Pacific Journal of Mathematics

IN THIS ISSUE—

Vol. 15, No. 1 **September, 1965**

Printed in Japan by International Academic Printing Co., Ltd., Tokyo Japan

PACIFIC JOURNAL OF MATHEMATICS
Vol. 15, No. 1, 1965

UNIMODULAR SOLUTIONS OF INFINITE SYSTEMS OF LINEAR EQUATIONS

DONALD C. BENSON

It is well known that if a series of real numbers $\sum_{n=1}^{\infty} a_n$ converges, but not absolutely, then for any b, there exists a sequence $\{x_i\}, x_i = \pm 1$, such that $\sum_{n=1}^{\infty} a_n x_n = b$. In § 1, a criterion is given on a system of denumerably many equations of this type, with real coefficients, so that solutions $x_i = \pm 1$ exist for arbitrary right hand sides. A sequence $\{x_i\}$ such that $x_i = \pm 1$ will be called unimodular. In § 2, there results are extended to finite systems, and it is shown that an infinite system has unimodular solutions for arbitrary right hand sides if and only if every finite subsystem has this property. § 3 shows that if a system satisfies the criterion of § 1, then, in a certain sense, "almost any" sequence $\{x_i\}$, $x_i = \pm 1$, "satisfies" the system for any choice of right hand sides. In § 4, conditions are given whereby infinite systems can be constructed which satisfy the criterion of § 2. It follows, for example, that the system

$$\sum_{j=1}^{\infty} (-1)^{[j/2^i]} j^{-\alpha} x_j = b_i \,, \qquad i = 1, 2, \cdots; 0 < \alpha \leqq 1$$

has solutions ($x_i = \pm 1$) for any b_i ($i = 1, 2, \cdots$). The b_i are allowed to be real numbers or $\pm\infty$.

1. The main theorem. THEOREM 1. *Let a_{ij} ($i, j = 1, 2, 3, \cdots$) be real numbers such that there exist x_{jkl} ($j = 1, 2, \cdots; k = 0, 1, 2, \cdots; l = 1, 2, \cdots$) which satisfy the following conditions:*

1. *Each x_{jkl} is equal to $+1$ or -1.*
2. *$\sum_{j=1}^{\infty} a_{ij} x_{jkl}$ converges for all i such that $i \neq k$ and $i \leqq l$.*
3. *$\sum_{j=1}^{\infty} a_{ij} x_{jil}$ diverges to $+\infty$.*

Then, for any sequence $\{b_i\}$, the infinite system of equations

$$\sum_{j=1}^{\infty} a_{ij} x_j = b_i \tag{1}$$

can be solved such that for each i, $x_i = \pm 1$. Here, b_i is allowed to be either a real number or $\pm\infty$.

Proof. If $k \neq i \leqq l$, for any $\varepsilon > 0$ there exists $N(\varepsilon; i, k, l)$ such

Received January 22, 1964.

that

$$\left|\sum_{j=m+1}^{n} a_{ij}x_{jkl}\right| < \varepsilon \text{ and } |a_{in}| < \varepsilon/2 , \tag{2}$$

provided $m, n > N(\varepsilon; i, k, l)$.

We define the solution $\{x_i\}$ inductively along with positive integers M_{nm} which will be defined whenever n is a positive integer and m is a nonnegative integer such that $m \leqq n$. The ordered pairs (n, m) are ordered lexicographically, i. e., $(n, m) < (n_1, m_1)$ if and only if either $n < n_1$, or $n = n_1$ and $m < m_1$. The induction will be with respect to this order.

The following definitions will be used with $m \leqq n$ and $i \leqq n$:

$$B_{inm} = \begin{cases} b_i \text{ if } b_i \text{ is finite} \\ \pm n \text{ if } b_i = \pm\infty \text{ and } i \leqq m , \\ \pm(n-1) \text{ if } b_i = \pm\infty \text{ and } i > m \end{cases} \tag{3}$$

$$A_{inm} = \begin{cases} b_i \text{ if } b_i \text{ is finite} \\ B_{inm} \text{ if } i \neq m , \\ \pm(n-1) \text{ if } i = m \text{ and } b_i = \pm\infty \end{cases} \tag{4}$$

$$\delta_{inm} = \begin{cases} (2(m-i)+1)/n^2 \text{ if } i < m \\ 2(n-i)/(n-1)^2 + 2m/n^2 \text{ if } i \geqq m . \end{cases} \tag{5}$$

Let us suppose that positive integers M_{nm} have been defined for $(n, m) \leqq (s, t)$, and x_i for $i \leqq M_{st}$ such that the following conditions are satisfied:

(A) $M_{nm} < M_{pq}$ if and only if $(n, m) < (p, q)$.

(B) $M_{n0} \geqq N(1/n^2; i, k, n)$ for all $i, k \leqq n$ $(i \neq k)$.

(C) $\left|\sum_{j=1}^{M_{nm}} a_{mj}x_m - B_{mnm}\right| < 1/n^2$ where $m \neq 0$.

(D) If $i \leqq n-1$ and $M_{nm} \geqq p > \begin{cases} M_{n,m-1} \text{ if } m \neq 0 \\ M_{n-1,n-1} \text{ if } m = 0, \end{cases}$

then

(D1) $A_{inm} - \delta_{inm} < \sum_{j=1}^{p} a_{ij}x_j < B_{inm} + \delta_{inm}$ if $A_{inm} \leqq B_{inm}$

and

(D2) $B_{inm} - \delta_{inm} < \sum_{j=1}^{p} a_{ij}x_j < A_{inm} + \delta_{inm}$ if $B_{inm} \leqq A_{inm}$.

We wish to determine M_{uv} where (u, v) is the immediate successor of (s, t), and x_i, $i = M_{st}+1, \cdots, M_{uv}$, such that the conditions (A)–(D) are valid for all $(n, m) \leqq (u, v)$. There are two cases to consider. Either we have $s = t$, or $s > t$.

Case I. $s = t$. In this case the immediate successor of (s, t) is $(s+1, 0)$. Putting $M_{s+1,0}$ equal to the largest of the numbers $M_{ss}+1$ and $N(1/(s+1)^2; i, k, s+1)$ for all $i, k \leqq s+1$ $(i \neq k)$, we see that (A) and (B) will be satisfied.

We now put $x_j = x_{jos}$, $(j = M_{ss}+1, \cdots, M_{s+1.0})$. Condition (C) remains satisfied because the newly defined quantities do not occur in (C). Condition (B) holds with $n = s$ by the inductive assumption. Therefore, $|\sum_{j=k+1}^{l} a_{ij}x_{jos}| < 1/s^2$, provided $k, l > M_{ss}$ and $i \leqq s$. We have $\delta_{iss} = (2(s-i)+1)/s^2$, $\delta_{i,s+1,0} = 2(s+1-i)/s^2$, and hence $\delta_{i,s+1,0} - \delta_{iss} = 1/s^2$. From the equality $A_{iss} = B_{iss} = A_{i,s+1,0} = B_{i,s+1,0}$ for $i < s$, we see that (D) holds with $i < s$. It must be shown that (D) holds with $i = s$. We have also $B_{sss} = B_{s,s+1,0} = A_{s,s+1,0}$. Recall that

$$\left|\sum_{j=M_{ss}}^{l} a_{ij}x_{jos}\right| < 1/s^2 .$$

Since $\delta_{s,s+1,0} = 2/s^2$ and (C) holds with $i = s$, the result follows, namely that (D) holds with $i = s$. This disposes of Case I.

Case II. $s > t$. The immediate successor of (s, t) is $(s, t+1)$. We use the fact that $\sum_{j=1}^{\infty} a_{t+1,j}x_{j,t+1,s} = +\infty$.

Subcase IIA. For some $l > M_{st}$ we have

$$\left|\sum_{j=1}^{l} a_{t+1,j}x_{jos} - B_{t+1,s,t+1}\right| < 1/s^2 .$$

In this case, we put $l = M_{s,t+1}$ and $x_j = x_{jos}$, $j = M_{st}+1, \cdots, M_{s,t+1}$.

Subcase IIB. If the above never happens, then

$$\sum_{j=1}^{l} a_{t+1,j}x_{jos} - B_{t+1,s,t+1}$$

must keep the same sign for all $l > M_{st}$, because, for $j > M_{st}$, we have $|a_{t+1,j}| < 1/2s^2$, from inductive assumption (B).

Let $\sigma = \pm 1$, depending upon whether the sign stays $+$ or $-$. Because the series $\sum_{j=1}^{\infty} a_{t+1,j}x_{j,t+1,s}$ diverges to $+\infty$, there exists $K > M_{st}$ such that

$$(6) \qquad \sum_{j=M_{st}+1}^{l} a_{t+1,j}x_{jos} + \sum_{j=M_{st}+1}^{l} a_{t+1,j}x_{j,t+1,s} > 0$$

for all $l > K$. Let K_0 be the smallest number K with this property. We put $x_j = x_{jos}$, $M_{st} < j < K_0$, and $x_j = -x_{j,t+1,s}$, $j = K_0, \cdots, M_{s,t+1}$, where the integer $M_{s,t+1}$ will now be defined. Because $\sum_{j=1}^{\infty} a_{t+1,j}x_{j,t+1,s} = +\infty$, and $|a_{t+1,j}| < 1/2s^2$ for $j > M_{st}$, there exist integers $M > M_{st}$

such that (C) is satisfied for $n = s$, $m = t + 1$ and $M_{s,t+1} = M$. Let $M_{s,t+1}$ be the smallest integer with the above property.

It will be shown that conditions (A)–(D) hold for both subcases. Conditions (A) and (B) are evident. Condition (C) follows immediately from the construction above. Condition (D) is somewhat more difficult.

It will be shown first that (D) holds with $i = t + 1$, $n = s$, $m = t + 1$. We may suppose $i \leqq n - 1$, i.e., $t + 1 \leqq s - 1$. We have

$$\delta_{t+1,s,t+1} - \delta_{t+1,s,t} = 2/s^2 \,, \tag{7}$$

and

$$B_{t+1,s,t} = A_{t+1,s,t} = A_{t+1,s,t+1} \,.$$

From inductive assumption (D), we have

$$\left|\sum_{j=1}^{M_{st}} a_{t+1,j}x_j - A_{t+1,s,t+1}\right| < \delta_{t+1,s,t+1} - 2/s^2 \,. \tag{8}$$

Thus, in Subcase IIA with $M_{st} < p \leqq M_{s,t+1}$, and in Subcase IIB, with $M_{st} < p < K_0 < M_{s,t+1}$, we have

$$\left|\sum_{j=1}^{p} a_{t+1,j}x_j - A_{t+1,s,t+1}\right| < \delta_{t+1,s,t+1} - 1/s^2 \,.$$

This disposes of Subcase IIA, because the above inequality is stronger than (D). It must be shown now that (D) holds for $K_0 \leqq p \leqq M_{s,t+1}$, with $i = t + 1$, $n = s$, $m = t + 1$.

From the inequalities,

$$\sigma \sum_{j=M_{st}+1}^{K_0} a_{t+1,j}x_{jos} + \sum_{j=M_{st}+1}^{K_0} a_{t+1,j}x_{j,t+1,s} > 0 \,,$$

$$\sigma \sum_{j=M_{st}+1}^{K_0-1} a_{t+1,j}x_{jos} + \sum_{j=M_{st}+1}^{K_0-1} a_{t+1,j}x_{j,t+1,s} \leqq 0 \,,$$

(where an empty sum is taken to equal zero), it follows that we have $(p \leqq M_{s,t+1})$

$$\begin{aligned} \sum_{j=M_{st}+1}^{K_0-1} a_{t+1,j}x_{jos} - \sum_{j=K_0}^{p} a_{t+1,j}x_{j,t+1,s} \\ &= \sigma \sum_{j=M_{st}+1}^{p} a_{t+1,j}x_j \\ &< - \sum_{j=M_{st}+1}^{p} a_{t+1,j}x_{j,t+1,s} + 2\,|\,a_{t+1,K_0}| \\ &< - \sum_{j=M_{st}+1}^{p} a_{t+1,j}x_{j,t+1,s} + 1/s^2 \end{aligned} \tag{10}$$

using (B) and the definition of x_j, $M_{st} < j \leqq M_{s,t+1}$. Now, from (6), we have

(11) $$-\sum_{j=M_{st}+1}^{p} a_{t+1,j}x_{j,t+1,s} < -\sigma \sum_{j=M_{st}+1}^{p} a_{t+1,j}x_{jos} .$$

Combining (1), (11) and (B), we have

(12) $$\sigma \sum_{j=M_{st}+1}^{p} a_{t+1,j}x_j < -\sigma \sum_{j=M_{st}+1}^{p} a_{t+1,j}x_{jos} + 1/s^2 < 2/s^2 .$$

From (8), we have

(13) $$\sigma \sum_{j=1}^{Mst} a_{t+1,j}x_j < \sigma A_{t+1,s,t+1} + \delta_{t+1,s,t+1} - 1/s^2 .$$

Putting (12) and (13) together, we have

(14) $$\sigma \sum_{j=1}^{p} a_{t+1,j}x_j < \sigma A_{t+1,s,t+1} + \delta_{t+1,s,t+1} .$$

From the definition of σ and the fact that if $A_{t+1,s,t+1} \neq B_{t+1,s,t+1}$, then $|A_{t+1,s,t+1} - B_{t+1,s,t+1}| = 1$, it follows that $\sigma(A_{t+1,s,t+1} - B_{t+1,s,t+1}) \geqq 0$. Thus (D1) or (D2) must be demonstrated, depending on whether $\sigma = -1$ or $+1$.

If $\sigma = -1$, we have from (14)

$$A_{t+1,s,t+1} - \delta_{t+1,s,t+1} < \sum_{j=1}^{p} a_{t+1,j}x_j$$

which is half of (D1). From the definition of $M_{s,t+1}$, we have

$$\sum_{j=1} a_{t+1,j}x_j < B_{t+1,s,t+1} + 1/s^2 < B_{t+1,s,t+1} + \delta_{t+1,s,t+1},$$

provided $M_{st} < p \leqq M_{s,t+1}$, which gives the remaining half of (D1). Similar considerations show that if $\sigma = 1$, then (D2) holds. This concludes the demonstration that (D) holds with $i = t + 1, n = s, m = t + 1$.

It remains to show that (D) holds for $i \neq t + 1, n = s, m = t + 1$, $i \leqq s - 1$.

We have $\delta_{i,s,t+1} - \delta_{i,s,t} = 2/s^2$, so that it is sufficient to show

(15) $$\left| \sum_{j=M_{st}+1}^{p} a_{ij}x_j \right| < 2/s^2$$

for $M_{st} < p \leqq M_{s,t+1}$.

In Subcase IIA, we have, using (B),

(16) $$\left| \sum_{j=M_{st}+1}^{p} a_{ij}x_j \right| = \left| \sum_{j=M_{st}+1}^{p} a_{ij}x_{jos} \right| < 1/s^2 .$$

In Subcase IIB with $p < K_0$, (15) is valid once again because (16) holds.

In Subcase IIB with $p \geqq K_0$, we have

$$\left|\sum_{j=M_{st}+1}^{p} a_{ij}x_j\right| \leqq \left|\sum_{j=M_{st}+1}^{K_0-1} a_{ij}x_{jos}\right| + \left|\sum_{j=K_0}^{p} a_{ij}x_{j,t+1,s}\right| < (1/s^2) + (1/s^2) \tag{17}$$

because of (B). This concludes the demonstration of Case II.

The definition by induction is completed by setting $M_{01} = N(1; 1, 0, 1)$ and observing that thereby (A)–(D) are satisfied with $n = 1, m = 0, i = 1$.

It remains to show the sequence $\{x_i\}$ constructed in this way satisfies the infinite system of equations (1). However, this follows from the fact that $\{x_i\}$ satisfies condition (D).

2. Systems with finitely many equations. The extension of Theorem 1 to systems with finitely many equations is accomplished by producing an infinite system which can be treated by Theorem 1 and which is equivalent to the given finite system.

THEOREM 2. *Let a_{ij} $(i = 1, \cdots, R; j = 1, 2, \cdots)$ be real numbers such that there exist x_{jk} $(j = 1, 2, \cdots; k = 0, 1, \cdots, R)$ which satisfy the following conditions:*

1. *Each x_{jk} is equal to $+1$ or -1.*
2. *$\sum_{j=1}^{\infty} a_{ij}x_{jk}$ converges for all i such that $i \neq k$.*
3. *$\sum_{j=1}^{\infty} a_{ij}x_{ji}$ diverges to $+\infty$.*

Then, for any numbers $b_1, \cdots, b_R$, each of which is a real number or $\pm\infty$, the equations

$$\sum_{j=1}^{\infty} a_{ij}x_j = b_i\,, \qquad (i = 1, \cdots, R) \tag{18}$$

can be solved such that for each i, $x_i = \pm 1$.

Proof. We construct an auxiliary infinite system of equations

$$\sum_{j=1}^{\infty} \alpha_{ij}\xi_j = \beta_i \qquad (i = 1, 2, \cdots)\,. \tag{19}$$

We define $\beta_{i+nR} = b_i$ for any nonnegative integer n, and

$$\alpha_{i+nR,l} = \begin{cases} a_{ik} \text{ if there exists } k > 0 \text{ such that} \\ \quad l = T(k + n - 1) + n - 1 \text{ and} \\ 0 \text{ otherwise,} \end{cases} \tag{20}$$

where $T(n) = n(n + 1)/2$ is the nth triangular number. The fact to be used about $T(n)$ is that each positive integer has one and only one representation in the form $T(k+n-1)+n-1=S(k,n)$, where k and n are positive integers.

Let us define $\xi_{jkl}=\xi_{jk}$ as follows:

$$(21)\quad \xi_{l,i+nR}=\begin{cases} x_{ki} \text{ if there exists } k>0 \text{ such that } l=S(k,n), \text{ and} \\ x_{k0} \text{ if } l=S(k,m),\ n\neq m>0\ . \end{cases}$$

Then we have

1. Each ξ_{jk} is equal to $+1$ or -1.
2. $\sum_{j=1}^{\infty}\alpha_{ij}\xi_{jk}$ converges for all i such that $i\neq k$.
3. $\sum_{j=1}^{\infty}\alpha_{ij}\xi_{ji}$ diverges to $+\infty$.

The hypotheses of Theorem 1 are satisfied, and therefore the system (19) has a solution $\{\xi_j\}$. Then $x_j=\xi_{T(j)}$, $j=1,2,\cdots$, is a solution of (18).

COROLLARY. *The system* (1), *with arbitrary right hand sides, has a unimodular solution if and only if every finite subsystem of* (1), *with arbitrary right hand sides has a unimodular solution.*

A system of nondenumerably many equations of the type described in Theorem 1 will never have unimodular solutions for all possible right hand sides, because the number of ways in which the right hand sides could be prescribed would have cardinality greater than C, whereas the cardinality of all unimodular sequences, $x_i=\pm1$, $i=1,2,\cdots$, is equal to C. (Here C denotes the cardinality of the continuum.)

3. The metric space $\mathscr{M}$. The set of sequences $\{x_i\}$, $x_i=\pm1$ form a complete metric space under the metric

$$d(\{x_i\},\{x_i'\})=1/l\ ,$$

where $l=\min\{i: x_i\neq x_i'\}$.

Let a_{ij} satisfy the hypotheses of Theorem 1. Let U_i, $i=1,2,\cdots$, be nonempty open sets of extended real numbers. (U_i may contain $+\infty$ or $-\infty$.) Let $\mathscr{N}_M$ be the set of sequences $\{x_i\}$ such that for all $N\geqq M$, $\sum_{j=1}^{N}a_{ij}x_j\notin U_i$ for some i $(0<i\leqq M)$.

$\mathscr{N}_M$ is closed. For suppose $\{x_i^n\}\in\mathscr{N}_M$ and $\lim_{n\to\infty}d(\{x_i^n\},\{x_i\})=0$. Also, suppose there exists $N\geqq M$ such that $\sum_{j=1}^{N}a_{ij}x_j\in U_i$ for each i $(0<i\leqq M)$. For sufficiently large n, we have $x_j=x_j^n$, $j=1,\cdots,N$, and hence we get $\sum_{j=1}^{N}a_{ij}x_j^n\in U_i$ $(0<i\leqq M)$, contrary to the assumption $\{x_i^n\}\in\mathscr{N}_M$.

$\mathscr{N}_M$ is nowhere dense. For suppose $\{x_i\}\in\mathscr{N}_M$. Let b_i be an arbitrary element of U_i. For any $P > 0$, there exists, because of Theorem 1, a sequence $\{x_i\}$ such that

1. $x'_i = x_i$, $i = 1, 2, \cdots, P$, and
2. $\sum_{j=1}^{\infty} a_{ij}x'_j = b_i$.

Clearly, $\{x'_i\} \notin \mathscr{N}_M$ and $d(\{x_i\}, \{x'_i\}) < 1/P$.

Thus the set $\bigcup_{M=1}^{\infty} \mathscr{N}_M = \mathscr{N}$ is of the first category, and since $\mathscr{M}$ is a complete metric space, $\mathscr{M} - \mathscr{N}$ is of the second category. We have proved the following:

LEMMA 1. *For any sequence $\{x_n\}$ in $\mathscr{M} - \mathscr{N}$ there exists an infinite monotone increasing sequence $\{N_k\}$ of positive integers such that for each k, $\sum_{j=1}^{N_k} a_{ij}x_j \in U_i$ for $i \leqq k$.*

For any sequence $\{b_i\}$ of extended real numbers we may take U_i^n as follows:

$$U_i^n = \begin{cases} \{x: |x - b_i| < 1/n\} & \text{if } b_i \text{ is finite} \\ \{x: \pm(x - b_i) > n\} & \text{if } b_i = \pm\infty \,. \end{cases}$$

By applying the lemma for each n to $\{U_i^n\}$ $0 < i < \infty$, we find that there exists a monotone increasing sequence of positive integers $\{S_k\}$ such that

$$\sum_{j=1}^{S_k} a_{ij}x_j \in U_i^k \qquad \text{for } i \leqq k \,. \tag{23}$$

From (22), it now follows that we have

$$\lim_{k\to\infty} \sum_{j=1}^{S_k} a_{ij}x_j = b_i \qquad \text{for every } i > 0 \,.$$

In summary, this proves the following:

THEOREM 3. *Let a_{ij} satisfy the hypotheses of Theorem 1. Then there exists a sequence $\{x_i\}$, $x_i = \pm 1$, with the following property. (Indeed, any sequence $\{x_i\}$ in the complete metric space $\mathscr{M}$, apart from a certain set of first category, has this property.) For any sequence $\{b_i\}$ of extended real numbers, there exists a sequence of positive integers $\{S_k\}$ such that for each i,*

$$\lim_{k\to\infty} \sum_{j=1}^{S_k} a_{ij}x_j = b_i \,.$$

4. Sufficient conditions. In this section we shall find sufficient conditions on the coefficients a_{ij} so that the hypotheses of Theorem 1 are satisfied.

THEOREM 4. *Let $\{a_i\}$ be a sequence of real numbers such that*

(i) $a_i > 0$

(ii) $\sum_{i=1}^{\infty} a_i = \infty$.

(iii) *For every* $k \geqq 1$, $a_i - a_{j+k}$ *is monotone decreasing in* i.

(iv) a_i *tends monotonically to zero as* $i \to \infty$.

Let $a_{ij} = (-1)^{[j/2^i]} a_j$. *Then* a_{ij} *satisfies the hypotheses of Theorem* 1.

Proof. We must find sequences $\{x_{jkl}\}$, $1 \leqq j < \infty$, $0 \leqq k < \infty$, $1 \leqq l < \infty$, such that conditions (1), (2) and (3) of Theorem 1 are fulfilled.

$k = 0$. First we show that by putting $x_{j0l} = 1$, the conditions are satisfied for $k = 0$. Condition (3) is fulfilled vacuously and condition (1) is trivial.

It will be shown that condition (2) holds, i.e., that $\sum_{j=1}^{\infty} (-1)^{[j/2^i]} a_j$ converges for each i. Let

(24) $$(-1)^k b_k = \sum_{j=k\cdot 2^i}^{(k+1)2^i - 1} (-1)^{[j/2^i]} a_j \ .$$

Then we have

(25) $$b_k = \sum_{j=k\cdot 2^i}^{(k+1)2^i - 1} a_j > 0 \ .$$

From (iv), b_k is monotone decreasing, and hence $\sum (-1)^k b_k$ converges. The condition (2) follows because

(26) $$\sum_{j=1}^{k} (-1)^j b_j = \sum_{j=0}^{(k+1)2^i - 1} (-1)^{[j/2^i]} a_j \ .$$

$k \neq 0$. Let $x_{jkl} = (-1)^{[j/2^k]}$. Since it is assumed that $\sum_{i=1}^{\infty} a_i = \infty$, we have $\sum_{j=1}^{\infty} a_{ij} x_{jil} = \infty$, and thus condition (3) holds.

We will show (2) holds and thereby complete the proof by showing that

(27) $$\sum_{j=1}^{\infty} (-1)^{[j/2^i] + [j/2^k]} a_j$$

converges if $i > k$. We have

(28) $$[j/2^i] + [j/2^k] = [(j + 2^i)/2^i] + [(j + 2^i)/2^i] - 1 - 2^{i-k}$$

and

(29) $$(-1)^{[j/2^i] + [j/2^k]} = -(-1)^{[(j+2^i)/2^i] + [(j+2^i)/2^k]} \ .$$

Putting

(30) $$(-1)^n c_n = \sum_{j=n\cdot 2^i}^{(n+1)2^i - 1} (-1)^{[j/2^i] + [j/2^k]} a_j \ ,$$

we have

$$c_n = \sum_{j=n\cdot 2^i}^{(n+1)2^i-1} (-1)^{[j/2^k]} a_j \\ = \sum_{j=n\cdot 2^i}^{n\cdot 2^i+2^k-1} \sum_{m=0}^{2^{i-k-1}-1} (a_{j+2m\cdot 2k} - a_{j+(2m+1)\cdot 2k}) . \tag{31}$$

Evidently c_n is positive. Also, c_n is monotone because, from (iii), $a_i - a_{i+2^k}$ is monotone decreasing in i. Thus $\sum_{n=1}^{\infty}(-1)^n c_n$ converges. Since we have

$$\sum_{j=1}^{n}(-1)^j c_j = \sum_{j=0}^{n\cdot 2^i-1}(-1)^{[j/2^i]+[j/2^k]} a_j \tag{32}$$

it follows that (27) converges if $i > k$. This concludes the proof of Theorem 4.

The sequence $a_i = 1/i^\alpha$, for positive $\alpha \leqq 1$, is an example satisfying (i)–(iv) of Theorem 4.

This result can be extended with the help of Abel's test for convergence.

THEOREM 5. *Let $\{a_i\}$ satisfy the hypotheses of Theorem 4. Let $\{v_{ij}\}$, $i, j = 1, 2, \cdots$ satisfy the following:*

1. $v_{ij} > 0$.
2. *For each i, $\{v_{ij}\}$ is monotone (increasing or decreasing) with respect to j.*
3. $\sum_{j=1}^{\infty} a_j v_{ij} = \infty$ *for each i.*

Then $(-1)^{[j/2^i]} a_j v_{ij}$ satisfies the hypotheses of Theorem 1.

Proof. We take the same definition for x_{jk} as in Theorem 4. Then $\sum_{j=1}^{\infty} a_{ij} x_{jk}$ converges for $i \neq k$ by Abel's test. Further, we have

$$\sum_{j=1}^{\infty} a_{ij} x_{jil} = \sum_{i=1}^{\infty} a_j v_{ji} = +\infty .$$

We obtain a result which allows us to transform any array of coefficients a_{ij} which satisfies the hypotheses of Theorem 1 into a different array satisfying the same conditions. First we need a lemma which is related to Abel's test for convergence.

LEMMA 2. *Let $\{v_i\}$ be a monotone decreasing sequence of real numbers which is bounded away from zero; i.e., there exists b such that $0 < b \leqq v_i$ for all i. Suppose $\sum_{i=1}^{\infty} a_i = +\infty$. Then $\sum_{i=1}^{\infty} v_i a_i = +\infty$.*

Proof. Let $s_n = \sum_{i=1}^{n} a_i$, and let $h_n = \inf_{k \geqq n} s_k$. We have, if $m \leqq p$,

$$\begin{aligned}(33)\qquad \sum_{i=1}^{p} a_i v_i &= s_1(v_1 - v_2) + \cdots + s_{p-1}(v_{p-1} - v_p) + s_p v_p \\ &\geqq h_1(v_1 - v_m) + h_m v_m \\ &\geqq -|h_1| v_1 + h_m b\ .\end{aligned}$$

The result follows because $h_m \to \infty$ as $m \to \infty$.

THEOREM 6. *Let a_{ij} satisfy the hypotheses of Theorem 1. Let v_{ij} satisfy the following:*

1. *There exist c_i such that $0 < c_i \leqq v_{ij}$ for all positive integers i and j.*
2. *For each i, $\{v_{ij}\}$ is monotone decreasing with respect to j.*

Then $a_{ij}v_{ij}$ satisfy the hypotheses of Theorem 1.

Proof. The conditions are satisfied by using the x_{jkl} which are assumed to exist in Theorem 1. We have that $\sum_{j=1}^{\infty} a_{ij}v_{ij}x_{jkl}$ converges if $i \neq k$ by Abel's test and diverges to $+\infty$ for $i = k$ by Lemma 2.

UNIVERSITY OF CALIFORNIA, DAVIS

PACIFIC JOURNAL OF MATHEMATICS
Vol. 15, No. 1, 1956

TRANSITIVE GROUPS OF COLLINEATIONS ON CERTAIN DESIGNS

RICHARD E. BLOCK

Let $M = (a_{ij})$ be an $m \times n$ matrix with entries in $\{1, -1\}$. Suppose that there is a positive integer d such that the inner product of every pair of distinct rows of M is $n - 2d$; this is equivalent to assuming that any two distinct rows have Hamming distance d, i.e. differ in exactly d places. The rows of M form the code words of a binary code; such a code is called a (binary) *constant-distance code*, of length n and distance d. Special cases of matrices which may be taken to be M are the Hadamard matrices, which are defined by the condition that $m = n = 2d$, and the incidence matrices (written with ± 1) of balanced incomplete block designs, which are characterized by the property that all column sums are equal and all row sums are equal.

Suppose that π is a permutation of $\{1, \cdots, n\}$ such that replacement, for $i = 1 \cdots, n$, of the $\pi(i)$th column of M by the ith column of M sends each row of M into a row of M. Then π induces a permutation of the rows of M. Call such a pair of permutations of the columns and of the rows a collineation of M, or of the code. We shall examine constant-distance codes with a group G of collineations which is transitive on the columns. We shall show that G has at most two orbits on the rows (just one orbit if and only if M comes from a balanced incomplete block design), and that if G is nilpotent then at most one of these orbits contains more than a constant row.

Moreover, it will be shown that this last conclusion need not hold if G is not assumed nilpotent; this will be done by giving an infinite class of Hadamard matrices with doubly transitive collineation groups.

One way of obtaining a constant-distance code with a transitive group on the columns is the following. Given a (cyclic) (v, k, λ) difference set, write a v-tuple of 1's and -1's with 1 in the k places which corresponds to elements of the difference set, and repeat this v-tuple s times to obtain a vs-tuple. The set of all cyclic permutations of this vs-tuple forms constant-distance code with v code words and distance $d = 2(k - \lambda)s$. Call such a code an *iterated difference set code*. The code is closed under the cyclic shift (the permutation $\pi = (1, 2, \cdots, vs)$ on the columns).

Our results imply that, conversely, any constant-distance code which is closed under the cyclic shift consists of repeated cyclic shifts of

Received December 20, 1963.

some single word, plus possibly a single constant word. The main part of the code is thus an iterated difference set code; the extra word can occur if and only if the parameters (v, k, λ) are of Hadamard type.

2. The number of orbits on the rows.

THEOREM 1. *Suppose that G is a group of collineations of a constant-distance code. If G is transitive on the columns then G has at most two orbits on the rows.*

Proof. Suppose that G has t orbits $T_1, \cdots, T_t$ on the rows. Then there are integers r_i such that each row in T_i has exactly r_i 1's, $i = 1, \cdots, t$. It follows that if α_i and α_j are rows and $\alpha_i \in T_i$, $\alpha_j \in T_j$, and if $c(\alpha_i, \alpha_j)$ is the number of places in which both α_i and α_j have 1, then $r_i + r_j = d + 2c(\alpha_i, \alpha_j)$, or $c(\alpha_i, \alpha_j) = (r_i + r_j - d)/2$. Let v_i denote the number of words in T_i. Since G is transitive on the columns, for each column there are the same number k_i of words in T_i with 1 in that place; we have $k_i = v_i r_i/n$, where n is the length of the words. Thus the words in T_i form the incidence matrix of a balanced incomplete block design with $\lambda = r_i - (d/2)$. Now suppose that $t \geqq 2$, that T_i and T_j are distinct orbits and that $\alpha \in T_j$. Counting in two ways the total number of times in which words in T_i have a 1 in the same place as a 1 in α, we have $v_i(r_i + r_j - d)/2 = r_j k_i$. Thus, since $k_i = v_i r_i/n$,

$$n\frac{(r_i + r_j - d)}{2} = r_i r_j \,. \tag{1}$$

Suppose that, $r_i \neq n$. Then for some prime p, with p^e and p^f the highest powers of p dividing n and r_i, respectively, one has $e > f$. Since $v_i r_i = n k_i$ and

$$r_i(k_i - 1) = \left(r_i - \frac{d}{2}\right)(v_i - 1)\,, \tag{2}$$

$p \nmid (v_i - 1)$ and $p^f \mid r_i - (d/2)$. If $r_i = r_j$ then the left side of (1) is divisible by p^{e+f}, the right side only by p^{2f}, a contradiction. Hence $r_i \neq r_j$ if $i \neq j$. Also $r_i \neq n/2$, since otherwise, by (1), $r_i = n/2 = d$ and $k_i = v_i/2$, contradicting (2). Thus r_j is uniquely determined in terms of r_i by (1). It follows that $t \leqq 2$, and the theorem is proved.

If there is only one orbit, then, as shown in the above proof, M is the incidence matrix of a balanced incomplete block design. The next result is the converse.

THEOREM 2. *Suppose that G is a group of collineations of a balanced incomplete block design. If G is transitive on the blocks then G is also transitive on the points.*

Proof. The incidence matrix of the design is a constant-distance code with $d = 2(r - \lambda)$. If G had two orbits on the points, then $r_1 = r_2 = r$. But by the proof of Theorem 1, $r_1 \neq r_2$, a contradiction. This proves Theorem 2.

COROLLARY 1. *Let G be a group of collineations of a constant-distance code. Suppose that G fixes c columns and is transitive on the remaining columns. Let q be the number of different c-tuples in the rows of the submatrix formed by the c fixed columns. Then G has at most 2 q orbits on the rows; if moreover the code corresponds to a balanced incomplete block design, then G has exactly q orbits on the rows (points).*

Proof. The set of rows with a given c-tuple in the fixed columns must be closed under G; deleting the fixed columns from these rows, one obtains a constant distance code with a transitive group of collineations. The result now follows immediately from Theorems 1 and 2.

These results are a partial generalization to nonsymmetric designs of a theorem proved by Dembowski [2], Hughes [3], and Parker [4], which says that for a symmetric design, the number of orbits on the points is the same as the number of orbits on the lines. However there are balanced incomplete block designs with a group of collineations which is transitive, even cyclic, on the points, but not transitive on the lines.

3. Codes with a nilpotent transitive group. In this section we assume that M is an $m \times n$ matrix whose rows form a constant-distance code with distance d, and that G is a group of collineations which is transitive on the columns. Let H denote the subgroup of G fixing the first column. We shall continue using the notation T_i, v_i, r_i and k_i introduced in the above proofs.

THEOREM 3. *Suppose that T_1 and T_2 are distinct orbits of G (on the rows). For $i = 1, 2$, take α_i in T_i and let S_i be the subgroup of G fixing α_i. Suppose that p is any prime such that the highest power p^j of p dividing n does not divide d. Then, either for $i = 1$ or 2, S_i contains the normalizer of a Sylow p-subgroup of G, $p \mid v_i - 1$, and $p^j \mid r_i$.*

Proof. If the orbit T_i is trivial (consists of a constant word) then $S_i = G$ and the conclusion is obvious. Thus suppose that both orbits are nontrivial. Take a prime p such that p^j, the highest power of p dividing n, does not divide d. Let p^e and p^f be the highest powers of p dividing r_1 and r_2, respectively; by choice of notation we may suppose that $e \leqq f$. By (1), $p^i \mid r_1 r_2$.

Suppose first that $p \nmid v_1 - 1$ and $p \nmid v_2 - 1$. Then by (2), $p^e \mid [r_1 - (d/2)]$ and $p^f \mid [r_2 - (d/2)]$, so that $p^f \mid (d/2)$ and $p^e \mid r_1 + r_2 - d$. If $p > 2$ then p^{j+e} divides the left side of (1) while p^{e+f} is the highest power of p dividing the right side; hence $f \geqq j$, so that $p^j \mid d$, a contradiction. If $p = 2$ then $p^{e-1} \mid [(r_1 + r_2 - d)/2]$ and p^{j+e-1} divides the left side of (1), so that $f \geqq i - 1$, $p^{j-1} \mid (d/2)$ and $p^j \mid d$, again a contradiction.

Hence $p \mid v_i - 1$ for some i, with $i = 1$ or 2. Then since $p \mid ([G : S_i] - 1)$, $p \nmid [G : S_i]$ and S_i contains a Sylow p-subgroup of G. Suppose that K is any subgroup of G, and consider the orbits of K when K is regarded as a permutation group on the columns. For each of these orbits there is an x in G such that the number of elements in the orbit is $[K : K \cap xHx^{-1}]$. If p^l is the highest power of p dividing $|H|$ then p^{j+l} is the highest power of p dividing $|G|$. Hence if K contains a Sylow p-subgroup of G then $p^j \mid [K : K \cap xHx^{-1}]$ for any x. Taking $K = S_i$ we see that $p^j \mid r_i$, since the set of places where α_i has 1 is a union of orbits of S_1 (on the columns). If $g \in G$ and $g \notin S_i$ then $g\alpha_i \neq \alpha_i$, and gS_ig^{-1} is the subgroup of G fixing $g\alpha_i$. If moreover gS_ig^{-1} contains a Sylow p-subgroup of S_i, then p^j divides the number of elements in each orbit (on the columns) of $S_i \cap gS_ig^{-1}$. But the set of places where α_i and $g\alpha_i$ disagree is a union of orbits of $S_i \cap gS_ig^{-1}$, so that $p^j \mid d$, a contradiction. Therefore no Sylow p-subgroup of S_i is contained in a conjugate of S_i. Suppose that P is a Sylow p-subgroup of S_i (and so also of G), and that $x \in N_G(P)$, the normalizer of P. If $x \notin S_i$ then $xS_1x^{-1} \neq S_i$ but $P = xPx^{-1} \subseteqq xS_ix^{-1}$, a a contradiction. Hence $N_G(P) \subseteqq S_i$, and the theorem is proved.

COROLLARY 2. *If G is a nilpotent group of collineations of M which is transitive on the columns, then either G is transitive on the rows or one of the two orbits of G on the rows consists of one trivial row.*

Proof. Unless M has only the two trivial rows, there is a prime p such that the highest power of p dividing n does not divide d. Since a Sylow p-subgroup of a nilpotent group is normal, if G is not transitive on the rows then by Theorem 3, G fixes a row. This proves the result.

Now suppose the constant distance code is closed under the cyclic shift $\pi = (1, 2, \cdots, n)$. If α is a code word with r ones, then α must be periodic of (minimal) period v, a divisor of n; write $v = n/s$.

A single period of α gives a (v, k, λ) difference set with $k = r/s$ and $\lambda = [r - (d/2)]/s$. Thus the set of cyclic shifts $\pi^t\alpha$ or α forms an s-times iterated (v, k, λ)-difference set code; solving $k(k-1) = \lambda(v-1)$ for s, one has $s = n + [2r(r-n)/d]$. By Corollary a, either this set is the entire code or there is one more word, with all 1's or all -1's. If the extra word has all -1's then $r = d$, $\lambda = d/2s$, and from $k(k-1) = \lambda(v-1)$ one obtains $n/s = 2d/s$. Hence, with $d/2s = u$, one would have $v = 4u - 1$, $k = 2u$ and $\lambda = u$. If on the other hand the extra word has all 1's, then we have the complement of a code of the above type, and $v = 4u - 1$, $k = 2u - 1$ and $\lambda = u - 1$.

The above characterization of constant-distance code closed under the cyclic shift was conjectured by the writer and proved independently at the same time by the writer [1] and R.C. Titsworth [5]. Titsworth's proof uses arguments on polynominals dividing $x^n - 1$.

3. Hadamard matrices and codes with two orbits. In this section we give a class of Hadamard matrices with doubly transitive collineation groups, and use these matrices to obtain a class of constant-distance codes with a transitive group on the columns for which the conclusion of Corollary 2 does not hold.

Let A be the Hadamard matrix of order 4 with 1 on the diagonal, -1 elsewhere, and let $B = B(s)$ be the tensor product of s copies of A.

THEOREM 4. *For any s, the group G of collineations of $B(s)$ is doubly transitive on the columns (and also on the rows).*

Proof. Denote the rows and columns of B by s-tuples, so that

$$b_{i_1}\cdots, i_s; j_1, \cdots, j_s = a_{i_1, j_1}a_{i_2, j_2}\cdots a_{i_s, j_s}\,.$$

The result is obvious when $s = 1$. Suppose $s = 2$. We shall show that the subgroup H of G fixing the column (1, 1) is transitive on the remaining columns. If τ_1 and τ_2 are any permutations on four letters then the permutation of columns sending (i_1, i_2) to $(\tau_1(i_1), \tau_2(i_2))$ is a collineation of B, sending row (i_1, i_2) to row $(\tau_1(i_1), \tau_2(i_2))$; denote this collineation by (τ_1, τ_2). It can be verified that the product of four transpositions of columns $\sigma = ((1, 4)\,(2, 3))((4, 1)\,(3, 2))((1, 3)\,(2, 4))((3, 1)\,(4, 2))$ is a collineation of B; also, $\sigma \in H$. Taking σ and its products with various (τ_1, τ_2), we see that all columns other than (1, 1) form a single orbit of H. Moreover some (τ_1, τ_2) moves column (1, 1), so that G is transitive, and hence doubly transitive. Now suppose that $s > 2$. If τ is a collineation of $B(2)$ and if a set of two column coordinates of $B(s)$ is given, then a collineation of $B(s)$ is obtained by applying τ to the given column coordinates while keeping the remaining ones fixed. Using this type of collineation, we see that the subgroup of G fixing column $(1, \cdots, 1)$ is transitive on the remaining columns. Hence G is always doubly transitive on the columns, and, by symmetry, also on the rows. This completes the proof.

COROLLARY 3. *For every power* 4^s *of* $4 (s > 1)$, *there is a constant-distance code with* 4^s *words of length* $4^s - 1$, *such that the group of collineations is transitive on the columns but has two nontrivial orbits on the rows.*

Proof. The matrix $B(s)$ is Hadamard, and hence its rows form a constant-distance code. Complement the rows with $a+1$ in column $(1, \cdots, 1)$ and then delete this column. What remains is still a constant-distance code; call it C. The subgroup of G fixing $(1, \cdots, 1)$ clearly gives a group of collineations of C which is transitive on the columns. Moreover the set of uncomplemented rows is closed under the group, so the group has two nontrivial orbits. This completes the proof.

Let G and H continue to have the same meanings as in Theorem 4. It follows from Corollary 2 and the proof of Corollary 3 that H is not nilpotent. However it can actually be shown that the subgroup K of H fixing column $(1, 2)$ is isomorphic to S_6, being generated by σ and certain (τ_1, τ_2)'s. Hence when $s = 2$, G has order $16 \cdot 15 \cdot 720$. Also it follows that if $s > 1$ then G contains a subgroup isomorphic to S_6 which fixes $2 \cdot 4^{s-2}$ columns.

REFERENCES

1. R.E. Block, *Difference sets, block designs, and constant distance codes,* Space Programs Summary No. 37-22, Vol. IV, Jet Propulsion Laboratory, California Institute of Technology, August 31, (1963), 137-138.
2. P. Dembowski, *Verallgemeinerungen von Transitivitätsklassen endlicher projektiver Ebenen,* Math. Zeit. **69** (1958), 59-89.
3. D.R. Hughes, *Collineations and generalized incidence matrixes,* Trans. Amer. Math. Soc. **86** (1957), 284-286.
4. E.T. Parker, *On collineations of symmetric designs,* Proc. Amer. Math. Soc. **8** (1957), 350-351.
5. R.C. Titsworth, *Binary cyclic constant-distance codes,* Space Programs Summary No. 37-22, Vol. IV, Jet Propulsion Laboratory, California Institute of Technology, August 31, (1963), 147, 152-153.

CALIFORNIA INSTITUTE OF TECHNOLOGY

PACIFIC JOURNAL OF MATHEMATICS
Vol. 15, No. 1, 1965

EXISTENCE OF BEST RATIONAL TCHEBYCHEFF APPROXIMATIONS

BARRY BOEHM

Some conditions are given which guarantee the existence of best Tchebycheff approximations to a given function f by generalized rational functions of the form

$$r(x) = \frac{a_1 g_1(x) + \cdots + a_n g_n(x)}{b_1 h_1(x) + \cdots + b_m h_m(x)}$$

The principal theorem states that such a best Tchebycheff approximation exists whenever $f, g_1, \cdots, g_n, h_1, \cdots, h_m$ are bounded continuous functions, defined on an arbitrary topological space X, and the set $\{h_1, \cdots, h_m\}$ has the dense nonzezo property on X: if $b_1, \cdots, b_n$ are real numbers not all zero, then the function $b_1 h_1 + \cdots + b_m h_m$ is different from zero on a set dense in X. An equivalent statement is that the set $\{h_1, \cdots, h_m\}$ is linearly independent on every open subset of X.

Further theorems assure the existence of best weighted Tchebycheff approximations and best constrained Tchebycheff approximations by generalized rational functions and by approximating functions of other similar forms.

Terminology. Let X be an arbitrary topological space, and let $C[X]$ be the linear space of functions f continuous on the space X, normed with the *Tchebycheff norm*

$$||f||_T = \sup_{x \in X} |f(x)| .$$

In this paper, we investigate the conditions necessary to guarantee the existence of a best approximation to functions $f \in C[X]$ by rational combinations of functions $g_1, \cdots, g_n, h_1, \cdots, h_m \in C[X]$. Such functions have the form

$$r_\gamma = \frac{a_1 g_1 + \cdots + a_n g_n}{b_1 h_1 + \cdots + b_m h_m} ,$$

where $\gamma = (a_1, \cdots, a_n, b_1, \cdots, b_m)$ is a vector in the closed set Γ_{n+m} of all real $(n + m)$-tuples satisfying

Received March 12, 1964. The material in this paper was included in the author's Doctoral Dissertation submitted to the University of California, Los Angeles under the guidance of Professor E.W. Cheney. The research was supported in part by The RAND Corporation under U.S. Air Force Project RAND (Contract No. AF 49 (638)-700) and in part by the U.S. Air Force Office of Scientific Research (Contract No. AF-AFOSR-77-63).

$$|b_1| + \cdots + |b_m| = 1 .$$

One such condition is that r_γ be well defined at points x_0 such that

$$b_1 h_1(x_0) + \cdots + b_m h_m(x_0) = 0 ;$$

thus, we shall restrict our attention to sets of functions $\{h_1, \cdots, h_m\}$ for which we can guarantee a unique definition of $r_\gamma(x_0)$.

A set of functions $\{h_1, \cdots, h_m\}$ is said to have the *dense nonzero property* on X if, for any $\gamma \in \Gamma_{n+m}$, the function

$$b_1 h_1 + \cdots + b_m h_m$$

is different from zero on a set Y_γ dense in X. (An equivalent statement is that the set $\{h_1, \cdots, h_m\}$ is linearly independent on all open subsets of X.) If this is the case, the function r_γ is well defined on the set Y_γ; to define r_γ uniquely at points $x_0 \in X - Y_\gamma$, we set

$$r_\gamma(x_0) = \limsup_{x \in Y_\gamma, x \to x_0} r_\gamma(x) .$$

We could define $r_\gamma(x_0)$ by a liminf operation just as well; all that is necessary is to define the function r_γ uniquely, and in such a way that if the limit

$$\lim_{x \in Y_\gamma,\ x \to x_0} r_\gamma(x)$$

exists, it is equal to $r_\gamma(x_0)$. Thus, if $\{h_1, \cdots, h_m\}$ has the dense nonzero property on X, the *generalized rational function* r_γ is uniquely defined on X for all $\gamma \in \Gamma_{n+m}$.

For each set $\{g_1, \cdots, g_n, h_1, \cdots, h_m\}$ such that $\{h_j\}$ has the dense nonzero property on X, let R denote the set of generalized rational functions

$$R = \{r_\gamma : \gamma \in \Gamma_{n+m}\} .$$

Then for each $f \in C[X]$ there exists a real number dist (R, f) representing the distance from f to the set R:

$$\text{dist}\,(R, f) = \inf_{r_\gamma \in R} \| f - r_\gamma \| .$$

If there exists a function $r_{\gamma*} \in R$ such that

$$\| f - r_{\gamma*} \| = \text{dist}\,(R, f) ,$$

then $r_{\gamma*}$ is called a *best rational approximation* to f, and dist (R, f), is the *error of the best rational approximation*.

After a brief survey in 2 and 3 of previous existence results and nonexistence phenomena, we demonstrate in § 4 that under the conditions prescribed above, there exists for every $f \in C[X]$ a best rational approximation $r_{\gamma *}$. Some extensions and specializations of this existence theorem, including its relation to the nonexistence phenomena of § 3, will be given in § 5. In § 6, we present some existence theorems for two other approximating families similar in nature to the family of rational approximations.

2. Previous results. The special case $m=1$, $h_1(x)=1$ corresponds to approximation by generalized polynomials $a_1g_1 + \cdots + a_ng_n$; it has been the subject of much fruitful study due to the feature of linearity in the coefficients a_i. An existence theorem was obtained in this case for Tchebycheff approximation of continuous functions f by algebraic polynomials

$$g_i(x) = x^{i-1}$$

by Borel in 1905 [2]; his proof was extended by Achieser [1] to arbitrary elements g_i in a normed linear space S.

Results are more sparse for the general rational problem $(m > 1)$ in which the coefficients do not enter linearly. Walsh obtained in 1931 [6] an existence theorem for ratios of polynomials of the same degree defined on a perfect set X in the complex plane.

THEOREM (Walsh). *For any $f \in C[X]$, X a perfect set in the complex plane, there exists a best Tchebycheff approximation $r_{\gamma *}$ to f among all rational functions of the form*

$$r_\gamma(x) = \frac{a_0 + a_1x + \cdots + a_nx^n}{b_0 + b_1x + \cdots + b_nx^n}$$

for $\gamma \in \Gamma_{2n+2}$.

Walsh also proved in [6] a similar existence theorem for L^p norms. Achieser gives in [1] an incomplete proof of theorem above for ratios of polynomials of arbitrary degrees on an interval $[a, b]$ of the real line. Cheney and Loeb [3] have recently obtained a similar theorem for rational trigonometric approximation.

Furthermore, the Achieser and Cheney-Loeb theorems show that with no loss of generality the denominator of the best approximation may be assumed to be strictly positive on the interval of definition.

3. Nonexistence phenomena. Some of the possible pitfalls in the existence problem are illustrated by the following two examples of nonexistence phenomena. In the first example, we consider the problem of approximating $f(x)=x$ in the Tchebycheff sense by a rational function of the form

$$r_\gamma(x) = \frac{a_1 x^2}{b_1 + b_2 x}$$

on the interval [0, 1], with the additional condition that the denominator be strictly positive on [0, 1]. Here, however, by setting $a_1 = b_2 = 1$ and letting $b_1 \downarrow 0$, we see that $\operatorname{dist}(R, f) = 0$, although no allowable $r_\gamma \in R$ achieves this minimum distance.

The second example shows that difficulties may arise when the dense nonzero property is violated. Consider the problem of approximating $f(x) = (x-1)(x-2)/2$ in the Tchebycheff sense by a rational function of the form

$$r_\gamma(x) = \frac{a_1}{b_1 + b_2 x},$$

with the three points 0, 1, 2 comprising X. Since $f(0) = 1$, $f(1) = f(2) = 0$, we see that the deviation of the approximation $\varepsilon/(x + \varepsilon)$ from f on X is no greater than $\varepsilon/(1 + \varepsilon)$, which can be made arbitrarily small by making ε small. Thus $\operatorname{dist}(R, f) = 0$, although again no choice of $r_\gamma \in R$ achieves this minimum.

4. An existence theorem.

We shall find it convenient to state part of the theorem as a separate lemma.

LEMMA 1. *If $f, h_1, \cdots, h_m$ are bounded functions on X, an arbitrary topological space, such that the set $\{h_j\}$ has the dense nonzero property on X, and if the set of functions $\{g_1, \cdots, g_m\}$ is linearly independent on X, then any sequence $\{\gamma_k\}$ of vectors in Γ_{n+m} such that*

$$\lim_{k\to\infty} \| r_{\gamma_k} - f \| = \inf_{\gamma \in \Gamma_{n+m}} \| r_\gamma - f \| = \operatorname{dist}(R, f),$$

has a cluster point $\gamma_0 \in \Gamma_{n+m}$.

Proof. (i). Define the functions $A = \sum a_i g_i$, $B = \sum b_j h_j$, with $\sum |b_j| = 1$; define A_k and B_k similarly. The boundedness of the h_j implies for any B that

$$\| B \| \leqq N = \max \| h_j \| ;$$

the linear independence of the set $\{g_i\}$ implies the existence of a positive number δ such that

$$\sum |a_i| = 1 \text{ implies } \| A \| \geqq \delta.$$

It is clear that for sufficiently large K, $k \geqq K$ implies

$$\operatorname{dist}(R, f) + 1 \geqq \| r_{\gamma_k} \| \geqq \frac{\| A_k \|}{M}$$

Hence, for $k \geqq K$

$$\| A_k \| \leqq N [\operatorname{dist}(R, f) + 1] ,$$

and by the definition of the number δ, for $k \geqq K$,

$$\sum_{i=1}^{n} | a_{j_k} | \leqq M = \frac{N}{\delta} [\operatorname{dist}(R, f) + 1] .$$

Thus, for $k \geqq K$, $\{\gamma_k\}$ is restricted to the compact set

$$\{\gamma \colon \sum | a_i | \leqq M, \ \sum | b_j | = 1\} .$$

By the Bolzano-Weierstrass theorem, then, the sequence $\{\gamma_k\}$ has a cluster point $\gamma_0 \in \Gamma_{n+m}$.

THEOREM 1. *If $f, g_1 \cdots, g_n, h_1, \cdots, h_m$ are bounded functions in $C[X]$, X an arbitrary topological space, and if the set $\{h_j\}$ has the dense nonzero property on X, then there exists a best rational Tchebycheff approximation $r_{\gamma*}$ to f on X.*

Proof. (i) Select a maximal linearily independent subset $\{g_1, \cdots, g_p\}$ among the functions g_i, and let $d = \operatorname{dist}(R, f)$. Then, any sequence $\{\gamma_k\}$ of vectors $\gamma_k \in \Gamma_{p+m}$ such that

$$\| r_{\gamma_k} - f \| \leqq d + 1/k$$

has by Lemma 1 a cluster point $\gamma_0 = (a_{10}, \cdots, a_{p0}, b_{10}, \cdots, b_{m0}) \in \Gamma_{p+m}$. We shall show that

$$\| r_{\gamma_0} - f \|_T = d .$$

Clearly, since $\gamma_0 \in \Gamma_{p+m}$, we need only show

$$\| r_{\gamma_0} - f \|_T \leqq d .$$

Since the set of functions $\{h_j\}$ has the dense nonzero property on X, the set Y_{γ_0} of points x at which the denominator $B_0(x)$ is different from zero, is dense in X. At points $x \in Y_{\gamma_0}$, we have for each k

$$\begin{aligned} | r_{\gamma_0}(x) - f(x) | &\leqq | r_{\gamma_0}(x) - r_{\gamma_k}(x) | + | r_{\gamma_k}(x) - f(x) | \\ &\leqq | r_{\gamma_0}(x) - r_{\gamma_k}(x) | + d + 1/k . \end{aligned}$$

As the functions h_j are bounded on X,

$$B_k \xrightarrow[k\to\infty]{} B_0$$

uniformly on X. Since $B_0(x) \neq 0$ for $x \in Y_{\gamma_0}$, this implies

$$\frac{A_k(x)}{B_k(x)} \xrightarrow[k\to\infty]{} \frac{A_0(x)}{B_0(x)}$$

for $x \in Y_{\gamma_0}$. Hence, for $x \in Y_{\gamma_0}$,

$$\lim_{k\to\infty} | r_{\gamma_0}(x) - r_{\gamma_k}(x) | = 0 ,$$

and thus

$$| r_{\gamma_0}(x) - f(x) | \leqq d .$$

It remains only to obtain this inequality for points $x_0 \in X - Y_{\gamma_0}$.

(ii). By the definition of the rational functions r_γ, we have for $x_0 \in X - Y_{\gamma_0}$ that

$$r_{\gamma_0}(x_0) = \lim_{x\in Y_{\gamma_0}} \sup_{x\to x_0} r_\gamma(x) .$$

Thus, there exists a sequence $\{x_\nu\}$ of points in Y_{γ_0} such that

$$| r_{\gamma_0}(x_0) - r_{\gamma_0}(x_\nu) | \leqq 1/\nu$$
$$| f(x_0) - f(x_\nu) | \quad \leqq 1/\nu$$

(since also $f \in C[X]$). Hence,

$$| r_{\gamma_0}(x_0) - f(x_0) | \leqq | r_{\gamma_0}(x_0) - r_{\gamma_0}(x_\nu) | + | r_{\gamma_0}(x_\nu) - f(x_\nu) |$$
$$+ | f(x_\nu) - f(x_0) | \leqq 1/\nu + d + 1/\nu .$$

Since the left hand side of this inequality is independent of ν, it follows for $x_0 \in X - Y_{\gamma_0}$ that

$$| r_{\gamma_0}(x_0) - f(x_0) | \leqq d .$$

Therefore $|| r_{\gamma_0} - f ||_T \leqq d$, implying, since $\gamma_0 \in \Gamma_{p+m}$, that $|| r_{\gamma_0} - f ||_T = d$, showing that indeed there exists a best approximation $r_{\gamma^*} = r_{\gamma_0}$ to f.

5. Extensions and specializations. Theorem 1 can be extended to the problem of weighted Tchebycheff approximation, in which the distance between f and r_γ is measured by the functional

$$|| s(r_\gamma = f) ||_T$$

for some prescribed weighting function $s \in C[X]$. This problem is equivalent to that of approximating the function sf by rational combinations of the functions sg_1 and h_j; existence of a best approximation is thus guaranteed whenever the products sf and sg_i are bounded functions and the functions h_j satisfy the hypotheses of Theorem 1.

Also, the proof of Theorem 1 is valid if the coefficients γ are restricted to a closed set $C_{n+m} \subset \Gamma_{n+m}$ containing at least one *feasible vector* γ^0 such that

$$\| s(r_{\gamma^0} - f) \|_T < \infty .$$

A slight but straightforward modification of step (ii) of Lemma 1 is needed if no vectors of the form $(0, \cdots, 0, b_1, \cdots, b_m)$ are in C_{n+m}.

Thus, the following theorem holds.

THEOREM 2. *If $f, s, g_1, \cdots, g_n, h_1, \cdots, h_m \in C[X]$ are such that the functions $sf, sg_1, \cdots, sg_n$ are bounded on X, an arbitrary topological space, and the set $\{h_j\}$ has the dense nonzero property on X, then for any closed set $C_{n+m} \subset \Gamma_{n+m}$ of coefficient vectors including a feasible vector γ^0, there exists a best weighted rational Tchebycheff approximation $r_{\gamma*}$ to f, such that*

$$\| s(r_{\gamma*} - f) \|_T = \inf_{\gamma \in C_{n+m}} \| s(r_\gamma - f) \|_T .$$

If the closed set of coefficients C_{n+m} of form

$$C_{n+m}(\varepsilon) = \{\gamma \in \Gamma_{n+m} : | \sum b_j h_j(x) | \geqq \varepsilon > 0, \ x \in X\}$$

is nonempty, we can obtain existence theorems with much weaker hypotheses on the functions involved, since in this case the set Y_{γ_0} comprises all of X, and step (ii) of Theorem 1, the only step requiring the continuity of f, s, g_1, and h_j, is not required in the proof. Hence, the following theorem holds in an arbitrary normed linear space.

THEOREM 3. *If the functions $f, s, g_1, \cdots, g_n, h_1, \cdots, h_m$ are such that $sf, sg_1, \cdots, sg_n, h_1, \cdots, h_m$ are bounded on X, an arbitrary set of points x, and if the set $C_{n+m}(\varepsilon) \subset \Gamma_{n+m}$ is nonempty, then there exists a best weighted rational approximation $r_{\gamma*}$ to f such that*

$$\| s(r_{\gamma*} - f) \| = \inf_{\gamma \in C_{n+m}(\varepsilon)} \| s(r_\gamma - f) \| .$$

Let us now consider the nonexistence examples of § 3 in the light of the above existence theorems. The first example can be handled by Theorem 1 by allowing the denominator $b_1 + b_2 x$ to have its zero at a point $x_0 \in [0, 1]$, and defining $a_1 x_0^2/(b_1 + b_2 x_0)$ by a limsup operation, which reduces in this case to a limit opreration. Thus, the function x^2/x is an acceptable rational function in Theorem 1, and is indeed the best approximation $r_{\gamma*}$.

The second example cannot be handled by Theorem 1 since the dense nonzero property is violated. A weaker result can be given for both examples by Theorem 3, however, by considering only those rational functions such that $b_1 + b_2x \geqq \varepsilon$; i.e., $\gamma \in C_3(\varepsilon)$. With this modification, a best approximation $r_{\gamma*}$ exists in the first example and is at least as good as $x^2/(\varepsilon + x)$; hence the error

$$\text{dist}(R, f) \leqq \varepsilon/(\varepsilon + 1)$$

can be made as small as desired by taking ε small enough. In the second example, $r_{\gamma*}$ again exists and is at least as good as $\varepsilon/(\varepsilon + x)$; thus again

$$\text{dist}(R, f) \leqq \varepsilon/(\varepsilon + 1) .$$

In practical problems, placing such a "floor" under the denominator function and slightly above zero is often a reasonable thing to do, as the inequality constraint $B(x) \geqq \varepsilon$ is no harder to deal with than $B(x) > 0$.

In most continuous rational Tchebycheff approximation problems, the existence of a best approximation is guaranteed by Theorems 1 and 2, as sets of functions with the dense nonzero property are fairly common. They include all linearly independent sets of functions analytic on a perfect set X, and all sets of piecewise analytic functions on X which are linearly independent on each component of analyticity.

An independent result similar to Theorem 1 has been obtained recently by Newman and Shapiro [4]. Their existence theorem is stated for functions defined on a compact Hausdorff space X, and thus does not cover such problems as the approximation of functions continuous and bounded on the positive real axis by functions of the form

$$r_\gamma(x) = \frac{\sum a_i e^{-\lambda_i x}}{\sum b_i e^{-\mu_i x}}$$

for λ_i, $\mu_j \geqq 0$, a problem handled by Theorem 1. Rice in [5] has also obtained independently a somewhat similar existence theorem for the interval $[0, 1]$, under the assumption that the denominator possess only a finite set of zeros.

6. Existence theorems for other approximating families. The fact that best approximations exist among rational functions with coefficients in a closed set allows us, with the aid of the following lemma, to state some theorems assuring the existence of best approximations in other approximating families.

LEMMA 2. *The set of all vectors* $(c_{11}, \cdots, c_{1m}, c_{21}, \cdots, c_{nm})$ *such that* $c_{ij} = a_i b_j$ *for real numbers* a_i, b_j, *is closed.*

The proof of this lemma is straightforward, and is omitted here.

The following theorem follows directly from Lemma 2 and Theorem 3, with $m = 1$, $n = pq$, and $g_\nu = u_i v_j$, since the set of numerator coefficients $c_\nu = a_i b_j$ is closed.

THEOREM 4. *If the functions* $f, s, u_1, \cdots, u_p, v_1, \cdots, v_q$ *are such that the products* $sf, su_1v_1, \cdots, su_pv_q$ *are bounded on* X, *an arbitrary set of points* x, *then there exists a best approximation*

$$P^* = (a_1^* u_1 + \cdots + a_p^* u_p)(b_1^* v_1 + \cdots + b_q^* v_q)$$

to the function f, *such that*

$$\| s(P^* - f) \| = \inf_{a_i, b_j} \| s[(\sum a_i u_i)(\sum b_j v_j) - f] \| .$$

In a similar fashion, a theorem can be established on the existence of best approximations by finite products of generalized polynomials of the form

$$P = (\sum a_{i1} g_{i1})(\sum a_{i2} g_{i2}) \cdots (\sum a_{in} g_{in}) .$$

In particular, if the component polynomials are of the form $ax + b$, we have the following corollary.

COROLLARY 4a. *Any function* f *bounded on a compact domain* X *on the real line has, among all polynomials* P_n *of degree* n *having only real roots, a best approximation* P_n^*.

The next theorem follows from Lemma 2 and Theorem 2; a similar theorem can be based on Lemma 2 and Theorem 3.

THEOREM 5. *If the functions*

$$f, s, u_1, \cdots, u_p, v_1, \cdots, v_q, h_1, \cdots, h_m \in C[X]$$

are such that the products of $sf, su_1h_1, \cdots, su_ph_m, sv_1, \cdots, sv_q$ *are bounded on* X, *an arbitrary topological space, and the set* $\{h_j\}$ *has the dense nonzero property on* X, *then there exists a best weighted Tchebycheff approximation*

$$P^* = a_1^* u_1 + \cdots + a_p^* u_p + \frac{d_1^* v_1 + \cdots + d_q^* v_q}{b_1^* h_1 + \cdots + b_m^* h_m}$$

to the function f, *such that*

$$\| s(P^* - f) \|_T = \inf_{a_i\ b_j\ d_k} \| s\Big(\sum a_i u_i + \frac{\sum d_k v_k}{\sum b_j h_j} - f\Big) \|_T .$$

References

1. N. I. Achieser, *Theory of approximation,* Frederick Ungar Publishing Company, New York, 1956.
2. E. Borel, *Lecons sur les Fonctions des Variables Reelles,* Paris, 1905.
3. E. W. Cheney, and H.L. Loeb, *Generalized rational approximation,* J. Soc. Indust. Appl. Math., Series B (Numerical Analysis) **1** (1964), 11-25.
4. D. J. Newman, and H. S. Shapiro, *Approximation by generalized rational functions,* Proceedings of Conference on Approximation, Birkhäuser Verlag, 1964 (to appear).
5. J. R. Rice, *On the existence of best Tchebycheff approximations by general rational functions,* Abstract 63T-331, Notices Amer. Math. Soc. **10** (1963), 576.
6. J. L. Walsh, *The existence of rational functions of best approximation,* Trans. Amer. Math. Soc. **33** (1931), 668-689.

THE RAND CORPORATION

PACIFIC JOURNAL OF MATHEMATICS
Vol. 15, No. 1, 1965

A NOTE ON HAUSDORFF'S SUMMATION METHODS

J. P. BRANNEN

If $\{a_n\}$ is a moment sequence and (Δa) is the difference matrix having base sequence $\{a_n\}$, then (Δa) is symmetric about the main diagonal if and only if the function $\alpha(x)$ such that $a_n = \int_0^1 x^n d\alpha(x)$, $n = 0, 1, 2, \cdots$, is symmetric in the sense that $\alpha(x) + \alpha(1 + x) = \alpha(1) + \alpha(0)$ except for at most countably many x in $[0, 1]$. This property is related to the "fixed points" of the matrix H, where HaH is the Hausdorff matrix determined by the moment sequence $\{a_n\}$.

In each of the papers [2], [3] and [5], there is reference to difference matrices of the form

$$(\Delta d) = \begin{bmatrix} \Delta^0 d_0 & \Delta^0 d_1 & \Delta^0 d_2 & \vdots \\ \Delta^1 d_0 & \Delta^1 d_1 & \Delta^1 d_2 & \\ \Delta^2 d_0 & \Delta^2 d_1 & \Delta^2 d_2 & \\ & \cdots & & \end{bmatrix}$$

where $\{d_n\}$ is a moment sequence, $\Delta^0 d_n = d_n$, $n = 0, 1, 2, \cdots$ and $\Delta^m d_n = \Delta^{m-1} d_n - \Delta^{m-1} d_{n+1}$, for $n = 0, 1, 2, \cdots$ and $m = 1, 2, 3, \cdots$. In [2], Garabedian and Wall discussed the importance of (Δd) having the property of being symmetric about the main diagonal, i.e. $\Delta^m d_n = \Delta^n d_m$. They also showed that if $\{d_n\}$ is a totally monotone sequence, then (Δd) is symmetric about the main diagonal if and only if the function $f(x)$ which generates $\{d_n\}$ has a certain type continued fraction expansion.

In this paper, the symmetry of (Δd) is investigated with the restriction of total monotonicity removed and a collection of necessary and sufficient conditions are given, Theorem 3, for moment sequences in general. A relation is established between the symmetry of (Δd) and the "fixed points" of the difference matrix

$$(1) \qquad H = \begin{bmatrix} \binom{0}{0} & & \\ \binom{1}{0} & -\binom{1}{1} & \\ \binom{2}{0} & -\binom{2}{1} & \binom{2}{2} \\ & \cdots & \end{bmatrix}.$$

Received March 15, 1964. This work was performed under the auspices of the United States Atomic Energy Commission.

2. Notation, definitions, and examples. Except for some notation and definitions introduced for convenience, the notation and definitions of this paper will follow [6].

NOTATION. If $\{d_n\}$ is an infinite sequence, d^* and d' denote respectively the diagonal and column matrices determined by $\{d_n\}$.

DEFINITION 1. If $\{d_n\}$ is a number sequence such that for some function $f(x)$ on $[0, 1]$,

$$d_p = \int_0^1 x^p df(x) = \int_0^1 (1-x)^p df(x) \; ; \qquad p = 0, 1, 2, \cdots ,$$

then $\{d_n\}$ is called a symmetric moment sequence.

The Cesàro moment sequence $1, \frac{1}{2}, \frac{1}{3}, \cdots$ provides an example of a moment sequence satisfying Definition 1 since for $p = 0, 1, 2, \cdots$

$$(2) \qquad c_p = \int_0^1 x^p dx = x^{p+1}/p + 1\Big]_0^1$$
$$= \int_0^1 (1-x)^p dx = -(1-p)^{p+1}/p + 1\Big]_0^1 = \frac{1}{p+1} .$$

DEFINITION 2. If A is a semi-infinite, lower triangular, matrix having inverse and $\{a_n\}$ and $\{d_n\}$ are sequences such that $A^{-1}d^*Aa' = A^{-1}a^*Ad'$, then $\{a_n\}$ and $\{d_n\}$ are symmetric relative to A.

The Cesàro moment sequence $1, \frac{1}{2}, \frac{1}{3}, \cdots, c_p$ of (2), and the sequence $1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \cdots$ are symmetric relative to the matrix H of (1).

3. THEOREMS. LEMMA. *Suppose $\{s_n\}$ is a sequence such that $s_p \neq 0$ for $p = 0, 1, 2, \cdots$ and suppose that A is a semi-infinite matrix having inverse such that $As' = s'$; then,*

(i) $A^{-1}s' = s'$,

(ii) *$\{x_n\}$ and $\{s_n\}$ are symmetric with respect to A if and only if $Ax' = x'$, and*

(iii) *if $A^{-1}a^*As' = A^{-1}s^*Aa'$ and $A^{-1}b^*As' = A^{-1}s^*Ab'$, then $A^{-1}b^*Aa' = A^{-1}a^*Ab'$.*

Proof. (i) is obvious. For the proof of (ii), we first suppose $\{x_n\}$ is symmetric with $\{s_n\}$ relative to A so that $A^{-1}x^*As' = A^{-1}s^*Ax'$. Multiplying both sides on the left by A and using $As' = s'$ gives $x^*s' = s^*Ax'$. Under the hypothesis, s^* has inverse s^{*-1} so that

$$(3) \qquad s^{*-1}x^*s' = s^{*-1}s^*Ax' = Ax' .$$

Since $x^*s' = s^*x'$, it follows from (3) that $x' = Ax'$.

On the other hand, if $Ax' = x'$,

$$A^{-1}x^*As' = A^{-1}x^*s' \tag{4}$$

and

$$A^{-1}s^*Ax' = A^{-1}s^*x' \; .$$

Since $s^*x' = x^*s'$, it follows from (4) that x and s are symmetric relative to A.

For the proof of (iii), we suppose that $a' = s^{*-1}a^*s'$ and $b' = s^{*-1}b^*s'$, from which it follows that

$$A^{-1}a^*Ab' = A^{-1}a^*s^{*-1}b^*s' \tag{5}$$

and

$$A^{-1}b^*Aa' = A^{-1}b^*s^{*-1}a^*s' \; . \tag{6}$$

Since diagonal matrices permute, it follows that (5) and (6) are equal establishing (iii).

THEOREM 1. *If $\{b_n\}$ is a moment sequence, i.e.,*

$$b_p = \int_0^1 x^p dg(x) \; , \tag{7}$$

$\{b_n\}$ and the Cesàro sequence (2) *are symmetric relative to H if and only if $\{b_n\}$ is a symmetric moment sequence.*

Proof. Let

$$f_n(x) = \begin{cases} \sum_{p=0}^{n-1} \binom{n}{p}(-1)^p x^p & \text{for } n = 2, 4, 6, \cdots \\ \sum_{p=0}^{n-1} \binom{n}{p}(-1)^p x^p - 2x^n & \text{for } n = 1, 3, 5, \cdots \; . \end{cases}$$

Clearly, if $\{t_n\}$ is any number sequence, $Ht' = t'$ if and only if

$$\sum_{p=0}^{n-1} \binom{n}{p}(-1)^p t_p = 0 \qquad \text{for } n = 2, 4, 6, \cdots$$

and

$$\sum_{p=0}^{n-1} \binom{n}{p}(-1)^p t_p - 2t_n = 0 \qquad \text{for } n = 1, 3, 5, \cdots \; .$$

Thus if $\{b_n\}$ is defined as in (7), $Hb' = b'$ if and only if

$$\int_0^1 f_n(x) dg(x) = 0 \qquad \text{for } n = 1, 2, 3, \cdots \; . \tag{8}$$

But, $f_n(x) = (1-x)^n - x^n$ for $n = 1, 2, 3, \cdots$ so that

$$(9) \qquad \int_0^1 f_n(x)dg(x) = \int_0^1 (1-x)^n dg(x) - \int_0^1 x^n dg(x)\ ,$$

and consequently (8) holds if and only if $\{b_n\}$ is a symmetric moment sequence. It follows from (9) and (2) that $Hc' = c'$ and from the preceding Lemma that $\{b_n\}$ and $\{c_n\}$ are symmetric relative to H.

Conversely, if $\{b_n\}$ and $\{c_n\}$ are symmetric relative to H, it follows that $Hb' = b'$, and if $\{b_n\}$ is defined as in (7), then $\{b_n\}$ is a symmetric moment sequence.

THEOREM 2. *If $g(x)$ is of bounded variation on* $[0, 1]$ *and* $\{z_n\}$ *is the moment sequence determined by $g(x)$, the following two statements are equivalent*:

(i) $\{z_n\}$ *is a symmetric moment sequence, and*

(ii) *there do not exist uncountably many x in* $[0, 1]$ *for which* $g(x) + g(1-x) \neq g(1) + g(0)$.

Proof. Suppose (i). Then let $u = 1 - x$ so that,

$$z_p = \int_0^1 (1-x)^p dg(x) = \int_0^1 u^p dg(1-x) = -\int_0^1 u^p dg(1-u)\ .$$

Thus, $\int_0^1 (1-x)^p dg(x) = -\int_0^1 x^p dg(1-x)$ so that for $p = 0, 1, 2, \cdots$,

$$(10) \qquad \int_0^1 x^p d[g(x) + g(1-x)] = 0\ .$$

Since $g(x) - g(1-x)$ is of bounded variation on $[0, 1]$, (10) implies that for every $k(x)$ continuous on $[0, 1]$, $\int_0^1 k(x)d[g(x) + g(1-x)] = 0$. This, [4, p. 69], implies (ii). Reversing the steps leading to (10) shows that (ii) implies (i).

An interesting example of a function satisfying (ii) is provided by Evans in [1].

THEOREM 3. *Suppose $g(x)$ is of bounded variation on* $[0, 1]$ *and suppose* $\{a_n\}$ *is the moment sequence generated by $g(x)$. The following statements are equivalent*:

(i) $\{a_n\}$ *is a symmetric moment sequence,*

(ii) $Ha' = a'$,

(iii) $\{a_n\}$ *and the Cesàro moment sequence* $\{c_n\}$ *are symmetric relative to H, and*

(iv) *the difference matrix* (Δa) *having base sequence* $\{a_n\}$ *is symmetric about the main diagonal.*

Proof. Theorem 1 implies the equivalence of (i), (ii), and (iii).

(i) implies (iv) provided

$$(11)\quad \int_0^1 x^m(1-x)^n dg(x) = \int_0^1 x^n(1-x)^m dg(x) \qquad \text{for } m, n = 0, 1, 2, \cdots .$$

Let $u = 1 - x$ so that $\int_0^1 x^m(1-x)^n dg(x) = \int_1^0 (1-u)^m u^n dg(1-u)$. Thus (11) may be rewritten as

$$(12)\qquad -\int_0^1 (1-x)^m x^n dg(1-x) = \int_0^1 x^n(1-x)^m dg(x)$$
$$= \int_0^1 x^n(1-x)^m d[g(x) + g(1-x)] = 0 .$$

That (12) is the case for $\{a_n\}$ a symmetric moment sequence follows from (ii) of Theorem 2. (iv) implies (ii) since (iv) implies that $a_n = \Delta^n A_0$, which is the same as saying that $Ha' = a'$. Thus the equivalence of the four statements is established.

I am grateful to Professor H. S. Wall for some comments which have been of considerable value in the preparation of this paper.

References

1. G. C. Evans, *Calculation of moments for a Cantor-Vitali function*, Sixth Herbert Ellsworth Slaught Memorial Paper, Mathematical Association of America, (1957), 22-27.

2. H. L. Garabedian and H. S. Wall, *Hausdorff methods of summation and continued fractions*, Trans. Amer. Math. Soc. **48** (1940), 185-207.

3. H. L. Garabedian, Einar Hille, and H. S. Wall, *Formulations of the Hausdorff inclusion problem*, Duke Math. J. **8** (1941), 193-213.

4. T. H. Hildebrandt, *Introduction to the theory of integration*, Academic-Press, New York, 1963.

5. W. T. Scott and H. S. Wall, *The transformation of series and sequences*, Trans. Amer. Math. Soc. **51** (1942), 255-279.

6. H. S. Wall, *Analytic theory of continued fractions*, van Nostrand, New York, 1948.

SANDIIA LABORATORY

PACIFIC JOURNAL OF MATHEMATICS
Vo. 15, No. 1, 1965

TOPOLOGICAL SEMILATTICES ON THE TWO-CELL

To Professor A. D. Wallace on his 60th birthday

DENNISON R. BROWN

Topological lattices on the n-cell have been studied by L.W. Anderson, A.D. Wallace, A.L. Shields, and L.E. Ward, Jr. In particular, these authors have papers setting forth conditions under which a topological lattice on the two-cell is topologically isomorphic to the product lattice $I \times I$.

The primary purpose of this paper is the investigation of topological semilattices (commutative, idempotent topological semigroups) on the two-cell which retain the lattice like property that for each element x, $\{y: x \leqq y\}$ is a connected set. Specifically, it is shown that any such entity is the continuous homomorphic image of one of a fixed pair of semilattices on the two-cell, where the choice of domain depends on the location of the zero element.

It is also proved that a TSL on the two-cell has an identity (a unique maximal element) and $\{y: x \leqq y\}$ connected for each element x if and only if it is the continuous homomorphic image of $I \times I$. Also, if $\{y: x \leqq y\}$ is connected for each element x, then S, a TSL on the two-cell, is generated by its boundary B in the sense that $B^2 = S$.

Semilattices on the n-cell are also discussed. Let S be such an object with boundary B. It is proved that if x is a maximal element of S, then $x \in B$. If S has an identity, 1, and T is a continuum chain from 1 to 0, then $S = BT$.

Finally, let S be a continuum TSL with 1 and let A be the subset defined by $x \in A$ if and only if $\{y: x \leqq y\}$ is connected. Then (1) $x \in A$ if and only if there is a continuum chain from 1 to x; and(2) A is a nondegenerate continuum sub-TSL of S.

Topological lattices on the n-cell have been studied in [1], [6], and in [8]. In particular, these papers set forth conditions under which a topological lattice on the two-cell is iseomorphic (topologically isomorphic) to the product lattice $I \times I$.

The primary purpose of this paper is the investigation of topological semilattices (commutative, idempotent topological semigroups) on the two-cell which retain the lattice-like property that for each element x, $M(x)$ is a connected set (see below). Specifically, we show that any such entity is the continuous homomorphic image of one of a fixed pair of semilattices, where the choice of domain depends upon the location of the zero.

Received March 12, 1964. This work was supported in part by NSF contract GP-1637.

Section 3 discusses semilattices on the n-cell. The role of the boundary sphere in determining the multiplication is seen to be quite important.

The next section is the main body of the paper. In addition to the theorems indicated above, we prove that any topological semilattice on the two-cell which has a unique maximal element and all $M(x)$ connected must be the continuous homomorphic image of $I \times I$. In particular, any topological lattice has these properties. We also show that, if each $M(x)$ is connected, then a topological semilattice S on the two-cell is generated by its boundary B in the sense that $B^2 = S$.

In § 5 we prove that if S is a compact, connected topological semilattice with identity, then the subset of elements x such that $M(x)$ is connected is a compact connected subsemilattice of S.

We are indebted to Professors H. Cohen and R.J. Koch for their helpful comments and support.

2. Preliminaries. A *topological semilattice* (hereafter TSL) is a pair $(S, \leqq)$ such that S is a Hausdorff topological space, $\leqq$ is a continuous semilattice ordering on S. Equivalently, S is a commutative, idempotent topological semigroup with $x \leqq y$ if and only if $xy = x$.

An element x of S is *maximal* if it is dominated by no other element of S; that is, $xy = x$ implies $y = x$. A *minimal* element is defined dually. It is well known that a compact TSL has maximal elements and a unique minimal element. For $x \in S$, let $M(x) = \{y: x \leqq y\}$, $L(x) = \{y: y \leqq x\}$. It is easy to verify that $L(x) = Sx$. These are closed subsemilattices of S [10].

A *chain* is a totally ordered subset of S. Of primary interest here are compact, connected chains; in case S is metric these are known to be arcs [13] and will be referred to henceforth as *arc chains*.

The following theorem, due to Koch [4], is stated without proof.

THEOREM A. *If S is a compact, connected, metric TSL with zero* (0), *then every $x \in S$ is connected to* 0 *by an arc chain.*

If S is as stated in Theorem A and has also $M(x)$ connected for each $x \in S$, then, by replacing S by $M(x)$ in the theorem, is may be seen that any pair of comparable elements in S is connected by an arc chain.

