$'s are mutually orthogonal in $L^2_{\mathscr{H}}$ for $m = 0, \pm 1, \pm 2, \cdots$ and the orthogonal sum $\mathscr{M} = \sum_{m \geqq 0} \mathscr{N}_m$ is a simply invariant subspace of $L^2_{\mathscr{H}}$ without remote past whose range function is the given $\mathscr{J}$. Thus we have

THEOREM 6. *A measurable range function is of constant dimension a.e. if and only if it is the range function of a simply invariant subspace without remote past.*

4. The modification employed in § 2 for discussing the line case in the scalar context carries over without change to the vector situation and extends Theorems 3–5 to $L^2_{\mathscr{H}}$ over the line. Theorem 6 remains true but needs to be discussed anew; we omit the details.

5. Let m be a regular Baire measure on a locally compact space X and P a subspace of $L^\infty(dm)$ which is weak* dense. The reasoning given in § 2–3 shows that the doubly invariant subspaces $\mathscr{M}$ of $L^2(dm)$ are the subspaces of the form $C_E L^2(dm)$, $E \subset X$ measurable. Using this we wish to prove the following

THEOREM 7. *Let $\mathscr{N}$ be a subspace of $L^p(dm)$ which is invariant under multiplication by functions in P and which is closed if $1 \leqq p < \infty$ and weak* closed if $p = \infty$. Then $\mathscr{N} = C_E L^p(dm)$ for some measurable subset E of X.*

---

[1] This construction resulted from a discussion with Professor Ju-kwei Wang.

Proof.

Case (i) $1 \leqq p < 2$:

Let $\mathscr{M} = \mathscr{N} \cap L^2(dm)$. Then $\mathscr{M}$ is a doubly invariant subspace of $L^2(dm)$ and so $\mathscr{M} = C_E L^2(dm)$ for some measurable subset E. We shall show that $\mathscr{N} = C_E L^p(dm)$.

Let $f \in \mathscr{N}$ and $f = f_1 f_2$ be any factorization for f as a product of an L^μ function and L^2 function where $(1/\mu)+(1/2)=(1/p)$, for instance $f_2 = |f|^{p/2}$ and $f_1 = (\text{sgn.} f)\,|f|^{1-(p/2)}$. Let P_a be the subalgebra generated by P and constants in $L^\infty(dm)$. The closed subspace $[f_2 P_a]_2$ generated by $f_2 P_a$ in $L^2(dm)$ is doubly invariant and hence $[f_2 P_a]_2 = C_{E_2} L^2(dm)$ for some $E_2 \subset X$. Now

$$f_1 C_{E_2} \in f_1 C_{E_2} L^2(dm) = f_1 [f_2 P_a]_2 \subset [f_1 f_2 P_a]_p \subset \mathscr{N}$$

Trivially $f_1 C_{E_2} \in L^\mu(dm) \subset L^2(dm)$. Hence

$$f_1 C_{E_2} \in \mathscr{N} \cap L^2(dm) = \mathscr{M} = C_E L^2(dm)\,.$$

Let $f_1 C_{E_2} = C_E g$, $g \in L^2(dm)$. Then $g \in L^\mu(dm)$. So

$$f = f_1 f_2 = f_1 C_{E_2} g',\ g' \in L^2(dm) = C_E g \cdot g' \in C_E L^p(dm)\,.$$

This shows $\mathscr{N} \subset C_E L^p(dm)$. The reverse inclusion is immediate from the invariance of $\mathscr{N}$. Hence $\mathscr{N} = C_E L^p(dm)$ in this case.

Case (ii). $2 < p \leqq \infty$:

Let $\mathscr{N}' = \{f \mid f \in L^{p'}, f \perp \mathscr{N}\}$ where $(1/p') + (1/p) = 1$. Then $\mathscr{N}'$ is a doubly invariant subspace of $L^{p'}$ and $1 \leqq p' < 2$. Hence $\mathscr{N}' = C_{E'} L^{p'}$ for some $E' \subset X$. Then $\mathscr{N} = C_E L^p$ where $E = X - E'$.

References

1. J. Dixmier, *Les Algèbres D'operateurs Dans L'espace Hilbertien* (Algèbres De Von Neumann) Gauthier Villars, Paris, 1957.

2. H. Helson and D. Lowdenslager, *Invariant subspaces,* Proc. Int. Symp. on linear spaces, Jerusalem, 1960, 251-262.

3. P. D. Lax, *Translation Invariant spaces,* Proc. Int. Symp. on linear spaces, Jerusalem, 1960, 299-306.

PANJAB UNIVERSITY, CHANDIGARH, INDIA
UNIVERSITY OF CALIFORNIA, BERKELEY U.S.A.

ON THE EXTENSIONS OF LATTICE-ORDERED GROUPS

J. ROGER TELLER

1. Introduction. Throughout this paper $A = 0, a, b, \cdots, \Delta = \theta, \alpha, \beta, \cdots$ and G will be abelian partially ordered groups (p.o. groups). G is a *p.o. extension* of A by Δ if there is an order preserving homomorphism (o-homomorphisn) π of G onto Δ with kernel A such that π induces an o-isomorphism of G/A with Δ, (i.e. $\pi(g) > \theta$ implies $g + A$ contains a positive element). If A and Δ are lattice ordered groups (l-groups) then G is an *l-extension* if G is an l-group, π is an l-homomorphism and π induces an l-isomorphism between G/A and Δ. In this case A is an l-ideal of G.

If G is a p.o. extension of A by Δ then for each $\alpha \in \Delta$ choose $r(\alpha) \in G$ such that $\pi(r(\alpha)) = \alpha$ and $r(\theta) = 0$. Define

$$f(\alpha, \beta) = -r(\alpha + \beta) + r(\alpha) + r(\beta) \quad \text{for all } \alpha, \beta \in \Delta$$

and

$$Q_\alpha = \{a \in A \mid r(\alpha) + a \geqq 0\} \quad \text{for } \alpha \in \Delta^+ = \{\delta \in \Delta \mid \delta \geqq \theta\} .$$

Then the following conditions are satisfied for all α, β, γ in Δ.

(i) $f(\alpha, \beta) = f(\beta, \alpha)$

(ii) $f(\alpha, \theta) = f(\theta, \alpha) = 0$

(iii) $f(\alpha, \beta) + f(\alpha + \beta, \gamma) = f(\alpha, \beta + \gamma) + f(\beta, \gamma)$.

Moreover, for $\alpha, \beta \in \Delta^+$ we have

(iv) $Q_\alpha \neq \phi$

(v) $Q_\alpha + Q_\beta + f(\alpha, \beta) \subseteqq Q_{\alpha+\beta}$

(vi) $Q_\theta = A^+$.

Conditions (iv)–(vi) are due to L. Fuchs and can be derived from the results in [5].

Now if $\bar{G} = A \times \Delta$ and we define $(a, \alpha) + (b, \beta) = (a + b + f(\alpha, \beta), \alpha + \beta)$ and (a, α) positive if $\alpha \in \Delta^+$ and $a \in Q_\alpha$, then the mapping $(a, \alpha) \to r(\alpha) + a$ is an o-isomorphism of $\bar{G}$ onto G. In what follows we usually identify G and $\bar{G}$.

Conversely, if we are given $A, \Delta, f: \Delta \times \Delta \to A$ and $Q: \Delta^+ \to$ {subsets of A} such that f and Q satisfy (i)–(vi) then $\bar{G}$ is a p.o. extension of A by Δ and the mapping $(a, \alpha) \to \alpha$ is the corresponding o-homomorphism.

Two p.o. extensions $G = (A, \Delta, f, Q)$ and $G' = (A, \Delta, f', Q')$ are *o-equivalent* if there is a function $t: \Delta \to A$ such that

Received July 25, 1963. This research was supported in part by grant No. 21447 from the National Science Foundation and represents a portion of the author's dissertation. The author wishes to express his appreciation to Professor L. Fuchs, who suggested the problem, and Professor P. F. Conrad for their help in preparing this paper.

$$f'(\alpha, \beta) = f(\alpha, \beta) - t(\alpha + \beta) + t(\alpha) + t(\beta)$$

and

$$Q'_\alpha = -t(\alpha) + Q_\alpha .$$

This is equivalent to the fact that there exists an o-isomorphism of G onto G' that induces the identity on A and $G/A = \Delta$.

In Theorem 1 we give necessary and sufficient conditions that a p.o. extension $G = (A, \Delta, f, Q)$ be an l-extension. If G is an l-extension such that for each $\alpha \in \Delta^+$, Q_α is a principal dual ideal, that is, generated by a single element, then Lemma 2.2 shows G is o-equivalent to the cardinal sum $A \boxplus \Delta$. We show in Lemma 2.3, if A is a lexicographic extension of an l-ideal B (notation: $A = \langle B \rangle$) then for each $\alpha \in \Delta^+$, $Q_\alpha = A$ or Q_α is a principal dual ideal. Theorem 2 shows that if G is an l-extension of $A = \langle B \rangle$ then G contains an l-ideal $H \cong A \boxplus J$, $J \subseteqq \Delta$ and G is an l-extension of H by the ordered group (o-group) Δ/J. In addition if Δ is an o-group then $G = \langle A \boxplus J \rangle$.

Theorem 3 gives a method of constructing l-extensions from an abelian extension $G = (A, \Delta, f)$ that depends only on the cardinal summands of A.

In § 4 we use the above to investigate those l-extensions of an l-group A with a finite basis. We show that to an o-equivalence every l-extension of such an l-group A by an l-group Δ is determined by a meet-preserving homomorphism of the semigroup Δ^+ to the semigroup of all cardinal summands of A such that $f(\alpha, \beta) \in H_{\alpha+\beta}$.

2. Extensions of l-groups.

A subset Q of A is a *dual ideal* if $a \in Q$ and $b \geqq a$ implies $b \in Q$.

LEMMA 2.1. *If A is an l-group and $Q \subseteqq A$ is a dual ideal that satisfies*

(*) $\quad Q \cap (b + A^+)$ *has a smallest element for all* $b \in A$,

then Q is a sublattice of A. Thus Q is a lattice dual ideal.

Proof. Let $a, b \in Q$, then $a \vee b \in Q$ since Q is a dual ideal. Also, $a, b \in Q \cap [(a \wedge b) + A^+]$ so by (*) there is an element $x \in Q \cap [(a \wedge b) + A^+]$ such that $x \leqq a$ and $x \leqq b$. Hence, $x \leqq a \wedge b$ so $a \wedge b \in Q$ and Q is a sublattice of A as desired.

If E is a subset of A then the dual ideal generated by E (notation: $DI(E)$) is $\{x \in A \mid x \geqq y$ for some $y \in E\}$. If a dual ideal is generated by a single element we say the dual ideal is *principal*.

THEOREM 1. *Suppose A and Δ are l-groups and $G = (A, \Delta, f, Q)$ is a p.o.-extension of A by Δ. Then G is an l-extension if and only if*

(1) *if* $\alpha \wedge \beta = \theta$ *then* $Q_\alpha \cap [Q_\beta + b + f(\alpha - \beta, \beta)]$ *has a smallest element for all* $b \in A$,

and

(2) $$Q_\alpha + Q_\beta + f(\alpha, \beta) = Q_{\alpha+\beta} \textit{ for } \alpha, \beta \in \Delta^+.$$

Proof. Let G be an l-extension. Suppose $b \in A$ and $\alpha, \beta \in \Delta^+$ are such that $\alpha \wedge \beta = \theta$. Let $\gamma = \alpha - \beta$. For $a \in A$, the mapping of $(a, \alpha) \to \alpha$ is an l-homomorphism so $(b, \gamma) \vee (0, \theta) = (d, \alpha)$ where $d \in A$. Now $(d, \alpha) \geqq (0, \theta)$ implies $d \in Q_\alpha$ and $(d, \alpha) \geqq (b, \gamma)$ implies $(0, \theta) \leqq (d, \alpha) - (b, \gamma) = [d - b - f(\gamma, \beta), \beta]$ so $d - b - f(\gamma, \beta) \in Q_\beta$. Hence, $d \in Q_\alpha \cap [Q_\beta + b + f(\alpha - \beta, \beta)]$. If $c \in Q_\alpha \cap [Q_\beta + b + f(\alpha - \beta, \beta)]$ then a similar argument shows $(c, \alpha) \geqq (b, \gamma)$ and $(c, \alpha) \geqq (0, \theta)$. Hence, $(c, \alpha) \geqq (d, \alpha)$ and $c \geqq d$. Therefore, d is the smallest element in $Q_\alpha \cap [Q_\beta + b + f(\alpha - \beta, \beta)]$ and (1) holds.

To show (2) let $\alpha, \beta \in \Delta^+$. If either $\alpha = \theta$ or $\beta = \theta$ then (2) is trivial, so suppose $\alpha > \theta$ and $\beta > \theta$. Since G is a p.o.-extension we have $Q_\alpha + Q_\beta + f(\alpha, \beta) \subseteqq Q_{\alpha+\beta}$. For the reverse containment, let $x \in Q_{\alpha+\beta}$, $y \in Q_\alpha$, $b = x - y - f(\alpha, \beta)$ and $(a, \beta) = (b, \beta) \vee (0, \theta)$. Now $(c, \alpha + \beta) \geqq (0, \theta)$ if and only if $c \in Q_{\alpha+\beta}$; $(c, \alpha + \beta) \geqq (b, \beta)$ if and only if $c \in Q_\alpha + b + f(\alpha, \beta)$. On the other hand, since $(a, \beta) = (b, \beta) \vee (0, \theta)$, $c \in Q_{\alpha+\beta} \cap [Q_\alpha + b + f(\alpha, \beta)]$ if and only if $c \in Q_\alpha + a + f(\alpha, \beta)$. Hence $Q_{\alpha+\beta} \cap [Q_\alpha + b + f(\alpha, \beta)] = Q_\alpha + a + f(\alpha, \beta)$ and by (1) a is the smallest element in $Q_\beta \cap (Q_\theta + b)$. Therefore,

$$[Q_\alpha + b + f(\alpha, \beta)] \cap Q_{\alpha+\beta}$$
$$= Q_\alpha + f(\alpha, \beta) + [Q_\beta \cap (Q_\theta + b)] \subseteqq Q_\alpha + f(\alpha, \beta) + Q_\beta \,.$$

By the choice of b, $x \in [Q_\alpha + b + f(\alpha, \beta)] \cap Q_{\alpha+\beta}$ and $Q_\alpha + Q_\beta + f(\alpha, \beta) = Q_{\alpha+\beta}$.

For the sufficiency assume (1) and (2) hold and suppose $(b, \beta) \in G$ and that (b, β) is not comparable with $(0, \theta)$. Let c be the smallest element in $Q_{\beta\vee\theta} \cap [Q_{-(\beta\wedge\theta)} + b + f(\beta, -(\beta \wedge \theta))]$. Then $(c, \beta \vee \theta) \geqq (0, \theta)$ and (b, β). If $(a, \alpha) \geqq (b, \beta), (0, \theta)$ then $a \in Q_\alpha \cap [Q_{\alpha-\beta} + b + f(\alpha - \beta, \beta)]$. Condition (1) implies (*) so $Q_{\alpha-(\beta\vee\theta)}$ is a sublattice of A and from (2) we can derive the equality,

$$Q_\alpha \cap [Q_{\alpha-\beta} + b + f(\alpha - \beta, \beta)] = [Q_{\alpha-(\beta\vee\theta)} + f(\alpha - (\beta \vee \theta), \beta \vee \theta)]$$
$$+ \{Q_{\beta\vee\theta} \cap [Q_{-(\beta\wedge\theta)} + b + f(\beta, -(\beta \wedge \theta))]\} \,.$$

Since c was chosen as the smallest element we have $a \in Q_{\alpha-(\beta\vee\theta)} + f(\alpha - (\beta \vee \theta), \beta \vee \theta) + c$ and therefore $(a, \alpha) \geqq (c, \beta \vee \theta)$. Hence, $(c, \beta \vee \theta) = (b, \beta) \vee (0, \theta)$ and G is an l-extension of A by Δ. It can be shown that conditions (1) and (2) are equivalent to those given by L. Fuchs [5]. The entire proof was given so that this paper will be more self-contained.