A space S is homotopically trivial if $\pi_i(S) = 0$, $i > 0$ where $\pi_i(S)$ is the ith homotopy group of S. The following result extends slightly a theorem of Anderson and Ward [2].

THEOREM B. *If S is an arcwise connected idempotent semigroup* *with* (0), *then* S *is homotopically trivial.*

Proof. Let $f: I_1 \times \cdots \times I_n \to S$ with $f(\text{Bndry}\,(I \times \cdots \times I) = 0$. Define $g: (I_1 \times \cdots \times I_n) \times I \to S$ by

$$g(x_1, \cdots, x_n, r) = f(x_1, \cdots, x_n)f(x_1 - x_1 r, \cdots, x_n - x_n r) .$$

Henceforth, the letter I will be reserved to represent the TSL on the arc [0, 1] defined by $xy = \min(x, y)$, where the ordering is that inherited from the real numbers. Any arc chain is iseomorphic to I [5].

3. **Semilattices on the n-cell.** Throughout this section, S represents a TSL whose underlying space is an n-cell, and B the boundary $n - 1$ sphere of S. If S has an identity, 1, then 1 is clearly the unique maximal element of S. It is well-known [7] that $1 \in B$. The following order—theoretic version of the maximum modulus theorem generalizes this statement.

THEOREM 1. *Let x be a maximal element of S. Then $x \in B$.*

Proof. By the maximality of x, and Theorem B, $S \backslash \{x\}$ is a subsemilattice, homotopically trivial. Hence $x \notin S \backslash B$.

In [3] and in [5] it was shown that, under certain conditions, the multiplication in S is determined by that in B together with that in a certain arc subsemigroup. The next theorem is of a similar nature.

THEOREM 2. *Let S have a* 1, *and let T be any arc chain from* 1 *to* 0. *Then $S = BT$.*

Proof. Since $B \subset BT$, it suffices to show BT is contractible. Since T is an interval, define $g: (BT) \times T \to BT$ by $g(bt, r) = btr$. Then $g(bt, 1) = bt$, and $g(bt, 0) = 0$. The function g is clearly continuous, hence the proof is complete.

4. **Semilattices on the 2-cell.** The following lemmas will be useful in the sequel.

LEMMA 1. *Let S be a topological semilattice in which, for each x, $M(x)$ is a connected set. Let $f: S \to T$ be a continuous homomorphism of S onto T. Then, for each $y \in T$, $M(y)$ is connected; furthermore f is a monotone map.*

Proof. The continuous homomorphic image of an arc chain is clearly an arc chain, although possibly degenerate. Let $y \in T$, $z \in M(y)$, $y = f(a)$, $z = f(b)$. Then $y = yz = f(a)f(b) = f(ab)$. Let C be an arc chain in S from b to ab. Then $f(C)$ is again such from z to y.

Now let $X = f^{-1}(y)$, $a, b \in X$. Let C, D be arc chains from a to ab, b to ab respectively. Then $C \cup D \subseteqq X$, hence X is connected.

LEMMA 2. *Let S be a two cell, $S = C \cup D$, $C \cap D = \square$, where C and D are arc wise connected sets. Let B be the boundary of S and suppose $B \cap C \neq \square \neq B \cap D$. Then $C \cap B$ and $D \cap B$ are each connected sets.*

Proof. If $C \cap B$ is not connected, then clearly neither is $D \cap B$. Choose a, b from different components of $C \cap B$, and let T be an arc in C connecting a, b. Choose d, e from different components of $D \cap B$, so that $\{a, b\}$ separates d and e in B, and let J be an arc in D connecting d, e. Then J and T cannot be disjoint [13], which contradicts $C \cap D = \square$.

In the remainder of this section, S will represent a TSL whose underlying space is a two-cell, and B the boundary circle of S.

COROLLARY. *Let $a, b \in B$, with $M(ab)$ a connected set. Decompose B into arcs P, Q with $P \cap Q = \{a, b\}$. Then either $P \subseteqq M(ab)$ or $Q \subseteqq M(ab)$.*

Proof. $M(ab)$ is arcwise connected by Theorem A. On the other hand, any $x \in S \backslash M(ab)$ can be connected to 0 by an arc chain T in S. Clearly $T \cap M(ab) = \square$. Hence $S \backslash M(ab)$ is arcwise connected. By the lemma above, $M(ab) \cap B$ is therefore connected and the result follows.

Methods used in portions of the proof of the following theorem are similar to those used in [8].

THEOREM 3. *Suppose S has a 1. These are equivalent:*
 (i) *for each x, $M(x)$ is a connected set;*
 (ii) *B is the union of two maximal arc chains of S;*
 (iii) *S is the continuous homomorphic image of $I \times I$.*

Proof. (i) implies (ii). Fix $a \in B$, $a \neq 1$. By the above corollary, $M(a)$ must contain one of the boundary arcs between a and 1. Designate this arc by Q and let $p \in Q$, $p \neq a$. Let P be the boundary arc between p and 1 which is contained in Q. Then $P \subseteqq M(p)$, for if not then $a \in M(p)$, which is false. It follows that any element of Q

compares with p, and hence Q is totally ordered. Let J be a maximal chain in B, with a, $1 \in J$. By continuity of multiplication, J is closed and therefore proper in B. Let $t \in B\backslash J$, and let K be a maximal chain in B, with t, $1 \in K$. By the maximality of J and K, $B = J \cup K$. From the anti-symmetry of the relation ($\leqq$), J and K have the same minimal element, z. Now $M(z)$ is a compact, connected semilattice, homotopically trivial by Theorems A and B. Since $B \subseteqq M(z)$, $M(z) = S$, and hence $z = 0$. The arc chains J, K are thus maximal in B. If $c \in S\backslash B$, then $M(c) \cap J$ and $L(c) \cap J$ are closed, disjoint subsets of J, and hence fail to exhaust J. Then maximality of J, K in S is now immediate.

(ii) implies (iii). Let R be the closed ideal $(I \times \{0\}) \cup (\{0\} \times I)$ of $I \times I$; and let M be the Rees quotient $(I \times I)/R$. This TSL on the 2-cell has the properties that every nonzero element of M is represented uniquely as a product of two boundary elements, one from each of the maximal chains composing the boundary of M (if $a \in C$, then $a = a \cdot 1$), and $ab = 0$ implies either $a = 0$ or $b = 0$. Denote the boundary of M by $C = V \cup W$, and the boundary of S by $B = J \cup K$, with V, W, J, K maximal arc chains.

Let $f_1\colon V \to J$ and $f_2\colon W \to K$ be iseomorphisms. Define $f^*\colon M \to S$ by $f^*(x) = f_1(a)f_2(b)$, where $ab = x$, $a \in V$, $b \in W$. The only element of M which has a nonunique representation in this manner is 0; but $ab = 0$ requires that one of a, $b = 0$. Hence f^* is well defined, and the following diagram is commutative:

$$\begin{array}{ccc} M & \xrightarrow{\quad f^* \quad} & S \\ \uparrow & & \uparrow \\ V \times W & \xrightarrow{\; f_1 \times f_2 \;} & J \times K \end{array}$$

Here, the vertical arrows represent the respective multiplication functions. Since these functions, together with f_1 and f_2, are continuous, and V and W are compact, it follows that f^* is continuous.

Next, let $a = v_1w_1$, $b = v_2w_2$ be elements of M. Then

$$\begin{aligned} f^*(ab) &= f^*(v_1w_1v_2w_2) = f^*((v_1v_2)(w_1w_2)) = f_1(v_1v_2)f_2(w_1w_2) \\ &= f_1(v_1)f_1(v_2)f_2(w_1)f_2(w_2) = f^*(v_1w_1)f^*(v_2w_2) = f^*(a)f^*(b) . \end{aligned}$$

Hence f^* is a homomorphism.

Finally, $f^*(M)$ is a compact connected TSL containing B; by Theorems A and B it follows that $f^*(M) = S$. The natural map of $I \times I$ onto $(I \times I)/R$ is now composed with f^* to obtain the desired result.

(iii) implies (i). Clearly $I \times I$ has $M(x)$ connected for each x. By Lemma 1, S has this property also.

COROLLARY 3.1. *If S has a 1 and $M(x)$ connected for each $x \in S$, then $0 \in B$.*

COROLLARY 3.2. *If S is a topological lattice on the 2-cell, then S is the continuous (semilattice) homomorphic image of $I \times I$.*

Proof. As a topological lattice, S has a 1 and has $M(x)$ connected for each $x \in S$.

THEOREM 4. *If $M(x)$ is connected for each $x \in S$, then $B^2 = S$.*

Proof. Suppose $0 \in B$. It will first be shown that $B^3 \subseteqq B^2$. Let $a, b, c \in B$. In order to prove that $abc \in B^2$, it suffices to assume that a, b, c are distinct and nonzero. Assume also that these points are named so that $0, b$ are in different components of $B \backslash \{a, c\}$. By the corollary to Lemma 2, one component of $B \backslash \{a, c\}$ lies in $M(ac)$. Since $0 \notin M(ac)$, it follows that $b \in M(ac)$, hence $abc = ac$ and $B^3 \subseteqq B^2$. Hence $B^4 = B^3B \subseteqq B^2B \subseteqq B^2$. Since B^2 is a compact connected TSL, and $B \subseteqq B^2$, it follows from Theorems A and B that $B^2 = S$.

Now suppose $0 \in S \backslash B$, and again select a, b, c distinct elements of B. If any of ab, ac, bc, $abc \in B$, then immediately $abc \in B^2$. Similarly, it may be assumed that $a \notin M(bc)$, $b \notin M(ac)$, $c \notin M(ab)$. By the corollary to Lemma 2, it follows that $B = M(ab) \cup M(ac) \cup M(bc)$. But the latter subset is included in $M(abc)$. Since $M(abc)$ is a compact, connected TSL and $B \subseteqq M(abc)$, it again follows that $M(abc) = S$, and therefore $abc = 0$. It has now been shown that $B^3 = B^2 \cup \{0\}$; thus $B^2 \cup \{0\}$ is compact and connected. Furthermore, $(B^2 \cup \{0\})^2 \subseteqq B^4 \cup \{0\} \subseteqq B^3 \cup \{0\}$; hence $B^2 \cup \{0\}$ is a subsemilattice containing B. This yields $B^2 \cup \{0\} = S$. But B^2 is compact, hence $0 \in B^2$. Consequently $B^3 \subseteqq B^2$, and as before, $B^2 = S$.

The next pair of theorems shows that the structure of S when $0 \in B$ is essentially different from that occurring when $0 \in S \backslash B$. Let $T = \{(x, y) \in I \times I : x + y \leqq 1\}$. Note T is a subsemilattice of $I \times I$.

THEOREM 5. *The semilattice S has $0 \in B$ and $M(x)$ connected for each x if and only if S is a continuous homomorphic image of T.*

Proof. As in Theorem 3, let $R = (I \times \{0\}) \cup (\{0\} \times I)$. Let $N = T/R$, the Rees quotient of T. Let D be the boundary circle of N. Note that every nonzero element of N has a unique representation as a product of two not necessarily distinct elements of D, and that $ab = 0$ implies $a = 0$ or $b = 0$. Now let $f: D \to B$ homeomorphically, with $f(0) = 0$, and extend f to $f^*: N \to S$ by $f^*(c) = f(a)f(b)$, where $a, b \in D$. As in Theorem 3, f^* is well defined, and the following diagram is commutative:

$$\begin{array}{ccc} N & \xrightarrow{\ f^*\ } & S \\ \uparrow & & \uparrow \\ D \times D & \xrightarrow{\ f \times f\ } & B \times B \end{array}$$

Vertical arrows represent the respective multiplication functions of N and S. Since these functions and f are continuous and D is compact, f^* is continuous. By Theorem 4, f^* maps N onto S.

It remains to show that f^* is a homomorphism. To this end, let $c = ab$, $a = pq$, $b = rs$, with p, q, r, s distinct nonzero elements of D. Some unique pair of p, q, r, s separates the remaining pair from 0 on D; suppose p, r and 0 lie in different components of $D\backslash\{q, s\}$. By the corollary to Lemma 2, $\{p, r\} \subseteqq M(qs)$. Hence $ab = pqrs = qs$, and $f^*(c) = f^*(ab) = f^*(qs) = f(q)f(s)$. On the other hand, since f is a homeomorphism on D, it follows that $\{f(p), f(r)\}$ and $\{f(0) = 0\}$ lie in different components of $B\backslash\{f(q), f(s)\}$. Hence $\{f(p), f(r)\} \subseteqq M(f(q)f(s))$ and therefore $f(q)f(s) = f(p)f(q)f(r)f(s) = f^*(pq)f^*(rs) = f^*(a)f^*(b)$. The argument is similar if p, q, r, s occur in a different order in D. This portion of the proof is now complete.

The converse follows from the fact that T has $M(x)$ connected for each x and Lemma 1, together with the fact that a monotone map of a a two-cell onto a two-cell must take boundary onto boundary [12].

Now, let W be the disk of radius one, centered at the origin of a plane, and let F be the boundary circle of W. If x, $y \in F$, let xy be the midpoint of the chord joining x and y. This is transparently continuous, and note that a nonzero point of W is uniquely represented as the product of two boundary points. To extend the multiplication to all of W let $a = wx$, $b = yz$ where a, $b \in W$ and $w, x, y, z \in F$; set $ab = 0$ if each boundary arc containing w, x, y and z has length $\geqq \pi$, and otherwise let ab be the product of those two of the four elements w, x, y and z whose distance apart is a maximum. Again continuity is obvious, as is the fact that multiplication is commutative and idempotent.

It is certainly desirable to give an alternative, order-theoretic description of W. For each $x \in W$ let $L(x)$ be the intersection of all circular disks which are tangent to F, which contain x, and whose boundaries contain 0. This is a semilattice partial order and $L(x) \cap L(y)$ is precisely $L(z)$ for that $z(= xy) \in L(x) \cap L(y)$ which is at maximum distance from 0. In this manner, it if easily seen that $M(x)$ is connected for each $x \in W$. Note that, if a, b, c, $d \in F$, then $abcd \neq 0$ if and only if one of these elements, say a, has the property that b, c, d lie in the same component of $F\backslash\{a, -a\}$, where $(-a)$ represents the

[1] The author is indebted to the referee for improving the description of the semigroup W.

point of F antipodal to a.[1]

THEOREM 6. *Let* $0 \in S \backslash B$. *Then* $M(x)$ *is a connected set for each* $x \in S$ *if and only if* S *is the continuous homomorphic image of* W.

Proof. One implication is an immediate consequence of Lemma 1, since W has $M(x)$ connected for all x. Therefore, assume that S has $M(x)$ connected for all x, $0 \notin B$. It will be shown that S is the continuous homomorphic image of W.

First, suppose that f is a continuous map of F onto B. Define $f^*: W \rightarrow S$ by $f^*(x) = f(a)f(b)$, where $x = ab$, a, $b \in F$. Recall that, if $x \neq 0$, then this representation of x is unique, and $f^*(x)$ is therefore well defined. On the other hand 0 may be expressed only as the product of any pair of antipodal elements of F. Hence, in order that f^* be well defined, $ab = 0$, a, $b \in F$ must imply $f(a)f(b) = 0$ in S. The construction of a continuous map f with this property is the major portion the proof. For $x \in B$, define $A(x) = \{y \in B : xy = 0\}$. In the sequel, the expression $[a, b]$, a, $b \in B$, will represent the counterclockwise arc of B from a to b.

(A) *For each* $x \in B$, $A(x)$ *is a continuum; further, there exist* y, $z \in B$ *such that* $yz = 0$, $x \in [y, z]$ *and* $A(x) \cap [z, y]$ *is nonempty.* By Theorem 4, $B^2 = S$; hence there exist y, $z \in B$ such that $yz = 0$. Let $x \in [y, z]$, $t \in [z, y]$. By the corollary to Lemma 2, then either $y \in M(xt)$ or $z \in M(xt)$. Hence $x[z, y] \subseteqq L(y) \cup L(z)$. Since $x[z, y]$ is connected and $L(y) \cap L(z) = \{0\}$, it follows that for some $t \in [z, y]$, $xt = 0$; hence $A(x) \neq \square$. Next, let $a, b \in A(x)$, $x \in [a, b]$, $t \in [b, a]$. Again by the corollary to Lemma 2, either $a \in M(xt)$ or $b \in M(xt)$. Say $a \in M(xt)$; then $xt = a(xt) = (ax)t = 0$. Therefore $[b, a] \subseteqq A(x)$; by using the compactness of B to obtain a maximal interval, it may be seen that $A(x)$ is an (possibly degenerate) arc in B.

(B) *There exist* $a_0, a_1 \in B$ *such that* $a_0a_1 = 0$ *and for every* $x \in (a_0, a_1)$, $A(x) \subseteqq (a_1, a_0)$. Let $a \in B$; there exists $a_1 \in A(a)$ such that $[a, a_1] \cap A(a) = \{a_1\}$ and there exists $a_0 \in A(a_1)$ such that $[a_0, a_1] \cap A(a_1) = \{a_0\}$. Then $a_0a_1 = 0$ and we observe that $a \in [a_1, a_0]$. It $x \in (a_0, a_1)$ then it is obvious that $a_1x \neq 0 \neq ax$. Now by (A), $A(x)$ meets $[a_1, a]$ and, since it is connected, $A(x) \subset (a_1, a) \subset (a_1, a_0)$.

(C) *Let* a, b, c, d, e *be five elements of* B *occurring in counter clockwise order as listed; suppose also that* $ac = 0 = bd$ *and that* $be \neq 0 \neq ce$. *Then* $A(e) \subseteqq (b, c)$. For, from $ac = 0$, $ec \neq 0$, $e \in (c, a)$, we have $A(e) \subset (e, c)$ by (A). Similarly, from $bd = 0$, $eb \neq 0$, $e \in (d, b)$, we have $A(e) \subset (b, e)$. Therefore $A(e) \subset (e, c) \cap (d, b) = (b, c)$.

The function $f: F \rightarrow B$ will now be defined.

(D) Choose elements a_0, a_1 of B as in part B above. For convenience, the antipodal point of $x \in F$ is denoted by $-x$. Fix any $x \in F$ and define $f(x) = a_0$. $f(-x) = a_1$. F is now decomposed into the two closed intervals $[x, -x]$ and $[-x, x]$, while $B = [a_0, a_1] \cup [a_1, a_0]$. The scheme is now as follows: f will map a dense subset of $[x, -x]$ onto a dense subset of $[a_0, a_1]$, and a dense subset of $[-x, x]$ onto a dense subset of $[a_1, a_0]$ in an order preserving manner; furthermore $f(y)f(-y)$ will be 0 for every y in either the dense subset of $[x, -x]$ or the dense subset of $[-x, x]$. The function f will then be extended through standard methods into a continuous map of F onto B; that $f(y)f(-y) = 0$ for all $y \in F$ will be a consequence of this method of construction.

For ease of notation, set $-x = x_1 = y_0$, $x_0 = x = y_1$ in F and $a_1 = b_0$, $b_1 = a_0$ in B. Now $F = [x_0, x_1] \cup [y_0, y_1]$, $B = [a_0, a_1] \cup [b_0, b_1]$. Let a_{01} be the mid-point of $[a_0, a_1]$, x_{01} the mid-point of $[x_0, x_1]$, $y_{01} (= -x_{01})$ that of $[y_0, y_1]$. Define $f(x_{01}) = a_{01}$. By part B), $A(a_{01}) \subseteq (b_0, b_1)$; let $b_{01} \in A(a_{01})$ and define $f(y_{01}) = b_{01}$.

(E) Next, let b_{001} be the mid-point of $[b_0, b_{01}]$, y_{001} that of $[y_0, y_{01}]$, and define $f(y_{001}) = b_{001}$. Let $x_{001} = -y_{001}$. It is necessary to map x_{001} into some point a_{001} of the interval $[a_0, a_{01}]$. To this end, suppose that $a_0 b_{001} = 0 = a_{01} b_{001}$. Then by (A) above $[a_0, a_{01}] \subseteq A(b_{001})$. In this case choose a_{001} to be the mid-point of $[a_0, a_{01}]$. If, on the other hand, $a_0 b_{001} = 0 \neq a_{01} b_{001}$, then let $a_{001} = a_0$; if $a_0 b_{001} \neq 0 = a_{01} b_{001}$, let $a_{001} = a_{01}$. Finally, if $a_0 b_{001} \neq 0 \neq a_{01} b_{001}$, then, on applying (C) with $a = b_{01}$, $b = a_0$, $c = a_{01}$, $d = b_0$, $e = b_{001}$, it follows that $A(b_{001}) \subseteq (a_0, a_{01})$. In this case, choose a_{001} arbitrarily in $A(b_{001})$. Similarly, let b_{011} be the mid-point of $[b_{01}, b_1]$; by an argument similar to the one above, there exists $a_{011} \in [a_{01}, a_1]$ such that $a_{011} b_{011} = 0$. Choose the appropriate y_{011}, x_{011} in F and define $f(x_{011}) = a_{011}$, $f(y_{011}) = b_{011}$.

(F) In the next stage, mid-points a_{0001} of $[a_0, a_{001}]$, a_{0011} of $[a_{001}, a_{01}]$, a_{0101} of $[a_{01}, a_{011}]$ and a_{0111} of $[a_{011}, a_1]$ are chosen as images of the appropriate x_i. As many as two of the four intervals listed may be degenerate; it is still possible to choose a "mid-point". Suppose, for example, that $a_{001} = a_0 = a_{0001}$. Then $a_{0001} \cdot b_{001} = 0 = a_{0001} \cdot b_0$, hence by (E) above, b_{0001} may be chosen as the mid-point of $[b_0, b_{001}]$.

(G) At any stage, suppose a_i is the mid-point of $[a_j, a_k]$. It is then necessary that $b_i \in [b_j, b_k]$. By examining the products $a_i b_j$ and $a_i b_k$, b_i may be chosen precisely by means of the argument used in part (E). A dual argument is obvious in the event that the original choice of mid-point is from a subinterval of $[b_0, b_1]$, rather than $[a_0, a_1]$.

(H) By (G) it may be assumed that f has been defined on a dense subset D of F into B. The image subset $f(D)$ is dense in B, since lengths of complementary intervals clearly approach zero. Also, f is monotonic within $[x_0, x_1]$ and $[y_0, y_1]$. It is therefore possible to extend f to a continuous map of F onto W. Furthermore, choose $x \in F$. It must be shown that $f(x)f(-x) = 0$. Let $x \in [x_0, x_1]$. It may be assumed that $x \notin D$; let $\{x_i\} \to x$, $\{x_i\} \subseteqq D$. Then $\{-x_i\} \to -x$ and $\{-x_i\} \subseteqq D$. By the continuity of f, $\{f(x_i)\} \to f(x)$, $\{f(-x_i)\} \to f(-x)$. Finally, $\{0 = f(x_i)f(-x_i)\} \to f(x)f(-x)$, by the continuity of multiplication in S. Hence $f(x)f(-x) = 0$.

(I) From the discussion prior to (A), the function f^* is now well defined from W into S, and the following diagram is commutative:

$$\begin{array}{ccc} W & \xrightarrow{\ f^*\ } & S \\ \uparrow & & \uparrow \\ F \times F & \xrightarrow{\ f \times f\ } & B \times B \end{array}$$

Since F is compact, f^* is continuous; by Theorem 4, f^* maps W onto S. It remains to show that f^* is a homomorphism. Let $ab = c \in W$, and suppose $a = wx$, $b = yz$, with $w, x, y, z \in F$. Then $c = wxyz$. If $c \neq 0$, then recall that one of these factors of c, say w, must have the property that x, y, z are all in the same component of $F \backslash \{w, -w\}$. Suppose further that $\{x, z\} \subseteqq [w, y] \subseteqq [w, -w]$, where all intervals represented are counter clockwise. Then $c = wy$, hence $f^*(c) = f(w)f(y)$. On the other hand, since f is monotone on $[w, y]$, $\{f(x), f(z)\} \subseteqq [f(w), f(y)]$ in B, and $f(-w) \in [f(y), f(w)]$. If $f(y) = f(-w)$, then $f(w)f(y) = 0 = f^*(c) = f^*(a)f^*(b)$. If $f(y) \neq f(-w)$, then by the corollary of Lemma 2, $[f(w), f(y)] \subseteqq M(f(w)f(y))$, hence $f^*(a)f^*(b) = f(w)f(x)f(y)f(z) = f(w)f(y) = f^*(c)$. The other cases are handled similarly.

If $ab = c = 0$ in W, again with $a = wx$, $b = yz$, then it must be shown that $f(w)f(x)f(y)f(z) = 0$. Since $c = 0$, x, y, z cannot all be in the same component of $F \backslash \{w, -w\}$. Suppose y is in one component of $F \backslash \{w, -w\}$, and $\{x, z\}$ in the other. Then, within the component containing $\{x, z\}$, $-y$ must be separated from $-w$ by one of x, z; otherwise w, x, z, are in the same component of $F \backslash \{y, -y\}$. Suppose x separates $-y$ from $-w$. Then, applying the corollary to Lemma 2, $-y \in M(wx)$. Hence $f(-y) \in M(f(w)f(x))$, and therefore

$$\begin{aligned} f^*(a)f^*(b) &= f(w)f(x)f(y)f(z) = [f(-y)f(w)f(x)]f(y)f(z) \\ &= [f(-y)f(y)]f(w)f(x)f(z) = 0 = f^*(c) . \end{aligned}$$

The remaining cases are similar. This completes the proof.

5. **Remarks on the general case.** It is easy to construct a TSL in which $M(x)$ fails to be connected for some x. For example, let J be the arc subsemilattice of $I \times I$ consisting of

$$(\{0\} \times I) \cup (\{1\} \times I) \cup (I \times \{0\}) .$$

Then $M((0, y))$ is not connected for any $y > 0$. Similarly, the product TSL on the disk $J \times J$ contains points of this nature. For a more complicated example, let K be a subset of $I \times I$ defined as follows: Let a canonical Cantor set C be constructed on the arc I, and let D_i be the union of the open intervals deleted from C at the ith stage in its construction. Let $K_i = \{(x, y) : x \in D_i,\ (3^i - 1)/3^i \leqq y \leqq 1\}$. Let $K = (I \times I) \setminus \bigcup_i K_i$. Then K is a subsemilattice and is topologically a disk. Set $z = (0, 1)$. Then $M(z)$ is a Cantor set.

LEMMA 3. *Let S be any compact connected metric TSL with identity. Let $A = \{x : M(x)$ is a connected subset of $S\}$. Then $x \in A$ if and only if x lies on an arc chain containing* 1.

Proof. Suppose $x \in A$. Then $M(x)$ is a compact connected TSL, and by Theorem A, there exists an arc chain T from 1 to x. Conversely, let T be an arc chain from 1 to x. Clearly $T \subseteqq M(x)$. Let $y \in M(x)$. Then yT is connected, and contains x and y, and is a subset of $M(x)$. Hence every element of $M(x)$ is connected to x by a connected subset of $M(x)$, and therefore $M(x)$ is connected.

Recall that, if $\{A_n\}_{n \in \omega}$ is a collection of closed subsets of a space S, then $\lim\sup \{A_n\} = \{x \in S;$ if $x \in U$, U open in S, $n \in \omega$, then there exists $m > n$ such that $A_m \cap U \neq \Box\}$.

THEOREM 7. *Let S and A be as in Lemma* 3. *Then A is a compact connected subsemilattice of S containing* 0 *and* 1.

Proof. Clearly $0, 1 \in A$. Let $x, y \in A$, and let I, J be arc chains from 1 to x, y, respectively. Then $I \cup xJ$ is an arc chain from 1 to xy; by Lemma 3, $xy \in A$. Hence A is a subsemilattice. Furthermore, since $I \subseteqq A$, every element of A lies in a connected subset of A which also includes the element 1; hence A is connected. It remains to show that A is closed. Let $\{x_n\}$ be a sequence in A, and let $\{x_n\}$ converge to x. For each n, let T_n be an arc chain from 1 to x_n. Let $T = \lim\sup \{T_n\}$. The set T is known to be connected [12]. To see that $T \subseteqq M(x)$, choose $a \in T$, let $\{a_n\}$ cluster to a, $a_n \in T_n$. Then $\{x_n\} = \{a_n x_n\}$ clusters to ax, hence $ax = x$. Therefore $a \in M(x)$, and x is connected to 1 inside of $M(x)$. It now follows easily that every element of $M(x)$ lies in a connected set containing x within $M(x)$, hence $x \in A$ and the proof is complete.

In all examples on the 2-cell known to the author, A is also locally connected; it is conjectured that if S is locally connected, then A is also. Indeed, it may be that A is a homomorphic retract of S.

If S is not assumed to have a 1, none of the conclusions of the above theorem need hold. In particular, certain subsemilattices of the TSL $J \times J$ mentioned earlier in this section fail in these respects.

Bibliography

1. L.W. Anderson, *Topological lattices and* n-*cells,* Duke Math. J. **25** (1958), 205-208.
2. L.W. Anderson and L.E. Ward, Jr., *A structure theorem for topological lattices,* Proceeding of the Glasgow Mathematical Association, Vol. V (1961), 1-3.
3. A.L. Hudson, *Some semigroups on the two-cell,* Proc. Amer. Math. Soc. **10** (1959), 648-655.
4. R.J. Koch, *Arcs in partially ordered spaces,* Pacific J. Math. **9** (1959), 723-728.
5. P.S. Mostert and A.L. Shields, *On the structure of semigroups on a compact manifold with boundary,* Ann. Math. **65** (1957), 117-143.
6. A.L. Shields, *The n-cube as a product semigroup,* Michigan Math. J. **4** (1957), 165-166.
7. A.D. Wallace, *Cohomology, dimension, and mobs,* Summa Brasilia Math. **3** (1953), 43-54.
8. _______, *Factoring a lattice,* Proc. Amer. Math. Soc. **9** (1958), 250-252.
9. _______, A *theorem on acyclicity,* Bull. Amer. Math. Soc. **67** (1961), 123-124.
10. L. E. Ward, P*artially ordered topological spaces* Proc. Amer. Math. Soc. **5** (1954), 144-161.
11. L.E. Ward, Jr., *Concerning Koch's theorem on the existence of arcs,* to appear.
12. G.T. Whyburn, *Analytic topology,* New York, 1942.
13. R.L. Wilder, *Topology of manifolds,* New York, 1949.

The University of Tennessee

PACIFIC JOURNAL OF MATHEMATICS
Vol. 15, No. 1, 1965

SOME INEQUALITIES FOR SYMMETRIC MEANS

P. S. BULLEN

This paper was received before the synoptic introduction became a requirement.

1. In two recent papers, [3, 4], Everitt has generalised certain known inequalities, by replacing the known monotonicity of certain set (or sequence) functions by super-additivity; the sequence functions are zero if all the terms of the sequence are equal.

Included in the inequalities generalised is one due to Rado, [5, p. 61]. Bullen and Marcus, [1], recently proved a multiplicative analogue of this inequality and a generalisation to symmetric means. It is one of the intentions of this note to show that the corresponding sequence function, which is 1 when all the terms of the sequence are equal is logarithmically super-additive, (Corollary 5, below). Further properties of these sequence functions are then investigated.

2. $(a) = (a_1, \cdots, a_m)$ will denote an m-tple of positive numbers. $E_r(a)$, $1 \leqq r \leqq m$, is the rth elementary symmetric function of (a),

$$(1) \qquad E_r = E_r(a) = \sum \prod_{j=1}^{r} a_{i_j} , \qquad E_0 = 1 ,$$

the sum being over all r-tples, $i_1, \cdots, i_r$, such that $1 \leqq i_1 < \cdots < i_r \leqq m$. $P_r(a)$ is the mean of $E_r(a)$,

$$(2) \qquad P_r = P_r(a) = \binom{m}{r}^{-1} E_r .$$

If $m = n + q$, $(\bar{a}) = (a_1, \cdots, a_n)$, $(\tilde{a}) = (a_{n+1}, \cdots, a_{n+q})$ and correspondingly $\bar{E}_r = E_r(\bar{a})$, $\tilde{E}_r = E_r(\tilde{a})$, etc., if r has suitable values. When $r = 1$ the symmetric means are arithmetic means and will be written $P_1 = A_{n+q}$, $\bar{P}_1 = \bar{A}_n$, $\tilde{P}_1 = \tilde{A}_q$. Similarly, P_{n+q}, $\bar{P}_n$, $\tilde{P}_q$ are powers of geometric means and will be written G_{n+q}^{n+q}, $\bar{G}_n^n$ and $\tilde{G}_q^q$ respectively.

3. It is known, [5, p. 52] that

$$(3) \quad s < t \text{ implies } P_s^t \geqq P_t^s, \text{ with equality if and only if } a_1 = \cdots = a_m.$$

It is easily seen from (1) that

(i) if $s \leqq \min(n, q)$ then $E_s = \sum_{t=0}^{s} \bar{E}_{s-t}\tilde{E}_t$,

(ii) if $s > \max(n, q)$ then $E_s = \sum_{t=0}^{n+q-s} \bar{E}_{n-t}\tilde{E}_{s-n+t}$,

Received June 3, 1963. This research was supported by the U. S. Air Force Office of Scientific Research.

(iii) if $q < s \leqq n$ then $E_s = \sum_{t=0}^{q} \bar{E}_{s-t}\tilde{E}_t$.

Using these identities and (2) we have

LEMMA 1. (i) *If* $1 \leqq s \leqq n + q$

$$P_s = \sum_{t=u}^{v} \lambda_t^{(s)} \bar{P}_{s-t}\tilde{P}_t \,, \tag{4}$$

where $u = \max(s - n, 0)$, $v = \min(s, q)$, $\lambda_0^{(s)} = \lambda_s^{(s)} = 1$ *and* $t \neq 0, s$, $\lambda_t^{(s)} = \left[\binom{n}{s-u-t}\binom{q}{u+t}\right] \Big/ \binom{n+q}{s}$.

(ii) *In particular if* $a_{n+1} = \cdots = a_{n+q} = \beta$ *then*

$$P_s = \sum_{t=u}^{v} \lambda_t^{(s)} \bar{P}_{s-t}\beta^t \,, \tag{5}$$

and if in addition $a_1 = \cdots = a_n = \alpha$,

$$P_s = \sum_{t=u}^{v} \lambda_t^{(s)} \alpha^{s-t}\beta^t \,. \tag{6}$$

When $q = 1$ this reduces to formulae (2) and (4) of [1].

4. We are now in a position to state and prove

THEOREM 2. *Let* $1 \leqq r \leqq k \leqq n + q$ *and* $u = \max(r - n, 0)$, $v = \min(r, q)$, $w = \max(k - n, 0)$, $x = \min(k, q)$. *Then*

(i) *if* $v \leqq w$ *and* $r - u \leqq k - x$

$$\frac{P_r^{k/r}}{P_r} \geqq \frac{\bar{P}_{r-u}^{(k-x)/(r-u)}}{\bar{P}_{k-x}} \frac{\tilde{P}_v^{w/v}}{\tilde{P}_w} \,, \tag{7}$$

(ii) *if* $v \leqq w$

$$\frac{P_r^{k/r}}{P_k} \geqq \frac{\tilde{P}_v^{w/v}}{\tilde{P}_w} \,, \tag{8}$$

with equality in each case if and only if either $r = k$ *or* $a_1 = \cdots = a_{n+q}$.

Before proceeding with the proof it should be noted that the condition $v \leqq w$ becomes $r - u \leqq k - x$ if n and q are interchanged. So if $r - u \leqq k - x$ inequality (8) holds, with the role of n and q interchanged; or equivalently $P_r^{k/r}/P_k \geqq \bar{P}_{r-u}^{(k-x)/(r-u)}/\bar{P}_{k-x}$. If neither $v \leqq w$ nor $r - u \leqq k - x$ then, from (3), nothing is true. The condition $v \leqq w$ is equivalent to $\min(r, q) \leqq \max(k - r, 0)$ and for this either $r < q$ and $k \geqq n + r$ or $r \geqq q$ and $k = n + q$; that is $k \geqq n + v$. For both $v \leqq w$ and $r - n \leqq k - x$ either $r < \min(n, q)$ and $k \geqq r + \max(n, q)$ or $r \geqq \min(n, q)$ and $k = n + q$.

Proof of Theorem 2. If $r = k$ the results are trivial so assume $r < k$. Rewrite (7) as

$$L = \frac{P_k^r}{\bar{P}_{k-x}^r \tilde{P}_w^r} \leqq \frac{P_r^k}{\bar{P}_{r-u}^{[(k-x)/(r-u)]r} \tilde{P}_v^{wr/v}} = R \, .$$

By (4) with $s = r$

$$(9) \qquad P_r^k = \left(\sum_{t=u}^{v} \lambda_t^{(r)} \bar{P}_{r-t} \tilde{P}_t \right)^k .$$

Using (3) on each term of this sum

$$P_r^k \geqq \left(\sum_{t=u}^{v} \lambda_t^{(r)} \bar{P}_{r-u}^{(r-t)/(r-u)} \tilde{P}_v^{t/v} \right)^k .$$

By (6) the right hand side of this inequality is the kth power of the rth symmetric mean of $b_1, \cdots, b_{n+q}$ where $b_1 = \cdots = b_n = P_{r-u}^{1/(r-u)}$ and $b_{n+1} = \cdots = b_{n+q} = P_v^{1/v}$. Using (3), (6) and $r < k$ this gives

$$P_r^k \geqq \left(\sum_{t=w}^{x} \lambda_t^{(k)} \bar{P}_{r-u}^{(k-t)/(r-u)} \tilde{P}^{t/v} \right)^r$$
$$= \bar{P}_{r-u}^{[(k-x)/(r-u)]r} \tilde{P}^{wr/v} \left(\sum_{t=w}^{x} \lambda_t^{(k)} \bar{P}_{r-u}^{(x-t)/(r-u)} \tilde{P}_v^{(t-w)/v} \right)^r .$$

On rewriting we get,

$$R \geqq \left(\sum_{t=w}^{x} \lambda_t^{(k)} \bar{P}_{r-u}^{(x-t)/(r-u)} \tilde{P}_v^{(t-w)/v} \right)^r = S \, , \quad \text{say} .$$

Similarly by (4)

$$(10) \qquad P_k^r = \left(\sum_{t=w}^{x} \lambda_t^{(k)} \bar{P}_{k-t} \tilde{P}_t \right)^r .$$

Using (3) on each term of this sum gives

$$P_k^r \leqq \left(\sum_{t=w}^{x} \lambda_t^{(k)} \bar{P}_{k-x}^{(k-t)/(k-x)} \tilde{P}_w^{t/w} \right)^r$$
$$= \bar{P}_{k-x}^r P_w^r \left(\sum_{t=w}^{x} \lambda_t^{(k)} \bar{P}_{k-x}^{(x-t)/(r-x)} \tilde{P}_w^{(t-w)/w} \right)^r .$$

Rewriting we have that

$$L \leqq \left(\sum_{t=w}^{x} \lambda_t^{(k)} \bar{P}_{k-x}^{(x-t)/(k-x)} \tilde{P}_w^{(t-w)/w} \right)^r = T \, , \quad \text{say} .$$

By the condition in (i) and (3), $T \leqq S$, which proves (7). Some terms in the above proof become undefined in certain limiting cases. If they are defined to be 1 the proof is then correct. Finally, since $r < k$, the inequality is clearly strict when (3) is. This completes the proof of (i).

To prove (ii) the procedure is similar except that when (3) is applied to the right hand sides of (9) and (10) it is applied to the second part of each term only, that is to $\tilde{P}_t$. The analysis is then the same with (5) being used instead of (6).

COROLLARY 3.

$$\left(\frac{A_{n+q}}{G_{n+q}}\right)^{n+q} \geqq \left(\frac{\bar{A}_n}{\bar{G}_n}\right)^n \left(\frac{\tilde{A}_q}{\tilde{G}_q}\right)^q \tag{11}$$

with equality if and only if $a_1 = \cdots = a_{n+q}$.

Proof. From Theorem 2(i) with $r = 1$, $k = n + q$.

COROLLARY 4. *If* $1 \leqq r \leqq s \leqq n$ *then*

$$\frac{P_r^{s+1}}{P_{s+1}^r} \geqq \frac{\bar{P}_r^s}{\bar{P}_s^r}, \tag{12}$$

and in particular

$$\left(\frac{A_{n+1}}{G_{n+1}}\right)^{n+1} \geqq \left(\frac{\bar{A}_n}{\bar{G}_n}\right)^n, \tag{13}$$

with equality if and only if $a_1 = \cdots = a_{n+1}$.

Proof. From Theorem 2 (ii) with $k = s + 1$, $n = 1$. These results are those in [1].

Finally if $r\{(a)\} = r(a) = (A_m/G_m)^m$ then we have

COROLLARY 5. $\log r\{(\bar{a}) \cup (\tilde{a})\} \geqq \log r(\bar{a}) + \log r(\tilde{a})$.

5. The above inequalities (11), (13) and that due to Rado, [5, p. 61] can be further generalised by the use of weighted means. Let $(w) = (w_1, \cdots, w_n)$ be an m-tple of nonnegative numbers, not all zero. Define

$$W_r = \sum_{i=1}^{n} w_r, \qquad W_r > 0,$$

$$A_r = A_r^{(w)} = \frac{1}{W_r}\sum_{i=1}^{r} w_i a_i, \qquad G_r = G_r^{(w)} = \left(\prod_{i=1}^{r} a_i^{w_i}\right)^{1/W_r}.$$

It is known that

$$G_r \leqq A_r, \quad \text{with equality only when} \quad a_1 = \cdots = a_n. \tag{14}$$

A generalisation of Rado's inequality and (13) is given by

THEOREM 6.

$$W_n(A_n - G_r) \leqq W_{n+1}(A_{n+1} - G_{n+1}), \tag{15}$$

(16) $$\left(\frac{A_n}{G_n}\right)^{W_n} \leqq \left(\frac{A_{n+1}}{G_{n+1}}\right)^{W_{n+1}},$$

with equality if and only if $a_1 = \cdots = a_{n+1}$.

Proof. The proofs are exactly those of the special cases. As direct proofs were not given in [1] they will be given here. In particular the proof of (15) is simpler than that suggested in [5].

(15) is equivalent to $G_{n+1} \leqq (W_n/W_{n+1})G_n + (w_{n+1}/W_{n+1})a_{n+1} = U$, say.

(17) $$\begin{aligned} G_{n+1} &= G_n^{(W_n/W_{n+1})} a_{n+1}^{(w_{n+1})/(W_{n+1})} \\ &\leqq U \end{aligned}$$

by an application of (14).

Similarly (16) is equivalent to

$$A_{n+1} \geqq A_n^{(W_n/W_{n+1})} a_{n+1}^{(w_{n+1})/(W_{n+1})} = V, \quad \text{say}$$

but

(18) $$\begin{aligned} A_{n+1} &= \frac{W_n}{W_{n+1}} A_n + \frac{w_{n+1}}{W_{n+1}} a_{n+1} \\ &\geqq V, \end{aligned}$$

by an application of (14).

In a similar way we can prove

THEOREM 7.

$$W_{n+p}(A_{n+q} - G_{n+q}) \geqq \bar{W}_n(\bar{A}_n - \bar{G}_n) + \tilde{W}_q(\tilde{A}_q - \tilde{G}_q),$$

$$\left(\frac{A_{n+q}}{G_{n+q}}\right)^{W_{n+q}} \geqq \left(\frac{\bar{A}_n}{\bar{G}_n}\right)^{\bar{W}_n} \left(\frac{\tilde{A}_q}{\tilde{G}_q}\right)^{\tilde{W}_q},$$

with equality if and only if $a_1 = \cdots = a_{n+q}$.

Generalisations along the same lines are possible for the inequalities (7), (8) and (12). Suppose $(wa) = (w_1a_1, \cdots, w_ma_m)$; then define

$$F_r(a) = \frac{E_r(wa)}{E_r(w)},$$

a generalisation of $P_r(a)$, to which it reduces if $w_1 = \cdots = w_m \neq 0$. The two m-tples (a), (w) will be said to be similarly ordered if for all i, j, $a_i \leqq a_j$ $(a_i \geqq a_j)$ implies $w_i \leqq w_j$ $(w_i \geqq w_j)$.

THEOREM 8. *If* (a) *and* (w) *are similarly ordered then*

(i) $s<t$ *implies* $F_s^t > F_t^s$, *with equality if and only if* $a_1 = \cdots = a_m$.

(ii) *inequalities* (7), (8) *and* (12) *hold, subject to the relevant conditions, with* P *replaced by* F.

Proof. The proof of (i) is exactly that of (3), [5, p. 53]. Then the inequalities follow as before.

The requirement that (a) and (w) be similarly ordered is essential as the following example shows. If $(a) = (1, 1, 2)$ and $(w) = (2, 1, 1)$ then $F_1 < F_2^{1/2}$ but $F_2^{1/2} > F_3^{1/3}$. The extreme case $s = 1$, $t = m$ of (i) is a weaker form of (14) since $F_m^{1/m}$ is the unweighted geometric mean whereas F_1 is the weighted arithmetic mean with the larger numbers having the larger weights.

6. In recent papers Diananda, [2] and Kober [6], have investigated further properties of $A_n - G_n$. We will now prove multiplicative analogues of their results. Let $(w) = (w_1, \cdots, w_n)$, $w_i > 0$, $W_n = 1$ and define

$$R_n = R_n^{(w)} = \frac{A_n^{(w)}}{G_n^{(w)}} = \frac{A_n}{G_n}$$

$$L_n = \prod_{i,j=1}^{n} \left(\frac{a_i^{1/2} a_j^{-1/2} + a_i^{-1/2} a_j^{1/2}}{2} \right),$$

$$\Lambda_n = \prod_{i,j=1}^{n} \left(\frac{a_i^{1/2} a_j^{-1/2} + a_i^{-1/2} a_j^{1/2}}{2} \right)^{w_i w_j},$$

$w = \min(w_1, \cdots, w_n)$, $W = \max(w_1, \cdots, w_n)$.

THEOREM 9.

$$L_n^{w/(n-1)} \leqq R_n \leqq E_n^W, \tag{19}$$

$$\Lambda_n^{1/(1-w)} \leqq R_n \leqq \Lambda_n^{1/w}, \tag{20}$$

with equality if and only if $a_1 = \cdots = a_n$.

Proof. The proofs of (19) and (20) are similar so only that of (20) will be given. Writing $\alpha = 1/(1 - w)$ the left hand inequality in (20) can be rewritten as

$$G_n^{1-\alpha} \Pi_n^{\alpha} \leqq A_n, \tag{21}$$

where

$$\Pi_n = \prod_{i,j=1}^{n} \left(\frac{a_i + a_j}{2} \right)^{w_i w_j}.$$

The left hand side of (21) is equal to

$$\prod_{1 \leqq i < j \leqq n} \left(\frac{a_i + a_j}{2} \right)^{2\alpha w_i w_j} \prod_{i=1}^{n} a^{\{\alpha w_i^2 + (1-\alpha) w_i\}}.$$

Since $\alpha w_i^2 + (1 - \alpha) w_i \geqq 0$ and $\sum_{1 \leqq i < j \leqq n} 2\alpha w_i w_j + \sum_{i=1}^{n} \{\alpha w_i^2 + (1 - \alpha) w_i\} =$

1, an application of (14) gives (21).

The proof of the right hand inequality in (20) is slightly longer. The proof is by induction on n and the result is trivial when $n = 1$. By rewriting, the inequality is equivalent to

$$\beta_n(a) = \frac{A_n^w G_n^{1-w}}{\Pi_n} \leqq 1 . \tag{22}$$

Using (17) and (18) it is easy to show that

$$\beta_n(a) = \frac{\{(1-w_n)A_{n-1} + w_n a_n\}^w G_{n-1}^{(1-w_n)(1-w)} a_n^{w_n(1-w_n-w)}}{\Pi_{n-1} \cdot \prod_{i=1}^{n-1} \left(\frac{a_i + a_n}{2}\right)^{2w_i w_n}} .$$

In particular therefore, if $a_1 = \cdots = a_{n-1} = \alpha$,

$$\beta_n(a) = \frac{\{(1-w_n)\alpha + w_n a_n\}^w \alpha^{(1-w_n)(w_n-w)} a_n^{w_n(1-w_n-w)}}{\left(\frac{a_n + \alpha}{2}\right)^{2w_n(1-w_n)}} . \tag{23}$$

Further if $v = \min(w_1, \cdots, w_{n-1})$ then $v \geqq w$ and

$$\beta_{n-1}(a_1, \cdots, a_{n-1}) = \frac{A_{n-1}^v G_{u-1}^{1-v}}{\Pi_{n-1}^{1/(1-w_n)^2}} .$$

Now, since $1 - w_n - w \geqq 0$ and $w + (1-w_n)(1-w) + w_n(1-w_n-w) = 2w_n(1-w_n)$, an application of (14) to (23) demonstrates (22) in this special case.

If we now assume $\beta_{n-1} \leqq 1$ then

$$\beta_n \leqq \frac{\beta_n}{\beta_{n-1}^{(1-w_n)^2}} = \frac{\{(1-w_n)A_{n-1} + w_n a_n\}^w G_{n-1}^{(1-w_n)(1-w)-(1-w_n)^2(1-v)} a_n^{w_n(1-w_n-w)}}{A_{n-1}^{v(1-w_n)^2} \prod_{i=1}^{n} \left(\frac{a_i + a_n}{2}\right)^{2w_i w_n}}$$

$$\leqq \frac{\{(1-w_n)A_{n-1} + w_n a_n\}^w A_{n-1}^{(1-w_n)(w_n-w)} a_n^{w_n(1-w_n-w)}}{\prod_{i=1}^{n-1} \left(\frac{a_i + a_n}{2}\right)^{2w_i w_n}}$$

using (14). Without any loss of generality we can assume that $a_n = \max(a_1, \cdots, a_n)$, when in particular $a_n \geqq A_{n-1}$. Then

$$\beta_n(a) \leqq \frac{\{(1-w_n)A_{n-1} + w_n a_n\}^w A_{n-1}^{(1-w_n)(w_n-w)} a_n^{w_n(1-w_n-w)}}{\left(\frac{a_n + A_{n-1}}{2}\right)^{2w_n(1-w_n)}}$$

$$\leqq 1 ,$$

by the particular case (23) with $\alpha = A_{n-1}$.

The cases of equality are immediate.

It might be remarked that if W' is second largest and w' the second smallest of (w) tnen

$$1 \leqq L_n^{ww'} \leqq \varLambda_n \leqq L_n^{WW'} .$$

It is possible to generalise Hölder's inequality using Theorem 9.

THEOREM 10. *Let* $a_{ij} \geqq 0$ $(i = 1, \cdots, m,\ j = 1, \cdots, n)$ *and*

$$\sum_{i=1}^{m} a_{ij} = s_j > 0 \qquad (j = 1, \cdots, n) .$$

Then

$$D \prod_{j=1}^{n} s_j^{w_j} \geqq \sum_{i=1}^{m} \prod_{j=1}^{u} a_{ij}^{w_j} \geqq d \prod_{j=1}^{n} s_j^{w_j}$$

where

$$D = \min(1, L, \varLambda)$$
$$d = \max(l, \lambda)$$

and

$$L = \max_i L_n^{-(w/n-1)}(a_{i1}, \cdots, a_{in}) = \max_i L_{n\,i}^{-(w/n-1)}$$
$$= \max_i \varLambda_n^{-(1/1-w)}(a_{i1}, \cdots, a_{in}) = \max_i \varLambda_{n\,i}^{-(1/1-w)}$$
$$l = \min_i L_{n,i}^{-W} ,$$
$$\lambda = \min_i \varLambda_{n,i}^{-1/w} .$$

Proof. A simple modification of the usual proof [5, p. 23] using Theorem 9 instead of (14).

REFERENCES

1. P. Bullen and M. Marcus, *Symmetric means and matrix inequalities*, Proc. Amer. Math. Soc. **12** (1961), 286-290.
2. P. H. Diananda, *On some inequalities of H. Kober*, Proc. Camb. Phil. Soc. **59** (1963), 341-346.
3. W. N. Everitt, *On the Hölder inequality*, J. London Math. Soc. **36** (1961), 145-148.
4. ———, *On an inequality for the generalised arithmetic and geometric means*, Amer. Math. Monthly **70** (1963), 251-253.
5. Hardy, Littlewood and Polya, *Inequalities*, Cambridge, 1934.
6. H. Kober, *On the arithmetic and geometric means and on Hölder's inequality*, Proc. Amer. Math. Soc. **9** (1958), 452-459.

PACIFIC JOURNAL OF MATHEMATICS
Vol. 15, No. 1, 1965

ON ARITHMETIC PROPERTIES OF COEFFICIENTS OF RATIONAL FUNCTIONS

DAVID G. CANTOR

The purpose of this note is to prove the following generalization of a result of Polya:

THEOREM. **Let $\{a_n\}$ be a sequence of algebraic integers, and f a nonzero polynomial with complex coefficients. If $\sum_{n=0}^{\infty} f(n)a_n z^n$ is a rational function, then so is $\sum_{n=0}^{\infty} a_n z^n$.**

Polya [3] has proved that if $\sum_{n=0}^{\infty} na_n z^n$ is a rational function, then so is $\sum_{n=0}^{\infty} a_n z^n$. It follows immediately from Polya's result that if k is a rational integer and $\sum_{n=0}^{\infty} (n-k)a_n z^n$ is a rational function, then so is $\sum_{n=0}^{\infty} a_n z^n$. It is then easy to prove inductively, that if f is a polynomial with complex coefficients, all of whose roots are rational integers, and if $\sum_{n=0}^{\infty} f(n)a_n z^n$ is a rational function, then so is $\sum_{n=0}^{\infty} a_n z^n$.

Suppose K is an algebraic number field and $A \subset K$ is an ideal. If α and β are algebraic numbers in K, we say, as usual, that $\alpha \equiv \beta(A)$, if there exists a rational integer r, relatively prime to A, such that $r\alpha$ and $r\beta$ are algebraic integers and $(r\alpha - r\beta) \in A$. We say that A divides the numerator (denominator) of α if $\alpha \equiv 0(A)$ $((1/\alpha) \equiv 0(A))$. We denote the norm of the ideal A by NmA.

LEMMA 1. *Let K be an algebraic number field and $\alpha \in K$ an algebraic number. Then the set of those prime ideals of K which divide the numerator of some element of the sequence $\{k - \alpha : k = 1, 2, 3, \cdots\}$ is infinite.*

Proof. Suppose n is a rational integer such that $n\alpha$ is an algebraic integer, and suppose $P_1, P_2, \cdots, P_r$ are the only prime ideal divisors of the sequence $\{nk - n\alpha : k = 1, 2, 3, \cdots\}$. Now $Nm(nk - n\alpha)$ is a nonconstant polynomial $g(k)$ with rational integral coefficients. Hence for each rational integer k, there exist rational integers $s_1, s_2, \cdots, s_r$ such that $g(k) = \mp \prod_{i=1}^{r} (NmP_i)^{s_i}$. Thus there are only finitely many rational primes which divide some element of the sequence $\{g(k) : k = 1, 2, 3, \cdots\}$. But this is false [2, p. 82].

REMARK. A less elementary proof of Lemma 1 is obtained by observing that if P is a prime ideal with residue class degree 1, and not dividing the denominator of α, then there exists a rational integer

Received August 13, 1963.

n such that $n \equiv \alpha(P)$; since the set of such prime ideals has Dirichlet density 1, among all prime ideals, there are infinitely many of them.

LEMMA 2. *Suppose $\{a_n\}$ is a sequence of algebraic integers and α is an algebraic number. If $\sum_{n=0}^{\infty} (n - \alpha)a_n z^n$ is a rational function then so is $\sum_{n=0}^{\infty} a_n z^n$.*

Proof. Since $\sum_{n=0}^{\infty} (n - \alpha)a_n z^n$ is a rational function, there exist distinct nonzero algebraic numbers $\theta_1, \theta_2, \cdots, \theta_m$ and polynomials with algebraic coefficients $\lambda_1, \lambda_2, \cdots, \lambda_m$ such that

$$(1) \qquad (n - \alpha)a_n = \sum_{i=1}^{m} \lambda_i(n)\theta_i^n ,$$

for all $n \geqq n_0$, where n_0 is a rational integer. By replacing the sequence $\{a_n\}$ by the sequence $\{a_{n+n_0}\}$ if necessary, we may assume that (1) holds for all $n \geqq 0$. Let K be an algebraic number field which contains α, the coefficients of the λ_i, and the θ_i. Choose a rational integer k and a prime ideal $P \subset K$ such that P divides the numerator of $k - \alpha$ and does not divide the numerator or denominator of α, the θ_i, the differences $(\theta_i - \theta_j)\,(i \neq j)$, and the coefficients of the λ_i; by Lemma 1, there are infinitely many choices for the prime ideal P. Suppose that $NmP = p^f$ where p is a rational prime. We substitute $n = k + jp^f$ in (1), where j is a rational integer:

$$(k + jp^f - \alpha)a_n = \sum_{i=1}^{m} \lambda_i(k + jp^f)\theta_i^{k+jpf} .$$

Since $p^f \equiv 0(P)$ and $k \equiv \alpha(P)$, we obtain

$$0 \equiv \sum_{i=1}^{m} \lambda_i(\alpha)\theta_i^k \theta_i^{jpf}(P) .$$

But $\theta_i^{jpf} \equiv \theta_i^j(P)$, hence

$$(2) \qquad \sum_{i=1}^{m} \lambda_i(\alpha)\theta_i^{k+j} \equiv 0(P) .$$

The m equations obtained from (2) by successively substituting $j = 0, 1, 2, \cdots, m-1$ are linear in the $\lambda_i(\alpha)$ and have as determinant $\prod_{i=1}^{m} \theta_i^k$ times the Vandermonde determinant $\det \| \theta_i^j \|$, $1 \leqq i \leqq m$, $0 \leqq j \leqq m - 1$, which is not $\equiv 0(P)$, since P does not divide any of the θ_i or the differences $(\theta_i - \theta_j)\,(i \neq j)$. Hence

$$(3) \qquad \lambda_i(\alpha) \equiv 0(P), 1 \leqq i \leqq m .$$

By Lemma 1, (3) is true for infinitely many prime ideals P, hence $\lambda_i(\alpha) = 0, 1 \leqq i \leqq m$. It follows that the polynomials $\lambda_i(n)$ are divisible by $n - \alpha$. Put $\mu_i(n) = \lambda_i(n)/(n - \alpha)$; $\mu_i(n)$ is a polynomial with algebraic coefficients. By (1)

$$a_n = \sum_{i=1}^{m} \mu_i(n)\theta_i^n .$$

Thus $\sum_{n=0}^{\infty} a_n z^n$ is a rational function.

LEMMA 3. *Suppose $\{a_n\}$ is a sequence of algebraic numbers and f is a nonzero polynomial with complex coefficients. If $\sum_{n=0}^{\infty} f(n)a_n z^n$ is a rational function, then there exists a nonzero polynomial g with algebraic coefficients snch that $\sum_{n=0}^{\infty} g(n)a_n z^n$ is a rational function.*

Proof. There exist distinct nonzero complex numbers $\theta_1, \theta_2, \cdots, \theta_m$ and nonzero polynomials with complex coefficients $\lambda_1, \lambda_2, \cdots, \lambda_m$ such that

$$(4) \qquad f(n)a_n = \sum_{i=1}^{m} \lambda_i(n)\theta_i^n ,$$

for all large n. Without loss of generality, we may assume that (4) holds for all $n \geqq 0$. In what follows, all fields are considered as subfields of the field of complex numbers. Denote by Ω the field of algebraic numbers, and by L the smallest field which contains Ω, the θ_i, and all of the coefficients of the polynomials $f, \lambda_1, \lambda_2, \cdots, \lambda_m$.

Since L is finitely generated over Ω, it has a finite transcendence basis $x_1, x_2, \cdots, x_r$. Each of the θ_i, the coefficients of the λ_i, and the coefficients of f satisfies an irreducible polynomial equation whose coefficients are elements of $\Omega[x_1, x_2, \cdots, x_r]$. Let $h_1, h_2, \cdots, h_s$ be all of the nonzero coefficients of these polynomials; $h_1, h_2, \cdots, h_s$ are polynomials in $x_1, x_2, \cdots, x_r$ with coefficients in Ω. Since there are only finitely many such polynomials, there exist algebraic numbers $\xi_1, \xi_2, \cdots, \xi_r$ such that $h(\xi_1, \xi_2, \cdots, \xi_r) \neq 0, 1 \leqq i \leqq s$. The map $x_i \rightarrow \xi_i$ gives rise to a homomorphism of the ring $\Omega[x_1, x_2, \cdots, x_r]$ onto Ω, which is the identity on Ω. By the extension of place theorem [1, p. 8], this homomorphism can be extended to a place $\varphi : L \rightarrow \Omega$, which is the identity on Ω. If $\alpha \in L$, we denote by $\bar{\alpha}$ the image of α under φ and if b is a polynomial, $b(n) = \sum_{i=1}^{t} b_i n^i$ with coefficients $b_i \in L$, we denote by $\bar{b}$ the polynomial with $\bar{b}(n) = \sum_{i=1}^{t} \bar{b}_i n^i$. The θ_i and the coefficients of $f, \lambda_1, \lambda_2, \cdots, \lambda_m$ satisfy nonconstant polynomials $g_1, g_2, \cdots, g_v$ with nonzero constant term; the nonzero coefficients of these polynomials are the h_j. Under the place φ the h_j go into finite nonzero algebraic numbers $\bar{h}_j$. Hence the polynomial $\bar{g}_k$ has the same degree as g_k, all of its terms are finite, and its constant term is not zero $(1 \leqq k \leqq v)$. The $\bar{\theta}_i$ and the coefficients of $\bar{f}, \bar{\lambda}_1, \bar{\lambda}_2, \cdots, \bar{\lambda}_r$ are roots of these polynomials; hence the $\bar{\theta}_i$ are finite, nonzero algebraic numbers, and the $\bar{f}, \bar{\lambda}_1, \bar{\lambda}_2, \cdots, \bar{\lambda}_m$ are nonzero polynomials, with finite, algebraic coefficients. Applying the place φ to both terms in (4), and putting $\bar{f} = g$, yields, since $\bar{a}_n = a_n$

$$g(n)a_n = \sum_{i=1}^{m} \bar{\lambda}_i(n)\bar{\theta}_i^n .$$

Hence

$$\sum_{n=0}^{\infty} g(n)a_n z^n = \sum_{n=0}^{\infty} \sum_{i=1}^{m} \bar{\lambda}_i(n)\bar{\theta}_i^n z^n$$

is a rational function, and g is a nonzero polynomial with algebraic coefficients.

Proof of theorem. By Lemma 3, we may assume that f has algebraic integer coefficients. Let α be a root of f and $g(n) = f(n)/(n - \alpha)$; by the lemma of Gause, $g(n)$ is a polynomial with algebraically integral coefficients. Put $b_n = g(n)a_n$; $\{b_n\}$ is a sequence of algebraic integers and $\sum_{n=0}^{\infty} (n - \alpha)b_n z^n$ is a rational function. By Lemma 2, so is $\sum_{n=0}^{\infty} b_n z^n$. Proceeding inductively, on the degree of f, we see that $\sum_{n=0}^{\infty} a_n z^n$ is a rational function.

REMARK. By the Remark following Lemma 1, one can replace, in the theorem, the requirement that the a_n be integers, by the requirement that the set of prime ideal divisors of the denominators of the a_n has Dirichlet density less than 1 among all prime ideals.

Let $f(z) = \sum_{n=0}^{\infty} a_n z^n$, where the a_n are rational integers. Polya's theorem then asserts that if $f'(z)$ is a rational function, so is $f(z)$. The corresponding assertion of our generalization of Polya's theorem is: Let $f(z) = \sum_{n=0}^{\infty} a_n z^n$ be a power series with algebraically integral coefficients. If there exists a nonzero differential operator L, of the form $L = \sum_{i=0}^{r} c_i(zd/dz)^i$ (c_i complex numbers), such that Lf is a rational function, then so is $f(z)$.

REFERENCES

1. S. Lang, *Introduction to algebraic geometry*, New York, 1958.
2. T. Nagell, *Introduction to number theory*, New York, 1958.
3. G. Polya, *Arithmetische Eigenschaften der Reihenentwicklungen*, J. Reine u. angew. Math., **151** (1921), 1-31.

INSTITUTE FOR DEFENSE ANALYSES
UNIVERSITY OF WASHINGTON

PACIFIC JOURNAL OF MATHEMATICS
Vol. 15, No. 1, 1965

DEDEKIND DOMAINS AND RINGS OF QUOTIENTS

LUTHER CLABORN

We study the relation of the ideal class group of a Dedekind domain A to that of A_S, where S is a multiplicatively closed subset of A. We construct examples of (a) a Dedekind domain with no principal prime ideal and (b) a Dedekind domain which is not the integral closure of a principal ideal domain. We also obtain some qualitative information on the number of non-principal prime ideals in an arbitrary Dedekind domain.

If A is a Dedekind domain, S the set of all monic polynomials and T the set of all primitive polynomials of $A[X]$, then $A[X]_S$ and $A[X]_T$ are both Dedekind domains. We obtain the class groups of these new Dedekind domains in terms of that of A.

1. LEMMA 1-1. *If A is a Dedekind domain and S is a multiplicatively closed set of A such that A_S is not a field, then A_S is also a Dedekind domain.*

Proof. That A_S is integrally closed and Noetherian if A is, follows from the general theory of quotient ring formations. The primes of A_S are of the type PA_S, where P is a prime ideal of A such that $P \cap S = \phi$. Since height $PA_S =$ height P if $P \cap S = \phi$, $P \neq (0)$ and $P \cap S = \phi$ imply that height $PA_S = 1$.

PROPOSITION 1-2. If A is a Dedekind domain and S is a multiplicatively closed set of A, the assignment $C \rightarrow CA_S$ is a mapping of the set of fractionary ideals of A onto the set of fractionary ideals of A_S which is a homomorphism for multiplication.

Proof. C is a fractionary ideal of A if and only if there is a $d \in A$ such that $dC \subseteqq A$. If this is so, certainly $dCA_S \subseteqq A_S$, so CA_S is a fractionary ideal of A_S. Clearly $(B \cdot C)A_S = BA_S \cdot CA_S$, so the assignment is a homomorphism. Let D be any fractionary ideal of A_S. Since A_S is a Dedekind domain, D is in the free group generated by all prime ideals of A_S, i.e. $D = Q_i^{e_1} \cdots Q_k^{e_k}$. For each $i = 1, \cdots, k$ there is a prime P_i of A such that $Q_i = P_iA_S$. Set $E = P_1^{e_1} \cdots P_k^{e_k}$. Then using the fact that we have a multiplicative homomorphism of fractionary ideals, we get

Received December 13, 1963.

$$EA_S = (P_1A_S)^{e_1} \cdots (P_kA_S)^{e_k} = Q_1^{e_1} \cdots Q_k^{e_k}.$$

COROLLARY 1–3. *Let A be a Dedekind domain and S be a multiplicatively closed set of A. Let $\bar{C}$ (for C a fractionary ideal of A or A_S) denote the class of the ideal class group to which C belongs. Then the assignment $\bar{C} \to \overline{CA_S}$ is a homomorphism φ of the ideal class group of A onto that of A_S.*

Proof. It is only necessary to note that if $C = dA$, then $CA_S = dA_S$.

THEOREM 1–4. *The kernel of φ is generated by all $\bar{P}_\alpha$, where P_α ranges over all primes such that $P_\alpha \cap S \neq \phi$.*

If $P_\alpha \cap S \neq \phi$, then $P_\alpha A_S = A_S$. Suppose C is a fractionary ideal such that $\bar{C} = \bar{P}_\alpha$, i.e. $C = dP_\alpha$ for some d in the quotient field of A. Then $CA_S = dP_\alpha A_S = dA_S$, and thus $\overline{CA_S}$ is the principal class.

On the other hand, suppose that C is a fractionary ideal of A such that $CA_S = xA_S$. We may choose x in C. Then $C^{-1} \cdot xA$ is an integral ideal of A, and $(C^{-1} \cdot xA)A_S = A_S$. In other words, $C^{-1} \cdot xA = P_1^{f_1} \cdots P_l^{f_l}$, where $P_i \cap S \neq \phi$, $i = 1, \cdots, l$. Then $\bar{C} = \bar{P}_1^{-f_1}, \cdots, -\bar{P}_l^{-f_l}$, completing the proof.

EXAMPLE 1–5. There are Dedekind domains with no prime ideals in the principal class.

Let A be any Dedekind domain which is not a principal ideal domain. Let S be the multiplicative set generated by all Π_α, where Π_α ranges over all the prime elements of A. Then by Theorem 1–4, A_S will have the same class group as A but will have no principal prime ideals.

COROLLARY 1–6. *If A is a Dedekind domain which is not a principal ideal domain, then A has an infinite number of non-principal prime ideals.*

Proof. Choose S as in Example 1–5. Then A_S is not a principal ideal domain, hence has an infinite number of prime ideals, none of which are principal. These are of the form PA_S, where P is a (non-principal) prime of A.

COROLLARY 1–7. *Let A be a Dedekind domain with torsion class group and let $\{P_\alpha\}$ be a collection of primes such that the subgroup of the ideal class group of A generated by $\{\bar{P}_\alpha\}$ is not the entire* *class group. Then there are always an infinite number of non-principal primes not in the set* $\{P_\alpha\}$.

Proof. For each α, chose n_α such that $P_\alpha^{n_\alpha}$ is principal, say $= A\cdot a_\alpha$. Let S be the multiplicatively closed set generated by all a_α. By Theorem 1–4, A_S is not a principal ideal domain, hence A_S must have an infinite number of non-principal prime ideals by Corollary 1–6. These come from non-principal prime ideals of A which do not meet S. Each P_α does meet S, so there are an infinite number of non-principal primes outside the set $\{P_\alpha\}$.

COROLLARY 1–8. *Let A be a Dedekind domain with at least one prime ideal in every ideal class. Then for any multiplicatively closed set S, A_S will have a prime ideal in every class except possibly the principal class.*

Proof. By Corollary 1–3, every class of A_S is the image of a class of A. Let $\bar{D}$ be a non-principal class of A_S. $\bar{D} = \overline{CA_S}$, where C is a fractionary ideal of A. By assumption, there is a prime P of A such that $\bar{P} = \bar{C}$. If $PA_S = A_S$, then CA_S is principal and so $\bar{D}$ is the principal class of A_S. This is not the case, so PA_S is prime, and certainly $\overline{PA_S} = \overline{CA_S} = \bar{D}$.

EXAMPLE 1–9. There is a Dedekind domain which is not the integral closure of a principal ideal domain.

Let $A = Z[\sqrt{-5}]$. A is a Dedekind domain which is not a principal ideal domain. In A, $29 = (3 + 2\sqrt{-5})\cdot(3 - 2\sqrt{-5})$. It follows from elementary algebraic number theory that $\Pi_1 = 3 + 2\sqrt{-5}$ and $\Pi_2 = 3 - 2\sqrt{-5}$ generate distinct prime ideals of A. Let $S = \{\Pi_1^k\}_{k\geqq 0}$. Then A_S is by Theorem 1–4 a Dedekind domain which is not a principal ideal domain. Let F denote the quotient field of A and Q the rational numbers. A_S cannot be the integral closure of a principal ideal domain whose quotient field is F since principal ideal domains are integrally closed. If A_S were the integral closure of a principal ideal domain C with quotient field Q, then C would contain Z, and Π_1 and Π_2 would be both units or nonunits in A_S (since Π_1 and Π_2 are conjugate over Q). But only Π_1 is a unit in A_S.

REMARK 1–10. Example 1–9 settles negatively a conjecture in Vol. I of *Commutative Algebra* [2, p. 284]. The following conjecture may yet be true: Every Dedekind domain can be realized as an A_S, where A is the integral closure of a principal ideal domain in a finite extension field and S is a multiplicatively closed set of A.

2. Lemma 2–1. *Let A be a Dedekind domain. Let S be the multiplicatively closed set of $A[X]$ consisting of all monic polynomials of $A[X]$. Let T be the multiplicatively closed set of all primitive polynomials of $A[X]$ (i.e. all polynomials whose coefficients generate the unit ideal of A). Then $A[X]_S$ and $A[X]_T$ are both Dedekind domains.*

Proof. $A[X]$ is integrally closed and noetherian, and so both $A[X]_S$ and $A[X]_T$ are integrally closed and noetherian. Let P be a prime ideal of $A[X]$. If $P \cap A \neq (0)$, then $P \cap A = Q$ is a maximal ideal of A. If $P \neq QA[X]$, then passing to $A[X]/QA[X]$, it is easy to see that $P = QA[X] + f(X) \cdot A[X]$ where $f(X)$ is a suitably chosen monic polynomial of $A[X]$. In this case $P \cap S \neq \phi$, so $PA[X]_S = A[X]_S$. Thus if $P \cap A \neq (0)$ and $PA[X]_S$ is a proper prime of $A[X]_S$, then $P = QA[X]$ where $Q = P \cap A$. Then height $P =$ height $Q = 1$. If $P \cap A = (0)$, then $PK[X]$ is a prime ideal of $K[X]$ (where K denotes the quotient field of A). Certainly height $P =$ height $PK[X] = 1$, so in any case if a prime P of $A[X]$ is such that $P \cap S = \phi$, then height $P \leqq 1$. This proves that $A[X]_S$ is a Dedekind domain. Since $S \subseteqq T$, $A[X]_T$ is also a Dedekind domain by Lemma 1–1.

Remark 2–2. $A[X]_T$ is customarily denoted by $A(X)$ [1, p. 18]. For the remainder of this article, $A[X]_S$ will be denoted by A^1.

Proposition 2–3. *A^1 has the same ideal class group as A. In fact, the map $\bar{C} \to \overline{CA^1}$ is a one-to-one map of the ideal class group of A onto that of A^1.*

We can prove that $\bar{C} \to \overline{CA^1}$ is a one-to-one map of the ideal class of A into that of A by showing that if two integral ideals D and E of A are not in the same class, neither are DA^1 and EA^1. Suppose then that $\overline{DA^1} = \overline{EA^1}$. This implies that there are elements $f_i(X)$, $g_i(X)$, $i = 1,2$ in $A[X]$ with $g_i(X)$ monic for $i = 1,2$ such that

$$DA^1 \cdot \frac{f_1(X)}{g_1(X)} = EA^1 \cdot \frac{f_2(X)}{g_2(X)} .$$

Let a_i be the leading coefficient of $f_i(X)$ for $i = 1,2$, and let $d \in D$. Then we get a relation

$$d \cdot \frac{f_1(X)}{g_1(X)} = \frac{e(X)}{g(X)} \cdot \frac{f_2(X)}{g_2(X)} , \ g(X) \text{ monic},$$

where $e(X)$ can be chosen as a polynomial in $A[X]$ all of whose coefficients are in E. This leads to $d\, g_2(X) \cdot f_1(X) \cdot g(X) = e(X) \cdot f_2(X) \cdot g_1(X)$. The leading coefficient on the right is in $a_2 \cdot E$. This shows that $a_1 \cdot D$

$D \subseteqq a_2 \cdot E$. Likewise $a_2 \cdot E \subseteqq a_1 \cdot D$, thus $a_1 \cdot D = a_2 \cdot E$ and $\bar{D} = \bar{E}$.

To prove the map is onto, the following lemma is needed.

LEMMA 2–4. *Let A be a Dedekind domain with quotient field K. To each polynomial $f(X) = a_n X^n + \cdots - + a_o$ of $K[X]$ assign the fractionary ideal $c(f) = (a_n, \cdots, a_o)$. Then $c(fg) = c(f)\,c(g)$.*

Proof. Let V_p (for each prime P of A) denote the P-adic valuation of A. It is immediate that $V_p(c(f)) = \min\ V_p(a_i)$. Because of the unique factorization of fractionary ideals in Dedekind domains, it suffices to show that $V_p(c(fg)) = V_p(c(f)) + V_p(c(g))$ for each prime P of A. This will be true if the equation is true in each $A_p[X]$. But A_p is a principal ideal domain, and the well-known proof for principal ideal domains shows the truth of the lemma.

To complete Prop. 2–3, let P be a prime ideal of A^1. The proof of Lemma 2–1 shows that if $P \cap A \neq (0)$, then $P = QA^1$ where Q is a prime of A. Thus $\bar{P} = \overline{QA^1}$ and ideal classes generated by these primes are images of classes of A. Suppose now that P is a prime of A^1 such that $P \cap A = (0)$. Let $P^1 = P \cap A[X]$. Then $P^1 \cap A = (0)$, and $P^1 \cdot K[X]$ is a prime ideal of $K[X]$. Let $P^1 \cdot K[X] = f(X)K[X]$; we may choose $f(X)$ in $A[X]$. Let $C = c(f)$. Suppose that $g(X) \cdot f(X) \in A[X]$. Then because $c(fg) = (c(f)) + (c(g)) \geqq 0$ for all P, $g(X) \in C^{-1} \cdot A[X]$. Conversely if $g(X) \in C^{-1} \cdot A[X]$, then $g(X)\,f(X) \in A[X]$. Thus $P^1 = f(X)K[X] \cap A[X] = C^{-1} \cdot A[X] \cdot f(X)A[X]$, and $P = P^1 A^1 = C^{-1} \cdot A^1 \cdot f(X)A^1$. This gives finally that $\bar{P} = \overline{C^{-1}A^1}$, and the class is an image of a class of A under our map. Since the ideal class group of A^1 is generated by all $\bar{P}$ where P is a prime of A^1, this finishes the proof.

COROLLARY 2–5. *A^1 has a prime ideal in each ideal class.*

Proof. Let w be any nonunit of A. Then $(wX + 1)\,K[X] \cap A^1$ $(= (wX + 1)A^1)$ is a prime ideal in the principal class. Otherwise let C be any integral ideal in a nonprincipal class $\bar{D}^{-1}$. C can be generated by 2 elements, so suppose $C = (c_0, c_1)$; then $Q = (c_0 + c_1 X) \cdot K[X] \cap A^1$ is a prime ideal in $\overline{C^{-1}A^1} = \bar{D}$.

PROPOSITION 2–6. If A is a Dedekind domain, then $A(X)$ is a principal ideal domain.

Proof. Since $A(X) = A^1_T$, Corollary 1–3 and the proof of Corollary 2–5 show that each nonprincipal class of $A(X)$ contains a prime $QA(X)$, where Q is a prime ideal of A of the type $(c_0 + c_1 X)K[X] \cap A^1$. Clearly $Q \cap A[X] = (c_0 + c_1 X)K[X] \cap A[X] = C^{-1} \cdot A[X] \cdot (c_0 + c_1 X)A[X] \not\subseteqq$

$PA[X]$ for any prime P of A. Thus there is in $Q \cap A[X]$ a primitive polynomial of $A[X]$. Thus $QA(X) = A(X)$. Theorem 1–4 now implies that every class of A becomes principal in $A(X)$, i.e. $A(X)$ is a principal ideal domain.

REMARK 2–7. Proposition 2–6 is interesting in light of the fact that the primes of $A(X)$ are exactly those of the form $PA(X)$, where P is a prime of A [1, p. 18].

REMARK 2–8. If the conjecture given in Remark 1–10 is true for a Dedekind domain A, it is also true for A^1. For suppose $A = B_M$, where M is a multiplicatively closed set of B and B is the integral closure of a principal ideal domain B_0 in a suitable finite extension field. Let S, S^1, and T be the set of monic polynomials in $A[X]$, $B[X]$, and $B_0[X]$ respectively. Then $A^1 = A[X]_S = (B_M[X])_S = (B[X]_M)_S = (B[X])_{\langle M,S\rangle} = (B[X]_{S^1})_{\langle M,S\rangle}$. The last equality holds because $S^1 \subseteqq S \subseteqq \langle M, S\rangle$. It is easy to see that $B[X]_{S^1}$ is the integral closure of the principal ideal domain $B_0[X]_T$ in $K(X)$, where K is the quotient field of B.

REFERENCES

1. M. Nagata, *Local rings*, New York, Interscience Publishers, Inc. (1962).

2. O. Zariski and P. Samuel, *Commutative algebra*, Vol. I, Princeton, D. Van Nostrand Company (1958).

CORNELL COLLEGE

PACIFIC JOURNAL OF MATHEMATICS
Vol. 15, No. 1, 1965

HOMOTOPY COMMUTATIVITY AND THE MOORE SPECTRAL SEQUENCE

ALLAN CLARK

This paper initiates the study of strong homotopy commutativity is both geometric and algebraic contexts in order to correct an error in a paper of J. C. Moore.

The difficulty [8, § 7, Theorem II] lies in the tacit assumption here (and in the remark following Proposition 7.1) that the multiplication $m: X \times X \to X$ on an H-space X induces a morphism of H-spaces $\Omega X \times \Omega X \to \Omega X$ where ΩX denotes the associative loop space of X defined in [6, Chapter 2]. Unfortunately the situation is more complex than this. A morphism of H-spaces $\Omega m: \Omega(X \times X) \to \Omega X$ is induced by the product m on X. However for associative loop spaces, $\Omega X \times \Omega X$ is not the same as $\Omega(X \times X)$, although it has the same homotopy type. Moreover there is no obvious morphism of H-spaces from $\Omega X \times \Omega X$ to $\Omega(X \times X)$ with which the induced morphism Ωm could be composed.

There are three ways to resolve this problem so that the proofs involved can be carried through. One is to assume that X is associative, use the product induced from X as the product in the ordinary loop space, and to take the usual loop product as the morphism of H-spaces. Another way would be to use the product induced on the ordinary loop space as a morphism of Stasheff's A_∞ structures [10]. The third way, the one used in this paper, is to show that there is a strongly homotopy multiplicative map of H-spaces $\Omega X \times \Omega X \to \Omega X$, and that this is sufficient for the proofs desired. The second and third alternatives are homotopy equivalent, and the third is preferred in order to use the bar construction rather than the less familiar tilde construction of Stasheff [10].

The exposition is organized as follows: § 1 sets up the geometry and discusses strong homotopy commutativity; § 2 recalls the bar construction and the definition of the Moore spectral sequence; § 3 defines the algebraic analogue of strong homotopy commutativity for a differential graded algebra A and uses it to construct a product in the bar construction $\bar{B}(A)$, and consequently to introduce a Hopf algebra structure into the Moore spectral sequence; § 4 proves a homology suspension theorem for a contractible fibre space over an H-space, a slight improvement on a theorem of Browder [1, Theorem 5.13] which contains the original result of Moore in which the trouble began.

The comments of the referee resulted in substantial improvement

Received November 14, 1963 and in revised form March 5, 1964. The author was partially supported by a National Science Foundation grant.

in the author's exposition, and the author is indebted to W. Browder for helpful conversations.

1. The geometric situation. X will denote a pathwise connected and simply connected H-space with multiplication $m: X \times X \to X$ which has a two-sided unit element e. In addition we assume that X has *finite homological type*, that is, that the singular homology groups of X are finitely generated in every degree.

ΩX will denote the associative loop space of X as defined in [6, Chapter 2], and $\phi: \Omega X \times \Omega X \to \Omega X$ will denote the loop multiplication. ϕ gives ΩX the structure of an associative H-space with unit element Ωe, the unique loop of length zero. Clearly $\Omega X \times \Omega X$ is also an associative H-space with multiplication $(\phi \times \phi)(1 \times T \times 1)$ where T denotes the standard twisting map. Note that (unlike the ordinary loop space) $\Omega X \times \Omega X$ is not the same as $\Omega(X \times X)$.

DEFINITION 1.1. A map $f: Y \to Z$ where Y and Z are associative H-spaces is *strongly homotopy multiplicative* if there exist maps

$$M_n: Y \times (I \times Y)^n \to Z$$

for every nonnegative integer n, such that $M_0 = f$, and such that

$$M_n(y_0, t_1, y_1, \cdots, t_n, y_n) = \begin{cases} M_{n-1}(y_0, t_1, \cdots, t_{i-1}, y_{i-1}y_i, t_{i+1}, \cdots, t_n, y_n) & \text{for } t_i = 0 \\ M_{i-1}(y_0, t_1, \cdots, t_{i-1}, y_{i-1})M_{n-i}(y_i, t_{i+1}, \cdots, t_n, y_n) & \text{for } t_i = 1 \,. \end{cases}$$

The definition is due to Sugawara [11].

If X is an associative H-space, then B_X will denote the classifying space of X as constructed in Dold and Lashof [3].

THEOREM 1.2. (Sugawara [11, Lemma 2.2]). *If $f: Y \to Z$ is a strongly homotopy multiplicative map, then f induces a map $B_f: B_Y \to B_Z$.*

DEFINITION 1.3. An associative H-space X is *strongly homotopy commutative* if there exists a strongly homotopy multiplicative map $f: X \times X \to X$, such that $f(e, x) = x = f(x, e)$.

THEOREM 1.4. (Sugawara [11, Theorem 4.3]). *If X is an H-space with associative multiplication, then B_X is an H-space if and only if the multiplication on X is strongly homotopy commutative.*

COROLLARY 1.5. *X is an H-space if and only if ΩX is strongly homotopy commutative.*

PROPOSITION 1.6. For any pathwise connected spaces $X_1, \cdots, X_m$

there exists a strongly homotopy multiplicative map

$$\psi\colon \Omega X_1 \times \cdots \times \Omega X_m \to \Omega(X_1 \times \cdots \times X_m) .$$

Proof. $\Omega(X_1 \times \cdots \times X_m)$ may be considered to be the subspace of $\Omega X_1 \times \cdots \times \Omega X_m$ consisting of m-tuples of loops of equal length. If $(\omega_1, \cdots, \omega_m) \in \Omega X_1 \times \cdots \times \Omega X_m$ and $\omega_i\colon [0, s_i] \to X_i$, then we define $\psi(\omega_1, \cdots, \omega_m) = (\bar{\omega}_1, \cdots, \bar{\omega}_m)$ where $\bar{\omega}_i = \omega_i\{s - s_i\}$ where $s = \max\{s_i\}$ and $\{s - s_i\}$ denotes the constant loop of length $s - s_i$.

The homotopies M_n are complicated to define. For $i = 0, \cdots, n$ suppose that $\omega^i = (\omega^i_1, \cdots, \omega^i_m) \in \Omega X_1 \times \cdots \times \Omega X_m$ and $\omega^i_j\colon [0, s^i_j] \to X_j$. Then we want to define

$$M_n(\omega^0, t_1, \omega^1, \cdots, t_n, \omega^n) = \psi(\omega_1(t), \cdots, \omega_m(t))$$

for $t = (t_1, \cdots, t_n) \in I^n$, the unit n-cube, and with

$$\omega_j(t) = \omega^0_j \cdot \{t_1\delta^1_j(t)\} \cdot \omega^1_j \cdot \cdots \cdot \{t_n\delta^n_j(t)\} \cdot \omega^n_j .$$

(As above $\{r\}$ denotes the constant loop of length r and $\cdot$ denotes the loop product.) Then for $t_k = 1$ we must have that the loops $\omega^0_j \cdot \{t_1\delta^1_j(t)\} \cdot \omega^1_j \cdot \cdots \cdot \omega^{k-1}_j \cdot \{\delta^k_j(t)\}$ have the same lengths for different j's. Setting $\lambda^k_j(t) = s^0_j + t_1\delta^1_j(t) + \cdots + t_{k-1}\delta^{k-1}_j(t) + s^{k-1}_j$ and $\lambda^k(t) = \max(\lambda^k_j(t))$, we must have $\delta^k_j(t) = \lambda^k(t) - \lambda^k_j(t)$. This gives an inductive definition for $\delta^k_j(t)$ and we note that $\delta^k_j(t)$ actually depends only on $t_1, \cdots, t_{k-1}$. This completes the definition of M_n.

REMARK. If $\omega^i_j = \Omega e$ for all $j \neq k$, then, setting $s = s^k_1 + \cdots + s^k_n$,

$$\begin{aligned} M_n(\omega^0, t_1, \cdots, t_n, \omega^n) &= \psi(\Omega e, \cdots, \Omega e, \omega^k_1 \cdots \omega^k_n, \Omega e, \cdots, \Omega e) \\ &= (\{s\}, \cdots, \{s\}, \omega^k_1 \cdots \omega^k_n, \{s\}, \cdots, \{s\}) . \end{aligned}$$

COROLLARY 1.7. *If X is an H-space, there are strongly homotopy multiplicative maps*

$$\psi^{(k)}\colon (\Omega X)^k \to \Omega X .$$

Proof. Let $m^k\colon X^k \to X$ be given by some fixed way of associating the product of k elements of X. (Unless otherwise specified we shall assume that $m^k(x_1, \cdots, x_k) = x_1(x_2(\cdots(x_{k-1}(x_k))\cdots))$. Let M^k_n denote the homotopies defined above for the map $\psi\colon (\Omega X)^k \to \Omega(X^k)$. Then $\psi^{(k)} = \Omega(m^k) \circ \psi$ and the homotopies for $\psi^{(k)}$ are given by $\tilde{M}^k_n = \Omega(m^k) \circ M^k_n$. When $k = 2$, this shows that ΩX is strongly homotopy commutative.

2. The bar construction. For the convenience of the reader and to fix notation we recall the principal parts of the bar construction (Eilenberg and MacLane [4]).

K will denote a commutative ring with unit and A will denote an associative DGA algebra over K with augmentation $\varepsilon: A \to K$. Then $\bar{A} = \operatorname{Ker} \varepsilon$ and $s\bar{A}$ denotes the *suspension* of $\bar{A}$, the graded module formed from $\bar{A}$ by raising degrees by one. $\bar{B}_n(A) = (s\bar{A})^n$, the n-fold tensor power of $s\bar{A}$, with the convention $(sA)^0 = K$. The (*normalized*) *bar construction* $\bar{B}(A)$ is the graded K-module with component $\bar{B}_n(A)$ in degree n, and with the obvious augmentation. Elements of $\bar{B}_n(A)$ are written as linear combinations of elements $[a_1 | \cdots | a_n] = [a_1] \otimes \cdots \otimes [a_n]$ where $[a_i]$ denotes the suspension of $a_i \in \bar{A}$. Then $\bar{B}(A)$ is graded by assigning the element $[a_1 | \cdots | a_n]$ *external degree* n and *internal degree* $m = \sum_{i=1}^{n} \deg(a_i)$, and bidegree (n, m). It will be convenient to abbreviate by $\bar{B}^n(A)$ the graded module with component $\bar{B}_k(A)$ in (external) degree k for $0 \leqq k \leqq n$, and the 0 module in all other degrees. As a differential K-module $\bar{B}(A)$ has a total differential $d_T = d_E + d_I$ where the *external and internal differentials*, d_E and d_I are given by the formulas

$$d_E([a_1 | \cdots | a_n]) = \sum_{i=1}^{n-1} (-1)^{\sigma(i)} [a_1 | \cdots | a_i a_{i+1} | \cdots | a_n]$$

$$d_I([a_1 | \cdots | a_n]) = \sum_{i=1}^{n} (-1)^{\sigma(i-1)} [a_1 | \cdots | da_i | \cdots | a_n]$$

in which $\sigma(i) = \deg([a_1 | \cdots | a_i])$ and $a_i a_{i+1}$ and da_i indicate the product and differential taken in A. $\bar{B}(A)$ is an associative co-algebra in a natural way with coproduct

$$\Delta([a_1 | \cdots | a_n]) = \sum_{i=0}^{n} [a_1 | \cdots | a_i] \otimes [a_{i+1} | \cdots | a_n]$$

where in the extreme terms $i = 0$ and $i = 1$, $[\ ] = 1 \in \bar{B}_0(A) = K$.

When K is a principal ideal domain, the homology of the bar construction $\bar{B}(A)$ will be denoted $\operatorname{Tor}^A(K, K)$, and extension of the usual use of this notation. (See Moore [8].)

If $\bar{B}(A)$ is filtered by external degree, we obtain the *Moore spectral sequence*, $\{E^r(A), d^r\}$, in which $E^1 \approx \bar{B}(H(A))$, and in which $E^2(A) \approx \operatorname{Tor}^{H(A)}(K, K)$ provided that $H_n(A)$ is K-projective for all n. If $H(A)$ is of finite type, then $E^r \Rightarrow E^\infty \approx E^0(\operatorname{Tor}^A(K, K))$, the graded module associated with the filtration induced on $\operatorname{Tor}^A(K, K)$.

3. Homotopy commutativity and the bar construction.

In this section are given algebraic analogues for some parts of the geometric situation discussed in § 1 and a product is introduced into the Moore spectral sequence turning it into a spectral sequence of Hopf algebras.

DEFINITION 3.1. Let A and A' be associative DGA algebras over K.

A homomorphism of DGA modules over K, $h: A \to A'$, is *strongly homotopy multiplicative* if there exists for each nonnegative integer n, a homomorphism of K-modules of degree n,

$$h_n: A \otimes \cdots (n+1) \cdots \otimes A \to A'$$

such that $h_0 = h$ and

$$\begin{aligned} dh_n(a_1 \otimes \cdots \otimes a_{n+1}) &= \sum_{i=1}^{n+1} (-1)^{\sigma(i-1)} h_n(a_1 \otimes \cdots \otimes da_i \otimes \cdots \otimes a_{n+1}) \\ &+ \sum_{i=1}^{n} (-1)^{\sigma(i)} [h_{n-1}(a_1 \otimes \cdots \otimes a_i a_{i+1} \otimes \cdots \otimes a_{n+1}) \\ &- h_i(a_1 \otimes \cdots \otimes a_{i+1}) h_{n-i+1}(a_{i+2} \otimes \cdots \otimes a_{n+1})] \end{aligned}$$

where as before $\sigma(i) = \deg([a_1 | \cdots | a_i])$.

Clearly this condition implies that h is (chain) homotopy multiplicative A homomorphism of algebras, $h: A \to A'$, is automatically strongly homotopy multiplicative taking h_n to be the zero homomorphism for $n > 0$.

PROPOSITION 3.2. If $h: A \to A'$ is strongly homotopy multiplicative, then h induces a homomorphism of DGA coalgebras over K

$$\bar{B}(h): \bar{B}(A) \to \bar{B}(A') .$$

Furthermore $\bar{B}(h)$ induces a homomorphism of spectral sequences

$$E^r(h): E^r(A) \to E^r(A')$$

such that $E^1(h) = \bar{B}(h_*)$ where $h_*: H(A) \to H(A')$. In other words $E^1(h)$ is given by

$$E^1(h)([x_1 | \cdots | x_n]) = [h_*(x_1) | \cdots | h_*(x_n)] .$$

Proof. Let $S(n, k)$ denote the set of k-tuples of nonnegative integers whose sum is n. Then

$$\begin{aligned} &\bar{B}(h)([a_1 | \cdots | a_n]) \\ &\quad = \sum_{k=1}^{n} \sum_{S(n,k)} [h_{i_1-1}(a_1 \otimes \cdots \otimes a_{i_1}) | \cdots | h_{i_k-1}(a_{n-i_k+1} \otimes \cdots \otimes a_n)] . \end{aligned}$$

All the properties required of $\bar{B}(h)$ are easily checked by direct computation.

DEFINITION 3.3. An associative DGA algebra A is said to be *strongly homotopy commutative* if there is a strongly homotopy multiplicative homomorphism $h: A \otimes A \to A$ such that $h(a \otimes 1) = a = h(1 \otimes a)$.

PROPOSITION 3.4. If A is a strongly homotopy commutative, associative DGA algebra, then $\bar{B}(A)$ is a DGA Hopf algebra, and the terms of the Moore spectral sequence $E^r(A)$ are Hopf algebras provided the ground ring K is a field. Furthermore as $1 \in A_0$ is a unit for the map $h: A \otimes A \to A$, the Hopf algebras $E^r(A)$ are commutative, associative, and have a unit.

Proof. The shuffling map $\Sigma: \bar{B}(A) \otimes \bar{B}(A) \to \bar{B}(A \otimes A)$ is a homomorphism of coalgebras and therefore $\Phi = B(h)\Sigma$ provides a product for $\bar{B}(A)$ which is a homomorphism of coalgebras, and hence $\bar{B}(A)$ becomes a Hopf algebra. Let $E^0(A)$ denote the associated graded DGA Hopf algebra for the filtration of $\bar{B}(A)$ by external degree. Then $E^0(A) \approx \bar{B}(A)$ as a coalgebra, and as $h(a \otimes 1) = a = h(1 \otimes a)$, it follows that $E^0(\Phi)$ is just the well known shuffle product of Eilenberg and MacLane [4]. The remaining conclusions follow immediately since the shuffle product is commutative, associative, and has a unit.

COROLLARY 3.5. *If X is a pathwise connected and simply connected H-space of finite homological type, then there exists a spectral sequence of commutative and associative Hopf algebras with unit over Z_p, $\{E^r(\Omega X; Z_p), d^r\}$, such that*

$$E^2(\Omega X; Z_p) \approx \mathrm{Tor}^{H_*(\Omega X; Z_p)}(Z_p, Z_p) \qquad (\textit{as Hopf algebras})$$
$$E^\infty(\Omega X; Z_p) \approx E^0(H_*(X; Z_p)) \qquad (\textit{as Hopf algebras})$$

where $E^0(H_*(X; Z_p))$ denotes the associated graded Hopf algebra under a filtration of $H_*(X; Z_p)$.

Proof. Let $A = C_N(\Omega X; Z_p)$, the normalized singular chains of ΩX mod p. Then by 1.7 ΩX is strongly homotopy commutative. It will follow that A is strongly homotopy commutative in the algebraic sense if we set (for $a_i \in A \otimes A$)

$$h_n(a_1 \otimes \cdots \otimes a_{n+1}) = \tilde{M}^2_{n\#}(a_1 \otimes e \otimes \cdots \otimes e \otimes a_{n+1})$$

where $\#$ indicates the induced chain homomorphism and e denotes the singular 1-chain of I given by the identity map on I.[1] From Moore [8, Theorem 7.1] it follows that $H_*(X; Z_p) \approx \mathrm{Tor}^A(Z_p, Z_p)$ as Z_p-coalgebras, and the rest is immediate from 3.4 by setting $E^r(\Omega X; Z_p) = E^r(A)$.

REMARK. The results of this paper up to this point could be generalized as follows: an H-space Y is *n-homotopy commutative* if and only if there exists for $k = 0, 1, \cdots, n$ maps

$$M^2_k: (Y \times Y) \times (I \times Y \times Y)^k \to Y$$

[1] The only nontriviality involved is to know that $h_n(a_1 \otimes \cdots \otimes a_{n+1}) = 0$ when each $a_i = c_i \otimes 1$ (or $1 \otimes c_i$) and this follows from the remark which precedes 1.7.

with the appropriate properties as given in § 1. Then strongly homotopy commutative would be the same as ∞-homotopy commutative. The definitions and proofs above could be modified to obtain a Hopf algebra structure in the terms $E^1(Y; Z_p), \cdots, E^{n+1}(Y; Z_p)$ of the Moore spectral sequence converging to $E^0(H_*(B_Y; Z_p))$ using as hypothesis only that Y is n-homotopy commutative.

4. The suspension theorem. Using the results above we prove a suspension theorem which is a slight improvement on Browder [1, Theorem 5.13]. The notation and terminology is that of [5]; in particular Q is the functor which assigns to an algebra its module of decomposable elements, and P is the functor which assigns to a coalgebra its submodule of primitive elements. $\Gamma(x)$ denotes the ring with divided powers of x as defined in [5, Chapter 5] and $\gamma_k(x) = x^k/k!$.

If $\sigma_*: H_*(\Omega X; Z_p) \to H_*(X; Z_p)$ denotes the suspension in homology mod p for the contractible fibre space of paths over X as defined in [7], then Ker σ_* contains the decomposable elements of $H_*(\Omega X; Z_p)$ and Im σ_* is contained in the submodule of primitive elements of $H_*(X; Z_p)$ and σ_* has degree 1. (See [7] for proofs.) Therefore σ_* induces in each degree $i \geqq 1$ a homomorphism

$$s_i: Q(H_i(\Omega X; Z_p)) \to P(H_{i+1}(X; Z_p))$$

from the indecomposable elements of degree i of $H_*(\Omega X; Z_p)$ to the primitive elements of degree $i+1$ of $H_*(X; Z_p)$. Our suspension theorem will be a statement about the s_i. The philosophy of the proof will be a bit more clear if the reader bears in mind that taking $A = C_N(\Omega X; Z_p)$, $H_*(X; Z_p) \approx \mathrm{Tor}^A(Z_p, Z_p) \equiv H(\bar{B}(A))$ and under this isomorphism the suspension is given by $x \to [x]$ on the chain level. More precisely if $\sigma: A \to \bar{B}(A)$ is given by $\sigma(x) = [x]$, then we have a commutative diagram

$$\begin{array}{ccc} H_*(\Omega X; Z_p) & \xrightarrow{\sigma_*} & H_*(X; Z_p) \\ \approx\downarrow & & \downarrow\approx \\ H(A) & \xrightarrow[\sigma_*]{} & H(\bar{B}(A)) . \end{array}$$

In fact this is just [8, Proposition 7.2] applied to the case at hand.

THEOREM 4.1. *Let X be a pathwise connected and simply connected H-space of finite homological type and let s_i denote the homology mod p suspension in degree i as defined above. Then*

(a) *if $p \neq 2$, s_i is a monomorphism unless $i = (2mp^k + 2)p^q - 2$ for $k > 0, q > 0$, and $Q(H_{2m}(\Omega X; Z_p)) \neq 0$, or unless $i = (2m+2)p^q - 2$ for $q > 0$ and $Q(H_{2m+1}(\Omega X; Z_p)) \neq 0$;*

(b) *if $p \neq 2$, s_i is an epimorphism unless $i = 2mp^k + 1$ for $k > 0$*

and $Q(H_{2m}(\Omega X; Z_p)) \neq 0$;

(c) *if* $p = 2$, s_i *is a monomorphism unless* $i = 2^q(2^k m + 2) - 2$, *for* $q > 0, k > 0$, *and* $Q(H_m(\Omega X; Z_2)) \neq 0$;

(d) *if* $p = 2$, s_i *is an epimorphism unless* $i = 2^k m + 1$ *for* $k > 0$ *and* $Q(H_m(\Omega X; Z_2)) \neq 0$.

Proof. Since X has finite homological type, ΩX does also, and ΩX is pathwise connected. $H_*(\Omega X; Z_p)$ is a commutative and associative Hopf algebra over Z_p and is connected. From the Borel decomposition theorem for Hopf algebras it follows that as an algebra, $H_*(\Omega X; Z_p)$ is a tensor product of exterior, polynomial and truncated polynomial algebras, each with a single generator. Since $\mathrm{Tor}^A(K, K)$ commutes with tensor products (as a functor of A) [2, Chapter XI, Theorem 3.1, p. 209], to compute $E^2 \approx \mathrm{Tor}^{H_*(\Omega X; Z_p)}(Z_p, Z_p)$ in the spectral sequence of 3.5, we need only compute on sample factors. The results are listed in the table below. The first entry is given by [8, Proposition 4.1], and the others admit similar and very simple proofs.

	A	$\mathrm{Tor}^A(Z_p, Z_p)$
	$E(x, 2m+1)$	$\Gamma(sx, 1, 2m+1)$
$p \neq 2$	$L(x, 2m)$	$E(sx, 1, 2m)$
	$L(x, 2m)/(x^{p^k})$	$E(sx, 1, 2m) \otimes \Gamma(tx, 2, 2mp^k)$
	$E(x, m) = L(x, m)/(x^2)$	$\Gamma(sx, 1, m)$
$p = 2$	$L(x, m)$	$E(sx, 1, m)$
	$L(x, m)/(x^{2^k})$ $(k > 1)$	$E(sx, 1, m) \otimes \Gamma(tx, 2, 2^k m)$

where $E(x, n)$, $L(x, n)$, and $\Gamma(x, n)$ indicate exterior, polynomial, and divided polynomial algebras on a single generator of degree n. In the right hand columm bidegrees are specified, and all entries are Hopf algebras with primitive generators. sx and tx indicate the suspension and transpotence of x. (For the definition of transpotence see [9].)

By induction on r we shall prove the following statements for E^r:

(1) The generators of odd degree have external degree 1 and are primitive.

(2) Generators of even degree are primitive if the external degree is 1 and nonprimitive for external degree greater than 2, and the nonprimitive generators have the form γ_{p^q} where x is a primitive generator.

(3) If $\gamma_k(x) \neq 0$, but $\gamma_{k+1}(x) = 0$, then $k < r$.

(4) As a differential Hopf algebra E^r is the tensor product of differential Hopf algebras with differential identically zero, and differential Hopf algebras of the form $E(x, 1, m) \otimes \Gamma(y, u, v)$ where $d^r(\gamma_{p^q}(y)) =$

x and where (consequently) $r = up^{q-1}$ for $u = 1$ or $u = 2$.

Clearly (4) implies (1), (2), and (3) for E^{r+1}, and we need only show that (1), (2), and (3) imply (4). Suppose that d^r does not vanish completely. Let z be a generator of minimal degree such that $d^r z \neq 0$. Since $r \geqq 2$ and d^r has bidegree $(-r, r-1)$ it must be that $z = \gamma_{p^q}(y)$ where y has external degree $u = 1$ or $u = 2$. Furthermore z has even total degree and $d^r z$ is therefore a primitive element of odd degree and therefore has external degree 1. We may therefore assume without loss of generality that $x = d^r z$ is a generator. It follows that $r = up^q - 1$. Then setting $A = E(x, 1, m) \otimes \Gamma(y, u, v)$ we have that $E^r \approx A \otimes B$ where B is a differential Hopf algebra satisfying (1), (2), and (3) and the same argument may be repeated for B, and so on until E^r is exhausted. A small modification is necessary in the case that $p = 2$ and $x = d^r z$ generates a divided polynomial algebra $\Gamma(x, 1, m)$, but we leave this to the reader. It should also be remarked that in the notation of [5], the algebra B is usually denoted by $E^r//A$.

Since the ground ring is the field Z_p, we have $E^\infty \approx H_*(X; Z_p)$ as a Z_p-module. Then the following diagram is commutative

$$\begin{array}{ccc} H_*(\Omega X; Z_p) & \xrightarrow{\sigma_*} & H_*(X; Z_p) \\ & \searrow^{\sigma_*^\infty} & \downarrow \approx \\ & & E^\infty(\Omega X; Z_p) = E^0(H_*(X; Z_p)) \end{array}$$

where σ_*^∞ is induced from the algebraic suspension σ defined above. Since the primitive elements of $H_*(X; Z_p)$ are mapped into (but not necessarily onto) primitive elements of $E^\infty(\Omega X; Z_p)$ by the vertical isomorphism, information about σ_* may be obtained from information about σ_*^∞ which can be calculated routinely using (4) above.

Evidently by purely algebraic considerations of the filtration on the bar construction it could be shown that the vertical isomorphism is a homomorphism of coalgebras. Since the primitive submodule of E^∞ gives us an upper bound on the primitive elements of $H_*(X; Z_p)$, we have not tried to carry this out.

Added in proof. Judging by P. J. Hilton's review [Math. Reviews, 29, 2809] a similar proof of a theorem similar to 4.1 appears in Samuel Gitler's paper, *Spaces fibered by H-spaces* (Spanish), Bol. Soc. Mat. Mexicana (2) 7 (1962), 71–84. On the evidence of the review, it appears that Gitler's proof contains the algebraic version of Moore's error.

Bibliography

1. W. Browder, *On differential Hopf algebras*, Trans. Amer. Math. Soc. **107** (1963), 153-176.
2. H. Cartan and S. Eilenberg, *Homological Algebra*, Princeton University Press, 1956.
3. A. Dold and R. Lashof, *Principal quasifibrations and fibre homotopy equivalence of bundles*, Illinois J. Math. **3** (1959), 285-305.
4. S. Eilenberg and S. MacLane, *On the groups $H(\pi, n)$, I*, Ann. of Math. **58** (1953), 55-106.
5. J. Milnor and J. C. Moore, *On the structure of Hopf Algebras*, Princeton, 1959 (mimeograph).
6. J. C. Moore, *Seminar on algebraic homotopy theory*, Princeton 1955/56 (mimeograph).
7. ———, *La Suspension*, Seminaire H. Cartan-J. C. Moore, **12** (1959/60), expose 6.
8. ———, *Algebre homologique et homologie des espaces classifiants*, Seminaire H. Cartan-J. C. Moore, **12** (1959/60), expose 7.
9. Seminaire H. Cartan, *Algebres d'Eilenberg-MacLane et homotopie*, **7** (1954/55).
10. J. D. Stasheff, *Homotopy associativity of H-spaces, I, II*, Trans. Amer. Math. Soc. **108** (1963), 275-292, 293-312.
11. M. Sugawara, *On the homotopy-commutativity of groups and loop spaces*. Mem. Coll. Sci. Univ. Kyoto Ser. A, Math. **33** (1960/61), 251-259.

Brown University

PACIFIC JOURNAL OF MATHEMATICS
Vol. 15, No. 1, 1965

THE ASYMPTOTIC NATURE OF THE SOLUTIONS OF CERTAIN LINEAR SYSTEMS OF DIFFERENTIAL EQUATIONS

A. DEVINATZ

Suppose $y'(t) = [A + V(t) + R(t)]y(t)$ is a system of differential equations defined on $[0, \infty)$, where A is a constant matrix, $V(t) \to 0$ as $t \to \infty$ and the norms of the matrices $V'(t)$ and $R(t)$ are summable. If the roots of the characteristic polynomial of A are simple, then under suitable conditions on the real parts of the roots of the characteristic polynomials of $A + V(t)$ a theorem of N. Levinson gives an asymptotic estimate of the behavior of the solutions of the differential system as $t \to \infty$. In this paper Levinson's theorem is improved by removing the condition that the characteristic roots of A are simple. Under suitable conditions on $V(t)$ and $R(t)$ and the characteristic roots of $A + V(t)$, which reduce to Levinson's conditions when the characteristic roots of A are simple, asymptotic estimates are obtained for the solutions of the given system.

The proof given here, with essential modifications, will follow the proof given by Levinson [3] [2, p. 92]. One interest in the improved theorem is in its application to the problem of finding the deficiency index of an ordinary self-adjoint differential operator, which will appear in a subsequent paper. We shall establish the following.

THEOREM.[1] *Let A be a constant $n \times n$ matrix whose minimal polynomial is of degree n and is of the form*

$$\chi(\lambda) = \prod_{k=1}^{m} (\lambda - \lambda_k)^{n_k},\ \lambda_j \neq \lambda_k \ \textit{for}\ j \neq k,\ \sum_{k=1}^{m} n_k = n\ .$$

Let $q + 1 = \max n_k$, $V(t)$ an $n \times n$ matrix with $(q + 1)$-times continuously differentiable elements satisfying $t^{2q} | v_{ij}^{(r)}(t) |^{1/r} \in L^1$ for $1 \leqq r \leqq q + 1$ and $V(t) \to 0$ as $t \to \infty$. Let the roots of $\det (A + V(t) - \lambda I) = 0$ be $\{\lambda_k(t)\}_1^m$ and for $t \geqq \tau_0$ we suppose the minimal polynomial of $A + V(t)$ is

$$\chi(\lambda, t) = \prod_{k=1}^{m} (\lambda - \lambda_k(t))^{n_k}\ ,$$

where $\lambda_k(t) \to \lambda_k$ as $t \to \infty$. For a given k, let

Received February 3, 1964. Research supported by NSF Grant G 24834.

[1] If A is an $n \times n$ matrix with entries a_{ij} we shall write $| A | = \sum_{ij} | a_{ij} |$. If x is a vector with entries x_i we shall write $| x | = \sum_i | x_i |$.

$$d_{kj}(t) = Re(\lambda_k(t) - \lambda_j(t)) ,$$

and suppose that all $j, 1 \leqq j \leqq n$, *fall into one of two classes* I_1 *and* I_2, *where* $j \in I_1$, *if and only if* $t^{-q} \exp \int_0^t d_{kj} \to \infty$ *as* $t \to \infty$ *and*

$$(|t - \tau|^q + 1) \exp - \int_\tau^t d_{kj} < M < \infty \quad \textit{for} \quad t \geqq \tau \geqq 0 ,$$

$j \in I_2$ *if and only if* $\int_\tau^t d_{kj} < \log M$ *for* $t \geqq \tau \geqq 0$. *Let* $R(t)$ *be a matrix valued function with measurable elements such that* $t^{2q} |R(t)| \in L^1$. *Let* $\{q_{kj}; 1 \leqq j \leqq n_k\}$ *be a set of "principal vectors" for* λ_k; *i.e.,* $q_{kj} = (A - \lambda_k I)^{n_k - j} g_{kn_k}$, $(A - \lambda_k I)^{n_k - 1} g_{kn_k} \neq 0$ *and* $(A - \lambda_k I)^{n_k} g_{kn_k} = 0$. *Then, given the differential equation*

$$y'(t) = [A + V(t) + R(t)]y(t) \tag{1.1}$$

there exists a t_0 *and a fundamental system of solutions* $\{y_{kj}(t); 1 \leqq j \leqq n_k, 1 \leqq k \leqq m\}$ *such that*

$$\left[\frac{t^{j-1}}{(j-1)!} \exp \int_{t0}^t \lambda_k(\tau) d\tau\right]^{-1} y_{kj}(t) - q_{kj} \to 0, t \to \infty .$$

2. We begin the proof by first considering a differential system of the form

$$y'(t) = (A(t) + R(t))y(t) , \tag{2.1}$$

where $A(t)$ is a matrix with blocks $\{J_j(t)\}_1^m$ down the main diagonal and zeros elsewhere, $J_j(t)$ being an $n_j \times n_j$ matrix with the same number $\lambda_j(t)$ down the main diagonal, 1 down the superdiagonal and zeros elsewhere, and $R(t)$ has measurable entries with $t^{2q} |R(t)| \in L^1$, where $q + 1 = \max \{n_j, 1 \leqq j \leqq m\}$.

One fundamental matrix Ψ for the system

$$y'(t) = A(t)y(t) \tag{2.2}$$

has blocks $\{P_j\}_1^m$ down the main diagonal and zeros elsewhere, where P_j is an $n_j \times n_j$ matrix of the form

$$P_j(t) = \exp \int_{t_0}^t \lambda_j \begin{bmatrix} 1 & t & t^2/2! & \cdots & t^{n_j-1}/(n_j-1)! \\ 0 & 1 & t & \cdots & t^{n_j-2}/(n_j-2)! \\ \vdots & & \ddots & & \\ \vdots & & & \ddots & \\ 0 & \cdot & \cdot & 0 & 1 \end{bmatrix} . \tag{2.3}$$

This may be checked by a direct computation. Again, it may be easily checked that

$$P_j^{-1}(t) = \exp - \int_{t_0}^t \lambda_j \begin{bmatrix} 1 & -t & t^2/2! & -t^3/3! \cdots (-1)^{n_j-1}t^{n_j-1}/(n_j-1)! \\ 0 & 1 & -t & t^2/2! \cdots (-1)^{n_j-2}t^{n_j-2}/(n_j-2)! \\ \vdots & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot \quad 0 \quad 1 \end{bmatrix}$$

and

(2.4)

$$P_j(t)P_j^{-1}(\tau) = \exp \int_{\tau}^t \lambda_j \begin{bmatrix} 1 & (t-\tau) & (t-\tau)^2/2! \cdots (t-\tau)^{n_j-1}/(n_j-1)! \\ 0 & 1 & (t-\tau) \quad \cdots (t-\tau)^{n_j-2}/(n_j-2)! \\ \vdots & \cdot & \cdot \quad \cdot \\ 0 & \cdot & \cdot \quad \cdot \quad 0 \quad 1 \end{bmatrix}.$$

Let us fix k and let Ψ_1 be that matrix with zeros everywhere except for diagonal blocks $\{P_j; j \in I_1\}$, where each such P_j has the same position as in the matrix Ψ. Let Ψ_2 be the corresponding type matrix with diagonal blocks $\{P_j; j \in I_2\}$. Clearly $\Psi = \Psi_1 + \Psi_2$.

Let e_i be the vector with jth component equal to δ_{ij}, δ_{ij} being the Kronecker symbol. Now set $i = l + \sum_{j=1}^{k-1} n_j$, where $1 \leqq l \leqq n_k$, and consider the equation

$$\phi(t) = \Psi(t)e_i + \int_{t_0}^t \Psi_1(t)\Psi^{-1}(\tau)R(\tau)\phi(\tau)d\tau - \int_t^\infty \Psi_2(t)\Psi^{-1}(\tau)R(\tau)\phi(\tau)d\tau . \tag{2.5}$$

It may be checked by a straightforward computation that, at least formally, ϕ is a solution to (2.1). Hence, if it can be shown that a solution to (2.5) exists, where the integrands are in L^1, then this solution will also be a solution to (2.1).

We proceed by successive approximations. Choose $\phi^0 = 0$ and hence $\phi^1 = \Psi(t)e_i$. It follows that

$$|\phi^1 - \phi^0| \leqq \left[\exp \int_{t_0}^t Re\lambda_k\right] \sum_{j=0}^{l-1} t^j/j! \tag{2.6}$$

Now, the matrix $\Psi_1(t)\Psi^{-1}(\tau)$ has blocks along the main diagonal which are zero in those positions for which $j \in I_2$ and of the form (2.4) in those positions for which $j \in I_1$. Hence, using the hypothesis of the theorem of § 1, for $t_0 \leqq \tau \leqq t$ we have

(2.7)

$$|\Psi_1(t)\Psi^{-1}(\tau)R(\tau)| \leqq C[|t-\tau|^q + 1] \exp\left(-\int_\tau^t d_{kj}\right) \exp\left(\int_\tau^t Re\lambda_k\right) |R(\tau)|$$
$$\leqq CM|R(\tau)| \exp \int_\tau^t Re\lambda_k ,$$

where C is a suitable constant dependent only of q. In the same way, for $t \leqq \tau < \infty$,

$$(2.8) \qquad |\Psi_2(t)\Psi^{-1}(\tau)R(\tau)| \leqq CM[|t-\tau|^q + 1]\,|R(\tau)|\exp -\int_t^\tau Re\lambda_k .$$

Using the estimates (2.6), (2.7) and (2.8) we arrive at the estimate

$$(2.9) \qquad \begin{aligned} &|\phi^2 - \phi^1| \exp -\int_{t_0}^t Re\lambda_k \\ &\leqq CM\left\{\int_{t_0}^t |R(\tau)| \sum_{j=0}^{l-1} \tau^j/j!d\tau + \int_t^\infty |R(\tau)|[|t-\tau|^q + 1] \sum_{j=0}^{l-1} \tau^j/j!d\tau\right\}. \end{aligned}$$

Now using the fact that $\tau^{2q}|R(\tau)| \in L^1$ we can choose t_0 so large so that

$$(2.10) \qquad |\phi^2 - \phi^1| \exp -\int_{t_0}^t Re\lambda_k \leqq 1/2 \quad \text{for} \quad t \geqq t_0 .$$

Using (2.7), (2.8) and (2.10) and proceeding by induction we find that for $j \geqq 1$,

$$(2.11) \qquad \begin{aligned} &|\phi^{j+1} - \phi^j| \exp - \int_{t_0}^t Re\lambda_k \\ &\leqq (1/2)^{j-1} CM\left\{\int_{t_0}^t |R(\tau)|\,d\tau + \int_t^\infty [|t-\tau|^q + 1]\,|R(\tau)|\,d\tau\right\} \\ &\leqq (1/2)^j . \end{aligned}$$

This means that there exists a function ϕ so that on every compact subinterval of $[t_0, \infty)$, ϕ^j goes uniformly to ϕ, and indeed, using (2.6),

$$(2.12) \qquad |\phi - \phi^j| \leqq (1/2)^{j-1} \exp\int_{t_0}^t Re\lambda_k,\ |\phi| \leqq C[t^q + 1]\exp\int_{t_0}^t Re\lambda_k .$$

The estimates (2.12) taken together with the estimate (2.8) shows that the integrands in (2.5) are in L^1 and that indeed ϕ is a solution of that equation.

We claim that

$$(2.13) \qquad [\phi(t) - \Psi(t)e_i] \exp - \int_{t_0}^t \lambda_k \to 0 \quad \text{as} \quad t \to \infty .$$

To show this, it is enough to show that

$$(2.14) \qquad \exp\left(-\int_{t_0}^t Re\lambda_k\right)\int_{t_0}^t \Psi_1(t)\Psi^{-1}(\tau)R(\tau)\phi(\tau)d\tau \to 0 \quad \text{as } t \to \infty, \text{ and}$$

$$(2.15) \qquad \exp\left(-\int_{t_0}^t Re\lambda_k\right)\int_t^\infty \Psi_2(t)\Psi^{-1}(\tau)R(\tau)\phi(\tau)d\tau \to 0 \quad \text{as } t \to \infty .$$

Using (2.12) and (2.8) we see that the norm of (2.15) is less than or equal to

$$C^2M\int_t^\infty [|t-\tau|^q+1][\tau^q+1]|R(\tau)|\,d\tau,$$

which goes to zero as $t\to\infty$. To prove (2.14) we use the fact that $t^{-q}\exp\int_{t_0}^t d_{kj}\to\infty$. Choose t_1 so that $CM\int_{t_1}^\infty |R(\tau)||\phi(\tau)|\,d\tau<\varepsilon$. Then the norm of (2.14) is less than or equal to

$$\varepsilon+\exp\left(-\int_{t_0}^t Re\lambda_k\right)|\Psi_1(t)|\int_{t_0}^{t_1}|\Psi^{-1}(\tau)||R(\tau)||\phi(\tau)|\,d\tau\,.$$

Now,

$$\exp\left(-\int_{t_0}^t Re\lambda_k\right)|\Psi_1(t)|\leqq Ct^q\sum_{j\in I_1}\exp-\int_{t_0}^t d_{kj}\to 0\quad\text{as } t\to\infty\,.$$

Hence we see that (2.14) is valid.

The vector $\left[\exp-\int_{t_0}^t\lambda_k\right]\Psi(t)e_i$ has the entry $t^{l-j-1}/(l-j-1)!$ in the $i+j$ position, $0\leqq j\leqq l-1$, and zero elsewhere. Hence

$$\left\{\frac{t^{l-1}}{(l-1)!}\exp\int_{t_0}^t\lambda_k\right\}^{-1}\phi(t)-e_i\to 0\quad\text{as } t\to\infty\,. \tag{2.16}$$

Let us designate the solution we have obtained in the previous considerations by ϕ_i. Then the set of solutions $\{\phi_i\}_1^n$ is a fundamental system for (2.1). Indeed, it is clear that the determinant of the matrix Φ with the vectors ϕ_i as columns is nonzero for t sufficiently large.

3. In order to use the results of § 2 to prove the theorem of § 1 it will be necessary to establish the following.

LEMMA. *Suppose the matrix $A+V(t)$ satisfies the conditions of the theorem of* § 1. *Then for all sufficiently large t there exists a differentiable and invertible matrix $P(t)$ such that $t^{2q}|P^{-1}(t)P'(t)|\in L^1$, $P(t)[A+V(t)]P^{-1}(t)$ is a Jordan canonical form, $P(t)\to P$ and $P^{-1}(t)\to P^{-1}$ as $t\to\infty$, where PAP^{-1} is a corresponding Jordan canonical form for A, and the columns of P^{-1} are a given set of principal vectors for A.*

Proof. Let $\lambda_1,\lambda_2,\cdots,\lambda_m$ be the distinct eigenvalues of A. Since the coefficients of the characteristic polynomial of $A+V(t)$ are continuous functions of t in a neighborhood of ∞, using the hypothesis of the theorem, there exists a neighborhood of ∞ so that $A+V(t)$ has eigenvalues $\lambda_1(t),\cdots,\lambda_m(t)$ which are continuous for all t in that neighborhood. In particular, this means that $\lambda_k(t)\to\lambda_k$ as $t\to\infty$.

In fact, for t sufficiently large, each $\lambda_k(t)$ is $(q+1)$-times continuously differentiable. To see this, we consider the characteristic polynomial

$$(3.1)\qquad F(\lambda, t) = \sum_{j=0}^{n} f_j(t)\lambda^{n-j} = (-1)^n \prod_{j=1}^{m} (\lambda - \lambda_j(t))^{n_j} ,$$

where $f_j(t)$ is $(q+1)$-times continuously differentiable. If we set $G_k(\lambda, t) = \partial^{n_k - 1}F(\lambda, t)/\partial\lambda^{n_k - 1}$, then $G_k(\lambda_k(\tau), \tau) = 0$, but $\partial G_k(\lambda_k(\tau), \tau)/\partial\lambda \neq 0$. Hence, the implicit function theorem tells us that there exists a neighborhood about τ and a $(q+1)$-times continuously differentiable function μ_k, defined in this neighborhood, so that $\mu_k(\tau) = \lambda_k(\tau)$ and $G_k(\mu_k(t), t) = 0$. Moreover, if any other continuous function satisfies the last two conditions, then this other function coincides with μ_k in some neighborhood of τ. Hence $\lambda_k(t) = \mu_k(t)$ in some neighborhood of τ, which proves our assertion.

Let $\{q_{kj}; 1 \leqq j \leqq n_k\}$ be a given set of principal vectors for λ_k and let Q be the matrix whose columns are $\{q_{11}, \cdots, q_{1n_1}, q_{21}, \cdots, q_{2n_2}, \cdots, q_{m1}, \cdots, q_{mn_m}\}$, in the given order. Then, since the minimal and characteristic polynomials of A are of the same degree, $Q^{-1}AQ$ is in the Jordan canonical form (see e.g. [1], Ch. XVII). If V_k is the subspace generated by $\{g_{kj}; 1 \leqq j \leqq n_k\}$, then A is reduced by V_k. Hence, if we set

$$\pi_k(A) = \prod_{j \neq k} (A - \lambda_j)^{n_j} ,$$

then this matrix is reduced by V_k and the restriction of $\pi_k(A)$ to V_k has an inverse. Let us set $h_k = \pi_k^{-1}(A)q_{kn_k}$, where by $\pi_k^{-1}(A)$ we mean the inverse of the restriction of $\pi_k(A)$ to V_k.

Let us write the minimal polynomial, $\chi(\lambda, t)$, of $A + V(t)$ as

$$\chi(\lambda, t) = (\lambda - \lambda_k(t))^{n_k}\pi_k(\lambda, t) ,$$

where

$$\pi_k(\lambda, t) = \prod_{j \neq k} (\lambda - \lambda_j(t))^{n_j} .$$

Set $q_{kn_k}(t) = \pi_k(A + V(t), t)h_k$; then since $\pi_k(A + V(t), t) \to \pi_k(A)$ as $t \to \infty$, it follows that if we set

$$q_{kj}(t) = (A + V(t) - \lambda_k(t))^{n_k - j}q_{kn_k}(t)$$

the set $\{q_{kj}(t)\}_1^{n_k}$ forms a set of principal vectors for the eigenvalue $\lambda_k(t)$, provided t is sufficiently large. Indeed for t sufficiently large,

$$(A + V(t) - \lambda_k(t))^{n_k - 1}q_{kn_k}(t) \neq 0 ,$$

but

$$(A + V(t) - \lambda_k(t))^{n_k}q_{kn_k}(t) = \chi(A + V(t), t)h_k = 0 .$$

If $Q(t)$ is the matrix whose columns are the vectors

$$\{q_{11}(t), \cdots, q_{1n_1}(t), q_{21}(t), \cdots, q_{2n_2}(t), \cdots, q_{m1}(t), \cdots, q_{mn_m}(t)\},$$

in the order given, then $Q^{-1}(t)[A + V(t)]Q(t)$ is in the Jordan canonical form ([1]).

Notice that the elements of $Q(t)$ are polynomial functions in $\{\lambda_k(t)\}_1^m$ and the elements of $A + V(t)$, and hence the elements of $Q^{-1}(t)$ are rational functions in these variables, where the denominator of each rational function is $\det Q(t)$. Hence, if we set $P(t) = [\det Q(t)]Q^{-1}(t)$, then the elements of $P(t)$ are polynomials in the previously mentioned variables and $P(t)[A + V(t)]P^{-1}(t)$ is in the Jordan canonical form. Further, from the assumptions of the lemma, and the manner of construction of $Q(t)$, it is clear that $Q(t) \to Q$, where $Q^{-1}AQ$ is in the Jordan canonical form. Hence $P(t) \to P$, where PAP^{-1} is in the Jordan canonical form.

Since $P^{-1}(t) \to P^{-1}$, it is clear that $P^{-1}(t)$ is bounded in a neighborhood of infinity. Hence, if we can show that $t^{2q} | P'(t) | \in L^1$ we will have proved the lemma. The elements of $P'(t)$ are linear functions of $\{\lambda'_k(t)\}_1^m$ and $\{v'_{ij}(t)\}$ (the entries of $V'(t)$) with coefficients which are bounded in a neighborhood of infinity. Since, by hypothesis $t^{2q} | v'_{ij}(t) | \in L^1$, if we can show that $t^{2q} | \lambda'_k(t) | \in L^1$ we will be done.

Use (3.1) to obtain

$$\begin{aligned}\frac{\partial^{n_k} F(\lambda_k(t), t)}{\partial t^{n_k}} &= \sum_{j=1}^{n} f_j^{(n_k)}(t)\lambda_k^{n-j}(t) \\ &= (-1)^{n+n_k}\Big[n_k! \prod_{j \neq k} (\lambda_k(t) - \lambda_j(t))^{n_j}\Big][\lambda'_k(t)]^{n_k} .\end{aligned}$$

Since $\prod_{j \neq k} (\lambda_k(t) - \lambda_j(t))^{n_j}$ is uniformly bounded away from zero and $\lambda_k(t)$ is bounded, in a neighborhood of ∞, it follows that there exists a constant N such that

$$(3.2) \qquad | \lambda'_k(t) | \leqq N\Big[\sum_1^n | f_j^{(n_k)}(t) |\Big]^{1/n_k} \leqq N \sum_1^n | f_j^{(n_k)}(t) |^{1/n_k} .$$

Each function f_j is the sum of suitably signed products of elements of $A + V(t)$. A typical term in the sum representing f_j is say $a_1(t) \cdots a_j(t)$, where $a_i(t)$ is an entry of $A + V(t)$. The n_k derivative of this product is given by

$$\sum C_{i_1,\cdots,i_j} a_1^{(i_1)}(t) \cdots a_j^{(i_j)}(t) ,$$

where $C_{i_1,\cdots,i_j}$ are the constants which appear in the multinomial expansion of $(x_1 + \cdots + x_j)^{n_k}$ and the sum is taken over all j-tuples of nonnegative integers, $(i_1, \cdots, i_j)$, whose sum is n_k. Hence if

$$t^{2q} \mid a_1^{(i_1)} \cdots a_j^{(i_j)} \mid^{1/n_k} \in L^1 \tag{3.3}$$

it will follow that $t^{2q} \mid \lambda'_k \mid \in L^1$ and hence $t^{2q} \mid P'(t) \mid \in L$.

If $\sum_{r=1}^{j} i_r = n_k$, we may apply Holder's inequality to get,

$$\int_{t_0}^{\infty} t^{2q} \left| \prod_{r=1}^{j} a_r^{(i_r)} \right|^{1/n_k} \leqq \prod_{r=1}^{j} \left[\int_{t_0}^{\infty} t^{2q} \mid a_r^{(i_r)} \mid^{1/i_r} \right]^{i_r/n_k}, \tag{3.4}$$

where we make the convention that if $i_r = 0$, then

$$\| a_r \|_\infty = \sup_{t \geqq t_0} \mid a_r(t) \mid = \left[\int_{t_0}^{\infty} t^{2q} \mid a_r^{(i_r)} \mid^{1/i_r} \right]^{i_r/n_k}.$$

From the hypothesis of the lemma it follows from (3.4) that (3.3) is satisfied and hence lemma is proved.

4. Using the results of § 2 and § 3 it is now an easy matter to finish the proof of the theorem stated in § 1. Make the transformation $x(t) = P(t)y(t)$ in (1.1) and we get the equation

$$x' = [P(A + V)P^{-1} - P^{-1}P' + PRP^{-1}]x. \tag{4.1}$$

The matrix $P(A + V)P^{-1}$ is in the Jordan form of the matrix $A(t)$ of (2.1) and $t^{2q} \mid PRP^{-1} - P^{-1}P' \mid \in L^1$. Hence, we may apply the results of § 2 and for $i = l + \sum_{j=1}^{k-1} n_j$, $1 \leqq l \leqq n_k$, we find a solution x_i such that

$$\left[\frac{t^{l-1}}{(l-1)!} \exp \int_{t_0}^{t} \lambda_k \right]^{-1} x_i(t) - e_i \to 0 \quad \text{as } t \to \infty.$$

Hence, if $y_i(t) = P^{-1}(t)x_i$, we get

$$\left[\frac{t^{l-1}}{(l-1)!} \exp \int_{t_0}^{t} \lambda_k \right]^{-1} y_i(t) - P^{-1}e_i \to 0 \quad \text{as } t \to \infty,$$

where $P^{-1} = \lim_{t \to \infty} P^{-1}(t)$.

The vector $P^{-1}e_i$ is the ith column of P^{-1} which by Lemma 3 can be taken to be the given principal vector q_{kl}. Since the vectors $\{q_{kl}; 1 \leqq l \leqq n_k, 1 \leqq k \leqq m\}$ are linearly independent, the vectors $\{y_i(t)\}_1^m$ form a fundamental set of solutions of (1.1). This completes the proof of the theorem.

Note added in proof. The theorem of this paper can be generalized in the following way. Using the same notation as in the theorem let p be a real number satisfying the inequality $0 \leqq p \leqq q$. Suppose further that for each given k all integers j, $1 \leqq j \leqq n$, fall into two classes I_1 and I_2 where I_1 is the same as in the hypothesis of the theorem but now I_2 is the collection of j so that

$$(|t-\tau|^p+1)\exp\int_\tau^t d_{kj} < M < \infty \quad \text{for} \quad t \geqq \tau \geqq 0\,.$$

Then under the hypothesis that $t^{2q-p}\,|\,v_{ij}^{(r)}(t)\,|^{1/r}$, $1 \leqq r \leqq q+1$, and $t^{2q-p}\,|\,R(t)\,|$ are summable, the conclusion of the theorem holds. The proof of the generalized theorem follows the proof given in the text *mutatis mutandis*.

REFERENCES

1. E. T. Browne, *Introduction to the theory of determinants and matrices*, Chapel Hill, 1958.
2. E. A. Coddington and N. Levinson, *Theory of ordinary differential equations*, New York, 1955.
3. N. Levinson, *The asymptotic nature of the solutions of linear systems of differential equations*, Duke Math. J. **15** (1948), 111–126.

WASHINGTON UNIVERSITY
ST. LOUIS, MISSOURI

PACIFIC JOURNAL OF MATHEMATICS
Vol. 15, No. 1, 1965

APPROXIMATION BY CONVOLUTIONS

R. E. EDWARDS

This paper is concerned mainly with approximating functions on closed subsets P of a locally compact Abelian group G by absolute-convex combinations of convolutions $f * g$, with f and g extracted from bounded subsets of conjugate Lebesgue spaces $L^p(G)$ and $L^{p'}(G)$. It is shown that the Helson subsets of G can be characterised in terms of this approximation problem, and that the solubility of this problem for P is closely related to questions concerning certain multipliers of $L^p(G)$. The final theorem shows in particular that the P. J. Cohen factorisation theorem for $L^1(G)$ fails badly for $L^p(G)$ whenever G is infinite compact Abelian and $p > 1$.

1. The Approximation Problem.

(1.1) Throughout this note, G denotes a locally compact Abelian group and X its character group. For the most part we shall be concerned with the possibility of approximating functions on closed subsets P of G by absolute-convex combinations

$$\sum_{r=1}^{n} \alpha_r(f_r * g_r) , \tag{1}$$

of convolutions $f * g$, where f and g are selected freely from bounded subsets of conjugate Lebesgue spaces $L^p(G)$ and $L^{p'}(G)$ $(1/p + 1/p' = 1)$. In the sums (1), the number n of terms is variable, whilst the complex coefficients α_r are subject to the condition

$$\sum_{r=1}^{n} | \alpha_r | \leqq 1 . \tag{2}$$

Accordingly, if the f_r and g_r are respectively free to range over subsets A and B of $L^p(G)$ and $L^{p'}(G)$, the allowed sums (1) compose precisely the convex, balanced envelope of

$$A * B = \{ f * g : f \in A, g \in B \} .$$

We denote by $C_0(G)$ the Banach space of continuous, complex-valued functions on G which tend to zero at infinity, the norm being $\| u \| = \sup \{ | u(x) | : x \in G \}$. The space $C_0(P)$ is defined similarly, P replacing G throughout. If G (or P) is compact, the restriction that the functions tend to zero at infinity becomes void; we then write $C(G)$ (or $C(P)$) in place of $C_0(G)$ (or $C_0(P)$).

It is well-known that if $1 < p < \infty$ then $f * g \in C_0(G)$ whenever

Received February 5, 1964.

$f \in L^p(G)$ and $g \in L^{p'}(G)$, so that restriction from G to P results in a member of $C_0(P)$.

(1.2) Given an exponent p satisfying $1 < p < \infty$ and a closed subset P of G, we shall consider the following assertion:—

(A_P^p) To each member u of a second category subset of $C_0(P)$ corresponds a number $K = K(P,p,u) < \infty$ such that u is the uniform limit on P of absolute-convex combinations (1), the f_r and g_r being subject to the restrictions

$$\| f_r \|_p \leqq \sqrt{K}, \; \| g_r \|_{p'} \leqq \sqrt{K} .$$

It is evident that (A_P^p) and ($A_P^{p'}$) are equivalent assertions. Furthermore, only a little reflection is required to see that (A_P^1) is true for every P, so that the restriction $1 < p < \infty$ is reasonable. With this restriction on p, (A_P^p) signifies that each u belonging to the said second category set belongs to the closed, convex, balanced envelope in $C_0(P)$ of $A * B$, where A and B are respectively the closed balls in $L^p(G)$ and $L^{p'}(G)$ of radius $\sqrt{K}$ (which a priori may depend upon u).

(1.3) As well shall see, the truth or falsity of (A_P^p) is equivalent to an assertion about bounded measures supported by P which may conveniently be expressed by regarding such a measure as a multiplier (or centraliser) of ($L^p(G)$.

We denote by $M(G)$ the space of bounded, complex (Radon) measures on G; it may be regarded as the dual of $C_0(G)$. Furthermore, $M(P)$ may be thought of as the subset of $M(G)$ composed of measures $\mu \in M(G)$ whose supports are contained in P.

Each $\mu \in M(G)$ generates a multiplier T_μ of $L^p(G)$ defined by $T_\mu f = \mu * f$ for $f \in L^p(G)$. In general, by a multiplier of $L^p(G)$ is meant a continuous endomorphism of $L^p(G)$ which commutes with translations. Each multiplier T of $L^p(G)$ has a norm

$$\| T \| = \sup \{ \| Tf \|_p : \| f \|_p \leqq 1 \} .$$

Accordingly we may define $N_p(\mu)$ for $\mu \in M(G)$ as the norm of T_μ regarded as a multiplier of $L^p(G)$.

It is easily seen that

$$N_p(\mu) \leqq \| \mu \| , \tag{3}$$

equality holding if $p = 1$ (and hence also if $p = \infty$).

Although, as will be seen in (2.3), the norms $N_p(\mu)$ and $\| \mu \|$ are not generally equivalent on $M(G)$ when $1 < p < \infty$, yet equivalence may obtain on $M(P)$ for suitable closed subsets P of G. In fact, as the next theorem shows, the suitable sets P are just those for which the assertion (A_P^p) is true. When $p = 2$ one obtains in this way a new characterisation of the so-called Helson subsets of G; see (1.6) *infra.* A further link between (A_P^p) and properties of certain sets of multipliers of $L^p(G)$ is expressed in Theorem (2.1).

(1.4) THEOREM. *Let P be a closed subset of G, and let $1 < p < \infty$. Then (A_P^p) is true if and only if there exists a number $k = (P, p) < \infty$ such that*

$$\| \mu \| \leqq k.N_p(\mu) \,, \tag{4}$$

for each $\mu \in M(P)$.

Proof. Suppose first that (4) holds for $\mu \in M(P)$. This signifies that

$$\| \mu \| \leqq k.\ \mathrm{Sup} \left| \int_G (\mu * f)\, g dx \right| ,$$

the supremum being taken over those f and g lying respectively in the unit balls in $L^p(G)$ and $L^{p'}(G)$. Since

$$\int_G (\mu * f)\, g dx = \int_G (\check{f} * g)\, d\mu \,,$$

where $\check{f}(x) = f(-x)$, it follows that

$$\| \mu \| \leqq \mathrm{Sup} \left\{ \left| \int_G (f * g)\, d\mu \right| : \| f \|_p \leqq \sqrt{K},\ \| g \|_{p'} \leqq \sqrt{K} \right\} .$$

From this it follows that for each $u \in C_0(P)$ one has

$$\left| \int_G u d\mu \right| \leqq \mathrm{Sup} \left| _G (f * g)\, d\mu \right| , \tag{5}$$

where now f and g vary subject to the conditions

$$\| f \|_p \leqq \sqrt{K}.\ \sqrt{\| u \|}\,,\ \| g \|_{p'} \leqq \sqrt{K}.\sqrt{\| u \|} \,. \tag{6}$$

Now (5), combined with the Bipolar Theorem, shows that u belongs to the closed, convex, balanced envelope in $C_0(P)$ of the functions $f * g$ (or, more precisely, their restrictions to P), where f and g are subject to (6). Thus the assertion (A_P^p) is true for each $u \in C_0(P)$, with

$$K(P,p,u) \leqq k(P,p).\, \| u \| \,.$$

Conversely, suppose that (A_P^p) is satisfied. Let Σ denote the set of $u \in C_0(P)$ for which $K(P,p,u)$ exists finitely, so that Σ is a second category subset of $C_0(P)$. For a given $u \in \Sigma$, the set of admissible numbers $K(P,p,u)$ is easily seen to be closed. Denote by S the set of $u \in \Sigma$ for which the infimum of this set of admissible values of $K(P,p,u)$ is at most unity. Thus S consists precisely of those $u \in C_0(P)$ which are limits in $C_0(P)$ of sums (1), wherein

$$(7) \qquad \| f_r \|_p \leqq 1, \ \| g_r \|_{p'} \leqq 1 .$$

It is almost evident that S is closed, convex, and balanced in $C_0(P)$. Moreover, Σ is the union of the sets nS $(n = 1, 2, \cdots)$. Since Σ is second category in $C_0(P)$, it follows that S must be a neighbourhood of zero in $C_0(P)$. Consequently, $\Sigma = C_0(P)$ and, for some $r > 0$, each $u \in C_0(P)$ satisfying $\| u \| \leqq r$ is the limit in $C_0(P)$ of sums (1) with the f_r and g_r subject to (7). Then, however, each $u \in C_0(P)$ belongs to the closed, convex, balanced envelope in $C_0(P)$ of the set of convolutions $f * g$ with

$$\| f \|_p \leqq r^{-1/2} \sqrt{\| u \|}\,, \quad \| g \|_{p'} \leqq r^{-1/2} \sqrt{\| u \|}\,.$$

For $\mu \in M(P)$ it is therefore the case that

$$\begin{aligned} \| \mu \| &= \operatorname{Sup}\left\{ \left| \int_G u d\mu \right| : \| u \| \leqq 1 \right\} \\ &\leqq \operatorname{Sup}\left\{ \left| \int_G (f * g) d\mu \right| : \| f \|_p \leqq r^{-1/2}, \| g \|_{p'}, \leqq r^{-1/2} \right\}. \end{aligned}$$

Using again the relation

$$\int_G (f * g)\, d\mu = \int_G (\mu * \check{f})\, g dx ,$$

it appears that

$$\begin{aligned} \| \mu \| &\leqq r^{-1/2}. \operatorname{Sup}\left\{ \| \mu * f \|_p : \| f \|_p \leqq r^{-1/2} \right\} \\ &= r^{-1}. N_p(\mu) , \end{aligned}$$

which is (4), with $k = r^{-1}$. The proof is thus complete.

(1.5) REMARK. It has appeared in the course of the preceding proof that, if the approximation specified in (A_P^p) is possible for each member of a second category subset of $C_0(P)$, then it is indeed possible for each $u \in C_0(P)$, and this with a value of $K(P,p,u)$ not exceeding $K_0(P,p). \| u \|$.

(1.6) *The case $p = 2$: relation with Helson sets.* When $p = 2$ it is a simple consequence of the Parseval formula and Plancherel's theorem that

$$\begin{aligned} N_2(\mu) &= \| \hat{\mu} \| \\ &= \operatorname{Sup}\left\{ | \hat{\mu}(\xi) | : \xi \in X \right\}, \end{aligned}$$

where

$$\hat{\mu}(\xi) = \int_G \overline{\xi(x)}\, d\mu(x) ,$$

is the Fourier-Stieltjes transform of μ. Reference to Rudin [4], p.115, Theorem 5.6.3 shows then that as a Corollary to Theorem (1.4) one obtains the fact that (A_P^2) is true for a closed set $P \subset G$ if and only if P is a Helson subset of G. (Rudin assumes his Helson sets to be compact, but this restriction is unnecessary in the present connection.)

From the case $p = 2$ of Theorem (1.4) we may also derive a known property of Helson subsets of discrete groups G. (For historical reasons, Helson subsets of discrete groups are often termed Sidon sets; see [4], Section 5.7.)

(1.7) COROLLARY. *Suppose that G is discrete and that P is a Helson (or Sidon) subset of G. Then each bounded, complex-valued function on P is the restriction to P of the Fourier-Stieltjes transform of some measure on the (compact) character group X.* (Cf. [4], p. 121, *Theorem* 5.7.3(d).)

Proof. Let $B(P)$ be the superspace of $C_0(P)$ formed of all bounded, complex-valued functions on P. On $B(P)$ take the topology of pointwise convergence on P. Let T denote the linear mapping of $M(X)$ into $B(P)$ which assigns to $\lambda \in M(X)$ the function $T\lambda$ defined by

$$T\lambda(x) = \int_X \xi(x) d\lambda(\xi) .$$

It is evident that T is continuous for the weak topology $t = \sigma(M(X), C(X))$ on $M(X)$. For any $k > 0$, the set

$$S_k = \{\lambda \in M(X) : \| \lambda \| \leqq k\} ,$$

is compact for t, so that its image $T(S_k)$ is compact, and therefore closed, in $B(P)$. It will therefore suffice to show that, for some $k > 0$, $T(S_k)$ is dense in

$$V = \{ v \in B(P) : \| v \| \leqq 1 \} ;$$

and this will certainly be the case if $T(S_k)$ is shown to be dense in the closed unit ball $V_0 = V \cap C_0(P)$ in $C_0(P)$.

Suppose then that $u \in V_0$. Since P is a Helson set, (1.5) affirms the existence of a number $k = K_0(P,2)$ such that u is the limit, uniformly on P, and so a fortiori in the sense of the pointwise topology, of functions (1) with $\| f_r \|_2 \leqq \sqrt{K}$ and $\| g_r \|_2 \leqq \sqrt{K}$. By the Plancherel theory, these approximating functions form a sequence $(u_s)_{s=1}^{\infty}$, each term of which is expressible in the form

$$u_s(x) = \int_X \xi(x) F_s(\xi) d\xi = T\lambda_s (x) ,$$

where $\lambda_s \in M(X)$ is defined by $d\lambda_s(\xi) = F_s(\xi)d\xi$, and where

$$F_s = \sum_{r=1}^{n} \alpha_r^{(s)} \hat{f}_r^{(s)} . \hat{g}_r^{(s)} ,$$

so that

$$\|\lambda_s\| = \int_X |F_s(\xi)| d\xi \leqq \sum_{r=1}^{n_s} |\alpha_r^{(s)}| \cdot \|\hat{f}_r^{(s)}\|_2 \cdot \|\hat{g}_r^{(s)}\|_2$$
$$\leqq \sum_{r=1}^{n_s} |\alpha_r^{(s)}| \cdot \sqrt{k} \cdot \sqrt{k} \leqq k .$$

Thus $u_s \in T(S_k)$ for each s, which shows that each $u \in V_0$ belongs to the closure in $B(P)$ of $T(S_k)$, as we wished to show.

2. *Falsity of* (A_G^p). It is not altogether trivial to decide whether or not (A_G^p) is true. By expressing this assertion in terms of multipliers of $L^p(G)$, we shall show that (A_G^p) is false at any rate whenever $1 < p < \infty$ and G is infinite compact Abelian. The same conclusion is derivable without explicit mention of multipliers; see Remark (3.2) *infra*.

Let us denote by $m^p(G)$ the set of all multipliers of $L^p(G)$. As observed in (1.3), we may regard $M(G)$ as a subset of $m^p(G)$. The next theorem makes reference to the so-called weak and uniform operator topologies on $m^p(G)$, and for brevity we shall label these "W.O.T." and "U.O.T." respectively.

(2.1) THEOREM. *If P is a closed subset of G, the following four statements are equivalent:—*

(i) *$M(P)$ is closed in $m^p(G)$ for the* U.O.T.;

(ii) *$M(P)$ is sequentially closed in $m^p(G)$ for the* W.O.T.;

(ii′) *$M(P)$ contains the closure in $m^p(G)$, relative to the* W.O.T., *of any N_p-bounded subset of $M(P)$;*

(iii) *there exists a number $k = k(P,p) < \infty$ such that*

$$\|\mu\| \leqq k.N_p(\mu) ,$$

for $\mu \in M(P)$, i.e., by Theorem (1.4), (A_P^p) *is true.*

Proof. Since P is closed, $M(P)$ is in any case complete for the norm $\|\mu\|$. Since $m^p(G)$ is complete for the U.O.T., $M(P)$ is complete for N_p if and only if (i) holds. In any case, $N_p(\mu) \leqq \|\mu\|$. These remarks, combined with the Inversion Theorem for Banach spaces, show that (i) and (iii) are equivalent.

It is evident that (ii) implies (i). Also, since any sequence in $M(P)$ which is convergent for the W.O.T. is N_p-bounded (a direct application of the uniform boundedness principle), (ii′) implies (ii). It therefore remains only to show that (iii) implies (ii′).

Suppose then that (μ_i) is an N_p-bounded net in $M(P)$ such that $\lim_i T_{\mu_i} = T$ in the W.O.T.: we have to show that $T = T_\mu$ for some $\mu \in M(P)$. Now, since (iii) is true by hypothesis, $\mathrm{Sup}_i \| \mu_i \| < \infty$. Hence the net (μ_i) has a weak limiting point $\mu \in M(G)$. Since P is closed, μ necessarily belongs to $M(P)$. The definition of the weak topology on $M(G)$ ensures that, for each $f \in L^p(G)$ and each $g \in L^{p'}(G)$, the number $\int_G (\mu * f) g dx$ is a limiting point of the numerical net

$$\left(\int_G (\mu_i * f) g dx\right) = \left(\int_G (T_{\mu_i} f)\, g dx\right).$$

But this last net is convergent to $\int (Tf) g dx$. It follows that $Tf = \mu * f$ for each $f \in L^p(G)$, i.e., $T = T_\mu \in M(P)$, which is what we wished to prove.

(2.2) Remark. It is simple to verify that if $\mu \in M(P)$, then the multiplier T_μ has the property that $T_\mu f$ is, for each $f \in L^p(G)$, the limit of linear combinations of translates $f(x\text{-}a)$ of f with $a \in P$. Problem: Is it true that conversely any $T \in m^p(G)$, which is so approximable, is the limit in the W.O.T. of multipliers T_μ with μ ranging over some N_p-bounded subset of $M(P)$? The answer is affirmative if $P = G$ is compact, as will appear in the proof immediately below.

(2.3) Corollary. *Suppose that G is infinite compact Abelian. Then (A_G^p) is false for every p satisfying* $1 < p < \infty$.

Proof. Let us show first that any $T \in m_p(G)$ is the limit in the W.O.T. of an N_p-bounded net (μ_i) in $M(G)$. Take any base (U_i) of compact neighbourhoods of zero in G, and choose for each i a non-negative, continuous function h_i on G with support contained in U_i and such that $\int h_i dx = 1$. Then $\lim_i h_i * f = f$ in $L^p(G)$ for each $f \in L^p(G)$, so that

$$T_x = \lim_i T(h_i * f) = \lim_i Th_i * f = \lim_i k_i * f,$$

where $k_i = Th_i \in L^p(G)$ and

$$\begin{aligned}\| k_i * f \|_p = \| T(h_i * f)\|_p &\leqq \| T \| \cdot \| h_i * f \|_p \\ &\leqq \| T \| \cdot \| f \|_p .\end{aligned}$$

Let $\mu_i \in M(G)$ be defined by $d\mu_i(x) = k_i(x) dx$. Then $N_p(\mu_i) \leqq \| T \|$, and $\lim_i T_{\mu_i} f = \lim_i k_i * f = f$ in $L^p(G)$. Thus $\lim_i T_{\mu_i} = T$ in the

W.O.T. (even in the strong operator topology), and the net (μ_i) is N_p-bounded. This verifies our claim.

This being so, Theorem (2.1) shows that it is now sufficient to show that $M(G) \neq m^p(G)$, when G and P satisfy the stated conditions. To this end, we choose and fix any infinite Sidon subset S of X, and aim to show that corresponding to any bounded-complex-valued function b on X which vanishes on $X \cap S'$ there is a multiplier $T \in m^p(G)$ for which

$$(Tf)^\wedge(\xi) = b(\xi)\hat{f}(\xi) \qquad (\xi \in X)\,. \tag{8}$$

Indeed, if $1 < p \leqq 2$, this follows from the substance of p. 130 of [4]. If, on the other hand, $2 < p < \infty$ there is by that same token a multiplier T_1 of $L^{p'}(G)$ such that (8) is true with T_1 in place of T, and it then suffices to take for the desired T the adjoint of T_1.

If the multiplier T defined by (8) were of the form T_μ with $\mu \in M(G)$, then (8) would entail that

$$\hat{\mu}(\xi) = b(\xi) \qquad (\xi \in X)\,. \tag{9}$$

Since therefore $\hat{\mu}$ vanishes off S, the lemma immediately below would combine with (9) to show that

$$\Sigma_{\xi \in S} \mid b(\xi) \mid^2 < \infty\,. \tag{10}$$

However, S being infinite, we are at liberty to suppose that (10) is false, in which case T is not of the form T_μ. Thus $M(G)$ is a proper subset of $m^p(G)$, and the proof is complete.

(2.4) *Let G be a compact Abelian group and S a Sidon subset of X. If $\mu \in M(G)$ is such that*

$$\hat{\mu}(\xi) = 0 \qquad (\xi \in X \cap S')\,, \tag{11}$$

then μ is absolutely continuous (relative to Haar measure on G) and its Radon-Nikodym derivative h belongs to $L^q(G)$ for every finite q. In particular,

$$\Sigma_{\xi \in S} \mid \hat{\mu}(\xi) \mid^2 < \infty\,.$$

Proof. It is known ([4], p. 128, Theorem 5.7.7) that

$$\| t \|_q \leqq B_q \| t \|_1\,, \tag{12}$$

for every $q < \infty$ and every trigonometric polynomial t on G for which $\hat{t}(\xi) = 0$ for $\xi \in X \cap S'$, the number B_q being independent of t. On the other hand one may select in many ways a net (t_i) of trigonometric polynomials on G such that $\lim_i t_i * \mu = \mu$ weakly in $M(G)$ and $C \equiv$

$\sup_i \| t_i \|_1 < \infty$. The inequality (12) applies to $t_i * \mu$ and gives

$$\| t_i * \mu \|_q \leqq B_q \| t_i * \mu \|_1 \leqq B_q C \| \mu \| .$$

Supposing that $q > 1$, it follows that the net $(t_i * \mu)$ has a weak limiting point h_q in $L_q(G)$ and, since $t_i * \mu \to \mu$ weakly in $M(G)$, μ can be none other than the measure defined by $d\mu(x) = h_q(x)dx$. Putting $h = h_2 \in L_2(G)$, it is seen that $h_q = h$ a.e. for each $q > 1$, so that $h \in L^q(G)$ for every finite q. This h is, modulo negligible functions, the Radon-Nikodym derivative of μ, and the lemma is established.

3. Impossibility of factorisation in $L^p(G)$, $p > 1$. It was shown by P.J. Cohen [1] that each $h \in L^1(G)$ can be factorised as $f * g$ with f and g in $L^1(G)$. Now, if $p > 1$, $L^p(G)$ is an algebra under convolution if G is compact (and, if Abelian as we assume throughout, in no other cases). The next theorem, still concerned with approximation by sums of the type (1), though now with different restrictions on the f_r and g_r, shows that Cohen's result is very far from being extendible to $L^p(G)$ with $p > 1$.

(3.1) THEOREM. *Let G be infinite compact Abelian, and let $1 < p \leqq \infty$. Let Σ denote the set of functions h in $L^p(G)$ with the following property:— There exists a number $R = R(p, h) < \infty$ such that h is the weak limit in $M(G)$ of finite sums*

$$\sum_{r=1}^{n} f_r * g_r , \tag{13}$$

subject to the condition

$$\sum_{r=1}^{n} \| f_r \|_p \cdot \| g_r \|_p \leqq R . \tag{14}$$

Then Σ is a first category subset of $L^p(G)$.

Note. In the statement of Theorem (3.1) we are regarding $L^p(G)$ as a subset of $M(G)$, identifying a function $f \in L^p(G)$ with the measure μ defined by $d\mu(x) = f(x)dx$.

Proof. Take again an infinite Sidon subset S of X. Since $p > 1$ there exists ([4], p. 130) a number $c = c(p, S)$ such that

$$\| \hat{f} \|_{2,S} \equiv [\Sigma_{\xi \in S} | \hat{f}(\xi) |^2]^{1/2} \leqq c. \| f \|_p ,$$

for each $f \in L^p(G)$. If k is a sum of the type (13), then $\hat{k} = \sum_{r=1}^{n} \hat{f}_r \cdot \hat{g}_r$ and so, by the Cauchy-Schwarz inequality,

$$\Sigma_{\xi\in S}\,|\,\hat{k}(\xi)\,| \leqq \sum_{r=1}^{n} \|\,\hat{f}_r\,\|_{2,S}\cdot\|\,\hat{g}_r\,\|_{2,S}\ ,$$
$$\leqq c^2 \sum_{r=1}^{n} \|\,f_r\,\|_p\cdot\|\,g_r\,\|_p\ ,$$
$$\leqq c^2 R\ ,$$

the last step by virtue of (14). Consequently, the inequality

(15) $$\Sigma_{\xi\in S}\,|\,\hat{h}\,(\xi)\,| < \infty\ ,$$

is satisfied by each $h \in \Sigma$.

If $\sum$ were second category in $L^p(G)$, an argument similar to that used in the proof of Theorem (1.4) would show that

(16) $$\Sigma_{\xi\in S}\,|\,\hat{h}(\xi)\,| \leqq c'\,\|\,h\,\|_p\ ,$$

for each $h \in L^p(G)$, c' being independent of h. This in turn would entail the existence of a measure $\mu \in M(G)$ (actually a function in $L^{p'}(G)$ if $p < \infty$) such that

$$\hat{\mu}(\xi) = \begin{cases} 1 & \text{if } \xi \in S\ , \\ 0 & \text{if } \xi \in X \cap S'\ . \end{cases}$$

But this would contradict Lemma (2.4). Thus Σ must be a first category subset of $L^p(G)$, as asserted.

(3.2) REMARK. The preceding proof can be modified slightly to show that $\Sigma \cap C(G)$ is a first category subset of $C(G)$, thus providing an alternative proof of Corollary (2.3).

(3.3) REMARK. The final phase of the preceding proof, leading from (16) to the contradiction, may be completed without reference to Lemma (2.4), and is in fact quite independent of the notion of Sidon sets and their properties. This is shown by the following lemma.

(3.4) LEMMA. *Let G be compact Abelian. If S is a subset of X such that*

(17) $$\Sigma_{\xi\in S}\,|\,\hat{u}(\xi)\,| < \infty\ ,$$

holds for each u in a second category subset of $C(G)$, then S is necessarily finite.

Proof. The hypothesis entails (cf. the proof of Theorem (1.4)) the existence of a number c'' such

$$\Sigma_{\xi\in S}\,|\,\hat{u}(\xi)\,| \leqq c''\,\|\,u\,\|\ ,$$

for each $u \in C(G)$. This and the Riesz theorem combine to show that to each bounded, complex-valued function b on S corresponds a measure $\mu \in M(G)$ for which

$$\hat{\mu}(\xi) = b(\xi) \text{ for } \xi \in S, \ \hat{\mu}(\xi) = 0 \text{ for } \xi \in X \cap S'.$$

This μ is uniquely determined by b and the mapping T which carries b into μ is an algebraic isomorphism of the algebra $B(S)$ of all bounded, complex-valued functions on S (with the sup norm and pointwise product) into the convolution algebra $M(G)$. By Theorem 1 of [2], this entails that $B(S)$ is of finite dimension, so that S must be finite.

(3.5) REMARK. Yet another way of deriving a contradiction from (16), or from the apparently weaker variant (17), is to invoke a known theorem which says that if S is a Sidon subset of the character group of a compact Abelian group G, then for any given $v \in l^2(S)$ there exists $u \in C(G)$ such that $\hat{u}(\xi) = v(\xi)$ for $\xi \in S$. For the circle group this is established by Rudin ([5], 5.1 and 5.3), though the result for Hadamard sets S of integers is much older; and for general G it follows from Theorem 5.7.7 of [4] together with a result due to Hewitt and Zuckerman ([3], Theorem 8.6) which applies even to non-Abelian compact G.

REFERENCES

1. Paul J. Cohen, *Factorization in group algebras*, Duke Math. J. **26** (1959), 199-205.
2. R. E. Edwards, *On functions which are Fourier transforms*, Proc. Amer. Math. Soc. **5** (1954), 71-78.
3. Edwin Hewitt and H. S. Zuckerman, *Some theorems on lacunary Fourier series, with extensions to compact groups*, Trans. Amer. Math. Soc. **93** (1959), 1-19.
4. Walter Rudin, *Fourier analysis on groups*, Interscience Publishers (1962).
5. ——, *Trigonometric series with gaps*, J. Math. and Mech. **9** (1960), 203-228.

INSTITUTE OF ADVANCED STUDIES
AUSTRALIAN NATIONAL UNIVERSITY

PACIFIC JOURNAL OF MATHEMATICS
Vol. 15, No. 1, 1965

DECOMPOSITION THEOREMS FOR FREDHOLM OPERATORS

T. W. GAMELIN

This paper is devoted to proving and discussing several consequences of the following decomposition theorem:

Let A and B be closed densely-defined linear operators from the Banach space X to the Banach space Y such that $D(B) \supseteqq D(A)$, $D(B^*) \supseteqq D(A^*)$, the range $R(A)$ of A is closed, and the dimension of the null-space $N(A)$ of A is finite. Then X and Y can be decomposed into direct sums $X = X_0 \oplus X_1$, $Y = Y_0 \oplus Y_1$, where X_1 and Y_1 are finite dimensional, $X_1 \subseteqq D(A)$, $X_0 \cap D(A)$ is dense in X, and (X_0, Y_0) and (X_1, Y_1) are invariant pairs of subspaces for both A and B. Let A_i and B_i be the restrictions of A and B respectively to X_i. For all integers k, $(B_0A_0^{-1})^k(0) \subseteqq R(A_0)$, and

$$\dim (B_0A_0^{-1})^k(0) = k \dim (B_0A_0^{-1})(0) = k \dim N(A_0) .$$

Also, the action of A_1 and B_1 from X_1 to Y_1 can be given a certain canonical description.

The object of this paper is to study the operator equation $Ax - \lambda Bx = y$, where A and B are (unbounded) linear operators from a Banach space X to a Banach space Y. In § 1, an integer $\mu(A:B)$ is defined, which expresses a certain interrelationship between the null space of A and the null space of B. In § 1 and 2, decomposition theorems are proved which refine theorem 4 of [2]. The theorems allow us to split off certain finite dimensional invariant pairs of subspaces of X and Y so that A and B are well-behaved with respect to $\mu(A:B)$ on the remainder.

In § 4, the stability of these decompositions under perturbation of A by λB is investigated. In § 5, relations between the dimensions of certain subspaces of X and Y are given, and a formula for the Fredholm index of $A - \lambda B$ is obtained. These extend results of Kaniel and Schechter [1], who consider the case $X = Y$ and B the indentity operator.

It should be noted that the results of Kaniel and Schechter referred to here follow from theorems 3 and 4 of [2]. The results of this paper properly refine Kato's results only when the null space of B is not $\{0\}$.

1. We will be considering linear operators T defined on a dense linear subset $D(A)$ of a Banach space X, and with values in a Banach space Y. $N(T)$ and $R(T)$ will denote the null space and range of T respectively, while $\alpha(T)$ is the dimension of $N(T)$, and $\beta(T)$ is the

Received February 10, 1964.

codimension of $\overline{R(T)}$ in Y. T is a Fredholm operator if T is closed, $R(T)$ is closed, and both $\alpha(T)$ and $\beta(T)$ are finite. The index of a Fredholm operator is the integer.

$$\kappa(T) = \alpha(T) - \beta(T) .$$

Let P be a subspace of X, Q a subspace of Y. (P, Q) is an *invariant pair of subspaces* for T if $T(P \cap D(T)) \subseteqq Q$.

Standing assumptions: In the remainder of the paper, A and B are closed linear operators from X to Y, $D(A)$ is dense in X, $D(B) \supseteqq D(A)$, and $D(B^*) \supseteqq D(A^*)$; A is semi-Fredholm, in the sense that $R(A)$ is closed and $\alpha(A) < \infty$.

The assumption $D(B^*) \supseteqq D(A^*)$ seems necessary for the proof of the decomposition theorems. It is often met when A and B are differential operators on some domain in Euclidean space, and the order of B is less than the order of A. It is always met when B is bounded.

The linear manifolds $N_k = N_k(A:B)$ and $M_k = M_k(A:B)$ are defined by induction as follows:

$$\begin{aligned} N_1 &= N(A) \\ N_k &= A^{-1}(BN_{k-1}) , \qquad k > 1 \\ M_k &= BN_k . \end{aligned}$$

N_k and M_k are increasing sequences of linear manifolds in X and Y respectively.

The smallest integer n such that N_n is not a subset of $B^{-1}R(A)$ will be denoted by $\nu(A:B)$. If N_n is a subset of $B^{-1}R(A)$ for all n, then we define $\nu(A:B) = \infty$. (cf. [2])

The dimension of N_k will be denoted by $\pi_k = \pi_k(A:B)$, and the dimension of M_k by $\rho_k = \rho_k(A:B)$. Then $\pi_1 = \alpha(A)$, and, in general, $\pi_k \leqq k\alpha(A)$. $\mu(A:B)$ will denote the first integer n such that $\pi_n < n\alpha(A)$. It $\pi_n = n\alpha(A)$ for all intergers n, then we define $\mu(A:B) = \infty$.