An l-group G is a *cardinal sum* of l-ideals $A_1, A_2, \cdots, A_n$ (notation: $G = A_1 \boxplus \cdots \boxplus A_n$) if G is the direct sum (notation: $G = A_1 \oplus A_2 \oplus \cdots \oplus A_n$) of the A_i and if for $a_i \in A_i$, $a_1 + \cdots + a_n \geqq 0$ if and only if $a_i \geqq 0$ for $i = 1, \cdots, n$. It can be shown that a direct sum of l-ideals of an l-group is actually the cardinal sum. G is a *lexico-extension* of an l-group A (notation: $G = \langle A \rangle$) if A is an l-ideal of G, G/A is an o-group, and every positive element in G but not in A exceeds every element in A. In this case we note that if $a + A < b + A$ in G/A then each element of $b + A$ exceeds every element of $a + A$.

LEMMA 2.2. *Suppose G is an l-extension of A by Δ.*

(a) *If $Q_\alpha = A$ for all $\theta \neq \alpha \in \Delta^+$ then $G = \langle A \rangle$.*

(b) *If Q_α is a principal dual ideal for each $\alpha \in \Delta^+$ then G is o-equivalent to the cardinal sum, $A \boxplus \Delta$, of A and Δ.*

Proof. Let G be an l-extension of A by Δ.

(a) If $Q_\alpha = A$ for all $\theta \neq \alpha \in \Delta^+$, then every positive element of $G \backslash A$ exceeds every element of A. From (1) it follows that Δ is an o-group and therefore $G = \langle A \rangle$.

(b) If Q_α is a principal dual ideal for each $\alpha \in \Delta^+$, let x_α be the generator of Q_α. By (2) we have $x_\alpha + x_\beta + f(\alpha, \beta) = x_{\alpha+\beta}$. Let $H = A \boxplus \Delta$, then $H = (A, \Delta, f' \equiv 0, Q' \equiv A^+)$ is an l-extension of A by Δ. Define $t' \colon \Delta^+ \to A$ as $t'(\alpha) = x_\alpha$. Then t' induces a function $t \colon \Delta \to A$ and it follows that for $\alpha, \beta \in \Delta$

$$0 = f'(\alpha, \beta) = f(\alpha, \beta) - t(\alpha + \beta) + t(\alpha) + t(\beta)$$

and

$$A^+ = Q'_\alpha = -t(\alpha) + Q_\alpha \quad \text{for } \alpha \in \Delta^+ .$$

Hence G and H are o-equivalent l-extensions.

LEMMA 2.3. *Let $A = \langle B \rangle$, $A \neq B$ and $G = (A, \Delta, f, Q)$ be an l-extension. Then for $\alpha \in \Delta^+$ either $Q_\alpha = A$ or Q_α is a principal dual ideal.*

Proof. If A is an o-group, $\alpha \in \Delta^+$ and $Q_\alpha \neq A$ then there is $b \in A$ such that $b < a$ for all $a \in Q_\alpha$. Hence, $(b, \alpha) \vee (0, \theta) = (c, \alpha)$ implies c is the smallest element in Q_α and therefore Q_α is a principal dual ideal.

If A is not an o-group then $B \subset A$ and A/B is an o-group. Suppose $\alpha \in \Delta^+$ and $Q_\alpha \neq A$, then there is $0 > b \in A \backslash B$ such that $b + B \neq x + B$ for all $x \in Q_\alpha$. For suppose for each $0 > b \in A \backslash B$ there is an $x \in Q_\alpha$ such that $b + B = x + B$, then $b + h \in Q_\alpha$ for some $h \in B$. Now for any $c \in A$ there is $0 > a \in A \backslash B$ such that $a + B < c + B$ so $c > a + h$ which implies $c \in Q_\alpha$. Thus $Q_\alpha = A$, a contradiction.

Now $Q_\alpha \cap (b + Q_\theta)$ must have a smallest element so it suffices to show $Q_\alpha \subseteqq b + Q_\theta$. To this end let $x \in Q_\alpha$. If $x + B \leqq b + B$ then either $x + B < b + B$ which implies $x < b$ and $b \in Q_\alpha$ or $x + B = b + B$. Both cases lead to contradictions so $x + B > b + B$ which implies $x > b$ and $x \in b + Q_\theta$. The proof is complete.

COROLLARY 2.1. *If* $A = \langle B \rangle$ *then* (1) *may be replaced by*

(1′) *If* $\alpha, \beta \in \Delta^+$ *and* $\alpha \wedge \beta = \theta$ *then either* Q_α *and* Q_β *are principal dual ideals or* Q_α *is principal and* $Q_\beta = A$.

Proof. If G is an l-extension and $\alpha, \beta \in \Delta^+$ such that $\alpha \wedge \beta = \theta$ then (1) implies $Q_\alpha \cap Q_\beta$ must have a smallest element and (1′) follows from Lemma 2.3. Conversely, if x is the smallest element in Q_α, y the smallest in Q_β and $b \in A$ then $x \vee (y + b + f(\alpha - \beta, \beta)$ is the smallest in $Q_\alpha \cap [Q_\beta + b + f(\alpha - \beta, \beta)]$. If $Q_\beta = A$ then x is the smallest and if $Q_\alpha = A$, $y + b + f(\alpha - \beta, \beta)$ is the smallest.

From the above it follows that if $A = \langle B \rangle$ and Δ is an o-group then (1) may be replaced by

(1″) For each $\alpha \in \Delta^+$, $Q_\alpha = A$ or Q_α is a principal dual ideal.

From (2) of Theorem 1 we have: The only l-extensions of $A = \langle B \rangle$ by an Archimedean o-group Δ are o-isomorphic to the cardinal extension or the lexico-extension.

THEOREM 2. *Let* $A = \langle B \rangle$ *and* Δ *be l-groups and* $G = (A, \Delta, f, Q)$ *be an l-extension. Then* G *contains an l-ideal* H *which is o-isomorphic to* $A \boxplus J$, $J \subseteqq \Delta$, *and* G *is an l-extension of* H *by the o-group* Δ/J.

Proof. By Lemma 2.3 either $Q_\alpha = A$ or Q_α is principal for all $\alpha \in \Delta^+$. Let $J^+ = \{\alpha \in \Delta^+ \mid Q_\alpha \neq A\}$. Then by (2) of Theorem 1, J^+ is a convex subsemigroup of Δ^+. Let J be the l-ideal of Δ generated by J^+ and let $H = (A, J, f', Q')$ where $f' = f \mid (J \times J)$ and $Q'_\alpha = Q_\alpha$, $\alpha \in J^+$. Then H is an l-ideal of G and Q'_α is a principal dual ideal for all $\alpha \in J^+$. Therefore by Lemma 2.2, we have H o-isomorphic to $A \boxplus J$.

By way of contradiction, if Δ/J is not an o-group then there are $X, Y \in (\Delta/J)^+$ such that $X \wedge Y = J$. Let $X = \alpha + J$, $Y = \beta + J$ then $X \wedge Y = (\alpha + J) \wedge (\beta + J) = (\alpha \wedge \beta) + J = J$ so $\alpha \wedge \beta \in J$. Now $\alpha = (\alpha \wedge \beta) + \gamma$, $\beta = (\alpha \wedge \beta) + \delta$ where $\gamma \wedge \delta = \theta$ and $\gamma, \delta \notin J$, hence $Q_\gamma = A = Q_\delta$. This contradicts Corollary 2.1. Thus Δ/J is an o-group.

Finally, the natural mappings induce an o-isomorphism of G/H onto Δ/J. Hence, G is an l-extension of H by the o-group Δ/J.

We note that if $\alpha \in \Delta^+ \backslash J^+$ then $Q_\alpha = A$ so if $0 < g \in G \backslash H$ then $g > a$ for all $a \in A$.

COROLLARY 2.2. *If Δ is an o-group and $G = (A, \Delta, f, Q)$ is an l-extension then $G = \langle A \boxplus J \rangle$.*

Proof. If Δ is an o-group then $\Delta = \langle J \rangle$. The corollary follows from the results of Conrad [3, p 235] since $A \boxplus J$ contains all the nonunits of G.

We note that if G is an l-group with two disjoint elements but not three then G is an l-extension of an o-group by an o-group and hence we have the structure theorem of Conrad and Clifford [4] for the abelian case.

3. l-extensions with each Q_α generated by a coset of an l-ideal. Throughout this section we will consider those l-extensions $G = (A, \Delta, f, Q)$ where, for each $\alpha \in \Delta^+$, $Q_\alpha = DI(x_\alpha + H_\alpha)$, H_α an l-ideal of A.

LEMMA 3.1. *Suppose $G = (A, \Delta, f, Q)$ is an l-extension of the above type. Then there is an l-extension $G' = (A, \Delta, f', Q')$ o-equivalent to G with $Q'_\alpha = DI(H_\alpha)$ for each $\alpha \in \Delta^+$.*

Proof. If G is an l-extension and $Q_\alpha = DI(x_\alpha + H_\alpha)$ for each $\alpha \in \Delta^+$, then there is a mapping $t: \Delta^+ \longrightarrow A$ defined as $t'(\alpha) = x_\alpha$. Since each $\alpha \in \Delta$ has a unique representation $\alpha = \alpha^+ - \alpha^-$ where $\alpha^+ = \alpha \vee \theta$, $\alpha^- = -(\alpha \wedge \theta)$, we can extend t' to a mapping $t: \Delta \longrightarrow A$ by defining $t(\alpha) = t'(\alpha^+) - t'(\alpha^-)$.

Let $f'(\alpha, \beta) = f(\alpha, \beta) - t(\alpha + \beta) + t(\alpha) + t(\beta)$ and $Q'_\alpha = -t(\alpha) + Q_\alpha$. It is easily verified that f' and Q' satisfy conditions (i)–(vi) so $G' = (A, \Delta, f', Q')$ is a p.o. extension of A by Δ. From Theorem 1 it follows that G' is an l-extension. Clearly, G' is o-equivalent to G and $Q' = DI(H_\alpha)$.

For those l-extensions G of A by Δ with Q_α as above the question of o-equivalence leads to an investigation of the l-ideals of A. To show this we need the following.

LEMMA 3.2. *If A is an l-group, H and K l-ideals of A and $DI(y + H) = DI(z + K)$ then $y + H = z + K$ and $H = K$.*

Proof. Suppose $DI(y + H) = DI(z + K)$ where H and K are l-ideals of A. If $x = z - y$ then $DI(H) = DI(x + K)$. Since $H \subseteqq DI(x + K)$, $0 \in DI(x + K)$. If $0 \notin x + K$ then $0 > x + k$, $k \in K$ so $x + K$ contains a negative element. Since $DI(H)$ is a semigroup, $2(x + k) \in DI(x + K)$

so $2x + 2k \geqq x + 1, 1 \in K$. Hence, $x + (2k - 1) \geqq 0$. This is a contradiction since $x + K$ can contain no positive elements. Thus $0 \in x + K$ and $x \in K$. Moreover, we have $DI(H) = DI(K)$ which implies $H = K$. For if $H \neq K$ then, without loss of generality, there is $0 > h \in H \backslash K$. But $h \in DI(K)$ so $h > k \in K$. Hence, $0 > h > k$, and by convexity $h \in K$, a contradiction. Thus, $H = x + K = z - y + K$ and $y + H = z + K$.

Now if $G = (A, \Delta, f, Q)$ and $G' = (A, \Delta, f', Q')$ are two l-extensions with Q_α and Q'_α generated by l-ideals H_α and H'_α of A, then G and G' are o-equivalent if and only if there is a function $t: \Delta \rightarrow A$ such that

$$f'(\alpha, \beta) = f(\alpha, \beta) - t(\alpha + \beta) + t(\alpha) + t(\beta)$$
$$H'_\alpha = H_\alpha \text{ and } t(\alpha) \in H'_\alpha .$$

The question at this point is which l-extensions will have Q_α generated by a coset of an l-ideal. We give a partial answer to this question in the next section.

We complete this section by giving a method for the construction of l-extensions of l-groups.

THEOREM 3. *Suppose A and Δ are l-groups and $G = (A, \Delta, f)$ is an abelian extension of A by Δ. For each $\alpha \in \Delta^+$, let H_α be a cardinal summand of A such that*

(1*) *if $\alpha \wedge \beta = \theta$ then $H_\alpha \cap H_\beta = 0$*

(2*) *$H_\alpha + H_\beta = H_{\alpha+\beta}$ and $f(\alpha, \beta) \in H_{\alpha+\beta}$.*

If $Q_\alpha = DI(H_\alpha)$ then $G = (A, \Delta, f, Q)$ is an l-extension of A by Δ.

Proof. Clearly (iv) is satisfied and for any $\alpha \in \Delta^+$, (2*) implies $H_\theta \subseteqq H_\alpha$. From (1*) it follows that $H_\theta = 0$. Thus $Q_\theta = A^+$ and (vi) is satisfied. Moreover, from (2*) we have $DI(H_\alpha + H_\beta + f(\alpha, \beta)) = DI(H_{\alpha+\beta})$ so $DI(H_\alpha) + DI(H_\beta) + f(\alpha, \beta) = DI(H_{\alpha+\beta})$ and (2) of Theorem 1 holds.

If $\alpha \wedge \beta = \theta$ then $H_\alpha \cap H_\beta = 0$ so $H_{\alpha+\beta} = H_\alpha \oplus H_\beta$ and since H_α and H_β are l-ideals we have $H_{\alpha+\beta} = H_\alpha \boxplus H_\beta$. Since $H_{\alpha+\beta}$ is a cardinal summand we conclude $A = H_{\alpha+\beta} \boxplus D = H_\alpha \boxplus H_\beta \boxplus D$ where D is an l-ideal of A. Suppose $b \in A$ and $b + f(\alpha - \beta, \beta) = (a_1, a_2, a_3)$ where $a_1 \in H_\alpha, a_2 \in H_\beta$ and $a_3 \in D$. We show $(a_1, 0, a_3 \vee 0)$ is the smallest element in

$$Q_\alpha \cap (b + f(\alpha - \beta, \beta) + Q_\beta) = DI(H_\alpha) \cap DI(b + f(\alpha - \beta, \beta) + H_\beta) .$$

Now $(a_1, 0, a_3 \vee 0) \geqq (a_1, 0, 0)$ so $(a_1, 0, a_3 \vee 0) \in DI(H_\alpha)$. Also $(a_1, 0, a_3) = (a_1, 0, a_3) = (a_1, a_2, a_3) - (0, a_2, 0)$ so $(a_1, 0, a_3) \in b + f(\alpha - \beta, \beta) + H_\beta$ and $(a_1, 0, a_3 \vee 0) \in DI(b + f(\alpha - \beta, \beta) + H_\beta)$. If

$$(u, v, w) \in DI(H_\alpha) \cap DI(b + f(\alpha - \beta, \beta) + H_\beta)$$

then $u \geqq h_\alpha \in H_\alpha, v \geqq 0$ and $w \geqq 0$. Also $u \geqq a_1, v \geqq a_2 + h_\beta$ where

$h_\beta \in H_\beta$ and $w \geqq a_3$. Hence, $(u, v, w) \geqq (a_1, 0, a_3 \vee 0)$ and $(a_1, 0, a_3 \vee 0)$ is the smallest element in $Q_\alpha \cap (b + f(\alpha - \beta, \beta) + Q_\beta)$. Thus G is an l-extension of A by Δ.