In general, $\mu(A:B) \geqq \nu(A:B) + 1$. This inequality is trivial if $\nu = \infty$. If $\nu < \infty$, then $M_{\nu-1} \subseteqq R(A)$, while $M_\nu \nsubseteq R(A)$. Consequently, $\pi_{\nu+1} < \pi_\nu + \alpha(A) \leqq (\nu + 1)\alpha(A)$, and so $\mu(A:B) \leqq \nu + 1$.

We define $\sigma_k(A:B) = \pi_k - \pi_{k-1}$. Then σ_k is the dimension of the quotient space N_k/N_{k-1}. $\{\sigma_k\}$ is a decreasing sequence of nonnegative integers, and so the limit

$$\sigma(A:B) = \lim_{k\to\infty} \sigma_k(A:B) \qquad \text{exists.}$$

If $\mu(A:B) = \infty$, then $\sigma(A:B) = \alpha(A)$.

2. THEOREM 1. *Assume, in addition to the standing assumptions on A and B, that $\nu(A:B) = \infty$. Then X and Y can be decomposed*

into direct sums

$$X = X_0 \oplus X_1$$
$$Y = Y_0 \oplus Y_1 ,$$

where X_1 *and* Y_1 *are finite dimensional,* $X_1 \subseteqq D(A)$, $X_0 \cap D(A)$ *is dense in* X_0, *and* (X_0, Y_0) *and* (X_1, Y_1) *are invariant pairs for both* A *and* B. *If* A_i *and* B_i *are the restrictions of* A *and* B *respectively to* X_i, *then* $\mu(A_0, B_0) = \infty$, *while* A_1 *and* B_1 *map* X_1 *onto* Y_1.

Furthermore, X_1 and Y_1 can be decomposed as direct sums

$$X_1 = P_1 \oplus \cdots \oplus P_p$$
$$Y_1 = Q_1 \oplus \cdots \oplus Q_p ,$$

where A_1 and B_1 map P_j onto Q_j. Bases $\{x_j^i : 1 < i \leqq \eta(j)\}$ and $\{y_j^i : 1 \leqq i \leqq \eta(j) - 1\}$ can be chosen for P_j and Q_j respectively so that

$$Ax_j^{i+1} = Bx_j^i = y_j^i , \qquad 1 \leqq i \leqq \eta(j) - 1$$
$$Ax_j^1 = 0 = Bx_j^{\eta(j)} .$$

Although the decomposition is not, in general, unique, the integers p and $\eta(j)$, $1 \leqq j \leqq m$, are uniquely determined by A and B. In fact,

$$p = \alpha(A) - \sigma(A : B) .$$

Proof. Let $n = \alpha(A)$, and suppose that $\{z_1^1, \cdots z_n^1\}$ is a basis for $N(A)$. Since $\nu(A : B) = \infty$, z_j^i can be chosen by induction so that $Az_j^i = Bz_j^{i-1}$. $\{z_j^i : 1 \leqq j \leqq n,\ 1 \leqq i \leqq m\}$ is a spanning set for N_m, while $\{Bz_j^i : 1 \leqq j \leqq n,\ 1 \leqq i \leqq m\}$ is a spanning set for M_m. Also, $\{z_i^m : 1 \leqq i \leqq n\}$ span N_m modulo N_{m-1}.

Recall that $\sigma_m = \sigma(m) = \dim(N_m/N_{m-1})$. By induction, the order of the z_j^i can be chosen so that $\{z_{n-\sigma(m)+1}^m, \cdots, z_n^m\}$ span N_m modulo N_{m-1}. Then

$$G_m = \{z_j^i : n - \sigma(i) + 1 \leqq j \leqq n,\ 1 \leqq i \leqq m\}$$

is a basis for N_m.

Let $\eta(j)$ be the greatest integer k such that $z_j^k \in G_k$. If $z_j^k \in G_k$ for all k, let $\eta(j) = \infty$. Then $1 \leqq \eta(1) \leqq \eta(2) \leqq \cdots \leqq \eta(n)$. Let p be the greatest integer k such that $\eta(k) < \infty$. By definition of σ, it is clear that

$$p = \alpha(A) - \sigma .$$

Suppose $1 \leqq j \leqq p$. $z_j^{\eta(j)+1}$ is linearly dependent on the set $G_{\eta(j)+1}$, and so we can write

$$z_j^{\eta(j)+1} = \sum \alpha_{ik} z_k^i ,$$

where the sum is taken over all pairs of integers (i, k), with the understanding that $z_k^i = 0$ if $i \leqq 0$ and $\alpha_{ik} = 0$ if $z_k^i \notin G_{\eta(j)+1}$. For $-1 \leqq q \leqq \eta(j)$ define

$$x_j^{\eta(j)-q} = z_j^{\eta(j)-q} - \sum \alpha_{ik} z_k^{i-q-1}.$$

For $0 \leqq q \leqq \eta(j)$,

$$\begin{aligned} Bx_j^{\eta(j)-q} &= Bz_j^{\eta(j)-q} - \sum \alpha_{ik} Bz_k^{i-q-1} \\ &= Az_j^{\eta(j)-q+1} - \sum \alpha_{ik} Az_k^{i-q} \\ &= Ax_j^{\eta(j)-q+1} \end{aligned}$$

In particular, $Bx_j^{\eta(j)} = 0$.

Since the sum for $x_j^{\eta(j)-q}$ involves $z_j^{\eta(j)-q}$ only in the first term, the $z_{j(j)-q}^{\eta}$ may be replaced by the $x_j^{\eta(j)-q}$, $0 \leqq q \leqq \eta(j)$, to obtain another basis for $N_{\eta(j)+1}$. Repeating this process for $1 \leqq j \leqq p$, and making other appropriate replacements, we arrive at vectors x_j^i such that.

(1) $$x_1^1, \cdots, x_1^n \text{ are a basis for } N(A)$$

(2) $$Bx_j^i = Ax_j^{i+1} , \qquad 1 \leqq i \leqq \eta(j)$$

(3) $$Bx_j^{\eta(j)} = 0 , \qquad 1 \leqq j \leqq p .$$

For convenience, it is assumed that

(4) $$x_j^i = 0 \qquad \text{if } i > \eta(j) .$$

If $1 \leqq j \leqq p$, let P_j be the subspace of X with basis $\{x_j^1, \cdots, x_j^{\eta(j)}\}$. Let Q_j be the subspace of Y with basis $\{y_j^1, \cdots, y_j^{\eta(j)-1}\}$, where $y_j^i = Bx_j^i = Ax_j^{i+1}$. Let $X_1 = P_1 \oplus \cdots \oplus P_p$ and $Y_1 = Q_1 \oplus \cdots \oplus Q_p$. Then X_1 and Y_1 satisfy all the conclusions of the theorem. To conclude the proof, it suffices to produce complementary subspaces to X_1 and Y_1 which also form an invariant pair.

We will construct functionals

$$\begin{aligned} &\{g_j^i : 1 \leqq i \leqq \eta(j) , \qquad 1 \leqq j \leqq p\} \text{ on } X \text{ and} \\ &\{f_j^i : 1 \leqq i \leqq \eta(j) - 1 , \ 1 \leqq j \leqq p\} \text{ on } Y \text{ such that} \end{aligned}$$

the f_j^i are in the domain of A^* and

(5) $$g_j^{i+1} = A^* f_j^i , \qquad 1 \leqq i \leqq \eta(j) - 1$$

(6) $$g_j^i = B^* f_j^i , \qquad 1 \leqq i \leqq \eta(j) - 1$$

(7) $$f_j^i(y_k^q) = \delta_{iq}\delta_{jk} , \qquad \begin{aligned} &1 \leqq j,\ k \leqq n \\ &1 \leqq q \leqq i \end{aligned}$$

$$(8) \qquad g_j^i(x_k^q) = \delta_{iq}\delta_{jk}\,, \qquad \begin{array}{l} 1 \leqq j,\ k \leqq n \\ 1 \leqq q \leqq i\,. \end{array}$$

Let $g_j^{\eta(j)}$ be any functional on X which satisfies (8). The other g_j^i will be chosen by induction.

Suppose that f_k^q and g_k^q are chosen, for $q > i \geqq 1$, to satisfy (5) through (8). By (8), g_k^{i+1} is orthogonal to $N(A)$, and so g_k^{i+1} is in the closure of $R(A^*)$. Since $R(A)$ is closed, $R(A^*)$ is closed, and there is an $f_k^i \in D(A^*)$ for which $A^*f_k^i = g_k^{i+1}$. Let $g_k^i = B^*f_k^i$. Then (5) and (6) hold by definition.

To verify (7), we have for $q \leqq i$,

$$\begin{aligned} f_j^i(y_k^q) &= f_j^i(Ax_k^{q+1}) \\ &= (A^*f_j^i)(x_k^{q+1}) \\ &= g_j^{i+1}(x_k^{q+1}) = \delta_{iq}\delta_{jk}\,. \end{aligned}$$

(8) is an immediate consequence of (7).

$$\begin{aligned} \text{Let } X_0 &= \cap\,\{N(g_j^i)\colon 1 \leqq i \leqq \eta(j)\,, \qquad 1 \leqq j \leqq p\} \\ Y_0 &= \cap\,\{N(f_j^i)\colon 1 \leqq i \leqq \eta(j) - 1\,, \qquad 1 \leqq j \leqq p\}\,. \end{aligned}$$

From (7) and (8), it is clear that $X_0 \cap X_1 = \{0\}$ and $Y_0 \cap Y_1 = \{0\}$. Since the codimension of X_0 in X is no greater than the number of functionals g_j^i defining it, and since this number is the dimension of X_1, we must have $X = X_0 \oplus X_1$. Similarly, $Y = Y_0 \oplus Y_1$.

Suppose $x \in D(A) \cap X_0$. Then $f_j^i(Ax) = (A^*f_j^i)(x) = g_j^{i+1}(x) = 0$, and so $Ax \in Y_0$. Similarly, $Bx \in Y_0$, and $(X_0,\ Y_0)$ is an invariant pair for both A and B.

Since $(X_0,\ Y_0)$ and $(X_1,\ Y_1)$ are invariant pairs, $N_k(A:B) \cap X_0 = N_k(A_0:B_0)$. For k sufficiently large, $X_1 \subseteqq N_k(A:B)$, and so

$$\begin{aligned} \dim\{N_{k+1}(A_0:B_0)/N_k(A_0:B_0)\} &= \dim\{N_{k+1}(A:B)/N_k(A:B)\} \\ &= \sigma \\ &= \alpha(A) - p \\ &= \alpha(A_0)\,. \end{aligned}$$

This can occur only if $\dim N_k(A_0:B_0) = k\alpha(A_0)$ for all integers k. Hence $\mu(A_0:B_0) = \infty$.

3. Let $(P,\ Q)$ be an invariant pair of finite dimensional subspaces for A and B. $(P,\ Q)$ is an *irreducible invariant pair of type* ν if there are bases $\{x_i\}_{i=1}^n$ for P and $\{y_i\}_{i=1}^n$ for Q such that $Bx_i = y_i$, $Ax_1 = 0$, and $Ax_i = y_{i-1}$, $2 \leqq i \leqq n$.

$(P,\ Q)$ is an *irreducible invariant pair of type* μ if there are bases $\{x_i\}_{i=1}^n$ for P and $\{y_i\}_{i=1}^{n-1}$ for Q such that

$$\begin{aligned} Ax_1 &= 0 = Bx_n \\ Ax_{i+1} &= y_i = Bx_i\,, \qquad 1 \leqq i \leqq n-1\,. \end{aligned}$$

(P, Q) is an *irreducible invariant pair of type* μ^* if there are bases $\{x_i\}_{i=1}^{n-1}$ for P and $\{y_i\}_{i=1}^{n}$ for Q such that

$$\begin{aligned} Bx_i &= y_i\,, \qquad 1 \leqq i \leqq n-1 \\ Ax_i &= y_{i+1}\,, \qquad 1 \leqq i \leqq n-1\,. \end{aligned}$$

(P, Q) is an *invariant pair of type* ν if $P = P_1 \oplus \cdots \oplus P_k$ and $Q = Q_1 \oplus \cdots \oplus Q_k$, where (P_j, Q_j) is an irreducible invariant pair of type ν, $1 \leqq j \leqq k$. *Invariant pairs of type* μ or *type* μ^* are defined similarly.

It is straightforward to verify that if (P, Q) is an (irreducible) invariant pair of type $\mu(A:B)$ (resp. $\mu^*(A:B)$), then (P, Q) is an (irreducible) invariant pair of type $\mu(A-\lambda B:B)$ (resp. $\mu^*(A-\lambda B:B)$), for all complex numbers λ. If (P, Q) is an invariant pair of type μ, then $\nu(A \mid P, B \mid P) = \infty$ and $\mu((A \mid P)^*, (B \mid P)^*) = \infty$. If (P, Q) is of type μ^*, then $\nu(A \mid P, B \mid P) = \infty$ and $\mu(A \mid P, B \mid P) = \infty$.

THEOREM 2. *Suppose A and B satisfy the standing hypothesis. Then there exist decompositions*

$$\begin{aligned} X &= X_0 \oplus X_1 \oplus X_2 \\ Y &= Y_0 \oplus Y_1 \oplus Y_2 \end{aligned}$$

Where (X_0, Y_0) is an invariant pair, (X_1, Y_1) is an invariant pair of type μ, and (X_2, Y_2) is an invariant pair of type ν. If A_0 and B_0 are the restrictions of A and B respectively to X_0, then $\nu(A_0, B_0) = \infty$ and $\mu(A_0, B_0) = \infty$.

Proof. Theorem 2 follows from Theorem 1 and Kato's Theorem 4 [1], after it is noted that the latter theorem, although stated only for bounded operators B, is valid under the less restrictive assumption that $D(B^*) \supseteqq D(A^*)$.

THEOREM 3. *In addition to the standing hypotheses, suppose that A is a Fredholm operator. Then there exist decompositions*

$$\begin{aligned} X &= X_0 \oplus X_1 \oplus X_2 \oplus X_3 \\ Y &= Y_0 \oplus Y_1 \oplus Y_2 \oplus Y_3\,, \end{aligned}$$

where each (X_i, Y_i) is an invariant pair, (X_1, Y_1) is of type μ, (X_2, Y_2) is of type ν, and (X_3, Y_3) is of type μ^*. If A_0 and B_0 are the restrictions of A and B to X_0, then $\nu(A_0 : B_0) = \infty$, $\mu(A_0 : B_0) = \infty$, $\mu(A_0^* : B_0^*) = \infty$, and *$\nu(A_0^* : B_0^*) = \infty$.*

If $X^* = X_0^* \oplus X_1^* \oplus X_2^* \oplus X_3^*$ and $Y^* = Y_0^* \oplus Y_1^* \oplus Y_2^* \oplus Y_3^*$ are the corresponding decompositions of the adjoint spaces, then (Y_1^*, X_1^*) is an invariant pair of type $\mu_*(A^* : B^*)$, (Y_2^*, X_2^*) is an invariant pair of type $\nu(A^* : B^*)$, and (Y_3^*, X_3^*) is an invariant pair of type $\mu(A^* : B^*)$.

Proof. In view of Theorem 2, we may assume that $\mu(A : B) = \infty$ and $\nu(A : B) = \infty$. Then $\nu(A^* : B^*) = \infty$, and we can proceed to decompose X^* and Y^*, as in the proof of Theorem 1. The only difficulty encountered is to produce vectors x_j^i to span X_3 which actually lie in $D(A)$. An induction argument similar to that used in Theorem 1 to produce the f_j^i and g_j^i can also be employed in this case.

4. Let $\Phi^+(A : B)$ be the set of complex numbers λ such that $A - \lambda B$ is a closed operator from $D(A)$ to Y, and such that $R(A - \lambda B)$ is closed and $\alpha(A - \lambda B) < \infty$. $\Phi^+(A : B)$ is an open subset of the complex plane which, by assumption, contains the point $\lambda = 0$.

For all $\lambda \in \Phi^+(A : B)$, Theorems 1 and 2 are applicable to the operators $A - \lambda B$ and B. Also, for $\lambda \in \Phi^+(A : B)$ we define

$$\begin{aligned}\sigma_k(\lambda) &= \sigma_k(A - \lambda B : B)\\ \pi_k(\lambda) &= \pi_k(A - \lambda B : B)\\ \rho_k(\lambda) &= \rho_k(A - \lambda B : B)\\ \sigma(\lambda) &= \sigma(A - \lambda B : B)\,.\end{aligned}$$

THEOREM 4. *Let A and B satisfy the standing hypotheses. There exists a decomposition*

$$\begin{aligned}X &= X_0 \oplus X_1\\ Y &= Y_0 \oplus Y_1\end{aligned}$$

such that (X_0, Y_0) is an invariant pair, and (X_1, Y_1) is an invariant pair of type $\mu(A - \lambda B : B)$ for all complex numbers λ. If A_0 and B_0 are the restrictions of A and B to X_0, then $\mu(A_0 - \lambda B_0 : B_0) = \infty$ for all $\lambda \in \Phi^+(A : B)$ satisfying $\nu(A - \lambda B : B) = \infty$.

Proof. The points $\lambda \in \Phi^+(A : B)$ for which $\nu(A - \lambda B : B) < \infty$ form a discrete subset of $\Phi^+(A : B)$, and so there is a $\lambda' \in \Phi^+$ such that $\nu(A - \lambda' B : B) = \infty$. Let $X = X_0 \oplus X_1$ be the decomposition of Theorem 1 with respect to $A - \lambda' B$ and B. Then (X_1, Y_1) is an invariant pair of type $\mu(A - \lambda B : B)$ for all complex numbers λ, as remarked earlier.

If $\lambda \in \Phi^+(A : B)$ and $\nu(A - \lambda B : B) = \infty$, then X_0 and Y_0 cannot be decomposed further as in Theorem 1, for such a decomposition would violate the fact that $\mu(A_0 - \lambda' B_0 : B) = \infty$. Hence $\nu(A - \lambda B : B) =$

∞ implies $\mu(A_0 - \lambda B_0 : B_0) = \infty$.

Let D be the subset of $\Phi^+(A:B)$ of complex numbers λ for which $\nu(A - \lambda B : B) < \infty$. D is a discrete subset of $\Phi^+(A:B)$ with no limit points in $\Phi^+(A:B)$(cf [1]).

THEOREM 5. *$\mu(A - \lambda B : B)$ is a constant, either finite or infinite, for $\lambda \in \Phi^+(A:B) - D$.*

Proof. In view of Theorem 4, it suffices to prove the theorem when A and B are operators in an invariant pair of type μ. For this, it suffices to look at an irreducible invariant pair of type μ. This case is easy to verify.

THEOREM 6. *$\sigma(\lambda)$ is constant on each component of $\Phi^+(A:B)$.*

Proof. It suffices to show that $\sigma(\lambda)$ is constant in a neighborhood of an arbitrary point $\lambda' \in \Phi^+(A:B)$. Let $X = X_0 \oplus X_1 \oplus X_2$ and $Y = Y_0 \oplus Y_1 \oplus Y_2$ be the decomposition of Theorem 2 with respect to $A - \lambda' B$ and B. Then $\nu(A_0 - \lambda B_0 : B_0) = \infty$ for λ near λ', and so $\sigma(\lambda) = \alpha(A_0 - \lambda B_0)$ for λ near λ'. By Theorem 3, [2], $\alpha(A_0 - \lambda B_0) = \alpha(A_0 - \lambda' B_0)$ for λ near λ'.

5. Let $X = X_0 \oplus X_1 \oplus X_2$ and $Y = Y_0 \oplus Y_1 \oplus Y_2$ be the decompositions of Theorem 2 with respect to A and B. Let $\pi_k = \pi_k^0 + \pi_k^1 + \pi_k^2$ and $\rho_k = \rho_k^0 + \rho_k^1 + \rho_k^2$ be the corresponding decompositions of π_k and ρ_k. Assume that r is chosen small that $0 < |\lambda| < r$ implies $\lambda \in \Phi^+(A:B)$ and $\nu(A - \lambda B : B) = \infty$. Then $\pi_k^0(\lambda) = k\sigma(\lambda)$ for $|\lambda| < r$. If k is sufficiently large,

$$\pi_k^1(\lambda) = \dim X_1 , \qquad |\lambda| < r$$

$$\pi_k^2(\lambda) = \begin{cases} \dim X_2, & \lambda = 0 \\ 0 , & 0 < |\lambda| < r . \end{cases}$$

Also, $\rho_k^0(\lambda) = k\sigma(\lambda)$ for $|\lambda| < r$. For k sufficiently large,

$$\rho_k^1(\lambda) = \dim Y_1$$

$$\rho_k^2(\lambda) = \begin{cases} \dim Y_2 , & \lambda = 0 \\ 0 , & 0 < |\lambda| < r . \end{cases}$$

We define, for any $\lambda \in \Phi^+(A:B)$,

$$\pi(\lambda) = \lim_{k \to \infty} [\pi_k(\lambda) - k\sigma(\lambda)] \tag{1}$$

$$\rho(\lambda) = \lim_{k = \infty} [\rho_k(\lambda) - k\sigma(\lambda)] \tag{2}$$

$\pi(\lambda)$ and $\rho(\lambda)$ correspond to $\tau(\lambda)$ defined in [1]. From the preceding, we deduce that

$$\pi(\lambda) = \begin{cases} \dim X_1 , & 0 < |\lambda| < r \\ \dim (X_1 \oplus X_2) , & \lambda = 0 \end{cases} \tag{3}$$

$$\rho(\lambda) = \begin{cases} \dim Y_1 , & 0 < |\lambda| < r \\ \dim (Y_1 \oplus Y_2) , & \lambda = 0 . \end{cases} \tag{4}$$

From these formulae, it follows that

$$\alpha(A - \lambda B) = \sigma(\lambda) + \pi(\lambda) - \rho(\lambda) , \qquad 0 < |\lambda| < r , \tag{5}$$

for both sides of this expression are equal to

$$\alpha(A_0 - \lambda B_0) + \dim X_1 - \dim Y_1 .$$

We will assume in the remainder of the discussion that A is a Fredholm operator. The set of complex numbers λ such that $A - \lambda B$ is a Fredholm operator will be denoted by $\Phi(A : B)$. $\Phi(A : B)$ is an open subset of the complex plane, and consists of the union of those components of $\Phi^+(A : B)$ for which $R(A - \lambda B)$ is of finite codimension in Y, i.e., for which $\alpha(A^* - \lambda B^*) < \infty$.

The quantities $\pi_k^*(\lambda) = \pi_k(A^* - \lambda B^* : B^*)$, $\rho_k^*(\lambda)$, $\sigma^*(\lambda)$, $\pi^*(\lambda)$ and $\rho^*(\lambda)$ are then well-defined for $\lambda \in \Phi(A : B)$. The formula for the adjoint operators corresponding to (5) is

$$\alpha(A^* - \lambda B^*) = \sigma^*(\lambda) + \pi^*(\lambda) - \rho^*(\lambda) , \qquad 0 < |\lambda| < r . \tag{6}$$

Since $\alpha(A^* - \lambda B^*) = \beta(A - \lambda B)$, we have

$$\begin{aligned} \kappa(A - \lambda B) = {} & (\sigma(\lambda) - \sigma^*(\lambda)) \\ & + (\pi(\lambda) - \pi^*(\lambda)) - (\rho(\lambda) - \rho^*(\lambda)) \qquad 0 < |\lambda| < r . \end{aligned} \tag{7}$$

In view of the decomposition of Theorem 3, the jump discontinuity of π^* at $\lambda = 0$ is equal to that of π at $\lambda = 0$, i.e., they are both equal to $\dim X_2 = \dim Y_2$. Hence (7) holds also for $\lambda = 0$, and we arrive at the following theorem.

THEOREM 7. *For all* $\lambda \in \Phi(A : B)$,

$$\kappa(A - \lambda B) = (\sigma(\lambda) - \sigma^*(\lambda)) + (\pi(\lambda) - \pi^*(\lambda)) - (\rho(\lambda) - \rho^*(\lambda)).$$

Analogous formulae can be written down if it is assumed, further, that B is a Fredholm operator. If $M(B) = \{0\}$ and $R(B)$ is dense in Y_1 then $\rho(\lambda) = \rho^*(\lambda) = \pi(\lambda) = \pi^*(\lambda) = 0$, and Theorem 7 reduces to

$$\kappa(A - \lambda B) = \sigma(\lambda) - \sigma^*(\lambda), \qquad \lambda \in \Phi(A : B) . \tag{8}$$

This latter formula is due to Kaniel and Schechter [1], when $X = Y$ and B is the identity operator.

BIBLIOGRAPHY

1. Kaniel and Schechter, *Spectral theory for Fredholm operators*, Comm. on Pure and Applied Math., vol. 16, no. 4 (1963), 423-448.
2. Kato, *Perturbation theory for nullity, deficiency and other quantities of linear operators*, J. d'Analyse Math. **VI** (1958), 261-322.

PACIFIC JOURNAL OF MATHEMATICS
Vol. 15, No. 1, 1965

ON THE INVARIANT MEAN ON TOPOLOGICAL SEMIGROUPS AND ON TOPOLOGICAL GROUPS

EDMOND GRANIRER

Let S be a topological semigroup and $C(S)$ be the space of bounded continous functions on S. The space of translation invariant, bounded, linear functionals on $C(S)$ and its connection with the structure of S, are investigated in this paper. For topological groups G, not necessarily locally compact, the space of bounded, linear, translation invariant functionals, on the space $UC(G)$ of bounded uniformly continuous functions, is also investigated and its connection with the structure of G pointed out. The obtained results are applied to the study of the radical of the convolution algebra $UC(G)^*$ (for locally compact groups, or for subgroups of locally convex linear topological spaces) and some results which seem to be unknown even when G is taken to be the real line are obtained.

The topological semigroup S is assumed to have a separately continuous multiplication, and $C(S)$ is given the usual *sup* norm. $C(S)^*$ will denote the conjugate Banach space of $C(S)$. If $a \in S$ and f is any function on S then f_a is defined by $f_a(s) = f(as)$ for $s \in S$. $\varphi \in C(S)^*$ is said to be left invariant if $\varphi(f_a) = \varphi(f)$ for each f in $C(S)$ and a in S. $J_c l(S)$ will denote the space of left invariant elements of $C(S)^*$. A topological semigroup is said to be left amenable as a discrete semigroup if there is a linear functional $\varphi \neq 0$ on $m(S)$ (the space of all real bounded functions on S with the usual sup. norm) which satisfies $\varphi(f_a) = \varphi(f)$ for each a in S and f in $m(S)$ and $\varphi(f) \geqq 0$ if $f \geqq 0$. An analogous definition holds for the right amenable case. A topological semigroup is said to be amenable as a discrete semigroup if it is right and left amenable as a discrete semigroup.

The following are results of I. S. Luthar [12]:

(1) If S is an abelian topological semigroup with a compact ideal then $\dim J_c l(S) = 1$

(2) If G is an abelian topological group having a certain property P (Any noncompact locally compact group or any nonzero subgroup of a linear convex topological vector space has this property see [12] p. 406) then $\dim J_c l(G) \geqq 2$.

We say that a subset S_0 of the semigroup S is a left-ideal group if S_0 is a group when endowed with the multiplication induced from S

Received December 2, 1963. Supported in part by National Science Foundation Grant 19869.

and ss_0 belongs to S_0 for any s in S and s_0 in S_0. If S is also a topological space then $S_0 \subset S$ is a compact left-ideal group if it is a left-ideal group and a compact subset of S.

The following theorem is proved in Ch. IV of this paper:

THEOREM IV-1. *Let S be a topological semigroup (with only separately continuous multiplication and no separation axioms) containing exactly $n(0 < n < \infty)$ compact left-ideal groups. Then* $\dim J_c l(S) = n$.

If S is abelian and contains a compact ideal then as known and directly shown, S contains a unique group and compact ideal (see the argument in [12] at the top of p. 404) and so $\dim J_c l(S) = 1$, which yields Luthar's first result.

When considering this Theorem IV-I one is tempted to conjecture that its converse if true i.e.

(A) If S is a topological semigroup and $\dim J_c l(S) = n$ $0 < n < \infty$, then S contains exactly n compact left-ideal groups[1].

This conjecture, even when allowing S to be a topological semigroup with jointly continuous multiplication and S to be a Hausdorff regular topological space, cannot be true as the following simple example shows:

E. Hewitt (see [22]) has constructed a regular Hausdorff space S such that the only real continuous functions on it are the constant functions. Define in this space S the following multiplication: $ab = a$ for any $a, b \in S$. If $F: S \times S \to S$ is defined by $F(a, b) = ab = a$ and $U \subset S$ is open then $F^{-1}(U) = \{(a, b); ab \in U\} = \{(a, b); a \in U\} = U \times S$ which is surely open in $S \times S$. Therefore multiplication in S is jointly continuous and S is a Hausdorff regular topological space. But $C(S)$ is one dimensional and so $C(S)^*$ is one dimensional. Moreover, if we define $\varphi(f) = f(a)$ for each f in $C(S)$ and some fixed $a \in S$ then $\varphi \neq 0$ is easily seen to be left invariant. Thus $\dim J_c l(S) = 1$. But S does

[1] This conjecture made by I. S. Luthar for the abelian case (see [**12**] p. 403) and believed to be true by this author for completely regular topological semigroups, is not true even for abelian topological groups. In fact let G be a pseudocompact noncompact abelian topological group and A a translation invariant nonnegative linear functional on $C(G)$ such that $\|A\| = 1$. By Theorem 4.1 of W. Comfort and K. Ross (see [**23**] G) is totally bounded and each f in $C(G)$ is uniformly continuous and therefore has a unique uniformly continuous extension $\bar{f}$ to the compact topological group $\bar{G}$ (the completion of G). Conversely any $\bar{f} \in C(\bar{G})$ is the uniformly continuous extension of a unique $f \in C(G)$. Define now the linear functional $\bar{A}$ on $C(\bar{G})$ by $\bar{A}\bar{f} = Af$. It is not hard to show now, after using heavily the Comfort-Ross theorem, that $\bar{A}$ is translation invariant (with respect to the elements of $\bar{G}$) nonnegative and $\|\bar{A}\| = 1$. Therefore $\bar{A}\bar{f} = \int \bar{f} dm$ where m is the unique normalized Haar measure on $\bar{G}$. This shows that $\dim J_c l(G) = 1$ while G is not compact. Many thanks are due to W. Comfort and K. Ross for kindly letting this author have a preprint of their paper.

not contain any proper left ideal since $Sb = S$ for any $b \in S$. And S is neither a group nor is it compact (For a compact hausdorff space S, $C(S)$ even separates points). Nevertheless, in certain cases, statement (A) holds true. The following theorem is proved in Ch. II of this paper:

THEOREM II-2. *Let S be a countable topological semigroup which is left amenable as a discrete semigroup and which is a T_1 regular topological space* (and therefore completely regular).

Then $\dim J_c l(S) = n$, $n < \infty$, if and only if then S contains exactly n *finite* left-ideal groups[2].

Consider now G to be a topological group and denote by $LUC(G) \subset C(G)$ the space of left uniformly continuous functions on the group G. Let $J_u l(G) \subset LUC(G)^*$ be defined as:

$$\{\varphi;\ \varphi(f_a) = \varphi(f) \text{ for each } f \text{ in } LUC(G) \text{ and } a \text{ in } G\}\,.$$

Also, recall that at least any abelian or solvable or locally finite group G, is left amenable as a discrete group. (see Day [4] for these and more examples). We can now state our next result:

THEOREM III-2. *Let G be a separable locally compact hausdorff topological group which is amenable as a discrete group. Then*

(1) *Either* $\dim J_u l(G) = 1$ *or* $\dim J_u l(G) = \infty$ *and* $\dim J_u l(G) = 1$ *if and only if G is compact.*

(2) *Either* $\dim J_c l(G) = 1$ *or* $\dim J_c l(G) = \infty$ *and* $\dim J_c l(G) = 1$ *if and only if G is compact.*

THEOREM III-3. *Let G be any separable* (*not necessarily closed*) *subgroup of locally convex linear topological space. Then*

(1) *Either* $\dim J_u l(G) = 1$ *or* $\dim J_u l(G) = \infty$ *and* $\dim J_u l(G) = 1$ *if and only if* $G = \{0\}$.

(2) *Either* $\dim J_c l(G) = 1$ *or* $\dim J_c l(G) = \infty$ *and* $\dim J_u l(G) = 1$ *if and only if* $G = \{0\}$.

From these theorems it is obvious that for both the considered groups $\dim J_u l(G) = \dim J_c l(G)$ invariably holds. An example of a countable abelian topological group in which $\dim J_u l(G) = 1$ while $\dim J_c l(G) = \infty$ is given in Ch. III. This example uses heavily the theorems on countable topological semigroups obtained in Ch. II.

Separable topological groups G which are amenable as discrete groups and have a certain property B (G has property B means that G admits

[2] One cannot hope for much more than this theorem. In fact an example of a locally compact abelian topological semigroup (with jointly continuous multiplication) for which statement A does not hold true for any n can be given.

a real left uniformly continuous *unbounded* function. Noncompact locally compact groups, nonzero subgroups of locally convex linear topological spaces and groups which admit a right invariant unbounded metric have this property.) are considered in Ch. III and for them it is proved that $\dim J_u l(G) = \infty$ and $\dim J_c l(G) = \infty$ (see Theorem (III. 1)). It should be remarked here, that our results neither imply, nor are implied by Luthar's results in [12]. They improve Luthar's results in the case where G is separable and either locally compact or a subgroup of a locally convex linear topological space (and also in certain other cases) but they do not deal at all with the non separable case.

We consider further in this paper the Banach space $LUC(G)^*$ (i.e. the conjugate of $LUC(G)$). As known and easily seen $LUC(G)^*$ becomes a Banach algebra under convolution as multiplication (while convolution in $C(G)^*$ cannot generally be defined, as known). If we denote by $R(G)$ the radical of the Banach algebra $LUC(G)^*$ (which may not be commutative though G is so) then the following results are obtained, as immediate consequences of our work:

THEOREM. *If G is a separable, noncompact, locally compact topological group which is amenable as a discrete group, then the radical $R(G)$, of $LUC(G)^*$ is infinite dimensional (see Theorem* III-6)

Combining this theorem with a known result, to be found in Rudin [15], which asserts that if G is compact abelian then $C(G)^*$ is semisimple one gets.

THEOREM III-4. *Let G be a separable abelian locally compact topological group. Then either $R(G) = \{0\}$ or $R(G)$ is infinite dimensional. Moreover $R(G) = \{0\}$ if and only if G is compact*[3].

THEOREM III-5. *Let G be a separable subgroup of a locally convex linear topological space. Then either $R(G) = \{0\}$ or $R(G)$ is infinite dimensional. Moreover $R(G) = \{0\}$ if and only if $G = \{0\}$.*

If we take G to be the real line R and therefore $LUC(G) = UC(R)$ to be the space of real uniformly continuous bounded functions on R then the algebra $UC(R)^*$, with convolution as multiplication, has as infinite dimensional radical. It is not hard to see that this holds true also for the complex valued uniformly continuous functions on R. Even this result for the real line seems to be unknown.

[3] It can be proved that $R(G) = \{0\}$ for any compact topological group G. Therefore Theorem III-4 holds true for any separable locally compact G, which is amenable as a discrete group. Thanks are due to Professor M. Rajagopalan for communicating this fact to me.

In the end it is a pleasure for me to thank Ranga R. Rao for the friendly and fruitful conversations I had with him. It was in fact his idea to use the functions $\{f_n\}$ in the proof of Theorem III-1.

Some notations. S is a topological semigroup if it has an associative multiplication and is a topological space (with no separation axioms) and for any fixed a in S the mappings $s \to as$ and $s \to sa$ are continuous from S to S. (i.e. multiplication is only separately continuous). We do not assume that $(x, y) \to xy$ from $S \times S \to S$ is continuous. As remarked in [19] p. 64 the multiplicative semigroup or linear continuous operators on a Banach space with the weak operator topology is only separately continuous.

G is a topological group if it is a group, has a Hausdorff topology and $(x, y) \to xy^{-1}$ from $G \times G \to G$ is continuous (i.e. in this case jointly continuous multiplication.)

If S is a set then $l_1(S), m(S)$ are defined as usual (see Day [5] p. 28) and if S has a topology then $C(S)$ is again defined as usual (see introduction). We stress that we deal only with real valued bounded functions in this paper. If X, Y are normed spaces then X^*, Y^* are their respective conjugate Banach spaces and if $T: X \to Y$ is linear then $T^*: Y^* \to X^*$ denotes the conjugate of T (see [5] pp. 14-17.)

If $A \subset S$ then 1_A is the function whose value is one an A and zero otherwise (when no ambiguity may arise, 1 will denote the constant one function on S, i.e. 1_S). If A, B are subsets of S then $A - B$ will invariably mean the set of points of A which are not in B.

If f is a function on S and $a \in S$ then f_a, f^a are defined by $(f_a)(s) = f(as)$ and $(f^a)(a) = f(sa)$ for each s in S. A linear manifold (which means the same as a linear subspace or in short a subspace) $L \subset m(S)$ is left invariant if $f_a \in L$ for each $f \in L$. In this case $\varphi \in L^*$ is left invariant if $\varphi(f_a) = \varphi(f)$ for each f in L and a in S. If L contains the constant functions then $\varphi \in L^*$ is called a mean if $\varphi(f) \geqq 0$ for $f \geqq 0$ in L and $\varphi(1_S) = 1$. $\varphi \in L^*$ is called a *finite mean of* L^* if there is a finite subset $\{a_1, \cdots, a_n\} \subset S$, and nonnegative $\alpha_1, \cdots, \alpha_n$ with $\Sigma\alpha_i = 1$ such that $\varphi(f) = \sum_{i=1}^{n} \alpha_i f(a_i)$ for each $f \in L$.

If S is a topological semigroup then $J_c l(S) = \{\varphi \in C(S)^*; \varphi(f_a) = \varphi(f)$ for each $f \in C(S)$ and $a \in S\}$ and $Jl(S) = \{\varphi \in m(S)^*; \varphi(f_a) = \varphi(f)$ for each $f \in m(S)$ and $a \in S\}$. For "left-ideal group" or "compact left-ideal group" see the introduction. A finite left ideal group is a left ideal group which contains a finite number of elements. If X is a Banach space and $Y \subset X$ a subspace then we write $\dim Y = n$ if Y is n dimensional, $0 \leqq n < \infty$, and $\dim Y = \infty$ if Y is not finite dimensional.

If X is a Banach space with conjugate space X^* then the w^* topology in X^* (sometimes called the X topology of X^*) is defined as in Day [5] p. 17.

A nonempty class F of subsets of a set S is called a field (σ-field) if it is closed under complementation and under the operation of taking finite (countable) unions.

II. The invariant mean on countable topological semigroups

The main theorem of this chapter is Theorem 2. The main tool for its proof is Theorem 1. The proof of Theorem 1 uses basically the same idea as the proof of Theorem (5.1) of [6]. It yields though a simpler proof even for the discrete case than Theorem (5.1) of [6].

DEFINITION 1. Let S be a semigroup. Define $l_a: m(S) \to m(S)$ by $l_a f = f_a$ for any a in S. If $L_0 \subset m(S)$ is a left invariant manifold then define $l_a^0: L_0 \to L_0$ by $l_a^0 f = f_a$ for any a in S and f in L_0. Denote in this case $\mathscr{L}_a = l_a^*: m(S)^* \to m(S)^*$, $\mathscr{L}_a^0 = (l_a^0)^*: L_0^* \to L_0^*$ and

$$J_0 l(S) = \{\varphi \in L_0^*; \mathscr{L}_s^0 \varphi = \varphi \text{ for each } s \in S\} .$$

THEOREM 1. *Let S be a left amenable semigroup and $L_0 \subset m(S)$ be a left invariant linear manifold containing the constants. Assume that there is a sequence $\{s_n\}_1^\infty \subset S$ such that*

$$\{\varphi \in L_0^*; \mathscr{L}_{s_n}^0 \varphi = \varphi, n = 1, 2, \cdots\} = J_0 l(S) .$$

If $\dim J_0 l(S) < \infty$ then each left invariant mean $\varphi \in L_0^*$ is a w^*-sequential limit of finite means, in other words there is a sequence of finite means φ_n in L_0^* such that $\varphi(f) = \lim_{n \to \infty} \varphi_n(f)$ for each $f \in L_0$.

REMARK 1. If we do not assume the existence of a countable sequence $\{s_n\} \subset S$ as above then the theorem does not remain true as is shown by the following example: Let G be an abelian compact hausdorff nonseparable topological group and let $L_0 = C(G)$. Then $\dim J_0 l(G) = 1$. Let $\varphi_0 \in L_0^*$ be the left invariant mean represented by the normalized Haar measure on G. Assume that $\varphi_0(f) = \lim_{n \to \infty} \varphi_n(f)$ for each $f \in C(G)$ where φ_n are finite means i.e., $\varphi_n(f) = \sum_{j=1}^k \alpha_j f(g_j)$ where α_j, g_j and k depend on n, $\alpha_j \geqq 0$ and $\sum \alpha_j = 1$.

If we call $\sigma(\varphi_n) = \{g_1, \cdots, g_k\}$, then $A = \bigcup_{n=1}^\infty \sigma(\varphi_n)$ is countable and therefore the group generated by A is countable and therefore the closure of this group, say G_0, is a closed separable subgroup of G. Since G is nonseparable $G \neq G_0$. But if $f \in C(G)$ satisfies $f(g) \geqq 1$ for $g \in G_0$ then $\varphi_n(f) \geqq 1$ since $\sigma(\varphi_n) \subset G_0$. Therefore $\varphi_0(f) \geqq 1$ which shows by [8] p. 248 that $\mu(G_0) = 1$ where μ is the normalized Haar measure on G. But if $a \in G$ and $a \notin G_0$ then $aG_0 \cap G_0 = \emptyset$ and so $1 = \mu(G) \geqq \mu(aG_0) + \mu(G_0) = 2$, which is a contradiction.

Thus the above theorem is not true if we do not assume the existence of the above sequence $\{s_n\}$. This is the reason why Luthar, in his theorem about the uniqueness of the invariant mean on an abelian semigroup, see [11] and this author, in proving the theorem about the finite dimensionality of the set of invariant means on a semigroup, (see [6]), had to handle first the case in which the semigroup was countable and only afterwards, by using arguments involving much more the algebraic properties of semigroups, to handle the uncountable case (which is not yet proved in its due generality).

Proof of the Theorem. Let $\varphi_0 \in L_0^*$ be a left invariant mean. Let $\psi \in m(S)^*$ be a norm preserving extension of φ_0. Since $1 \in L_0$ and φ_0 is a mean one has: $1 = \|\varphi_0\| = \varphi_0(1) = \psi(1)$. But $\|\psi\| = \|\varphi_0\|$ and so $1 = \|\psi\| = \psi(1)$. This implies as known that $\psi(f) \geqq 0$ if $f \geqq 0$. (In fact if $f \in m(S), 1 \geqq f \geqq 0$, would be such that $\psi(f) < 0$ then $\|1 - f\| \leqq 1$ and $\|\psi\| \geqq \psi(1 - f) = \psi(1) - \psi(f) > 1$) and therefore ψ is a mean. If ν is a left invariant mean on $m(S)$ then $\varphi_0' = \nu \odot \psi$ is a left invariant mean on $m(S)$ (see Day [4] p. 526–527 and p. 529 Cor. 2) which is an extension of φ_0. In fact, if $f \in L_0$ then $(\nu \odot \psi)(f) = \nu(h)$ where $h(s) = \psi(l_s f) = \varphi_0(l_s f) = \varphi_0(f)$. Thus $h(s)$ is constant on S and takes only the value $\varphi_0(f)$. Hence $\nu(h) = \varphi_0(f)$, since ν is a mean. (We notice that we could have applied an invariant extension theorem of R. J. Silverman see [16] in order to get immediately the existence of φ_0' but we prefered the above simple argument).

Let now $\{\varphi_\alpha'\}$ be a net of finite means in $m(S)^*$ such that $w^*\text{-}\lim_\alpha \varphi_\alpha' = \varphi_0'$ and $\lim_\alpha \|\mathscr{L}_s\varphi_\alpha' - \varphi_\alpha'\| = 0$ for each s in S. (see [6] p. 44, (5.8)*). If $\varphi_\alpha \in L_0^*$ is the restriction of φ_α' to L_0 then since $\lim_\alpha \varphi_\alpha'(f) = \varphi_0'(f)$ for each $f \in m(S)$ we get that $\lim_\alpha \varphi_\alpha(f) = \varphi_0(f)$ for each $f \in L_0$ and thus $w^*\text{-}\lim \varphi_\alpha = \varphi_0$ (in L_0^*). Moreover if $f \in L_0$ and $\|f\| \leqq 1$,

$$\begin{aligned} |(\mathscr{L}_s{}^0\varphi_\alpha - \varphi_\alpha)f| &= |\varphi_\alpha(l_s^0 f - f)| = |\varphi_\alpha'(l_s f - f)| \\ &= |(\mathscr{L}_s\varphi_\alpha' - \varphi_\alpha')f| \leqq \|\mathscr{L}_s\varphi_\alpha' - \varphi_\alpha'\| \to 0 \end{aligned}$$

for each s in S. This implies that $\lim_\alpha \|\mathscr{L}_s{}^0\varphi_\alpha - \varphi_\alpha\| = 0$ (where the norm now is that of L_0^*) for each s in S.

Let now $S(\varphi_0, 1/n) = \{\varphi \in L_0^*; \|\varphi - \varphi_0\| < 1/n\}$ and let V_n be a sequence of convex w^* neighborhoods of φ_0 which are w^*-closed such that $V_{n+1} \subset V_n$ for $n = 1, 2, \cdots$ and

$$\varphi_0 \in V_n \cap J_0 l(S) \subset S\left(\varphi_0, \frac{1}{n}\right) \cap J_0 l(S) .$$

The choice of such V_n's is possible since $J_0 l(S)$ is finite dimensional (see [6] p. 44 (5.5)* and p. 45). There is now an α'_n such that $\alpha \geqq \alpha'_n$ implies $\|\mathscr{L}^0_{s_i}\varphi_\alpha - \varphi_\alpha\| < 1/n$ for $i = 1, 2, \cdots, n$.

Since φ_0 is a w^* limit point of the net $\{\varphi_\alpha\}$ there is an $\alpha_n \geqq \alpha'_n$ such that $\varphi_{\alpha_n} \in V_n$. Write $\varphi_{\alpha_n} = \varphi_n$ and let ψ_0 be some w^*-limit point of the net $\{\varphi_n\}$. The set of means of L_0^* can be written as

$$\bigcap_{f \in L_0^*\ f \geqq 0} \{\varphi \in L_0^*;\ \|\varphi\| \leqq 1 \text{ and } \varphi(f) \geqq 0\}$$

and so is w^* compact. This shows the existence of such a ψ_0 (and so ψ_0 is even a mean). Moreover, if $f \in L_0$ $\|f\| \leqq 1$ and s_j is fixed then

$$\begin{aligned} |(\mathscr{L}^0_{s_j}\psi_0 - \psi_0)f| &\leqq |\mathscr{L}^0_{s_j}(\psi_0 - \varphi_n)f| \\ &\quad + |(\mathscr{L}^0_{s_j}\varphi_n - \varphi_n)f| + |(\varphi_n - \psi_0)f| \\ &\leqq |(\psi_0 - \varphi_n)l^0_{s_j}f| \\ &\quad + \|\mathscr{L}^0_{s_j}\varphi_n - \varphi_n\| + |(\varphi_n - \psi_0)f| \,. \end{aligned}$$

If $\varepsilon > 0$ is given then there is an $n_0 \geqq j$ such that $1/n_0 < \varepsilon/3$ and therefore for $n \geqq n_0$, $\|\mathscr{L}^0_{s_j}\varphi_n - \varphi_n\| < \varepsilon/3$. Since ψ_0 is a w^*-limit point of $\{\varphi_n\}$, there is an $n_1 \geqq n_0$ such that $|(\psi_0 - \varphi_{n_1})l^0_{s_j}f| < \varepsilon/3$ and $|(\varphi_n - \psi_0)f| < \varepsilon/3$. Thus $\mathscr{L}^0_{s_j}\psi_0 = \psi_0$ for each j and using the assumption of our theorem we get that $\psi_0 \in J_0l(S)$. But ψ_0 is also a w^* limit point of the sequence $\{\varphi_n\}_{n=k}^\infty \subset V_k$. Since V_k is w^* closed $\psi_0 \in V_k$ for each k. Thus $\psi_0 \in V_k \cap J_0l(S) \subset S(\varphi_0, 1/k) \cap J_0l(S)$. This shows that $\|\psi_0 - \varphi_0\| < 1/k$ for each k and so $\varphi_0 = \psi_0$. Therefore the sequence $\{\varphi_n\} \subset L_0^*$ has the unique w^*-limit point φ_0. Therefore $\lim_{n\to\infty} \varphi_n(f) = \varphi_0(f)$ for each $f \in L$ (see [6] p. 43 and replace there $m(G)$ by L_0). This finishes the proof of our theorem.

REMARK 2. $J_0l(S)$ coincides with the linear manifold spanned by the left invariant means in $J_0l(S)$. Since if $\varphi \in J_0l(S)$ and $\psi \in m(S)^*$ is any extension of φ and if ν is any left invariant mean of $m(S)^*$ then $\varphi' = \nu \odot \psi \in m(S)^*$ is a left invariant extension of $\varphi \in J_0l(S)$ (see begining of proof of the preceeding theorem). But by [6] p. 55 footnote 5 there are left invariant means φ'_1, φ'_2 in $m(S)^*$ such that $\varphi' = \alpha\varphi'_1 - \beta\varphi'_2$. If φ_i is the restriction of φ'_i to L_0 then $\varphi = \alpha\varphi_1 - \beta\varphi_2$ and φ_i are left invariant means of L_0^*.

DEFINITION 2. If X is a topological space then $A \subset X$ is called a Z-set if $A = \{x; f(x) = 0\}$ for some $f \in C(X)$, F_X will denote the field generated by the Z-sets and B_X is the σ-field generated by the Z-sets (or the σ-field of Baire subsets of X).

LEMMA 1. *Let S be a countable topological semigroup which is left amenable as a discrete semigroup. If the set of left invariant* *elements of* $C(S)$, $J_c l(S)$, *is finite dimensional then each left invariant mean* φ_0 *of* $C(S)^*$ *can be represented by a regular countable additive measure* m_0 *on* B_S.

Proof. Let $\varphi \in C(S)^*$ be a left invariant mean. Taking in the previous theorem $L_0 = C(S)$ we get that there is a sequence of finite means $\{\varphi_n\}$ such that $\lim_{n->\infty} \varphi_n(f) = \varphi(f)$ for each $f \in C(S)$.

If $a \in S$ then let m_a be the countable additive measure defined on B_S by: $m_a(B) = 1$ if and only if $a \in B$. m_a is regular and countably additive and since any finite mean can be represented by a linear combination of m_a's we get that φ_n are represented by countable additive regular measures m_n on B_S. Thus for each $f \in C(S)$

$$\varphi_0(f) = \lim_{n-\infty} \int f dm_n .$$

Applying now A. D. Alexandroff's theorem (for statement and proof see Varadarajan [17] p. 68-69 Theorem 19) there exists a *countably* additive measure m_0 on F_S such that

$$\varphi_0(f) = \int f dm_0 \text{ for each } f \in C(S) .$$

By a known theorem m_0 can be uniquely extended to a countably additive measure on B_S. (see [17] p. 45 Thm 18). By the second part of [17] Thm. 18 p. 45 this m_0 is even regular.

REMARK 3. Applying now the uniqueness part of Alexandroff's theorem on the representation of linear functionals by measures, (see Alexandroff [1] or Varadarajan [17] p. 39 Thm 5) we get that for any Z-set Z_0 one has $m_0(Z_0) = \inf \{\varphi_0(f); f \geqq 1_{Z_0}, f \in C(S)\}$.

THEOREM 2. *Let S be a countable topological semigroup which is left amenable as a discrete semigroup and which is a* T_1 *and regular topological space* (for definition see [10] p.113). *Then* $\dim J_c l(S) = n$, $n < \infty$, *if and only if S contains exactly n finite left-ideal groups.*

REMARK 4. (a) If φ is any invariant mean on $m(S)$ then its restriction to $C(S)$ is an invariant mean of $C(S)^*$. Thus in any case $\dim J_c l(S) \geqq 1$ (if S is left amenable as a discrete semigroup).
(b) Two different left-ideal groups are disjoint (each one is a minimal left ideal).

Proof of Theorem. S being countable is Lindelöf and being also regular is normal (see Kelley [10] p. 113) We show now that any closed $F \subset S$ is a Z-set.

Let $S - F = \{s_1, s_2, \cdots\}$ and let $f_n \in C(S)$ satisfy $0 \leqq f_n \leqq 1$ and $f_n(F) = 0$ while $f_n(s_n) = 1$ (Uryson's lemma).

Let $f(s) = \sum_{n=1}^{\infty} (1/2^n) f_n(s)$. Then $f \in C(S)$ and $\{s; f(s) = 0\} = \{F\}$. (This is the standard well known proof that any closed G_δ in a normal space is a Z-set).

Let φ_0 be a left invariant mean on $C(S)$ and let m_0 be the regular countably additive measure such that

$$\varphi_0(f) = \int f dm_0 \text{ for each } f \in C(S) .$$

If $S = \{t_1, t_2, \cdots\}$, then $1 = m_0(S) = \sum_{n=1}^{\infty} m_0(\{t_1\})$. Therefere there is some $a \in S$ (one of the t_i's) such that $m_0(\{a\}) > 0$. Now for any finite subset $F \subset S$

$$(1) \qquad \begin{aligned} m_0(\{sF\}) &= \inf \{\varphi_0(f); f \geqq 1_{sF}\} = \inf \{\varphi_0(f_s); f \geqq 1_{sF}\} \\ &\geqq \inf \{\varphi_0(h); h \geqq 1_F\} = m_0(F) . \end{aligned}$$

And the inequality is true since $f \geqq 1_{sF}$ implies that $f_s(t) = f(st) \geqq 1$ for $t \in F$ i.e. $f_s \geqq 1_F$.

Therefore if $a \in S$ satisfies $m_0(\{a\}) > 0$ and $s \in S$ we have $m_0(\{sa\}) \geqq m_0(\{a\}) > 0$. This shows that Sa is a finite left ideal (since $m_0(S) = 1$). If $A \subset Sa$ is a minimal left ideal then for $b \in A$, $Ab \subset A$ and since Ab is a left ideal, $Ab = A$. If we denote $A = \{b_1, \cdots, b_N\}$, the above shows that for each pair i, j, $1 \leqq i, j \leqq N$, there is some k, $1 \leqq k \leqq N$, such that $b_k b_i = b_j$. Taking $F = \{b_i\}$ in the inequality (1) we get that $m_0(\{b_j\}) = m_0(\{b_k b_i\}) \geqq m_0(\{b_i\}) > 0$ and interchanging i and j we get that $m_0(\{b_j\}) = m_0(\{b_i\}) > 0$ for each b_i, b_j in the finite minimal left ideal A, i.e. $m_0(\{b_1\}) = m_0(\{b_2\}) = \cdots = m_0(\{b_N\})$. If now b is any element of A then $m_0(bA) \geqq m_0(A) = N m_0(\{bb_1\})$. But $bA \subset A$ and therefore $m_0(bA) = j m_0(\{bb_1\})$ where j is the number of differents elements in bA. Thus $j = N$ and $bA = A$. This shows that A is a finite minimal left ideal which satisfies for each $b \in A$ that $bA = Ab = A$. This shows that A is a finite left ideal group.

If $s \in S$ and e is the identity of A then $sA = (se)A = A$ since $se \in A$. Thus $sA = A$ so that any finite left-ideal group is also what is (unnecessarily) called in [6] p. 34 a (l.i.l.c). (Also, obviously, any finite group and (l.i.l.c) is a left-ideal group.) Now the number of finite left-ideal groups in S is less than or equal to n (where $\dim J_c l(S) = n$) since if $A_1, \cdots, A_n, A_{n+1}$ would be finite left-ideal groups and we would define $\varphi_i \in C(S)$ by $\varphi_i(f) = [1/N(A_i)] \sum_{s \in A_i} f(s)$ where $N(A_i)$ is the number of elements of A_i then as easily checked φ_i is a left invariant mean on $C(S)$ (since $sA_i = A_i$ for each $s \in S$). But $\varphi_1, \cdots, \varphi_{n+1} \in C(S)^*$ are linearly independent. In fact if $\sum_1^{n+1} \alpha_i \varphi_i = 0$ and if we define f_i' on $\bigcup_1^{n+1} A_j$ by $f_i'(s) = 1$ for $s \in A_i$ and $f_i'(s) = 0$ if $s \in A_j$ for $j \neq i$ then we can, by Tietze's extension theorem find an extension $f_i \in C(S)$ of f_i'. For this f_i we have $0 = \sum \alpha_j \varphi_j(f_i) = \alpha_i \varphi_i(f_i) = \alpha_i$ which shows that $\alpha_i = 0$ so that $\dim J_c l(G) \geqq n+1$, which contradicts our assumption. Thus there are at most n finite left-ideal groups in S.* If $m(1 \leqq m \leqq n)$ is the number of the finite left-ideal groups in S then we get by [6] p. 34 Thm. 3.1 and p. 36 Remark 3.2 and [6] p. 55 footnote 5 that $\dim Jl(S) = m$ where $Jl(S)$ is the set of left invariant elements of $m(S)^*$. But any $\varphi \in J_c l(S) \subset C(S)^*$ has an extention $\varphi' \in Jl(S) \subset m(S)^*$ (see beginning of proof to Thm (II.1)). Thus if $\varphi_1, \cdots, \varphi_n$ are n linearly independent elements of $J_c l(S)$ and $\{\varphi_1', \cdots, \varphi_n'\} \subset Jl(S) \subset m(S)^*$ are extensions of $\varphi_1, \cdots \varphi_n$ respectively then $\varphi_1', \cdots, \varphi_n'$ are also linearly independent in $m(S)^*$. Since if $\sum_1^n \alpha_i \varphi_1' = 0$ then for each $f \in C(S) \subset m(S)$ we would have $\sum_1^n \alpha_i \varphi_i(f) = 0$ which would imply that $\alpha_1 = \alpha_2 = \cdots \alpha_n = 0$. Therefore $\{\varphi_1', \cdots, \varphi_n'\} \subset Jl(S)$ are linearly independent which shows that $m \geqq n$ and S contains exactly n finite left-ideal groups.

REMARK 5. We also proved at the end of this theorem that $\dim J_c l(S) = n$ implies $\dim Jl(S) = n$ where S is countable and left amenable as a discrete semigroup. That this does not hold true for noncountable S is shown by the following example: Let G be an abelian compact Hausdorff topological group which is not finite. Then by Theorem B of [6] p. 32 we get that $\dim Jl(G) = \infty$ while $\dim J_c l(G) = 1$ (The Haar measure is unique). In other words the restriction of the infinite dimensional space $Jl(G) \subset m(G)^*$ to $C(G)$ forms an one dimensional subspace of $C(G)^*$ which coincides with $J_c l(G)$. The end of the proof of our preceeding theorem shows that this cannot happen if G is countable.

COROLLARY 1. *Let S be a countable T_1 regular topological semigroup which is left amenable as a discrete semigroup. If S has left cancellation then* $\dim J_c l(G) = n$ $(n < \infty)$ *if and only if S is finite and is the union of n finite disjoint left-ideal groups.* $\dim J_c l(G) = 1$ *if and only if then G is a finite group.*

Proof. At the end of the last theorem it was in fact shown that $n = \dim J_c l(S) = \dim Jl(S)$ where $Jl(S)$ is the set of left invariant

* We could also proceed as follows: Let m, $1 \leqq m < \infty$, be the number of finite left ideal groups of S and let A be a compact left ideal group of S. Then A is a countable group and has a compact hausdorff topology in which multiplication is separately continuous. Hence by the theorem of Ellis (see Ellis [21] or Glicksberg-Deleeuw [19] p.p. 64-65 and p.p. 94-96) A is a compact topological group which is countable. Hence A has to be finite (since if m is its normalised Haar measure then $m\{a\} > 0$ for some a in A, hence $m(A) = \infty$, if A is infinite, which cannot be.). Therefore S contains, in our case exactly m compact left ideal groups. By Theorem IV-1 of the present paper $\dim J_c 1(S) = m$ which finishes the proof.

elements of $m(S)^*$. Applying Thm E of [6] p. 49 and remembering footnote 5 on p. 55 of [6] we get this corollary.

REMARK 6. If G is a discrete amenable group and $G' \subset G$ a subgroup then there exists a linear positive isometry from $Jl(G')$ into $Jl(G)$ (see Day [4] p. 534). Therefore, the assumption that $\dim Jl(G) = n$ implies that $\dim Jl(G') \leqq n$. If G is a topological group and $G' \subset G$ a subgroup then there does not generally exist a linear isometry from $J_c l(G')$ to $J_c l(G)$. In fact let G be a compact abelian hausdorff topological group. Then $\dim J_c l(G) = 1$. If now $G' \subset G$ is any countable (not finite) subgroup then G' being abelian, is amenable as a discrete group and satisfies all the assumptions of our previous corollary. Therefore $\dim J_c l(G') = \infty$, which shows that there cannot exist an isometry from $J_c l(G')$ into $J_c l(G)$. This theorem of Day was the main tool to pass from the countable case to the uncountable case when dealing with discrete groups (see [6] p. 46 proof of Cor (5.3)). The above example shows that this important tool is not more available when dealing with topological groups.

III. The invariant mean on separable topological groups

The main theorem of this chapter is Theorem 1. We have to restrict ourselves to topological groups rather than topological semigroups since our method works only for left uniformly continuous functions and on semigroups there may not be any uniformity at all which is consistent with the algebraic structure.

DEFINITION 1. Let G be a topological group and $U \subset G$ a neighborhood of the identity. We say that U totally covers G if $G \subset \bigcup_{i=1}^{k} Ua_i$: for some finite subset $\{a_1, \cdots, a_k\} \subset G$. (We should have said that U left totally covers G but we drop the "left" since we do not deal at all with the "right" case.)

We say that the topological group G has property (B) if it has a neighborhood of the identity U such that none of its powers totally covers G (or in other words for each n and each finite subset $\{a_1, \cdots, a_k\} \subset G$, $G - \bigcup_{i=1}^{k} U^n a_i \neq \varnothing$.)

REMARK 1.

(a) A noncompact locally compact group has property B since if U is a compact neighborhood of the identity then U^n is compact for each n and so $\bigcup_{i=1}^{k} U^n a_i$ is compact and therefore does not cover the whole of G.

(b) Any subgroup $G \neq \{0\}$ of a hausdorff locally convex linear topological space E has property B. Since if $0 \neq a \in G$ and f is a linear continuous functional on E such that $f(a) \neq 0$ then let $U = \{y \in G; |f(y)| < 1\}$. Then the n^{th} power of U is defined by $U^n = U + \cdots + U$ (n times). Thus if $y \in U^n$ then $y = u_1 + \cdots + u_n$ with $u_i \in U$ and $|f(y)| \leqq n$. Therefore $y \in \bigcup_{i=1}^k U^n + g_i$ implies that

$$|f(y)| \leqq \max_{1 \leqq i \leqq k} |f(g_i)| + n = K .$$

But there is a positive integer j which satisfies $|f(ja)| = j . |f(a)| > K$. Since $ja \in G$ this implies that G is not included in $\bigcup (g_i + U^n)$ so that G has property (B).

(C) Any topological metric group G which admits a right invariant *non bounded* metric (i.e. its topology can be given by a right invariant metric d such that for any $K>0$ there are $a, b \in G$ satisfying $d(a, b) > K$), has property B. It should be pointed out that any metric topological group admits as known (see G. Birkhof [2] or Kakutani [9]) a right invariant metric. Therefore the real requirement is that the metric should be unbounded. (If G is totally bounded and metric then any admissible invariant metric is bounded).

Assume that G admits a right invariant unbounded metric d. If e is the identity element of G then let $U = \{g; d(e, g) < 1\}$. Then for $u \in U^n$ $d(e, u) \leqq n$. This is true for $n = 1$. Assume that it holds for $n - 1$. If $u \in U^n$ then $u = u_1 u_2 \cdots u_n$ with $u_i \in U$. Then

$$\begin{aligned} d(u_1 u_2 \cdots u_n, e) &\leqq d(u_1 u_2 \cdots u_n, u_2 \cdots u_n) + d(u_2 \cdots u_n, e) \\ &\leqq d(u_1, e) + n - 1 \leqq n \end{aligned}$$

since d is right invariant. If $G \subset \bigcup_{i=1}^k U^n a_i$ then any $g \in G$ satisfies $g \in U^n a_i$ for some $1 \leqq i \leqq k$ and so $g = v a_i$ with $v \in U^n$. Thus

$$d(e, g) \leqq d(e, a_i) + d(a_i, v a_i) \leqq K + d(e, v) \leqq K + n = K_1$$

where $K = \max \{d(e, a_i), 1 \leqq i \leqq k\}$. But the metric d is unbounded and therefore there are $a, b \in G$ such that $d(e, ba^{-1}) = d(a, b) > K_1$ which is a contradiction. (As we see here it is enough that $d(x, y)$ should be a continuous unbounded right invariant pseudometric on G and it is not necessary that d, generates the topology of G)[4]

The following lemma is needed in what follows:

LEMMA 1. *Let G be a separable hausdorff topological group having property* (B) *and let $\{p_j\}_1^\infty$ be dense in G. Then for any open sym-*

---

[4] The following example of a group with property (B) seems to have some interest. Consider the space $L_p(0, 1)$, for $0 < p < 1$, with the metric $\int_0^1 |x(t) - y(t)|^p dt$. The function $F(x) = \int_0^1 |x(t)|^p dt$ is uniformly continuous and $F(nx) = n^p \int_0^1 |x(t)|^p dt \to \infty$ if $n \to \infty$ and $x \neq 0$. As known there is no nonzero continuous linear functional (or even character) on $L_p(0, 1)$, for $0 < p < 1$, and hence it is not even a locally convex linear topological space (see M. M. Day Bull. Amer. Math. Soc. **46** (1940), 816–823).

metric neighborhood of the identity U none of whose powers totally cover G there exist a left uniformly continuous nonnegative function F on G such that

$$\{g;\, 0 \leqq F(g) \leqq k\} = F^{-1}([0, k]) \subset \bigcup_{j=1}^{k+2} U_{p_j}^{2(k+2)} \qquad \text{for } k = 1, 2, 3, \cdots$$

Proof. Let $\{p_n\}$ be a countable dense subset of G. We define an increasing sequence of open subsets of G in the following way:

$$A_1 = Up_1 .$$

As well known $\bar{A}_1 = \cap VA_1$ where V ranges over all the neighborhoods of e and therefore $\bar{A}_1 \subset U^2 p_1$. Let

$$A_2 = U(\bar{A}_1 \cup Up_2) .$$

We get immediately that $Up_1 \cup Up_2 \subset A_2$ and $U\bar{A}_1 \subset A_2$ and

$$\begin{aligned} \bar{A}_2 &\subset U[U(\bar{A}_1 \cup Up_2)] \\ &\subset U^2(U^2 p_1 \cup Up_2) \subset U^4 p_1 \cup U^3 p_2 \subset U^4 p_1 \cup U^4 p_2 . \end{aligned}$$

Assume now that $A_1, A_2, \cdots, A_n$ have been chosen such that

$$Up_1 \cup Up_2 \cdots \cup Up_j \subset A_j \subset \bar{A}_j \subset U^{2j} p_1 \cup U^{2j} p_2 \cdots \cup U^{2j} p_j$$

and $U\bar{A}_{j-1} \subset A_j$ for each $j \leqq n - 1$ then we chose

$$A_n = U(\bar{A}_{n-1} \cup Up_n) .$$

We have that $(Up_1 \cup Up_2 \cdots \cup Up_n) \subset (A_{n-1} \cup Up_n) \subset A_n$ and that

$$\begin{aligned} U\bar{A}_{n-1} \subset A_n \subset \bar{A}_n &\subset U^2(\bar{A}_{n-1} \cup Up_n) \\ &\subset U^2[U^{2(n-1)} p_1 \cup \cdots \cup U^{2(n-1)} p_{n-1} \cup U^{2(n-1)} p_n] \\ &\subset U^{2n} p_1 \cup \cdots \cup U^{2n} p_n . \end{aligned}$$

In short our sequence of open substs A_n satisfies the following

(III. 1) $$\bigcup_{n=1}^{\infty} A_n = G$$

($\bigcup_{i=1}^{\infty} Up_i \subset \bigcup_{n=1}^{\infty} A_n$ and $G = \bigcup_{i=1}^{\infty} Up_i$ since otherwise there would be some $a \in G$ such that $a \notin Up_i$ for each i i.e., $p_i \notin U\alpha$ for each i (U is open symetric) which cannot be since $\{p_i\}$ is dense in G.)

(III. 2) $$U\bar{A}_n \subset A_{n+1} \subset \bigcup_{i=1}^{n+1} U^{2(n+1)} p_i .$$

We can also assume that $A_n - A_{n-1} \neq 0$ for each n (Where for $A, B \subset G$, $A - B$ are the elements of A which are not in B). (Since otherwise we would choose $A_{n_1} = A_1$, n_2 to be the first $n > n_1$ for which $A_n - A_{n_1} = \varnothing$ and if n_{k-1} was already chosen then let n_k be the first $n > n_{k-1}$ for which $A_n - A_{n_{k-1}} \neq \emptyset$. There is such a n_k since U^k does not totally cover G for any k. Obviously the sequence $A'_k = A_{n_k}$ would satisfy (III. 1) (III. 2) in addition to $A'_k - A'_{k-1} \neq \emptyset$)

It is proved in A. Weil [18] p. 13 that if E is a uniform space and V' a neighborhood of the diagonal in $E \times E$ and if $p_0 \in E$ then there exists a uniformly continuous function $f: E \to [0, 1]$ such that $f(p) = 0$ and $f(q) = 1$ for $q \in E - V'(p)$. (where $V'(p) = \{q \in E; (p, q) \in V'\}$ and for $A \subset E$, $V'_\alpha(A) = \bigcup_{p \in A} V'_\alpha(p)$). But moreover, if we chose a fixed sequence of symmetric neighborhoods of the diagonal (i.e., elements of the uniformity) in $E \times E$ say V'_n which satisfy $V'_{n+1} V'_{n+1} \subset V'_n$ (for notation see [18] A. Weil) for $n = 0, 1, 2 \cdots$ and $V'_0 \subset V'$ then the function $f: E \to [0, 1]$ can even be chosen to satisfy the condition $|f(q) - f(r)| < 1/2^{m-1}$ whenever $(p, q) \in V'_m$ (see [18] p. 14). We notice also that the sequence V'_n is not dependent upon p. But the same proof yields actually more: If V' is a neighborhood of the diagonal (a member of the uniformity) and the sequence V'_n is chosen as above and if P is any subset of E then there exists a uniformly continuous $f: E \to [0, 1]$ such that $f(p) = 0$ if $p \in P$ and $f(q) = 1$ if $q \in E - V'(P)$. Returning now to our group we consider its left uniformity i.e. the uniformity whose elements are all the sets of the form $V' = \{(p, q); q \in Vp\}$ where $p, q \in G$ and V ranges over all the neighborhoods of e. Let V_n, $n = 0, 1, 2 \cdots$ be a fixed sequence of symmetric neighborhoods of e in G such that $V_0 = U$ and $V_{n+1} V_{n+1} \subset V_n$ for each n. Then $V'_n = \{(p, q); q \in V_n p\} \subset G \times G$ are symmetric elements of the uniformity which satisfy $V'_{n+1} V'_{n+1} \subset V'_n$ (since for each $p \in G$ $(V'_{n+1} V'_{n+1})(p) = V'_{n+1}(V_{n+1} p) \subset V_n p = V'_n(p)$). Therefore since $V'_0(\bar{A}_k) = V_0 \bar{A}_k = U\bar{A}_k \subset A_{k+1}$ there exists a left uniformly continuous function $f_k: G \to [0,1]$ such that $f_k(A_k) = 0$ and $f_k(G - U\bar{A}_k) = 1$ which implies that $f_k(G - A_{k+1}) = 1$. Moreover if $(p, q) \in V'_m$ i.e. if $q \in V_m p$ then $|f_k(p) - f_k(q)| < 1/2^{m-1}$ for each k.

Consider now the sequence of functions

$$h_k(g) = f_k(g) + k - 1 \qquad \text{for } k = 1, 2, 3, \cdots$$

We have:

$$h_1(g) = \begin{cases} 0 \text{ on } \bar{A}_1 \\ 1 \text{ on } G - A_2 \end{cases};$$

$$h_2(g) = \begin{cases} 1 \text{ on } \bar{A}_2 \\ 2 \text{ on } G - A_3 \end{cases}$$

$$h_k(g) = \begin{cases} k - 1 \text{ on } \bar{A}_k \\ \quad k \quad \text{ on } G - A_{k+1} \end{cases}$$

and also

(III. 3) $$|h_k(p) - h_k(q)| = |f_k(p) - f_k(q)| < 1/2^{m-1}$$

$q \in V_m P$ uniformly in k. (Our sequence of symmetric neighborhoods V_m is the same for all A_k).

Define now the required function F on G as follows:

$$\text{(III. 4)} \qquad F(g) = \begin{cases} h_1(g) & \text{for } g \in A_2 \\ h_2(g) & \text{for } g \in A_3 - A_2 \\ h_k(g) & \text{for } g \in A_{k+1} - A_k \text{ if } k \geqq 2 \,. \end{cases}$$

Since $A_k \subset A_{k+1}$ and $\bigcup_{n=1} A_k = G$, F is a well defined and real valued function on G which satisfies that $\{g: 0 \leqq F(g) \leqq k\} \subset A_{k+2}$, since if $g \notin A_{k+2}$ then $g \in A_n - A_{n-1}$ for some $n > k + 2$ and so

$$F(g) = h_n(g) \geqq n - 1 \geqq k + 1 \,.$$

Therefore by (III. 2) $F^{-1}([\mathrm{O}k]) \subset \bigcup_{i=1}^{k+2} U^{2(k+2)} p_i$. We also notice that $F(g)$ is not bounded since $A_k - A_{k-1} \neq \emptyset$ and for $g \in A_k - A_{k-1}$, $F(g) = h_k(g) \geqq k - 1$. We prove now that F is left uniformly continuous:

If $\varepsilon > 0$ is given then there exists an m such that $2^{-m+2} < \varepsilon$. We shall show that for any $p, q \in G$ such that $q \in V_m p$, $|F(p) - F(q)| < \varepsilon$. Assume therefore that $q \in V_m p$. If p and q are both in $A_{k+1} - A_k$ for some $k \geqq 2$ or are both in A_2 we can immediately conclude from (III. 3), (III. 4) that: $|F(p) - F(q)| = |h_k(p) - h_k(q)| < 1/2^{m-1} < \varepsilon$ where $k = 1$ if p and q are both in A_2.

If the above is not the case then let i be the first index for which $p \in A_i$ and j be the first index for which $q \in A_j$. Assume that $i < j$. Since $q \in V_m p \subset Up \subset UA_i \subset A_{i+1}$ (see (III. 2)) we have that $j = i + 1$ and $q \in A_{i+1}$ (we can assume that $i \geqq 2$ since if $i = 1$ then $p, q \in A_2$ and we alredy dealt with this case). Thus $p \in A_i - A_{i-1}$ and

$$q \in A_{i+1} - A_i \subset G - A_i \quad \text{and} \quad q \in V_m p \,.$$

Therefore:

$$\begin{aligned} |F(q) - F(p)| &= h_i(q) - h_{i-1}(p) \\ &= h_i(q) - (i-1) + (i-1) - h_{i-1}(q) \\ &= h_i(q) - h_i(p) + h_{i-1}(q) - h_{i-1}(p) \end{aligned}$$

since:

$$h_{i-1}(g) = \begin{cases} i - 2 \text{ on } \bar{A}_{i-1} \\ i - 1 \text{ on } G - A_i \end{cases}$$

and

$$h_i(g) = \begin{cases} i - 1 & \text{on } \bar{A}_i \\ \quad i & \text{on } G - A_{i+1} \,. \end{cases}$$

Therefore, remembering that $q \in V_m p$ and applying (III. 4), we get:

$$\begin{aligned} |F(q) - F(p)| &\leqq |h_i(q) - h_i(p)| + |h_{i-1}(q) - h_{i-1}(p)| \\ &\leqq 1/2^{m-1} + 1/2^{m-1} < \varepsilon \,. \end{aligned}$$

If $j < i$ then, remembering that $q \in V_m p$ if and only if $p \in V_m q$ (V_m is symmetric), and interchanging p and q we get that $|F(p) - F(q)| < \varepsilon$ for this case also.

COROLLARY. *A topological group G has property B if and only if there exists a left uniformly continuous real valued unbounded function on G*[5].

Proof. If G has property B then the function $F(g)$ of the preceeding lemma is unbounded and left uniformly continuous.

Conversely if $F(g)$ is an unbounded left uniformly continuous function on G there is a neighborhood of the identity u such that $|F(a) - F(b)| < 1$ if $b \in Ua$ for any $a, b \in G$. We show now that if $b \in U^n a$ then $|F(a) - F(b)| \leqq n$. Assume that this is true for n. If $b \in U^{n+1}a$ then $b \in U(U^n a)$ and so there is some $c \in U^n a$ such that $b \in Uc$. Therefore $|F(b) - F(c)| < 1$ and so

$$|F(b) - F(a)| \leqq |F(b) - F(c)| + |F(c) - F(a)| \leqq 1 + n .$$

Assume now now that $G = \bigcup_{i=1}^{k} U^n a_i$ where $a_i \in G$. If now $g \in G$ then $g \in U^n a_i$ for some $1 \leqq i \leqq k$ and so $|F(g) - F(a_i)| \leqq n$ which implies that

$$|F(g)| \leqq n + \max_{1 \leqq i \leqq k} |F(a_i)| .$$

This contradicts the assumption that $F(g)$ is not bounded.

DEFINITION 2. We denote by $LUC(G) \subset C(G)$ the norm closed subspace of $C(G)$ of left uniformly continuous functions on G, i.e. $f \in C(G)$ is in $LUC(G)$ if and only if for each $\varepsilon > 0$ there is a neighborhood of the identity, V in G such that $|f(vg) - f(g)| < \varepsilon$ for each $v \in V$ and $g \in G$.

---

[5] This corollary is an immediate consequence of a theorem of M. Atsuji (see Canad. J. Math. **13** (1961), p. 661) who proved that it holds true for any uniform space. Thanks are due to K. Ross and W. Comfort for communicating it to me. The above corollary (which is not used in what follows) gives in fact a characterization of what may be called "uniformly pseudocompact groups" i.e. groups for which every uniformly continuous real function is bounded. It states: Each left uniformly continuous real function on the topological group G is bounded if and only if each neighborhood of the identity has some power which totally covers G (see def. 1 of this ch.). The following example of an abelian metric group for which every uniformly continuous real function is bounded but the group is not totally bounded (i.e. its completion is not compact) has been given by W. W. Comfort and K. A. Ross in [**23**]. Let $G = T^{\aleph_0}$ (where T is the circle group) and define for $x = \{x_n\}$, $y = \{y_n\}$ in G, $x \cdot y = \{x_n y_n\}$. The metric d is defined by $d(x, y) = \sup \{|x_n - y_n|; n = 1, 2, \cdots\}$. These remarks and the above corollary are given here only for the general information of the reader and are not used later on. The lemma preceeding the above corollary is though, used heavily in what follows.

$J_u l(G)$ will denote the space of left invariant element of $LUC(G)^*$. Since $LUC(G)$ will play the role of L_0 of Theorem II-1, define $l_a^0 f = f_a$ for a in S and f in $LUC(G)$. Also $\mathscr{L}_a^0 = (l_a^0)^*$.

THEOREM 1. *Set G be a separable hausdorff topological group which is amenable as discrete group and satisfies property* (B). *Then $J_u l(G)$ is infinite dimensional. As an immediate consequence $J_c l(G)$ is infinite dimensional.*

Proof. We remark first that $LUC(G)$ is a left invariant subspace of $m(G)$ containing the constant functions, since if $f \in LUC(G)$ and $a \in G$ then let U be a neighborhood of the identity e of G such that $|f(ug) - f(g)| < \varepsilon$ for each u in U and g in G. Then $|f(uag) - f(ag)| < \varepsilon$ for u in U and g in G. If V is a neighborhood of e such that $aV \subset Ua$ then

$$|(l_a f)(vg) - (l_a f)(g)| = |f(avg) - f(ag)| < \varepsilon$$

for each v in V and g in G which shows that $l_a f \in LUC(G)$.

G is amenable as a discrete group and therefore there exists a left invariant mean μ on $m(G)$. The restriction of μ to $LUC(G)$ is a left invariant mean. Therefore in any case $\dim J_u l(G) \geqq 1$. Assume now that $\dim J_u l(G) = n$ where $0 < n < \infty$. We shall show that in this case G has not property (B). Let $\{p_n\}$ be a countable dense subset of G and let $\varphi \in LUC(G)^*$ satisfy $\|\varphi\| = 1$ and $\mathscr{L}_{p_n}^0 \varphi = \varphi$ for $n = 1, 2, \cdots$. Let $a \in G$, then for $f \in LUC(G)$, $\varphi(l_{p_n}^0 f) = \varphi f$ and so:

$$\begin{aligned}(\mathscr{L}_a^0 \varphi - \varphi) f &= \varphi(l_a^0 f - f) \\ &= \varphi[(l_a^0 - l_{p_n}^0) f] + \varphi(l_{p_n}^0 f - f) \\ &= \varphi[(l_a^0 - l_{p_n}^0) f] \,.\end{aligned}$$

But for any $\varepsilon > 0$ there is a neighborhood V of e such that $|f(vg) - f(g)| < \varepsilon$ for g in G and $v \in V$, i.e. surely $|f(vag) - f(ag)| < \varepsilon$ whenever $v \in V$. Thus for any $b \in Va$ we have that $\|(l_b^0 - l_a^0) f\| < \varepsilon$. Since p_n is dense in G there is some p_j in Va. For this p_j we can write $|(\mathscr{L}_a^0 \varphi - \varphi) f| \leqq |\varphi(l_a^0 - l_p^0)(f)| \leqq \|(l_a^0 - l_{p_n}^0) f\| < \varepsilon$. This shows that $\varphi \in J_u l(G)$ or that $\{\varphi \in LUC(G)^*; \mathscr{L}_g^0 \varphi = \varphi \text{ for } g \in G\} = \{\varphi \in LUC(G)^*; \mathscr{L}_{p_n}^0 \varphi = \varphi, n = 1, 2, \cdots\}$ Denoting $L_0 = LUC(G)$ we can apply Theorem II-1 to get that for any left invariant mean φ of $LUC(G)^*$ there exists a sequence of finite means $\{\varphi_n\}$ such that $\lim_{n \to \infty} \varphi_n(f) = \varphi(f)$ for each $f \in LUC(G)$. We choose φ as a *two sided* invariant *mean on* $m(G)$. (see [4] p. 529) This φ will be fixed till the end of the proof. Then the restriction of this φ (which we again denote by this same φ) to $LUC(G)$ will be at least a left invariant mean on $LUC(G)$. Therefore $\varphi(f) = \lim_{n \to \infty} \varphi_n(f)$ for each $f \in LUC(G)$ where φ_n is a sequence of finite means of $LUC(G)^*$[6]. Let U be a neighborhood of e such that none of its powers totally covers G. We may assume that U is symmetric (since any neighborhood of e included in U also has this property). If $A \subset G$ we shall write $\varphi(A)$ instead of $\varphi(1_A)$ (we remember that $\varphi(f)$ is defined for any f in $m(G)$). We shall show at first that $\varphi(U^n) > 0$ for some integer $n > 0$. This will immediately yield that U^{2n} totally covers G, which is the desired contradiction.

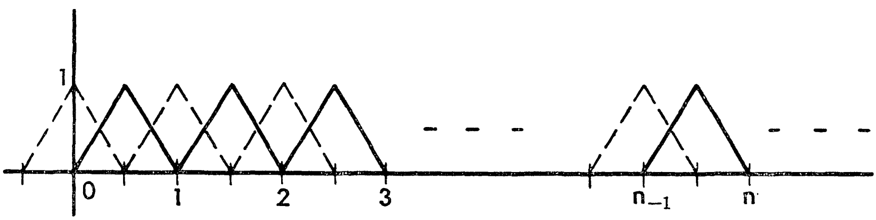
Define the following bounded uniformly continuous functions on the real line:

$$f_n(x) = \begin{cases} 1 - 2\left|x - \left(n - \frac{1}{2}\right)\right| & \text{if } n - 1 \leqq x \leqq n \\ 0 & \text{otherwise} \end{cases}$$

and

$$g_{n+1}(x) = \begin{cases} 1 - 2\,|x - n| & \text{if } n - \frac{1}{2} \leqq x \leqq n + \frac{1}{2} \\ 0 & \text{otherwise}\,. \end{cases}$$

Since the functions $f_i(x)$, $f_j(x)$ (or $g_i(x)$, $g_j(x)$) have disjoint carriers if $i \neq j$ the two functions $f(x) = \sum_1^\infty f_n(x)$ and $g(x) = \sum_1^\infty g_n(x)$ are well defined, their graph is plotted:



[6] This does not imply that φ can be represented by a countably additive measure on the Baire field of G. Consider in fact the following example: Let G be the additive group of rationals with the metric $|r_2 - r_1|$ and let a be an irrational number. Let r_n be a sequence of rationals converging to a and let m_n be the point measure concentrated at r_n. Then $\lim_{n\to\infty}\int f dm_n = \lim_{n\to\infty} f(r_n) = Af$ exists for each uniformly continuous bounded f (and equals $\tilde{f}(a)$ where $\tilde{f}$ is the uniformly continuous extension of f to the whole real line). Assume now that $Af = \int f dm$ for some countably additive real valued measure m on B_G and consider the sequence of uniformly continuous functions defined for x in G by

$$f_n(x) = \begin{cases} 1 - n\,|x - a| & \text{if } |x - a| \leqq \frac{1}{n} \\ 0 & \text{otherwise.} \end{cases}$$

Then $f_n(x) \downarrow 0$ for each $x \in G$ and $|f_n(x)| \leqq 1$. Therefore $Af_n = \int f_n dm \to 0$ by Lebesgue's bounded convergence theorem. But $Af_n = \lim_{k\to\infty} f_n(r_k) = 1$ for each n, which cannot be.

If though, G would be a locally compact group then the above relation $\varphi(f) = \lim_{n\to\infty} \varphi_n(f)$ would imply that φ can be represented by a countably additive measure on B_G. (see Dieudonne: Sur le produit de composition Compositio Math. 1954 p. 28). In this particular case the proof of our theorem could be simplified.

where $f(x)$ is represented by the solid line while $g(x)$ by the interrupted line. If $\{a_n\}$ is any bounded sequence of reals then it may easily be proved that both $\sum a_n f_n(x)$ and $\sum a_n g_n(x)$ are bounded uniformly continuous functions on the real line. Therefore if $F(g)$ is the left uniformly continuous real valued function on the group G which satisfies.

$F^{-1}([Ok]) \subset \bigcup_{j=1}^{k+2} U^{2(k+2)} p_j$ (see Lemma III-1) then surely $\sum a_n f_n(F(g))$ and $\sum a_n g_n(F(g))$ will both be bounded left uniformly continuous functions on G. But since $\sum_1 f_n(x) + \sum_1 g_n(x) \geqq d \geqq 0$ for some $d > 0$, for each $x \geqq 0$ we have that $\sum f_n(F(g)) + \sum g_n(F(g)) \geqq d \geqq 0$ for each g of G. Therefore $\varphi[\sum f_n(F(g)) + \sum g_n(F(g))] > 0$ and so either

$$\varphi[\sum f_n(F(g))] > 0 \quad \text{or} \quad \varphi[\sum g_n(F(g))] > 0 .$$

Assume therefore that $\varphi[\sum f_n(F(g))] > 0$ (for the other case the proof is similar) and define the following linear positive functionals on the Banach space m of all the bounded real sequences $\{a_k\}$ (with the sup norm):

$$\varphi'_n[a_k] = \varphi_n\left[\sum_{k=1}^{\infty} a_k f_k(F(g))\right]$$

and

$$\varphi'\{a_k\} = \varphi\left[\sum_{k=1}^{\infty} a_k f_k(F(g))\right]$$

where φ_n is the sequence of finite means of $LUC(G)^*$ which satisfies $\lim_{n->} \varphi_n(f) = \varphi(f)$ for each f in $LUC(G)$. But for any $f \in LUC(G)$.

$\varphi_n(f) = \sum_{i=1}^{j} \alpha_i f(g_i)$ where $\alpha_i \geqq 0 \sum \alpha_i = 1$ (and j, $\{\alpha_i\}$ and $g_i \in G$ depend on φ_n.) Therefore as is easily seen

$$\varphi'_n\{1\} = \varphi_n\left(\sum_{k=1}^{\infty} f_k(F(g))\right) = \sum_{k=1}^{\infty} \varphi_n f_k(F(g)) = \sum_{k=1}^{\infty} \varphi'_n\{1_k\}$$

where $\{1\} \in m$ is the sequence whose constant value is 1 and $\{1_k\} \in m$ is the sequence which is identically zero except at the place k where it is 1. This shows that $\varphi'_n \in Q[l_1] \subset m^*$ where l_1 is the Banach space of all the absolute convergent real sequences $\{b_i\}$ with norm $\sum |b_i|$ and $Q: l_1 \to l_1^{**} = m^*$ is the natural maping from the Banach space l_1 into its second adjoint. (see Day [5] pp. 29–30). But l_1 is weakly sequentially complete ([5] p. 33 Cor. 3) and therefore $Q[l_1]$ is w^*-sequentially complete in m^*. (for notation see Day [5] p. 17). Therefore we have the following situation: If $n \to \infty$ then

$$\varphi'_n\{a_k\} = \varphi_n[\sum a_k f_k(F(g)] \to \varphi[\sum a_k f_k(F(g))] = \varphi'\{a_k\} .$$

Thus $\varphi' \in l_1$ which immediately implies that $\varphi'\{1\} = \sum_{k=1}^{\infty} \varphi'\{1_k\}$. But by definition $\varphi'\{1_k\} = \varphi[f_k(F(g))]$. Thus

$$0 < \varphi[\sum f_k(F(g))] = \varphi'\{1\} = \sum \varphi'\{1_k\} = \sum \varphi f_k(F(g))$$

and since $f_k(F(g)) \geqq 0$ for each g in G and $\varphi \geqq 0$ we have that for at least one $k > 0$, $\varphi[f_k(F(g))] = c > 0$. Now

$$\{t; f_k(t) > 0\} \subset [k-1, k] \subset [0, k]$$

and so

$$\{g; f_k(F(g)) > 0\} \subset \{g; F(g) \in [0, k]\} = F^{-1}[0, k] \subset \bigcup_{j=1}^{k+2} U^{2(k+2)}p_j .$$

But we can easily find (as in elementary integration theory) a function of the form $h(g) = \sum_1^l \alpha_i 1_{A_i}(g) \in m(G)$ such that $\alpha_i \geqq 0$ $\alpha_l = 0, A_1, \cdots, A_{l-1}$ form a partition of $\{g; f_k(F(g)) > 0\}$ and $A_l = G - \{g; f_k(F(g)) > 0\}$ and $0 \leqq f_k(F(g)) - h(g) < c/3$. If we remember now that φ is defined on all of $m(G)$ (and we have used till now only its restriction to $LUC(G)$) we can write

$$\begin{aligned} c &= \varphi[f_k(F(g))] \\ &= \varphi[f_k(F(g)) - h(g)] + \varphi[h(g)] \leqq c/3 + \varphi[h(g)] . \end{aligned}$$

Therefore $\varphi(h) > 0$ which implies immediately that $\varphi(A_i) > 0$ for some $1 \leqq i \leqq l-1$. Since $A_i \subset \bigcup_{j=1}^{k+2} U^{2(k+1)}p_j$ we get that $\varphi(U^{2(k+2)}p_j) > 0$ for some j and using the fact that φ is also a *right* invariant mean we get that $\varphi(U^{2(k+2)}) > 0$ (Remember that $\varphi(A) = \varphi(1_A) = \varphi(1_A^g) = \varphi(1_{Ag^{-1}})$ for any $g \in G$ and $A \subset G$).

Let now $V = U^{2(k+2)}$. We shall prove that $V^2 = U^{4(k+2)}$ totally covers G, which will contradict the assumption that no power of U totally covers G. U is symmetric and therefore so is V and $\varphi(V) > 0$. Assume that V^2 does not totally cover G. Then we chose an infinite sequence of elements $\{a_n\} \subset G$ this way: $a_1 = e$. Since $G \neq V^2 a_1$ let $a_2 \notin V^2 a_1$. Thus $Va_2 \cap Va_1 = \varnothing$ (since $V^{-1} = V$). If $a_1, \cdots, a_{n-1}$ have been chosen such that $Va_i \cap Va_j = \varnothing$ if $i \neq j$ and $1 \leqq i, j \leqq n-1$ then since $G \neq \bigcup_1^{n-1} V^2 a_i$ there is some element $a_n \notin \bigcup_1^{n-1} V^2 a_i$. Thus $a_n \notin V^2 a_i$ for each $1 \leqq i \leqq n-1$ and so $Va_n \cap Va_i = \varnothing$ for $1 \leqq i \leqq n-1$. Therefore for any $n > 0$

$$1 = \varphi(g) \geqq \varphi(Va_1) + \varphi(Va_2) + \cdots + \varphi(Va_n) = n\varphi(V) .$$

This shows that $\varphi(V) = 0$ which is a contradiction and so $V^2 = U^{4(k+2)}$ totally covers G. This proves that $J_u l(G)$ is infinite dimensional. As an immediate consequence one gets that $J_c l(G)$ is infinite dimensional as follows: G is amenable and so surely $\dim J_c l(G) \geqq 1$ and $\dim J_u l(G) \geqq 1$. Assume now that $\dim J_c l(G) = n$, $n < \infty$. We show that this implies that $\dim J_u l(G) \leqq n$, which cannot be.

If $\psi \in J_u l(G)$ then it has a left invariant extension $\psi'' \in m(G)^*$ (see

Remark II-2). The restriction ψ' of this ψ'', to $C(G)$ is left invariant and so any $\psi \in J_u l(G)$ has an extension $\psi' \in J_c l(G)$. If $\dim J_c l(G) = n$, $n < \infty$, and $\{\psi_1, \cdots, \psi_{n+1}\} \subset J_u l(G)$ would be linearly independent then let $\{\psi'_1, \cdots, \psi'_{n+1}\} \subset J_c l(G)$ be respective extensions. Then $\sum_1^{n+1} \alpha_1 \psi'_i = 0$ for some reals α_i would surely imply that $\sum_1^{n+1} \alpha_i \psi_i(f) = 0$ for any f in $LUC(G)$ and so $\alpha_i = 0$ for $1 \leqq i \leqq n + 1$. Therefore $\dim J_u l(G) \leqq n$ which cannot be.*

THEOREM 2. *Let G be a separable locally compact hausdorff topological group which is amenable as a discrete group. Let $J_c l(G) \subset C(G)^*$ be the space of left invariant elements of $C(G)^*$, $J_u l(G) \subset LUC(G)^*$ be the space of left invariant elements of $LUC(G)^*$. Then*

(1) *Either* $\dim J_c l(G) = 1$ *or* $\dim J_c l(G) = \infty$ *and furthermore* $\dim J_c l(G) = 1$ *if and only if G is compact.*

(2) *Either* $\dim J_u l(G) = 1$ *or* $\dim J_u l(G) = \infty$ *and furthermore* $\dim J_u l(G) = 1$ *if and only if G is compact.*

REMARK 2. (a) The reader should remember that at least any abelian or solvable, or locally finite group is amenable as a discrete group. (see Day [4] pp. 516–518 for these and more examples)

(b) This theorem is not known even for the real line R. It asserts that $C(R)^*$ ahd $LUC(R)^*$ both have an infinite dimensional subspace of invariant elements.

Proof of theorem. G is amenable and so the restriction of any left invariant mean to $C(G)$ or $LUC(G)$ is a left invariant mean of $C(G)$ or $LUC(G)$. Thus $\dim J_c l(G) \geqq 1$ and $\dim J_u l(G) \geqq 1$ in any case.

If G is compact then $LUC(G) = C(G)$ as well known (see A. Weil [18]) and there is a unique left invariant mean on $C(G)$ (which is represented by the normalized Haar measure on G). Thus by the Remark II-2 we get that $\dim J_u l(G) = \dim J_c l(G) = 1$.

Assume now that $\dim J_u l(G) = n$, $n < \infty$. Then G is compact (since otherwise it would be noncompact locally compact and therefore would satisfy property B and by the previous theorem would satisfy $\dim J_u l(G) = \infty$) Therefore $n = 1$. Thus $\dim J_u l(G)$ can be either 1 or ∞ and $\dim J_u l(G) = 1$ if and only if G is compact. Using in the same way the previous theorem one immediately gets the remaining part of this theorem. Remembering that any nonzero subgroup of the

---

* In fact if A is any left invariant subspace of $m(G)$ containing $LUC(G)$ and $J_A 1(G)$ the space of left invariant elements of A^*, then as above, $\dim J_u 1(G) \leqq \dim J_A 1(G)$ which shows that Theorem III-1 holds true $C(G)$ is replaced by A. All the following theorems involving $C(G)$ could be shown to hold true when $C(G)$ is replaced by A. We could take as A, for instance, the space of all bounded Baire measurable functions on G.

additive group of a hausdorff locally convex linear topological space has property (B) (see, Remark III-1 (b)) and using in the same way Theorem III-1 one immediately obtains.

THEOREM 3. *Let G be any separable subgroup of the additive group of a hausdorff locally convex linear topological space. Then*

(1) *Either* $\dim J_u l(G) = 1$ *or* $\dim J_u l(G) = \infty$ *and furthermore* $\dim J_u l(G) = 1$ *if and only if* $G = \{0\}$.

(2) *Either* $\dim J_c l(G) = 1$ *or* $\dim J_c l(G) = \infty$ *and furthermore* $\dim J_c l(G) = 1$ *if and only if* $G = \{0\}$.

EXAMPLE 1. From the above theorems it follows that for separable locally compact groups (which are amenable as discrete groups) and for separable subgroups of a hausdorff locally convex linear topological space $\dim J_c l(G) = \dim J_u l(G)$ invariably holds. We give now on example of an abelin countable hausdorff topological group which satisfies $\dim J_u l(G) = 1$ while $\dim J_c l(G) = \infty$. Let G' be a compact abelian separable metric group which is not finite and let $d(x, y)$ be an admissible invariant metric on G'. Then $f \in LUC(G')$ if and only if f is uniformly continuous on G' as a metric space with the metric d. Let $\{g_1 g_2, \cdots\}$ be a countable dense subset of G' and let G be the group generated by $\{g_1 g_2, \cdots\}$. Then G is a countable Hausdorff abelian topological group and therefore G is T_1 and regular (even completely regular see [18] p. 13). Therefore G is amenable as a discrete group and hence we can apply Corollary II-2 to get that $\dim J_c l(G) = \infty$.

Consider now $LUC(G)$. Any f in $LUC(G)$ has a unique uniformly continuous extension $f' \in C(G')$ such that $\sup_{g \in G} |f(g)| = \sup_{g \in G'} |f'(g)|$. But any $f' \in C(G')$ is uniformly continuous on the (compact) metric space (G', d) and therefore its restriction to G is uniformly continuous on (G, d). Thus $T: C(G') \rightarrow LUC(G)$ defined by $(Tf)(g) = f(g)$ for g in G is a positive linear isometry *onto* $LUC(G)$. Therefore $T^*: LUC(G)^* \rightarrow C(G')^*$ is an isometry. Since $\dim J_c l(G' = 1$ it will be enough to show that $T^*\varphi \in J_c l(G')$ for any $\varphi \in J_u l(G)$.

Let $l'_a: C(G') \rightarrow C(G')$ be defined by $l'_a f = f_a$ for $a \in S$ and $l^0_a: LUC(G) \rightarrow LUC(G)$ be defined by $l^0_a f = f_a$ for $a \in S$. If $g, a \in G \subset G'$ then

$$T(l'_a f)(g) = (l'_a f)(g) = f(ag) = l^0_a(Tf)(g) .$$

Thus $T(l'_a f) = l^0_a(Tf)$ if $a \in G$. Let now $\varphi \in J_u l(G)$ and $a \in G$ then for $f \in C(G')$ $(T^*\varphi)(l'_a f) = \varphi(Tl'_a f) = \varphi(l^0_a Tf) = \varphi(Tf) = T^*\varphi(f)$.

If $a \notin G$ but $a \in G'$ then there is a sequence $\{a_n\} \subset G$ such that $d(a_n a) \rightarrow 0$. Since d is an invariant metric we have that $d(a_n g, ag) = d(a_n a) \rightarrow 0$ for any g in G. But any $f \in C(G')$ is uniformly continuous, which means that for $\varepsilon > 0$ there is a $\delta > 0$ such that if $d(x, y) < \delta$ then $|f(x) - f(y)| < \varepsilon$. If therefore n_0 is such that $n \geqq n_0$ implies

$d(a_n a) < \delta$ then $|(l'_{a_n} - l'_a)f(g)| = |f(a_n g) - f(ag)| < \varepsilon$. This shows that $\|(l'_{a_n} - l'_a)f\| \to 0$ if $n \to \infty$. Thus

$$(T^*\varphi)(l'_a f) = \lim_{n\to\infty} (T^*\varphi)(l'_{a_n} f) = (T^*\varphi)f$$

since $a_n \in G$. Therefore $T^*\varphi \in J_c l(G')$. As one can easily see the condition that G' is metric is not essential and may easily be dropped. Also instead of G' being abelian we may require that G' is amenable as a discrete group and therefore we get:

COROLLARY 1. *If G_1 is a compact hausdorff topological group which is amenable as a discrete group and $G \subset G_1$ is any countable (not finite) subgroup then* $\dim J_u l(G) = 1$ *while* $\dim J_c l(G) = \infty$.

We may remark that we take G' of the preceeding example to be the closure of G in G_1 and we remember that G' as a subgroup of an amenable group is also amenable as a discrete group. (see Day [4] p. 516 (D)).

Applications: The Banach Algebra $LUC(G)^*$. Let G be a topological group and define in $LUC(G)^*$ (where $LUC(G) \subset C(G)$ are the left uniformly continuous functions with the sup. norm) the following multiplication: If $\varphi, \psi \in LUC(G)^*$ then for $f \in LUC(G)$ $[\varphi \odot \psi](f) = \varphi(y)$ where $y(h) = \psi(l^0_h f)$ for $h \in G$. (And $l^0_a: LUC(G) \to LUC(G)$ is defined by $l^0_a f = f_a$ for $a \in G$). The function y belongs to $LUC(G)$. In fact $|y(h)| \leqq \|\psi\| \|l^0_h f\| \leqq \|\psi\| \|f\|$ and so y is bounded, but moreover, $y(h)$ is left uniformly continuous. This is true since for any $\varepsilon > 0$ there is a neighborhood of the identity V such that $|f(vg) - f(g)| < \varepsilon$ for each g in G and v in V. In other words $\|l^0_v f - f\| < \varepsilon$ for each v in V. Thus

$$\begin{aligned} |y(vh) - y(h)| &= |\psi l^0_{vh} f - \psi l^0_h f| = |\psi(l^0_{vh} f - l^0_h f)| \\ &= |\psi^0_{l_h}(l^0_v f - f)| \leqq \|\psi\| \|l^0_v f - f\| < \|\psi\| \varepsilon \end{aligned}$$

for each h in G. Therefore this multiplication is at least well defined. But moreover, it renders $LUC(G)^*$ a Banach algebra as easily shown and known. In fact if $\varphi, \psi \in LUC(G)^*$ and $f \in LUC(G)$ then

$$|(\varphi \odot \psi)(f)| = |\varphi_h(\psi l^0_h f)| \leqq \|\varphi\| \|\psi l^0_h f\|$$

(where φ_h means φ with respect to the variable $h \in G$ and $\|\psi l^0_h f\| = \sup_{h\in G} |\psi l^0_h f|$). But $|\psi l^0_h f| \leqq \|\psi\| \|l^0_h f\| \leqq \|f\| \|\psi\|$. The associative law is also easily proved. In fact if $\lambda, \mu, \nu \in LUC(G)^*$ and $f \in LUC(G)$ then $[\lambda \odot (\mu \odot \nu)](f) = \lambda_a[(\mu \odot \nu) l^0_a f] = \lambda_a[\mu_b[\nu l^0_b(l^0_a f)]]$. But

$$[(\lambda \odot \mu) \odot \nu] f = (\lambda \odot \mu)_b[\nu l^0_b f] = (\lambda \odot \mu)(y)$$

where $y(g) = \nu l_g^0 f$ for each $g \in G$. But $(l_a^0 y)(g) = y(ag) = \nu(l_{ag}^0 f) = \nu(l_g^0 l_a^0 f)$. Therefore $(\lambda \odot \mu)y = \lambda_a[\mu l_a^0 y] = \lambda_a[\mu_b(\nu l_b^0 l_a^0 f)]$ which implies that $\odot$ is associative. The distributive laws are also easily proved. The following should be noted here: In $C(G)^*$ we cannot define the same multiplication as above since if $\varphi, \psi \in C(G)^*$, $f \in C(G)$ and G is not compact then $y(h) = \psi(f_h)$ is not generally a continuous function of h. In fact the following nice result has been established by Chivukula R. Rao, for groups G with an invariant metric: If $f \in C(G)$ satisfies for each $\psi \in C(G)^*$ that $\psi f_g = y(g) \in C(G)$ then f is uniformly continuous (see C. R. Rao [13] p. 17 thm 2). As an immediate consequence of our work combined with a result proved in Rudin [15], one gets the following results: (Denote by $R(G)$ the radical of the algebra $LUC(G)^*$.)

THEOREM 4. *Let G be a separable abelian locally compact hausdorff topological group. Then either $R(G) = \{0\}$ or $R(G)$ is infinite dimensional. Moreover; $R(G) = \{0\}$ (i.e. $LUC(G)^*$ is semisimple) if and only if G is compact.*

We need the following lemma whose proof is essentially known (see Civin-Yood [3], p. 849)

LEMMA. *Let G be a topological group and $J_u l(G) \subset LUC(G)^*$ be the space of left invariant elements and let*

$$J_1 = \{\varphi \in J_u l(G);\ \varphi(1_G) = 0\}\ .$$

Then J_1 is a two sided ideal and $J_1^2 = \{0\}$.

Proof. If $\mu, \nu \in J_1$ and $f \in LUC(G)$ then $y(h) = \nu l_h^0 f = \nu(f)$ for each $h \in G$ i.e., $y(h) = \nu(f) \cdot 1_G$. Therefore $\mu \odot \nu(f) = \mu(\nu(f) \cdot 1_G) = \nu(f)\mu(1_G) = 0$. This shows that $J_1^2 = \{0\}$. Let now $\varphi \in LUC(G)^*$, $\nu \in J_u l(G)$ and $f \in LUC(G)$. Then $y(h) = \nu(f) 1_G(h)$. Thus $\varphi \odot \nu(f) = \varphi(y) = \varphi(1_G) \cdot \nu(f)$. In other words

(III. 5) $\qquad \varphi \odot \nu = c \cdot \nu$ where $c = \varphi(1_G)$ is a constant .

If $\nu \in J_1 \subset J_u l(G)$ then $(\varphi \odot \psi) 1_G = c \cdot \nu(1_G) = 0$ and so $\varphi \odot \psi \in J_1$. Therefore J_1 is a left ideal. Moreover if $a \in G$ then

$$\nu \odot \varphi(l_a^0 f) = \nu_h(\varphi(l_h^0 l_a^0 f)) = \nu_h(\varphi l_{ah}^0 f)\ .$$

But if we define now $y(h) = \varphi(l_h^0 f)$ then $(l_a^0 y)(h) = y(ah) = \varphi(l_{ah}^0 f)$. Therefore $\nu_h(\varphi l_{ah}^0 f) = \nu(l_a^0 y) = \nu(y) = \nu(\varphi l_h^0 f) = \nu \odot \varphi(f)$ which proves that $\nu \odot \varphi$ is left invariant. But since $l_h^0 1_G = 1_G$ and $\varphi(l_h^0 1_G) = \varphi(1_G) = c$, we immediately get that $\nu \odot \varphi(1_G) = \nu(c 1_G) = c \cdot \nu(1_G) = 0$. Therefore

$\nu \odot \varphi \in J_1$ which finishes the proof of this lemma.

REMARK. The above lemma implies as well known that $J_1 \subset R(G)$ for any topological group G.

Proof of Theorem 4. Denote by $M_u l(G)$ the set of left invariant means of $LUC(G)^*$ and let $\varphi_0 \in M_u l(G)$ be fixed. Then obviously

$$M_u l(G) - \varphi_0 = \{\varphi - \varphi_0, \varphi \in M_u l(G)\} \subset J_1 \subset R(G)$$

since $\varphi(1_G) = 1$ for each $\varphi \in M_u l(G)$. But as pointed out in the Remark (II-2) the linear manifold spanned by $M_u l(G)$ coincides with $J_u l(G)$. Assume now that $\dim R(G) = n$ where $0 \leqq n < \infty$, then $\dim J_u l(G) = \dim M_u l(G) < \infty$. This implies by Theorem (III. 2) that G is compact. But by Rudin [15] if G is any compact abelian topological group then $C(G)^*$ with the above defined multiplication is semisimple. Since for compact G, $C(G) = LUC(G)$ we get that $R(G) = \{0\}$. Therefore either $R(G) = \{0\}$ or $\dim R(G) = \infty$. And $R(G) = \{0\}$ if and only if G is compact.

THEOREM 5. *If G is separable subgroup of a locally convex linear topological space then either $R(G) = \{0\}$ or $\dim R(G) = \infty$. Moreover $R(G) = \{0\}$ if and only if $G = \{0\}$.*

Proof. As in the previous theorem if $\dim R(G) = n$ where $0 \leqq n < \infty$ then $\dim J_u l(G) < \infty$ which implies by Theorem (III. 3) that $G = \{0\}$. But if $G = \{0\}$ then surely $R(G) = \{0\}$. Which finishes the proof of this theorem.

THEOREM 6. *Let G be a separable hausdorff topological group which is amenable as a discrete group. If G has property* (B) *then* $\dim R(G) = \infty$.

Proof. As above $M_u l(G) = \varphi_0 \subset J_1 \subset R(G)$. But by Theorem (III, 1) $\dim J_u l(G) = \infty$ and since $M_u l(G)$ spans $J_u l(G)$, $\dim M_u l(G) = \infty$ which proves this theorem.

REMARK. (a) If $LUC(G)^*$ contains two distinct left invariant means φ_1 and φ_2 then the algebra $LUC(G)^*$ is not commutative since $\varphi_1 \odot \varphi_2 = \varphi_2$ and $\varphi_2 \odot \varphi_1 = \varphi$. Therefore if G is even a commutative noncompact locally compact separable group, then $LUC(G)^*$ is not commutative.

(b) If $L_c UC(G)$ is the Banach space of bounded complex valued left uniformly continuous functions on G and the algebra $L_c UC(G)^*$ is defined as above then Theorems III-4, III-5, III-6 hold true also for $L_cUC(G)^*$. Since any $\varphi \in J_1$ can be extended to $L_cUC(G)$ by defining for $f, g \in LUC(G)$ $\varphi(f + ig) = \varphi(f) + i\varphi(g)$. If $J_1' \subset L_cUC(G)^*$ is the set of all such extensions of elements of $J_1 \subset LUC(G)^*$ then $J_1' \subset R_c(G)$ where $R_c(G)$ denotes the radical of $L_cUC(G)^*$. From here one immediately gets that Theorem 4 holds also for the complex case.

IV. The invariant mean on semigroups containing compact groups and left ideals

The main theorem of this chapter is Theorem IV-1. The following lemma is essentially known and we need it in the special form appearing here.

LEMMA 1. *Let S be a topological semigroup which contains a compact left-ideal group A_0. If $\{A_\alpha; \alpha \in I\}$ is the set of all compact left-ideal groups of S then $A = \bigcup_{\alpha \in I} A_\alpha$ is a right minimal ideal. Moreover if e_α is the identity of the group A_α then for any $a \in A$, $e_\alpha a = a$. Also for any $t \in S$, $tA_\alpha = A_\alpha$.*

REMARK. A_α as groups and left ideals are minimal left ideals and therefore are disjoint.

Proof. Let $s \in S$. Then $A_\alpha s$ is a minimal left ideal since if $L \subset A_\alpha s$ is a left ideal and $as \in L$ with $a \in A_\alpha$ then $A_\alpha s = (A_\alpha a)s \subset L$ (since A_α is a group). Thus $A_\alpha s = L$ is a minimal left ideal. But $A_\alpha sa \subset A_\alpha$ for any $a \in A_\alpha$ and therefore $A_\alpha sas \subset A_\alpha s$. Since $A_\alpha s$ is a minimal left ideal $(A_\alpha s)as = A_\alpha s$. If $t \in S$ then $tA_\alpha = t(e_\alpha A_\alpha) = (te_\alpha)A_\alpha = A_\alpha$ since $te_\alpha \in A_\alpha$, which is a group. In particular for $a \in A_\alpha$ $as(A_\alpha s) = A_\alpha s$. In other words for any $b \in A_\alpha s$, $b(A_\alpha s) = A_\alpha s = (A_\alpha s)b$ holds which proves that the semigroup $A_\alpha s$ is in fact a group. Thus $A_\alpha s$ is a left ideal and group which as a continuous immage of A_α is also a compact subset of S. Therefore $A_\alpha s = A_\beta \subset A$ for some $\beta \in I$. Thus for any $s \in S, As = \bigcup_{\alpha \in I} A_\alpha s \subset A$ which shows that A is a right ideal.

Let now R be any right ideal of S and $r \in R$. Then $A_\alpha = rA_\alpha \subset R$ for each $\alpha \in I$. This shows that $A \subset R$ (i.e. that A is included in each right ideal of S) and in particular that A is a minimal right ideal. Now if e_α, e_β are the identities of A_α, A_β respectively then $e_\alpha \cdot e_\beta \in A_\beta$ and

$$(e_\alpha e_\beta)(e_\alpha e_\beta) = e_\alpha(e_\beta(e_\alpha e_\beta)) = e_\alpha(e_\alpha e_\beta) = e_\alpha e_\beta .$$

Thus $e_\alpha e_\beta$ is an idempotent of the group A_β and therefore $e_\alpha e_\beta = e_\beta$ for any $\alpha, \beta \in I$. If now $a \in A$ then $a \in A_\beta$ for some $\beta \in I$ and therefore $e_\alpha a = e_\alpha(e_\beta a) = e_\beta a = a$.

REMARK. In semigroup terminology this shows that A is the Suschevitch kernel of the semigroup S.

If $\varphi \in C(S)^*$ then $\varphi \geqq 0$ (is positive) if $\varphi(f) \geqq 0$ for each $f \geqq 0$, $f \in C(S)$. An operator $T: C(S)^* \to C(S')^*$ is called positive if $T\varphi \geqq 0$ whenever $\varphi \geqq 0$.

LEMMA 2. *Let S be the semigroup of* Lemma (IV, 1) *and $\pi: C(S) \to C(A)$ be defined by $(\pi f)(a) = f(a)$ for a in A. Then $\pi^*: C(A)^* \to C(S)^*$ is a linear positive isometry such that $\pi^*[J_c l(A)] = J_c l(S)$. Moreover $\pi^{*^{-1}}: J_c l(S) \to J_c l(A)$ is also positive.*

Proof. π maps $C(S)$ onto $C(A)$ since if $h \in C(A)$ then define $\tilde{h} \in C(S)$ by

$$\tilde{h}(s) = h(e_\alpha s) , \tag{IV. 1}$$

where e_α is the identity of the group A_α for some fixed $\alpha \in I$. If $s \in A$ then by the proceeding lemma $e_\alpha s = s$ and so $\tilde{h}(s) = h(s)$. Also $s \to e_\alpha s$ is a continuous map from S to A (with the relative topology) since if $0'$ is open in A then $0' = 0 \cap A$ with 0 open in S and (since A is a right ideal) $\{s; e_\alpha s \in 0'\} = \{s; e_\alpha s \in 0\}$ which is open by the continuity of the left multiplication. Since $h \in C(A)$ we get that $h(e_\alpha s) = \tilde{h}(s) \in C(S)$. We also remark that if $\| h \| \leqq 1$ then $\| \tilde{h} \| \leqq 1$ and so π maps the unit ball of $C(S)$ *onto* the unit ball of $C(A)$. Also if $f \geqq 0$ then $\pi f \geqq 0$ and $\pi(1_S) = 1_A$. Therefore if $\varphi \in C(A)^*$ then

$$\begin{aligned} \| \pi^*\varphi \| &= \sup_{\|f\| \leqq 1\ f \in C(S)} |(\pi^*\varphi)f| = \sup_{\|f\| \leqq 1\ f \in C(S)} |\varphi(\pi f)| \\ &= \sup_{\|h\| \leqq 1\ h \in C(A)} |\varphi(h)| = \|\varphi\| . \end{aligned}$$

Therefore $\pi^*: C(A)^* \to C(S)^*$ is a positive linear isometry into $C(S)^*$. We shall show that it maps $J_c l(A)$ *onto* $J_c l(S)$. If $s \in S, a \in A$ then let $l'_a: C(A) \to C(A)$ and $l_s: C(S) \to C(S)$ be defined by: $l'_a h = h_a$ and $l_s f = f_s$ for and $a \in A$ and s in S. Let $\mathscr{L}_s = l_s^*$. Then $(\pi l_a f)(b) = (l_a f)(b) = f(ab) = (\pi f)(ab) = (l'_a(\pi f))(b)$ for each $a, b \in A$. Thus $\pi l_a f = l'_a \pi f$ for each $f \in C(S)$ and so for any $a \in A$ and $\varphi \in J_c l(A)$ and $f \in C(S)$:

$$(\pi^*\varphi)(l_a f) = \varphi(\pi l_a f) = \varphi(l'_a \pi f) = \varphi(\pi f) = (\pi^*\varphi)(f) .$$

Thus $\mathscr{L}_a(\pi^*\varphi) = \pi^*\varphi$ for each $a \in A$. If now $s \in S$ and $a \in A$ then

$$\mathscr{L}_s(\pi^*\varphi) = \mathscr{L}_s(\mathscr{L}_a(\pi^*\varphi)) = \mathscr{L}_{sa}(\pi^*\varphi) = \pi^*\varphi$$

since $sa \in A$. Thus $\pi^*: J_c l(A)) \to J_c l(S)$ is a linear positive isometry into. We prove now that π^* maps $J_c l(A)$ onto $J_c l(S)$.

Let $\varphi \in J_c l(S)$ and let $f \in C(S)$ satisfy $f(a) = 0$ for each $a \in A$. Then for $a \in A$ we have $(l_a f)(s) = f(as) = 0$, since A is a right ideal.

Thus $\varphi(f) = \varphi(l_a f) = \varphi(0) = 0$. Therefore if $f_1, f_2 \in C(S)$ satisfy $f_1(a) = f_2(a)$ for each $a \in A$ then $f_1(a) - f_2(a) = 0$ for $a \in A$ and so $\varphi(f_1) = \varphi(f_2)$. In other words if $h \in C(A)$ and $\tilde{h} \in C(S)$ is *any* extension of h to all of S then $\varphi(h)$ does not depend on the particular extension $\tilde{h} \in C(S)$ of $h \in C(A)$. Therefore $\varphi' \in C(A)^*$ defined for $h \in C(S)$ by

$$\varphi'(h) = \varphi(\tilde{h})$$

where $\tilde{h} \in C(S)$ is any extension of h to all of S, is at least well defined. Moreover if $\varphi \geqq 0$ and $h \geqq 0$ then the extension $\tilde{h} \in C(S)$ defined above (IV. 1) satisfies $\tilde{h}(s) = h(e_\alpha s) \geqq 0$ and so $\varphi'(h) = \varphi(h) \geqq 0$. This shows that if $\varphi \geqq 0$ then $\varphi' \geqq 0$. It is easily checked that φ' is linear. Also if $\| h \| \leqq 1$ then the extension defined by IV-1 satisfies $\| \tilde{h} \| \leqq 1$ and thus

$$| \varphi'(h) | = | \varphi(\tilde{h}) | \leqq \| \varphi \| \, \| \tilde{h} \| \leqq \| \varphi \| .$$

This shows that $\varphi' \in C(A)^*$.

We show now that $\varphi' \in J_c l(A)$. Let $a \in A$ be fixed. Then

$$\varphi'(l'_a h) = \varphi(\widetilde{l'_a h})$$

where $\widetilde{l'_a h}$ is *any* extension, in $C(S)$, of $l'_a h \in C(A)$. But $a \in A_{\alpha_0}$ for some $\alpha_0 \in I$ and the function defined by

$$\text{(IV. 2)} \qquad (\widetilde{l'_a h})(s) = (l'_a h)(e_{\alpha_0} s)$$

is a bounded continuous extension of $l'_a h \in C(A)$. (where e_{α_0} is the identity of A_{α_0}). And for each $s \in S$:

$$\widetilde{l'_a h}(s) = (l'_a h)(e_{\alpha_0} s) = h(a e_{\alpha_0} s) = h(as) .$$

But if $\tilde{h} \in C(S)$ is any extension of h to all of S then, since A is a right ideal, we get

$$h(as) = \tilde{h}(as) = (l_a \tilde{h})(s) .$$

Therefore

$$\widetilde{l'_a h}(s) = (l_a \tilde{h})(s)$$

where $\widetilde{l'_a h}$ is the extension defined by (IV. 2) while $\tilde{h} \in C(S)$ is any extension of h. Therefore

$$\varphi'(l'_a h) = \varphi(\widetilde{l'_a h}) = \varphi(l_a \tilde{h}) = \varphi(\tilde{h}) = \varphi'(h) .$$

This shows that $\varphi' \in J_c l(A)$. Moreover $\pi^* \varphi' = \varphi$. In fact if $f \in C(S)$ then $(\pi^* \varphi')(f) = \varphi'(\pi f) = \varphi(f)$ since f is obviously an extension of $\pi f \in C(A)$. Therefore $\pi^*: J_c l(A) \to J_c l(S)$ is a positive linear isometry onto and positive elements in $Jl(S)$ have positive preimages in $J_c l(A)$

or in other words $\pi^{*^{-1}}: J_c l(S) \to J_c l(A)$ is a linear positive isometry onto.

REMARK. We notice that we do not assume any separation axioms about the topological space A. We shall show in what follows that in fact we can assume about A that it is even a hausdorff space (and even that $C(A)$ separates points).

In fact define in A the following equivalence relation: If $a, b \in A$ then $a \sim b$ if and only if $x(a) = x(b)$ for each $x \in C(A)$. Obviously this is an equivalence relation but moreover $\sim$ is even a congruence, i.e., if $a \sim b$ then $ca \sim cb$ and $ac \sim bc$ for each $c \in A$. This is true since for any $x \in C(A)$

$$x(ca) = x_c(a) = x_c(b) = x(cb)$$

and

$$x(ac) = x^c(a) = x^c(b) = x(bc) .$$

Let A' be the collection of all equivalence classes of A and for each $a \in A$ let a' be the equivalence class containing a. Define in A' the multiplication $a' \cdot b' = (ab)'$. Since $\sim$ is a congruence this multiplication is well defined and renders A' a semigroup. (see Lyapin [20] p. 361–362). Thus $\psi: A \to A'$ defined by $\psi(a) = a'$ is a homomorphism of A onto A'. Define now in A' the quotient topology this way: $U' \subset A'$ is open if and only if $\psi^{-1}(U') \subset A$ is open. Thus $\psi: A \to A'$ is a continuous homomorphism and so $A'_\alpha = \psi(A_\alpha)$ are compact. Moreover if $a \in A_\alpha$ then $A'_\alpha = \psi(A_\alpha) = \psi(aA_\alpha) = \psi(a)\psi(A_\alpha) = a'A'_\alpha$ and in the same way $A'_\alpha a' = A'_\alpha$ which shows that A'_α is a group. Also if $b \in A$ then $b'A'_\alpha = \psi(b)\psi(A_\alpha) = \psi(bA_\alpha) = \psi(A_\alpha) = A'_\alpha$ which shows that A'_α is a left ideal.

But moreover, A' with the above defined quotient topology has separately continuous multiplication. In fact if U' is an open set in A' and $a_0 \in A$ then we have to show that $0' = \{c'; a'_0 c' \in U'\}$ is open in A' or that

$$0 = \psi^{-1}(0') = \{c; (a_0 c)' \in U'\} = \{c; a_0 c \in \psi^{-1}(U')\}$$

is open in A. But since ψ is continuous $\psi^{-1}(U)$ is open in A and since left multiplication by a_0 is continuous, we get that $\{c; a_0 c \in \psi^{-1}(U')\}$ is open in A. In the same way one shows that right multiplication in A' is continuous. Define now the map $\tilde{\psi}: C(A') \to C(A)$ by $(\tilde{\psi}x')(a) = x'(\psi a) = x'(a')$ for each $a \in A$. Since $\psi(A) = A'$, $\tilde{\psi}$ is a linear positive isometry (i.e., if $x' \geqq 0$ then $\tilde{\psi}(x') \geqq 0$) into $C(A)$. But we notice now that each $x \in C(A)$ gives raise to an $x' \in C(A')$ by defining: $x'(a') = x(a)$ where a is any representative of the equivalence class $a' \in A'$. Since x

is constant on equivalence classes, x' is well defined and $x' \in C(A')$, since if V is an open set of reals then

$$\psi^{-1}\{a'; x'(a') \in V\} = \{a; x(a) = x'(a') \in V\} = x^{-1}(V)$$

which is open in A since $x \in C(A)$. Also $(\tilde{\psi}x')(a) = x'(\psi(a)) = x'(a') = x(a)$. This shows that $\tilde{\psi}\colon C(A') \to C(A)$ is onto. It is immediate now that A' is hausdorff. In fact if $a', b' \in A'$ are such that $a' \neq b'$ then there is an $x \in C(A)$ such that $x(a) \neq x(b)$ i.e. $x'(a') \neq x'(b')$ so that $C(A')$ even separates points.

LEMMA 3. *$\tilde{\psi}^*\colon C(A)^* \to C(A')^*$ is a linear positive isometry such that $\tilde{\psi}^*[J_c l(A)] = J_c l(A')$. $\tilde{\psi}^{*-1}\colon J_c l(A') \to J_c l(A)$ is also positive.*

Proof. Since $\tilde{\psi}\colon C(A) \to C(A')$ is a positive isometry onto we immediately get that $\psi^*\colon C(A)^* \to C(A')^*$ is a linear positive isometry. Let now $l'_{a'}\colon C(A') \to C(A')$ be given by $(l'_{a'}x')(c') = x'(a'c')$ for each $c' \in A'$ and $l_a\colon C(A) \to C(A)$ by $l_a x = x_a$, As known and easily checked $l_a(\tilde{\psi}x') = \tilde{\psi}(l'_{a'}x')$, which shows that if $\varphi \in J_c l(A)$ then:

$$(\tilde{\psi}^*\varphi)(l'_{a'}x') = \varphi\tilde{\psi}(l'_{a'}x') = \varphi(l_a\tilde{\psi}x') = \varphi(\tilde{\psi}x') = (\tilde{\psi}^*\varphi)(x') .$$

Therefore $\tilde{\psi}^*[J_c l(A)] \subset J_c l(A')]$.

If now $\varphi' \in J_c l(A')$ then let $\varphi \in C(A)^*$ be defined, for $x \in C(A)$, by $\varphi(x) = \varphi'(x')$ where $x' \in C(A')$ is given by $x'(a') = x(a)$ for each $a \in A$. Then $(l_a x)'(b') = (l_a x)(b) = x(ab) = x'((ab)') = x'(a'b') = (l'_{a'}x')(b')$. Thus

$$\varphi(l_a x) = \varphi'((l_a x)') = \varphi'(l'_{a'}x') = \varphi'(x') = \varphi(x) .$$

Therefore $\varphi \in J_c l(A)$. But $(\tilde{\psi}x')(a) = x'(a') = x(a)$ and thus

$$(\tilde{\psi}^*\varphi)(x') = \varphi(\tilde{\psi}x') = \varphi(x) = \varphi'(x') .$$

This shows that $\tilde{\psi}^*\varphi = \varphi'$ i.e., that $\tilde{\psi}^*[J_c l(A)] = J_c l(A')$. We also notice that if $\varphi' \geqq 0$ then $\varphi \geqq 0$ and so positive elements in $Jl(A')$ have positive preimages and so ψ^{*-1} is also positive.

REMARKS. We notice that A'_α is a group which is a compact hausdorff topological space with separately continuous multiplication and therefore by Ellis theorem (see Ellis [21] or Glicksberg Deleeuw [19] p. 64–65 and p. 94–96) each A'_α is a compact Hausdorff *topological group*. (i.e. the mapping $(a, b) \to ab^{-1}$ from $A'_\alpha \times A'_\alpha$ into A'_α is continuous).

THEOREM 1. *Let S be a topological semigroup (only with separately continuous multiplication) and let S contain exactly n compact left-ideal groups $A_1, \cdots, A_n$. Then* $\dim J_c l(S) = n$ *and* $J_c l(S)$ *is* *spanned by the left invariant means.*

Proof. If $A = \bigcup_{i=1}^n A_i$ then $\pi^*: J_c l(A) \to J_c l(S)$ is a positive isometry onto (and so maps left invariant means into left invariant means). If A' is the semigroup of Lemma (IV. 3) then $\tilde{\psi}^{*-1}: J_c l(A') \to J_c l(A)$ is a linear positive isometry onto and so it is enough to show that $\dim J_c l(A') = n$ and that $J_c l(A')$ is spanned by the set of left invariant means. We recall now that $A' = \bigcup_{i=1}^n A'_i$ is a Hausdorff topological space and that A'_i are compact topological groups and left ideals and therefore disjoint. Thus A' is a compact hausdorff semigroup and multiplication is (at least) separately continuous. In what follows we shall drop the prime and write A, A_i instead of $A', A'_i \cdot A, A_i$ are compact hausdorff. But A_i as the complement of the compact set $\bigcup_{j \neq i} A_j$, is also open. Therefore $1_{A_i} \in C(A)$. Hence if $f \in C(A)$ then $f(a) = \sum f(a) 1_{A_i}(a)$ for each $a \in A$ and $f \cdot 1_{A_i} \in C(A)$. Moreover if $h \in C(A_i)$ then $\tilde{h}$ defined by $\tilde{h}(a) = h(ae_i)$ for each $a \in A$ is an extension of h to all of A and $\tilde{h} \in C(A)$. Furthermore, if $h \geqq 0$ then $\tilde{h} \geqq 0$ and if $\| h \| \leqq 1$ then $\| \tilde{h} \| \leqq 1$. Let $\pi_i: C(A) \to C(A_i)$ be defined by $(\pi_i f)(a) = f(a)$ for $a \in A_i$. If $a \in A_i$ then let $l^i_a: C(A_i) \to C(A_i)$ be defined by $l^i_a h = h_a$ for $a \in A_i$. Also, $l_a: C(A) \to C(A)$ is defined by $l_a f = f_a$ for any a in A. Let $l^*_a = \mathscr{L}_a$. Then as easily checked: $\pi_i l_a f = l^i_a \pi_i f$ for each $f \in C(A)$ and a in A_i.

Let now $\varphi'_i \in C(A_i)^*$ be the linear positive functional of norm one represented by the normalized Haar measure on the compact hausdorff topological group A_i. Define $\varphi_i \in C(A)^*$ by

(IV. 3) $$\varphi_i(f) = \varphi'_i(\pi_i f) \qquad \text{for each } f \in C(A) .$$

Then we get immediately that $\varphi_i \geqq 0$, $\varphi_i(1_A) = 1$ and that $\varphi_i(1_{A_i}) = 1$ while $\varphi_i(1_{A_k}) = 0$ if $i \neq k$. Thus for any $a \in A_i$:

$$\varphi_i(l_a f) = \varphi'_i(\pi_i l_a f) = \varphi'_i(l^i_a \pi_i f) = \varphi'_i(\pi_i f) = \varphi_i(f) .$$

Therefore $\mathscr{L}_a \varphi_i = \varphi_i$ for each $a \in A_i$. If now $c \in A$ and $a \in A_i$ then $\mathscr{L}_c \varphi_i = \mathscr{L}_c \mathscr{L}_a \varphi_i = \mathscr{L}_{ca} \varphi_i = \varphi_i$ since $ca \in A_i$. Therefore φ_i is a left invariant mean in $C(A)^*$. Also $\varphi_1, \cdots, \varphi_n$ are linearly independent (since if $\sum \alpha_i \varphi_i = 0$ then $\alpha_k = (\sum \alpha_j \varphi_j)(1_{A_k}) = 0$). It remains to show that $\varphi_1, \cdots, \varphi_n$ span $J_c l(A)$.

If $h \in C(A_i)$ and if $\tilde{h} \in C(A)$ is any extension of h (for instance $h(c) = h(ce_i)$ for each $c \in A$) then let $P_i(h) = \tilde{h} \cdot 1_{A_i} \in C(A)$. In other words $P_i h \in C(A)$ equals h on A_i and 0 outside A_i. Thus $P_i: C(A_i) \to C(A)$ and as easily checked:

(IV. 4) $$\pi_i P_i h = h \qquad \text{for } h \in C(A_i)$$

and

(IV. 5) $$P_i\pi_i f = f1_{A_i} \qquad \text{for } f \in C(A) .$$

If $a \in A_i$ and $h \in C(A_i)$ then

$$P_i(l_a^i h)(b) = \begin{cases} 0 & \text{if } b \notin A_i \\ (l_a^i h)(b) = h(ab) & \text{if } b \in A_i \end{cases} .$$

Moreover,

$$l_a(P_i h)(b) = (P_i h)(ab) = \begin{cases} 0 & \text{if } ab \notin A_i \\ h(ab) & \text{if } ab \in A_i \end{cases} .$$

But $ab \in A_i$ if and only if $b \in A_i$ (if $b \notin A_i$ then $b \in A_j$ for $j \neq i$ and so $ab \in A_j$) and $ab \notin A_i$ if and only if $b \notin A_i$. This shows that

(IV. 6) $$P_i(l_a^i h) = l_a(P_i h) \qquad \text{for each } h \in C(A_i) \text{ and } a \in A_i .$$

Let $\varphi \in J_c l(A) \subset C(A)^*$ and define $\psi_i \in C(A_i)^*$ by

(IV. 7) $$\psi_i(h) = \varphi(P_i h) .$$

If $a \in A_i$ then by IV-6, IV-7: $\psi_i(l_a^i h) = \varphi(P_i l_a^i h) = \varphi(l_a P_i h) = \varphi(P_i h) = \psi_i(h)$ which shows that ψ_i is a left invariant functional in $C(A_i)$. Therefore, (by the uniqueness of the Haar measure) we get that $\psi_i = \alpha_i \varphi_i'$ for some real number α_i. Therefore if $f \in C(A)$ then using IV-5, IV-7 and IV-3 one gets:

$$\begin{aligned} \varphi(f) &= \sum \varphi(f1_{Ai}) = \sum \varphi(P_i\pi_i f) \\ &= \sum \psi_i(\pi_i f) = \sum \alpha_i \varphi_i'(\pi_i f) = \sum \alpha_i \varphi_i(f) . \end{aligned}$$

Thus $\varphi = \sum \alpha_i \varphi_i$ which finishes the proof. As a special case one gets the following theorem of I. S. Luthar (see [12] p. 403).

THEOREM. *If S is an abelian topological semigroup which contains a compact ideal then* $\dim J_c l(S) = 1$.

Proof. As in Luthar's proof if I is a compact ideal of S and $I_1, \cdots, I_n$ are closed ideals of S contained in I then $I_1 \cdots I_n \subset \bigcap_{j=1}^n I_j \neq \emptyset$. Therefore the family F of all closed ideals of S contained in I has the finite intersection peoperty and so $A = \bigcap_{I' \in F} I' \neq \emptyset$. Thus A is a compact ideal. If $a \in A$ then $aA \subset A$ is a compact ideal and so $aA = A$ which shows that A is a group. If now A_1 is any other compact ideal and group of S then $A_1 A \subset A \cap A_1 \neq \emptyset$ and if $a \in A \cap A_1$ then $A = Aa = A_1 a = A_1$ which shows that S contains exactly one ideal and compact group. Using Theorem (IV. 1) we get that $\dim J_c l(S) = 1$ or that $C(S)$ admits a unique invariant mean.

REFERENCES

1. A. D. Alexandroff, *Additive set functions in abstract spaces* (Russian), Math. Sbornic **8** (50) 1940, 307-348; **9** (51) 1941, 563-628; **13** (55) (1943), 169-238.

2. G. Birkoff, Compositio Math. (3) (1936), 424-430.

3. P. Civin-B. Yood, *The second conjugate space of a Banach Algebra as an algebra,* Pacific J. Math, **11** (1961), 847-870.

4. M. M. Day, *Amenable semigroups,* Illinois J. Math. **1** (1957), 509-544.

5. ———, *Normed linear spaces,* Berlin, Springer, 1958.

6. E. Granirer, *On amenable semigroups with a finite dimensional set of invariant means I, II,* Illinois J. Math. **7** (1963), 32-48 and 49-58.

7. ———, *A theorem on amenable semigroups,* Trans. Amer. Math. Soc. (to appear)

8. P. R. Halmos, *Measure theory,* New York, Van Nostrand (1959).

9. S. Kakutani, *Über die Metrisation der topologixhen Gruppen Proc,* Imp. Acad. Jap. **12** (1936), 82-84.

10. L. Kelley, *General topology,* New York, Van Nostrand, 1955.

11. Indar S. Luthar, *Uniqueness of the invariant mean on an abelian semigroup,* Illinois J. Math. **3** (1959), 28-44.

12. ———, *Uniqueness of the invariant mean on an abelian topological semigroup,* Trans Amer. Math. Soc. (1962), 403-411.

13. Chivulkula R. Rao, *In variant means on spaces of continuous or measuarable functions,* thesis submitted to the University of Illinois, Urbana, Illinois.

14. W. G. Rosen, *On invariant means over compact semigroups* Proc. Amer. Math. Soc. **7** (1956), 1076-1082.

15. W. Rudin, *Fourier Analysis on groups,* Interscience Publishers, 1962.

16. R. J. Silverman, *Invariant linear functions,* Trans Amer. Math. Soc. **81** (1956), 411-424.

17. V. S. Varadarajan, *Measures on topological spaces* (Russian) Math. Sbornic, **55** (97) No. 1 (1961), 35-100.

18. A. Weil, *Sur les espaces a structure uniforme at sur la topologic generale,* Paris, Herman, Cie 1938.

19. K. Deleew, I. Glicksberg, *Applications of almost periodic compactifications,* Acta. Math. **105** (1961), 63-97.

20. E. Lyapin, *Semigroups* (Russian), Moskow, 1960.

21. R. Ellis, *Locally compact transformation groups,* Duke Math. J. **24** (1957), 119-120.

22. E. Hewitt, *On two problems of Urisohn,* Ann. of Math. (2) **47** (1946), 503-509.

23. W. W. Comfort and K. A. Ross, *Pseudocompactness and uniform continuity in topological groups* (to appear in the Pacific J. Math.)

THE UNIVERSITY OF ILLINOIS
URBANA, ILLINOIS

PACIFIC JOURNAL OF MATHEMATICS
Vol. 15, No. 1, 1965

CLOSED VECTOR FIELDS

N. HICKS

We study closed vector fields on a semi-Riemannian manifold. In particular, we study the differential geometry of the submanifolds determined by a nonvanishing closed field. Expressions are computed for the Weingarten map, the mean curvature, the Riemannian curvature, and the Laplacian of the square of the length of the field. Thus we obtain a necessary and sufficient condition that the constant hypersurface of a nontrivial harmonic function be a minimal surface. We obtain conditions that imply the classical Codazzi-Mainardi equations hold. We obtain conditions that imply the existence of a representation of the manifold as a cross product in which one factor is a real line. Finally, various special cases are examined.

1. Notation. Let M be a connected C^∞ semi-Riemannian manifold with metric tensor $\langle\ ,\ \rangle$ and Riemannian connexion D [see Helgason 4 or Hicks 7 for definitions]. We summarize the properties of D and some associated concepts we shall use. The operator D assigns to each pair of C^∞ vector fields X and Y on an open set U of M, a C^∞ vector field $D_X Y$ called the covariant derivative of Y in the direction X. If X, Y, and Z are C^∞ fields on U and f a C^∞ function (real valued) on U then we have the following relations between vector fields on U:

$$\begin{aligned} D_X(Y+Z) &= D_X Y + D_X Z \\ D_{(X+Y)}Z &= D_X Z + D_Y Z \\ D_{fX} Y &= f D_X Y \\ D_X(fY) &= (Xf)Y + f D_X Y \\ \operatorname{Tor}(X, Y) &= D_X Y - D_Y X - [X, Y] \\ R(X, Y)Z &= D_X D_Y Z - D_Y D_X Z - D_{[X,Y]}Z\,. \end{aligned}$$

We call Tor the torsion on D and R the curvature of D. Since D is Riemannian, Tor $= 0$, and D is compatible with the metric tensor, thus

$$\begin{aligned} D_X Y - D_Y X &= [X, Y] \\ X\langle Y, Z\rangle &= \langle D_X Y, Z\rangle + \langle Y, D_X Z\rangle\,. \end{aligned}$$

We extend the operator D_X, as usual, to be a complete derivation on the tensor algebra over M. If $T^{r,s}$ denotes the set of r-contravariant and s-covariant tensors on M, then D_X: $T^{r,s} \to T^{r,s}$. If $f \in T^{0,0}$, then $D_X f = Xf$. If $Y \in T^{1,0}$, then $D_X Y$ is given by the connexion. If

Received January 16, 1964. This work was supported in part by NSF Grants G 23842 and GP 88.

$w \in T^{0,1}$, then $(D_X w)(Y) = X(w(Y)) - w(D_X Y)$. The last equality contains the seeds of what is meant by a complete derivation which we explain. Having defined D_X on functions, fields, and 1-forms if $\phi \in T^{r,s}$, $w_i \in T^{0,1}$ for $i = 1, \cdots, r$, and $Y_j \in T^{1,0}$ for $j = 1, \cdots, s$, then

$$\begin{aligned} X\phi(w_1, \cdots, w_r, Y_1, \cdots, Y_s) &= (D_X\phi)(w_1, \cdots, w_r, Y_1, \cdots, Y_s) \\ &+ \sum_i \phi(w_1, \cdots, w_{i-1}, D_X w_i, w_{i+1}, \cdots, w_r, Y_1, \cdots, Y_s) \\ &+ \sum_j \phi(w_1, \cdots, w_r, Y_1, \cdots, Y_{j-1}, D_X Y_j, Y_{j+1}, \cdots, Y_s) , \end{aligned}$$

where all terms are well-defined except the first term on the right side of the equation.

The symbol Δ will denote the general covariant differentiation operator $\Delta: T^{r,s} \to T^{r,s+1}$ which is induced by D. Using the above notation, $(\Delta\phi)(w_1, \cdots, w_r, Y_1, \cdots, Y_r, X) = (D_X\phi)(w_1, \cdots, w_r, Y_1, \cdots, Y_s)$.

Our study will concern *linear transformation valued tensors* on M (tensor fields of type 1, 1). For completeness, we define a linear transformation valued tensor A on an open set U of M to be a mapping that assigns to each point m in U, a linear transformation $A_m: M_m \to M_m$, where M_m is the tangent space at m. We say A is C^∞ if it maps C^∞ fields on U into C^∞ fields; then if X is a C^∞ field on U then the field $(A(X))_m = A_m(X_m)$ is C^∞ on U. We define the vector valued 2-form Tor_A by

$$\mathrm{Tor}_A(X, Y) = D_X A(Y) - D_Y A(X) - A[X, Y]$$

and let $\mathrm{tr}\, A$ and $\det A$ denote the trace and determinant functions on A, respectively.

We will use G to denote the nonsingular linear transformation induced by the metric tensor that maps M_m onto M_m^* for each m. Thus if X is in M_m then $G(X)(Y) = \langle X, Y\rangle$ for Y in M_m; or $G(X) = C_X\langle\ ,\ \rangle = \langle X,\ \rangle$ where C_X is contraction by X in the first covariant slot. We also use the symbol G for the inverse of G. Thus we think of G as a "switch map" and let the argument it is applied to tell us which map is being used. A vecter field X will be called *closed* (or *exact*) if $G(X)$ is closed (or exact), and X is *geodesic* if $D_X X = 0$. A vector X is *nonsingular* (not light-like) if $\langle X, X\rangle \neq 0$. If $\theta \in T^{r,s}$ with $r > 0$, then the *divergence of* θ is the tensor $\operatorname{div}\theta \in T^{r-1,s}$ defined by $\operatorname{div}\theta = \mathrm{tr}\,\Delta\theta$, where the trace is taken on the last covariant slot and last contravariant slot. If $Z_1, \cdots, Z_n$ is a base field of independent C^∞ vector fields on an open set U in M and $z_1, \cdots, z_n$ is the dual base of 1-forms, then

$$\begin{aligned} &(\operatorname{div}\theta)(w_1, \cdots, w_{r-1}, Y_1, \cdots, Y_s) \\ &\quad = \sum_{j=1}^{n} (\Delta\theta)(w_1, \cdots, w_{r-1}, z_j, Y_1, \cdots, Y_s, Z_j) . \end{aligned}$$

If $f \in T^{0,0}$, then the *gradient of f*, grad f, is the vector field $G(df)$, so $\langle \text{grad } f, X\rangle = Xf$, and the *Laplacian of f*, $\Delta_2 f$, is the function div (grad f). A function f is *harmonic* if $\Delta_2 f = 0$, and a field T is *conservative* if div $T = 0$.

2. Operators associated with a vector field. Let T be a C^∞ vector field on M. On each tangent space M_m, we define linear maps A_T, B_T, and C_T by

$$A_T(X) = D_X T,\ B_T(X) = D_X(D_T T)\ , \quad \text{and} \quad C_T(X) = R(X, T)T\ .$$

These maps are C^∞ since D and T are C^∞. Let U be the open set of points in M where $\langle T, T\rangle$ does not vanish. On U, we define the $C^\infty(n-1)$ dimensional distribution $T^\perp$ by

$$(T^\perp)_p = [X \in M_p : \langle X, T\rangle = 0]\ .$$

From the definition of the curvature R we have

$$C_T = B_T - A_T^2 + [A_T, D_T]$$

where

$$[A_T, D_T](X) = A_T(D_T X) - D_T(A_T X)$$

and thus $[A_T, D_T]$ is a linear transformation valued tensor. By the standard symmetry properties of the four covariant Riemann Christoffel tensor, the map C_T is symmetric (self-adjoint), and we call it the *Ricci map associated with T*. The trace of C_T is the *Ricci curvature of T*, which we denote by Ric (T, T).

Following Bochner [1], we say a field T is *restrained* if $\Delta_2\langle T, T\rangle < 0$ at some point or T has constant length. Bochner has shown that every field on a compact manifold is restrained, and in the noncompact case, a field is restrained if its length attains a relative maximum at some point.

Our main interests in this study are the cases when A_T is symmetric, or equivalently, T is closed. Since the gradient of any C^∞ function is a closed field, many closed fields exist.

PROPOSITION 1. For any field T, tr A_T = div T and tr $[A_T, D_T] = -T(\text{div } T)$. If $T = \text{grad} f$, then the Laplacian of f is the trace of A_T.

Proof. Let $Z_1, \cdots, Z_n$ be a set of nonsingular orthonormal vector fields belonging to a Riemannian normal coordinate system at a point m in M and let $w_1, \cdots, w_n$ be the dual 1-forms of this base. Thus if $e_i = \langle Z_i, Z_i\rangle$, then

$$\text{tr } A_T = \Sigma e_i \langle D_{Z_i} T, Z_i\rangle = \Sigma w_i(D_{Z_i} T) = \text{tr } \Delta(T)\ ,$$

and using the fact that $D_T Z_i = 0$ at m for any T,

$$\begin{aligned}\operatorname{tr}[A_T, D_T] &= \Sigma\langle A_T D_T Z_i - D_T A_T Z_i, Z_i\rangle e_i \\ &= \Sigma\langle D_{D_T Z_i} T - D_T D_{Z_i} T, Z_i\rangle e_i \\ &= -\Sigma T\langle D_{Z_i} T, Z_i\rangle e_i + \Sigma\langle D_{Z_i} T, D_T Z_i\rangle e_i \\ &= -T(\operatorname{tr} A_T)\,.\end{aligned}$$

PROPOSITION 2. For any field T,

$$\operatorname{Ric}(T, T) = \operatorname{tr} C_T = \operatorname{tr} B_T - \operatorname{tr} A_T^2 - T(\operatorname{div} T)\,.$$

Proof. Using the fields Z_i in the above proof,

$$\operatorname{tr} C_T = \Sigma\langle R(Z_i, T)T, Z_i\rangle e_i = \operatorname{Ric}(T, T)\,,$$

and the rest of the proposition follows from the linearity of the trace.

PROPOSITION 3. For any field T, T has constant length if and only if (Image $A_T) \subset T^{\perp}$.

Proof. For any vector X,

$$X\langle T, T\rangle = 2\langle D_X T, T\rangle = 2\langle A_T(X), T\rangle\,.$$

3. The symmetric case. Throughout this section we assume T is a closed field, or equivalently, A_T is symmetric (by the following proposition).

THEOREM 1. *A field T is closed if and only if A_T is symmetric. If T is closed, then $T^{\perp}$ is integrable on U.*

Proof. If X and Y are fields, then

$$\begin{aligned}(dG(T))(X, Y) &= X\langle T, Y\rangle - Y\langle T, X\rangle - \langle T, [X, Y]\rangle \\ &= \langle D_X T, Y\rangle - \langle D_Y T, X\rangle + \langle T, D_X Y - D_Y X - [X, Y]\rangle \\ &= \langle A_T X, Y\rangle - \langle A_T Y, X\rangle\,,\end{aligned}$$

since the torsion of D is zero.

If X and Y belong to $T^{\perp}$, then

$$\begin{aligned}\langle [X, Y], T\rangle &= \langle D_X Y - D_Y X, T\rangle \\ &= X\langle Y, T\rangle - \langle Y, D_X T\rangle - Y\langle X, T\rangle + \langle X, D_Y T\rangle \\ &= \langle X, A_T Y\rangle - \langle Y, A_T X\rangle = 0\end{aligned}$$

since $\langle Y, T\rangle \equiv \langle X, T\rangle \equiv 0$. Thus $T^{\perp}$ is involutive or integrable (see Chevalley [2]).

In the special case $T = \text{grad} f$, then the integral manifolds of $T^\perp$ on U are precisely the hypersurfaces on which f is constant. We next investigate the geometry of an integral manifold M' of $T^\perp$ through a point m in U. Since T is normal to M', we use T to frame M' locally (see Hicks [6]). Let e be the function on U which is plus or minus one according as $\langle T, T\rangle$ is positive or negative, respectively.

THEOREM 2. *Let L be the Weingarten map on M' and take X in $(M')_m$.*

$$L(X) = [e\langle T, T\rangle]^{-3/2}[e\langle T, T\rangle A_T(X) - e\langle T, A_T(X)\rangle T]$$

and the mean curvature H of M' is given by

$$H = \text{tr}\, L = |T|^{-1}[\text{div}\, T - T \log |T|]$$

where $|T| = [e\langle T, T\rangle]^{1/2}$ is the length of T. Thus M' is minimal if and only if $\text{div}\, T = T \log |T|$.

Proof. Let $N = [e\langle T, T\rangle]^{-1/2}\, T$ be the unit normal so

$$L(X) = D_X N = -[e\langle T, T\rangle]^{-3/2} e\langle A_T X, T\rangle T + [e\langle T, T\rangle]^{-1/2} A_T X\,.$$

To compute $\text{tr}\, L$, let $Z_1, \cdots, Z_{n-1}$ be a nonsingular orthonormal base of $(M')_m$ and let $Z_n = N$. Letting $e_i = \langle Z_i, Z_i\rangle$, then

$$\begin{aligned} H = \text{tr}\, L &= \sum_{j=1}^{n-1} \langle LZ_j, Z_j\rangle e_j \\ &= [e\langle T, T\rangle]^{-1/2} \sum_{1}^{n-1} \langle A_T Z_j, Z_j\rangle e_j\,. \end{aligned}$$

But

$$\begin{aligned} \langle A_T Z_n, Z_n\rangle e_n &= \langle D_N T, N\rangle e = \langle D_T T, T\rangle / \langle T, T\rangle \\ &= (1/2)(T\langle T, T\rangle)/\langle T, T\rangle = (1/2) T \log e\langle T, T\rangle\,. \end{aligned}$$

Hence, $H = (e\langle T, T\rangle)^{-1/2}[\text{tr}\, A_T - T \log |T|]$.

COROLLARY 1. *The constant hypersurfaces of a nonconstant harmonic function are minimal surfaces if and only if the gradient of the function has constant length along its integral curves.*

Proof. Let f be harmonic and $T = \text{grad} f$. Then T is closed and $\text{tr}\, A_T = \text{div}\, T = 0$. Hence $H = 0$ if and only if $\langle D_T T, T\rangle = 0$ or $T\langle T, T\rangle = 0$.

COROLLARY 2. *Let T be a unit field on M which is closed. Then* *the total curvature and mean curvature of the integral manifolds of* $T^{\perp}$ *are given by* $K = \det A_T$ *and* $H = \operatorname{div} T$. *Indeed,* $S = A_T$ *if and only if* T *is a unit field.*

The first corollary above suggests the definition of a *minimal harmonic function* as a harmonic function whose constant hypersurfaces are minimal surfaces. This class of harmonic functions has not been examined as yet, as far as we know, nor has the above result (Corollary 1) been proven before.

PROPOSITION 4. Let $\phi = \langle T, T\rangle$. Then $\operatorname{grad}\phi = 2D_T T$, which implies B_T is symmetric, and

$$\Delta_2\phi = 2\operatorname{tr} B_T = 2[\operatorname{Ric}(T, T) + \operatorname{tr} A_T^2 + T(\operatorname{div} T)]$$

while

$$(\Delta^2\phi)(Z, Y) = 2\langle B_T Z, Y\rangle .$$

Proof. Consider

$$(\Delta\phi)X = X\langle T, T\rangle = 2\langle D_X T, T\rangle = 2\langle X, D_T T\rangle .$$

Hence $\operatorname{grad}\phi = 2D_T T$, and $\Delta_2\phi = \operatorname{div}\operatorname{grad}\phi = 2\operatorname{tr} B_T$. The last expression for the Laplacian of ϕ follows from Proposition 2.

Finally,

$$(\Delta^2\phi)(Z, Y) = [D_Y(\Delta\phi)]Z = 2Y\langle Z, D_T T\rangle - 2\langle D_Y Z, D_T T\rangle = 2\langle Z, B_T Y\rangle .$$

We have immediately a slight generalization of a result of Bochner [1].

COROLLARY 1. *Let* T *be a closed field such that* $\operatorname{div} T$ *is constant along the integral curves of* T. *If* T *is restrained, then* $\operatorname{Ric}(T, T) < 0$ *at some point of* M *or* $\operatorname{Ric}(T, T) \leqq 0$ *on all of* M. *On a compact manifold whose Ricci curvature is always positive there can be no nontrivial closed field* T *with* $T(\operatorname{div} T) = 0$. *On a compact manifold whose Ricci curvature is nonnegative any nontrivial closed field* T *with* $T(\operatorname{div} T) = 0$ *must be a global parallel field with constant length, zero Ricci curvature, and* $A_T = 0$ (see Proposition 6).

Proof. In these cases,

$$\operatorname{Ric}(T, T) = (1/2)\Delta_2\phi - \operatorname{tr} A_T^2$$

which proves the first two statements immediately. If T is restrained, as in the last statement, then we force $\operatorname{Ric}(T, T) \equiv 0$ and T to have constant length since $R(T, T) < 0$ at any point is impossible. Thus ϕ

is constant, $\Delta_2\phi = 0$, and $\operatorname{tr} A_T^2 = 0$ which implies all the eigenvalues of A_T are zero, so $A_T = 0$.

COROLLARY 2. *A nontrivial closed field has constant length on a semi-Riemannian manifold if and only if its integral curves are geodesics.*

Proof. This is trivial since $\operatorname{grad} \phi = 2D_T T$.
The following result applies to any vector field.

PROPOSITION 5. The integral curves of a field T are reparameterizations of geodesics if and only if $D_T T = gT$ for some real valued C^∞ function g.

Proof. If the field fT is geodesic (f never vanishes), $0 = D_{fT} fT = f[(Tf)T + fD_T T]$ and $g = -T(\log f)$. Conversely, if $D_T T = gT$ then along each integral curve of T we need only solve the linear equation $(Tf) + fg = 0$ to obtain f for which fT is geodesic.

COROLLARY. *If T is closed, nonvanishing, and $D_T T = gT$ then* $\operatorname{Ric}(T) = g \operatorname{div} T - \operatorname{tr} A_T^2 + T(g - \operatorname{div} T)$.

We now study the case when T has constant length on the hypersurfaces M'.

THEOREM 3. *The following four statements are equivalent on the set U:*

(a) *A_T is invariant on $T^\perp$.*
(b) *T has constant length on any M'.*
(c) *$D_T T$ is orthogonal to $T^\perp$.*
(d) *$[T, X]$ is in $T^\perp$ if X in $T^\perp$.*

Proof. If X is in $T^\perp$ then $X\langle T, T\rangle = 2\langle A_T X, T\rangle = 2\langle X, D_T T\rangle$ which shows (a), (b), and (c) are equivalent. Also

$$\begin{aligned}\langle A_T X, T\rangle = \langle X, A_T T\rangle &= T\langle X, T\rangle - \langle D_T X, T\rangle \\ &= -\langle D_X T + [T, X], T\rangle ,\end{aligned}$$

where we extend X to be a C^∞ field in $T^\perp$. Hence $2\langle A_T X, T\rangle = \langle [X, T], T\rangle$ which shows (a) is equivalent to (d).

THEOREM 4. *If one of the statements in Theorem 3 holds and T does not vanish, then the integral curves of T are reparameterizations of geodesics,* $\operatorname{grad} \phi = 2D_T T = (T \log e\phi)T$, *and the vector* $\operatorname{grad} \phi$ *has* *constant length on* M', *i.e.* $T \log \phi$ *is constant on* M'. *Moreover, the mean curvature of* M' *is constant if and only if* $\operatorname{div} T$ *is constant on* M'.

Proof. Letting $\operatorname{grad} \phi = fT$ then $T\phi = 2\langle D_T T, T\rangle = \langle fT, T\rangle = f\phi$. If $T \neq 0$, then $\phi \neq 0$, so $f = (T\phi)/\phi = T \log e\phi$. The integral curves of T are reparameterizations of geodesic by Proposition 5.

Letting X be a C^∞ field in $T^\perp$, then

$$Xf = XT(\log \phi) = [X, T] \log \phi + T(X \log \phi) = 0$$

since $[X, T]$ is in $T^\perp$ and ϕ is constant on M'.

The last statement of the conclusion follows from Theorem 2.

COROLLARY. *If* $\operatorname{grad} \phi$ *does not vanish on* M, *then the hypersurfaces* M' *are precisely the constant hypersurfaces of* ϕ *if and only if one of the statements in Theorem* 3 *is true.*

We return to the study of the geometry of the hypersurface M'. Recall the fact that if L is the Weingarten map of an oriented nonsingular hypersurface in a semi-Riemannian manifold, then the Codazzi-Mainardi equations hold on the hypersurface if and only if $\operatorname{Tor}_L = 0$. In the following theorem, we write $A_T = \Delta T$ which is admissable by the identification of linear transformations with tensors of type 1,1.

THEOREM 5. *On the set* U, *the following three statements are equivalent:*

(a) *The Codazzi-Mainardi equations hold on* M'.

(b) $\operatorname{Tor}_{\Delta T} = 0$ *on vectors in* $T^\perp$.

(c) $R(X, Y)T = 0$ *for all* X, Y *in* $T^\perp$.

Proof. Let D' be the induced Riemannian covariant differentiation on M', thus for fields X and Y in $T^\perp$,

$$D_X Y = D'_X Y - \langle LX, Y\rangle rN$$

by the Gauss equation (see Hicks [7]), where $r = \langle N, N\rangle = e$.

Using the Gauss equation and Theorem 2, a straightforward computation yields,

$$\begin{aligned}\operatorname{Tor}_L(X, Y) &= D'_X(LY) - D'_Y(LX) - L([X, Y]) \\ &= [e\langle T, T\rangle]^{-1/2} \operatorname{Tor}_{\Delta T}(X, Y) \\ &\quad - [e\langle T, T\rangle]^{-3/2} e\langle \operatorname{Tor}_{\Delta T}(X, Y), T\rangle T \\ &= |T|^{-1} \operatorname{Tor}_{\Delta T}(X, Y)\ ,\end{aligned}$$

since $\mathrm{Tor}_{\Delta T}(X, Y) = D_X A_T Y - D_Y A_T X - A_T[X, Y] = R(X, Y)T$ and $\langle R(X, Y)T, T\rangle = 0$ by the skew-symmetry of the covariant Riemann-Christoffel curvature tensor. Thus $\mathrm{Tor}_{\Delta T}(X, Y)$ has no component orthogonal to M' and the conclusion now follows.

THEOREM 6. *On the set U, let P be a two dimensional subspace of M' with nonsingular orthonormal base X, Y Then*

$$K(P) = K'(P) - [e\langle T, T\rangle\langle X, X\rangle\langle Y, Y\rangle]^{-1}[\langle A_T X, X\rangle\langle A_T Y, Y\rangle - \langle A_T X, Y\rangle^2]$$

relates the Riemannian curvature of P with respect to M and M'.

Proof. The general Gauss curvature equation (see Hicks [6]) states that

$$\tan R(X, Y)Z = R'(X, Y)Z - r(\langle LY, Z\rangle LX - \langle LX, Z\rangle LY) .$$

Using Theorem 2, a straightforward computation yields the result.

COROLLARY. *If M is Riemannian and $T = \mathrm{grad}\, f$, m in U, and $x, y, \cdots$ are a set of Riemann normal coordinates at m such that $\partial/\partial x$ and $\partial/\partial y$ span the subspace P in M'_m, then*

$$K(P) = K'(P) - \left[\frac{\partial^2 f}{\partial x^2}\frac{\partial^2 f}{\partial y^2} - \left(\frac{\partial^2 f}{\partial x \partial y}\right)^2\right]\Big/ |\mathrm{grad}\, f|^2$$

at m.

Proof. Let $X = \partial/\partial x$ and $Y = \partial/\partial y$. Then $\langle A_T X, Y\rangle_m = \langle D_X T, Y\rangle = X\langle T, Y\rangle = X_m(Yf)$ since $(D_X Y)_m = 0$.

We now show the tensor $\mathrm{Tor}_{\Delta T}$ represents a condition on the holonomy of the distribution $T^\perp$.

THEOREM 7. *Let M be Riemannian, complete, connected, and simply connected. Let T be a nonvanishing closed field such that ΔT has no torsion. Then M is diffeomorphic to a product $M' \times R$, where M' is the $(n-1)$ dimensional integral submanifold of $T^\perp$ through a point m in M and R is the real line. Hence the orbit space M/T is diffeomorphic to M'.*

Proof. Since M is simply connected its restricted homogeneous holonomy group is equal to its homogeneous holonomy group H. The Lie algebra of H is generated by the linear transformations $R(X, Y)$ on M_m for all vectors X and Y in M_m (see Nomizu [8]). Since $\mathrm{Tor}_{\Delta T} = 0$, $R(X, Y)T = 0$ for all X and Y hence $R(X, Y)$ is invariant on $T^\perp$.

Since H is contained in the special orthogonal group $SO(n, R)$, which is compact, the exponential map is onto. If h is in H, then $h = \exp R(X, Y)$ for some X and Y in M_m, and thus $h(T^\perp)$ is contained in $T^\perp$. We now apply the result of DeRham [3] to get $M = M' \times N$. Since M is Riemannian and complete, N is diffeomorphic to the real line or the one dimensional torus. Since M is simply connected, N is diffeomorphic to R.

4. **Special cases.** We conclude with some special cases that follow from the above results. We will always assume the field T is nontrivial, nonsingular, and closed.

PROPOSITION 6. If $A_T \equiv 0$, then T is a geodesic field with constant length, zero divergence, and zero Ricci curvature. If T lies in the plane section P then $K(P) = 0$. Thus there is no pair of conjugate points along the geodesics determined by T. The distribution $T^\perp$ is integrable and its integral manifolds M' are flatly imbedded in M (i.e. $L \equiv 0$ on M'). Hence M' is a geodesic submanifold of M. If M is Riemannian, complete, and simply connected, then M is isometric to the product $M' \times R$.

PROPOSITION 7. If $B_T \equiv 0$ and T is geodesic then T has constant length c and $\text{Ric}(T) = -\text{tr}\, A_T^2 - T(\text{div}\, T)$. When M' is defined it has total curvature zero and mean curvature $(1/c)\,\text{div}\, T$. If M' is defined and flat everywhere, then $A_T \equiv 0$ and Proposition 6 is applicable.

PROPOSITION 8. If $B_T \equiv 0$ and the integral curves of T are reparameterizations of geodesics with $D_T T = gT$, then at points where g and T do not vanish, M' is flat and the Ricci curvature of T is zero.

In proving Proposition 8 one shows at points in U where g does not vanish then $A_T T = D_T T = (\text{div}\, T)T$ by applying Proposition 6 to $D_T T$. Furthermore, at such points $0 = B_T T = [T(\text{div}\, T) + (\text{div}\, T)^2]T$ so $\text{tr}\, A_T^2 = (\text{div}\, T)^2 = -T(\text{div}\, T)$ and $\text{Ric}(T) = 0$.

PROPOSITION 9. If $B_T \equiv 0$ and the integral curves of T are not reparameterizations of geodesics, then Proposition 6 may be applied to $D_T T$. Moreover $T^2\langle T, T\rangle$ is constant, hence there can be at most one point on each integral curve of T where the length of T has a critical point. If the integral curves of T are parametrically complete (defined for all parameter values), then M cannot be compact.

Notice in Proposition 9 the length of T is not constant along any of its integral curves, for $0 = T\langle T, T\rangle = 2\langle D_T T, T\rangle$ implies $D_T T = gT$ by Theorem 4, which implies the integral curves of T are geodesics by Proposition 5.