We note that, since any two representations of an l-group as a cardinal sum have a common refinement, the cardinal summands of an l-group form an additive semigroup closed with respect to intersection. That is, if $H = A \boxplus A'$ and $H = B \boxplus B'$ then $A = (A \cap B) \boxplus (A \cap B')$, $A' = (A' \cap B) \boxplus (A' \cap B')$ and $B = (A \cap B) \boxplus (A' \cap B)$. Thus $H = A \boxplus A' = (A + B) \boxplus (A' \cap B')$. Hence, $A + B$ is a cardinal summand of G.

4. Extensions of l-groups with a finite basis. An element g of an l-group G is *basic* if $0 < g$ and $\{x \in G \mid 0 < x \leqq g\}$ is ordered. A subset S of G is a *basis* for G if S is a maximum set of disjoint elements and each $g \in S$ is basic. Conrad [2] has shown that an l-group A with a finite basis of n elements is a lexico-sum of n ordered subgroups. In particular, A is the cardinal sum of two l-groups each with a basis of fewer than n elements, or A is a lexico-extension of such an l-group. In this section we are concerned with l-extensions of l-groups with finite bases.

LEMMA 4.1. *Suppose A has a finite basis and $G = (A, \Delta, f, Q)$ is an l-extension of A. Then for $\alpha \in \Delta^+$, $Q_\alpha = DI(x_\alpha + H_\alpha)$ where H_α is an l-ideal of A.*

Proof. Let A have a basis of n elements. The proof is by induction on n.

It follows from Lemma 2.3 that we need only consider $A = B \boxplus C$ and if $n = 1$ then $H_\alpha = A$ or $H_\alpha = 0$.

So suppose the theorem is true for all l-groups with a basis of fewer than n elements. Let $\varphi: A \to B$ and $\psi: A \to C$ be the projections. Now B has a basis of fewer than n elements and $G' = (B, \Delta, \varphi f, \varphi Q)$ is an l-extension of B so by induction $\varphi Q_\alpha = DI(x + M)$ where $x \in B$ and M is an l-ideal of B. Similarly, $\psi Q_\alpha = DI(y + N)$ where $y \in C$ and N is an l-ideal of C. Since Q_α is a sublattice of A, a straight forward argument shows $Q_\alpha = DI((x + y) + (M + N))$ and $M + N$ is an l-ideal of A. The proof is complete.

The following theorem shows that for an l-group A with a finite basis every l-extension G of A by an l-group Δ is o-equivalent to an l-extension constructed by the method described in Theorem 3. That is, to an o-equivalence, every such l-extension is determined by a meet-preserving homomorphism from the semigroup Δ^+ to the semigroup of all cardinal summands of A such that $f(\alpha, \beta) \in H_{\alpha+\beta}$.

In what follows we may, by Lemmas 3.1 and 4.1, assume for each $\alpha \in \Delta^+$ that $Q_\alpha = DI(H_\alpha)$.

THEOREM 4. *If A has a finite basis and $G = (A, \Delta, f, Q)$ is an l-extension of A by an l-group Δ then, for $\alpha, \beta \in \Delta^+$*

(a) *if $\alpha \wedge \beta = \theta$ then $H_\alpha \cap H_\beta = 0$*

(b) *$H_\alpha + H_\beta = H_{\alpha+\beta}$ and $f(\alpha, \beta) \in H_{\alpha+\beta}$*

(c) *H_α is a cardinal summand of A.*

Proof. Let A have a finite basis of n elements and G be an l-extension. By (1) if $\alpha \wedge \beta = \theta$ then $Q_\alpha \cap Q_\beta$ must have a smallest element w. Since $0 \in Q_\alpha \cap Q_\beta$, $w \leqq 0$ and therefore $w \in H_\alpha \cap H_\beta$. If $H_\alpha \cap H_\beta \neq 0$ then there is $h \in H_\alpha \cap H_\beta$ such that $h < w$ and $h \in Q_\alpha \cap Q_\beta$, a contradiction. Thus (a) holds.

From (2) we have

$$DI(H_\alpha) + DI(H_\beta) + f(\alpha, \beta) = DI(H_{\alpha+\beta})$$

so

$$DI(H_\alpha + H_\beta + f(\alpha, \beta)) = DI(H_{\alpha+\beta}) .$$

Thus by Lemma 2.3, $H_\alpha + H_\beta = H_{\alpha+\beta}$ and $f(\alpha, \beta) \in H_{\alpha+\beta}$ and (b) holds.

Now if $A = \langle B \rangle$ then for each $\alpha \in \Delta^+$, $H_\alpha = 0$ or $H_\alpha = A$ and (c) follows in a trivial way. So suppose $A = B \boxplus C$ and (c) is true for all l-groups with a basis of fewer then n elements. If $\varphi: A \rightarrow B$ and $\psi: A \rightarrow C$ are the projections then $G' = (B, \Delta, \varphi f, \varphi Q)$ and $G'' = (C, \Delta, \psi f, \psi Q)$ are l-extensions where $\varphi Q_\alpha = DI(\varphi H_\alpha)$ and $\psi Q_\alpha = DI(\psi H_\alpha)$. Hence, by induction, φH_α is a cardinal summand of B and ψH_α is a cardinal summand of C and we have $A = B \boxplus C = \varphi H_\alpha \boxplus M \boxplus \psi H_\alpha \boxplus N = \varphi H_\alpha \boxplus \psi H_\alpha \boxplus M \boxplus N = H_\alpha \boxplus M \boxplus N$ where M is an l-ideal of B and N is an l-ideal of C.

Using the results of Conrad [3, p. 223] we conclude that the minimal cardinal summands of an l-group A with a finite basis are those l-ideals of A that are lexico-extensions and are not bounded in A.

Added in Proof. The results of this paper have been extended by the author to include central extensions G of an abelian l-group A by an arbitrary l-group Δ. For central extensions, Theorem 1 (1) reads: if $\alpha \wedge \beta = \theta$ then $Q_\alpha \cap [Q_\beta + b + f(\beta, \alpha - \beta)]$ has a smallest element for all $b \in A$. In Theorem 2, G/H is still o-isomorphic to the o-group Δ/J but G need not be a central extension of H by Δ/J. The remaining results are unchanged for central extensions.

REFERENCES

1. G. Birkhoff, *Lattice Theory*, Revised Edition, Amer. Math. Soc., Colloquium Publications **25** (1948).

2. P. Conrad, *The structure of a lattice-ordered group with a finite number of disjoint elements*, Michigan Math. J., **7** (1960), 171–180.

3. ———, *Some structure theorems for lattice-ordered groups*, Trans. Amer. Math.

Soc., **99** (1961), 212-240.
4. P. F. Conrad and A. H. Clifford, *Lattice-ordered groups having at most two disjoint elements*, Proc. Glasgow Math. Assn., **4** (1960), 111-113.
5. L. Fuchs, *The extensions of partially ordered groups*, Acta. Math. Acad. Sci. Hungar., **1** (1950), 118-124.

THE TULANE UNIVERSITY OF LOUISIANA

UNIMODULAR GROUP MATRICES WITH RATIONAL INTEGERS AS ELEMENTS

R. C. THOMPSON

1. Introduction. Let G be a finite group of order n with elements $g_1, g_2, \cdots, g_n$. Let

$$x_{g_i}, \qquad 1 \leqq i \leqq n \tag{1}$$

be variables in one-to-one correspondence with the elements of G. The $n \times n$ matrix

$$X = (x_{g_i g_j^{-1}})_{1 \leqq i, j \leqq n} \tag{2}$$

is called the group matrix for G. If numerical values are substituted for the variables (1) in X, we say X is a group matrix for G. In this paper we study group matrices which have rational integers as elements. Let A' denote the transpose of the matrix A. A generalized permutation matrix is a square matrix with only 0, 1, -1 as elements and having exactly one nonzero element in each row and in each column. A square matrix A is said to be unimodular if the determinant of A is ± 1. The result obtained in this paper is the following theorem.

THEOREM. *Let G be a finite solvable group. Let A be a unimodular matrix of rational integers such that $B = AA'$ is a group matrix for G. Then $A = A_1T$ where A_1 is a unimodular group matrix of rational integers for G and T is a generalized permutation matrix.*

This theorem has already been proved for cyclic groups in [1] and for abelian groups in [2]. The present proof is a modification of the proof in [2].

2. Proof of the theorem. Let

$$1 = H_0 \subset H_1 \subset H_2 \subset \cdots \subset H_{m-1} \subset H_m = G \tag{3}$$

be an ascending chain of subgroups of G, where each H_{i-1} is normal in H_i with cyclic factor group H_i/H_{i-1} of order n_i, $1 \leqq i \leqq m$. We let $n_0 = 1$, so that H_i has order $n_0 n_1 \cdots n_i$. In order to simplify the proof we take the elements of G in a particular order. This will not affect the theorem as a reordering of the elements of G changes the group matrix X to PXP' for P a permutation matrix. Thus let

Received August 20, 1963.

H_i be generated by the elements of H_{i-1} and an element a_i such that the coset a_iH_{i-1} has order n_i. By induction we define column vectors V_i of the elements of H_i. We let

$$V_0 = (1) \tag{4}$$

be the one row column vector whose only element is the identity of G. Suppose

$$V_{i-1} = (h_1, h_2, \cdots, h_t)' \tag{5}$$

with

$$t = n_0 n_1 \cdots n_{i-1} , \tag{6}$$

has been defined, where $h_1, h_2, \cdots, h_t$ are the ordered elements of H_{i-1}. For any $g \in G$ let

$$g V_{i-1} = (gh_1, gh_2, \cdots, gh_t)' ,$$
$$V_{i-1}g = (h_1g, h_2g, \cdots, h_tg)' .$$

Then define V_i to be the column vector

$$V_i = \begin{bmatrix} V_{i-1} \\ a_i V_{i-1} \\ a_i^2 V_{i-1} \\ \cdots \\ a_i^{n_i-1} V_{i-1} \end{bmatrix} . \tag{7}$$

For an arbitrary finite group G with ordered elements $g_1, g_2, \cdots, g_n$ we define the *left regular representation* of G by the matrix equations

$$(gg_1, gg_2, \cdots, gg_n) = (g_1, g_2, \cdots, g_n)P^L(g) , \qquad g \in G .$$

Here $P^L(g)$ is a permutation matrix depending on the element $g \in G$. It is straightforward to check that the matrix X of (2) is given by

$$X = \sum_{g \in G} x_g P^L(g) .$$

The set of all $P^L(g)$ for $g \in G$ is denoted by $L(G)$.

We define the *right regular representation* of G by

$$(g_1g, g_2g, \cdots, g_ng)' = P(g)(g_1, g_2, \cdots, g_n)' , \qquad g \in G .$$

The set of all permutation matrices $P(g)$ for $g \in G$ is denoted by $R(G)$.

The group ring of the left (right) regular representation is the set of all linear combinations of the $P^L(g)$ $(P(g))$ for $g \in G$, and is denoted by $L^*(G)$ $(R^*(G))$. Thus the matrix (2) is the typical member of $L^*(G)$. The following two known facts are vital for the proof of our theorem:

(i) any matrix in $L^*(G)$ commutes with any matrix in $R^*(G)$;

(ii) any matrix that commutes with all the matrices in $R(G)$ is a member of $L^*(G)$.

NOTATION. We let diag $(X_1, X_2, \cdots, X_k)_k$ denote the direct sum of the square matrices $X_1, X_2, \cdots, X_k$:

$$\operatorname{diag}(X_1, X_2, \cdots, X_k)_k = \begin{bmatrix} X_1 & 0 & 0 & \cdots & 0 \\ 0 & X_2 & 0 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdots & 0 \\ 0 & 0 & 0 & \cdots & X_k \end{bmatrix}.$$

We set $[X_1]_1 = X_1$. If $k > 1$ and $X_1, X_2, \cdots, X_k$ are square matrices of the same size, we set

$$[X_1, X_2, \cdots, X_k]_k = \begin{bmatrix} 0 & X_1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & X_2 & 0 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdots & \cdot \\ 0 & 0 & 0 & 0 & \cdots & X_{k-1} \\ X_k & 0 & 0 & 0 & \cdots & 0 \end{bmatrix}.$$

We construct certain of the matrices in $R(G)$, where now the elements of G are ordered according to (4), (5), (6), (7). Let i be fixed, $1 \leqq i \leqq m$. Since H_{i-1} is normal in H_i, $V_{i-1}a_i = a_i P_{i-1}(a_i) V_{i-1}$ where $P_{i-1}(a_i)$ is a $t \times t$ permutation matrix (t as in (6)). Then, since

$$(8) \qquad a_i^{n_i} \in H_{i-1} ,$$

and because of (7), $V_i a_i = P_i(a_i) V_i$, where $P_i(a_i)$ is permutation matrix with the structure

$$(9) \qquad P_i(a_i) = [P_{i-1}(a_i), P_{i-1}(a_i), \cdots, P_{i-1}(a_i), \bar{P}_{i-1}(a_i)]_{n_i} .$$

In (9), $\bar{P}_{i-1}(a_i)$ is another $t \times t$ permutation matrix.

Because of (7), we also have for any $g \in H_{i-1}$, that $V_i g = P_i(g) V_i$, where the permutation matrix $P_i(g)$ has the structure

$$(10) \qquad P_i(g) = \operatorname{diag}(P_{i-1}(g), P_{i-1}(g), \cdots, P_{i-1}(g))_{n_i} , \qquad g \in H_{i-1} .$$

In (10), $P_i(g)$ is a block scalar matrix. The diagonal blocks $P_{i-1}(g)$ have dimensions $t \times t$. Furthermore, as g runs over the elements of H_{i-1}, $P_{i-1}(g)$ runs over all the matrices of $R(H_{i-1})$. Since H_i is generated by H_{i-1} and a_i, the matrices $P_i(g)$ for $g \in H_{i-1}$ and $P_i(a_i)$ generate $R(H_i)$.

Because of the ordering of the elements of G, the following block scalar matrices:

$$Q(g) = \operatorname{diag}(P_i(g), \cdots, P_i(g))_u \,, \qquad g \in H_{i-1} \text{ or } g = a_i \,, \tag{11}$$

$$u = n/tn_i \,, \tag{12}$$

are the matrices in $R(G)$ determined by the $g \in H_{i-1}$ and by $g = a_i$. Here $Q(g)$ is $n \times n$.

We now prove our theorem by the following induction argument. Suppose for a fixed i, $1 \leqq i \leqq m$, that $B = AA'$ and that

$$AQ(g) = Q(g)A \,, \qquad \text{for any } g \in H_{i-1} \,. \tag{13}$$

(In particular this is satisfied if $i = 1$ since then the only such $Q(g)$ is I_n, the $n \times n$ identity matrix.) We shall then show that a generalized permutation matrix T exists such that $B = (AT)(AT)'$ and such that $ATQ(g) = Q(g)AT$ for any $g \in H_{i-1}$ and for $g = a_i$, and so, in consequence, for any $g \in H_i$. Thus the induction will eventually yield a generalized permutation matrix T_1 such that $B = (AT_1)(AT_1)'$ and such that $AT_1Q(g) = Q(g)AT_1$ for any $g \in G$. It will now follow from (ii) that $AT_1 \in L^*(G)$, and the proof will be complete.

Hence assume $B = AA'$ where A satisfies (13). Partition

$$A = (A_{\alpha,\beta}) \,, \qquad 1 \leqq \alpha, \beta \leqq v = n_i u \,, \tag{14}$$

into blocks of dimensions $t \times t$. As $Q(g)$ for $g \in H_{i-1}$ is a block scalar matrix with the blocks $P_{i-1}(g)$ of $R(H_{i-1})$ on the main block diagonal, it follows from (ii) and (13) that each

$$A_{\alpha,\beta} \in L^*(H_{i-1}) \,, \qquad 1 \leqq \alpha, \beta \leqq v \,. \tag{15}$$

Since $B \in L^*(G)$, $BQ(a_i) = Q(a_i)B$ so that if

$$M = A^{-1}Q(a_i)A \,, \tag{16}$$

then,

$$MM' = I_n \,. \tag{17}$$

As A is unimodular the elements of M are integers. Hence (17) implies that M is a generalized permutation matrix. Partition A, A^{-1}, $Q(a_i)$, and M into $t \times t$ blocks. As each block of A lies in $L^*(H_{i-1})$ and as A^{-1} is a polynomial in A, each of the $t \times t$ blocks of A, of A^{-1}, and of $Q(a_i)$ is a linear combination of a finite number of $t \times t$ permutation matrices. Therefore each $t \times t$ block of M is a linear combination of a finite number of $t \times t$ permutation matrices. A permutation matrix is *doubly stochastic* in the sense that the sums across each row and down each column all have a common value.