BIBLIOGRAPHY

1. S. Bochner, *Vector fields and Ricci curvature*, Bull. Amer. Math. Soc. **52** (1946), 776-797.

2. C. Chevalley, *Theory of lie groups*, Princeton Univ. Press, Princeton, 1949.

3. G. DeRham, *Sur la reductibilité d'un espace de Riemann*, Comment. Math. Helv. **26** (1952), 328-344.

4. S. Helgason, *Differential geometry and symmetric spaces*, Academic Press, New York, 1962.

5. R. Hermann, *On geodesics that are also orbits*, Bull. Amer. Math. Soc. **66** (1960), 91-93.

6. N. Hicks, *Submanifolds of semi-Riemannian manifolds*, Rend. Cir. Mat. Palermo, (to appear).

7. N. Hicks, *Notes on differential geometry*, D. Van Nostrand, Princeton, (to appear).

8. K. Nomizu, *Lie groups and differential geometry*, Math. Soc. Japan Publ. 2, 1956.

THE UNIVERSITY OF MICHIGAN

PACIFIC JOURNAL OF MATHEMATICS
Vol. 15, No. 1, 1965

DOUBLY STOCHASTIC OPERATORS OBTAINED FROM POSITIVE OPERATORS

CHARLES HOBBY AND RONALD PYKE

A recent result of Sinkhorn [3] states that for any square matrix A of positive elements, there exist diagonal matrices D_1 and D_2 with positive diagonal elements for which $D_1 A D_2$ is doubly stochastic. In the present paper, this result is generalized to a wide class of positive operators as follows.

Let $(\Omega, \mathfrak{A}, \lambda)$ be the product space of two probability measure spaces $(\Omega_i, \mathfrak{A}_i, \lambda_i)$. Let f denote a measurable function on $(\Omega, \mathfrak{A})$ for which there exist constants c,C such that $0 < c \leqq f \leqq C < \infty$. Let K be any nonnegative, two-dimensional real valued continuous function defined on the open unit square, $(0,1) \times (0,1)$, for which the functions $K(u,\cdot)$ and $K(\cdot,v)$ are strictly increasing functions with strict ranges $(0,\infty)$ for each u or v in $(0,1)$. Then there exist functions $h\colon \Omega_1 \to E_1$ and $g\colon \Omega_2 \to E_1$ such that

$$\int_{\Omega_2} f(x, v)\, K(h(x), g(v))\, d\lambda_2(v) = 1 = \int_{\Omega_1} f(u,y)\, K(h(u), g(y))\, d\lambda_1(u)\ ,$$

almost everywhere — (λ).

Let $(\Omega, \mathfrak{A}, \lambda)$ be the product space of two probability measure spaces $(\Omega_i, \mathfrak{A}_i, \lambda_i)$. Let f denote a measurable function on $(\Omega, \mathfrak{A})$ for which there exist constants c,C such that

$$0 < c \leqq f \leqq C < \infty\ . \tag{1}$$

Let K be any nonnegative, real valued continuous function defined on the open unit square, $(0,1) \times (0,1)$, for which the functions $K(u,\cdot)$ and $K(\cdot,v)$ are strictly increasing functions with strict ranges $(0, \infty)$ for each u or v in $(0,1)$.

In what follows, h and g will denote measurable, real valued, functions defined on Ω_1, and Ω_2, respectively. Whenever well defined, set

$$\begin{aligned} R(x\colon h,g) &= \int_{\Omega_2} f(x,v)\, K(h(x), g(v))d\lambda_2(v) \\ C(y\colon h,g) &= \int_{\Omega_1} f(u,y)\, K(h(u), g(y))d\lambda_1(u) \end{aligned} \tag{2}$$

for $(x,y) \in \Omega$.

Received February 13, 1964. The authors research was supported in part by the National Science Foundation, Grant G-25211 and Grant G-19915.

For a fixed choice of h, g we can think of R and C as defining positive operators. The main result of this paper is that R and C can be made doubly stochastic by choosing h and g appropriately. One immediate consequence of this result is a recent theorem of Sinkhorn [3] on doubly stochastic matrices.

THEOREM. *There exist functions* $h: \Omega_1 \to (0,1)$ *and* $g: \Omega_2 \to (0,1)$ *for which*

$$R(x: h,g) = 1 = C(y: h,g) \ , \tag{3}$$

almost everywhere $- (\lambda)$.

Proof. We shall obtain h and g as the limits of two sequences of functions, $\{h_n\}$ and $\{g_n\}$. The h_n and g_n are defined recursively as follows.

Set $h_0(x) = \alpha$ for all $x \in \Omega_1$, where α is any number in (0,1). If h_n has been defined, let g_n be the function defined by the equation $C(y: h_n, g_n) = 1$. That is, $g_n(y)$ is the solution of the equation

$$1 = \int_{\Omega_1} f(x,y)\, K(h_n(x), g_n(y)) d\lambda_1(x) \ . \tag{4}$$

This solution exists and is unique since $C(y: h_n, t)$ is a strictly increasing continuous function of t with range $(0, \infty)$. Furthermore, g_n is easily seen to be measurable if h_n is measurable (certainly the case for h_0), since $\{y \in \Omega_2: g_n(y) \leqq t\} = \{y \in \Omega_2: C(y: h_n, t) \geqq 1\}$ and since $C(y: h_n, t)$ is a measurable function of y for each fixed t. By Fubini's theorem

$$\int_{\Omega_1} R(x: h_n, g_n) d\lambda_1(x) = \int_{\Omega_2} C(y: h_n, g_n) d\lambda_2(y) = 1. \tag{5}$$

Thus if $R(x: h_n, g_n) \geqq 1$ for all x in Ω_1, then $R(x: h_n, g_n) = 1$ almost everywhere $- \lambda_1$, and the proof is complete. If for some $x \in \Omega_1$, $R(x: h_n, g_n) < 1$, we define $h_{n+1}(x)$ to be the numbert for t which $R(x: t, g_n) = 1$. The existence and uniqueness of $h_{n+1}(x)$ follow from our assumptions on K. We set $h_{n+1}(x) = h_n(x)$ at every x where $R(x: h_n, g_n) \geqq 1$. Just as for g_n, we see that h_{n+1} is measurable (since g_n is measurable).

Let $A_n = \{x \in \Omega_1 \mid R(x: h_n, g_n) \leqq 1\}$. If for some $n \geqq 0$, $\lambda_1(A_n) = 1$ we stop our iteration since this implies that $R(x: h_n, g_n) = 1$ a.e. $-\lambda_1$, so we can take h_n and g_n to be h and g of the theorem. We shall assume henceforth that $\lambda_1(A_n) < 1$ for every n.

Observe that $h_{n+1}(x) \geqq h_n(x)$ for every x, thus

$$1 = C(y: h_n, g_n) \leqq C(y: h_{n+1}, g_n) \ . \tag{6}$$

Consequently $g_{n+1}(y) \leqq g_n(y)$ for every y. It follows from this monotonicity that the limits $h = \lim_{n\to\infty} h_n$ and $g = \lim_{n\to\infty} g_n$ exist. We shall now show that this choice of h and g satisfies the theorem.

By our construction, $\{A_n\}$ is a nondecreasing sequence of sets. Set $A = \lim_{n\to\infty} A_n$. Since $\lambda_1(A_n) < 1$, the complementary set A_n^c is a set of positive measure for each n. For $x \in A_n^c$, $h_n(x) = \alpha$ whence

$$1 \leqq R(x\colon h_n, g_n) = \int_{\Omega_2} f(x,y)\, K(\alpha, g_n(y)) d\lambda_2(y)$$
$$\leqq C\int_{\Omega_2} K(\alpha, g_n(y)) d\lambda_2(y) .$$

This inequality holds for each n, so one may take limits to obtain

$$1 \leqq C\int_{\Omega_2} K(\alpha, g(y)) d\lambda_2(y) .$$

Thus there are positive numbers r and σ such that $\lambda_2\{y \in \Omega_2\colon g(y) \geqq r\} > \sigma$. Then for arbitrary n and $x \in A_n$,

$$1 \geqq c\int_{\Omega_2} K(h_n, g_n) d\lambda_2(y) \geqq c\sigma K(h_n(x), r) .$$

Hence, by taking limits on n, one obtains $1 \geqq c\sigma K(h(x), r)$ for each $x \in A$. Let t be a number for which $1 = c\sigma K(t, r)$. Then $h(x) \leqq t$ for $x \in A$, and $h(x) = \alpha$ for $x \in A^c$, whence $h(x) \leqq \beta = \max(\alpha, t) < 1$ for all $x \in \Omega_1$. But for all $y \in \Omega_2$ and all n,

$$1 = \int_{\Omega_1} f(x,y)\, K(h_n(x), g_n(y)) d\lambda_1(x)$$
$$\leqq CK(\beta, g_n(y)) ,$$

thus $g(y) \geqq \gamma > 0$ where γ satisfies $C^{-1} = K(\beta, \gamma)$.

The import of the above is that the set $\{(h_n(x), g_n(y))\colon (x,y) \in \Omega, n \geqq 0\}$ is contained in a compact subset of the interior of $[0,1] \times [0,1]$, on which K is continuous, and hence bounded. Therefore, by the Lebesgue dominated convergence theorem

$$1 = \lim_{n\to\infty} C(y\colon h_n, g_n) = \int_{\Omega_1} f(x,y)\, K(h(x), g(y)) d\lambda_1(x)$$

and

$$1 = \lim_{n\to\infty} R(x\colon h_{n+1}, g_n) = \int_{\Omega_2} f(x,y)\, K(h(x), g(y)) d\lambda_2(y) ,$$

for $x \in A$. Moreover

$$1 \leqq \lim_{n\to\infty} R(x\colon h_n, g_n) = \int_{\Omega_2} f(x,y)\, K(h(x), g(y)) d\lambda_2(y) ,$$

for $x \notin A$. But an inequality here on a set of positive λ_1-measure is impossible by (5), thereby completing the proof.

COROLLARY (Sinkhorn [3]). *Let* $A = (a_{ij})$ *be an* m *by* m *matrix of positive elements. There exist diagonal matrices* D_1 *and* D_2 *of positive diagonal elements for which the matrix* D_1AD_2 *is doubly stochastic.*

Proof. In the above theorem let $\Omega_1 = \Omega_2 = \{1,2, \cdots, m\}$ and let $\lambda_1 = \lambda_2$ be the uniform measure, $\lambda_1(\{j\}) = 1/m$. Set $K(u,v) = uv(1-u)^{-1}(1-v)^{-1}$ and $f(i,j) = a_{ij}$. By the theorem there exist functions h and g such that

$$m^{-1}\sum_{i=1}^{m} a_{ij}h(i)g(j)\,[1-h(i)]^{-1}\,[1-g(j)]^{-1} = 1$$
$$= m^{-1}\sum_{j=1}^{m} a_{ij}h(i)g(j)\,[1-h(i)]^{-1}[1-g(j)]^{-1}\,.$$

The corollary is then proved if one lets $d_{1i} = m^{-1/2}[1-h(i)]^{-1}h(i)$ and $d_{2i} = m^{-1/2}[1-g(i)]^{-1}g(i)$ be the diagonal elements of D_1 and D_2 respectively.

The above result for symmetric matrices has also been obtained by Marcus and Newman [1] and Maxfield and Minc [2].

The application which motivated Sinkhorn's theorem was the case in which A is the matrix of maximum likelihood estimates of a stochastic transition matrix P of a Markov Chain. When it is further known that P is actually doubly stochastic, then Sinkhorn's result shows that numbers $\{x_1, \cdots, x_n; y_1, \cdots, y_n\}$ exist such that A can be renormalized by dividing the ith row by x_i and the jth column by y_j to obtain a doubly stochastic matrix. However, if one considers the maximum likelihood equations for the restricted case in which P is known to be doubly stochastic one observes that the proper normalized form of A (relative to the maximum likelihood approach) is a doubly stochastic matrix $B = (b_{ij})$ with $b_{ij} = a_{ij}(x_i + y_j)^{-1}$. The existence of such a normalization follows straightforwardly from the proof of the above theorem. To see this, consider the function $K(u,v) = [v^{-1} - (1-u)^{-1}]^{-1}$ defined on the triangular region $u > 0$, $v > 0$, $u + v < 1$. This function is nonnegative and continuous on this triangle. Moreover, both $K(u,\cdot)$ and $K(\cdot,v)$ are strictly increasing functions wherever defined and the ranges of $K(u,\cdot)$ and $K(\cdot,v)$ are respectively $(0,\infty)$ and $(v[1-v]^{-1},\infty)$ for each fixed u and v. Let λ_1 and λ_2 be the same discrete measures as used in the proof of the above corollary. The functions $R(x\colon h_n,g_n)$ and $C(y\colon h_n,g_n)$ then become finite sums. The only change required in the proof is that one must show that the points $(h_n(x),g_n(y))$, for all $n \geqq 1$ and all x and y, are well defined and contained in a compact subset of the domain of K. That this is true follows from the assumptions on the monotonicity, continuity and range of K, combined with the fact that the integrals are finite sums. Actually, because of these properties, it is clear that $K(h_n(x), g_n(y))$ is bounded by mc^{-1} for all n and y.

REFERENCES

1. M. Marcus and M. Newman. *The permanent of a symmetric matrix*, to appear. [cf. Notices of Amer. Math. Soc. **8** (1961), Abstract 587-85, p. 595.]
2. H. Minc and J. E. Maxfield, *A doubly stochastic matrix equivalent to a given matrix*, Notices of Amer. Math. Soc. **9** (1962), Abstract 592-81, p. 309.
3. R. Sinkhorn, *A relationship between arbitrary positive matrices and doubly stochastic matrices*, Ann. Math. Stat. **35** (1964), 876-879.

UNIVERSITY OF WASHINGTON

PACIFIC JOURNAL OF MATHEMATICS
Vol. 15, No. 1, 1965

CONCERNING PERIODIC SUBADDITIVE FUNCTIONS

R. F. JOLLY

The author investigates those subsets M of the complex plane with the group property that M is closed with respect to complex multiplication. In particular if M is closed, bounded and has for its boundary a curve given in polar form by $\rho(\theta) = r(\theta) \exp(i\theta)$ where r is a positive continuous function with period 2π, then r is characterized by these requirements, together with the additional condition that r be submultiplicative. If $f(x) = -\log r(x)$, the corresponding conditions on f are: f is a continuous nonnegative subadditive function with period 2π.

Some relations between the roots (zeros) and periods of subadditive functions are discussed and in particular, it is shown that: if f is a continuous subadditive function not identically zero, with period 1 and with a root c (i.e., $f(c) = 0$), then c is a rational number m/n (in lowest terms), $f(0) = 0$ and f has period $1/n$.

For each positive number c and function f on the set of all numbers, a type of polygonal approximation $P(c, f)$ is defined such that if f is continuous, $\lim P(c, f) = f$ uniformly over every bounded number set as $c \to 0$. If f is subadditive, $P(c, f)$ is subadditive. The subadditive $P(c, f)$ are characterized in terms of their slopes. Since a change of scale does not affect the subadditive property, the author studies functions with period 1 rather than those with period 2π. For each positive integer n, the collection K_n of all functions $P(1/n, f)$ for all continuous subadditive functions f with period 1, is shown to have a finite basis. In fact, K_n forms a function cone with finitely many extremal elements (the basis). While an explicit representation is not given, the proof shows how these extremal elements may be constructed.

Several examples are given to illustrate some pathological cases. The methods of this paper may easily be applied to the solution of certain other functional inequalities with corresponding restrictions.

1. Introduction.

The statement that M is a *G-set* means that M is a point set in the complex plane such that if P is in M and Q is in M, then the product PQ is in M, i.e., M has the group property that it is

Received March 5, 1964. Presented to the American Mathematical Society on August 29, 1963. Part of this work is contained in the author's dissertation at the University of Texas. The author is deeply grateful to Professor H. S. Wall for his suggestions about the preparation of this manuscript.

closed with respect to multiplication. For example, the set of all points[1] z such that $|z| < 1$ is a *G-set*; for each number k, the intersection of the interiors of the circles with radius $(1 + k^2)^{1/2}$ and centers ik and $-ik$ is a *G-set*; and the union of the coordinate axes is a *G-set* which contains no domain (§ 6). The statement that M is a *simple G-set* means that M is a *G-set* which is closed, bounded and which has for its boundary a curve given in polar form by $\rho(\theta) = r(\theta) \exp(i\theta)$ where r is a positive continuous function of period 2π. The function r is completely characterized by these requirements together with the additional condition that for each number x and number y, $r(x)\, r(y) \leqq r(x + y)$. If $f(x) = -\log r(x)$, the corresponding conditions on f are: f is continuous, nonnegative and of period 2π and moreover, for each number x and number y, $f(x + y) \leqq f(x) + f(y)$. Hence, the determination of all simple G-sets resolves itself into the problem of the determination of all continuous nonnegative subadditive functions with period 2π (§ 2).

For each number p, let F_p denote the collection of all functions f which are subadditive on the set of all numbers and which have the property that if x is a number, $f(x + p) = f(x)$. Let F denote the collection of all continuous functions in F_1. The statement that c is a *root* of f means $f(c) = 0$. An *anchored* function is one with zero for a root. In § 3, some relationships are shown between the roots and periods of subadditive functions. In particular, it is shown that if f is a function in F not identically zero and c is a root of f, then c is a rational number m/n (in lowest terms) and f is an anchored function with period $1/n$ (Theorem 5).

Since a change of scale does not affect the subadditive property (Lemma 3), we study the functions in F_1 instead of those in $F_{2\pi}$ and thereby simplify the notation. For each positive number c and function f on the set of all numbers, a type of *polygonal approximation* $P(c, f)$ is defined (§ 4) such that if f is continuous, $\lim_{c\to 0} P(c, f) = f$ uniformly over every bounded number set. These polygonal approximations to functions in F_0 are themselves in F_0 (Theorem 6) and are characterized in terms of their slopes (Theorem 7). It is then shown that for each positive integer n, the collection K_n of all functions $P(1/n, f)$ for all functions f in F, has a finite *basis* in the sense that there is an integer $M(n)$ and a sequence α_n of $M(n)$ elements of K_n such that a function g belongs to K_n if, and only if, g is the sum of a linear combination of the functions of α_n with nonnegative coefficients (Theorem 9). These polygonal subadditive functions are then used to characterize F as the collection to which f belongs if, and only if, f

[1] In this paper, the word *number* shall be used to denote a real number and the word *point* shall mean a point of the complex plane.

can be approximated uniformly by linear combinations of the functions of α_1, α_2, α_3, $\cdots$ with nonnegative coefficients.

There are simple G-sets which are not convex and in fact, which have no tangent at any point (Theorem 12). Some other examples, which show the difficulty in obtaining nontrivial characterizations of G-sets, are given in § 6. One example is a countable G-set dense in the plane.

2. Boundaries of simple G-sets. Throughout this section, assume that r is a positive continuous function with period 2π and let D denote the closed and bounded set with boundary $z = r(\theta) \exp(i\theta)$.

THEOREM 1. *The following two statements are equivalent:*

(i) *D is a simple G-set.*

(ii) *For each number α and each number β, $r(\alpha)r(\beta) \leqq r(\alpha + \beta)$.*

Proof. To show that statement (ii) implies statement (i), let $P = c \exp(i\alpha)$ and $Q = d \exp(i\beta)$ where $0 \leqq c < r(\alpha)$, $0 \leqq d < r(\beta)$ and $r(\alpha)r(\beta) \leqq r(\alpha + \beta)$. Hence $PQ = cd \exp(i(\alpha + \beta))$ and PQ is in D.

To show that statement (i) implies statement (ii), assume (i) and $r(\alpha)r(\beta) > r(\alpha + \beta)$. Let $P = (r(\alpha) - \delta) \exp(i\alpha)$ and $Q = (r(\beta) - \delta) \exp(i\beta)$ where $\delta = [r(\alpha)r(\beta) - r(\alpha + \beta)]/[r(\alpha) + r(\beta)] > 0$. Note that $r(\alpha) - \delta = [r^2(\alpha) + r(\alpha + \beta)]/[r(\alpha) + r(\beta)] > 0$. Therefore P is in D. Likewise $r(\beta) - \delta = [r^2(\beta) + r(\alpha + \beta)]/[r(\alpha) + r(\beta)]$ and Q is in D. Since P is in D and Q is in D. PQ is in D. Therefore $r(\alpha + \beta) > (r(\alpha) - \delta)(r(\beta) - \delta)$ but $(r(\alpha) - \delta)(r(\beta) - \delta) > r(\alpha)r(\beta) - \delta(r(\alpha) + r(\beta)) = r(\alpha + \beta)$. This is a contradiction.

THEOREM 2. *Suppose that D is a G-set. Then the following two statements are equivalent:*

(i) *Each point of D is the product of two points of D.*

(ii) $r(0) = 1$.

Proof. r is continuous and has period 2π, therefore there is a number $0 \leqq w < 2\pi$ such that for any number α, $r(\alpha) \leqq r(w)$. Since $r^2(w) \leqq r(2w) \leqq r(w)$, $r(w) \leqq 1$.

To show that statement (ii) implies statement (i), suppose $Z = d \exp(i\alpha)$ where $0 \leqq \mathrm{d} < r(\alpha)$. From the preceding $d < r(\alpha) \leqq r(w) \leqq 1$, hence $d < (1/2)(1 + d) < 1$ and $2d/(1 + d) < 1$. Let $W = (1/2)(1 + d) \exp(i\alpha)$ and $U = 2d/(1 + d)$. Then $Z = WU$.

To show that statement (i) implies statement (ii), assume (i) is true. Let us first show that $r(w) = 1$. Suppose $r(w) < 1$. Let $Z = (1/2)[r(w) + r^2(w)] \exp(iw)$. By (i) there is a point $W = c \exp(i\alpha)$ and a point $U = d \exp(i\beta)$ where $0 \leqq c < r(\alpha)$, $0 \leqq d < r(\beta)$ and $Z =$

WU. Note that $0 \leqq c < r(\alpha) \leqq r(w)$ and $0 \leqq d < r(\beta) \leqq r(w)$. Therefore $cd < r^2(w) < (1/2)[r(w) + r^2(w)] < r(w)$. Hence $Z \neq WU$ which is a contradiction. Therefore $r(w) = 1$.

Since $1 = r(w)r(kw) \leqq r((k+1)w) \leqq r(w) = 1$ when $r(kw) = 1$, it follows by induction that if n is a positive integer, $r(nw) = 1$.

Since r is continuous and periodic, r is uniformly continuous. Suppose $\varepsilon > 0$. Then there is a number $\delta > 0$ such that if $|x - y| < \delta$, $|r(x) - r(y)| < \varepsilon$. For infinitely many positive integers m and n, $|(w/2\pi) - (m/n)| < n^{-2} < \delta/(2\pi n)$. For such integers m and n, $|r(nw) - r(2\pi m)| < \varepsilon$, $r(nw) = 1$ (previously proven) and $r(2\pi m) = r(0)$ (r has period 2π). Hence, $1 - r(0) < \varepsilon$. Therefore $r(0) = 1$.

THEOREM 3. *If g is a continuous periodic submultiplicative function, either g is a positive function or $g \equiv 0$.*

Proof. For each number x, $0 \leqq g(x/2)\, g(x/2) \leqq g(x)$. Since g is continuous, either $g \equiv 0$ or there is some segment containing no root of g. Suppose $a < b$ and the segment (a, b) contains no root of g. If $a < x < b$ and $n > 0$, then $na < nx < nb$ and if $g(kx) \neq 0$ and $g(x) \neq 0$, then $0 < g(x)g(kx) \leqq g((k+1)x)$. It follows by induction that for every positive integer n, the segment (na, nb) contains no root of g. Since g has period p, for some positive number p, and there is a positive integer N such that $N(b - a) > p$, there is no root of g and hence g is positive.

THEOREM 4. *Suppose f and g are functions such that $f = -\log g$. Then the following two statements are equivalent*:

(i) *g is positive and submultiplicative.*

(ii) *f is subadditive.*

The proof is omitted.

3. Roots and periods of subadditive functions.

Let us now show how the roots of subadditive functions are related to their periods and in particular, what happens in the continuous case.

Note that if p and $-p$ are both roots of the function f of F_0, then f is an anchored function with period p. This is shown by the inequalities

$$0 = f(p) \leqq f(p) + f(0) = f(0) = f(p - p) \leqq f(p) + f(-p) = 0$$
$$f(x) = f(x + p - p) \leqq f(x + p) + f(-p) \leqq f(x) + f(p) = f(x).$$

On the other hand, the example

$$f(x) = \{1, \text{ if } x \leqq \pi/2;\ |\sin x|, \text{ if } \pi/2 \leqq x\}$$

shows that a continuous element of F_0 may have infinitely many roots without being either anchored or periodic. However, if f is a nonnegative continuous element of F_0 with both a positive and negative root, then f is anchored and periodic. This is easily shown by letting p and q denote respectively the smallest positive and largest negative roots of f (these obviously exist). But $q < p + q < p$ and $0 \leqq f(p + q) \leqq f(p) + f(q) = 0$. Hence $f(p + q) = 0$ and $p + q = 0$.

THEOREM 5. *Suppose f is in F and c is a number different from 0 such that $f(c) = 0$. Then*

(i) *f is anchored;*

(ii) *if c is irrational, $f \equiv 0$;*

(iii) *if c is a rational number m/n (in lowest terms), then f has period $1/n$.*

Proof. First we show that there is no number x such that $f(x) < 0$. Suppose there were such an x. Then $f(0 + 0) \leqq f(0) + f(0)$ implies $0 \leqq f(0)$. Hence $x \neq 0$. Since $f((k + 1)x) \leqq f(x) + f(kx)$, it follows by induction that $f(nx) \leqq nf(x)$ for every positive integer n. As f is continuous and periodic, there is a number M such that if w is a number, $|f(w)| < M$. Let m denote an integer greater than $M/|f(x)|$. Therefore $m|f(x)| > M$ but if $f(x) < 0$, $|f(mx)| = -f(mx) \geqq -mf(x) = m|f(x)|$ which is a contradiction.

Since f has period 1, assume $0 < c < 1$. It was previously shown that f is nonnegative and if n is a positive integer, $f(nx) \leqq nf(x)$. Hence $f(nc) = 0$ for every positive integer n. Following the line of argument used in Theorem 2, the fact that f is uniformly continuous may now be used to show that $f(0) = 0$ and therefore f is anchored.

To show (ii), assume c is irrational and $f \neq 0$. Again following a line of argument used in Theorem 2, the fact that the multiples of c modulo 1 are dense in the interval $[0, 1]$ gives a contradiction since f is continuous, has period 1 and $f(nc) = 0$ for every positive integer n.

To show (iii), assume c is the rational number m/n in lowest positive terms. There exists an integer k and an integer p such that $km - np = 1$. Hence $km/n = (1 + np) = p + 1/n$ and $0 = f(km/n) = f(p + 1/n) = f(1/n)$. For each number w, $f(w + 1/n) \leqq f(w) + f(1/n) = f(w)$. Therefore it follows by induction that for every number x and positive integer s, $f(x + s/n) \leqq f(x + 1/n) \leqq f(x)$. But if $s = n$, $f(x) = f(x + n/n) \leqq f(x + 1/n) \leqq f(x)$. Therefore f has period $1/n$.

4. Certain polygonal approximations to subadditive functions. Let us start with some elementary properties of subadditive functions and follow this with a definition and some properties of a certain type of polygonal function.

LEMMA 1. *If* $a > 0$, $b > 0$, f *is in* F_p *and* g *is in* F_p, *then* $af + bg$ *is in* F_p.

LEMMA 2. *If for each positive integer* n, f_n *is in* F_p *and* g *is a function such that for each number* x, $f_n(x)$ *converges to* $g(x)$, *then* g *is in* F_p.

LEMMA 3. *If* c *is a number,* f *is in* F_{cp} *and* g *is the function such that for every number* x, $g(x) = f(cx)$, *then* g *is in* F_p.

Proofs to Lemmas 1, 2 and 3 may be found in [3, Chap. VII].

DEFINITION. For each positive number c and function f defined on the set of all numbers, let $P(c, f)$ denote the function h such that (i) if n is a positive integer and $nc - c \leqq x \leqq nc$, then $h(x) = m_n(x - nc) + f(nc)$ where $m_n = [f(nc) - f(nc - c)]/c$ and (ii) if k is a negative integer and $kc \leqq x \leqq kc + c$, then $h(x) = -m_k(x - kc) + f(kc)$ where $m_k = -[f(kc + c) - f(kc)]/c$. Also let $m_0 = f(0)/c$.

DEFINITION. For each function h defined on the set of all numbers, let h^* denote the transformation from the set of all ordered number pairs into a number set such that for every ordered number pair (x, y), $h^*(x, y) = h(x) + h(y) - h(x + y)$.

THEOREM 6. *If* f *is in* F_0 *and* c *is a number then* $P(c, f)$ *is in* F_0.

This theorem is equivalent to Theorem 8 of [2].

NOTATION. When n is negative, let

$$\sum_{p=0}^{n} m_p = \sum_{p=n}^{0} m_p = \sum_{p=0}^{|n|} m_{-p} .$$

THEOREM 7. *If* f *is a function defined on the set of all numbers and* c *is a positive number, then* $P(c, f)$ *is in* F_0 *if, and only if, for every integer* n *and integer* k,

$$\sum_{p=0}^{n+k} m_p \leqq \sum_{p=0}^{n} m_p + \sum_{p=0}^{k} m_p .$$

Proof. For each positive integer n, $m_n = [f(nc) - f(nc - c)]/c$. Hence $cm_n = f(nc) - f(nc - c)$ and $f(nc) = cm_n + f(nc - c)$. It follows by induction that

$$f(nc) = c \sum_{p=1}^{n} m_p + f(0) = c \sum_{p=0}^{n} m_p .$$

For each negative integer k, $m_k = -[f(kc + c) - f(kc)]/c$. Hence $cm_k = -f(kc + c) + f(kc)$ and $f(kc) = cm_k + f(kc + c)$. It follows by induction that

$$f(kc) = c \sum_{p=k}^{-1} m_p + f(0) = c \sum_{p=0}^{k} m_p .$$

In the proof of Theorem 6, it is shown that $P(c, f)$ is in F_0 if, and only if, for every integer n and integer k, $f(nc) + f(kc) - f((n + k)c) \geqq 0$. Hence the theorem follows.

This theorem can be used to derive several of the well-known theorems concerning the rate of growth of subadditive functions. Notice that one could easily restrict the domain of f to the positive or negative numbers. Note also the obvious fact that if f has period nc, then $\sum_{p=1}^{n} m_p = 0$.

THEOREM 8. *If f is a function defined on the set of all numbers and $\{c_n\}$ is a number sequence converging to 0 such that $\{P(c_n, f)\}$ converges pointwise to f, then f is in F_0 if, and only if, $P(c_n, f)$ is in F_0 for every positive integer n.*

Proof. From Theorem 6, it follows that if f is in F_0 and n is a positive integer, then $P(c_n, f)$ is in F_0.

Under the hypothesis of the theorem, if $P(c_n, f)$ is in F_0 for every positive integer n, then by Lemma 2, f is in F_0.

5. **Periodic polygonal subadditive functions.** A type of polygonal approximation to elements of F is described and these are shown to have a finite basis. In fact, these polygonal approximations (for a fixed n) form a function cone with finitely many extremal elements (the basis). While an explicit representation is not given, the proof of Theorem 9 shows how these extremal elements may be constructed.

THEOREM 9. *For each positive integer n, there is an integer $M(n)$ and a finite sequence $\{\alpha_{np}\}$ with $M(n)$ terms such that*

(i) *if p is an integer and $1 \leqq p \leqq M(n)$, then for some function f in F, $\alpha_{np} = P(1/n, f)$ and*

(ii) *if g is in F and h is the function $P(1/n, g)$, then there is a sequence $\{a_p\}$ of nonnegative numbers such that $h = \sum_{p=1}^{M(n)} a_p \alpha_{np}$.*

Proof. Suppose n is a positive integer. Let F' denote the collection to which h belongs if, and only if, there is a function f in F

such that $h = P(1/n, f)$. Note that if h is in F', h is continuous, $h(1 + p/n) = f(1 + p/n) = f(p/n) = h(p/n)$ for each integer p and hence, h has period 1. By Lemma 1, if $a \geqq 0$, g is in F', and h is in F', then $ag + h$ is in F'. It follows by induction that any linear combination with nonnegative coefficients of functions in F' is itself in F'.

Let Z_0 denote the collection of all points $(p/n, k/n)$ for $p = 1, 2, 3, \cdots, n$ and $k = 1, 2, 3, \cdots, n$. For each h in F', $h^*(x + 1, y) = h^*(x, y + 1) = h^*(x, y)$. Making use of part of the proof of Theorem 6, h^* is nonnegative if, and only if, h^* is nonnegative at every point of Z_0. Each point $(p/n, k/n)$ of Z_0 such that $h^*(p/n, k/n) = 0$, is called a *zero* of h^*. If h is in F', h is said to be *fundamental* only if for each function g in F' such that $h - g$ is in F', there is a nonnegative number c such that $g = ch$.

Let us now show that when h is in F', the statement that $h^* \equiv 0$ is equivalent to the statement that $h \equiv 0$. If $h \equiv 0$, then $h^* \equiv 0$. Suppose $h^* \equiv 0$. Then for every number x, $h(2x) = 2h(x)$. Therefore by induction $h(x) = (2x)/2 = \cdots = h(n2^n)/2^n$ for every positive integer n. Since h is continuous and periodic, there is a number B such that for any number w, $| h(w) | < B$. As $0 \leqq h(x) < B/2^n$ for every positive integer n, $h \equiv 0$.

Next let us show that if f is in F' and g is in F' then the statement that $f^* = g^*$ is equivalent to the statement that $f = g$. If $f = g$, then $f^* = g^*$. If $f^* = g^*$ and $h = f - g$, then $h^* = (f - g)^* = f^* - g^* \equiv 0$ but from the preceding $h \equiv 0$ and hence $f = g$.

Next let us show that the function h in F' is fundamental if, and only if, it is true that if g is in F' and every zero of h^* is a zero of g^*, then every zero of g^* is a zero of h^*.

Case 1. Suppose h is not fundamental. Then there is a function g in F such that $h - g$ is in F'; yet there is no nonnegative number c such that $g = ch$. Note that there is no zero z of h^* which is not a zero of g^* as $(h - g)^*$ would be negative at z, which is impossible by a previous result. There is a least upper bound c of all numbers d such that $h - dg$ is in F'. By Lemma 2, $h - cg$ is in F', $h - cg \not\equiv 0$ by assumption. If $h^* - cg^*$ is positive at every point z of Z_0 which is not a zero of h^*, then there is a number $d > c$ such that $h^* - dg^*$ is positive at every such point z, which would contradict the fact that c is the largest number such that $h - cg$ is in F'. Hence $h - cg$ is a function such that every zero of h^* is a zero of $(h - cg)^*$ but some zero of $(h - cg)^*$ is not a zero of h^*. Note that since Z_0 is finite, it now follows by induction that there is some fundamental function f such that every zero of h^* is a zero of f^*. For each h in F' such that h^* has a zero, let Z_h denote the set to which z belongs only if z is a zero of h^*.

Case 2. Suppose g is in F', $g \not\equiv 0$ and Z_h is a proper subset of Z_g. There is a positive number c such that the product of c and the maximum value of g^* on Z_0 is less than the smallest positive value of h^* on Z_0. Therefore $h^* - cg^*$ is nonnegative at each point of Z_0; consequently $h - cg$ is in F'. But if $h - dg$ is in F' for some number d, then, as there is some zero z of g^* which is not a zero of h^*, $h^* - dg^*$ is positive at z and $h - dg \not\equiv 0$. Therefore h is not foundamental.

Let C denote the collection to which f belongs only if f is a fundamental function such that $\sum_{p=1}^{n} f(p/n) = 1$. For each function h in F except 0, there is a function f in C such that Z_h (if it exists) is a subset of Z_f. If f is a fundamental function and g is a fundamental function such that $Z_f = Z_g$ then by a previous argument, there is a positive number c such that $f = cg$. Therefore if f and g are fundamental functions in C, $Z_f \neq Z_g$ and neither is a subset of the other. Since Z_0 is finite, C is finite. Let $M(n)$ denote the number of functions in C and arrange these functions in a sequence $\{\alpha_{np}\}$.

It has been previously shown that any linear combination with nonnegative coefficients of elements of F is itself an element of F. Hence there remains only to show that every element h of F' can be represented as a linear combination with nonnegative coefficients of the functions $\{\alpha_{np}\}$. Let $h_1 = a_1\alpha_{n1}$ where a_1 is the largest number c such that $h - c\alpha_{n1}$ is in F'. For each positive integer $p \leqq M(n)$, let $h_p = h_{p-1} + a_p\alpha_{np}$ where a_p is the largest number c such that $h - h_{p-1} - c\alpha_{np}$ is in F'.

$$h_{M(n)} = \sum_{p=1}^{M(n)} a_p\alpha_{np} .$$

Let

$$g = h - h_{M(n)} = h - \sum_{p=1}^{M(n)} a_p\alpha_{np} .$$

Unless $g \equiv 0$ there is an integer k such that Z_g is a subset of $Z_{\alpha_{nk}}$. There is a largest number d such that $g - d\alpha_{nh}$ is in F'. $h_h = h_{k-1} + a_k\alpha_{nk}$ where a_k is the largest number c such that $h - h_{k-1} - c\alpha_{nk}$ is in F'. But if $g - d\alpha_{nk}$ is in F', then $(h - h_{n-1} - c\alpha_{nk}) - d\alpha_{nk}$ is in F' which is a contradiction. Therefore $g \equiv 0$.

THEOREM 10. *There is a sequence* $\{\alpha_n\}$ *of functions in* F *such that* f *belongs to* F *if, and only if, there exists a sequence* $\{g_n\}$ *converging uniformly to* f *such that for each positive integer* n, *there is a sequence* $\{a_{np}\}$ *of nonnegative numbers such that* $g_n = \sum_{p=1}^{n} a_{np}\alpha_p$.

Proof. Let $\{\alpha_n\}$ denote a sequence of functions in F such that for each positive integer n and positive integer $p \leqq M(n)$, there is a positive integer k such that $\alpha_k = a_{np}$ where α_{np} is the pth term of the sequence $\{\alpha_{np}\}$ of Theorem 9.

If f is in F, the sequence $\{D(1/n, f)\}$ converges uniformly to f. For each positive integer k, there is an integer $n(k)$ and a sequence $\{a_{n(k)p}\}$ of nonnegative numbers such that $P(1/k, f) = \sum_{p=1}^{n(k)} a_{n(k)p}\alpha_p$ (by Theorem 9). Let $g_m = \sum_{p=1}^{n(k)} a_{n(k)p}\alpha_p$ if $n(k) \leqq m < n(k+1)$. Then $\{g_m\}$ converges uniformly to f.

By using the fact that the sum of two continuous functions with period 1, is a continuous function with period 1, Lemma 1, and induction, it follows that if for some sequence $\{a_p\}$ of nonnegative numbers and some integer n, $g_n = \sum_{p=1}^{n} a_{np}\alpha_p$, then g is in F. If $\{g_n\}$ converges uniformly to f, then f is a continuous function with period 1 and by Lemma 2, f is in F.

6. Some examples and comments.

The examples in Theorem 11 are typical of the fundamental anchored polygonal elements in F; the example in Theorem 12 shows that some functions in F are pathological.

THEOREM 11. *Suppose* $0 < k < 1$ *and* f *is the function with period* 1 *such that if* $0 \leqq x \leqq k$, $f(x) = x(1-k)$ *and if* $k \leqq x \leqq 1$, $f(x) = k(1-x)$. *Then* f *is in* F.

This theorem can be shown by computing f^*. It is quite easy to establish that $f^* \geqq 0$.

It follows from Theorem 11 and a well-known characterization of continuous convex functions on an interval that if n is a positive integer and f is a nonnegative convex function on $[0, 1/n]$, then f can be extended to be subadditive with period $1/n$. This result should appear in the Pacific Journal in a paper by Richard Laatsch using different methods (private communication).

THEOREM 12. *There exists a totally nondifferentiable function in* F.

Proof. It follows from Theorem 11 that $y = |\operatorname{Arcsin}(\sin \pi x)|$ represents a function in F. In a different setting and using a different notation, there is a proof in [1, p. 115] that the function f defined by $f(x) = \sum_{n=1}^{\infty} |\operatorname{Arcsin}(\sin 2^n\pi x)| (\pi 2^n)$ is totally nondifferentiable. By Theorem 10, f is in F.

Notice that the graph of $Z = f(\theta)\exp(i\theta)$ forms the boundary of a simple G-set with no tangent at any point.

As might be expected, when one considers G-sets which are not necessarily simple, one finds some very complicated examples. The following should illustrate this and also, should show some of the aspects of G-sets.

Suppose M is a point set and M' is the set to which Z belongs if, and only if, for some point W of M and number $0 \leqq t \leqq 1$, $Z = tW$, i.e., M' is the smallest *star-shaped* set about the origin which covers M. A slight modification of the argument for Theorem 1 would show that if M is a G-set, then M' is a G-set. Moreover, a modification of the argument for Theorem 2 would show that if M is a G-set, then M is bounded if, and only if, M is a subset of the unit disc, i.e., if Z is in M, $|Z| \leqq 1$.

That the set M' need not contain a domain can be seen by taking, for some positive integer n, M to be then nth roots of 1.

Even when M' is a bounded G-set, there is no requirement that its boundary be the graph of $Z = r(\theta) \exp(i\theta)$ for some positive continuous function with period 2π. This can be seen by taking M to be, for some number $k \neq 0$, the graph of $Z = \exp((k + i)\theta)$ for $0 \leqq \theta < 2\pi$.

Suppose R is a number set and g is a function defined on R such that if x is in R and y is in R, then $x + y$ is in R and $g(x)g(y) = g(x + y)$. Then if M denotes the set of all points $Z = g(x) \exp(ix)$ for all numbers x in R, M is a G-set. It is well-known that such sets R, and even that countable sets R, exist along with additive functions f defined on R which are dense in the plane. If $g(x) = \exp(f(x))$ then the corresponding set would be dense in the plane.

The methods of this paper may easily be applied to certain other functional inequalities. For example, analogous theorems hold for the solutions to $f(2x) \leqq 2f(x)$ and in most cases the arguments do not need to be changed.

REFERENCES

1. R. P. Boas, *A primer of real functions*, Mathematical Association of America, Carus Monographs, Vol. 13, 1960.

2. Andrew Bruckner, *Minimal superadditive extensions of superadditive functions*, Pacific J. Math., **10** (1960), 1155-1162.

3. E. Hille, and R. Phillips, *Functional analysis and semi-groups*, Amer. Math. Soc., Colloquium Publications, Vol. 31, 1957.

UNIVERSITY OF TEXAS AND
UNIVERSITY OF CALIFORNIA, RIVERSIDE

PACIFIC JOURNAL OF MATHEMATICS
Vol. 15, No. 1, 1965

WAVE OPERATORS AND UNITARY EQUIVALENCE

TOSIO KATO

This paper is concerned with the wave operators $W_\pm = W_\pm(H_1, H_0)$ associated with a pair H_0, H_1 of selfadjoint operators. Let (M) be the set of all real-valued functions ϕ on reals such that the interval $(-\infty, \infty)$ has a partition into a finite number of open intervals I_k and their end points with the following properties: on each I_k, ϕ is continuously differentiable, $\phi' \neq 0$ and ϕ' is locally of bounded variation. Theorem 1 states that, if $H_1 = H_0 + V$ where V is in the trace class T, then $W'_\pm \pm W_\pm(\phi(H_1), \phi(H_0))$ exist and are complete for any $\phi \in (M)$; moreover, $M'_\pm$ are "piecewise equal" to $W_\pm$ (in the sense to be specified in text). Theorem 2 strengthens Theorem 1 by replacing the above assumption by the condition that $\phi_n(H_1) = \phi_n(H_0) + V_n$, $V_n \in T$, where $\phi_n \in (M)$ and ϕ_n is univalent on $(-n, n)$ for $n = 1, 2, 3, \ldots$. As corollaries we obtain many useful sufficient conditions for the existence and completeness of wave operators.

1. Introduction. The present paper is a continuation of earlier papers of the author on the theory of wave and scattering operators and the related theory of unitary equivalence of selfadjoint operators.

We begin with a brief review of relevant definitions and known results (see Kato [4, 5] and Kuroda [6]), adding some new definitions for convenience. Let $\mathfrak{H}$ be a Hilbert space and let H be a selfadjoint operator in $\mathfrak{H}$ with the spectral representation $H = \int \lambda dE(\lambda)$. A vector $u \in \mathfrak{H}$ is *absolutely continuous* (*singular*) with respect to H if $(E(\lambda)u, u)$ is absolutely continuous (singular) in λ (with respect to the Lebesgue measure). The set of all $u \in \mathfrak{H}$ which are absolutely continuous (singular) with respect to H forms a subspace of $\mathfrak{H}$, which we call the *absolutely continuous* (*singular*) *subspace* with respect to H and denote by $\mathfrak{H}_{ac}(\mathfrak{H}_s)$. These two subspaces are orthogonal complements to each other and reduce H. The part of H in $\mathfrak{H}_{ac}(\mathfrak{H}_s)$ is called the *absolutely continuous* (*singular*) part of H and is denoted by H_{ac} (H_s)

Let H_j, $j = 0, 1$, be two selfadjoint operators in $\mathfrak{H}$ with the spectral representation $H_j = \int \lambda dE_j(\lambda)$, and let P_j be the projection on the absolutely continuous subspace $\mathfrak{H}_{j,ac}$ with respect to H_j. If one or both of the strong limits

$$(1.1) \qquad W_\pm = W_\pm(H_1, H_0) = s - \lim_{t\to\pm\infty} \exp(itH_1)\exp(-itH_0)P_0$$

exist(s), it is (they are) called the (*generalized*) *wave operator*(*s*).

Received January 8, 1964. This work was sponsored (in part) by Office of Naval Research Contract 222 (62).

W_+ is, whenever it exists, a partial isometry on $\mathfrak{H}$ with initial set $\mathfrak{H}_{0,ac}$ and final set $\mathfrak{M}_+$ contained in $\mathfrak{H}_{1,ac}$. $\mathfrak{M}_+$ reduces H_1, and the part of H_1 in $\mathfrak{M}_+$ is unitarily equivalent to $H_{0,ac}$, with

$$E_1(\lambda)W_+ = W_+E_0(\lambda) , \qquad -\infty < \lambda < +\infty \tag{1.2}$$

The wave operator W_+ will be said to be *complete* if the final set $\mathfrak{M}_+$ coincides with $\mathfrak{H}_{1,ac}$.

W_+ has the property that, whenever $W_+(H_1, H_0)$ and $W_+(H_2, H_1)$ exist, then $W_+(H_2, H_0)$ exists and is equal to $W_+(H_2, H_1)W_+(H_1, H_0)$. If both $W_+(H_1, H_0)$ and $W_+(H_0, H_1)$ exist, then they are complete and are the adjoints to each other.

Similar results hold for W_+ replaced by W_-.

If $H_1 - H_0$ is small in the sense that $H_1 = H_0 + V$ with V belonging to the trace class $\boldsymbol{T}$ of operators on $\mathfrak{H}$, then both $W_\pm(H_1, H_0)$ exist and are complete. The main purpose of the present paper is to prove some generalizations of this theorem, which involve what we shall call the *principle of invariance of wave operators*. Roughly speaking, this principle asserts that the wave operators $W_\pm(\phi(H_1),\phi(H_0))$ exist for an "arbitrary" function ϕ and are independent of ϕ for a wide class of functions ϕ. Its precise formulation is given in Theorems 1 and 2 proved below.

The proof of these theorems is rather simple, depending essentially on a single inequality proved in a previous paper (Kato [5]). It will be noted that Theorem 2 contains as special cases most of the sufficient conditions for the existence and completeness of wave operators obtained in recent years (see Kuroda [6, 7], Birman [1, 2], Birman-Krein [3].

2. Principle of invariance of wave operators. Consider the wave operators $W_\pm(\phi(H_1), (\phi(H_0))$ where ϕ is a real-valued, Borel measurable function on $(-\infty, +\infty)$. The principle of invariance asserts that these wave operators do not denoted on ϕ. Of course certain restrictions must be imposed on ϕ and on the relation between H_0 and H_1 To this end it is convenient to introduce a certain class of functions.

DEFINITION. A real-valued function ϕ on $(-\infty, +\infty)$ is said to be of class (M) if the whole interval $(-\infty, +\infty)$ has a partition into a finite number of open intervals I_k, $k = 1, \cdots, r$, and their end points with the following properties: on each I_k, ϕ is strictly monotone and differentiable, with the derivative ϕ' continuous, $\phi' \neq 0$ and (locally) of bounded variation. $\{I_k\}$ will be called a system of intervals associated with ϕ (such a system is not unique).

THEOREM 1. *Let H_0, H_1 be selfadjoint operators such that $H_1 =$*

$H_0 + V$ with $V \in \boldsymbol{T}$. If ϕ is of class (M), $W'_{\pm} = W_{\pm}(\phi(H_1), \phi(H_0))$ exist and are complete. Furthermore, $W'_{\pm}$ are "piecewise equal" either to $W_{\pm} = W_{\pm}(H_1, H_0)$ or to $W_{\mp}$, in the sense that

$$(W'_{\pm} - W_{\pm})E_0(I_k) = 0 \text{ or } (W'_{\pm} - W_{\mp})E_0(I_k) = 0, k = 1, \cdots, r,$$

according as ϕ is increasing or decreasing on I_k. In particular, $W'_{\pm} = W_{\pm}(W'_{\pm} = W_{\mp})$ if ϕ is increasing (decreasing) in each I_k, $k= 1, \cdots, r$. (Here $\{I_k\}$ is a system of intervals associated with $\phi \in (M)$ and $E_0(I) = E_0(\beta - 0) - E_0(\alpha)$ if $I = (\alpha, \beta)$.)

Proof. It is known (see Kato [5]) that $W_{\pm}$ exist under the assumptions of the theorem.

We take a fixed I_k and assume that ϕ is increasing on I_k. We use the inequality (2.9) of the paper cited, which reduces for $s = 0$ to

$$(2.1) \qquad \|(W_+ - 1)x\| \leqq (8\pi m^2 \| V\|_1)^{1/4} \times \left(\int_0^{+\infty} \| |V|^{1/2} \exp(-itH_0)x\|^2 dt\right)^{1/4},$$

where $x \in \mathfrak{H}_{0,ac}$ is subjected to the condition that $d(E_0(\lambda)x, x)/d\lambda \leqq m^2$ almost everywhere. Here $|V|$ is the nonnegative square roof of V^2 and $\|V\|_1$ denotes the trace norm of V, which is finite by assumption.

Now let $u \in \mathfrak{H}_{0,ac}$ be such that $E_0(I_k)u = u$ and $d(E_0(\lambda)u, u)/d\lambda \leqq m^2$. We note that such u with finite m^2 form a dense subset of $E_0(I_k)\mathfrak{H}_{0,ac} = E_0(I_k)P_0\mathfrak{H}$ (see a similar proposition in loc. cit. when I_k is the whole interval). If we set $x = \exp(-is\phi(H_0))u$, we have $(E_0(\lambda)x, x) = (E_0(\lambda)u, u)$ so that the assumptions on x stated above are satisfied. Hence (2.1) gives

$$(2.2) \qquad \|(W_+ - 1)\exp(-is\phi(H_0))u\| \leqq (8\pi m^2 \|V\|_1)^{1/4}\eta(s)^{1/4},$$

$$(2.3) \qquad \eta(s) = \int_0^{+\infty} \| |V|^{1/2} \exp(-itH_0 - is\phi(H_0))u\|^2 dt = \sum_{n=1}^{\infty} |c_n| \int_0^{+\infty} |(\exp(-itH_0 - is\phi(H_0))u, f_n)|^2 dt,$$

where $\{f_n\}$ is a complete orthonormal system of eigenvectors of V and the c_n are the associated eigenvalues.

The integrals on the right of (2.3) have the form (A1) of Appendix, where $w(\lambda)$ is to be replaced by $d(E_0(\lambda)u, f_n)/d\lambda$ which belongs to $L^2(I_k)$ with L^2-norm not exceeding m (see loc. cit.). Therefore, each term on the right of (2.3) tends to 0 for $s \to +\infty$ (Lemma A3, Appendix). On the other hand, the series on the right of (2.3) is majorized by the convergent series $2\pi m^2 \sum |c_n| = 2\pi m^2 \|V\|_1$. Hence $\eta(s) \to 0$ for $s \to +\infty$ and the left member of (2.2) must tend to 0 for $s \to +\infty$. Since $(W_+ - 1)\exp(-it\phi(H_0))$ is uniformly bounded and the set of u

with the above properties is dense in $E_0(I_k)P_0\mathfrak{H}$ as remarked above, it follows that $(W_+ - 1)\exp(-is\phi(H_0))P_0E_0(I_k) \to 0$ strongly for $s \to +\infty$. But we have $W_+ \exp(-is\phi(H_0)) = \exp(-is\phi(H_1))W_+$ by (1.2). On multiplying the above result from the left with $\exp(is\phi(H_1))$, we thus obtain

$$(2.4) \qquad s - \lim_{s\to+\infty} \exp(is\phi(H_1))\exp(-is\phi(H_0))P_0E_0(I_k)$$
$$= W_+P_0E_0(I_k) = W_+E_0(I_k) \qquad \text{if } \phi \text{ is increasing on } I_k .$$

Similarly we can show that

$$(2.4') \qquad s - \lim_{s\to+\infty} \exp(is\phi(H_1))\exp(-is\phi(H_0))P_0E_0(I_k) = W_-E_0(I_k)$$
$$\text{if } \phi \text{ is decreasing on } I_k.$$

Since $P_0E_0(\lambda)$ is continuous in λ, we have $\sum_k P_0E_0(I_k) = P_0$. Adding (2.4) or (2.4′) for $k = 1, \cdots, r$, we thus arrive at the result

$$(2.5) \qquad s - \lim_{s\to+\infty} \exp(is\phi(H_1))\exp(-is\phi(H_0))P_0 = \sum_{k=1}^{r} W_{(\pm)}E_0(I_k) ,$$

where $W_{(\pm)}$ means that $W_+(W_-)$ should be taken if ϕ is increasing (decreasing) on I_k.

(2.5) shows that the wave operator $W_+(\phi(H_1), \phi(H_0))$ exists and is equal to the right member; it should be noted that the absolutely continuous subspace for $\phi(H_0)$ is identical with $\mathfrak{H}_{0,ac} = P_0\mathfrak{H}$ (Lemma A5, Appendix). Similar results hold for $W_-(\phi(H_1), \phi(H_0))$; we have only to take the opposite choice for $W_{(\pm)}$ in (2.5). These wave operators are complete since they also exist when H_0 and H_1 are exchanged.

3. Generalization. Let us consider a question which is in a sense converse to Theorem 1. Suppose $\psi(H_1) - \psi(H_0)$ belongs to $\boldsymbol{T}$ for some function ψ; then do the wave operators $W_\pm(H_1, H_0)$ exist?

The answer to this question is quite simple if ψ is of class (M) and, in addition, *univalent*. Then the inverse function exists, with domain Δ consisting of a finite number of open intervals and a finite number of points. This inverse function can be extended to a function $\hat{\psi}$ of class (M) by setting, for example, $\hat{\psi}(\lambda) = \lambda$ on the complement of Δ. Therefore, $W_\pm(H_1, H_0) = W_\pm(\hat{\psi}(\psi(H_1)), \hat{\psi}(H_0)))$ exist and are complete by Theorem 1.

If ψ is not univalent, we do not know whether the same resalts hold. But we can show that this is true if there is an *approximate univalent sequence* $\{\psi_n\}$ of functions of class (M) such that $\psi_n(H_1) - \psi_n(H_0) \in \boldsymbol{T}$. We call $\{\psi_n\}$ an approximate univalent sequence if ψ_n is univalent on $(-n, n)$, $n = 1, 2, \cdots$

More generally, we can prove

THEOREM 2. *Let H_0, H_1 be selfadjoint and let there exist an approximate univalent sequence $\{\psi_n\}$ of functions of class (M) such that $\psi_n(H_1) = \psi_n(H_0) + V_n$ with $V_n \in \boldsymbol{T}, n = 1, 2, \cdots$ Then, for any $\phi \in (M)$, the wave operators $W'_\pm = W_\pm(\phi(H_1), \phi(H_0))$ exist and are complete. In particular, $W_\pm = W_\pm(H_1, H_0)$ exist and are complete. $W'_\pm$ are piecewise equal either to $W_\pm$ or to $W_\mp$ in the sense stated in Theorem 1.*

Proof. I. The restriction of ψ_n to $(-n, n)$ has inverse function, which can be extended to a $\hat{\psi}_n \in (M)$ in the same way as above.

Set $\phi_n = \phi \circ \hat{\psi}_n \circ \psi_n$; then $\phi_n(\lambda) = \phi(\lambda)$ for $\lambda \in (-n, n)$, and $\phi_n \in (M)$ by Lemma A4 (Appendix). We define the following selfadjoint operators, all functions of H_j, $j = 0, 1$:

$$
\begin{aligned}
&\psi_n(H_j) = L_{nj} , &&(\hat{\psi}_n \circ \psi_n)(H_j) = H_{nj} , \\
&\phi_n(H_j) = K_{nj} = \int \lambda dF_{nj}(\lambda) , &&\phi(H_j) = K_j = \int \lambda dF_j(\lambda) .
\end{aligned} \tag{3.1}
$$

Since $K_{nj} = (\phi \circ \hat{\psi}_n)(L_{nj})$ by operational calculus (see Stone [8], Theorem 6.9), where $\phi \circ \hat{\psi}_n \in (M)$ and $L_{n1} = L_{n0} + V_n$, $V_n \in \boldsymbol{T}$, it follows from Theorem 1 that $W'_{n\pm} = W_\pm(K_{n1}, K_{n0})$ exist and are complete.

II. For any function ψ of class (M), $\psi(\pm\infty) = \lim_{\lambda \to \pm\infty} \psi(\lambda)$ exist (the values $\pm\infty$ being permitted for these limits). Thus $\phi_n(\pm\infty)$ and $(\hat{\psi}_n \circ \psi_n)(\pm\infty)$ exist. By replacing $\{\phi_n\}$ by a suitable subsequence (and correspondingly for $\{\psi_n\}$ and $\{\hat{\psi}_n\}$), we may assume that $\alpha_\pm \lim_{n\to\infty}\phi_n(\pm\infty)$ and $\beta_\pm = \lim_{n\to\infty} (\hat{\psi}_n \circ \psi_n)(\pm\infty)$ exist ($\pm\infty$ being permitted for these limits).

Let J be an open interval such that $\alpha_\pm$ and $\phi(\pm\infty)$ are exterior to J, and let $S = \phi^{-1}(J)$, $S_n = \phi_n^{-1}(J)$. S and S_n are unions of a finite number of open intervals and of points. Since $K_j\phi(H_j)$ and $K_{n_j} = \phi_n(H_j)$, we have (we denote by $E_j(S)$ the spectral measure determined from $\{E_j(\lambda)\}$)

$$
F_j(J) = E_j(S) , \qquad F_{nj}(J) = E_j(S_n) , \qquad j = 0, 1. \tag{3.2}
$$

S is bounded since $\phi(\pm\infty)$ are exterior to J. Similarly, S_n is bounded if n is sufficiently large, since $\alpha_\pm$ are exterior to J.

Take an n so large that S_n is bounded and $S \subset (-n, n)$. Since $\phi_n(\lambda) = \phi(\lambda)$ for $\lambda \in (-n, n)$, we have $S = (-n, n) \cap S_n$. Further take an $m > n$ such that $S_n \subset (-m, m)$. We have $S = (-m, m) \cap S_m$ as above, so that $S_m \cap S_n = S_m \cap (-m, m) \cap S_n = S \cap S_n = S$. Hence

$$
\begin{aligned}
F_{nj}(J)F_{mj}(J) &= F_j(S_n)E_j(S_m) \\
&= E_j(S_n \cap S_m) = E_j(S) = F_j(J) .
\end{aligned} \tag{3.3}
$$

III. Now we have, for any $u \in \mathfrak{H}_{0,ac} = P_0\mathfrak{H}$,

$$(3.4)\qquad \begin{aligned} &\exp(itK_{n1})(1-F_{n1}(J))\exp(-itK_{n0})P_0F_0(J)\\ &=(1-F_{n1}(J))\exp(itK_{n1})\exp(-itK_{n0})P_0F_0(J)\\ &\to(1-F_{n1}(J))W'_{n+}F_0(J) \qquad \text{strongly for } t\to+\infty\,. \end{aligned}$$

Since $(1-F_{n1}(J))W'_{n+}=W'_{n+}(1-F_{n0}(J))$ by (1.2) applied to W'_{n+}, and since $F_0(J)\leqq F_{n0}(J)$ by (3.3), the last member of (3.4) vanishes. On the other hand $\exp(-itK_{n0})F_0(J)=\exp(-itK_0)F_0(J)$ since $\phi_n(\lambda)=\phi(\lambda)$ for $\lambda\in(-n,n)$ and $F_0(J)=E_0(S)\leqq E_0((-n,n))$. On multiplying (3.4) from the left by $\exp(-itK_{n1})$, we thus obtain

$$(3.5)\qquad s-\lim_{t\to+\infty}(1-F_{n1}(J))\exp(-itK_0)P_0F_0(J)=0\,.$$

The same is true when n is replaced by the $m>n$ considered above. Now multiply the latter from the left by $F_{n1}(J)$ and add to (3.5). In view of (3.3), we then obtain

$$(3.6)\qquad s-\lim_{t\to+\infty}(1-F_1(J))\exp(-itK_0)P_0F_0(J)=0\,.$$

Multiply again (3.6) from the left by $\exp(itK_1)$; then

$$(3.7)\qquad \begin{aligned} &s-\lim_{t\to+\infty}\exp(itK_1)\exp(-itK_0)P_0F_0(J)\\ &=s-\lim_{t\to+\infty}F_1(J)\exp(itK_{n1})\exp(-itK_{n0})P_0F_0(J)\\ &=F_1(J)W'_{n+}F_0(J)\,, \end{aligned}$$

where we have again used the relation

$$\exp(-itK_0)F_0(J)=\exp(-itK_{n0})F_0(J)$$

and similarly $\exp(itK_1)F_1(J)=\exp(itK_{n1})F_1(J)=F_1(J)\exp(itK_{n1})$.

(3.7) shows that $\lim_{t\to+\infty}\exp(itK_1)\exp(-itK_0)u$ exists and is equal to $F_1(J)W'_{n+}u$ whenever u belongs to $P_0F_0(J)\mathfrak{H}$, where J is any interval with the four points $\alpha_\pm$ and $\phi(\pm\infty)$ in its exterior. Since such u forms a dense set in $P_0\mathfrak{H}$, the existence of $W'_+=W_+(K_1,K_0)$ has been proved. The existence of W'_- can be proved in the same way. Since K_0 and K_1 can be exchanged, all these wave operators are complete.

Incidentally, it follows from (3.7) that $W'_+u=F_1(J)W'_{n+}u$ for $u\in P_0F_0(J)\mathfrak{H}$. But $\|W'_+u\|=\|u\|=\|W'_{n+}u\|$ since W'_+ and W'_{+n} are isometric on $P_0\mathfrak{H}$. Since $F_1(J)$ is a projection, we must have $W'_+u=W'_{n+}u$. Similar result holds for W'_-. Thus

$$(3.8)\qquad (W'_\pm-W'_{n\pm})F_0(J)=0\,.$$

Note that this is true for sufficiently large n (depending on J).

IV. To prove the piecewise equality of $W'_\pm$ and $W_\pm$ or $W_\mp$, let I_k be one of the intervals associated with $\phi\in(M)$. We may assume that $\phi' > 0$ on I_k; we have to show that $(W'_\pm - W_\pm)E_0(I_k) = 0$. For this it suffices to show that $(W'_\pm - W_\pm)E_0(I) = 0$ for any finite subinterval I of I_k; we may further assume that $\beta_\pm$ are exterior to I and $\alpha_\pm$, $\phi(\pm\infty)$ are exterior to the interval $\phi(I)$.

We set $J = \phi(I)$ and apply the preceding results to J. Since $S = \phi^{-1}(J) \supset I$, we have $E_j(I) \leqq E_j(S) = F_j(J)$ and hence by (3.8)

$$(W'_\pm - W'_{n\pm})E_0(I) = 0 \tag{3.9}$$

for sufficiently larg n.

We have similar results when $\phi(\lambda)$ is replaced by the identity function λ (since $\beta_\pm$ and $\pm\infty$ are exterior to I). Then $W'_\pm$, $W'_{n\pm}$ are to be replaced respectively by $W_\pm = W_\pm(H_1, H_0)$ and $W_{n\pm} = W_\pm(H_{n1}, H_{n0})$. Thus

$$(W_\pm - W_{n\pm})E_0(I) = 0 \tag{3.10}$$

for sufficiently large n.

We may assume that n is so large that $I \subset (-n, n)$. I can be expressed as the union of a finite number of subintervals $\varDelta_p$ (and a finite number of points) in each of which ψ_n is monotonic. Then $\hat{\psi}_n$ is monotonic on $\varDelta'_p = \psi_n(\varDelta_p)$ since ψ_n is univalent on $(-n, n)$. $\phi \circ \hat{\psi}_n$ is also monotonic on $\varDelta'_p$ since $\phi' > 0$ on $\hat{\psi}_n(\varDelta'_p) = \varDelta_p$; it is increasing or decreasing with $\hat{\psi}_n$. Since $K_{nj} = (\phi \circ \hat{\psi}_n)(L_{nj})$, $H_{nj} = \hat{\psi}_n(L_{nj})$ and $L_{n1} = L_{n0} + V_n$, $V_n \in \boldsymbol{T}$, it follows from Theorem 1 that $(W'_{n\pm} - W_{n\pm})E_0(\varDelta_p) = 0$; note that $E_0(\varDelta_p) \leqq E_0(\psi_n^{-1}(\varDelta'_p)) = G_0(\varDelta'_p)$ where $\{G_0(\lambda)\}$ is the resolution of the identity for $L_{n0} = \psi_n(H_0)$. Adding the results obtained for $p = 1, 2, \cdots$, we have

$$(W'_{n\pm} - W_{n\pm})E_0(I) = 0\,. \tag{3.11}$$

The desired result $(W'_\pm - W_\pm)E_0(I) = 0$ follows from (3.9), (3.10) and (3.11).

4. **Applications.** A number of sufficient conditions for the existence and completeness of wave operators can be deduced from Theorem 1 or 2. We shall mention only a few.

(a) Let neither H_0 nor H_1 have the eigenvalue 0. If $H_1^{-p} = H_0^{-p} + V$ with $V \in \boldsymbol{T}$ for some *odd* integer p, then $W_\pm(\phi(H_1), \phi(H_0))$ exist and are complete for any $\phi \in (M)$.

The proof follows by applying Theorem 2 with $\psi_n = \psi$ (independent of n) where $\psi(\lambda) = \lambda^{-p}$ for $\lambda \neq 0$ and $\psi(0) = 0$.

(b) In (a) we may allow *even* integers p if we assume in addition that H_0 and H_1 are nonnegative.

In this case we need only to replace the above ψ by $\psi(\lambda) = (\operatorname{sign}\lambda)\,|\lambda|^{-p}$ for $\lambda \neq 0$.

(c) Let $(H_1 - \zeta)^{-1} - (H_0 - \zeta)^{-1} \in \boldsymbol{T}$ for some nonreal complex number ζ. Then $W_\pm(\phi(H_1), \phi(H_0))$ exist and are complete for any $\phi \in (M)$.

For the proof we first note that, if the assumption is true for some $\zeta = \zeta_0$, then it is true also for all nonreal ζ. This can be seen first for $|\zeta - \zeta_0| < |\operatorname{Im}\zeta_0|$ by considering the Neumann series for the resolvents. The result can then be extended to all ζ of the half-plane $(\operatorname{Im}\zeta)(\operatorname{Im}\zeta_0) > 0$ by a standard procedure. The other half-plane can be taken care of by considering the adjoints.

Set now $\psi_n(\lambda) = -i[(n - i\lambda)^{-1} - (n + i\lambda)^{-1}] = 2\lambda(n^2 + \lambda^2)^{-1}$. It follows from the above remark that $\psi_n(H_1) - \psi_n(H_0) \in \boldsymbol{T}$. But it is easy to see that $\{\psi_n\}$ is an approximate univalent sequence of functions of class (M). Hence the proposition follows by Theorem 2.

(b) It should be remarked that the existence of $W_\pm(\phi(H_1), \phi(H_0))$ implies the existence of

$$s - \lim_{n\to\pm\infty} U_1^n U_0^{-n} = W_\pm(H_1, H_0)\,, \tag{4.1}$$

where $U_j = (H_j - i)(H_j + i)^{-1}$ is the Cayley transform of H_j. In fact, $U_j = \exp(i\phi(H_j))$ where $\phi(\lambda) = -2 \operatorname{arccot}\lambda$, and ϕ belongs to (M), being strictly increasing on $(-\infty, +\infty)$.

Appendix. We prove here some lemmas which are used in the text.

LEMMA A1. *Let f, g be complex-valued, continuous functions on a closed interval $[a, b]$. Let f be of bounded variation with total variation V_f. Let $G(\lambda) = \int_a^\lambda g(\lambda)d\lambda$ and let $M_G = \max|G(\lambda)|$, $M_f = \max|f(\lambda)|$. Then $\left|\int_a^b f(\lambda)g(\lambda)d\lambda\right| \leqq (M_f + V_f)M_G$.*

The proof is simple and will be omitted.

LEMMA A2. *Let ϕ be a real-valued differentiable function on $[a, b]$ such that the derivative ϕ' is continuous, positive and of bounded variation. We have for any $t, s > 0$*

$$\left|\int_a^b \exp(it\lambda - is\phi(\lambda))d\lambda\right| \leqq \frac{2(c + V_{\phi'})}{c(t + cs)}\,,$$

where $c = \min\phi'(\lambda) > 0$ and $V_{\phi'}$ is the total variation of ϕ'.

Proof. The integral in question is equal to

$$\int_a^b i(t + s\phi'(\lambda))^{-1}(d/d\lambda) \exp(-it\lambda - is\phi(\lambda))d\lambda .$$

We apply Lemma A1 to estimate this integral, setting $f(\lambda) = i(t + s\phi'(\lambda))^{-1}$ and $g(\lambda) = (d/d\lambda) \exp(-it\lambda - is\phi(\lambda))$. Then $M_f = (t + cs)^{-1}$, $M_G \leqq 2$ and it is easily seen that $V_f \leqq sV_{\phi'}/(t + cs)^2 \leqq V_{\phi'}/c(t + cs)$. This proves the desired inequality.

LEMMA A3. *Let ϕ be of class (M) with an associated system of of intervals $\{I_k\}$ (see definition in text). For a fixed k, let $w \in L^2(I_k)$. If ϕ is increasing on I_k, we have*

$$\text{(A1)} \qquad \int_0^{+\infty} dt \left| \int_{-\infty}^{+\infty} \exp(-it\lambda - is\phi(\lambda))w(\lambda)d\lambda \right|^2 \longrightarrow 0 , \qquad s \to +\infty .$$

If ϕ is decreasing on I_k, (A1) is true if $\int_0^{+\infty} dt$ is relpaced by $\int_{-\infty}^0 dt$.

Proof. We may assume that $w \in L^2(-\infty, +\infty)$, on setting $w(\lambda) = 0$ for λ outside I_k. Let H be the selfadjoint operator $Hu(\lambda) = \lambda u(\lambda)$ acting in $L^2(-\infty, +\infty)$, and let U be the unitary operator defined by the Fourier transformation. The inner integral of (A1) represents the function $(U \exp(-is\phi(H))w(t)$, and the left member of (A1) is equal to $\| EU \exp(-is\phi(H))w \|^2$, where E is the projection of $L^2(-\infty, +\infty)$ onto the subspace consisting of all functions that vanish on $(-\infty, 0)$. Thus (Al) is equivalent to that $EU \exp(-is\phi(H))w \to 0$, $s \to +\infty$. Since $EU \exp(-is\phi(H))$ is uniformly bounded with norm $\leqq 1$, it suffices to prove (A1) for all w belonging to a fundamental subset of $L^2(I_k)$. Thus we may restrict ourselves to considering only characteristic functions w of closed finite subintervals $[a, b]$ of I_k.

Assume that ϕ is increasing on I_k. If we denote by $v_s(t)$ the inner integral of (A1) for the characteristic function w of $[a, b] \subset I_k$, we have by Lemma A2

$$|v_s(t)| \leqq \frac{2(c + V_{\phi'})}{c(t + cs)} \text{ so that } \int_0^{+\infty} |v_s(t)|^2 dt \leqq \frac{4(c + V_{\phi'})^2}{c^3 s} \longrightarrow 0$$

for $s \to +\infty$, where c is the minimum of $\phi'(\lambda)$ on $[a, b]$ and $V_{\phi'}$ is the total variation of ϕ' on $[a, b]$. A similar proof applies to the case $\phi' < 0$ on I_k, with $\int_0^{+\infty} dt$ replaced by $\int_{-\infty}^0 dt$.

LEMMA A4. *Let ϕ, ψ be of class (M). Then the composed function $\phi \circ \psi$ also belongs to (M), and there exists a system of intervals associated with $\phi \circ \psi$ such that, in each interval of the system, both ψ and $\phi \circ \psi$ are monotonic.*

Proof. Let $\{I_k\}$ and $\{J_h\}$ be systems of intervals associated with ϕ and ψ, respectively. For each h, ψ maps J_h one-to-one onto an open interval J'_h. Let J_{kh} be the inverse image under this map of $J'_h \cap h_k$. Obviously all J_{kh} are open and mutually disjoint, and cover the whole interval $(-\infty, +\infty)$ except for a finite number of points. It is easy to see that $\phi\circ\psi$ is monotonic and continuously differentiable on each J_{kh}, with $(\phi\circ\psi)'(\lambda) = \phi'(\psi(\lambda))\psi'(\lambda)$. Furthermore, $(\phi\circ\psi)'$ is locally of bounded variation on J_{kh}, for the same is true with ϕ' and ψ' by assumption. The intervals J_{kh} form a system stated in the lemma.

LEMMA A5. *Let ϕ be of class (M). For any selfadjoint operator H, the absolutely continuous subspace for $\phi(H)$ is identical with the absolutely continuous subspace for H.*

Proof. Let $H = \int \lambda dE(\lambda)$, $\phi(H) = \int \lambda dF(\lambda)$ be the spectral representations of the operators considered. We denote by $E(S)$, $F(S)$ the spectral measures constructed from $\{E(\lambda)\}$, $\{F(\lambda)\}$, respectively. For any Borel subsets S of the real line, we have $F(S)=E(\phi^{-1}(S))$. If $|S|=0$ (we denote by $|S|$ the Lebesgue measure of S), then $|\phi^{-1}(S)| = 0$ by the properties of $\phi\in(M)$, so that $F(S)u = 0$ if u is absolutely continuous with respect to H. On the other hand, $F(\phi(S)) = E(\phi^{-1}(\phi(S)) \geqq E(S)$. If $|S|=0$, we have $|\phi(S)|=0$ so that $\|E(S)u\| \leqq \|F(\phi(S))u\| = 0$ if u is absolutely continuous with respect to $\phi(H)$. This proves the lemma.

BIBLIOGRAPHY

1. M. Sh. Birman, *On the conditions for the existence of wave operators,* Doklady Akad. Nauk USSR **143** (1962), 506-509.

2. ———, *A test for the existence of wave operators,* Doklady Akad. Nauk USSR **147** (1962), 1008-1009.

3. M. Sh. Birman, and M.G. Krein, *On the theory of wave and scattering operators,* Doklady Akad. Nauk USSR **144** (1962), 475-478.

4. T. Kato, *On finite-dimensional perturbations of self-adjoint operators,* J. Math. Soc. Japan **9** (1957), 239-249.

5. ———, *Perturbation of continuous spectra by trace class operators,* Proc. Japan Acad. **33** (1957), 260-264.

6. S. T. Kuroda, *On the existence and the unitary property of the scattering operators,* Nuovo Cimento **12** (1959), 431-454.

7. ———, *Perturbation of continuous spectra by unbounded operators, I and II,* J. Math. Soc. Japan **11** (1959), 247-262; **12** (1960), 243-257.

8. M. H. Stone, *Linear transformations in Hilbert space and their applications to analysis,* AMS Colloq. Publ., 1932.

PACIFIC JOURNAL OF MATHEMATICS
Vol. 15, No. 1, 1965

INFINITE SUMS IN ALGEBRAIC STRUCTURES

PAUL KATZ AND ERNST G. STRAUS

The purpose of this note is an outline of an algebraic theory of summability in algebraic structures like abelian groups, ordered abelian groups, modules, and rings. "Infinite sums" of elements of these structures will be defined by means of homomorphisms satisfying some weak requirements of permanency which hold in all usual linear summability methods. It will turn out that several elementary well known theorems from the theory of infinite series, proved ordinarily by methods of analysis, (i.e. by use of some concept of a limit) are consequences of algebraic properties.

1. Definitions and existence theorems. Let G be an abelian group with a ring T operating from the left; we assume, without loss of generality, that T contains the integers. Denote by G^ω the strong direct sum of countably many copies of G, i.e., the set of all infinite sequences $s = (g_i)_{i=1}^\infty = (g_1, g_2, \cdots, g_i, \cdots)$ of elements of G, with the natural definitions of addition and of left multiplication by elements of T. Let Γ be the weak direct sum of countably many copies of G, i.e., the subgroup of G^ω consisting of all infinite sequences with at most a finite number of coordinates different from 0 (the neutral element of G). For $s=(g_1, g_2, \cdots, g_i, \cdots) \in G^\omega$, denote by s' the element $(0, g_1, g_2, \cdots, g_{i-1}, g_i, \cdots)$; s' will be called the translate of s.

DEFINITION 1. The T-subgroup S of G^ω will be called admissible if

$$(1) \qquad \Gamma \subset S$$

and if

(2) $s \in S$ if and only if $s' \in S$, where s' is the translate of s.

Obviously, both Γ and G^ω are admissible, and any subset K of G^ω can be completed in a unique way to a minimal admissible subgroup containing K.

DEFINITION 2. Let S be admissible, and φ a T-homomorphism $S \to G$ with the following properties:

$$(3) \qquad \varphi(g, 0, 0, \cdots) = g\,, \qquad (g \in G)$$

and

$$(4) \qquad \varphi(s) = \varphi(s')\,, \qquad (s \in S)\,.$$

Received August 21, 1963, and in revised form February 11, 1964.

Then φ will be called a *summation method on G with domain S*, and we shall refer to it briefly as the *summation method* $[S, \varphi]$.

Using the fact that S is admissible, and by properties (3) and (4) of the homomorphism φ, it follows immediately that

$$\varphi(g_1, g_2, \cdots, g_n, 0, 0, \cdots) = \sum_{i=1}^{n} g_i$$

for any summation method $[S, \varphi]$. Therefore, the (unique) summation method $[\Gamma, \varphi]$ shall be called *trivial*. Furthermore, by (3), φ is always a T-homomorphism *onto* G.

We ask first the following question:

When does there exist a summation method containing in its domain a given $s \in G^{\omega}$?

Denote by $s^{(n)}$ for any integer n the nth translate of s, i.e.

$$s^{(0)} = s$$

$$s^{(n)} = (s^{(n-1)})' = (\underbrace{0, \cdots, 0}_{n}, g_1, g_2, \cdots) \qquad \text{for } n > 0$$

$$s^{(n)} = (g_{-n+1}, g_{-n+2}, \cdots) \qquad \text{for } n < 0 .$$

Let S be the minimal admissible subgroup of G^{ω} containing s. The elements of S have the form

$$\sum_{k=-m}^{n} t_k s^{(k)} + \gamma , \tag{5}$$

where m, n are nonnegative integers, the t_k are arbitrary elements of T, and γ is an element of Γ. This representation of an element of S is not necessarily unique, and a T-homomorphism $\varphi: S \to G$ has to be independent of it. But, since all summation methods agree on Γ, one has to answer first the question when an expression of type (5) will be in Γ. We may evidently assume $m = 0$ (changing s if necessary); hence we shall study linearly independent expressions of type

$$l_i = \sum_{k=0}^{i} t_{ik} s^{(k)} = \gamma_i \in \Gamma \qquad (i = 0, 1, 2, \cdots) \tag{6}$$

where $t_{ik} \in T$. For each i, the coefficients t_{ii} appearing in (6) form a left ideal T_i of T, and $T_0 \subset T_1 \subset \cdots$

We now *assume that T satisfies an ascending chain condition*, so that each T_i is finitely generated, and that there exists an index m such that $T_m = T_{m+1} = \cdots$. Let t_{ii}^{j} $(j = 1, \cdots, n_i)$ be a system of generators of T_i $(i = 1, \cdots, m)$. Then a finite system of equations

$$(6') \qquad l_i^j = \sum_{k=0}^{i} t_{ik}^j s^{(k)} = \gamma_i^j \in \Gamma \,, \qquad j = 1, \cdots, n_i; \quad i = 1, \cdots, m$$

implies all equations of (6) in the sense that each l_i in (6) is a linear combination over T of the l_i and their translates.

A summation method φ on G with domain S exists if and only if $\varphi(s)$ satisfies all the equations

$$(6'') \qquad \left(\sum_{k=0}^{i} t_{ik}^j\right)\varphi(s) = \varphi(\gamma_i^j) \,, \qquad j = 1, \cdots, n_i; \quad i = 1, \cdots, m$$

where the right side is independent of φ, since on Γ the homomorphism φ is the ordinary sum of finitely many elements of G. Once $\varphi(s)$ is determined it extends by linearity (over T) to all of S.

This may be generalized easily for any finite number of elements $s_1, s_2, \cdots, s_r$. Assume a summation method φ defined for the minimal admissible subgroup Γ_1 containing s_1. We can now obtain a finite system of relations of the type (6′), with s replaced by s_2, and Γ by Γ_1. This leads to a system of necessary and sufficient conditions for $\varphi(s_2)$ compatible with $\varphi(s_1)$, which is analogous to (6″) (the right side there being already defined by the previous step). Proceed by induction.

As a consequence we can prove the following existence theorem:

THEOREM 1. *For any abelian group $G \neq \{0\}$ with ring of operators T satisfying an ascending chain condition, there exists a nontrivial summation method.*

Proof. Let $g \in G$ be $\neq 0$. Define $s = (g_n)_{n=1}^{\infty}$ by

$$g_n = \begin{cases} g \text{ if } n = 2^k \\ 0 \text{ otherwise} . \end{cases}$$

Let S be the minimal admissible subgroup containing s, and $\bar{g}$ any element of G such that $tg = 0$ implies $t\bar{g} = 0$ for all $t \in T$ (for example $\bar{g} = g$). Then obviously the only relations of type (6) are of the form $ts = 0$ (because $tg = 0$), so that (6″) reduces to $t\varphi(s) = 0$ whenever $tg = 0$. These conditions are satisfied by setting $\varphi(s) = \bar{g}$.

REMARK 1. From the $2^{\aleph_0}$ sequences in G^{ω} whose elements are g or 0 one can pick a subset R, of power $2^{\aleph_0}$ so that any relation $\sum_{j=0}^{m}\sum_{i=1}^{n} t_{ij} r_i^{(j)} \in \Gamma$ for elements t_{ij} of T and $r_i \in R$ implies $t_{ij} g = 0$ for all t_{ij}. Thus we can define $2^{2^{\aleph_0}}$ different summation methods for the least admissible S which contains R by setting $\varphi(r)$ to be 0 or g arbitrarily for each $r \in R$, and then extending φ to all of S by linearity (over T).

On the other hand, in a nontrivial group no summation method can assign a sum to all the sequences of elements of the group.

THEOREM 2. *Let $G \neq 0$ be a T-group and $g_i \in G$, $(i = 1, \cdots, n)$ such that $\sum_{i=1}^{n} g_i \neq 0$. Then there exists no summation method defined for*

$$s = (g_1, g_2, \cdots, g_n, g_1, g_2, \cdots, g_n, g_1, \cdots) .$$

Proof. $s^{(n+1)} - s = (g_1, g_2, \cdots, g_n, 0, 0, \cdots)$ would lead to

$$\varphi(s^{(n+1)} - s) = \varphi(s^{(n+1)}) - \varphi(s) = \varphi(s) - \varphi(s) = 0 = g_1 + g_2 + \cdots + g_n ,$$

a contradiction.

THEOREM 3. *If $\varphi_1, \varphi_2, \cdots, \varphi_n$ are summation methods on G with domain S, and $e_1, e_2, \cdots, e_n$ are T-endomorphisms of G so that $e_1 + e_2 + \cdots + e_n = 1$, then $e_1\varphi_1 + e_2\varphi_2 + \cdots + e_n\varphi_n$ is a summation method on G with domain S.*

Proof. Let $\varphi = e_1\varphi_1 + e_2\varphi_2 + \cdots + e_n\varphi_n$. Then φ is obviously a T-homomorphism $S \to G$. Since $\varphi_i(s') = \varphi_i(s)$, the same is true for φ, and for a $g \in G$ we have $\varphi(g, 0, 0, \cdots) = g$.

THEOREM 4. *Let $[S_1, \varphi_1]$, $[S_2, \varphi_2]$ be two summation methods on G which agree on $D_0 = S_1 \cap S_2$. Then there is a summation method φ on G with domain $S = S_1 + S_2$, such that $\varphi \mid S_i = \varphi_i$ for $i = 1, 2$.*

Proof. The group S is evidently admissible. Denote $D_i = (S_i \backslash D_0) \cup \{0\}$, $i = 1, 2$. Then any $s \in S$ can be written (not necessarily uniquely)

$$(7) \qquad s = d_0 + d_1 + d_2, \ d_i \in D_i, \ i = 0, 1, 2 .$$

Define φ by

$$\varphi(s) = \varphi_1(d_0) + \varphi_1(d_1) + \varphi_2(d_2) .$$

This definition is independent of the representation (7), since if $s = \bar{d}_0 + \bar{d}_1 + \bar{d}_2$ with $\bar{d}_i \in D_i$, then $A = \varphi(d_0 + d_1 + d_2) - \varphi(\bar{d}_0 + \bar{d}_1 + \bar{d}_2) = \varphi_1(d_0 - \bar{d}_0) + \varphi_1(d_1 - \bar{d}_1) + \varphi_2(d_2 - \bar{d}_2)$. The element $d_2 - \bar{d}_2$ is in S_2, but since $d_2 - \bar{d}_2 = \bar{d}_0 - d_0 + \bar{d}_1 - d_1$, it is in D_0, and therefore $\varphi_2(d_2 - \bar{d}_2) = \varphi_2(\bar{d}_0 - d_0 + \bar{d}_1 - d_1)$. Hence $A = \varphi_1(d_0 - \bar{d}_0) + \varphi_1(d_1 - \bar{d}_1) + \varphi_1(\bar{d}_0 - d_0 + \bar{d}_1 - d_1) = 0$. A similar reasoning is needed in order to show that $\varphi(s + \bar{s}) = \varphi(s) + \varphi(\bar{s})$ for s, $\bar{s} \in S$, since the sum of two representations of type (7) is generally not of the same type. Property (3) of φ is obvious, since $\Gamma \subset D_0$, and (4) follows easily, since (7) implies $s' = d'_0 + d'_1 + d'_2$, where $d'_i \in D_i$, $i = 0, 1, 2$. Since the decomposition (7) can be extended to ts, φ is a T-homomorphism, which finishes the proof.

REMARK 2. On the other hand, if $[S_1, \varphi_1]$ and $[S_2, \varphi_2]$ are summation methods which do not agree on $S_1 \cap S_2$, then there need not exist a summation method for the admissible subgroup $S_1 + S_2$. Take S_1 and φ_1 as S and φ in Theorem 1, and define $s_2 = (g_n^*)_{n=1}^{\infty}$ by

$$g_n^* = \begin{cases} 0 \text{ if } n = 2^k \\ g \text{ otherwise} . \end{cases}$$

Again, if $\bar{\bar{g}}$ is any element of G such that $tg = 0$ implies $t\bar{\bar{g}} = 0$ for any $t \in T$, then $\varphi_2(s_2) = \bar{\bar{g}}$ is a valid definition that can be extended to a summation method on the minimal admissible subgroup S_2 containing s_2. But $S_1 + S_2$ can not be the domain of any summation method, since it contains the element $(g, g, g, \cdots)$, in contradiction to the construction in Theorem 2.

REMARK 3. Let $(G_\alpha)_{\alpha \in A}$, where A is a set of indices, be a family of abelian groups with operators T; assume that S_α is an admissible subgroup of G_α^ω and that φ_α is a summation method on G_α with domain S_α for each $\alpha \in A$. Consider the (weak or strong) direct sum $G = \bigoplus_{\alpha \in A} G_\alpha$. Then it is easily shown that $S = \bigoplus_{\alpha \in A} S_\alpha$ is admissible for G, and that $\varphi = (\varphi_\alpha)_{\alpha \in A}$ is a summation method with domain S on G. It is clear that $[S, \varphi]$ is nontrivial if and only if at least one of the summation methods $[S_\alpha, \varphi_\alpha]$ is nontrivial.

2. Subgroups and ideals. To each subgroup H of G we associate the (left) annihilator ideal T_H of T consisting of all $t \in T$ such that $tH = 0$. If H is a T-subgroup of G, then T_H is a two-sided ideal, since $0 = t_H(tH) = (t_H t)H$ for every $t_H \in T_H$ and $t \in T$. Clearly $T_{H^\omega} = T_H$.

Let $[S, \varphi]$ be a summation method on G, and let H be a T-subgroup of G. Then $\varphi(S \cap H^\omega) = H_1$ is a T-subgroup of G which contains H. We call this group the *$[S, \varphi]$-extension of H*. It is easy to see that *if H_1 is an $[S, \varphi]$-extension of H, then $T_{H_1} = T_H$*; since $H_1 \supset H$, we obviously have $T_{H_1} \subset T_H$. On the other hand, $T_{H_1} \supset T_{H^\omega} = T_H$. From this, it follows:

THEOREM 5. *If H is a maximal T-subgroup for the annihilator ideal T_H, then H has no proper $[S, \varphi]$-extensions.*

THEOREM 6. *Let H_1 be a denumerable T-subgroup of G, and*

$H_2 > H_1$ a T-subgroup of G of cardinality not greater than $2^{\aleph_0}$ such that $T_{H_1} = T_{H_2}$. Then there is a summation method $[S, \varphi]$ on G so that H_2 is the $[S, \varphi]$-extension of H_1.

Proof. Let $\{h_1, h_2, \cdots\}$ be an enumeration of H_1, and let M be an increasing sequence of integers. Define sequences $s_{M,i} = (g_{n,i})_{n=1}^{\infty}$ by

$$g_{n,i} = \begin{cases} h_i \text{ if } n = 2^{p_i^m},\ m \in M,\ p_i = i\text{th prime} \\ 0 \text{ otherwise .} \end{cases}$$

It is easy to find (see Remark 1) a set $\mathfrak{M}$ of $2^{\aleph_0}$ sequences M such that any relation of the form $\sum_{r,j} t_{rj} s_{M,j}^{(r)} \in \Gamma$ implies $t_{ij} s_{M,j}^{(r)} = 0$ for all r and j, which in turn implies that $t_{rj} \in T_{H_1}$. Now, let $\{h_\alpha^{(2)}\}_{\alpha \in A}$ be a minimal system of generators of H_2, that is $\sum_\alpha t_\alpha h_\alpha^{(2)} = 0$ (finite sum) if and only if $t_\alpha h_\alpha = 0$ for all α. For any choice of the subsystem M_α of $\mathfrak{M}$ the definition $(s_{M,\alpha}) = h_\alpha^{(2)}$ for $\alpha \in A$ yields a summation method on the minimal admissible subgroup S of G^ω containing all the s_{M_α}.

REMARK 4. The restrictions on the cardinalities of H_1 and H_2 can be removed if we allow summation methods using, instead of G^ω, the strong direct sum G^ξ, where ξ is an arbitrary infinite ordinal.

EXAMPLE 1. Let G be a finite abelian group, and T the ring of integers modulo the minimal annihilator N of G. To each subgroup H of G corresponds the ideal generated by its minimal annihilator. Clearly, to every divisor D of N, there corresponds a unique maximal subgroup H_D of G with minimal annihilator D. Each subgroup of G can be $[S, \varphi]$-extended to exactly one H_D.

EXAMPLE 2. If G is the additive group of a ring R considered as the ring of operators T on G, then T-subgroups of G are the left ideals of R. Given now a subset $M \subset R$, it determines a left annihilator ideal T_M of M. Any finitely generated left ideal containing M whose annihilator is T_M can be represented as an $[S, \varphi]$-extension of the left ideal generated by M.

3. Ordered groups. Let G be an abelian group with a partial ordering relation $\geqq$ satisfying: (1) there is a semigroup $H \subset G$ containing the zero element and at least one element $\neq 0$, in which the binary reflexive and transitive relation $\geqq$ is defined; (2) if $h, h_1 \in H$ and $h > 0$, then $h_1 + h > h_1$; (3) the archimedean axiom: if $h_1, h_2 \in H$, $h_1 > 0$ and $h_2 > 0$, then there is a positive integer n such that $nh_1 > h_2$.

DEFINITION 3. Let G be a partially ordered abelian group. $s = (g_1, g_2, \cdots, g_n, \cdots) \in G^\omega$ will be called *positive* if $g_n \in H$ and $g_n \geqq 0$ for

$n = 1, 2, \cdots$, and if $g_{n_0} > 0$ for at least one index n_0. A summation method $[S, \varphi]$ will be called *positive* if $s \in S$ and s positive imply $\varphi(s) > 0$.

The positive elements of G^ω or of S evidently form a semigroup. Furthermore, if s is positive, so is its translate s'.

THEOREM 7. *Let G be a partially ordered abelian group, and $[S, \varphi]$ a positive summation method on G. If $s = (g_1, g_2, \cdots, g_n, \cdots) \in G$ is such that $g_{k_n} \geqq g > 0$ for infinitely many indices k_n, then $s \notin S$.*

Proof. The hypothesis implies that s is a positive element. Assume $s \in S$ and $\varphi(s) = \gamma$, then $0 > \gamma = \varphi(s) = \varphi(g_1, g_2, \cdots, g_{k_n}, 0, 0, \cdots) + \varphi(0, \cdots, 0, g_{k_n}, g_{k_n+1}, \cdots) = \sum_{i=1}^{k_n} g_i + \varphi(0, \cdots, 0, g_{k_n}, g_{k_n+1}, \cdots) > ng$ for each positive integer n. This contradicts the archimedean axiom.

COROLLARY 7.1. *There is no positive nontrivial summation method for the group of integers with their natural ordering.*

COROLLARY 7.2. *Let G be an abelian group with a linear ordering, and $[S, \varphi]$ a positive summation method on G. If $s = (g_1, g_2, \cdots, g_n, \cdots) \in S$ is positive, then* g.l.b $g_n = 0$ *and* $\varphi(s) \geqq$ l.u.b.$_{1 \leqq n < \infty} \sum_{i=1}^{n} g_i$.

From the last part of Corollary 7.2 it follows that if the partial sums of a "series" with positive terms are unbounded, then the "series" does not belong to the domain of any positive summation method.

THEOREM. 8. *Let G be a linearly ordered abelian group. Then there is a nontrivial positive summation method on G if and only if G contains an infinite sequence $g_1, g_2, \cdots$, of positive elements and an element g, such that $g_1 + \cdots + g_n \leqq g$ for all n.*

Proof. The necessity follows immediately from Corollary 7.2. To prove sufficiency, set $s = (s_n)_{n=1}^{\infty}$ and define

$$s_n = \begin{cases} g_k & \text{for } n = 2^k \\ 0 & \text{otherwise} . \end{cases}$$

Then the least admissible S which contains s has elements which can be expressed uniquely in the form

$$t = \gamma + \sum_{i=0}^{n} a_i s^{(i)}$$

where the a_i are integers and $\gamma = (\gamma_1, \gamma_2, \cdots, \gamma_m, 0, 0, \cdots) \in \Gamma$. $t \geqq 0$ implies $a_i \geqq 0$ and $\sum_{j=1}^{m} \gamma_j > -(\sum_{i=0}^{n} a_i)g$. Thus if we define $\varphi(s) = g$ we obtain

$$\varphi(t) = \sum_{j=1}^{m} \gamma_i + \left(\sum_{i=0}^{n} a_i\right)g$$

where $\varphi(t) \geqq 0$ whenever $t \geqq 0$, and φ can be extended in an obvious way to a summation method, and is nontrivial.

THEOREM 9. *Let G be a linearly ordered abelian group and $[S, \varphi]$ a positive summation mothod such that S contains all the positive elements $s = (g_i)_{i=1}^{\infty} \in G$ for which the "partial sums" $\sum_{i=1}^{n} g_i$ are bounded for all n. Then $\varphi(s) = \text{l.u.b.}_{1 \leq n < \infty} \sum_{i=1}^{n} g_i$ for any positive $s \in S$.*

Proof. By Corollary 7.2 we know that $\varphi(s) \geqq \bar{g} = \text{l.u.b.}_n \sum_{i=1}^{n} g_i$. Assume $\varphi(s) > \bar{g}$. Then $\varphi(0, \cdots, 0, g_N, g_{N+1}, \cdots) \geqq \varphi(s) - \bar{g} > 0$ for any N, and $g_N + g_{N+1} + \cdots + g_{N+k} < \bar{g}$ for all k. It follows that there is a greatest positive integer n_1 such that $(2n_1)(g_1 + \cdots + g_k) < \bar{g}$ for all k. Determine n_2 as greatest positive integer such that $(2n_2)(g_2 + \cdots + g_k) < \bar{g} - n_1 g$ for all k, etc. This defines a nondecreasing sequence of positive integers n_j with $n_j \to \infty$. Consider the element $\bar{s} = (n_j g_j)_{j=1}^{\infty} \in G^{\omega}$. It is obviously in S, since the partial sums $\sum_{j=1}^{r} n_j g_j$ are bounded for all r. On the other hand

$$\varphi(\bar{s}) > n_j(\varphi(s) - \bar{g})$$

for all j, which is in contradiction with the archimedean property of the order in G.

4. Limits.

DEFINITION 4. Let $[S, \varphi]$ be a summation method on the abelian group G. The sequence $\{g_1, g_2, \cdots, g_n, \cdots\}$ of elements of G will be called *$[S, \varphi]$-convergent to g* (notation: $g = \lim_{[S\,\varphi]} g_n$, or $g_n \xrightarrow[{[S,\varphi]}]{} g$) if (1) $s = (g_n - g_{n-1})_{n=1}^{\infty} \in S$, and (2) $\varphi(s) = g$. (Here $g_0 = 0$.)

The following properties are immediate:

THEOREM 10. (1) *The sequence $\{g, g, g, \cdots\}$ is $[S, \varphi]$-convergent to g for any $[S, \varphi]$.* (2) *If $g_n \xrightarrow[{[S,\varphi]}]{} g$ and $\bar{g}_n \xrightarrow[{[S,\varphi]}]{} \bar{g}$ then $g_n + \bar{g}_n \xrightarrow[{[S,\varphi]}]{} g + \bar{g}$.* (3) $\lim_{[S\,\varphi]}(-g_n) = -\lim_{[S\,\varphi]} g_n$. (4) *If $g = \lim_{[S,\varphi]} g_n$ and $h_1, h_2, \cdots, h_k$ are arbitrary elements of G, then the sequence $\{h_1, h_2, \cdots, h_k, g_1, g_2, \cdots, g_n, \cdots\}$ is $[S, \varphi]$-convergent to g.*

The last part of Theorem 10 implies that if $\lim_{[S,\varphi]} g_n = g$, then

$\{g_k, g_{k+1}, \cdots\}$ is $[S, \varphi]$-convergent to g, too.

An arbitrary subsequence of an $[S, \varphi]$-convergent sequence will not always be $[S, \varphi]$-convergent to the same limit, even if it is $[S, \varphi]$-convergent.

EXAMPLE 3. Let G be an abelian group with an element g of order > 2. Define S to be the minimal admissible subgroup of G^ω containing the element

$$s = (2g, -2g, 2g, -2g, \cdots) .$$

Since $s' + s = (2g, 0, 0, \cdots)$ we may define $\varphi(s) = g$. Then the sequence $\{2g, 0, 2g, 0, \cdots\}$ is $[S, \varphi]$-convergent to g, but the subsequence $\{2g, 2g, \cdots\}$ is $[S, \varphi]$-convergent to $2g$.

This example shows that it is not always possible to define a topology in G by means of $[S, \varphi]$-convergent sequences.

THEOREM 11. *Let G be an abelian group.* A *non-trivial summation method $[S, \varphi]$ on G, with the property that every subsequence of any $[S, \varphi]$-convergent sequence is $[S, \varphi]$-convergent to the same limit, exists if and only if G is infinite.*

Proof. Let G be finite. If a sequence of elements of G is not eventually constant, then two different elements must occur infinitely often. Hence no summation method $[S, \varphi]$ with the required property is possible.

Assume G infinite, and distinguish among the following cases:

(a) *G contains an element g of infinite order.* Let S be the minimal admissible subgroup of G^ω containing all the sequences $(n_i g)_{i=1}^\infty$ such that $\sum n_i$ converges p-adically to a rational integer n. Define then

$$\varphi((n_i g)_{i=1}^\infty) = ng .$$

(b) *There exists an element $g \neq 0$ of G of finite order divisible by arbitrarily high powers of some prime p.* Let M be the subgroup of the additive group of rationals, containing all the sequences $(p^{-k/n}a_n)_{n=1}^\infty$ where a_n and k_n are integers, such that $\sum_{n=1}^\infty p^{-k/n}a_n$ converges to a number of the form $p^{-k}a$, a and k integers. Let S be the minimal admissible subgroup of G^ω that contains the sequence $(p^{-k_n}a_n g)_{n=1}^\infty$, and define $\varphi((p^{-k_n}a_n g)_{n=1}^\infty) = p^{-k}ag$.

(c) *All elements of G are finite but not of bounded order, and no element of G is infinitely divisible* (by powers of some prime).

Define $G_n = n!G$; let S be the minimal admissible subgroup of G^ω consisting of the sequences $(g_n)_{n=1}^\infty$ so that there exists a g in G with $g - g_1 - g_2 - \cdots - g_n \in G_n$ for $n = 1, 2, \cdots$. Define $\varphi((g_n)_{n=1}^\infty) = g$.

(d) *All elements of G have bounded order* $\leqq m$. Then G must contain an infinite subgroup, all of whose elements have order p for some fixed prime p. Otherwise there would be a least divisor d of m for which there is an infinite subgroup G_1 of G such that $dG_1 = 0$. If d is composite, then for every prime divisor q of d the group qG_1 is finite, and hence the kernel of the homomorphism $G_1 \to qG_1$ is an infinite group G_2 with $qG_2 = 0$, contrary to the hypothesis.

Now, an infinite abelian group all of whose elements are of order p is the direct sum of infinitely many cyclic groups of order p, say $Z_1^{(p)} \oplus Z_2^{(p)} \oplus \cdots$. Let S be the minimal admissible subgroup of G^ω containing the sequences $(g_n)_{n=1}^\infty$ for which there exists a $g \in G$ such that $g - g_1 - \cdots - g_n \in Z_{n+1}^{(p)} \oplus Z_{n+2}^{(p)} \oplus \cdots$, and define $\varphi((g_n)_{n=1}^\infty) = g$.

UNIVERSITY OF WASHINGTON
UNIVERSITY OF CALIFORNIA LOS ANGELES

PACIFIC JOURNAL OF MATHEMATICS
Vol. 15, No. 1, 1965

ON AN EXTENSION OF THE PICARD-VESSIOT THEORY

H. F. KREIMER

In previous papers, the author has extended the Galois correspondences between differential Picard-Vessiot extensions and algebraic matrix groups to Picard-Vessiot extensions of a wider class of fields with operators, the so-called M-fields. In this paper, M-field extensions which generalize extensions by integrals and by exponentials of integrals are studied.

These fields are found to be simple field extensions and their structure in the case that the extension is algebraic is investigated. Under suitable restrictions on the fields of constants, the M-Galois groups of these fields are shown to be commutative. Criteria are established for such solution fields to be P–V extensions of M-fields of difference and differential type. An extension obtained by a finite sequence of algebraic extensions, extensions by integrals, and extensions by exponentials of integrals, is called a generalized Liouville extension. It is demonstrated that if the connected component of the identity element in the M-Galois group of a regular P–V extension is a solvable group, then the P–V extension is a generalized Liouville extension, and if a P–V extension is contained in a generalized Liouville extension then the connected component of the identity element in the M-Galois group of the P–V extension is solvable.

1. Terminology and notation are briefly considered in § 2, and a preliminary result on the constants of an algebraic M-extension of an M-field is obtained. The structure of solution fields analogous to extensions by integrals and criteria for the existence of P–V extensions of this type are determined in § 3, and a similar study of solution fields analogous to extensions by exponentials of integrals is made in § 4. In § 5, generalized Liouville extensions are defined, and solvability of the Galois group of a P–V extension is interpreted in terms of imbedding the extension in a generalized Liouville extension.

2. ***M*-rings.** The terminology and notaion of this paper are the same as in [6] and [7]. Let C be an associative, commutative coalgebra with identity over a ring W, which is freely generated as a W-module by a set M. If $w \to \bar{w}$ is a homomorphism of W into a ring S, let C^s be the S-module obtained from the W-module C by

Received February 21, 1963, and in revised form March 17, 1964. Part of a dissertation presented for the degree of Doctor of Philosophy in Yale University.

inverse transfer of the basic ring to S. If ρ is a homomorphism of a ring R into the algebra $(C^S)^* = \mathrm{Hom}_S(C^S, S)$; then for every $m \in M$ there is a mapping $a \longrightarrow a^\rho(m)$ of R into S, which will also be denoted by m, and the set of these mappings will be called an M-system of mappings of R into S. Let $m \longrightarrow \sum_{n\ p \in M} z_{mnp} n \otimes p$, where $m \in M$, $z_{mnp} \in W$, and $z_{mnp} = 0$ except for a finite number of elements n and p in M, be the coproduct mapping of C into $C \otimes_W C$; if $a, b \in R$ and $m \in M$, $(a+b)m = am + bm$ and $(ab)m = \sum_{n,p \in M} \overline{z_{mnp}}(an)(bp)$. An M-ring is a ring together with an M-system of mappings of the ring into itself. An M-ring of difference type is an M-ring in which the M-system of mappings consists of homomorphisms, and an M-ring of differential type is an M-ring in which the M-system of mappings consists of the identity automorphism and higher derivations of rank one or greater.

An element c of an M-ring R is a constant if $(ca)^\rho = c \cdot a^\rho$ for every $a \in R$. The following are equivalent:

(1) c is a constant of R,

(2) $c^\rho = c \cdot 1^\rho$,

(3) $(ca)m = c(am)$ for every $a \in R$ and $m \in M$,

(4) $cm = c(1m)$ for every $m \in M$.

The constants of R form a subring of R which contains the identity element of R and this subring will be denoted by R_c. Suppose $b, d \in R$ and d is a unit in R, then $bd^{-1} \in R_c$ if, and only if, $d(bm) = b(dm)$ for every $m \in M$. Consequently, if R is a field, so is R_c.

(2.1) LEMMA. *Let K be an M-field which is an M-extension of an M-field L. If K is an algebraic extension of L, then K_c is an algebraic extension of L_c.*

Proof. Suppose $d \in K_c$ and $f(x) = x^h + a_{h-1}x^{h-1} + \cdots + a_1x + a_0$ is the irreducible, monic polonomial over L for which d is a root. If $m \in M$, $0 = (f(d))m = (fm)(d)$, where $(fm)(x) = (1m)x^h + (a_{h-1}m)x^{h-1} + \cdots + (a_1m)x + a_0m$. But then $(fm)(x)$ must be a multiple of $f(x)$, thus $(fm)(x) = (1m)f(x)$ and $a_\alpha m = (1m)a_\alpha$ for $0 \leqq \alpha \leqq h-1$. Therefore, $a_\alpha \in L_c$ for $0 \leqq \alpha \leqq h-1$ and d is algebraic over L_c.

Let $S'(M)$ be the free semi-group with identity generated by the set M. Operations by elements of $S'(M)$ on an M-ring R are defined as follows: the identity element of $S'(M)$ operates on R as the identity automorphism of R, and any other element of $S'(M)$ operates on R as the resultant of the operations on R by its factors. If h is a positive integer, $r_1, r_2, \cdots, r_h$ are h elements of R, and $s_1, s_2, \cdots, s_h$ are h elements of $S'(M)$; denote by $W(r_1, r_2, \cdots, r_h; s_1, s_2, \cdots, s_h)$ the determinant:

$$\begin{vmatrix} r_1s_1 & r_1s_2 & \cdots & r_1s_h \\ r_2s_1 & r_2s_2 & \cdots & r_2s_h \\ \vdots & \vdots & & \vdots \\ r_hs_1 & r_hs_2 & \cdots & r_hs_h \end{vmatrix}$$

An M-field K which is an M-extension of an M-field L is a solution field over L if there exists a positive integer h and h elements $k_1, k_2, \cdots, k_h$ of K, such that $K = L\langle k_1, k_2, \cdots, k_h\rangle$ and, for some choice of h elements $t_1, t_2, \cdots, t_h$ in $S'(M)$, $W(k_1, k_2, \cdots, k_h; t_1, t_2, \cdots, t_h) = W_0 \neq 0$ while $W_0^{-1}W(k_1, k_2, \cdots, k_h; t_1, \cdots, t_{\alpha-1}, t_{\alpha+1}, \cdots, t_h, t) \in L$ for $1 \leqq \alpha \leqq h$ and $t = 1$ or $t = t_\beta m, m \in M$ and $1 \leqq \beta \leqq h$. The set of elements $k_1, k_2, \cdots, k_h$ is a fundamental set for K over L. K is a Picard-Vessiot extension of the M-field L if K is a solution field over L and, additionally, $K_c = L_c$ and L_c is an algebraically closed field.

3. Extensions by integrals.

(3.1) THEOREM. *Let K, L and L_0 be M-fields such that K is an M-extension of L and L is an M-extension of L_0, and assume there exists $k \in K$ such that $km - (1m)k = a_m \in L_0$ for every $m \in M$.*

(i) *$L\langle k\rangle$ is a solution field over L.*

(ii) *If $K_c = L\langle k\rangle_c$, then $L\langle k\rangle$ is invariant under M-automorphisms of K over L; and, if $L\langle k\rangle_c = L_c$, then the M-Galois group of $L\langle k\rangle$ over L is commutative.*

(iii) *As abstract fields, $L\langle k\rangle$ is a simple extension of L by adjunction of the element k.*

(iv) *$L\langle k\rangle_c = L_c$ if, and only if, $L\{k\}_c = L_c$.*

(v) *If k is algebraic over L but $k \notin L$ and $L\langle k\rangle_c = (L_0)_c$, L is a field of characteristic $p \neq 0$ and k is a root of an irreducible polynomial over L of the form $x^{p^h} + c_{h-1}x^{p^{h-1}} + \cdots + c_1x^p + c_0x + b$, where h is a positive integer, $c_\alpha \in (L_0)_c$ for $0 \leqq \alpha \leqq h-1$, and $b_m - (1m)b \in L_0$ for every $m \in M$.*

(vi) *If L is a field of characteristic zero and k is transcendental over L then $L\langle k\rangle_c = L_c$ if, and only if, there does not exist $b \in L$ such that $bm - (1m)b = a_m$ for every $m \in M$.*

(vii) *If $L\langle k\rangle$ is a P–V extension, such an extension is unique.*

Proof. (i) If $L\langle k\rangle = L$, then $L\langle k\rangle$ is trivially a solution field over L with fundamental set consisting of 1. Therefore, assume $k \notin L$. If $a_m = 0$ for every $m \in M$, then $k \in K_c$ and $L\langle k\rangle$ is a solution field over L with fundamental set consisting of k. If there exists $n \in M$ such that $a_n \neq 0$, then the determinant $\begin{vmatrix} 1 & k \\ 1n & kn \end{vmatrix} = a_n \neq 0$ while 1 and k are solutions of the equations $xm = a_m a_n^{-1}(xn) + ((1m) - (1n)a_m a_n^{-1})x$

and $(xn)m = (a_n m + \sum_m)a_n^{-1}(xn) + ((1n)m - (1n)(a_n m)a_n^{-1} - (1n)a_n^{-1}\sum_m)x$ where $\sum_m = \sum_{q,r\in M} z_{mqr}((1n)q)a_r$, for every $m \in M$. It is then readily established that $L\langle k\rangle$ is a solution field over L with fundamental set consisting of 1 and k.

(ii) An M-isomorphism φ of $L\langle k\rangle$ over L into K is completely determined by its action on k, and $(k\varphi - k)m = (km)\varphi - km = (a_m + (1m)k)\varphi - a_m - (1m)k = (1m)(k\varphi - k)$ for every $m \in M$. Therefore $k\varphi - k \in K_c$ or $k\varphi = k + c$ for some constant c. If $K_c = L\langle k\rangle_c$, then $L\langle k\rangle$ is invariant under M-automorphisms of K over L; and, if $L\langle k\rangle_c = L_c$, then the M-Galois group of $L\langle k\rangle$ over L is isomorphic to a subgroup of the additive group of constants of L.

(iii) The subring $L[k] \subseteq K$ of polynomials over L in k is an M-subring of K, and $L\langle k\rangle$ is simply the field of fractions of $L[k]$ in K. (See Corollary (4.2) of [6]).

(iv) If $L\langle k\rangle_c = L_c$ then certainly $L\{k\}_c = L_c$. If k is algebraic over L, then $L\langle k\rangle = L[k] = L\{k\}$ and the converse is true. Let k be transcendental over L. An element of $L\langle k\rangle$ may be represented as the ratio of a polynomial $f(k) \in L[k]$ and a monic polynomial $g(k) \in L[k]$. Suppose $f(k)\cdot(g(k))^{-1} \in K_c$ and is expressed in lowest terms, i.e., $f(k)$ and $g(k)$ are relatively prime. Then $g(k)\cdot((f(k))m) = f(k)\cdot((g(k))m)$ for every $m \in M$; and, were $(g(k))m \neq (1m)\cdot g(k)$ for some $m \in M$, then $f(k)\cdot(g(k))^{-1} = ((f(k))m - (1m)f(k))\cdot((g(k))m - (1m)g(k))^{-1}$. This last is impossible since the degree of $(g(k))m - (1m)g(k)$ is less than the degree of $g(k)$. Thus $(g(k))m = (1m)g(k)$ and $(f(k))m = (1m)f(k)$ for every $m \in M$, consequently $f(k), g(k) \in L\{k\}_c$. Therefore, if $L\{k\}_c = L_c$, then $f(k)\cdot(g(k))^{-1} \in L_c$.

(v) Suppose k is algebraic over L. If $L[y]$ is the ring of polynomials over L in an indeterminate y, determined as an M-extension of L by setting $ym = a_m + (1m)y$ for every $m \in M$; there is a canonical M-homomorphism η of $L[y]$ over L into K such that $y^\eta = k$. Let I be the kernel of η, and let $f(y)$ be the monic polynomial which generates I, i.e., the minimal polynomial for k over L. Because I is an M-ideal, $(f(y))m$ must be a multiple of $f(y)$ and computation shows that $(f(y))m = (1m)f(y)$, for every $m \in M$. Therefore $f(y) \in L[y]_c$. Suppose $L\langle k\rangle_c = L_c$ and $g(y) \in L[y]_c$. Then $g(k) \in L\langle k\rangle_c = L_c$, say $g(k) = c$, and k is a root of $g(y) - c$. Therefore $g(y) - c$ is a multiple of $f(y)$ and, if $g(y)$ has positive degree, it is not less than the degree of $f(y)$. Subsequently assume only that $L[y]_c$ contains polynomials of positive degree, and $f(y)$ is such a polynomial of least positive degree. If $d \in L_c$, $(y + d)m = (1m)(y + d) + a_m$ and $(f(y + d))m = (1m)f(y + d)$ for every $m \in M$. Therefore $f(y + d)$ and $f(y + d) - f(y)$ are elements of $L[y]_c$. The degree of $f(y+d)-f(y)$ is less than the degree of $f(y)$ and, therefore, cannot be positive. Thus $f(y+d)-f(y)=f(d)-f(0)$; but this identity can be valid only if $f(y)$ is a polynomial of degree not greater than one, or L is a field of characteristic $p \neq 0$ and $f(y)=b+\sum_{\alpha=0}^{h} c_\alpha y^{p^\alpha}$, where h is a nonnegative integer and b and c_α, $0 \leqq \alpha \leqq h$, are elements of L. If ρ is the representation of $L[y]$ in $(C^{L[y]})^*$ associated with the M-system of mappings on $L[y]$, then $y^\rho = a + y \cdot 1^\rho$ where a is that element of $(C^{L[y]})^*$ such that $a(m) = a_m$ for every $m \in M$. If L is a field of characteristic $p \neq 0$ and $f(y) = b+\sum_{\alpha=0}^{h} c_\alpha y^{p^\alpha}$, then $f(y)\cdot 1^\rho = (f(y))^\rho = b^\rho + \sum_{\alpha=0}^{h} c_\alpha^\rho a^{p^\alpha} + \sum_{\alpha=0}^{h} y^{p^\alpha}\cdot c_\alpha^\rho$. Therefore $c_\alpha \cdot 1^\rho = c_\alpha^\rho$ and $c_\alpha \in L_c$ for $0 \leqq \alpha \leqq h$ and, if $c_\alpha \in (L_0)_c$ for $0 \leqq \alpha \leqq h$, then $bm-(1m)b = -\sum_{\alpha=0}^{h}(c_\alpha^\rho a^{p^\alpha})(m) \in L_0$ for every $m \in M$. The assertion in (v) is now immediate.

(vi) Suppose L is a field of characteristic zero and k is transcendental over L. If $L\langle k\rangle_c \neq L_c$, then $L\{k\}_c \neq L_c$ and there is a polynomial over L in k of positive degree which belongs to $L\{k\}_c$. Let $f(k)$ be such a polynomial of least degree. By the argument in part (v), the degree of $f(k)$ is one. Then $f(k)$ generates a prime M-ideal I in $L\{k\}$ and $L\{k\}/I$ is M-isomorphic to L. If b is the image of $k+I$ under such an M-isomorphism, then $bm-(1m)b=a_m$ for every $m \in M$. Conversely, if there exists $b \in L$ such that $bm-(1m)b=a_m$ for every $m \in M$, then $k-b \in L\{k\}_c$ and $L\langle k\rangle_c \neq L_c$.

(vii) Let $L\langle k\rangle$ be a P–V extension of L and let $L\langle k'\rangle$ be a second P–V extension of L such that $k'm-(1m)k'=a_m$ for every $m \in M$. If k and k' are transcendental over L, there is an isomorphism φ of $L\langle k\rangle$ over L onto $L\langle k'\rangle$ such that $k^\varphi = k'$ and φ is an M-isomorphism. Suppose k is algebraic over L and either k' is transcendental over L or algebraic over L but of degree over L not less than the algebraic degree of k over L. If $f(x)$ is the monic minimal polynomial for k over L, then $f(k') \in L\langle k'\rangle_c = L_c$ by the argument in part (v); say $f(k)=d$. Then k' is a root of $f(x)-d$ and k' is algebraic over L with the same degree over L as k. If the degree of k over L is one, then $L\langle k\rangle = L = L\langle k'\rangle$. If the degree of k over L is greater than one, then L is a field of characteristic $p \neq 0$ and $f(x)=x^{p^h}+c_{h-1}x^{p^{h-1}}+\cdots+c_1x^p+c^0x+b$ where h is a positive integer and $c_\alpha \in L_c$ for $0 \leqq \alpha \leqq h-1$. Let c be a root in the algebraically closed field L_c of $x^{p^h}+c_{h-1}x^{p^{h-1}}+\cdots+c_1x^p+c_0x+d$. Then $f(k'+c)=f(k')-d=0$ and there is an isomorphism φ of $L\langle k\rangle$ over L onto $L\langle k'\rangle$ such that $k^\varphi = k'+c$. φ is an M-isomorphism.

(3.2) Corollary. *Let L be an M-field of characteristic zero such that L_c is algebraically closed, and let a_m, $m \in M$, be elements* *of L. There exists a P–V extension $L\langle k\rangle$ of L such that $km - (1m)k = a_m$ for every $m \in M$ if, and only if* (1) *there exists an element $b \in L$ such that $bm - (1m)b = a_m$ for every $m \in M$, in which case $L\langle k\rangle = L$, or* (2) *if $L[y]$ is the ring of polynomials over L in an indeterminate y, determined as an M-extension of L by setting $ym = a_m + (1m)y$ for $m \in M$, and $L(y)$ is the field of fractions of $L[y]$, then there is a structure of an M-field on $L(y)$ such that $L(y)$ is an M-extension of $L[y]$, in which case $L\langle k\rangle$ and $L(y)$ are M-isomorphic.*

Proof. If there exists $b \in L$ such that $bm - (1m)b = a_m$ for every $m \in M$, set $k = b$ to obtain a trivial P–V extension of L. If there does not exist $b \in L$ such that $bm - (1m)b = a_m$ for every $m \in M$, but there is a structure of an M-field on $L(y)$ such that $L(y)$ is an M-extension of $L[y]$; then $L(y)_c = L_c$ by part (vi) of Theorem (3.1) and, setting $k = y$, $L\langle k\rangle = L(y)$ is a P–V extension of L. The converse is immediate from parts (iii) and (v) of Theorem (3.1).

If L is an M-field of differential type and of characteristic zero such that L_c is algebraically closed, Corollary (3.2) may be applied to establish the existence of P–V extensions by adjunction of integrals.

(3.3) COROLLARY. *Let L be an M-field of difference type such that L_c is algebraically closed, and let a_m, $m \in M$, be elements of L. There exists a P–V extension $L\langle k\rangle$ of L such that $km - k = a_m$ for every $m \in M$ if, and only if, the characteristic is* 0 *or the following condition is fulfilled when the characteristic is $p \neq 0$: that there do not exist a nonnegative integer h, $c_\alpha \in L_c$ for $0 \leqq \alpha \leqq h$, and $b \in L$, such that $bm - b + \sum_{\alpha=0}^{h} c_\alpha (a_m)^{p^\alpha} = d_m \in L_c$ for every $m \in M$, where $\{d_m \mid m \in M\}$ is a finite set not equal to $\{0\}$.*

Proof. Let $L[y]$ and $L(y)$ be as in Corollary (3.2). The M-system of mappings on $L[y]$ consists of isomorphisms and these can be extended to $L(y)$, so that $L(y)$ is an M-field which is an M-extension of $L[y]$. Because of Corollary (3.2), only the case when L is a field of characteristic $p \neq 0$ need be considered. If $L[y]_c = L_c$, then $L(y)_c = L_c$ by part (iv) of Theorem (3.1) and, setting $y = k$, $L\langle k\rangle = L(y)$ is the desired P–V extension of L. If there exists an irreducible polynomial in $L[y]_c$ of positive degree, this polynomial generates a proper prime M-ideal I in $L[y]$. The M-field $L[y]/I$ is an algebraic extension of L and $(L[y]/I)_c = L_c$ by Lemma (1.1), since L_c is algebraically closed. Setting $k = y + I$, $L\langle k\rangle = L[y]/I$ is the desired P–V extension of L. Therefore, assume that there exist polynomials of positive degree in $L[y]_c$, let $f(y)$ be such a polynomial of least positive degree, but assume $f(y)$ is reducible. Analyzing $f(y)$ as in the proof of part (v) of Theorem (3.1), $f(y)$ must have the form $f(y) = b' + \sum_{\alpha=0}^{i} c'_\alpha y^{p^\alpha}$ where

i is a nonnegative integer and $c'_\alpha \in L_c$ for $0 \leqq \alpha \leqq i$. Let $g(y)$ be an irreducible monic polynomial which divides $f(y)$, and let ζ be a root of $g(y)$ in a splitting field for $f(y)$ over L. The roots of $f(y)$ are the elements $\zeta + e$ where e is a root of $f(y) - b'$ and lies in the algebraically closed field L_c. The roots of $g(y)$ are those elements $\zeta + e$ where e is a root of $g(\zeta + y)$. Let e and e' be roots of $g(\zeta + y)$; there is an automorphism of the splitting field of $f(y)$ over L which maps ζ to $\zeta + e'$ and its inverse maps $\zeta + e$ to $\zeta + e - e'$. Then $g(\zeta + e - e') = 0$, $e - e'$ is again a root of $g(\zeta + y)$, and the roots of $g(\zeta + y)$ form an additive subgroup of L_c. Therefore $g(\zeta + y)$ must be a p-polynomial over L_c, say $g(\zeta + y) = \sum_{\alpha=0}^{h} c_\alpha y^{p^\alpha}$ where h is a nonnegative integer and $c_\alpha \in L_c$ for $0 \leqq \alpha \leqq h$; and $g(y) = g(\zeta + (y - \zeta))$ must have the form $g(y) = b + \sum_{\alpha=0}^{h} c_\alpha y^{p^\alpha}$. Any irreducible monic polynomial which divides $f(y)$ will have the form $g(y + e)$ where e is a root in L_c of $f(y) - b'$. If $m \in M$; $(f(y))m = f(y)$, $(g(y))m = g(y + e_m) = g(y) + d_m$, and $bm + \sum_{\alpha=0}^{h} c_\alpha (a_m)^{p^\alpha} = b + d_m$, where $d_m = g(e_m) - b \in L_c$ and e_m is a root of $f(y) - b'$. Since $g(y)$ is a proper factor of $f(y)$, $g(y) \notin L[y]_c$ and $d_m \neq 0$ for some $m \in M$.

Conversely, assume there exist a nonnegative integer h, $c_\alpha \in L_c$ for $0 \leqq \alpha \leqq h$, and $b \in L$, such that $bm - b + \sum_{\alpha=0}^{h} c_\alpha (a_m)^{p^\alpha} = d_m \in L_c$ for every $m \in M$, where $\{d_m \mid m \in M\}$ is a finite set not equal to $\{0\}$. Let E be the additive subgroup of L_c generated by $\{d_m \mid m \in M\}$, let $g(y) = b + \sum_{\alpha=0}^{h} c_\alpha y^{p^\alpha}$, let $\bar{f}(y) = \prod_{e \in E} (y + e)$, and let $f(y) = \bar{f}(g(y))$. $\bar{f}(y)$ will be a p-polynomial over L_c, i.e. $\bar{f}(y)$ will be a finite linear combination over L_c of monomials y^{p^β}, β a nonnegative integer; $f(y)$ will have the form $f(y) = b' + \sum_{\alpha=0}^{i} c'_\alpha y^{p^\alpha}$ where i is a nonnegative integer and $c'_\alpha \in L_c$ for $0 \leqq \alpha \leqq i$; and $f(y) \in L[y]_c$. If the desired P–V extension $L\langle k\rangle$ existed, $f(k)$ would be an element of $L\langle k\rangle_c = L_c$. If c is a root in L_c of $f(y) - b' + f(k)$, then $f(k + c) = 0$ and some factor $g(k + c) + e = 0$. But then $0 = (g(k + c) + e)m = g(k + c) + e + d_m = d_m$ for every $m \in M$, contrary to the assumption that $\{d_m \mid m \in M\} \neq \{0\}$.

(3.4) COROLLARY. *Let L be an M-field such that the M-system of mappings on L consists of the identity automorphism m_0 and infinite higher derivations and L_c is algebraically closed. If a_m, $m \in M$ and $m \neq m_0$, are elements of L, there exists a P–V extension of differential type $L\langle k\rangle$ of L such that $km = a_m$ for every $m \in M$, $m \neq m_0$.*

Proof. Let $a_{m_0} = 0$, and let $L[y]$ and $L(y)$ be as in Corollary (3.2). The M-system of mappings on $L[y]$ consists of the identity automorphism m_0 and infinite higher derivations, and these can be extended to $L(y)$ so that $L(y)$ is an M-field of differential type which is an M-extension of $L[y]$. By repetition of the argument in the beginning of the proof of Corollary (3.3), only the case when L is a field of characteristic $p \neq 0$ and $L[y]_c$ contains polynomials of positive degree need be considered. Let $f(y) \in L[y]_c$ be a polynomial of positive degree, and let $g(y)$ be an irreducible factor of $f(y)$, say $f(y) = q(y)\cdot(g(y))^{h\cdot p^i}$ where h is a positive integer not divisible by p, i is a nonnegative integer, and $q(y)$ is not divisible by $g(y)$. Let $\{D_1, D_2, D_3, \cdots\}$ be an infinite higher derivation on $L[y]$ contained in the M-system of mappings on $L[y]$. If $D_0 = m_0$, $(g(y))D_0 = g(y)$. Let j be a positive integer and assume that $(g(y))D_\alpha$ is a multiple of $g(y)$ for $0 \leqq \alpha < j$. Observe that $(g(y))^{p^i}D_\alpha = 0$ for every positive integer α which is not divisible by p^i and $(g(y))^{p^i}D_{\alpha\cdot p^i} = ((g(y))D_\alpha)^{p^i}$ for every nonnegative integer α. Then $0 = (f(y))D_{j\cdot p^i}$ which is equal to a sum of terms divisible by $(g(y))^{h\cdot p^i}$ plus the term $hq(y)\cdot(g(y))^{(h-1)p^i}\cdot((g(y))D_j)^{p^i}$, and $(g(y))D_j$ must be divisible by $g(y)$. Consequently $g(y)$ generates a proper prime M-ideal I in $L[y]$, $L[y]/I$ is an algebraic extension of L and, setting $k = y + I$, $L\langle k\rangle = L[y]/I$ is the desired P–V extension of L.

4. Extensions by exponentials of integrals.

(4.1) THEOREM. *Let K, L, and L_0 be M-fields such that K is an M-extension of L and L is an M-extension of L_0, and assume there exists a nonzero $k \in K$ such that $km = a_m k$, where $a_m \in L_0$, for every $m \in M$.*

(i) *$L\langle k\rangle$ is a solution field over L.*

(ii) *If $K_c = L\langle k\rangle_c$, then $L\langle k\rangle$ is invariant under M-automorphisms of K over L; and, if $L\langle k\rangle_c = L_c$, then the M-Galois group of $L\langle k\rangle$ over L is commutative.*

(iii) *As abstract fields, $L\langle k\rangle$ is a simple extension of L by adjunction of the element k.*

(iv) *$L\langle k\rangle_c = L_c$ if, and only if, $L\{k\}_c = L_c$.*

(v) *If k is algebraic over L and $L\langle k\rangle_c = L_c$, then k is a root of an irreducible polynomial over L of the form $x^h + b$, where h is a positive integer, $b \neq 0$ and $(bm)b^{-1} \in L_0$ for every $m \in M$.*

(vi) *If $L\langle k\rangle$ is a P–V extension, such an extension is unique.*

Proof. (i) It is easily verified that $L\langle k\rangle$ is a solution field over L with fundamental set consisting of k.

(ii) An M-isomorphism φ of $L\langle k\rangle$ into K is completely determined by its action on k, and $k((k\varphi)m) = k((km)\varphi) = k((a_m k)\varphi) = (a_m k)\cdot(k\varphi) = (k\varphi)\cdot(km)$ for every $m \in M$. Therefore $(k\varphi)k^{-1} \in K_c$ or $k\varphi = ck$ for some nonzero constant c. If $K_c = L\langle k\rangle_c$, then $L\langle k\rangle$ is invariant under M-automorphisms of K over L; and, if $L\langle k\rangle_c = L_c$, then the M-Galois group of $L\langle k\rangle$ over L is isomorphic to a subgroup of the multiplicative group of nonzero constants of L.

(iii) The argument is the same as in part (iii) of Theorem (3.1).

(iv) If $L\langle k\rangle_c = L_c$ then certainly $L\{k\}_c = L_c$. If k is algebraic over L, then $L\langle k\rangle = L[k] = L\{k\}$ and the converse is true. Let k be transcendental over L. An element of $L\langle k\rangle$ may be represented as the ratio of a polynomial $f(k) \in L[k]$ and a nonzero polynomial $g(k) \in L[k]$ with either $f(0) = 1$ or $g(0) = 1$. Suppose $f(k)\cdot(g(k))^{-1} \in K_c$ and is expressed in lowest terms. Then $g(k)\cdot((f(k))m) = f(k)\cdot((g(k))m)$ for every $m \in M$; and, were $(g(k))m \neq (1m)g(k)$ for some $m \in M$, then $f(k)\cdot(g(k))^{-1} = ((f(k))m - (1m)f(k))\cdot((g(k))m - (1m)g(k))^{-1}$. This last is impossible, since it follows from the equations $f(0)\cdot((g(0))m - (1m)g(0)) = g(0)\cdot((f(0))m - (lm)f(0)) = 0$ that $(f(k))m - (1m)f(k)$ and $(g(k))m - (1m)g(k)$ are both divisible by k. Thus $(g(k))m = (1m)g(k)$ and $(f(k))m = (1m)f(k)$ for every $m \in M$, consequently $f(k), g(k) \in L\{k\}_c$. Therefore, if $L\{k\}_c = L_c$ then $f(k)\cdot(g(k))^{-1} \in L_c$.

(v) Suppose k is algebraic over L. If $L[y]$ is the ring of polynomials over L in an indeterminate y, determined as an M-extension of L by setting $ym = a_m y$ for every $m \in M$; there is a canonical M-homomorphism η of $L[y]$ over L into K such that $y^\eta = k$. Let I be the kernel of η. Since $k \neq 0$, $y \notin I$. Let $f(y)$ be a polynomial such that $f(y)$ generates I and $f(0) = 1$. Because I is an M-ideal, $(f(y))m$ must be a multiple of $f(y)$ and computation shows that $(f(y))m = (1m)f(y)$, for every $m \in M$. Therefore $f(y) \in L[y]_c$. Suppose $L\langle k\rangle_c = L_c$ and $g(y) \in L[y]_c$. Then $g(k) \in L\langle k\rangle_c = L_c$, say $g(k) = c$, and k is a root of $g(y) - c$. Therefore $g(y) - c$ is a multiple of $f(y)$ and if $g(y)$ has positive degree, it is not less than the degree of $f(y)$. Subsequently assume only that $L[y]_c$ contains polynomials of positive degree, and $f(y)$ is such a polynomial of least positive degree. If $b^{-1}y^h$ is the highest term of $f(y)$ and $m \in M$, then the identity $(f(y))m = (1m)f(y)$ implies $(b^{-1}y^h)m = (1m)b^{-1}y^h$. Therefore $b^{-1}y^h$ and $f(y) - b^{-1}y^h$ are elements of $L[y]_c$. The degree of $f(y) - b^{-1}y^h$ is less than the degree of $f(y)$ and, therefore, cannot be positive. Thus $f(y) - b^{-1}y^h = f(0) = c \in L_c$ or $f(y) = b^{-1}y^h + c$. Since $b^{-1}y^h \in L[y]_c$, $b(y^h m) = y^h(bm)$ or $(bm)b^{-1} = (y^h m)y^{-h} \in L_0$ for every $m \in M$. The assertion in (v) is now immediate.

(vi) Let $L\langle k\rangle$ be a P–V extension of L and let $L\langle k'\rangle$ be a second P–V extension of L such that $k' \neq 0$ and $k'm = a_m k'$ for every $m \in M$. If k and k' are transcendental over L, there is an isomorphism φ of $L\langle k\rangle$ over L onto $L\langle k'\rangle$ such that $k^\varphi = k'$ and φ is an M-isomorphism. Suppose k is algebraic over L and either k' is transcendental over L or algebraic over L but of degree over L not less than the algebraic degree of k over L. If $x^h + b$ is the minimal polynomial for k over L, then $b^{-1}(k')^h + 1 \in L\langle k'\rangle_c = L_c$ by the argument in part (v); say

$b^{-1}(k')^h + 1 = d$. Then k' is a root of $x^h + b(1-d)$ and $d \neq 1$. Let c be a root in the algebraically closed field L_c of $x^h - (1-d)^{-1}$. Then $(ck')^h + b = 0$ and there is an isomorphism φ of $L\langle k\rangle$ over L onto $L\langle k'\rangle$ such that $k^\varphi = ck'$. φ is an M-isomorphism.

(4.2) COROLLARY. *Let L be an M-field of difference type such that L_c is algebraically closed, and let a_m, $m \in M$, be elements of L. There exists a P–V extension $L\langle k\rangle$ of L such that $k \neq 0$ and $km = a_m k$ for every $m \in M$, if and only if, $a_m \neq 0$ for every $m \in M$ and there do not exist positive integers h and i and a nonzero $b \in L$, such that $bm = c_m(a_m)^h b$ for every $m \in M$, where c_m is an ith root of unity and some $c_m \neq 1$.*

Proof. If the desired P–V extension $L\langle k\rangle$ exists and $m \in M$, m is an isomorphism on $L\langle k\rangle$. Since $k \neq 0$, $km = a_m k \neq 0$ and $a_m \neq 0$. Therefore assume $a_m \neq 0$ for every $m \in M$. Let $L[y]$ be the ring of polynomials over L in an indeterminate y, determined as an M-extension of L by setting $ym = a_m y$ for every $m \in M$, and let $L(y)$ be the field of fractions of $L[y]$. The M-system of mappings on $L[y]$ consists of isomorphisms and these can be extended to $L(y)$, so that $L(y)$ is an M-field which is an M-extension of $L[y]$. If $L[y]_c = L_c$, then $L(y)_c = L_c$ by part (iv) of Theorem (4.1) and, setting $k = y$, $L\langle k\rangle = L(y)$ is the desired P–V extension of L. Suppose $f(y) \neq y$ is an irreducible polynomial in $L[y]_c$ of positive degree. $f(y)$ generates a proper prime M-ideal I in $L[y]$, $L[y]/I$ is an algebraic extension of L, $y \notin I$ and, setting $k = y + I$, $L\langle k\rangle = L[y]/I$ is the desired P–V extension of L. Consequently, assume that there exist polynomials of positive degree in $L[y]_c$, let $f(y)$ be such a polynomial of least positive degree, $f(y)$ may be chosen so that $f(0) \neq 0$, but assume $f(y)$ is reducible. Analyzing $f(y)$ as in the proof of part (v) of Theorem (4.1), $f(y)$ must have the form $(b')^{-1}y^i + c'$ where i is a positive integer. If $g(y)$ is an irreducible factor of $f(y)$ such that $g(0) = 1$, then $g(y)$ has the form $g(y) = b^{-1}y^h + 1$ where h is a positive integer, and all other such factors of $f(y)$ have the form $g(dy)$ where d is an ith root of unity in L_c. If $m \in M$; $(f(y))m = f(y)$, $(g(y))m = g(d_m y) = c_m^{-1}b^{-1}y^h + 1$ and $bm = c_m(a_m)^h b$, where $c_m = (d_m)^{-h}$ and d_m is an ith root of unity. Since $g(y)$ is a proper factor of $f(y)$, $g(y) \notin L[y]_c$ and $c_m \neq 1$ for some $m \in M$.

Conversely, assume there exist positive integers h and i and a nonzero $b \in L$, such that $bm = c_m(a_m)^h b$ for every $m \in M$, where c_m is an ith root of unity and some $c_m \neq 1$. Let $g(y) = b^{-1}y^h + 1$, and let $f(y)$ be the product of the distinct polynomials $g(dy)$ where d is an $h \cdot i$th root of unity. $f(y)$ will have the form $(b')^{-1}y^{h \cdot i} + 1$ and $f(y) \in L[y]_c$. If the desired P–V extension $L\langle k\rangle$ existed, $f(k) \neq 1$ would be an element of $L\langle k\rangle_c = L_c$. If c is a root in L_c of $y^{h\cdot i} - (1 - f(k))^{-1}$, then $f(ck) = 0$ and some factor $g(cdk) = 0$. But then $0 = (g(cdk))m = c_m^{-1}b^{-1}(cdk)^h + 1 = 1 - c_m^{-1}$ for every $m \in M$, contrary to the assumption that some $c_m \neq 1$.

(4.3) COROLLARY. *Let L be an M-field of differential type and of characteristic zero such that L_c is algebraically closed. If $m_0 \in M$ is the identity automorphism on L and a_m, $m \in M$ and $m \neq m_0$, are elements of L, there exists a P–V extension of differential type $L\langle k\rangle$ of L such that $k \neq 0$ and $km = a_m k$ for every $m \in M$, $m \neq m_0$.*

Proof. Let $a_{m_0} = 1$, and let $L[y]$ and $L(y)$ be defined as in the proof of Corollary (4.2). The M-system of mappings on $L[y]$ consists of the identity automorphism m_0 and higher derivations, and these can be extended to $L(y)$ so that $L(y)$ is an M-field of differential type which is an M-extension of $L[y]$. By repetition of the argument in the beginning of the proof of Corollary (4.2), only the case when $L[y]_c$ contains polynomials of positive degree need be considered. Let $f(y) \in L[y]_c$ be a polynomial of positive degree, choose $f(y)$ so that $f(0) \neq 0$, and let $g(y)$ be an irreducible factor of $f(y)$, say $f(y) = q(y)\cdot(g(y))^h$ where h is a positive integer and $q(y)$ is not divisible by $g(y)$. Let $\{D_\alpha\}$ be a higher derivation on $L[y]$ contained in the M-system of mappings on $L[y]$. If $D_0 = m_0$, $(g(y))D_0 = g(y)$. Let i be a positive integer not greater than the rank of $\{D_\alpha\}$ and assume that $(g(y))D_\alpha$ is a multiple of $g(y)$ for $0 \leqq \alpha < i$. Then $0 = (f(y))D_i$ which is equal to a sum of terms divisible by $(g(y))^h$ plus the term $hq(y)\cdot(g(y))^{h-1}\cdot((g(y))D_i)$, and $(g(y))D_i$ must be divisible by $g(y)$. Consequently $g(y)$ generates a proper prime M-ideal I in $L[y]$, $L[y]/I$ is an algebraic extension of $L, y \notin I$ and, setting $k = y + I$, $L\langle k\rangle = L[y]/I$ is the desired P–V extension of L.

(4.4) COROLLARY. *Let L be an M-field, such that the M-system of mappings on L consists of the identity automorphism m_0 and infinite higher derivations and L_c is algebraically closed. If a_m, $m \in M$ and $m \neq m_0$, are elements of L, there exists a P–V extension of differential type $L\langle k\rangle$ of L such that $k \neq 0$ and $km = a_m k$ for every $m \in M$, $m \neq m_0$.*

Proof. Because of Corollary (4.3), only the case where L is a field of characteristic $p \neq 0$ need be considered. Let $a_{m_0} = 1$, and let $L[y]$ and $L(y)$ be defined as in the proof of Corollary (4.2). The argument is then analogous to the proof of Corollary (3.4).

5. Generalized Liouville extensions.

(5.1) DEFINITION. An M-field K which is an M-extension of an M-field L is a generalized Liouville extension of L if there exists a positive integer i and $i+1$ intermediate M-subfields of K, $L = L_0 \subseteqq L_1 \subseteqq \cdots \subseteqq L_i = K$, such that for each integer α, $1 \leqq \alpha \leqq i$, there exists $k \in L_\alpha$ such that $L_\alpha = L_{\alpha-1}\langle k\rangle$ and

(1) L_α is an algebraic extension of $L_{\alpha-1}$, or

(2) $km - (1m)k = a_m \in L_{\alpha-1}$ for every $m \in M$, or

(3) $km = a_m k$ where $a_m \in L_{\alpha-1}$ for every $m \in M$.

If L is an M-subfield of an M-field K, let $A_K(L)$ denote the M-Galois group of K over L. If G is a subgroup of $A_K(L)$, let $I(G)$ denote the set of all elements of K left fixed by the automorphisms in G; $I(G)$ is an M-subfield of K and $L \subseteqq I(G) \subseteqq K$. Suppose K is a solution field over an M-field L such that $K_c = L_c$ and $k_1, k_2, \cdots, k_j$ is a fundamental set for K over L. If $\varphi \in A_K(L)$, then $k_\alpha\varphi = \sum_{\beta=1}^{j} c_{\alpha\beta}k_\beta$, $1 \leqq \alpha \leqq j$, where $(c_{\alpha\beta})_{1\leqq\alpha,\beta\leqq j}$ is a matrix over $K_c = L_c$, by Theorem (3.2) of [7]. The structure of $A_K(L)$ may be determined analogously to the analysis of the differential Galois group presented in Kaplansky's *An Introduction to Differential Algebra*[3]. The results needed in the sequel will be summarized here. $A_K(L)$ is an algebraic matrix group over L_c and the algebraic subgroups of $A_K(L)$ are the subgroups $A_K(L')$ where L' is an intermediate M-subfield of K, $L \subseteqq L' \subseteqq K$. If H is the connected component of the identity element of $A_K(L)$, then H is an algebraic subgroup of finite index in $A_K(L)$. Therefore $H = A_K(\bar{L})$ where $\bar{L} = I(H)$ and $\bar{L}$ is a finite dimensional algebraic extension of $I(A_K(L))$. Moreover, $\bar{L}$ is algebraically closed in K. Indeed, if $k \in K$ is algebraic over $\bar{L}$, then $A_K(\bar{L}\langle k\rangle)$ is an algebraic subgroup of finite index in H since the left cosets of H mod $A_K(\bar{L}\langle k\rangle)$ are in one-to-one correspondence with the distinct images of k under the automorphisms in H. Because H is connected, $A_K(\bar{L}\langle k\rangle) = H$ and $k \in \bar{L}$.

(5.2) THEOREM. *Let K be a P–V extension of an M-field L. If the connected component of the identity element in $A_K(L)$ is a solvable group, then K is a generalized Liouville extension of $I(A_K(L))$.*

Proof. Let H be the connected component of the identity element in $A_K(L)$ and let $\bar{L} = I(H)$. $\bar{L}$ is a finite dimensional algebraic extension of $I(A_K(L))$. Since H is a connected, solvable algebraic matrix group over the algebraically closed field L_c, a fundamental set $k_1, k_2, \cdots, k_j$ for K over L may be chosen so that the M-automorphisms of H are represented by triangular matrices, say $k_\alpha\varphi = \sum_{\beta=\alpha}^{j} c_{\alpha\beta}(\varphi)\cdot k_\beta$ for $\varphi \in H$ and $1 \leqq \alpha \leqq j$, where the coefficients $c_{\alpha\beta}(\varphi) \in L_c$. If $m \in M$ and

$\varphi \in H$, then $((k_j m)k_j^{-1})\varphi = ((k_j\varphi)m)(k_j\varphi)^{-1} = ((c_{jj}(\varphi)\cdot(k_j m))(c_{jj}(\varphi)\cdot k_j)^{-1} = (k_j m)k_j^{-1}$ and $(k_j m)k_j^{-1} \in \bar{L}$. Thus $k_j m = a_m k_j$ where $a_m \in \bar{L}$ for every $m \in M$. If $m \in M$ and $\varphi \in H$, let $k'_\alpha(m) = (k_\alpha k_j^{-1})m - (1m)k_\alpha k_j^{-1}$ and $c'_{\alpha\beta}(\varphi) = c_{\alpha\beta}(\varphi)\cdot(c_{jj}(\varphi))^{-1}$ for $\alpha \leqq \beta \leqq j-1$ and $1 \leqq \alpha \leqq j-1$; then

$$\begin{aligned}(k'_\alpha(m))\varphi &= ((k_\alpha\varphi)(k_j\varphi)^{-1})m - (1m)(k_\alpha\varphi)(k_j\varphi)^{-1} \\ &= \sum_{\beta=\alpha}^{j} c_{\alpha\beta}(\varphi)\cdot(c_{jj}(\varphi))^{-1}((k_\beta k_j^{-1})m - (1m)k_\beta k_j^{-1}) \\ &= \sum_{\beta=\alpha}^{j-1} c'_{\alpha\beta}(\varphi)\cdot k'_\beta(m)\ .\end{aligned}$$

By Theorem (3.2) of [7], K is finitely generated as an abstract field over $\bar{L} \supseteqq L$; therefore every intermediate subfield is also finitely generated over $\bar{L}$. Consequently, if L' is the M-subfield of K generated over $\bar{L}$ by the $k'_\alpha(m)$, $m \in M$ and $1 \leqq \alpha \leqq j-1$, then there are finitely many $m \in M$ such that L' is generated as an M-field over $\bar{L}$ by the $k'_\alpha(m)$ for these m and $1 \leqq \alpha \leqq j-1$. By induction on j, it may be assumed that L' is a generalized Liouville extension of $I(A_K(L))$. Since $(k_\alpha k_j^{-1})m - (1m)k_\alpha k_j^{-1} = k'_\alpha(m) \in L'$ for every $m \in M$ and $1 \leqq \alpha \leqq j-1$ while $k_j m = a_m k_j$ where $a_m \in \bar{L} \subseteqq L'$ for every $m \in M$, it follows that K is a generalized Liouville extension of $I(A_k(L))$.

In connection with this theorem, it should be noted that $I(A_K(L)) = L$ if K is a regular field extention of L. If K is an M-field of differential type, then $I(A_K(L)) = L$ provided only that K is a separable field extension of L.

(5.3) LEMMA. *Let K', K, L' and L be M-fields such that K' is an M-extension of L, K and L' are M-subfields of K' and contain L, and K' is generated by its subfields K and L'.*

(i) *If K is a solution field over L, K' is a solution field over L' and a fundamental set for K over L is a fundamental set for K' over L'.*

(ii) *If K and L' are linearly disjoint over L, there is a canonical isomorphism of $A_K(L)$ into $A_{K'}(L')$. Moreover, if K is a solution field over L, $K_c = L_c$, $K'_c = L'_c$, and $A_K(L)$ and $A_{K'}(L')$ are represented by matrices with respect to the same fundamental set for K over L and K' over L'; then this canonical isomorphism is the identity map on matrices.*

Proof. (i) The verification is immediate from the definition of solution field.

(ii) If K and L' are linearly disjoint over L, automorphisms of K over L extend uniquely to automorphisms of K' over L' and M-automorphisms of K over L extend to M-automorphisms of K' over

L', yielding an isomorphism of $A_k(L)$ into $A_{K'}(L')$. The remaining assertion is immediate.

The converse of theorem (5.2) is a consequence of

(5.4) THEOREM. *If K' is a generalized Liouville extension of an M-field L and K is an intermediate M-subfield of K' such that K is a P–V extension of L, then the connected component of the identity element in $A_K(L)$ is a solvable group.*

Proof. By Corollary (2.3) of [7], K and K'_c are linearly disjoint over $K_c = L_c$, whence K and $L(K'_c)$ are linearly disjoint over L. By Lemma (5.3), $K(K'_c)$ is a solution field over $L(K'_c)$; and there is a matrix representation for the algebraic group $A_K(L)$ over L_c, a matrix representation for the algebraic group $A_{K(K'_c)}(L(K'_c))$ over $K'_c \supseteqq L_c$, and a canonical isomorphism of $A_K(L)$ into $A_{K(K'_c)}(L(K'_c))$ which is the identity map on matrices. If H is the connected component of the identity element in $A_K(L)$, then H is an irreducible component of $A_K(L)$ and its image in $A_{K(K'_c)}(L(K'_c))$ is irreducible, hence connected, since L_c is algebraically closed. Therefore H is mapped into the connected component of the identity element in $A_{K(K'_c)}(L(K'_c))$, and it will suffice to prove the theorem under the assumptions that K is merely a solution field over L but $K'_c = L_c$.

Let $L = L_0 \subseteqq L_1 \subseteqq \cdots \subseteqq L_i = K'$ be as in definition (5.1), and let $k \in L_1$ be such that $L_1 = L\langle k\rangle$ and

(1) L_1 is an algebraic extension of L, or

(2) $km - (1m)k = a_m \in L$ for every $m \in M$, or

(3) $km = a_m k$ where $a_m \in L$, for every $m \in M$.

Be induction on i, it may be assumed that the connected component of the identity element in $A_{K\langle k\rangle}(L_1)$ is solvable. Let $\bar{L} = I(H)$, where again H denotes the connected component of the identity element in $A_K(L)$. K is a regular extension of $\bar{L}$, since $\bar{L}$ is algebraically closed in K and $\bar{L}$ is the fixed field of a group of automorphisms of K whence K is a separable extension of $\bar{L}$. If L_1 is an algebraic extension of L, then $\bar{L}\langle k\rangle$ is an algebraic extension of $\bar{L}$ and K and $\bar{L}\langle k\rangle$ are linearly disjoint over $\bar{L}$. The canonical isomorphism of $H = A_K(\bar{L})$ into $A_{K\langle k\rangle}(\bar{L}\langle k\rangle)$ given by lemma (5.3) must map H into the connected component of the identity element in $A_{K\langle k\rangle}(L_1)$, whence H is solvable.

Assume L_1 is not an algebraic extension of L. If k is transcendental over K, then $L_1 = L(k)$ and $K\langle k\rangle = K(k)$ by Theorems (3.1) and (4.1). K and L_1 are linearly disjoint over L, so again there is a canonical isomorphism of H into the connected component of the identity element in $A_{K\langle k\rangle}(L_1)$ and H is solvable. Suppose k is algebraic over K. If $km = a_m k$ where $a_m \in L$ for every $m \in M$, then $k^h + b = 0$ where h is a positive integer, $b \in K$ and again $(bm)b^{-1} \in L$ for every

$m \in M$. If $km - (1m)k = a_m \in L$ for every $m \in M$, then L is a field of characteristic $p \neq 0$ and $k^{p^h} + c_{h-1}k^{p^{h-1}} + \cdots + c_1k^p + c_0k + b = 0$ where h is a positive integer, $c_\alpha \in K_c = L_c$ for $0 \leqq \alpha \leqq h-1$, $b \in K$ and again $bm - (1m)b \in L$ for every $m \in M$. In either case $L\langle b\rangle$ is invariant under the automorphisms in $A_K(L)$ and $A_{L\langle b\rangle}(L)$ is commutative, by Theorems (3.1) and (4.1). Therefore, $A_K(L\langle b\rangle)$ is an invariant subgroup of $A_K(L)$ and the factor group, which is isomorphic to a subgroup of $A_{L\langle b\rangle}(L)$, is commutative. L_1 is an algebraic extension of $L\langle b\rangle$ and, by a preceding argument, the connected component of the identity element in $A_K(L\langle b\rangle)$ is canonically isomorphic to a subgroup of the connected component of the identity element in $A_{K\langle k\rangle}(L_1)$ and is solvable. Therefore H, the connected component of the identity element in $A_K(L)$, is solvable by Lemma (4.9) of [3].

References

1. R. M. Cohn, *Extensions of difference fields*, Amer. J. Math. **74** (1952), 507-530.

2. C. H. Franke, *Picard-Vessiot theory of linear homogeneous difference equations*, Trans. Amer. Math. Soc. **108** (1963), 491-515.

3. Irving Kaplansky, *An introduction to differential algebra*, Hermann, Paris, 1957.

4. E. R. Kolchin, *Algebraic matric groups and the Picard-Vessiot theory of homogeneous linear ordinary differential equations*, Annals of Math. **49** (1948), 1-42.

5. ———, *Picard-Vessiot theory of partial differential fields*, Proc. Amer. Math. Soc. **3** (1952), 596-603.

6. H. F. Kreimer, *The foundations for an extension of differential algebra*, Trans. Amer. Math. Soc., **III** (1964), 482-492.

7. ———, *An extension of differential Galois theory*, accepted for publication by Trans. Amer. Math. Soc.

8. Lawrence Markus, *Group theory and differential equations*, Lecture notes at the University of Minnesota, Institute of Technology, Department of Mathematics.

PACIFIC JOURNAL OF MATHEMATICS
Vol. 15, No. 1, 1965

ON A LINEAR FORM WHOSE DISTRIBUTION IS IDENTICAL WITH THAT OF A MONOMIAL

R. G. LAHA AND E. LUKACS

Several authors studied identically distributed linear forms in independently and identically distributed random variables. J. Marcinkiewicz considered finite or infinite linear forms and assumed that the random variables have finite moments of all orders. He showed that the common distribution of the random variables is then the Normal distribution. Yu. V. Linnik obtained some deep results concerning identically distributed linear forms involving only a finite number of random variables. The authors have investigated in a separate paper the case where one of the linear forms contains infinitely many terms while the other is a monomial. They obtained a characterization of the normal distribution under the assumption that the second moment of the random variable is finite. In the present paper we investigate a similar problem and do not assume the existence of the second moment.

1. We prove the following theorem:

THEOREM. *Let $\{X_j\}$ be a finite or denumerable sequence of independently and identically distributed nondegenerate random variables and let $\{a_j\}$ be a sequence of real numbers such that the sum $\sum_j a_j X_j$ exists*[1]. *Let $\alpha \neq 0$ be a real number such that*
(i) *the sum $\sum_j a_j X_j$ is distributed as αX_1*
(ii) $$\sum_j a_j^2 \geqq \alpha^2.$$
Then the common distribution of the X_j is normal.

REMARK. The converse statement is evidently true provided that $\sum_j a_j = \alpha$ if the sum $\sum_j a_j X_j$ contains more than two terms or $\mathscr{E}(X_j) = 0$ in case $\sum_j a_j X_j$ has only two terms.

In § 2 we prove three lemmas, the third of these has some independent interest. In § 3 the theorem is proved.

Received October 14, 1963, and in revised form March 10, 1964. The work of the first author was supported by the National Science Foundation under grant GP-96. The work of the second author was supported by the U.S. Air Force under grant AF-AFOSR-473-63.

[1] We say that the infinite sum $\sum_j a_j X_j$ exists, if it converges almost everywhere. It is known (see Loève [3] pg. 251) that for a series of independent random variables the concepts of convergence almost everywhere and weak convergence are equivalent.

2. Lemmas. We denote the common distribution of the random variable X_j by $F(x)$ and write $f(t)$ for the corresponding characteristic function.

LEMMA 1. *Suppose that all the conditions of the theorem except* (ii) *are satisfied. Then* $\sup_j |a_j| < |\alpha|$.

According to the assumptions we have

$$\prod_j f(a_j t) = f(\alpha t) . \tag{2.1}$$

We set $b_j = a_j/\alpha$ $(j = 1, 2, \cdots)$ and obtain

$$\prod_j f(b_j t) = f(t) . \tag{2.2}$$

The lemma is proven if we show that $|b_j| < 1$ for all j. First we note that if $|b_j| = 1$ for at least one value of j, then X_j has necessarily a degenerate distribution. We consider the case where $|b_k| > 1$ for at least one value k. We see then from (2.2) that

$$1 \geqq |f(b_k t)| \geqq |f(t)|$$

which means

$$1 \geqq |f(t)| \geqq |f(t/b_k)| \geqq |f(t/b_k^2)| \geqq \cdots \lim_{n\to\infty} |f(t/b_k^n)| = f(0) = 1 .$$

Therefore $|f(t)| \equiv 1$ and the distribution of X_j is again degenerate. We conclude therefore that

$$|b_j| < 1 \qquad (j = 1, 2 \cdots) \tag{2.3}$$

LEMMA 2. *Suppose that all the conditions of the theorem, except* (ii), *are satisfied then the function* $f(t)$ *has no real zeros.*

We first remark that the existence of the infinite sum $\sum_j a_j X_j$ implies that the sequence of random variables $S_N = \sum_{j=N+1}^{\infty} a_j X_j$ converges to zero (as $N \to \infty$) with probability 1. It follows from the continuity theorem that

$$\lim_{N\to\infty} \prod_{j=N+1}^{\infty} f(a_j t) = 1 \tag{2.4}$$

uniformly in every finite t-interval.

Let $\varepsilon > 0$ be an arbitrarily small number and let T be a positive number. It follows then from (2.4) that there exists an $N_0 = N_0(\varepsilon, T)$ such that for all $N \geqq N_0$ the inequality

$$\left| \prod_{j=N+1}^{\infty} f(b_j t) - 1 \right| \leqq \varepsilon \tag{2.5}$$

holds uniformly for $|t| \leqq T$.

We give an indirect proof of Lemma 2. Suppose that the function $f(t)$ has real zeros and let t_0 be one of the zeros of $f(t)$ which is closest to the origin. Then

$$\prod_j f(b_j t_0) = f(t_0) = 0 ,$$

so that either $f(b_j t_0) = 0$ for at least one value of j or the product is infinite and diverges to zero at the point $t = t_0$. The first case is impossible by virtue of (2.3) while the second contradicts the uniform convergence of the infinite product so that Lemma 2 is proven.

LEMMA 3. *Let $\{X_j\}$ be a finite or denumerable sequence of independently and identically distributed nondegenerate random variables and let $\{a_j\}$ be a sequence of real numbers such that the sum $\sum_j a_j X_j$ exists. Let $\alpha \neq 0$ be a real number such that* $\sup_j |a_j| < |\alpha|$. *Suppose that the sum $\sum_j a_j X_j$ has the same distribution as αX_1, then the common distribution of each X_j is infinitely divisible.*

To prove Lemma 3 we write (2.2) in the form[2]

$$f(t) = f(b_1 t) f(b_2 t) \cdots f(b_N t) \Phi_N(t) \tag{2.6}$$

where

$$\Phi_N(t) = \prod_{j=N+1}^{\infty} f(b_j t) \tag{2.7}$$

and where N is so large that the inequality (2.5) holds. Using (2.6) we see that

$$f(t) = \prod_{j=1}^{N} f(b_j^2 t) \prod_{\substack{j,k=1 \\ j>k}}^{N} [f(b_j b_k t)]^2 \left[\prod_{j=1}^{N} \Phi_N(b_j) \right] \cdot \Phi_N(t) . \tag{2.8}$$

We repeat this process n times and obtain

$$\begin{aligned} f(t) = & \left\{ \prod_{j_1+\cdots+j_N=n} [f(b_1^{j_1} \cdots b_N^{j_N} t)]^{(n;j_1\cdots j_N)} \right\} \\ & \cdot \left\{ \prod_{k=1}^{n} \prod_{j_1+\cdots+j_N=n-k} [\Phi_N(b_1^{j_1} \cdots b_N^{j_N} t)]^{(n-k;j_1\cdots j_N)} \right\} . \end{aligned} \tag{2.9}$$

Here all $j_k \geqq 0$ and $(m\,; j_1 \cdots j_N) = m!/j_1! \cdots j_N!$. Formula (2.9) indicates that the random variable X, whose characteristic function is $f(t)$, is the sum of $k_n = N^n + N^{n-1} + \cdots + N^2 + N + 1$ independent random

[2] If the sequence $\{X_j\}$ is finite then N is equal to the number of variables X_j so that $\Phi_N(t) \equiv 1$.

variables $X_{n,k}(k - 1, 2, \cdots, k_n)$, that is $X = \sum_{k=1}^{k_n} X_{n,k}$ for every n.

Such sequences of sums of independent random variables occur in the study of the central limit theorem, and we give next a few results which we wish to apply.

We say that the summands $X_{n,k}$ are uniformly asymptotically negligible (u.a.n), if $X_{n,k}$ converges in probability to zero, uniformly in k, as n tends to infinity; this means that for any $\varepsilon > 0$

$$\lim_{n\to\infty} \max_{1\leqq k\leqq k_n} P(|X_{n,k}| \geqq \varepsilon) = 0 . \tag{2.10}$$

It is known (see Loève [3] pg. 302) that condition (2.10) is equivalent to

$$\lim_{n\to\infty} \max_{1\leqq k\leqq k_n} |f_{n,k}(t) - 1| = 0 \tag{2.11}$$

uniformly in every finite t-interval.

Let $X_{n,k}$ $(k = 1, 2, \cdots, k_n)$ be, for each n, a finite set of independent random variables and suppose that the $X_{n,k}$ are u.a.n. Then the limiting distribution (as n tends to infinity) of the sums $\sum_{k=1}^{k_n} X_{n,k}$ is infinitely divisible.

For the proof we refer the reader to Loève [3] (pg. 309).

We turn now to the proof of Lemma 3 and show that the factors of (2.9) satisfy condition (2.11).

Let $\varepsilon > 0$ be an arbitrarily small number and $T > 0$. We see from (2.5) and (2.7) that we can select a sufficiently large N such that

$$|\Phi_N(t) - 1| \leqq \varepsilon \tag{2.12}$$

uniformly in $|t| \leqq T$. Since $|b_j| < 1$ we have

$$|b_1^{j_1} \cdots b_N^{j_N} t| < T$$

so that, according to (2.12),

$$|\Phi_N(b_1^{j_1} \cdots b_N^{j_N} t) - 1| \leqq \varepsilon \tag{2.13}$$

uniformly in $|t| \leqq T$ for the chosen value of N.

We consider next a typical factor $f(b_1^{j_1} \cdots b_N^{j_N} t)$ of the product in the first brace of formula (2.9). Here $j_1 + j_2 + \cdots + j_N = n$ and $j_k \geqq 0$ so that at least one of the j_k is positive. We show now that it is possible to choose an $n_0 = n_0(\varepsilon, T)$ such that for $n \geqq n_0$

$$\gamma_{j_1 \cdots j_N}(t) = |f(b_1^{j_1} \cdots b_N^{j_N} t) - 1| \leqq \varepsilon \tag{2.14}$$

uniformly in $|t| \leqq T$.

Clearly,

$$(2.15)\qquad \gamma_{j_1\cdots j_N}(t) \leqq \left|\int_{|x|\geqq A}\{\exp[ib_1^{j_1}\cdots b_N^{j_N}tx]-1\}dF(x)\right| + \left|\int_{|x|<A}\{\exp[ib_1^{j_1}\cdots b_N^{j_N}tx]-1\}dF(x)\right| .$$

We choose A so large that

$$(2.16)\qquad \left|\int_{|x|\geqq A}\{\exp[ib_1^{j_1}\cdots b_N^{j_N}tx]-1\}dF(x)\right| \leqq 2\int_{|x|\geqq A}dF(x) \leqq \frac{\varepsilon}{2} .$$

We note that

$$(2.17)\qquad \left|\int_{|x|<A}\{\exp[ib_1^{j_1}\cdots b_n^{j_N}tx]-1\}dF(x)\right| \leqq \left|b_1^{j_1}\cdots b_N^{j_N}\right| TA .$$

We select now an $n^* = n^*(j_1, \cdots, j_N; T, \varepsilon)$ so large that for $n \geqq n^*$ the inequality

$$(2.18)\qquad |b_1^{j_1}\cdots b_N^{j_N}|\, TA \leqq \frac{\varepsilon}{2}$$

holds. This is possible in view of (2.3). There are altogether N^n terms of the form $f(b_1^{j_1}\cdots b_N^{j_N}t)$ in (2.14) and we choose

$$(2.19)\qquad n_0 = n_0(\varepsilon, T) = \max_{j_1\cdots j_N} n^*(j_1, \cdots, j_N; T, \varepsilon) ;$$

then (2.14) follows from (2.16), (2.17), (2.18) and (2.19).

We see therefore that the set of independent random variables $X_{n,k}$ satisfies the u.a.n. condition (2.11). Therefore the distribution of X is infinitely divisible and Lemma 3 is proven.