As linear combinations of matrices doubly stochastic in this sense remain doubly stochastic, each $t \times t$ block of M is doubly stochastic. Let M_1 be a typical $t \times t$ block in M. Since M is a generalized permutation matrix, M_1 contains at most one nonzero element in each of its rows and columns. As M_1 is doubly stochastic, it now follows that M_1, if it is not the zero matrix, is either a permutation matrix or the negative of a permutation matrix. Since M is a generalized permutation matrix, it follows that, after partitioning into $t \times t$ blocks, M is a "generalized permutation matrix" in that it has exactly one nonzero block in each of its block rows and in each of its block columns. Each nonzero block is $\pm$ a permutation matrix.

There exists a permutation matrix R consisting of $t \times t$ blocks which are either the $t \times t$ zero matrix or I_t such that $R'MR$ is a direct sum of cycles. That is, $R'MR = \operatorname{diag}(E_1, E_2, \cdots, E_r)_r$ where

$$E_\delta = [E_{\delta,1}, E_{\delta,2}, \cdots, E_{\delta,e_\delta}]_{e_\delta}\,, \qquad 1 \leqq \delta \leqq r\,. \tag{18}$$

Here each $E_{\delta,\omega}$ is $\pm$ a $t \times t$ permutation matrix.

Note that $RQ(g) = Q(g)R$ for any $g \in H_{i-1}$ since each such $Q(g)$ is block scalar when partitioned into $t \times t$ blocks. Thus

$$ARQ(g) = Q(g)AR\,, \qquad \text{for any } g \in H_{i-1}\,,$$

and

$$(AR)^{-1}Q(a_i)AR = R'MR$$

is a direct sum of $E_1, E_2, \cdots, E_r$. Thus if we change notation and replace AR with A and $R'MR$ with M, we have (13), (14), (15), (16), (18) and

$$M = \operatorname{diag}(E_1, E_2, \cdots, E_r)_r\,.$$

Our immediate goal is to prove that each e_δ is n_i and that $r = u$. Because of (8)

$$\begin{aligned} M^{n_i} &= A^{-1}Q(a_i^{n_i})A \\ &= A^{-1}Q(g)A \qquad && \text{for some } g \in H_{i-1}\,, \\ &= Q(g) && \text{by (13)}\,. \end{aligned}$$

Hence each cycle E_δ of M has the property that

$$E_\delta^{n_i}$$

is block scalar. This is not possible if $e_\delta > n_i$. Hence each $e_\delta \leqq n_i$.

Counting rows in M we get $t(e_1 + e_2 + \cdots + e_r) = n$. If any $e_\delta < n_i$ we would have

(19) $$r > u\ .$$

Let $A_\alpha = (A_{\alpha,1}, A_{\alpha,2}, \cdots, A_{\alpha,v})$, $1 \leqq \alpha \leqq v$, be the block rows of A. For each fixed d such that $0 \leqq d < u$ it follows from (9), (11), and $Q(a_i)A = AM$ that

(20) $$P_{i-1}(a_i)A_{dn_i+k} = A_{dn_i+k-1}M\ , \qquad 2 \leqq k \leqq n_i\ .$$

Let $w_0 = 0$ and let $w_\delta = e_1 + e_2 + \cdots + e_\delta$ for $1 \leqq \delta \leqq r$. Then (20) implies than for $2 \leqq \mathrm{k} \leqq n_i$ and $0 \leqq \delta \leqq r-1$,

(21) $$\begin{aligned}&(A_{dn_i+k,w_\delta+1}, \cdots, A_{dn_i+k,w_\delta+1}) \\ &\quad = P_{i-1}(a_i)^{1-k}(A_{dn_i+1,w_\delta+1}, \cdots, A_{dn_i+1,w_\delta+1})E_{\delta+1}^{k-1}\ .\end{aligned}$$

For each fixed d, δ such that $0 \leqq d < u$, $0 \leqq \delta < r$, let $F_{d,\delta}$ be the submatrix of A containing the blocks $A_{\alpha,\beta}$ with $dn_i + 1 \leqq \alpha \leqq (d+1)n_i$ and $w_\delta + 1 \leqq \beta \leqq w_{\delta+1}$. Since each $A_{\alpha,\beta} \in L^*(H_{i-1})$, each row of a given $A_{\alpha,\beta}$ is a permutation of the first row of this $A_{\alpha,\beta}$. Since $P_{i-1}(a_i)$ and $E_{\delta+1}$ are generalized permutation matrices, this fact and (21) imply that each row of $F_{d,\delta}$ is a generalized permutation of the first row of $F_{d,\delta}$. Thus if we add all the columns of $F_{d,\delta}$ after the first to the first column of $F_{d,\delta}$ we produce a new matrix $\bar{F}_{d,\delta}$ in which the integers in the first column of $\bar{F}_{d,\delta}$ are all equal, modulo 2. Next add the first row of $\bar{F}_{d,\delta}$ to all the other rows of $\bar{F}_{d,\delta}$ to get a new matrix $\tilde{F}_{d,\delta}$. Then all the integers in the first column of $\tilde{F}_{d,\delta}$ below the top element are zero, modulo 2.

Now partition $A = (F_{d,\delta})$ into its blocks $F_{d,\delta}$. For each fixed $\delta, 0 \leqq \delta < r$, add to that column of A that intersects $F_{0,\delta}$ at the extreme left of $F_{0,\delta}$, all the other columns of A that intersect $F_{0,\delta}$. This produces a new matrix $\bar{A} = (\bar{F}_{d,\delta})$. For each fixed d, $0 \leqq d < u$, add the topmost row of $\bar{A}$ that intersects $\bar{F}_{d,0}$ to all the other rows of $\bar{A}$ that intersect $\bar{F}_{d,0}$. We get a new matrix $\tilde{A} = (\tilde{F}_{d,\delta})$. The r columns of $\tilde{A}$ that intersect $\tilde{F}_{0,\delta}$ at the extreme left of $\tilde{F}_{0,\delta}$, $0 \leqq \delta < r$, may now be regarded as vectors in a u dimensional vector space over the field of two elements. As $r > u$, these vectors are dependent and so $\tilde{A}$ (and hence A) is singular, modulo 2. This is a contradiction since the determinant of A is ± 1.

Consequently each $e_\delta = n_i$, $1 \leqq \delta \leqq r$, and $r = u$.

Now let $E_{p,q} = \varphi_{p,q}\bar{E}_{p,q}$ where $\varphi_{p,q} = \pm 1$ and $\bar{E}_{p,q}$ is a permutation matrix. Let δ be fixed, $1 \leqq \delta \leqq u$. Suppose that $P_{i-1}(a_i)$ has a one at position $(1, \omega)$ and let $\bar{E}_{\delta,1}$ have a one at position $(1, \mu)$. Let $K_{\delta,1}$ be the permutation matrix in $L(H_{i-1})$ with a one at position (μ, ω). ($K_{\delta,1}$ is the matrix in $L(H_{i-1})$ representing $h_\mu h_\omega^{-1}$; see (2) and (5).) Then $\tilde{E}_{\delta,1} = \bar{E}_{\delta,1}K_{\delta,1}$ has the same first row as $P_{i-1}(a_i)$. Similarly, by induction, we determine $K_{\delta,s}$ in $L(H_{i-1})$, $1 < s < n_i$, such that the permutation matrices

$$\tilde{E}_{\delta,s} = K'_{\delta,s-1}\bar{E}_{\delta,s}K_{\delta,s}\ , \qquad 1 < s < n_i\ ,$$

each have the same first row as $P_{i-1}(a_i)$. Then let

$$S_\delta = \operatorname{diag}\left(I_t,\ \varphi_{\delta,1}K_{\delta,1},\ \varphi_{\delta,1}\varphi_{\delta,2}K_{\delta,2},\ \cdots,\ \left(\prod_{j=1}^{n_i-1}\varphi_{\delta,j}\right)K_{\delta,n_i-1}\right)_{n_i}\ ,$$

and let $S = \operatorname{diag}(S_1, S_2, \cdots, S_u)_u$. Then

$$S'MS = \operatorname{diag}(\tilde{E}_1, \tilde{E}_2, \cdots, \tilde{E}_u)_u$$

where

$$(22) \qquad \tilde{E}_\delta = [\tilde{E}_{\delta,1}, \tilde{E}_{\delta,2}, \cdots, \tilde{E}_{\delta,n_i-1}, \pm\tilde{E}_{\delta,n_i}]_{n_i}\ , \qquad 1 \leqq \delta \leqq u\ .$$

In (22) each $\tilde{E}_{\delta,j}$, $1 \leqq j < n_i$, $1 \leqq \delta \leqq u$, is a permutation matrix with the same first row as $P_{i-1}(a_i)$ and each

$$\tilde{E}_{\delta,n_i}\ , \qquad 1 \leqq \delta \leqq u\ ,$$

is some unknown permutation matrix.

Now $SQ(g) = Q(g)S$ if $g \in H_{i-1}$ since S is block diagonal with its blocks in $L^*(H_{i-1})$ whereas $Q(g)$ for $g \in H_{i-1}$ is block scalar with its blocks in $R(H_{i-1})$. Thus if we change notation again and replace AS with A and $S'MS$ with M we retain the validity of (13) and (16) and now

$$(23) \qquad M = \operatorname{diag}(\tilde{E}_1, \tilde{E}_2, \cdots, \tilde{E}_u)_u\ .$$

Since for any $g \in H_{i-1}$, $a_i^{-1}ga_i = \bar{g} \in H_{i-1}$, it follows that for any $g \in H_{i-1}$ there exists a $\bar{g} \in H_{i-1}$ such that $Q(g)Q(a_i) = Q(a_1)Q(\bar{g})$. Hence, using (9), (10), and (11), we find

$$(24) \qquad P_{i-1}(g)P_{i-1}(a_i) = P_{i-1}(a_i)P_{i-1}(\bar{g})\ , \qquad g, \bar{g} \in H_{i-1}\ .$$

If we let $g \in H_{i-1}$ be such that $P_{i-1}(g)$ has a one at position $(1, \omega)$ then (24) says: row ω of $P_{i-1}(a_i)$ is determined in terms of row one of $P_{i-1}(a_i)$.

Now for $g \in H_{i-1}$:

$$\begin{aligned} Q(g)M &= Q(g)A^{-1}Q(a_i)A \\ &= A^{-1}Q(g)Q(a_i)A && \text{by (13)}\ , \\ &= A^{-1}Q(a_i)Q(\bar{g})A && \text{since } ga_i = a_i\bar{g}\ , \\ &= A^{-1}Q(a_i)AQ(\bar{g}) && \text{by (13)}\ , \\ &= MQ(\bar{g})\ . \end{aligned}$$

Hence, for fixed δ and j, $1 \leqq \delta \leqq u$, $1 \leqq j < n_i$, it now follows (using (10), (11), (22), and (23)) that

(25) $$P_{i-1}(g)\tilde{E}_{\delta,j} = \tilde{E}_{\delta,j}P_{i-1}(\bar{g}) , \qquad g, \bar{g} \in H_{i-1} .$$

As with (24), (25) determines each row of $\tilde{E}_{\delta,j}$ in terms of the first row of $\tilde{E}_{\delta,j}$. Consequently

(26) $$\tilde{E}_{\delta,j} = P_{i-1}(a_i) , \qquad 1 \leqq \delta \leqq u,\ 1 \leqq j < n_i .$$

We also have (8), hence

$$M^{n_i} = A^{-1}Q(a_i^{n_i})A = Q(a_i)^{n_i}$$

by (13). Hence, for each δ, $1 \leqq \delta \leqq u$,

(27) $$\tilde{E}_\delta^{n_i} = P_i(a_i)^{n_i} .$$

Each side of (27) is a block diagonal matrix. Equating the topmost diagonal blocks we get

$$\left[\prod_{j=1}^{n_i-1} \tilde{E}_{\delta,j}\right][\pm\tilde{E}_{\delta,n_i}] = P_{i-1}(a_i)^{n_i-1}\bar{P}_{i-1}(a_i) .$$

Hence, by (26),

$$\pm\tilde{E}_{\delta,n_i} = \bar{P}_{i-1}(a_i) , \qquad 1 \leqq \delta \leqq u .$$

We have now proved that $M = Q(a_i)$. Hence $Q(a_i)A = AQ(a_i)$. As indicated earlier, this is enough to complete the proof.

References

1. M. Newman and Olga Taussky, *On a generalization of the normal basis in abelian algebraic number fields*, Comm. Pure App. Math., **9** (1956), 85-91.

2. R. C. Thompson, *Normal matrices and the normal basis in abelian number fields*, Pacific J. Math., **12** (1962), 1115-1124.

THE UNIVERSITY OF BRITISH COLUMBIA,
VANCOUVER, CANADA

LEAST SQUARES AND INTERPOLATION IN ROOTS OF UNITY

J. L. WALSH AND A. SHARMA

We mention the Erdös-Turán theorem [2] that if $F(\theta)$ is a real continuous function with period 2π, and if $t_n(\theta)$ is the unique trigonometric polynomial of order n that coincides with $F(\theta)$ in $2n+1$ points equally spaced over an interval of length 2π, then $t_n(\theta)$ converges to $F(\theta)$ on that interval in the mean of second order. It is the purpose of the present note to prove the analogue in the complex domain, and to discuss some related remarks.

THEOREM 1. *Let the function $f(z)$ be analytic in D: $|z| < 1$, continuous in $D + C\,(C\colon |z| = 1)$, and let $p_n(z)$ be the polynomial of degree n coinciding with $f(z)$ in the $(n+1)$ st roots of unity. Then the sequence $p_n(z)$ converges to $f(z)$ on C in the mean of second order. Consequently we have*

$$\lim_{n\to\infty} p_n(z) = f(z) \qquad \textit{uniformly in } |z| \leqq r\,(<1)\,. \tag{1}$$

If we set

$$I_n = \int_C |f(z) - p_n(z)|^2 \, |dz|\,, \tag{2}$$

we have

$$p_n(z) \equiv \sum_{k=1}^{n+1} f(\omega^k) A_k(z)\,,$$

$$A_k(z) \equiv \frac{\omega^k(z^{n+1}-1)}{(n+1)(z-\omega^k)}\,, \quad \omega = e^{2\pi i/(n+1)}\,,$$

and shall show

$$\lim_{n\to=\infty} I_n = 0\,. \tag{3}$$

We introduce the notation

$$f(z) - t_n(z) \equiv \Delta(z)\,, \quad E_n = \max\,[|\,\Delta z\,|, z \text{ on } C]\,,$$

where $t_n(z)$ is the polynomial of degree n of best Tchebycheff approximation to $f(z)$ on C, and denote by $P_n(z)$ the polynomial of degree n that coincides with $\Delta(z)$ in the $(n+1)$ st roots of unity. Then we have $P_n(z) \equiv p_n(z) - t_n(z)$, whence

Received July 24, 1963. Research sponsored (in part) by U.S. Air Force, Office of Scientific Research. Abstract published in Notices of Amer. Math. Soc., vol. **10** (1963), p. 491.

$$I_n = \int_C |\Delta(z) - P_n(z)|^2 \, |dz|$$
$$\leqq 2\int_C |\Delta z|^2 \, |dz| + 2\int_C |P_n(z)|^2 \, |dz| = I_n + I_n'' .$$

There follow the relations $I_n' \leqq 4\pi E_n^2$,

$$I_n'' = 2\int_C \left| \sum_{k=1}^{n+1} \Delta(\omega^k) A_k(z) \right|^2 |dz|$$
$$\leqq 2 \sum_{k=1}^{n+1} \sum_{j=1}^{n+1} |\Delta(\omega^k)\bar{\Delta}(\omega^j)| \left| \int_C A_k(z)\bar{A}_j(z) \, |dz| \right|$$
$$\leqq 2E_n^2 \sum_{k=1}^{n+1} \sum_{j=1}^{n+1} \left| \int_C A_k(z)\bar{A}_j(z) \, |dz| \right| .$$

However, we have

$$A_k(z) \equiv \frac{\omega^k}{n+1}(z^n + \omega^k z^{n-1} + \cdots + \omega^{nk}) ,$$
$$\int_C A_k(z)\bar{A}_j(z) \, |dz| = \frac{2\pi\omega^{k-j}}{(n+1)^2}(1 + \omega^{k-j} + \omega^{2(k-j)} + \cdots + \omega^{n(k-j)})$$
$$= 2\pi\delta_{jk}/(n+1) ,$$

where δ_{jk} is the Kronecker δ. It is well known [4, Theorem 5, p. 36] that $E_n \to 0$ as $n \to \infty$, so (3) holds.

Equation (1) follows from (3) by the Cauchy integral formula

$$(4) \qquad [f(z) - p_n(z)]^2 = \frac{1}{2\pi i}\int_C \frac{[f(t) - p_n(t)]^2 \, dt}{t - z} , \quad |z| < 1 .$$

With the hypothesis of Theorem 1, the conclusion (1) is due to Fejér (1918). Theorem 1 is related to various other results concerning convergence of polynomials interpolating in roots of unity; for instance (Runge) if $f(z)$ in Theorem 1 is analytic in $|z| \leqq 1$, equation (1) holds uniformly in $|z| \leqq 1$. Further references to the subject are given by Curtiss [1].

There exist numerous other results, somewhat similar to Theorem 1, where now a sequence of polynomials $P_n(z, 1/z)$ of respective degrees n in z and $1/z$ converges on C in the second-order mean to a given function $f(z)$ defined merely on C. The function $f(z)$ can be expressed on C as $f(z) \equiv f_1(z) + f_2(z)$, where $f_1(z)$ is of the Hardy-Littlewood class H_2 and $f_2(z)$ of the analogous class G_2 for the region $|z| > 1$ (we suppose $f_2(\infty) = 0$; compare e.g. [4, § 6. 11]). Any function of class H_2 is orthogonal on C to any function of class G_2, so if we set $P_n(z, 1/z) \equiv p_n(z) + q_n(1/z)$, where $p_n(z)$ and $q_n(1/z)$ are polynomials of respective degrees n in their arguments, $q_n(0) = 0$, we have

$$(5) \qquad \lim_{n\to\infty} p_n(z) \equiv \frac{1}{2\pi i}\int_C \frac{f(t)dt}{t - z} \equiv f_1(z) \equiv \frac{1}{2\pi i}\int_C \frac{f_1(t)dt}{t - z}, \quad z \text{ interior to } C ,$$

$$(6) \qquad \lim_{n\to\infty} q_n(1/z) \equiv \frac{1}{2\pi i}\int_C \frac{f(t)dt}{t-z} \equiv f_2(z) \equiv \frac{1}{2\pi i}\int_C \frac{f_2(t)dt}{t-z}, \quad z \text{ exterior to } C,$$

with uniformity of approach for z on an arbitrary compact set in the respective regions. This remark concerning (5) and (6) applies for instance in the case of the Erdös-Turán theorem, where we set $F(\theta) \equiv f(e^{i\theta})$ and $t_n(\theta) \equiv p_n(e^{i\theta}, e^{-i\theta})$ on C.

A second remark concerning (5) and (6) is as follows. By the orthogonality relations we have for the second-order norms on C

$$\|f - P_n\|^2 = \|f_1 - p_n\|^2 + \|f_2 - q_n\|^2 .$$

Consequently the rapidity of convergence in the mean on C of p_n to f_1 and of q_n to f_2 is not less than that of P_n to f. If the positive numbers $\varepsilon_1, \varepsilon_2, \cdots$ are given and approach zero, there is a corresponding class K of functions $f(z)$ belonging to L_2 on C such that for each $f(z)$ there exist polynomials $P_n(z, 1/z)$ with

$$(7) \qquad \|f(z) - P_n(z, 1/z)\| = O(\varepsilon_n) ;$$

here the $P_n(z, 1/z)$ may be taken as the partial sums of the Fourier or Laurent development of $f(z)$ on C. It follows that every function $f(z)$ in K can be written $f(z) \equiv f_1(z) + f_2(z)$, with f_1 and f_2 in H_2 and G_2 respectively, where the partial sums $p_n(z)$ of the Taylor development of $f_1(z)$ satisfy

$$(8) \qquad \|f_1(z) - p_n(z)\| = O(\varepsilon_n)$$

and the partial sums $q_n(1/z)$ of the Laurent development of $f_2(z)$ satisfy

$$(9) \qquad \|f_2(z) - q_n(1/z)\| = O(\varepsilon_n) .$$

Thus f_1 and f_2 belong to K on C.

As a particular case of this application, we mention the class of functions $L(2, k, \alpha)$, $o < \alpha < 1$, namely the class of functions whose kth derivatives on C satisfy there a square-integrated Lipschitz condition of order α; this class was first studied by Hardy and Littlewood, theorems proved in detail by Quade [3]. An alternative definition of the class is (7) with $\varepsilon_n = 1/n^{k+\alpha}$. It follows that every function $f(z)$ of class $L(2, k, \alpha)$, $o < \alpha < 1$, can be expressed on C as $f_1(z) + f_2(z)$, where the latter two functions, of respective classes H_2 and G_2 satisfy (8) and (9) with the same values of ε_n; thus $f_1(z)$ and $f_2(z)$ likewise belong to $L(2, k, \alpha)$ on C. The case $\alpha = 1$ can be similarly treated, where the integrated Lipschitz conditions are replaced by the condition

$$(10) \qquad \int_0^{2\pi} |F(\theta + h) + F(\theta - h) - 2F(\theta)|^2 d\theta = O(h^2) ,$$

and $F(\theta) \equiv f^{(k)}(z)$ is continuous on C; this class was introduced by Zygmund, and is characterized by (7) with $\varepsilon_n = 1/n^{k+1}$. We have as before $f(z) \equiv f_1(z) + f_2(z)$ if $f(z)$ is given, and the corresponding classes of $f_1(z)$ and $f_2(z)$ are characterized by (8) and (9) with the same values of ε_n, and by (10). These classes of analytic functions are studied in [5].

Added in proof. A second proof of Theorem 1, due to G. H. Curtiss, will appear shortly.

References

1. J. H. Curtiss, *Interpolation with harmonic and complex polynomials to boundary values*, Jour. of Math. and Mech., **9** (1960), 167–192.

2. P. Erdös and P. Turán, *On interpolation, (I) quadrature and mean convergence in the Lagrange interpolation*, Annals of Math., **38** (1937), 142–155.

3. E. S. Quade, *Trigonometric approximation in the mean*, Duke Math. Jour., **3** (1937), 529–543.

4. J. L. Walsh, *Interpolation and approximation*, Amer. Math. Soc. Coll. Pubs., **20** (1960).

5. J. L. Walsh and H. G. Russell, *Integrated continuity conditions and degree of approximation by polynomials*, Trans. Amer. Math. Soc., **92** (1959), 355–370.

A JORDAN-HÖLDER THEOREM

CHARLES E. WATTS

1. The purpose of this note is to present a certain general theorem, of the Jordan-Holder type, for finite groups. This theorem, although a simple and natural extension of the classical theorem, has we believe passed unnoticed before. The technique of proof is foreign to the usual methods of finite group theory, but seems well-suited to the situation.

2. A nonempty class $\mathscr{D}$ of finite groups will be called a *genetic class* provided:

(1) If G_1 belongs to $\mathscr{D}$ and if G_2 is isomorphic to G_1, then G_2 belongs to $\mathscr{D}$.

(2) If G belongs to $\mathscr{D}$, then every normal subgroup and every quotient group of G also belongs to $\mathscr{D}$.

The following examples of genetic classes will be used as illustrations in the sequel:

The class $\mathscr{G}$ of all finite groups.
The class $\mathscr{E}$ of all one-element groups.
The class $\mathscr{A}$ of all finite abelian groups.
The class $\mathscr{O}$ of all groups of odd order.
The class $\mathscr{G}_n$ of all groups of order $\leqq n$.

Given any genetic class $\mathscr{D}$, we shall construct a "Grothendieck group" in the following way. Let Σ be the (countable) set of all isomorphism classes of finite groups, and let F be the free abelian group generated by Σ. If G is any finite group, its isomorphism class will be denoted by $[G]$, so that elements of F are finite sums

$$\Sigma\lambda_i[G_i], \quad \lambda_i \in Z ,$$

where Z denotes the ring of integers. We let $N(\mathscr{D})$ be the subgroup of F generated by all elements of the form

$$[G] - [H] - [G/H]$$

such that H is a normal subgroup of G and G/H belongs to the genetic class $\mathscr{D}$. Finally we set $K(\mathscr{D}) = F/N(\mathscr{D})$ and let k: $F \rightarrow K(\mathscr{D})$ be the natural epimorphism. Our object is to determine the structure of the abelian group $K(\mathscr{D})$.

3. Let $\mathscr{D}$ be a genetic class and let G be an arbitrary finite

Received April 3, 1963. This work was supported by the National Science Foundation under grant NSF-G 24155.

group. We say G is *$\mathscr{D}$-simple* provided that G has more than one element and that no proper quotient group of G belongs to $\mathscr{D}$, i.e., if H is normal in G and if G/H belongs to $\mathscr{D}$, then $H = G$ or $H = 1$. In particular, if G itself belongs to $\mathscr{D}$ and is $\mathscr{D}$-simple, then it is simple because of the second axiom for genetic classes. The following illustrations are based on the examples given in § 2 above.

G is $\mathscr{G}$-simple if and only if it is simple in the classical sense.

Every finite group is $\mathscr{E}$-simple.

G is $\mathscr{A}$-simple if and only if either G is cyclic of prime order, or else $\neq 1$ and equal to its commutator subgroup.

G is $\mathscr{O}$-simple if and only if G is simple and has odd order, or else has even order and no proper normal subgroups of odd index.

G is $\mathscr{G}_n$-simple if and only if G is simple and has order $\leqq n$ or else has order $> n$ and no proper normal subgroups of index $\leqq n$.

Having given the definition of $\mathscr{D}$-simplicity, we can now state the theorem referred to in § 1:

THEOREM. *Given any genetic class $\mathscr{D}$, $K(\mathscr{D})$ is the free abelian group freely generated by the elements $k\,[S]$, where S is $\mathscr{D}$-simple.*

4. In this section we begin to prove the theorem above and show its relation to the Jordan-Holder theorem.

Let $\mathscr{D}$ be a genetic class, G any finite group. If G is not $\mathscr{D}$-simple and if $G \neq 1$, we can find a normal subgroup G_1' of G such that $1 \neq G_1' \neq G$ and G/G_1' belongs to $\mathscr{D}$. Let G_1 be a maximal proper normal subgroup containing G_1'. Then G/G_1, being a quotient of G/G_1', is in $\mathscr{D}$ and is simple (and a fortiori $\mathscr{D}$-simple). Now if G_1 is not $\mathscr{D}$-simple and is $\neq 1$, we repeat the process to find a normal subgroup G_2 of G_1 such that G_1/G_2 is in $\mathscr{D}$ and is simple. Eventually we shall get a sequence

(1) $G = G_0, G_1, G_2, \cdots, G_n$,

where G_{i+1} is normal in G_i, where G_i/G_{i+1} is in $\mathscr{D}$ and is simple ($i = 0, 1, \cdots, n-1$), and where either G_n is $\mathscr{D}$-simple and not in $\mathscr{D}$ or else $G_n = 1$. Since G_i/G_{i+1} belongs to $\mathscr{D}$, we have

$$\begin{aligned} k[G] &= k[G_0] = k[G_0/G_1] + k[G_1] = \cdots \\ &= k[G_0/G_1] + k[G_1/G_2] + \cdots + k[G_{n-1}/G_n] + k[G_n] \,. \end{aligned}$$

Clearly if $G_n = 1$, then $k[G_n] = 0$. Thus we have shown that the elements $k[S]$, S $\mathscr{D}$-simple, generate the group $K(\mathscr{D})$.

We remark at this point that once we have shown the linear independence (over Z) of these generators, it will follow that the $\mathscr{D}$-simple groups G_i/G_{i+1}, G_n are uniquely determined (up to isomorphism) by G, and are independent of the sequence 1) used to compute them. Thus in the case $\mathscr{D} = \mathscr{G}$, we get precisely the classical Jordan-Holder theorem. In the general case, the groups G_i/G_{i+1} are of course among the composition factors of G, but the group G_n (if it is not 1) is something new. It is a subnormal subgroup of G which depends, up to isomorphism, only on G and on $\mathscr{D}$.

Continuing our digression from the proof, let us say that two finite groups G and G' are *$\mathscr{D}$-equivalent* if they represent the same element of $K(\mathscr{D})$. Thus G and G' are $\mathscr{G}$-equivalent if and only if they have the same composition factors, while to be $\mathscr{E}$-equivalent it is clear that they must be isomorphic. In general, the smaller the genetic class $\mathscr{D}$, the sharper is the notion of $\mathscr{D}$-equivalence.

5. We return to the proof of the theorem; it remains to show that the generators $k[S]$, S $\mathscr{D}$-simple, are linearly independent over Z. We shall show that for each $\mathscr{D}$-simple group S there exists an integer-valued function f defined on Σ (and depending on S) such that:

(1) $f[S] = 1$;
(2) $f[T] = 0$ if T is any $\mathscr{D}$-simple group not isomorphic to S;
(3) If H is a normal subgroup of G and if G/H is in $\mathscr{D}$, then $f[G] = f[H] + f[G/H]$.

Because of (3) such a function induces a homomorphism $K(\mathscr{D}) \to Z$, vanishing on $k[T]$ if T is as in (2), but equal to 1 on $k[S]$. The linear independence of the generators is an immediate consequence of the existence of such homomorphisms.

We construct f inductively. Let Σ_r be the set of isomorphism classes of groups of order $\leqq r$. Define $f = 0$ on Σ_1. Now suppose that f has been defined on Σ_r in such a way that (1), (2), (3) hold whenever S, T, G have orders $\leqq r$. Next suppose that G has order $r+1$. If G is $\mathscr{D}$-simple, then the value of $f[G]$ is forced by (1) or by (2). If G is not $\mathscr{D}$-simple, then it has a normal subgroup H with G/H in $\mathscr{D}$ and with H and G/H in Σ_r. Consequently, the value of $f[G]$ must be given by (3), and it remains to show that $f[H] + f[G/H]$ is independent of the choice of H as long as H has order $\leqq r$ and G/H is in $\mathscr{D}$.

Thus let K be another such subgroup. Then G/HK is in $\mathscr{D}$, since it is a quotient of G/H, and $H/H \cap K$ is in $\mathscr{D}$, since it is isomorphic to HK/H, which is normal in G/H. Hence using the Noether isomorphisms we get

$$\begin{aligned} & f[H] + f[G/H] = f[H] + f[G/HK] + f[HK/H] \\ & = f[H] + f[G/HK] + f[K/H \cap K] \\ & = f[H/H \cap K] + f[H \cap K] + f[G/HK] + f[K/H \cap K] . \end{aligned}$$

This last expression being symmetric in H and K, it follows that $f[H]+f[G/H]=f[K]+f[G/K]$. Thus we have shown how to extend f unambiguously to Σ_{r+1} in such a way that (1), (2), (3) still hold on this enlarged domain. Therefore f can be defined on all of Σ so as to have the desired properties, and this completes the proof.

UNIVERSITY OF ROCHESTER

ON SOME FINITE GROUPS AND THEIR COHOMOLOGY

KUNG-WEI YANG

The purposes of this paper are: (I) to characterize the finite groups whose 2-Sylow subgroups are not isomorphic to a generalized quaternion group and which have periodic cohomology of period 4, (II) to show that all possible cohomologies of such a group G can be realized by direct sums of G-modules which belong to a specific finite family of G-modules.

The author wishes to express his deep gratitude to Professor G. Whaples and Dr. K. Grant for many helpful suggestions and continual encouragement.

The reader is referred to [1, Ch. XII] for basic notions, definitions and notations concerning cohomology of finite groups. The only departure from [1, Ch. XII] is the following: we shall say that a finite group G has periodic cohomology of period k if k is the *least* positive integer such that $\hat{H}^k(G, Z)$ contains a maximal generator [1, pp. 260–261]. And to avoid typographical difficulties we will denote by $Z(l)$ the cyclic group of order l.

PROPOSITION I. *Let G be a finite group whose 2-Sylow subgroups are not isomorphic to a generalized quaternion group. Then G has periodic cohomology of period 4 if and only if G has a presentation*

$$G = \{\sigma, \tau \colon \sigma^s = 1, \tau^t = 1, \tau\sigma\tau^{-1} = \sigma^{-1}\}, \text{ with the conditions}$$

(i) *s is an odd integer >1,*
(ii) *t is a positive even integer prime to s.*

Proof. Let G be a finite group whose 2-Sylow subgroups are not isomorphic to a generalized quaternion group and which has periodic cohomology of period 4. It is well-known [1, Theorem 11.6, p. 262] that if a finite group has periodic cohomology (of finite period) every Sylow subgroup of the group is either cyclic or is a generalized quaternion group. Since we assume that the 2-Sylow subgroups of G are not isomorphic to a generalized quaternion group, every Sylow subgroup of G is cyclic. It is also well-known [6, Theorem 11, p. 175] that a finite group G containing only cyclic Sylow subgroups is metacyclic and has a presentation

$$G = \{\sigma, \tau \colon \sigma^s = 1, \tau^t = 1, \tau\sigma\tau^{-1} = \sigma^r\}, \text{ with the conditions}$$

Received August 21, 1963.

(1) $0 < s$, ($st =$ the order of the group G),
(2) $((r-1)t, s) = 1$
(3) $r^t \equiv 1 (\mathrm{mod}\, s)$, and conversely.
We observe that if $s = 1$ or $t = 1$ or $r = 1$ the finite group G is cyclic and G has periodic cohomology of period 2 (or 0). These cases are therefore excluded. On the other hand, once these exceptional cases are excluded G is no more a cyclic group and it will have periodic cohomology of period $\geqq 4$.

Notice that (1), (2) and (3) imply (i)

Let H be the subgroup of G generated by the element σ. H is clearly a cyclic normal subgroup of order s. And G/H is cyclic of order t. By condition (2), s and t are relatively prime to each other. We can therefore apply the decomposition theorem of Hochschild-Serre [2, Theorem 1, p. 127] and obtain

$$\hat{H}^k(G, K) \cong \hat{H}^k(G/H, K^H) \oplus (\hat{H}^k(H, K))^{G/H}, \tag{4}$$

for all $k > 0$ and for all G-module K. (For $k > 0$, $\hat{H}^k(G, K) = H^k(G, K)$). In particular, we have

$$\hat{H}^k(G, Z) \cong \hat{H}^k(G/H, Z) \oplus (\hat{H}^k(H, Z))^{G/H},$$

for $k > 0$. The G/H-operators on $\hat{H}^k(H, K)$ are explicitly described in [2, p. 117]. In particular, G/H-operators on $\hat{H}^k(H, Z)$ are induced by the automorphisms of H which are themselves induced, on H, by inner automorphisms of G. In the present situation, all such automorphisms of H are generated by the automorphism $f(\rho) = \rho^r (= \tau\rho\tau^{-1})$, where $\rho \in H$. The automorphism $f: H \to H$ induces an automorphism f^* of $\hat{H}^k(H, Z)$ [4, Lemma 3, p. 343] such that if $g_{2k} \in \hat{H}^{2k}(H, Z)$, then $f^*(g_{2k}) = r^k g_{2k}$. Therefore $\hat{H}^4(G, Z)$ has a maximal generator, i.e. G has periodic cohomology of period $\leqq 4$ if and only if $f^*(g) = g$ for all $g \in \hat{H}^4(H, Z)$. This is equivalent to
(5) $r^2 \equiv 1 (\mathrm{mod}\, s)$.
(We recall that $r = 1$ we excluded). An elementary number theoretic calculation shows that the only solution for r in (2) and (5) is $r \equiv -1 (\mathrm{mod}\, s)$. Therefore the number t in (3) is an even positive integer (if it is negative, we can present G by letting $\tau' = \tau^{-1}$). This shows that the finite group G has a presentation as mentioned above.

The converse of the proposition is obvious.

We know that if l is the order of the group G then for any G-module K all the cohomology groups $\hat{H}^k(G, K)$ $(-\infty < k < \infty)$ are of exponent l—that is, for all $g \in \hat{H}^k(G, K)$, $lg = 0$. Let

$$s = p_1^{u_1} \cdots p_h^{u_h},\ P_1 = \{p_1, \cdots, p_h\} \quad \text{and} \quad t = q_1^{v_1} \cdots q_e^{v_e},\ P_2 = \{q_1, \cdots, q_e\}$$

be decompositions of s and t into products of prime powers (where

$q_1 = 2$ and $v_1 \geqq 1$). It is obvious from (4) that a group with periodic cohomology of period 4 has P_2-period [1, Exercise 11, p. 265] equal to 2. Conversely, we have

PROPOSITION II. *Let G be a group having a presentation*

$G = \{\sigma, \tau: \sigma^s = 1, \tau^t = 1, \tau\sigma\tau^{-1} = \sigma^{-1}\}$ *with the conditions*

(i) *s is an odd integer >1.*

(ii) *t is a positive even integer prime to s.*

Let P_1, P_2 be as defined above. Then there exists a finite family of G-modules $\mathscr{F}$ such that given any sequence of abelian groups $A_k(-\infty < k < \infty)$ satisfying

(a) *each A_k is of exponent st,*

(b) *the sequence is periodic of period 4,*

(c) *the P_2-period of the sequence is equal to 2, then there exists a G-module M which is a direct sum of G-modules of $\mathscr{F}$ such that $\hat{H}^k(G, M) = A_k(-\infty < k < \infty)$.*

First we observe the following

LEMMA. *Let G be a finite group and let K be a G-module. Let S be a set of primes in the ring of integers Z and let Q(S) be the quotient ring* [5, *p.* 46] *of Z with respect to the multiplicative system generated by S. (As usual when Q(S) is considered as a G-module it is to be understood that G operates trivially on (the additive group of) Q(S)). Then*

$$\hat{H}^k(G, K \otimes Q(S)) \cong \hat{H}^k(G, K) \otimes Q(S)(-\infty < k < \infty),$$

where $\otimes = \otimes_Z$

The proof is immediate.

Proof of Proposition II. Let s, t, P_1, P_2 be as before. Let

$$s(i, \mu) = s/p_i^{\mu}(i = 1, \cdots, h, 0 \leqq \mu \leqq u_i),$$
$$t(i, \nu) = t/q_i^{\nu}(i = 1, \cdots, e, 0 \leqq \nu \leqq v_i).$$

Let $K^1(i, \mu) = \sum_{j=1}^{s(i,\mu)} Zx_j^{(i,\mu)}$ (direct sum on the symbols $x_j^{(i,\mu)}$)

$$K^2(i, \nu) = \sum_{j=1}^{t(i,\nu)} Zy_j^{(i,\nu)} \text{ (direct sum on the symbols } y_j^{(i,\nu)}).$$

Define G-operators on $K^1(i, \mu)$ and $K^2(i, \nu)$ by

$$\left.\begin{aligned} \sigma x_j^{(i,\mu)} &= x_{j+1}^{(i,\mu)} \\ \tau x_j^{(i,\mu)} &= x_{-j}^{(i,\mu)}, \end{aligned}\right. \quad \text{(subscripts are modulo } s(i, \mu))$$

$$\begin{aligned} \sigma y_j^{(i,\nu)} &= y_j^{(i,\nu)} \\ \tau y_j^{(i,\nu)} &= y_j^{(i,\nu)} . \end{aligned} \quad \text{(subscripts are modulo } t(i, \nu)) .$$

Let

$$\begin{aligned} M^1(i, \mu) &= K^1(i, \mu) \otimes Q((P_1 - \{p_i\}) \cup P_2), \\ M^2(i, \nu) &= K^2(i, \nu) \otimes Q(P_1 \cup (P_2 - \{q_i\})) . \end{aligned}$$

By (4), the above lemma and the fact that $(\hat{H}^{4k+2}(H, K^1(i, \mu))^{G/H} = (0)$, one shows

$$\begin{aligned} \hat{H}^{4k}(G, M^1(i, \mu)) &= Z(p_i^\mu) & \hat{H}^{4k}(G, M^2(i, \nu)) &= Z(q_i^\nu) \\ \hat{H}^{4k+1}(G, M^1(i, \mu)) &= (0) & \hat{H}^{4k+1}(G, M^2(i, \nu)) &= (0) \\ \hat{H}^{4k+2}(G, M^1(i, \mu)) &= (0) & \hat{H}^{4k+2}(G, M^2(i, \nu)) &= Z(q_i^\nu) \\ \hat{H}^{4k+3}(G, M^1(i, \mu)) &= (0) & \hat{H}^{4k+3}(G, M^2(i, \nu)) &= (0) \end{aligned}$$

The calculation is purely mechanical.

Now, let $0 \to I \to Z[G] \xrightarrow{\varepsilon} Z \to 0$, where $\varepsilon(\sum_{\sigma\in G} l_\sigma \sigma) = \sum_{\sigma \in G} l_\sigma$, $I =$ Ker(ε), and let $\mathscr{F}$ consist of

$$\begin{aligned} &I^k \otimes M^1(i, \mu)(k = 0, 1, 2, 3, i = 1, \cdots, h, 0 \leqq \mu \leqq u_i) \\ &I^k \otimes M^2(i, \nu)(k = 0, 1, i = 1, \cdots, e, 0 \leqq \nu \leqq v_i) , \end{aligned}$$

where $I^k = I \otimes \cdots \otimes I (k \text{ times})$, $I^\circ = Z$.

Suppose we are given a sequence of abelian groups $A_k(-\infty < k < \infty)$ satisfying conditions (a), (b), (c). Since by (a) each A_k is of exponent st, it follows from [3, Theorem 6, p. 17] that A_k is a direct sum of cyclic groups. Let nA denote the direct sum of n copies of A, where A is either an abelian group or a G-module and n is a cardinal number. Then we can write

$$A_k = \sum_{i=1}^{h} \sum_{0 \leqq \mu \leqq u_i} m(k, i, \mu) Z(p_i^\mu) \oplus \sum_{i=1}^{e} \sum_{0 \leqq \nu \leqq v_i} n(k, i, \nu) Z(q_i^\nu) ,$$

where $m(k, i, \mu) = m(k + 4, i, \mu)(i = 1, \cdots, h, 0 \leqq \mu \leqq u_i)$, $n(k, i, \nu) = n(k + 2, i, \nu)(i = 1, \cdots, e, 0 \leqq \nu \leqq v_i)$ and $m(k, i, \mu)$, $n(k, i, \nu)$ are cardinal numbers. Take

$$\begin{aligned} M = &\sum_{k=0}^{3} \sum_{i=1}^{h} \sum_{0 \leqq \mu \leqq u_i} m(k, i, \mu) I^k \otimes M^i(i, \mu) \\ &\oplus \sum_{k=0}^{1} \sum_{i=1}^{e} \sum_{0 \leqq \nu \leqq v_i} n(k, i, \nu) I^k \otimes M^2(i, \nu) . \end{aligned}$$

Observe that $\hat{H}^{k-l}(G, K) \cong \hat{H}^k(G, I^l \otimes K)$. Clearly $\hat{H}^k(G, M) = A_k$ $(-\infty < k < \infty)$.

REMARK. In a similar but much simpler fashion one can show that all possible cohomology of a cyclic group G can also be realized by direct sums of G-modules of a certain finite family of G-modules $\mathscr{F}'$.

Addendum to the paper

"On Some Finite Groups And Their Cohomology"

(Received October 11, 1963)

Let group G have a presentation

$$(*) \qquad G = \{\sigma, \tau : \sigma^s = 1, \tau^t = 1, \tau\sigma\tau^{-1} = \sigma^r\} ,$$

with the conditions

(i) $0 < s$

(ii) $((r-1)t, s) = 1$

(iii) $r^t \equiv 1 (\mathrm{mod}\ s)$

(iv) there exists a positive integer n such that n is the order to which r belongs to moduli p_i $(i = 1, \cdots, h)$ (i.e. n is the smallest positive integer such that $r^n \equiv 1(\mathrm{mod}\ p_i))$, where $s = p_1^{u_1} \cdots p_h^{u_h}$. Let s, t, P_1, P_2, be as defined before (here q_1 is not necessarily $=2$). It is clear from condition (iv) that G has P_1-period equal to $2n$ and P_2-period equal to 2.

Proposition III. *Let G be a group having a presentation $(*)$ with the conditions (i), (ii), (iii), (iv). Then there exists a finite family of G-modules $\mathscr{F}$ such that given any sequence of abelian groups $A_k(-\infty < k < \infty)$ satisfying the following conditions:*

(a) *each A_k is of exponent st*

(b) *the P_1-period (in the obvious sense) of the sequence is $2n$*

(c) *the P_2-period of the sequence is 2,*

there exists a G-module M, which is a direct sum of G-modules of $\mathscr{F}$ such that $\hat{H}^k(G, M) = A_k(-\infty < k < \infty)$.

Proof. Let $s(i, \mu)$, $t(i, \nu)$, $K^1(i, \mu)$, $K^2(i, \nu)$, be as defined in Proposition II, Define G-operators on $K^1(i, \mu)$ and $K^2(i, \nu)$ by

$$\left.\begin{aligned} \sigma x_j^{(i,\mu)} &= x_{j+1}^{(i,\mu)} \\ \tau x_j^{(i,\mu)} &= x_{r\cdot j}^{(i,\mu)} \end{aligned}\right. \quad \text{(subscripts are modulo } s(i, \mu)) ,$$

$$\left.\begin{aligned} \sigma y_j^{(i,\nu)} &= y_j^{(i,\nu)} \\ \tau y_j^{(i,\nu)} &= y_{j+1}^{(i,\nu)} \end{aligned}\right. \quad \text{(subscripts are modulo } t(i, \nu)) .$$

By condition (iv) we have

$$\hat{H}^{2nk+i}(H, K^1(i, \mu))^{G/H} = (0)(i = 1, 2, \cdots, 2n-1) .$$

The rest of the proof is parallel to that of Proposition II. $\mathscr{F}$ consists of G-modules

$$I^k \otimes M^1(i, \mu)(k = 0, 1, \cdots, 2n-1; i = 1, \cdots, h; \mu = 0, 1, \cdots, u_i)$$
$$I^k \otimes M^2(i, \nu)(k = 0, 1; i = 1, 2, \cdots, e; \nu = 0, 1, \cdots, v_i) .$$

Bibliography

1. H. Cartan and S. Eilenberg, *Homological Algebra,* Princeton, 1956.

2. G. Hochschild and J-P. Serre, *Cohomology of group extensions,* Trans. Amer. Math. Soc., **74** (1953), 110–134.

3. I. Kaplanksy, *Infinite Abelian Groups,* University of Michigan Press, Ann Arbor, 1956.

4. R. G. Swan, *The p-period of a finite group,* Ill. J. Math. **4** (1960), 341–346.

5. O. Zariski and P. Samuel, *Commutative Algebra,* Vol. 1, Van Nostrand, Princeton, 1958.

6. H. Zassenhaus, *The Theory of Groups,* 2nd ed., Chelsea, New York, 1958.

INDIANA UNIVERSITY

ON THE RING-LOGIC CHARACTER OF CERTAIN RINGS

ADIL YAQUB

Introduction. Boolean rings $(B, \times, +)$ and Boolean logics (= Boolean algebras) $(B, \cap, {}^*)$ though historically and conceptionally different, are equationally interdefinable in a familiar way [6]. With this equational interdefinability as motivation, Foster introduced and studied the theory of ring-logics. In this theory, a ring (or an algebra) R is studied modulo K, where K is an arbitrary transformation group in R. The Boolean theory results from the special choice, for K, of the "Boolean group," generated by $x^* = 1 - x$ (order 2, $x^{**} = x$). More generally, let $(R, \times, +)$ be a commutative ring with identity 1, and let $K = \{\rho_1, \rho_2, \cdots\}$ be a transformation group in R. The *K-logic* (*or K-logical algebra*) of the ring $(R, \times, +)$ is the (operationally closed) system $(R, \times, \rho_1, \rho_2, \cdots)$ whose class R is identical with the class of ring elements, and whose operations are the ring product "$\times$" of the ring together with the unary operations $\rho_1, \rho_2, \cdots$ of K. The ring $(R, \times, +)$ is called a *ring-logic*, mod K if (1) the "+" of the ring is *equationally* definable in terms of its K-logic $(R, \times; \rho_1, \rho_2, \cdots)$, and (2) the "+" of the ring is *fixed* by its K-logic. Of particular interest in the theory of ring-logics is the *normal group* D which was shown in [1] to be particularly adaptable to p^k-rings. Our present object is to extend further the class of ring-logics, modulo the normal group D itself. A by-product of this extension is the following result, namely, any finite commutative ring with zero radical is a ring-logic, mod D (see Corollary 8). Furthermore, in Corollary 10, we prove that, more generally, any (not necessarily finite) ring with unit which satisfies $x^n = x$(n fixed, $\geqq 2$) is a ring-logic (mod D). Finally, we compare the normal group with the so-called *natural* group in regard to the ring-logic character of a certain important class of rings (see section 3).

1. The finite field case. Let $(F_{p^k}, \times, +)$ be a Galois (finite) field with exactly p^k elements (p prime). Then, as is well known, F_{p^k} contains a multiplicative generator, ξ;

$$F_{p^k} = \{0, \xi, \xi^2, \cdots, \xi^{p^k-1} (=1)\} .$$

We now have the following (compare with [1]).

THEOREM 1. *Let F_{p^k} be a Galois field, and let ξ be a generator of F_{p^k}. Then the mapping $x \to x^\frown$ defined by*

Received August 16, 1963.

(1.1) $$x^{\frown} = \xi x + (1 + \xi x + \xi^2 x^2 + \cdots + \xi^{p^k-2} x^{p^k-2})$$

is a permutation of F_{p^k}, *with inverse given by*

(1.2) $$x^{\smile} = \xi^{p^k-2}(1 + x + x^2 + \cdots + x^{p^k-2}) + \xi^{p^k-2} x .$$

Furthermore, the permutation $\frown$ *is of period* p^k,

(1.3) $$x^{\frown p^k} = (\cdots (x^{\frown})^{\frown} \cdots)^{\frown} \quad (p^k\text{-iterations}) = x .$$

Proof. Since $a^{p^k-1} = 1$, $a \in F_{p^k}$, $a \neq 0$, therefore, by (1.1), $x^{\frown} = \xi x + \{[(1 - (\xi x)^{p^k-1}]/(1 - \xi x)\} = \xi x$, if $x \neq 0$ and $\xi x \neq 1$. Furthermore, by (1.1), $0^{\frown} = 1$ and $(1/\xi)^{\frown} = p^k \cdot 1 = 0$. Hence, $0^{\frown} = 1$, $1^{\frown} = \xi$, $\xi^{\frown} = \xi^2$, $(\xi^2)^{\frown} = \xi^3, \cdots, (\xi^{p^k-2})^{\frown} = 0$. This proves (1.3). To prove (1.2), observe that the right-side of (1.2) is equal to

$$\frac{1}{\xi} x + \frac{1}{\xi}\left\{\frac{1 - x^{p^k-1}}{1 - x}\right\} = \frac{1}{\xi} x , \qquad \text{if } x \neq 1 \text{ and } x \neq 0 .$$

Moreover, if $x \neq 0$ and $x \neq 1/\xi$, then $x^{\frown} = \xi x$ and hence $x^{\smile} = (1/\xi)x$. Since (1.2) clearly holds for $x = 0$, $x = 1/\xi$, and $x = 1$, therefore (1.2) is true for all elements of F_{p^k}, and the theorem is proved.

COROLLARY 2. *Under the permutation* $\frown$, F_{p^k} *suffers the cyclic permutation*

(1.4) $$(0, 1, \xi, \xi^2, \xi^3, \cdots, \xi^{p^k-2}) .$$

Following [1], we call $x^{\frown}$ the *normal negation* of x, and call the cyclic group D whose generator is $x^{\frown}$ the *normal group*. By Theorem 1, it is now clear that

$$D = D(\xi) = \{\text{identity}, \frown, \frown^2, \frown^3, \cdots, \frown^{p^k-1}\} .$$

As in [1], we define

(1.5) $$a \times_{\frown} b = (a^{\frown} \times b^{\frown})^{\smile} .$$

It is readily verified that

(1.6) $$a \times_{\frown} 0 = a = 0 \times_{\frown} a .$$

COROLLARY 3. *The elements of* F_{p^k} *are equationally definable in terms of the D-logic.*

Proof. By Corollary 2, it is easily seen that

$$(1.7)\qquad \begin{aligned} 0 &= xx^{\frown}x^{\frown 2}\cdots x^{\frown p^k-1} \\ 1 &= 0^{\frown} \\ \xi &= 1^{\frown} \\ \xi^2 &= \xi^{\frown} \\ &\cdots \\ \xi^{p^k-2} &= (\xi^{p^k-3})^{\frown}\,, \end{aligned}$$

and the corollary follows.

We recall from [3] the *characteristic function* $\delta_\mu(x)$, defined as follows: for a given $\mu \in F_{p^k}$,

$$(1.8)\qquad \delta_\mu(x) = \begin{cases} 1 & \text{if } x = \mu \\ 0 & \text{if } x \neq \mu\,. \end{cases}$$

In view of Corollary 2, it is easily seen that, for any given $\mu \in F_{p^k}$, there exists an integer r such that $\mu^{\frown r} = 0$. Then, clearly,

$$(1.9)\qquad \delta_\mu(x) = \delta_0(x^{\frown r}) \qquad \text{where } \mu^{\frown r} = 0\,.$$

Now, let $\sum^{\times}_{\alpha_i \in F} \alpha_i$ denote $\alpha_1 \times_{\frown} \alpha_2 \times_{\frown} \alpha_3 \cdots$, where $\alpha_1, \alpha_2, \alpha_3, \cdots$ are the elements of F. Then, by (1.6) and (1.8), we have the identity [3]

$$(1.10)\qquad f(x, y, \cdots) = \sum^{\times_\frown}_{\alpha,\beta,\cdots \in F_{p^k}} f(\alpha, \beta, \cdots)(\delta_\alpha(x)\delta_\beta(y)\cdots)\,.$$

In (1.10), $\alpha, \beta, \cdots$ range over all the elements of F_{p^k} while $x, y, \cdots$ are indeterminates over F_{p^k}. We shall use (1.9) and (1.10) presently.

LEMMA 4. *The characteristic functions $\delta_\mu(x)$, $\mu \in F_{p^k}$, are equationally definable in terms of the D-logic.*

Proof. Since $x^{p^k-1} = 1$, $x \neq 0$, $x \in F_{p^k}$, therefore, $\delta_0(x) = ((x^{p^k-1})^{\smile})^{p^k-1}$. Hence $\delta_0(x)$ is *equationally* definable in terms of the D-logic. Therefore, by (1.9), $\delta_\mu(x)$ is also equationally definable in terms of the D-logic, and the lemma is proved.

We are now in a position to prove the following.

THEOREM 5. *The Galois field $(F_{p^k}, \times, +)$ is a ring-logic* (mod D).

Proof. By (1.10), we have,

$$x + y = \sum^{\times_\frown}_{\alpha\,\beta \in F_{p^k}} (\alpha + \beta)(\delta_\alpha(x)\delta_\beta(y))\,.$$

Now, by Corollary 3, $\alpha + \beta$ is equationally definable in terms of the

D-logic. Moreover, by Lemma 4, each of the characteristic functions $\delta_\alpha(x)$ and $\delta_\beta(y)$ is equationally definable in terms of the D-logic. Hence the "+" of F_{p^k} is *equationally* definable in terms of the D-logic $(F_{p^k}, \times, \frown, \smile)$. Next, we show that $(F_k, \times, +)$ is *fixed* by its D-logic. Suppose then that there exists another ring $(F_{p^k}, \times, +')$, with the same class of elements F_{p^k} and the same "$\times$" as $(F_{p^k}, \times, +)$ and which has the *same logic* as $(F_{p^k}, \times, +)$. To prove that $+' = +$. Since both $(F_{p^k}, \times, +)$ and $(F_{p^k}, \times, +')$ have the *same* class of elements and the *same* "$\times$", it readily follows that $(F_{p^k}, \times +')$ is also a Galois field with exactly p^k elements. Since, up to isomorphism, there is *only one* Galois field with exactly p^k elements, therfore, $+' = +$, and the theorem is proved.

2. The General Case. In order to extend Theorem 5 to *any* finite commutative ring with zero radical, the following concept of independence, introduced by Foster [2], is needed.

DEFINITION. Let $\bar{A} = \{A_1, A_2, \cdots, A_n\}$ be a finite set of algebras of the same species S_p. We say that the algebras $A_1, A_2, \cdots, A_n$ are *independent* if, corresponding to each set $\{\varphi_i\}$ of expressions of species S_p $(i = 1, \cdots, n)$ there exists at least one expression ψ such that $\psi = \varphi_i \pmod{A_i}$ $(i = 1, \cdots, n)$. By an *expression* we mean some composition of one or more indeterminate-symbols $\xi, \cdots$ in terms of the primitive operations of $A_1, A_2, \cdots, A_n$; $\psi = \varphi \pmod{A}$ means that this is an identity of the algebra A.

We now examine the independence of the D-logics $(F_{p_i^{k_i}}, \times, \frown, \smile)$. Indeed, we have the following (compare with [2]).

THEOREM 6. *Let $p_1, \cdots, p_t$ be distinct primes. Then the D-logics $(F_{p_i^{k_i}}, \times, \frown, \smile)$ are independent.*

Proof. Let $n_i = p_i^{k_i}$, $F_i = F_{p_i^{k_i}} = \{0, 1, \lambda, \lambda^2, \cdots, \lambda^{n_i-2}\}$, $n = \max_{1 \leqq i \leqq t} \{n_i\}$, $N = \prod_{j=1}^t n_j$, $n_i N_i = N$, $E = \xi\xi^{\frown}\xi^{\frown 2} \cdots \xi^{\frown n-1}$.

It is easily seen, since the n_i's are *distinct prime* powers, that

$$|_i(\xi) = (E^{\frown N_i})^{n_i-1} = \begin{cases} 1 \pmod{F_i} \\ 0 \pmod{F_j} \end{cases} \quad (j \neq i) .$$

Now, to prove the indepedence of the logics $(F_i, \times, \frown, \smile)$ $(i = 1, \cdots, t)$ let $\varphi_1, \cdots, \varphi_t$ be any set of t expressions of species $\times, \frown, \smile$, i.e., primitive compositions of indeterminate-symbols in terms of the operations $\times, \frown, \smile$. Define an expression $K(\varphi_1, \cdots, \varphi_t)$ as follows (compare with [2]):

$$K(\varphi_1, \cdots, \varphi_t) = (\varphi_1 \cdot |_1(\xi)) \times_{\frown} (\varphi_2 \cdot |_2(\xi)) \times_{\frown} \cdots \times_{\frown} (\varphi_t \cdot |_t(\xi)) .$$

Then it is easily seen that $K(\varphi_1, \cdots, \varphi_t) = \varphi_i \pmod{F_i}$ $(i = 1, \cdots, t)$, since $a \times_{\frown} 0 = 0 \times_{\frown} a = a$, and the theorem is proved.

We shall now extend the concept of ring-logic to the direct sum of certain ring-logics. We denote the direct sum of A_1 and A_2 by $A_1 \oplus A_2$. The direct power A^m will denote $A \oplus A \oplus \cdots \oplus A$ (m summands).

THEOREM 7. *Let A be any subdirect sum with identity of (not necessarily finite) subdirect powers of the Galois fields $F_{p_i^{k_i}}$ $(i = 1, \cdots, t)$. Then A is a ring-logic (mod D).*

Proof. Let $q_1, \cdots, q_r$ be the *distinct* primes in $\{p_1, \cdots, p_t\}$. Since the Galois Fields $F_{p^{k_i}}$ and $F_{p^{k_j}}$ are both subfields of $F_{p^{k_i k_j}}$, it is easily seen that A is a *subring* of a direct sum of direct powers of $F_{q_i^{h_i}}$, $(i = 1, \cdots, r)$; i.e., A is a subring of $F_{q_1^{h_1}}^{m_1} \oplus \cdots \oplus F_{q_r^{h_r}}^{m_r}$ for some positive integers $h_1, \cdots, h_r$. Now, by Theorem 5, each $F_{q_i^{h_i}}$ is a ring-logic (mod D), and hence there exists a D-logical expression φ_i such that, for every $x_i, y_i \in F_{q_i^{h_i}}$ $(i = 1, \cdots, r)$,

$$x_i + y_i = \varphi_i(x_i, y_i; \times, \frown, \smile) .$$

Since, by Theorem 6, the D-logics $(F_{q_i^{h_i}}, \times, \frown, \smile)$ $(i = 1, \cdots, r)$ are independent, there exists a D-logical expression K such that

$$K = \begin{cases} \varphi_1 \pmod{F_{q_1^{h_1}}} \\ \cdots \\ \varphi_r \pmod{F_{q_r^{h_r}}} . \end{cases}$$

Therefore, for every $x_i, y_i \in F_{q_i^{h_i}}$ $(i = 1, \cdots, r)$,

$$x_i + y_i = \varphi_i = K(x_i, y_i; \times, \frown, \smile) .$$

Hence, the D-logical expression K represents the "+" of each $F_{q_i^{h_i}}$. Since the operations are component-wise in the direct sum $F_{q_1^{h_1}}^{m_1} \oplus \cdots \oplus F_{q_r^{h_r}}^{m_r}$, therefore, for all x, y in this direct sum, we have,

$$x + y = K(x, y; \times, \frown, \smile) .$$

Hence, *a fortiori*, the "+" of the *subring* A is *equationally* definable in terms of the D-logic.

Next, we show that A is *fixed* by its D-logic. Suppose there exists a "+′" such that $(A, \times, +')$ is a ring, with the same class of elements A and the same "×" as the ring $(A, \times, +)$, and which has the *same logic* $(A, \times, \frown, \smile)$ as the ring $(A, \times, +)$. To prove that $+' = +$. Now, since A is a subdirect sum of subdirect powers of $F_{p_i^{k_i}}$, therefore, a new "+′" in A defines and is defined by a new

"$+_1'$" in $F_{p_1^{k_1}}$, "$+_2'$" in $F_{p_2^{k_2}}$, $\cdots$, "$+_t'$" in $F_{p_t^{k_t}}$, such that $(F_{p_i^{k_i}}, \times, +_i')$ is a ring ($i = 1, \cdots, t$). Furthermore, the assumption that $(A, \times, +')$ has the same logic as $(A, \times, +)$ is equivalent to the assumption that each $(F_{p_i^{k_i}}, \times, +_i')$ has the same logic as $(F_{p_i^{k_i}}, \times, +)$ ($i = 1, \cdots, t$). Since, by Theorem 5, $(F_{p_i^{k_i}}, \times, +)$ is a ring-logic, and hence with its "$+$" fixed, it follows that $+_i' = +$ ($i = 1, \cdots, t$). Hence $+' = +$, and the theorem is proved.

Now, it is well known (see [4]) that any finite commutative ring with zero radical and with more than one element is isomorphic to the complete direct sum of a finite number of finite fields. Hence, Theorem 7 has the following

COROLLARY 8. *Any finite commutative ring with zero radical is a ring-logic* (mod D).

It is also well known (see [1; 5]) that every p-ring (p prime) is isomorphic to a subdirect power of F_p, and every p^k-ring (p prime) is isomorphic to a subdirect power of F_{p^k}. Hence, by letting $t = 1$ in Theorem 7, we obtain the following (compare with [1; 7])

COROLLARY 9. *Any p-ring with identity, as well as any p^k-ring with identity, is a ring-logic* (mod D).

Now, let n be a fixed integer, $n \geqq 2$. It is well known that a ring R which satisfies $x^n = x$ for all x in R is isomorphic to a subdirect sum of (not necessarily finite) subdirect powers of a *finite* set of Galois fields. Hence Theorem 7 has the following

COROLLARY 10. *Let R be a ring with unit such that $x^n = x$ for all x in R, where n is a fixed integer, $n \geqq 2$. Then R is a ring-logic* (mod D).

3. The natural group and the normal group. Let $(R, \times, +)$ be a commutative ring with unit 1. We recall (see [1]) that the *natural group* N is the group generated by $x^\wedge = x + 1$ (with inverse $x^\vee = x - 1$). In [7], it was shown that $(F_{p^k}, \times, +)$ is a ring-logic (mod N), and hence the "$+$" of F_{p^k} is equationally definable in terms of the N-logic $(F_{p^k}, \times, {}^\wedge)$. Moreover, by Theorem 5, $(F_{p^k}, \times, +)$ is a ring-logic (mod D), and hence the "$+$" of F_{p^k} is equationally definable in terms of the D-logic $(F_{p^k}, \times, \frown)$. Of the two rival logics, $(F_{p^k}, \times, \frown)$ requires only a knowledge of the multiplication table in F_{p^k} since, by Corollary 2, the effect of $\frown$ on F_{p^k} is the cyclic permutation $(0, 1, \xi, \xi^2, \cdots, \xi^{p^k-2})$. In this sense, the D-logical formula for the "$+$" of F_{p^k} is a *strictly multiplicative formula, and addition is thus* *equationally definable in terms of multiplication* whenever the generator ξ is chosen (compare with [1]). The situation is quite different in the case of the N-logical formula for the "+" of F_{p^k}, since the generator $x^\wedge = x + 1$ of the natural group N already involves a limited use of the addition table.

REFERENCES

1. A. L. Foster, *p^k-rings and ring-logics*, Ann. Scn. Norm Pisa, **5** (1951), 279-300.
2. ———, *The identities of—and unique subdirect factorization within—classes of universal algebas*, Math. Z., **62** (1955), 171-188.
3. ———, *Generalized Boolean theory of universal algebras, Part I*, Math. Z, **58** (1953), 306-336.
4. N. H. McCoy, *Rings and ideals*, Carus Math. Monog., **8** (1947).
5. N. H. McCoy and D. Montgomery, *A representation of generalized Boolean rings*, Duke Math J., **3** (1937), 455-459.
6. M. H. Stone, *The theory of representations of Boolean algebras*, Trans. Amer. Soc., **40** (1936), 37-111.
7. A. Yaqub, *On certain finite rings and ring-logics*, Pacific J. Math., **12** (1962), 785-790.

UNIVERSITY OF CALIFORNIA, SANTA BARBARA

A NOTE ON PSEUDO-CREATIVE SETS AND CYLINDERS

PAUL R. YOUNG

1. Notation and Definitions. We will use N to denote the set of all nonnegative integers. Unless specifically mentioned otherwise, all sets are considered subsets of N. If A is a set, $A' = N - A$. Since we consider only sets of nonnegative integers, we will not use Cartesian products of sets but will instead work with images of Cartesian products under some effective mapping. More specifically, if A and B are sets, let $A \otimes B = \{(a, b) \mid a \in A \text{ and } b \in B\}$. Let τ be any one-to-one effective mapping of $N \otimes N$ onto N. Then we define $A \times B$ to be $\tau(A \otimes B)$, and we abbreviate $\tau((a, b))$ to $\langle a, b\rangle$. (This is the notation introduced by Rogers in [4].) Given integers a and b we can always effectively find the integer $\langle a, b\rangle$, and given the integer $\langle a, b\rangle$ we can always effectively find a and b.

In [2], Myhill has called a set a cylinder if it is recursively isomorphic to $B \times N$ for some r.e. set B; however we will follow Rogers in calling a set, A, a cylinder if it is recursively isomorphic to $B \times N$ for any set B. Such a set A is called a cylinder of B.

For definitions of recursive, simple, and creative sets, see [3]. A noncreative, recursively enumerable (r.e.), set A has been called pseudo-creative if for every r.e. set $B \subset A'$ there is an infinite r.e. set $C \subset A'$ such that $B \cap C = \emptyset$. A nonrecursive r.e. set A has been called pseudo-simple if there is an infinite r.e. set $B \subset A'$ such that $A \cup B$ is simple. We will denote the class of all recursive sets by $\mathscr{C}_0$, the class of all simple sets by $\mathscr{C}_1$, the class of all pseudo-simple sets by $\mathscr{C}_2$, the class of all pseudo-creative sets by $\mathscr{C}_3$, and the class of all creative sets by $\mathscr{C}_4$. These classes are pairwise disjoint and every r.e. set falls into one of the classes ([2]).

Let A and B be sets. We write $A \leqq_1 B$ if there is a one-to-one recursive function such that $x \in A$ if and only if $f(x) \in B$, $A \leqq_m B$ if there is some recursive function g such that $x \in A$ if and only if $g(x) \in B$, and $A \leqq_{btt} B$ if A is reducible to B via bounded truth-tables. If there is no recursive function g such that $x \in A$ if and only if $g(x) \in B$, we write $A \not\leqq_m B$. If both $A \leqq_m B$ and $B \leqq_m A$, we write $A \equiv_m B$.

2. Introduction and preliminaries. In [2] it is shown that the class of pseudo-creative sets is nonempty by proving that the cylinder of any nonrecursive, noncreative, r.e. set is pseudo-creative. In this

Received June 28, 1963. Supported by National Science Foundation Fellowships. The results reported here are a portion of the author's doctoral dissertation written at M.I.T. under the direction of Prof. Hartley Rogers, Jr.

note we shall show that there is a pseudo-creative set which is not a cylinder, and we shall develop some related facts concerning the relation between pseudo-creative sets and cylinders.

LEMMA 1 (Myhill). *Every creative set is a cylinder. Every recursive set which is infinite and has an infinite complement is a cylinder. The empty set and N are cylinders. If A is pseudo-creative, pseudo-simple, or simple, then any cylinder of A is pseudo-creative. No simple set or pseudo-simple set is a cylinder. If A is r.e., then* $A \leqq_1 A \times N$ *and* $A \times N \leqq_m A$.

Proof. The proofs are straightforward and may be found in [2]. The requirement in the last assertion that A be r.e. may be omitted.

LEMMA 2. *Let A be a cylinder. Then there exists a one-to-one recursive function f such that* $x \in A$ *implies that* $\{x, f(x), f^2(x), f^3(x), \cdots\}$ *is an infinite r.e. subset of A, and* $x \in A'$ *implies that* $\{x, f(x), f^2(x), f^3(x), \cdots\}$ *is an infinite r.e. subset of* A'.

Proof. We may assume $A = B \times N$ for some set B. Define $f(\langle x, n\rangle) = \langle x, n+1\rangle$.

LEMMA 3 (Post-Shoenfield). *If B is a r.e. set and if* $A \leqq_{btt} B$ *where A is creative, then B is either creative or pseudo-creative.*

Proof. In [3] it is shown that B cannot be recursive or simple. In [5] it is shown that B cannot be pseudo-simple.

LEMMA 4. *Let* $A \in \mathscr{C}_i$, $B \in \mathscr{C}_j$, *and* $A \leqq_1 B$. *Then* $i \leqq j$.

Proof. The proof follows easily from the definitions and will be omitted.

LEMMA 5 (Fischer). *There is a simple set S such that* $S \times S \not\leqq_m S$.

Proof. See [1].

3. **Results.** An infinite set which contains no infinite r.e. subset is called immune ([3]).

LEMMA 6. *If A is a nonimmune infinite set, then* $A \times N \leqq_1 A \times A$.

Proof. Let B be an infinite r.e. subset of A and let g be a one-to-one recursive function whose range is B. Define $h(\langle a, b\rangle) = \langle a, g(b)\rangle$.

Then h is a one-to-one recursive function and $x \in A \times N$ if and only if $h(x) \in A \times A$.

COROLLARY 1. *Suppose S is simple or pseudo-simple. Then $S \times S$ is pseudo-creative.*

Proof. By Lemma 6, $S \times N \leqq_1 S \times S$. By Lemma 1 $S \times N$ is pseudo-creative and therefore by Lemma 4 $S \times S$ is either pseudo-creative or creative. Since $S \times S \leqq_{btt} S$, by Lemma 3 $S \times S$ cannot be creative.

THEOREM 1. *Let A be an infinite nonimmune set. Then $A \times A \leqq_m A$ implies that $A \times A$ is a cylinder.*

Proof. Suppose $A \times A \leqq_m A$ via the recursive function g. Define $h(\langle a, b\rangle) = \langle g(\langle a, b\rangle), \langle a, b\rangle\rangle$. Then $A \times A \leqq_1 A \times N$ via h. By Lemma 6, $A \times N \leqq_1 A \times A$. Thus $A \times A$ is recursively isomorphic to $A \times N$.

THEOREM 2. *Let A be any infinite r.e. set which is not pseudo-creative. Then $A \times A \leqq_m A$ if and only if $A \times A$ is a cylinder.*

Proof. In view of the preceding theorem, we need only prove that if $A \times A$ is a cylinder then $A \times A \leqq_m A$.

If A is creative or recursive so is $A \times A$, and in this case $A \times A \equiv_m A$ and $A \times A$ is a cylinder. Therefore we may assume that A is simple or pseudo-simple. Let $B \subset A'$ be a r.e. set such that $A \cup B$ is simple. (If A is simple, B is finite.) Let $B_0 = A \times N \cup N \times B$, and let $B_1 = N \times A \cup B \times N$. $B_0 \cup B_1$ is simple, for otherwise there is an infinite r.e. set $C \subset B_0' \cap B_1'$, and this implies that either $\{x \mid (\exists y)[\langle x, y\rangle \in C]\}$ is an infinite r.e. subset of $A' \cap B'$ or $\{y \mid (\exists x)[\langle x, y\rangle \in C]\}$ is an infinite r.e. subset of $A' \cap B'$.

Assume $A \times A$ is a cylinder and let f be the recursive function described in Lemma 2. (So $x \in A \times A$ implies that $\{x, f(x), f^2(x), \cdots\}$ is an infinite r.e. subset of $A \times A$ and $x \in (A \times A)'$ implies that $\{x, f(x), f^2(x), \cdots\}$ is an infinite subset of $(A \times A)'$.)

To obtain a many-one reduction of $A \times A$ to A: Given x, enumerate $\{x, f(x), f^2(x), f^3(x), \cdots\}$, B_0, and B_1. Since $B_0 \cup B_1$ is simple, we must eventually find an integer $\langle c, d\rangle$ either in $\{x, f(x), f^2(x), \cdots\} \cap B_0$ or in $\{x, f(x), f^2(x), \cdots\} \cap B_1$. In the former case define $g(x) = d$; in the latter case define $g(x) = c$. Then $x \in A \times A$ if and only if $g(x) \in A$.

We next modify Theorem 2 to characterize a class of pseudo-creative noncylinders.

COROLLARY 2. *Let A be a r.e. set which is not pseudo-creative. Then $A \times A$ is a pseudo-creative noncylinder if and only if $A \times A \not\leqq_m A$.*

Proof. If A is recursive, $A \times A$ is also recursive and $A \times A$ is many-one equivalent to A. If A is creative, since $A \leqq_1 A \times N \leqq_1 A \times A$, $A \times A$ is also creative and hence many-one equivalent to A. The corollary now follows from Theorem 2 and Corollary 1.

COROLLARY 3. *There exists a pseudo-creative set which is not a cylinder and which is bounded-truth-table reducible to a simple set.*

Proof. By Lemma 5 there is a simple set S such that $S \times S \not\leqq_m S$. Since $S \times S \leqq_{btt} S$, $S \times S$ is the desired set.

Our next theorem shows that Theorem 2 cannot be strengthened to include the pseudo-creative sets.

THEOREM 3. *There is a pseudo-creative set A such that $A \times A$ is a cylinder but $A \times A \not\leqq_m A$.*

Proof. Let S be a simple set such that $S \times S \not\leqq_m S$. Then

$$S \equiv_m S \times N \leqq_1 S \times S \leqq_1 (S \times S) \times N \leqq_1 (S \times N) \times (S \times N) .$$

Let $A = S \times N$. Then $A \times A$ is clearly a cylinder, but $A \times A \leqq_m A$ implies that $S \times S \leqq_m S$, a contradiction. Thus $A \times A \not\leqq_m A$, and by either Lemma 1 or Theorem 2, A is pseudo-creative.

Since any set is many-one equivalent to its cylinder and all creative sets are many-one equivalent, the cylinder of any pseudo-creative set is still pseudo-creative. Thus, since any set is one-to-one reducible to its cylinder, we might hope to subclassify the pseudo-creative sets into cylinders and noncylinders and obtain for the subclassification a result analogous to Lemma 4. In view of the following theorem, such an analogue fails.

THEOREM 4. *There exist pseudo-creative sets A and B such that A is a cylinder and $A \leqq_1 B$, but B is not a cylinder.*

Proof. Let $A = S \times N$ and $B = S \times S$ where S is a simple set such that $S \times S \not\leqq_m S$. By Theorem 2, $S \times S$ is not a cylinder, and by Lemma 6, $A \leqq_1 B$. By Lemma 1 A is pseudo-creative, and by Corollary 1, B is pseudo-creative.

REMARKS. 1. In another paper we shall show that there is a pseudo-creative set which is not a cylinder and which, in contrast to those pseudo-creative noncylinders constructed by using Theorem 2, is not bounded-truth-table reducible either to a simple set or to a pseudo-simple set.

2. The author does not know if there is a simple, pseudo-simple,

or pseudo-creative set A such that $A \times A \leqq_m A$. The question of whether such a set exists is equivalent to the following question: Is it true that if A is a r.e. set, then $A \times A \nleqq_m A$ if and only if A is either recursive or creative?

References

1. P. Fischer, *A note on bounded-truth-table reducibility*, Proc. Amer. Math. Soc., **14** (1963), 875–877.
2. J. Myhill, *Recursive digraphs, splinters and cylinders*, Math. Annalen, **133** (1959), 211–218.
3. E. L. Post, *Recursively enumerable sets of positive integers and their decision problems*, Bull. Amer. Math. Soc., **50** (1944), 284–316.
4. H. Rogers, Jr., *Recursive functions and effective computability*, To be published by McGraw Hill.
5. J. Shoenfield, *Quasicreative sets*, Proc. Amer. Math. Soc., **8** (1957), 964–967.

REED COLLEGE

Pacific Journal of Mathematics

Vol. 14, No. 2 June, 1964