Since $f(t)$ is an infinitely divisible characteristic function, it admits the Lévy-Khinchine representation

$$(2.20)\qquad \ln f(t) = i\alpha t - \beta t^2/2 + \int_{-\infty}^{-0}\left(e^{itx}-1-\frac{itx}{1+x^2}\right)\frac{1+x^2}{x^2}dG(x) + \int_{+0}^{\infty}\left(e^{itx}-1-\frac{itx}{1+x^2}\right)\frac{1+x^2}{x^2}dG(x)$$

where α and β are real numbers, $\beta \geqq 0$, and where $G(x)$ is a non-decreasing, right-continuous function such that $G(-\infty) = 0$ and $G(+\infty) = K < \infty$. Let now $f(t)$ be the characteristic function of an infinitely divisible symmetric distribution, so that $f(t) = f(-t)$. In this case one sees after some elementary transformations of the integrals in (2.20) that

$$(2.21)\qquad G(x) + G(-x-0) = C$$

for all $x \neq 0$. Using (2.20) and (2.21) we see that the characteristic function of a symmetric infinitely divisible distribution admits the representation

$$\ln f(t) = -\beta t^2/2 + \int_{+0}^{+\infty} (\cos tx - 1)\frac{1+x^2}{x^2}dH(x) \tag{2.22}$$

where

$$H(x) = \begin{cases} 2G(x) - C & \text{for } x > 0 \\ 0 & \text{for } x < 0\,. \end{cases} \tag{2.22a}$$

Thus $H(x)$ is a non decreasing, right-continuous, bounded function and $H(x)$ and $G(x)$ determine each other uniquely.

3. Proof of the theorem. We introduce the function

$$g(t) = f(t)f(-t) \tag{3.1}$$

and conclude from (2.2) that the relation

$$\prod_j g(b_j t) = g(t) \tag{3.2}$$

holds for all real t. Here $g(t)$ is the characteristic function of a symmetric distribution and is therefore a real and even function. It is no restriction to assume that

$$0 \leqq b_j < 1 \qquad (j = 1, 2, \cdots) \tag{3.3a}$$

where

$$\sum_{j=1}^{\infty} b_j^2 \geqq 1\,. \tag{3.3b}$$

According to (2.22) we have then the representation

$$\ln g(t) = -\beta t^2/2 + \int_{+0}^{\infty} (\cos tx - 1)\frac{1+x^2}{x^2}dH(x) \tag{3.4}$$

where $\beta \geqq 0$ and where $H(x)$ is a nondecreasing, right-continuous and bounded function. We use (3.4) and (3.3b) and obtain from (3.2) the relation

$$\begin{aligned} \sum_{j=1}^{\infty}\int_{+0}^{\infty} (\cos b_j tx - 1)\frac{1+x^2}{x^2}dH(x) \\ = K\frac{t^2}{2} + \int_{+0}^{\infty} (\cos tx - 1)\frac{1+x^2}{x^2}dH(x)\,. \end{aligned} \tag{3.5}$$

where

$$K = \beta\left[\sum_{j=1}^{\infty} b_j^2 - 1\right] \geqq 0\,.$$

We define the sequence $\{\psi_\nu(t)\}$ by

$$\psi_\nu(t) = \sum_{j=1}^{\nu} \int_{+0}^{\infty} (\cos b_j tx - 1) \frac{1+x^2}{x^2} dH(x) \tag{3.6}$$

so that

$$\lim_{\nu\to\infty} \psi_\nu(t) = \psi(t) = K\frac{t^2}{2} + \int_{+0}^{\infty} (\cos tx - 1)\frac{1+x^2}{x^2} dH(x) \tag{3.7}$$

for every real t.

Since $\psi(t)$ is the characteristic function of an infinitely divisible distribution it follows that $K \leqq 0$, so that we conclude from assumption (ii) that $K = 0$ and $\sum_{j=1}^{\infty} b_j^2 = 1$.

By a change of the variable of integration in (3.6) we obtain

$$\psi_\nu(t) = \int_{+0}^{\infty} (\cos tx - 1)\frac{1+x^2}{x^2}\left[\sum_{j=1}^{\nu} \frac{b_j^2 + x^2}{1+x^2} dH(x/b_j)\right].$$

We write

$$H_\nu(x) = \begin{cases} \int_{+0}^{x} \left[\sum_{j=1}^{\nu} \frac{b_j^2+y^2}{1+y^2} dH(y/b_j)\right] & \text{for } x > 0 \\ 0 & \text{for } x < 0\,. \end{cases} \tag{3.8}$$

Therefore we have, for every ν,

$$\psi_\nu(t) = \int_{+0}^{+\infty} (\cos tx - 1)\frac{1+x^2}{x^2} dH_\nu(x)\,. \tag{3.9}$$

It follows then from (3.7) and (3.8) that

$$\lim_{\nu\to\infty} H_\nu(x) = H(x) \tag{3.10}$$

for every x which is a continuity point of $H(x)$. The proof is carried in the same way in which the convergence theorem is proven (see Loève [3] pp. 300–301).

In view of (3.3a) we have

$$\frac{b_j^2 + y^2}{1+y^2} \geqq b_j^2 \qquad (j = 1, 2, \cdots)$$

so that we conclude from (3.8) that

$$H_\nu(x) \geqq \sum_{j=1}^{\nu} b_j^2 H\left(\frac{x}{b_j}\right) \tag{3.11}$$

for all ν.

It follows from (3.10) and (3.11) that

$$H(x) = \lim_{\nu\to\infty} H_\nu(x) \geqq \sum_{j=1}^{\infty} b_j^2 H\left(\frac{x}{b_j}\right)$$

for all $x > 0$ which are continuity points of $H(x)$.

Using equation (3.3b) we obtain

$$\sum_{j=1}^{\infty} b_j^2\left[H(x) - H\left(\frac{x}{b_j}\right)\right] \geqq 0 . \tag{3.12}$$

Since $H(x)$ is a nondecreasing function, we see from (3.3a) that

$$H(x) \leqq H\left(\frac{x}{b_j}\right) . \tag{3.13}$$

It follows from (3.12) and (3.13) that

$$H(x) = H\left(\frac{x}{b_j}\right)$$

for every $x > 0$ which is a continuity point of $H(x)$. Therefore

$$H(x) = H(+\infty) = C$$

for $x > 0$. We now turn to equation (3.4) and get

$$\ln g(t) = -\beta t^2/2 . \tag{3.14}$$

The statement of the theorem is an immediate consequence of (3.1) and of Cramér's theorem.

References

1. R. G. Laha-E. Lukacs, *On linear forms and stochastic integrals,* (to be published).
2. Yu. V. Linnik, *Linear forms and statistical criteria I, II,* Ukrainski Mat. Zurnal **5** (1953), 207-243 and 247-290. [English translation in Selected Translations in Mathematical Statistics and Probability vol. 3, pp. 1-90, American Math. Soc., Providence, R.I. (1962)].
3. M. Loève, *Probability Theory* (third edition), D. Van Nostrand, New York (1963).
4. J. Marcinkiewicz, *Sur une propriété de la loi de Gauss,* Math. Zeitschr. **44** (1949), 612-618.

THE CATHOLIC UNIVERSITY OF AMERICA

PACIFIC JOURNAL OF MATHEMATICS
Vol. 15, No. 1, 1965

SINGULARITIES OF SUPERPOSITIONS OF DISTRIBUTIONS

DONALD LUDWIG

Distributions of the form

$$F(x,\lambda) = \frac{1}{\Gamma\left(\frac{\lambda+1}{2}\right)} \int |f(x,u)|^{\lambda} g(x,u)du \tag{1}$$

are considered, where x and u belong to R^p and R^n respectively. The parameter λ is complex, and $F(x,\lambda)$ is evaluated for $Re(\lambda) < 0$ by analytic continuation. Such integrals arise in solution formulas for partial differential equations. In case $n = 1$ or $n = 2$, F is expressed in terms of homogeneous distributions of degree $> \lambda + \alpha$, where α is nonnegative and depends upon the geometry of the roots of f. The case of general n is also treated, in case the Hessian of f with respect to u is different from zero. The results lead to asymptotic expansions of analogous multiple integrals.

We assume that f and g are C^∞ real-valued functions, and we assume that the gradient of f with respect to x does not vanish in the region of $R^p \times R^n$ under consideration. Integration is taken over a compact region $U \subset R^n$, and we assume that g has its support in the interior of U. For $Re(\lambda) > 0$, the operation of F on a test function φ is defined by $I(\lambda) = \int F\varphi dx$. For other values of λ, $I(\lambda)$ is evaluated by an analytic continuation in λ. The factor $1/\Gamma[(\lambda + 1)/2]$ ensures that $I(\lambda)$ is an entire function of λ. We actually require only a finite number of derivatives of f and g, provided that $Re(\lambda)$ is bounded from below.

It is easy to see that, after a change of variables in x-space, $F_1(x_1, \lambda; x_2, \cdots, x_p) = F(x_1, x_2, \cdots, x_p, \lambda)$ is a distribution in x_1, with $x_2, \cdots, x_p$ regarded as parameters. In case $n = 1$ or $n = 2$, we show that F_1 may be expressed as a sum of homogeneous distributions, plus a smooth remainder. Each term in the expansion of F_1 is associated with a point or points where $f(x, u) = 0$ and $(\partial f/\partial u)(x, u) = 0$. Expressions such as $(\partial f/\partial x)$ and $(\partial f/\partial u)$ denote the gradients with respect to the x and u variables, respectively. In case $n = 1$, the most singular term of F_1 has the degree $\lambda + (1/m)$, if f has order m with respect to u at the corresponding point. In case $n = 2$, the degree of the most singular term of F_1 depends upon the geometry of the real roots

Received February 6, 1964. This research was supported in part by National Science Foundation Grant G-22982.

of f, regarded as functions of (u_1, u_2) for fixed x. The degree of the singularity varies between $\lambda + (1/m)$ and $\lambda + (2/m)$, if f has order m with respect to u at the point in question. The extreme values of the degree are assumed in case all roots of f are coincident, or distinct, respectively. We also consider higher values of n, in the case where the Hessian matrix $(\partial^2 f/\partial u_i \partial u_j)$ is nonsingular, which frequently arises in applications. In this case, the most singular part of F is homogeneous of degree $\lambda + (n/2)$.

Integrals of the form (1) arise in representations of solutions of hyperbolic partial differential equations, specifically the Herglotz—Petrovsky formula and its generalizations. (See I. M. Gelfand and G. E. Shilov [7] pp. 137–141, and R. Courant [2], pp. 727–733.) We shall apply the results of the present paper to the analysis of the singularities of fundamental solutions of linear hyperbolic equations in a forthcoming revision of [10].

Our results also have implications for the asymptotic behavior of single and double integrals, using a device of D. S. Jones and M. Kline [8]. Let

$$I(k) = \int \exp\,[ikf(u)]g(u)du \ .$$

Then

$$I(k) = \int e^{ikt}\, h(t)dt\ , \quad \text{where } h(t) = \int \delta(t - f(u))g(u)du\ .$$

Here δ represents the one-dimensional Dirac function. The behavior of $I(k)$ for large k is determined by the singularities of $h(t)$ (see A. Erdelyi [4], pp. 46–51.) But $h(t)$ is of the form (1), if we set $\lambda = -1$. For double integrals, our results extend those of D. S. Jones and M. Kline [8] and J. Focke [5] to give asymptotic expansions in cases where all derivatives of f of second order vanish at some point.

The outline of our work is as follows: the first section is devoted to preliminary remarks, which apply for any n. We show that F is a distribution in a single variable, and that singularities of F_1 at x_0 are associated with points u where $f(x_0, u) = 0$ and $(\partial f/\partial u)(x_0, u) = 0$. In the second section, we reduce the case $n = 1$ to consideration of an integral of the form

$$(2) \qquad I(x, \lambda, \alpha) = \gamma(\lambda)\int_0^a |\, x + u\,|^\lambda\, u^{\alpha-1} du\ ,$$

where α is a real number. Here and henceforth, we write $\gamma(\lambda) = 1/\Gamma[(\lambda + 1)/2]$. We analyze the singularities of (2) for arbitrary complex λ, and for $Re(\alpha) > 0$, using analytic continuation in both λ and α. The result is that $I(x, \lambda, \alpha)$ is the sum of a homogeneous distribution of degree $\lambda + \alpha$, and a smooth function. In the third section, we consider double integrals. We resolve the singularities of the zeros of f by a series of quadratic transformations, and reduce the problem to consideration of integrals of the form

$$(3) \qquad I(x, \lambda, \alpha, \beta) = \gamma(\lambda)\iint |x + u^\alpha v^\beta|^\lambda u^{\gamma-1}v^{\delta-1}g(u, v, \lambda)dudv \ .$$

In the fourth section, we expand (3) in powers of x. The integral is reduced to the form (2), or

$$(2') \qquad I'(x, \lambda, \alpha) = \gamma(\lambda)\int_0^a (x + u)^\lambda u^{\alpha-1} \log u du \ .$$

$I'(x, \lambda, \alpha)$ is just the derivative of $I(x, \lambda, \alpha)$ with respect to α. The fifth section is devoted to the simpler case of integrals where the Hessian of f with respect to u does not vanish. In this case, the leading singularity of F has degree $\lambda + (n/2)$.

Our procedures, especially in the case of double integrals, would be rather unwieldy for purposes of calculation. A simpler scheme is presented by G. F. D. Duff [3]. Our results may be regarded as a justification of certain of his methods. Our methods and results, especially in §§ 2 and 5, have much in common with L. Gårding [6].

1. General remarks. In this section, we shall first show that integrals of the form (1) define distributions in a single variable, with smooth (in distribution sense) dependence on the other variables as parameters. Then we show that the singularities of such integrals are associated with points where f and $\partial f/\partial u$ both vanish. This fact is the analog of the principle of stationary phase for asymptotic expansion of integrals.

To show that F, given by (1), is a distribution in one variable, we assume that $\partial f/\partial x_1$ is bounded away from zero in the region under consideration. Recalling our assumption that $(\partial f/\partial x) \neq 0$, we can arrange that $(\partial f/\partial x_1) \neq 0$ by taking a partition of unity in x, u space, and then rotating coordinates in x-space.

THEOREM 1.1. *If, for* $u \in U$, $a \leqq x_1 \leqq b$, *and for* $(x_2, \cdots, x_p)$ *belonging to an open subset of* R^{p-1}, *we have* $|(\partial f/\partial x_1)| \geqq \alpha > 0$, *and if* $\varphi(x_1) \in C^\infty$ *with support in* (a, b), *then* $I(\lambda)$, *given by the continuation of*

$$(1.1) \qquad I(\lambda) = \gamma(\lambda)\int F(x, \lambda)\varphi(x_1)dx_1 \ ,$$

depends continuously on φ *in the* C_0^∞ *topology, and smoothly on* $x_2, \cdots, x_p$. $I(\lambda)$ *is an entire analytic function of* λ. *We recall that*

$\gamma(\lambda) = 1/\Gamma[(\lambda + 1)/2]$.

Proof. We may rewrite (1.1) as a double integral, first choosing $Re(\lambda) > 0$. Then

$$I(\lambda) = \gamma(\lambda)\int_a^b\int_U |f(x, u)|^\lambda g(x, u) du \varphi(x_1) dx_1$$
$$= \gamma(\lambda)\int_U\int_a^b |f(x, u)|^\lambda g(x, u)\varphi(x_1) dx_1 du \ .$$

Now we introduce f as a variable of integration;

$$I(\lambda) = \gamma(\lambda)\int_U\int_\alpha^\beta |f|^\lambda \psi(f, u, x_2, \cdots, x_p) df du \ ,$$

where

$$\psi(f, u, x_2, \cdots, x_p) = \frac{g(X, u)\varphi(x_1)}{\dfrac{\partial f}{\partial x_1}(X, u)} \ ,$$

$X_j = x_j (j = 2, \cdots, p)$, and $X_1(f, u, x_2, \cdots, x_p)$ is defined by the relation $f(X, u) = f$. Clearly ψ and its derivatives with respect to $x_2, \cdots, x_p$ are in C_0^∞ with respect to f, depending continuously on φ in the topology of test functions. Hence it suffices to show that an integral of the form

$$J(\lambda) = \gamma(\lambda)\int |f|^\lambda \psi(f) df \ , \tag{1.2}$$

defines an analytic functional of ψ. Following I. M. Gelfand and G. E. Shilov [7], we write, with an arbitrary positive integer k,

$$J(\lambda) = \gamma(\lambda)\int_{-1}^{1} |f|^\lambda \left[\psi(f) - \sum_{j=0}^{k} \psi^{(j)}(0)\frac{f^j}{j!}\right] df$$
$$+ \gamma(\lambda)\sum_{j=0}^{k}\psi^{(j)}(0)\int_{-1}^{1}|f|^\lambda \frac{f^j}{j!} df + \gamma(\lambda)\int_{|f|\geqq 1}|f|^\lambda \psi(f) df \ .$$

The first and third terms are regular in λ for $Re(\lambda) > -k - 1$; the second term is easily evaluated as

$$\gamma(\lambda)\sum_{0\leqq 2l \leqq k}\psi^{(2l)}(0)\frac{1}{(\lambda + 2l + 1)(2l)!} \ .$$

Hence, since $\gamma(\lambda)$ has zeros for $\lambda = -2l - 1$, $l =$ integer $\geqq 0$, $J(\lambda)$ is an entire functional. Thus $I(\lambda)$ is also an entire functional.

According to the principle of stationary phase, the singularities of F arise from interior points where both f and $\partial f/\partial u$ vanish, or from boundary points where f vanishes and $\partial f/\partial u$ is normal to the boundary.

(See D. S. Jones and M. Kline [8].) We wish to consider only interior stationary points, and hence we assume that the support of $g(x, u)$ is in the interior of U.

THEOREM 1.2. *If the support of $g(x, u)$ is in the interior of U, and if, at a point x_0, $f(x_0, u)$ and $(\partial f/\partial u)(x_0, u)$ do not both vanish anywhere in U, then there exists a neighborhood of x_0 in which $F(x, \lambda)$ is smooth for all λ.*

Proof. Let $K = \inf_{u \in U} \{| f(x_0, u) |^2 + | (\partial f/\partial u)(x_0, u) |^2\}$. At each point $u_0 \in U$, we have either

(a) $| f(x_0, u_0) |^2 \geqq K/2$, or

(b) $| (\partial f/\partial u)(x_0, u_0) | \geqq K/2$.

Hence we can find a neighborhood of (x_0, u_0) in which either

(a) $| f |^2 > K/4$, or

(b) $| \partial f/\partial u |^2 > K/4$.

Such a neighborhood contains the product of an open ball $B(x_0) \subset R^p$, with center at x_0, and an open ball $B(u_0) \subset R^n$, with center at u_0. The set of such balls $B(u_0)$ forms an open covering of U, which can be reduced to a finite covering since U is compact. The intersection of the corresponding $B(x_0)$ is open. We denote this intersection by $C(x_0)$.

Thus, to each $u_0 \in U$ is associated an open set $N(u_0)$ in which either

(a) $| f |^2 > K/4$, or

(b) $| \partial f/\partial u |^2 > K/4$, for $x \in C(x_0)$, $u \in N(u_0)$.

We choose a C^∞ partition of unity subordinate to our finite covering of U. In sets of type (a), the integrand in (1) is C^∞ for $x \in C(x_0)$, for all λ. In sets of type (b), we may introduce f as variable of integration and proceed as in the proof of Theorem 1.1. Here x plays the role of a parameter. Thus integrals over sets of type (b) define functionals which are entire in λ, and which are C^∞ with respect to x.

2. **Single integrals.** In this section, we consider the case $n = 1$, i.e. where U is an interval of the real line. We shall obtain a description of the singularity of F near x_0, associated with a neighborhood of a point u_0 where $f(x_0, u_0) = 0$ and $(\partial f/\partial u)(x_0, u_0) = 0$. According to Theorem 1.2, every singularity of F corresponds to such a neighborhood. First we make a change of variables involving both x and u, and obtain an integral of the same type, where $f(x, u) = x_1 + u^m$. Theorem 2.1 states that, for fixed λ, $F(x, \lambda)$ is bounded if $g(x, u)$ vanishes sufficiently rapidly at $u = 0$. Thus, applying Taylor's theorem to g as function of u, we see that the singularities of F arise from terms of the form $\int | x_1 + u^m |^\lambda u^k du$. Finally, Theorems 2.2 and 2.3 show that

such an integral is the sum of a distribution homogeneous of degree $\lambda + (k+1)/m$ and a regular function.

Without loss of generality, we may assume that $x_0 = 0$ and $u_0 = 0$, and $(\partial f/\partial x_1)(0, 0) \neq 0$. We assume further that, at $(0, 0)$,

$$f = \frac{\partial f}{\partial u} = \cdots \frac{\partial^{m-1} f}{\partial u^{m-1}} = 0 , \qquad \frac{1}{m!} \frac{\partial^m f}{\partial u^m}(0, 0) \neq 0 .$$

We fix $x_2 = \cdots x_p = 0$, and denote x_1 by x. From Taylor's theorem,

$$\begin{aligned} f(x, u) &= f(0, u) + x e_1(x, u) \\ &= e_1(x, u)\left(x + \frac{f(0, u)}{e_1(x, u)}\right) . \end{aligned}$$

Here e_1 is a smooth function; $e_1(0, 0) = (\partial f/\partial x_1)(0, 0)$. Since f is of order m at the origin, we may write

$$f(x, u) = e_1(x, u)(x + u^m e_2(x, u)) ,$$

where $e_2(x, u)$ is smooth, and $e_2(0, 0) = \{[\partial^m f(0, 0)/\partial u^m]/[m!(\partial f/\partial x_1)(0,0)]\}$. If x and u are sufficiently small, the implicit function theorem implies that we may introduce a new variable of integration, $v = u \,|\, e_2(x, u) \,|^{1/m}$; thus we obtain

$$(2.1) \qquad \gamma(\lambda)\int | f |^\lambda g du = \gamma(\lambda)\int | x \pm v^m |^\lambda g_1(x, v; \lambda) dv ,$$

where

$$g_1(x, v; \lambda) = | e_1(x, u) |^\lambda g(x, u) \frac{du}{dv} .$$

By replacing x by $-x$ if necessary, we may bring (2.1) into the form where the plus sign holds.

Now we wish to apply Taylor's theorem to $g_1(x, v; \lambda)$, obtaining a polynomial in v, with a remainder which vanishes rapidly as $v \to 0$. First we show that, for fixed λ, the corresponding term in the expansion of F will be continuous, and can be made as smooth as desired.

THEOREM 2.1. *If $g(x, u; \lambda)$ has l derivatives with respect to u, and if $Re(\lambda) = \lambda_1 > -l - 1$, and if $m\lambda_1 + k + 1 > 0$, then*

$$(2.2) \qquad I(x, \lambda) = \gamma(\lambda)\int_0^a | x + u^m |^\lambda u^k g(x, u, \lambda) du$$

is continuous and bounded as a function of x.

Proof. We set $\xi = | x |^{1/m}$, and write $I = I_1 + I_2$, with

$$(2.3) \qquad I_1 = \gamma(\lambda)\int_0^{2\xi} | x + u^m |^\lambda u^k g(x, u, \lambda) du ,$$

$$I_2 = \gamma(\lambda)\int_{2\xi}^{a} | x + u^m |^\lambda u^k g(x, u, \lambda) du \ . \tag{2.4}$$

In (2.3), we introduce $u = \xi v$. Then

$$I_1 = \gamma(\lambda) | x |^{\lambda+(k+1)/m} \int_0^2 | \operatorname{sgn} x + v^m |^\lambda v^k g(x, \xi v, \lambda) dv \ .$$

Continuing this expression with respect to λ in the usual way (see proof of Theorem 1.1), we see that if $m\lambda_1 + k + 1 > 0$, I_1 is continuous and bounded. We may rewrite (2.4) as

$$I_2 = \gamma(\lambda)\int_{2\xi}^{a} \left| 1 + \frac{x}{u^m} \right|^\lambda u^{k+m\lambda} g(x, u; \lambda) du \ .$$

Hence,

$$| I_2 | \leqq \gamma(\lambda) \left| 1 - \frac{1}{2^m} \right|^{\lambda_1} \sup_{0 \leqq u \leqq a} | g(x, u; \lambda) | \int_0^a u^{k+m\lambda_1} du \ ,$$

which is clearly bounded if $k + m\lambda_1 + 1 > 0$. The continuity of I_2 follows similarly from the uniform continuity of the integrand.

We remark that smoothness of (2.2) for sufficiently large k follows from formal differentiation of (2.2), and application of Theorem 2.1.

Applying Taylor's theorem to $g_1(x, v; \lambda)$ appearing in (2.1), we see that

$$\gamma(\lambda)\int | f |^\lambda g(x, u, \lambda) du = \sum_{j=0}^{k} g^{(j)}(x, \lambda)\gamma(\lambda)\int | x + v^m |^\lambda v^j dv + \gamma(\lambda)\int | x + v^m |^\lambda v^{k+1} g_2(x, v, \lambda) dv \ . \tag{2.5}$$

Theorem 2.1 implies that the remainder is smooth in x for fixed λ, if k is sufficiently large. Evaluation of the singularities of F is therefore reduced to evaluation of the singularities of integrals of the form

$$I(x, \lambda) = m\gamma(\lambda)\int_0^a | x + v^m |^\lambda v^{n-1} dv \ . \tag{2.6}$$

A change of variables yields an integral of the form

$$I(x, \lambda) = \gamma(\lambda)\int_0^a | x + u |^\lambda u^{\alpha-1} du \ , \tag{2.7}$$

where $\alpha = (n/m)$.

In order to describe the singularities of (2.7) and related integrals, we shall require some facts about certain homogeneous distributions. We set

$$x_+ = \max(x, 0) \ , \qquad x_- = \max(-x, 0) \ .$$

LEMMA 2.1. *The functionals* $[1/\Gamma(\lambda+1)]x_+^\lambda$ *and* $[1/\Gamma(\lambda+1)]x_-^\lambda$ *are entire analytic functionals. Moreover,*

$$\begin{cases} \dfrac{1}{\Gamma(\lambda+1)} x_+^\lambda \Big|_{\lambda=-p} = \delta^{(p-1)}(x) & (p=1, 2, \cdots) \\ \dfrac{1}{\Gamma(\lambda+1)} x_-^\lambda \Big|_{\lambda=-p} = (-1)^{p-1}\delta^{(p-1)}(x) & (p=1, 2, \cdots) . \end{cases} \tag{2.8}$$

The proof is in I. M. Gelfand and G. E. Shilov [7], pp. 56–65. It is similar to the latter part of the proof of Theorem 1.1.

The following theorem leads immediately to results about (2.7).

THEOREM 2.2. *If* $Re(\alpha) > 0$, *the integral*

$$J_+(x, \lambda) = \int_0^a \frac{(x+u)_+^\lambda}{\Gamma(\lambda+1)} \frac{u^{\alpha-1}}{\Gamma(\alpha)} du \tag{2.9}$$

may be represented in the form

$$\begin{aligned} J_+(x, \lambda) = a_+(\lambda, \alpha) & \frac{x_+^{\lambda+\alpha}}{\Gamma(\lambda+\alpha+1)} \\ & + a_-(\lambda, \alpha) \frac{x_-^{\lambda+\alpha}}{\Gamma(\lambda+\alpha+1)} + R(x, \lambda, \alpha) . \end{aligned} \tag{2.10}$$

Here $R(x, \lambda, \alpha)$ *is a smooth function of* x *for small* x, *which is regular in* λ *and* α, *except for simple poles where* $\lambda + \alpha$ *is a nonnegative integer. The coefficients* a_+ *and* a_- *are regular except for simple poles where* $\lambda + \alpha$ *is an integer. The sum of the residues at the poles is zero, since* $J_+(x, \lambda)$ *is regular. We have*

$$a_+(\lambda, \alpha) = \frac{\sin \pi\lambda}{\sin \pi(\lambda+\alpha)} , \qquad a_-(\lambda, \alpha) = \frac{-\sin \pi\alpha}{\sin \pi(\lambda+\alpha)} . \tag{2.11}$$

We also have, for small x,

$$J_-(x, \lambda) = \int_0^a \frac{(x+u)_-^\lambda}{\Gamma(\lambda+1)} \frac{u^{\alpha-1}}{\Gamma(\alpha)} du = \frac{x_-^{\lambda+\alpha}}{\Gamma(\lambda+\alpha+1)} . \tag{2.12}$$

Proof. We shall use analytic continuation in λ and α. First we assume that $-1 < Re(\lambda) < -1/2$, $0 < Re(\alpha) < 1/2$. Then we may write

$$J_+(x, \lambda) = \int_0^\infty \frac{(x+u)_+^\lambda}{\Gamma(\lambda+1)} \frac{u^{\alpha-1}}{\Gamma(\alpha)} du + R(x, \lambda, \alpha) , \tag{2.13}$$

with

$$R(x, \lambda, \alpha) = -\int_a^\infty \frac{(x+u)_+^\lambda}{\Gamma(\lambda+1)} \frac{u^{\alpha-1}}{\Gamma(\alpha)} du . \tag{2.14}$$

The first integral in (2.13) may be treated by setting $u = |x| v$. The resulting coefficient of $|x|^{\lambda+\alpha}$ may be evaluated in terms of Γ-functions, to produce (2.11). To see that $R(x, \lambda, \alpha)$ is smooth in x, we introduce $v = (1/u)$ as variable of integration in (2.14); thus

$$R(x, \lambda, \alpha) = -\int_0^{1/a} \frac{(1+vx)_+^\lambda}{\Gamma(\lambda+1)} \frac{v^{-(\alpha+\lambda+1)}}{\Gamma(\alpha)} dv .$$

We may apply Taylor's theorem to $(1+vx)_+^\lambda$, obtaining a polynomial in vx, plus a remainder which vanishes rapidly for $v = 0$. Hence, the residues of R at its poles are powers of x, and the remainder is smooth in x.

Now we continue our representation (2.13) for $Re(\alpha) > 0$. Equation (2.9) shows that $J_+(x, \lambda)$ is regular for $-1 < \lambda < -1/2$ and $Re(\alpha) > 0$. On the other hand, the coefficients $a_\pm(x, \alpha)$ have simple poles for $\lambda + \alpha =$ integer. The residues at these poles are determined by the behavior at ∞ of the integrand in (2.13). Comparing (2.13) and (2.14), we see that the sum of the residues at the poles is zero.

Now we are ready to continue in λ, for fixed α, with $Re(\alpha) > 0$. First we assume that α is not an integer. From (2.10) and (2.11), it is apparent that the only possible singularities of the representation (2.10) are where $\lambda + \alpha$ is an integer. The case where $\lambda + \alpha$ is a nonnegative integer has already been discussed. If $\lambda + \alpha$ is a negative integer, then both $J(x, \lambda)$ and $R(x, \lambda, \alpha)$ are regular. It follows that the sum of the residues of

$$a_+ \frac{x_+^{\lambda+\alpha}}{\Gamma(x+\alpha+1)} \quad \text{and} \quad a_- \frac{x_-^{\lambda+\alpha}}{\Gamma(\lambda+\alpha+1)}$$

must be zero. This can be verified by a direct calculation, using Lemma 2.1.

If α is a positive integer, $\alpha = l$, we obtain

$$a_+(x, l) = (-1)^l , \qquad a_-(\lambda, l) = 0 .$$

In this case, R is regular in λ, because of the factor $1/\Gamma(\lambda+1)$.

The fact that $(x+u)_+^\lambda + (x+u)_-^\lambda = |x+u|^\lambda$ immediately implies

THEOREM 2.3. *If*

$$I(x, \lambda) = \gamma(\lambda) \int_0^a |x+u|^\lambda u^{\alpha-1} du , \tag{2.15}$$

we may write

$$I(x, \lambda) = b_+(\lambda, \alpha)\frac{x_+^{\lambda+\alpha}}{\Gamma(\lambda + \alpha + 1)} + b_-(\lambda, \alpha)\frac{x_-^{\lambda+\alpha}}{\Gamma(\lambda + \alpha + 1)} + R(x, \lambda, \alpha) . \tag{2.16}$$

Here

$$b_+(\lambda, \alpha) = \Gamma(\alpha)\gamma(\lambda)\Gamma(\lambda + 1)\frac{\sin \lambda\pi}{\sin \pi(\lambda + \alpha)} \tag{2.17}$$

$$b_-(\lambda, \alpha) = \Gamma(\alpha)\gamma(\lambda)\Gamma(\lambda + 1)\left[1 - \frac{\sin \pi\alpha}{\sin \pi(\lambda + \alpha)}\right], \tag{2.18}$$

and $R(x, \lambda, \alpha)$ is a smooth function of x, with poles if $\lambda + \alpha$ is a nonnegative integer.

REMARK. Equation (2.16) may be differentiated with respect to α, to obtain results for

$$\gamma(\lambda)\int_0^a | x + u |^\lambda u^{\alpha-1} \log u du .$$

We omit the calculation.

It may be useful to give our results for the leading, or most singular term in the expansion of (1) an explicit form. In this term, only the values of $(\partial f/\partial x_1)(0, 0) = b$, $(1/m!)(\partial^m f/\partial u^m)(0, 0) = c$, and $g(0, 0)$ enter. Taking the most singular term only,

$$F_1(x_1) \sim \gamma(\lambda)\int_{-a}^a | \pm bx + | c | u^m |^\lambda dug(0, 0) .$$

Setting $v = | c |^{1/m} u$, and $z = b \operatorname{sgn}(c)x$,

$$F_1(x_1) \sim \gamma(\lambda)\int_{-a_1}^{a_1} | z + v^m |^\lambda dv \frac{g(0, 0)}{| c | \frac{1}{m}} .$$

If m is even,

$$F_1(x_1) \sim 2\gamma(\lambda)\int_0^{a_1} | z + v^m |^\lambda dv \frac{g(0, 0)}{| c | \frac{1}{m}} ,$$

and if m is odd,

$$F_1(x_1) \sim \gamma(\lambda)\int_0^a (| z + v^m |^\lambda + | -z + v^m |^\lambda)dv \frac{g(0, 0)}{| c | \frac{1}{m}} .$$

These integrals may be evaluated by means of Theorem 2.3.

3. Reduction of double integrals to a standard form. We shall consider the integral (1), in the case $n = 2$. As before, the singularity of F near a given point x_0 is associated with points u_0 such that $f(x_0, u_0) = 0$ and $(\partial f/\partial u)(x_0, u_0) = 0$. Such points u_0 may be isolated, or may lie on a curve. In order to evaluate the contribution from a neighborhood of such a curve, we would have to cover it by a system of sufficiently small neighborhoods, taking particular notice of singular points of the curve, and then apply the theory of this section.

Without loss of generality, we may assume that $x_0 = 0$ and $u_0 = 0$. We set $f_0(u) = f(0, u)$. Our method consists in dividing the u-plane into regions, in such a way that distinct roots of f_0 appear in different regions. After a change of variables, f_0 may be represented as the product of a monomial and a nonvanishing function, in each region. The shapes of the regions involved are determined by the Puiseux expansions of the roots of f_0. We obtained the required regions by an iterative process. If f_0 is analytic, then the process will terminate. In fact, if distinct roots of f_0 have distinct Puiseux expansions, then the process will terminate if $f_0 \in C^\infty$. Since the process involves only a finite number of derivatives of f_0, it will terminate if f_0 has enough derivatives so that distinct roots have distinct truncated Puiseux expansions.

The integral over a single region assumes the form

$$\gamma(\lambda)\iint| x + u^\alpha v^\beta |^\lambda u^{\gamma-1}v^{\delta-1}g(u, v; \lambda, x)dudv \ . \tag{3.1}$$

Integrals of this form will be treated in § 4. Finally, (Lemma 3.1) we show that if f_0 has order m at the origin, then $\min(\gamma/\alpha, \delta/\beta) \geqq 1/m$.

As before, we assume that $(\partial f/\partial x_1)(0, 0) \neq 0$, we set $x_2 = x_3 = \cdots x_p = 0$, and we write $x_1 = x$. Then we may write

$$f(x, u) = f_0(u) + xe_1(x, u) = e_1(x, u)(x + f_0(u)E(x, u)) \ .$$

Functions denoted by e_i or E_i are different from zero at the origin. We first consider the simplest case, where the roots of f_0 have distinct tangents at the origin. We write $u_1 = u$, $u_2 = v$. Then $f_0(u_1, u_2) = P_m(u, v) + Q(u, v)$, where P_m is a homogeneous polynomial of degree m, and Q is of order $m + 1$ at the origin. By our assumption, the real roots of P_m are distinct. We introduce a partition of unity on the circle, symmetric about the origin, such that each function of the partition has its support in a region where either $P_m(\cos\theta, \sin\theta) \neq 0$, or $(\partial/\partial\theta)(P_m(\cos\theta, \sin\theta)) \neq 0$. Regions of the first type give rise to an integral of the form

$$\gamma(\lambda)\iint| x + r^m E_1(x, r, \theta) |^\lambda | e_1 |^\lambda g_1 r d\theta dr \ . \tag{3.2}$$

In regions of the second type, we may introduce $V = P_m(\theta) + rQ(r, \theta)$ as a variable of integration; we obtain an integral of the form

$$(3.3) \qquad \gamma(\lambda)\iint |x + r^m VE_1|^\lambda g_1(x, r, V, \lambda) r\,dV dr ,$$

if r is sufficiently small in the support of g_1.

Now we consider the general case, where P_m may have multiple roots. We shall obtain integrals similar to (3.2) and (3.3), which may be reduced to the form (3.1). By the term "sector" we shall mean a region generated by rotating a line about the origin. Thus a sector will consist of two wedge-shaped regions. By a "strip" we shall mean a region generated by displacement of a line parallel to the u-axis. By a "quadratic transformation" we shall mean a transformation of the form $u = u_1$, $v = u_1 v_1$. Under a quadratic transformation, a sector in the u, v plane which does not contain the v-axis is transformed into a strip in the u_1, v_1 plane. We shall be integrating over strips and sectors, and we would like to decompose an integral over a strip into a sum of integrals over sectors. We accomplish this by formally extending all integrations over the whole plane. First, we assume that the integrand in (1) has support in a finite disc about the origin. Given any open, finite covering of the unit circle, we can find a C^∞ partition of unity, such that each function $\varphi_j(\theta)$ has its support in one of the covering sets. The functions $\varphi_j(2\theta)$ provide a partition of unity which is constant on lines through the origin, and such that each function of the partition has its support in a sector. After rotation and application of a quadratic transformation, each of the functions φ_j will have support in a strip. Thus, after quadratic transformation, our original integral is transformed into a sum of integrals over strips. Integration over each strip may formally be extended over the whole plane, which in turn may be decomposed into sectors by a partition of unity. This process may be repeated as often as desired. In this way the burden of the complexities of the actual region of integration is thrown on the structure of the final partition of unity.

We cover each of the real roots of $P_m(u, v)$ by a sufficiently small open sector, and choose a covering of the remaining sectors which is finite and does not intersect the roots of P_m. We choose a partition of unity subordinate to this covering. Integrals over sectors which do not contain a root of P_m, or which contain a simple root of P_m, may be treated as before, leading to integrals of the form (3.2) or (3.3). A sector which contains a multiple root of P_m may be rotated so that the root coincides with the new u-axis. Under such a transformation, an expression of the form $u^\alpha v^\beta E(u, v)$, where $E(0, 0) \neq 0$, is transformed into a similar expression. Such expressions remain of the same type under a quadratic expression as well. Hence, after a rotation and a quadratic transformation, we have

$$E(u, v, x)f_0(u, v) = E(P_m + Q) = u_1^m E_1[P_{m_1}(u_1, v_1) + Q_1] .$$

Here we have divided $P_m + Q$ by u_1^m and collected terms of lowest degree in u_1 and v_1 to obtain $P_{m_1}(u_1, v_1)$. Observe that m_1 is less than or equal to the multiplicity of the root of P_m in question.

Now we apply a similar procedure to P_{m_1} instead of P_m. A second application of the procedure may result in an expression of the form

$$Ef_0 = u_2^\alpha v_2^\beta E_2[P_{m_2} + Q] ,$$

if v_1 divides P_{m_1}, but further applications of the procedure do not result in expressions of more complicated form. We temporarily halt our procedure if Ef_0 assumes the form

$$Ef_0 = u_l^\alpha v_l^\beta E_l(v_l + au_l + \cdots)^\gamma , \tag{3.4}$$

where $a \neq 0$. This situation will always occur if distinct roots of f_0 have distinct (truncated) Puiseux expansions. In particular, if f_0 is analytic, we may apply the Weierstrass preparation theorem to f_0 (see G. A. Bliss [1], pp. 53–55.) Thus after rotation,

$$f_0(u, v) = E(u, v)[u^m + a_1(v)u^{m-1} + \cdots + a_m(v)] ,$$

where $E(u, v)$ and $a_j(v)(j = 1, \cdots, m)$ are analytic and $E(0, 0) \neq 0$. Since the field of fractional power series is algebraically closed, the roots of f_0 may be expanded in Puiseux series (see R. Walker [12], pp. 97–102.) Thus after a finite number of quadratic transformations, the distinct roots of f_0 must belong to distinct sectors, and f_0 will appear as a product of powers of factors whose lowest term is of degree one, multiplied by a nonvanishing function, as shown in (3.4). Hence, after a final rotation and quadratic transformation, we are led to integrals of the form

$$\gamma(\lambda)\int | x + u^\alpha v^\beta E(u, v, x) |^\lambda u^{\gamma-1}v^{\delta-1}g(u, v, x, \lambda)dudv ,$$

where $\gamma \geqq 1$, $\delta \geqq 1$. The factor $u^{\gamma-1}v^{\delta-1}$ arises from the Jacobians of the quadratic transformations. Now, using the implicit function theorem, we set $v_1 = v\,| E(u, v, x) |^{1/\beta}$, for u, v, x sufficiently small and obtain an integral of the form

$$\gamma(\lambda)\iint | \pm x + u^\alpha v_1^\beta |^\lambda u^{\gamma-1}v_1^{\delta-1}g(u, v_1, x, \lambda)dudv_1 . \tag{3.5}$$

Thus after appropriate changes of variables, and a partition of unity, the evaluation of (1) is reduced to evaluation of integrals of the form (3.5).

We shall see in the next section that the leading singularity of (3.5) is determined by $\mu = \min(\gamma/\alpha, \delta/\beta)$.

LEMMA 3.1. *If f_0 has order m at the origin, and $g(u, v, s, \lambda)$ has the form $u^{\gamma-1}v^{\delta-1}g_1(u, v, x, \lambda)$, then in all integrals of the form (3.5) which arise by the preceding process, we have $\mu \geqq \min(\gamma/m, \delta/m)$.*

Proof. All integrals which arise are of the form

$$I_n = \gamma(\lambda)\iint |x + u^\alpha v^\beta E(P_l + Q)|^\lambda u^{\gamma-1}v^{\delta-1}g(u, v, x, \lambda)dudv .$$

For such an integral, we define $\mu_n = \min[\gamma/(\alpha + l), \delta/(\beta + l)]$. We show that μ is a nondecreasing function under rotations, quadratic transformations, and (clearly) if a monomial is factored out of $P_l + Q$. The only nontrivial case is a quadratic transformation. Under quadratic transformation,

$$I_{n+1} = \gamma(\lambda)\iint |x + u_1^{\alpha+\beta+l}v_1^\beta E(P_{l_1} + Q_1)|^\lambda u_1^{\gamma+\delta-1}v_1^{\delta-1}g(u_1, u_1v_1, x, \lambda)du_1dv_1 .$$

Hence, $\mu_{n+1} = \min[(\gamma + \delta)/(\alpha + \beta + l + l_1), \delta/(\beta + l_1)]$. Since $l_1 \leqq l$, we have $\delta/(\beta + l_1) \geqq \mu_n$, hence also $(\gamma + \delta)/(\alpha + \beta + l + l_1) \geqq \mu_n$.

4. Expansion of double integrals. In this section, we shall expand double integrals of the form

$$I(x) = \gamma(\lambda)\iint |x + u^\alpha v^\beta|^\lambda u^{\gamma-1}v^{\delta-1}g(u, v; \lambda, x)dudv , \tag{4.1}$$

in powers of x, using the results of § 2. First we prove Theorem 4.1, which asserts that $I(x)$ is continuous in x if $\gamma + \alpha Re(\lambda) > 0$ and $\delta + \beta Re(\lambda) > 0$. Thus if $g(u, v; \lambda, x)$ is written as a sum of functions, with remainder multiplied by a large power of both u and v, then the remainder will give rise to a continuous function of x. The major portion of this section is devoted to expansion of integrals of the form

$$J(x) = \gamma(\lambda)\iint_{\substack{u\geqq 0\\ v\geqq 0}} |x + u^\alpha v^\beta|^\lambda u^{\gamma-1}v^{\delta-1}g(v; \lambda, x)\varphi(u)dudv , \tag{4.2}$$

where $g(v)$ and $\varphi(u)$ have compact support, and $\varphi(\mu) \equiv 1$ for small u. The results are summarized as Lemma 4.2. An appropriate expansion of $g(u, v; \lambda, x)$, together with Lemma 4.2 then implies an expansion of (4.1) in powers of x, specified in Theorem 4.2. Finally, we give a more or less explicit formula for the coefficient of the leading or most singular term in the expansion of (4.1).

THEOREM 4.1. *If $\alpha, \beta \geqq 0$, $\gamma, \delta > 0$ and if $\gamma + \alpha Re(\lambda) > 0$ and*

$\delta + \beta Re(\lambda) > 0$, *and if* $g(u, v; \lambda)$ *has enough derivatives with respect to* u *and* v, *and has compact support in* u *and* v, *then the integral* (4.1) *is bounded and continuous in* x, *for small* x.

Proof. Let $\xi = |x|^{1/\alpha}$. We write

$$I(x) = \gamma(\lambda)\iint_{|u|\leqq k\xi} |x + u^\alpha v^\beta|^\lambda u^{\gamma-1}v^{\delta-1}gdudv$$
$$+ \gamma(\lambda)\iint_{u>k\xi} |x + u^\alpha v^\beta|^\lambda u^{\gamma-1}v^{\delta-1}gdudv ,$$
$$I(x) = I_1 + I_2 .$$

We shall specify k presently. In the first integral, we set $u = \xi\mu$. Then

$$I_1 = \gamma(\lambda)\iint_{|\mu|\leqq k} \xi^{\alpha\lambda+\gamma} |\pm 1 + \mu^\alpha v^\beta|^\lambda \mu^{\gamma-1}v^{\delta-1}gd\mu dv .$$

Now we choose k so small that $\pm 1 + \mu^\alpha v^\beta$ does not vanish for $|\mu| \leqq k$, if v is in the support of g. As in the proof of Theorem 2.1, it follows that I_1 is continuous in x.

In the second integral, we divide by $|u|^\alpha$; thus

$$I_2 = \gamma(\lambda)\iint_{|u|>k\xi} \left|\pm\frac{x}{|u|^\alpha} + v^\beta\right|^\lambda v^{\delta-1}gdv(\pm 1)\,|u|^{\alpha\lambda+\gamma-1}du .$$

Now we may apply Theorem 2.1 to the inner integral taken over v, since $x|u|^{-\alpha}$ is bounded. Since $\beta Re(\lambda) + \delta > 0$, the inner integral is continuous in x and u for u bounded away from zero, and bounded for u in the region of integration. Hence, since $\alpha Re(\lambda) + \gamma > 0$, the double integral is continuous in x.

We proceed to the statement and proof of Lemma 4.1. Starting with (4.2), we set $\mu = u^\alpha$, $\nu = v^\beta$, $p = \gamma/\alpha$, $q = \delta/\beta$, $r = 1/\beta$, $\varphi_1(\mu) = \varphi(\mu^{1/\alpha})$. Thus

$$J(x) = \gamma(\lambda)\iint_{\substack{\mu\geqq 0\\ \nu\geqq 0}} |x + \mu\nu|^\lambda \mu^{p-1}\nu^{q-1}g(\nu^r; x, \lambda)\varphi_1(\mu)\frac{d\mu d\nu}{\alpha\beta} .$$

We recall that $\varphi_1(\mu) \equiv 1$ for small μ. Introducing $w = \mu\nu$ as a new variable of integration, we have

$$J(x) = \gamma(\lambda)\int_0^\infty |x + w|^\lambda k(w; x, \lambda)dw , \tag{4.3}$$

where

$$k(w; x, \lambda) = \int_0^\infty g\left(\frac{w^r}{\mu^r}; x, \lambda\right)\varphi_1(\mu)w^{q-1}\mu^{p-q-1}\frac{d\mu}{\alpha\beta} . \tag{4.4}$$

For $w \neq 0$, the integral exists and is smooth, since g and φ_1 have compact support. For the same reason, k has compact support in w. It follows that the singularities of $J(x)$ for small x are determined by the behavior of $k(w; x, \lambda)$ for small w. This is precisely the statement that the singularities of (4.2) are associated with the u and v axes. Since x and λ play the role of parameters in the following, we shall usually not indicate their presence.

LEMMA 4.1. *For small w, the function $k(w)$, defined by* (4.4), *with $p, q, r > 0$, may be represented in the form*

$$k(w) = a_0 w^{p-1} + a_0^1 w^{p-1} \log w + \sum_{l=0}^{L} b_l w^{q+lr-1} + w^{q+(L+1)r-1} R(w^r) \,, \tag{4.5}$$

for any L, if g has enough derivatives. The coefficients a_0, a_0^1, b_l depend on x and λ, and are given by certain of the formulas (4.6–4.23). *The coefficient a_0^1 vanishes unless $p = q + Jr$, for some integer $J \geqq 0$. The remainder $R(w^r)$ is smooth for small values of its argument.*

Proof. We distinguish three cases:

(A) $q > p$,

(B) $q < p$, $p \neq q + Jr$ for any integer J, and

(C) $q = p + Jr$, $J =$ integer $\geqq 0$.

A. If $q > p$, we define

$$k_0(w) = w^{q-1} \int_0^\infty g\left(\frac{w^r}{\mu^r}\right) \mu^{p-q-1} \frac{d\mu}{\alpha\beta} \,.$$

If $w \neq 0$, this integral exists, since g has compact support. Making a change of variables,

$$k_0(w) = w^{p-1} \int_0^\infty g(\nu^r) \nu^{q-p-1} \frac{d\nu}{\alpha\beta} \,.$$

Hence, we have $k_0(w) = w^{p-1} a_0(x, \lambda)$, with

$$a_0(x, \lambda) = \int_0^\infty g(\nu^r; x, \lambda) \nu^{q-p-1} \frac{d\nu}{\alpha\beta} \,. \tag{4.6}$$

Now we may write

$$k(w) = a_0 w^{p-1} + w^{q-1} \int_0^\infty g\left(\frac{w^r}{\mu^r}\right) [\varphi_1(\mu) - 1] \mu^{p-q-1} \frac{d\mu}{\alpha\beta} \,. \tag{4.7}$$

We observe that $\varphi_1(\mu) - 1$ vanishes for small μ. Hence, g may be expanded in powers of w^r/μ^r, leading to an expansion of $k(w)$ in powers of w^r. We have

(4.8) $$b_l(x, \lambda) = \frac{1}{l!}\left(\frac{\partial}{\partial v}\right)^l g(v; x, \lambda)\Big|_{v=0} \int_0^\infty [\varphi_1(\mu) - 1]\mu^{p-q-lr-1} \frac{d\mu}{\alpha\beta} .$$

B. If $q < p$, $q + (J-1)r < p < q + Jr$, for some positive integer J, we write

$$g(v)\varphi(\mu) = \left(\sum_0^{J-1} g_l v^l\right)\varphi(u) + \left(g(v) - \sum_0^{J-1} g_l v^l\right) + \left(g(v) - \sum_0^{J-1} g_l v^l\right)(\varphi(u) - 1) ;$$

here

$$g_l(x, \lambda) = \frac{1}{l!}\left(\frac{\partial}{\partial v}\right)^l g(v; x, \lambda)\Big|_{v=0} .$$

Thus $k(w) = k_1(w) + k_2(w) + k_3(w)$, with

(4.9) $$k_1(w) = \int_0^\infty \sum_{l=0}^{J-1} g_l \frac{w^{lr}}{\mu^{lr}} \varphi_1(\mu) w^{q-1}\mu^{p-q-1} \frac{d\mu}{\alpha\beta} ,$$

(4.10) $$k_2(w) = \int_0^\infty \left[g\left(\frac{w^r}{\mu^r}\right) - \Sigma g_l \frac{w^{lr}}{\mu^{lr}}\right] w^{q-1}\mu^{p-q-1} \frac{d\mu}{\alpha\beta} ,$$

(4.11) $$k_3(w) = \int_0^\infty \left[g\left(\frac{w^r}{\mu^r}\right) - \Sigma g_l \frac{w^{lr}}{\mu^{lr}}\right] [\varphi_1(\mu) - 1]\mu^{p-q-1} \frac{d\mu}{\alpha\beta} .$$

The integral (4.9) exists, since for $0 \leqq l \leqq J - 1$, $p - lr - q > 0$, and φ_1 has compact support. In fact, we have

$$k_1(w) = \sum_{l-0}^{J-1} b_l w^{q+lr-1} ,$$

with

(4.12) $$b_l(x, \lambda) = g_l(x, \lambda) \int_0^\infty \varphi_1(\mu)\mu^{p-lr-q-1} \frac{d\mu}{\alpha\beta} , \qquad (0 \leqq l \leqq J - 1) .$$

The integral (4.10) exists for $w \neq 0$, since, for small μ, $g(w^r/\mu^r)$ vanishes, and $p - lr - q > 0$. For large μ, the quantity inside the brackets may be written as $w^J \mu^{-J} h(w/\mu)$, where h is a smooth function. Hence, k_2 is integrable at ∞. A change of variables shows that

$$k_2(w) = w^{p-1}\int_0^\infty [g(\nu^r; x, \lambda) - \Sigma g_l(x, \lambda)\nu^{lr}]\nu^{q-p-1} \frac{d\nu}{\alpha\beta} ;$$

thus

(4.13) $$a_0(x, \lambda) = \int_0^\infty [g(\nu^r; x, \lambda) - \Sigma g_l(x, \lambda)\nu^{lr}]\nu^{q-p-1} \frac{d\nu}{\alpha\beta} .$$

Finally, we observe that (4.11) may be written in the form

$$(4.14)\qquad k_3(w) = w^{q+Jr-1}\int_0^\infty h\left(\frac{w^r}{\mu^r}\right)[\varphi_1(\mu) - 1]\mu^{p-Jr-q-1}\frac{d\mu}{\alpha\beta}\,.$$

This integral may be treated in the same manner as (4.7).

C. If $p = q + Jr$, J is a nonnegative integer. This case is similar to the preceding one. We shall use the Heaviside function

$$H(1-v) = \begin{cases} 0 & \text{if } v \geqq 1 \\ 1 & \text{if } v < 1\,. \end{cases}$$

We may write

$$\begin{aligned}\varphi(u)g(v) = \sum_{l=0}^{J-1}(g_l v^l)\varphi(u) &+ \left[g(v) - \sum_{l=0}^{J-1} g_l v^l - H(1-v)g_J v^J\right] \\ &+ \left[g(v) - \sum_{l=0}^{J-1} g_l v^l - H(1-v)g_J v^J\right][\varphi(u) - 1] \\ &+ H(1-v)g_J v^J H(1-u) + H(1-v)g_J v^J[\varphi(u) - H(1-u)]\,.\end{aligned}$$

Thus $k(w) = \sum_{j=1}^5 k_j(w)$, with

$$(4.15)\qquad k_1(w) = \int_0^\infty\left(\sum_l g_l w^{lr}\mu^{-lr}\right)\varphi_1(\mu)w^{q-1}\mu^{p-q-1}\frac{d\mu}{\alpha\beta}\,,$$

$$(4.16)\qquad \begin{aligned}k_2(w) = \int_0^\infty &[g(w^r\mu^{-r}) - \Sigma g_l w^{lr}\mu^{-lr} \\ &- H(1 - w\mu^{-1})g_J w^{Jr}\mu^{-Jr}]w^{q-1}\mu^{p-q-1}\frac{d\mu}{\alpha\beta}\,,\end{aligned}$$

$$(4.17)\qquad \begin{aligned}k_3(w) = \int_0^\infty &[g(w^r\mu^{-r}) - \Sigma g_l w^{lr}\mu^{-lr} \\ &- H(1 - w\mu^{-1})g_J w^{Jr}\mu^{-Jr}](\varphi_1(\mu) - 1)w^{q-1}\mu^{p-q-1}\frac{d\mu}{\alpha\beta}\,,\end{aligned}$$

$$(4.18)\qquad k_4(w) = \int_0^\infty H\left(1 - \frac{w}{\mu}\right)g_J w^{Jr}\mu^{-Jr}H(1-\mu)w^{q-1}\mu^{p-q-1}\frac{d\mu}{\alpha\beta}\,,$$

$$(4.19)\qquad \begin{aligned}k_5(w) = \int_0^\infty &H(1 - w\mu^{-1})g_J w^{Jr}\mu^{-Jr} \\ &\times[\varphi_1(\mu) - H(1-\mu)]w^{q-1}\mu^{p-q-1}\frac{d\mu}{\alpha\beta}\,.\end{aligned}$$

The integral (4.15) is identical with (4.9); thus the coefficients $b_l(l = 0, \cdots, J-1)$ are given by (4.12). The integral (4.16) is similar to (4.10). By analogous reasoning, we conclude that $k_2(w) = w^{p-1}c_0(x, \lambda)$, with

$$(4.20)\quad c_0(x,\lambda) = \int_0^\infty [g(\nu^r) - \Sigma g_l \nu^{lr\;-Jr-} - H(1-\nu)g_J\nu^{Jr}]\nu^{-lr-^1}\frac{d\mu}{\alpha\beta}\,;$$

the coefficient $a_0(x, \lambda)$ will also involve a contribution from $k_5(w)$. In the integral (4.17), we observe that the integrand vanishes for small μ. Hence, for small w, $H(1 - w\mu^{-1}) \equiv 1$, in the region of integration. Thus, if $v^{J+1}h(v) = g(v) - \sum_0^J g_l v^l$, we may write

$$k_3(w) = w^{p+r-1}\int_0^\infty h(w^r\mu^{-r})[\varphi_1(\mu) - 1]\mu^{-r-1}\frac{d\mu}{\alpha\beta}, \tag{4.21}$$

for small w. Thus $k_3(w)$ may be expanded in powers of w^r, in the same manner as (4.7).

The integral (3.22) exists for $w \neq 0$, since integration may be taken over a finite segment excluding the origin. After a change of variables,

$$k_4(w) = -\frac{g_J}{\alpha\beta} w^{p-1}\log w .$$

Hence

$$a_0^1(x, \lambda) = -\frac{g_J}{\alpha\beta}(x, \lambda) . \tag{4.22}$$

Finally, for small w,

$$k_5(w) = w^{p-1}g_J\int_0^\infty [\varphi_1(\mu) - H(1 - \mu)]\mu^{-1}\frac{d\mu}{\alpha\beta} .$$

Combining this result with (4.20), we have

$$a_0(x, \lambda) = c_0(x, \lambda) + g_J(x, \lambda)\int_0^\infty \frac{\varphi_1(\mu) - H(1 - \mu)}{\mu}\frac{d\mu}{\alpha\beta} . \tag{4.23}$$

This completes the proof of Lemma 4.1.

Now we may apply Theorem 2.2 to the integral (4.3). Lemma 4.1 immediately implies

LEMMA 4.2. *$J(x)$, given by (4.3), has an expansion in distributions homogeneous of degrees $\lambda + p$, $\lambda + q + lr (0 \leqq l \leqq L)$, possibly including a term of the form $a_0^1 c_\pm x_\pm^{\lambda+p} \log|x|$.*

It follows that there is a similar expansion of $I(x)$, given by (4.1), provided that $g(u, v; x, \lambda)$ can be represented as a sum of terms of the form $g(v; x, \lambda)\varphi(u)$, plus a remainder multiplied by large powers of both u and v. We define the second difference quotient

$$\begin{aligned} g_{12}(u, v; x, \lambda) &= \int_0^1\int_0^1 \frac{\partial^2 g}{\partial u\partial v}(us, vt; x, \lambda)dsdt \\ &= \frac{1}{uv}[g(u, v) - g(u, o) - g(o, v) + g(o, o)] . \end{aligned}$$

Hence

$$g(u, v) = g(u, o) - g(o, v) - g(o, o) + uvg_{12}(u, v). \tag{4.24}$$

Clearly, g_{12} is smooth if g is smooth. Unfortunately, the terms on the right hand side of (4.24) do not have compact support in u and v. Although this difficulty could be circumvented by a systematic use of finite-part integrals, we prefer to work with functions with compact support.

Let φ be a C^∞ function with compact support, which is even, and such that $\varphi \equiv 1$ in a neighborhood of the origin. We define $h(u, v)$ by the equation

$$\begin{aligned} g(u, v) = g(u, o)\,\varphi(v) + g(o, v)\,\varphi(u) - g(o, o)\,\varphi(u)\,\varphi(v) \\ + uvh(u, s). \end{aligned} \tag{4.25}$$

Using (4.24), we may write

$$\begin{aligned} h(u, v) = \left(\frac{g(u, o) - g(o, o)}{u}\right)\left(\frac{1 - \varphi(v)}{v}\right) + \left(\frac{g(o, v) - g(o, o)}{v}\right)\left(\frac{1-\varphi(u)}{u}\right) \\ + g(o, o)\left(\frac{\varphi(u) - 1}{u}\right)\left(\frac{\varphi(v)-1}{v}\right) + g_{12}(u, v); \end{aligned}$$

hence h is a smooth function. We may apply the same process to $h(u, v)$, and thus obtain a remainder for g with the factor u^2v^2. The process will terminate only if g ceases to have the required derivatives.

We conclude that, after breaking the region of integration into quadrants, $I(x)$ may be represented as a sum of integrals of the form (4.2), plus a smooth remainder. Thus we have

THEOREM 4.2. *$I(x)$, given by* (4.1), *has an expansion in distributions homogeneous of degrees*

$$\lambda + \frac{\gamma + m}{\alpha}\,(0 \leqq m \leqq M),\ \lambda + \frac{\delta + l}{\beta}\,(0 \leqq l \leqq L)\ ,$$

plus terms of the form $d_\pm x_\pm^\sigma \log(x)$, in case $(\gamma + m)/\alpha = (\delta + l)/\beta = \sigma$ for certain l and m. The remainder has order greater than $\min[(\gamma + M)/\alpha), (\delta + L)/\beta)]$.

Now we shall compute the most singular term in the expansion of $I(x)$. We break the region of integration into quadrants, and evaluate the contribution from a single quadrant. The complete result would depend on the parity of $\alpha, \beta, \gamma, \delta$. As before, we write $p = \gamma/\alpha$, $q = \delta/\beta$. Observe that a lower bound on p and q is given by Lemma 3.1.

A. If $p < q$, we write, from (4.25),

$$g(u, v) = g(o, v)\,\varphi(u) + [g(u, o) - g(o, o)\,\varphi(u)]\,\varphi(v) + uvh(u, v).$$

Since $g(u, o) - g(o, o)\,\varphi(u)$ is smooth and vanishes for $u = o$, the leading term arises from $g(o, v)\,\varphi(v)$. From Lemma 4.1, we obtain

$$\begin{aligned} I_{++}(x) &= \gamma(\lambda) \iint_{\substack{u>0\\ v>0}} |\, x + u^{\alpha}v^{\beta}\,|^{\lambda} u^{\gamma-1} v^{\delta-1}\, g(u, v; x, \lambda)\,du\,dv \\ &= \gamma(\lambda) \int_0^{\infty} |\, x + w\,|^{\lambda}\,[a_0 w^{p-1} + o(w^{p-1})]\,dw; \end{aligned}$$

from (4.6) we have

$$a_0(x, \lambda) = \int_0^{\infty} g(o, v; x, \lambda)\, v^{\delta-\beta\gamma/\alpha}\, \frac{dv}{\alpha}.$$

The leading term of $I_{++}(x)$ is given by Theorem 2.3.

B. If $q < p$, we have, similarly,

$$I_{++}(x) = \gamma(\lambda) \int_0^{\infty} |\, x + w\,|^{\lambda}[b_0 w^{q-1}) + o(w^{q-1})]\,dw,$$

with

$$b_0(x, \lambda) = \int_0^{\infty} g(u, o; x, \lambda)\, u^{\gamma-\alpha\delta/\beta}\, \frac{du}{\beta}.$$

C. If $p = q$, we write

$$g(u, v) = g(u, o)\,\varphi(v) + g(o, v)\,\varphi(u) - g(o, o)\,\varphi(u)\,\varphi(v) + uvh(u, v).$$

Applying Lemma 4.1 to each of the first three terms, we obtain

$$I_{++}(x) = \gamma(\lambda) \int_0^{\infty} |\, x + w\,|^{\lambda}\,[a_0 w^{p-1} + a_0^1 w^{p-1} \log |\, w\,| + o(w^{p-1})]\,dw,$$

with

$$\begin{aligned} a_0(x, \lambda) &= \int_0^{\infty} [g(o, v) - H(1 - v)\, g(o, o)]v^{-1}\, \frac{dv}{\alpha} \\ &\quad + \int_0^{\infty} [g(u, o) - H(1 - u)\, g(o, o)]u^{-1}\, \frac{du}{\beta}, \end{aligned}$$

and

$$a_0^1(x, \lambda) = -\,\frac{g(o, o; x, \lambda)}{\alpha\beta}.$$

We remark that the preceding integrals are the finite parts of the integrals

$$\int_0^\infty g(o, v) \frac{dv}{\alpha v}, \text{ and } \int_0^\infty g(u, o) \frac{du}{\beta u}.$$

5. **Integrals with nonvanishing Hessian.** We consider integrals of the form

$$F(x, \lambda) = \gamma(\lambda) \int_U |f(x, u)|^\lambda g(x, u) du, \tag{5.1}$$

where $x \in X \subset R^p$, $u \in U \subset R^n$, and g has support in the interior of the bounded set U. We assume that the Hessian matrix $[\partial^2 f/(\partial u_i \partial u_k)]$ is nonsingular for all $x \in X$ and $u \in U$. In this case, a rather simple description of the singularity of F can be given, using only the results of § 2. Our method consists in a change of variables of integration, which enables us to write $f(x, u) = \tilde{f}(x) \pm U_1^2 \pm \cdots U_n^2$. An application of Theorems 2.2 and 2.3 then shows that F can be expressed in terms of $\tilde{f}_\pm^{\lambda+n/2}$. Similar results have been obtained by a number of authors, for example J. Leray [9], L. Gårding [6], G. F. D. Duff [3], D. Ludwig [10].

Theorem 1.2 implies that the singularities of F are associated with points x_0, u_0 where both $f(x_0, u_0) = 0$ and $[(\partial f/\partial u)(x_0, u_0)] = 0$. Thus we may analyse the singularity of F near x_0 by covering the associated point or points u_0 by a finite collection of sufficiently small neighborhoods and choosing a partition of unity. We shall assume that this has been done. The size of the neighborhoods will be determined from the following discussion.

Since the Hessian matrix is nonsingular, we may determine $u = u_0(X)$ from the equations $(\partial f/\partial u)(x, u) = 0$ in a neighborhood of x_0. We write $u = u_0(x) + v$, $f_1(x, v) = f(x, u_0(x) + v)$. We can perform a rotation in the v-space so that the matrix $[\partial^2 f_1/(\partial v_i \partial v_k)]$ is diagonal at $x = x_0$, $v = 0$. Now we determine $\tilde{v}_1(x, v_2, \cdots, v_n)$ from the equation $\partial f_1/\partial v_1 = 0$. Hence

$$f_1(x, v) = f_1(x, \tilde{v}_1, v_2, \cdots, v_n) + (v - \tilde{v}_1)^2 e_1(x, v),$$

where $e_1(x, v)$ does not vanish for x near x_0, if v is small. Applying this process to $v_2, \cdots, v_n$ in succession, we obtain

$$f_1(x, v) = f_1(x, 0) + \sum_{j=1}^n (v_j - \tilde{v}_j)^2 e_j(x, v),$$

for x near x_0, and for v sufficiently small. This type of result is known as Morse's lemma (see M. Morse [11].) We set

$$V_j = (v_j - \tilde{v}_j) | e_j(x, v) |^{1/2},$$

and

$$\tilde{f}(x) = f_1(x, 0) = f(x, u_0(x)).$$

Introducing V as variable of integration, we have F as a sum of integrals of the form

$$(5.2) \qquad I(x, \lambda) = \gamma(\lambda) \int |\tilde{f}(x) \pm V_1^2 \cdots \pm V_n^2|^\lambda g_1(x, V) dV.$$

We note that

$$g_1(x, 0) = g(x, u_0(x))\, 2^{n/2} \Delta^{-1/2},$$

where

$$\Delta = \left| \det \left(\frac{\partial^2 f_1}{\partial v_i \partial v_k} \right) \right|.$$

This integral could be handled by an application of Theorems 2.2 and 2.3 n times; we prefer to apply the theorems only twice. After rearrangement of indices, we may assume that

$$e_1(x, v) > 0, \cdots e_k(x, v) > 0,$$

$e_{k+1}(x, v) < 0, \cdots e_{k+l}(x, v) < 0$. Here $k + l = n$. We write

$$r_1^2 = V_1^2 + \cdots V_k^2;\; r_2^2 = V_{k+l}^2 + \cdots V_{k+l}^2\,.$$

Then

$$I(x, \lambda) = \gamma(\lambda) \iint \Big| \, |\tilde{f}(x) + r_1^2 - r_2^2|^\lambda$$
$$\times \iint g_1(x, r_1\omega_1, r_2\omega_2) d\omega_1 d\omega_2 r_1^{k-1} r_2^{l-1} dr_1 dr_2.$$

Here ω_1 and ω_2 represent the corresponding angular variables. Integrating first over these angular variables we obtain

$$I(x, \lambda) = \gamma(\lambda) \iint |\tilde{f}(x) + r_1^2 - r_2^2|^\lambda g_2(x, r_1^2, r_2^2) r_1^{k-1} r_2^{l-1} dr_1 dr_2.$$

We note that g_2 is regular in r_1^2 and r_2^2, and

$$g_2(x, 0, 0) = \frac{4(2\pi)^{n/2}}{\Gamma\left(\frac{k}{2}\right)\Gamma\left(\frac{l}{2}\right)} \Delta^{-1/2} g(x, u_0(x))\,.$$

Now we may expand g_2 in integral powers of r_1^2 and r_2^2; for fixed λ the remainder will be smooth in x if enough terms are taken. It therefore suffices to find the singularity of a single term of the form

$$(5.3)\qquad J(x, \lambda) = \gamma(\lambda) \iint_{\substack{s_1 \geqq 0 \\ s_2 \geqq 0}} | \tilde{f}(x) + s_1 - s_2 |^{\lambda} s_1^{k/2-1} s_2^{l/2-1} ds_1 ds_2 .$$

The leading term of $I(x, \lambda)$ will have precisely the form (5.3), multiplied by

$$\frac{(2\pi)^{n/2}}{\Gamma\left(\frac{k}{2}\right)\Gamma\left(\frac{l}{2}\right)} \Delta^{-1/2} g(x, u_0(x)) .$$

Now applying Theorem 2.3, we see that

$$J(x, \lambda) = \gamma(\lambda)\,\Gamma\left(\frac{l}{2}\right)\Gamma(\lambda + 1)\left[\frac{\sin \pi\left(\lambda + \frac{l}{2}\right) - \sin \pi \frac{l}{2}}{\sin \pi\left(\lambda + \frac{l}{2}\right)} I_+ + \frac{\sin \pi\lambda}{\sin \pi\left(\lambda + \frac{l}{2}\right)} I_-\right] + R(x, \lambda),$$

where

$$I_{\pm} = \frac{1}{\Gamma\left(\lambda + \frac{l}{2} + 1\right)} \int_{s_1 > 0} (\tilde{f}(x) + s_1)_{\pm}^{\lambda + l/2} s_1^{k/2-1} ds_1 ,$$

and $R(x, \lambda)$ is regular. Now applying Theorem 2.2 to $I_{\pm}$, we find that

$$\begin{aligned} J(x, \lambda) = {} & \Gamma\left(\frac{l}{2}\right)\Gamma\left(\frac{k}{2}\right)\gamma(\lambda)\Gamma(\lambda + 1) \\ & \times\left[\frac{\sin \pi\left(\lambda + \frac{l}{2}\right) - \sin \pi \frac{l}{2}}{\sin \pi\left(\lambda + \frac{n}{2}\right)} \frac{\tilde{f}_+^{\lambda + (n/2)}}{\Gamma\left(\lambda + \frac{n}{2} + 1\right)} \right. \\ & \left. + \frac{\sin \pi\left(\lambda + \frac{k}{2}\right) - \sin \pi \frac{k}{2}}{\sin \pi\left(\lambda + \frac{n}{2}\right)} \frac{\tilde{f}_-^{\lambda + (n/2)}}{\Gamma\left(\lambda + \frac{n}{2} + 1\right)}\right] + R_2(x, \lambda). \end{aligned}$$

Hence the leading term of $I(x, \lambda)$ is given by

$$(5.4)\qquad (2\pi)^{n/2} \Delta^{-1/2} g(x, u_0(x)) \times\left[d_+ \frac{\tilde{f}_+^{\lambda + n/2}}{\Gamma\left(\lambda + \frac{n}{2} + 1\right)} + d_- \frac{\tilde{f}_-^{\lambda + n/2}}{\Gamma\left(\lambda + \frac{n}{2} + 1\right)}\right],$$

with

$$(5.5) \qquad d_+ = \gamma(\lambda)\,\Gamma(\lambda+1)\left[\frac{\sin\pi\left(\lambda+\frac{l}{2}\right)-\sin\pi\frac{l}{2}}{\sin\pi\left(\lambda+\frac{n}{2}\right)}\right],$$

$$(5.6) \qquad d_- = \gamma(\lambda)\,\Gamma(\lambda+1)\left[\frac{\sin\pi\left(\lambda+\frac{k}{2}\right)-\sin\pi\frac{k}{2}}{\sin\pi\left(\lambda+\frac{n}{2}\right)}\right].$$

The coefficients $d_\pm$ have simple poles as functions of λ according to the following scheme:

If k and l are both even, there are poles if λ is of the form $-2q$, q integer $\geqq 0$.

If h and l are both odd, there are poles if $\lambda = -2q-1$, q integer $\geqq 0$.

If $k+l$ is odd, there are poles if $\lambda = q + 1/2$, q any integer.

Since $I(x, \lambda)$ is regular for all λ, of course the sum of the residues at these poles is zero.

Bibliography

1. G. A. Bliss, *Fundamental existence theorems*, New York, 1913.
2. R. Courant, *Methods of mathematical physics*, II, New York, 1962.
3. G. F. D. Duff, *On the Riemann matrix of a hyperbolic system*, MRC Tech. Rep. #246, U. of Wisconsin, 1961.
4. A. Erdélyi, *Asymptotic Expansions*, New York, 1956.
5. J. Focke, *Asymptotische Entwicklungen mittels der methode der stationären phase*; Berichte Säch. Akad. der Wiss. Leipzig, V. 101, Heft 3, 1954.
6. L. Gårding, *Transformation de Fourier des distributions homogènes*, Bull. Sci. Math. France (89), **4**, (1961), 381-428.
7. I. M. Gelfand and G. E. Shilov, *Verallgemeinerte Funktionen*, I. Berlin, 1960.
8. D. S. Jones and M. Kline, *Asymptotic expansion of multiple integrals and the method of stationary phase*, J. Math. Physics **37** (1958) 1-28.
9. J. Leray, *Le calcul différentiel et intégral sur une Variété analytique complexe*, Bull. Soc. Math. France **87** (1959) 6-180.
10. D. Ludwig, *The singularities of the Riemann function*, Rep. No. NYO—9351, Inst. of Math. Sci. New York U. 1961.
11. M. Morse, *The Calculus of variations in the large*, Providence, 1934.
12. R. J. Walker, *Algebraic curves*, Princeton, 1950.

Courant Institute of Mathematical Sciences,
New York University

PACIFIC JOURNAL OF MATHEMATICS
Vol. 15, No. 1, 1965

NORMS AND INEQUALITIES FOR CONDITION NUMBERS

ALBERT W. MARSHALL AND INGRAM OLKIN

The condition number c_φ of a nonsingular matrix A is defined by $c_\varphi(A) = \varphi(A)\varphi(A^{-1})$ where ordinarily φ is a norm. It was proved by O. Taussky-Todd that (c) $c_\varphi(A) \leqq c_\varphi(AA^*)$ when $\varphi(A) = (tr\, AA^*)^{1/2}$ and when $\varphi(A)$ is the maximum absolute characteristic root of A. It is shown that (c) holds whenever φ is a unitarily invariant norm, i.e., whenever φ satisfies $\varphi(A) > 0$ for $A \neq 0$; $\varphi(\alpha A) = |\alpha| \varphi(A)$ for complex α; $\varphi(A + B) \leqq \varphi(A) + \varphi(B)$; $\varphi(A) = \varphi(AU) = \varphi(AU)$ for all unitary U. If in addition, $\varphi(E_{ij}) = 1$, where E_{ij} is the matrix with one in the (i, j)th place and zeros elsewhere, then $c_\varphi(A) \geqq [c_\varphi(AA^*)]^{1/2}$. Generalizations are obtained by exploiting the relation between unitarily invariant norms and symmetric gauge functions. However, it is shown that (c) is independent of the usual norm axioms.

1. Introduction. The genesis of this study is the proposition that under certain conditions, the matrix AA^* is more "ill-conditioned" than A. More precisely, the condition number $c_\varphi(A)$ is defined for nonsingular matrices A as

$$c_\varphi(A) = \varphi(A)\varphi(A^{-1}) ,$$

where ordinarily φ is a norm. The statement concerning ill-conditioning of AA^* is the inequality

$$\text{(c)} \qquad c_\varphi(A) \leqq c_\varphi(AA^*) .$$

Where $\varphi(A)$ is the maximum absolute characteristic root of A and where $\varphi(A) = (tr\, AA^*)^{1/2}$, inequality (c) was proved by O. Taussky-Todd [7]. This raises the question of whether (c) is true for all norms. In this paper, we show that quite the contrary is true; (c) is independent of the usual norm axioms. However, we also prove that (c) does hold for a quite general class of norms.

In the course of proving these results, we obtain some inequalities for symmetric gauge functions, which may be of independent interest.

2. Gauge functions and matrix norms. We call φ a *matrix norm* if

$$\text{(aI)} \qquad \varphi(A) > 0 \quad \text{when } A \neq 0 ,$$

Received March 10, 1964.

(aII) $$\varphi(\alpha A) = |\alpha| \varphi(A) \qquad \text{for complex } \alpha \, ,$$

(aIII) $$\varphi(A + B) \leqq \varphi(A) + \varphi(B) \, .$$

In addition to these basic axioms, various other conditions are sometimes imposed:

(aIV) $$\varphi(E_{ij}) = 1 \, ,$$

where E_{ij} is the matrix with one in the (i, j)th position and zero elsewhere,

(aV) $$\varphi(AB) \leqq \varphi(A)\varphi(B) \, ,$$

(aVI) $$\varphi(A) = \varphi(UA) = \varphi(AU) \qquad \text{for all unitary matrices } U \, .$$

If φ satisfies aI, aII, aIII, and aVI, φ is called a *unitarily invariant norm*.

There is an important connection between unitarily invariant norms and symmetric gauge functions. A function Φ on a complex vector space is called a *gauge function* if

(bI) $$\Phi(u) > 0 \qquad \text{when } u \neq 0 \, ,$$

(bII) $$\Phi(\alpha u) = |\alpha| \Phi(u) \qquad \text{for complex } \alpha \, ,$$

(bIII) $$\Phi(u + v) \leqq \Phi(u) + \Phi(v) \, .$$

Often it is convenient to assume, in addition, that

(bIV) $$\Phi(e_i) = 1 \, ,$$

where e_i is the vector with one in the ith place and zero elsewhere. If, in addition to bI, bII, and bIII,

(bV) $$\Phi(u_1, \cdots, u_n) = \Phi(\varepsilon_1 u_{i_1}, \cdots, \varepsilon_n u_{i_n})$$

whenever $\varepsilon_j = \pm 1$ and $(i_1, \cdots, i_n)$ is a permutation of $(1, \cdots, n)$, then Φ is called a *symmetric gauge function.*

It was noted by Von Neumann [8] that a norm φ is unitarily invariant if and only if there exists a symmetric gauge function Φ such that $\varphi(A) = \Phi(\alpha)$ for all A, where $\alpha_1^2, \cdots, \alpha_n^2$ are the eigenvalues of AA^*.

If Φ is a symmetric gauge function and u, v satisfy $u_i \leqq v_i$, $i = 1, \cdots, n$, then it follows [6, p. 85] that

(2.1) $$\Phi(u_1, \cdots, u_n) \leqq \Phi(v_1, \cdots, v_n) \, .$$

If Φ is a symmetric gauge function satisfying bIV, then [6, p. 86]

(2.2) $$\max_i |u_i| \leqq \Phi(u_1, \cdots, u_n) \leqq \sum_{i=1}^{n} |u_i| \, .$$

If φ is the unitarily invariant matrix norm determined by Φ as above, then it follows that

$$\frac{\varphi(AB)}{\varphi(A)\varphi(B)} \leqq \frac{\sum_{i=1}^{n} \lambda_i(ABB^*A^*)}{[\max_i \lambda_i(AA^*)][\max_j \lambda_j(BB^*)]}$$
$$\leqq \frac{n \max_i \lambda_i(BB^*A^*A)}{[\max_i \lambda_i(AA^*)][\max_j \lambda_j(BB^*)]} \leqq n ,$$

where $\lambda_i(M)$ are the eigenvalues of M. Thus, for any $k \geqq n$, $k\varphi$ is a unitarily invariant matrix norm also satisfying aV. Of course, φ itself satisfies aIV (since Φ satisfies bIV), and this property is destroyed by the renormalization.

3. The condition number inequality.

THEOREM 3.1. *If φ is a unitarily invariant norm, then*

(c) $$c_\varphi(A) \leqq c_\varphi(AA^*) .$$

If Φ is a symmetric gauge function which determines φ, then we may rewrite (c) in the form

$$\Phi(\alpha_1, \cdots, \alpha_n)\Phi(\alpha_1^{-1}, \cdots, \alpha_n^{-1}) \leqq \Phi(\alpha_1^2, \cdots, \alpha_n^2)\Phi(\alpha_1^{-2}, \cdots, \alpha_n^{-2}) .$$

Thus, Theorem 3.1 is a very special case of

THEOREM 3.2. *If Φ is a symmetric gauge function, then $\Phi(\alpha_1^r, \cdots, \alpha_n^r)\Phi(\alpha_1^{-r}, \cdots, \alpha_n^{-r})$ is increasing in $r > 0$, where $\alpha_i > 0$.*

The proof of Theorem 3.2 is embodied in the lemmas below.

Following [2] we say $(a_1, \cdots, a_n)$ is *majorized* by $(b_1, \cdots, b_n)$, written $(a) \prec (b)$, if

(i) $a_1 \geqq \cdots \geqq a_n > 0$, $b_1 \geqq \cdots \geqq b_n > 0$,

(ii) $\sum_1^k a_i \leqq \sum_1^k b_i$, $\quad k = 1, \cdots, n-1$,

(iii) $\sum_1^n a_i = \sum_1^n b_i$.

LEMMA 3.3. *If* $(a) \prec (b)$, *and Φ is a symmetric gauge function, then*

(3.1) $$\Phi(a_1, \cdots, a_n) \leqq \Phi(b_1, \cdots, b_n) ,$$

(3.2) $$\Phi(a_1^{-1}, \cdots, a_n^{-1}) \leqq \Phi(b_1^{-1}, \cdots, b_n^{-1}) .$$

Proof. Proofs of (3.1) have been given by Fan [1] and Ostrowski

[3]; by an argument similar to that of Fan, we prove (3.2).

First, note that we can assume for h and j fixed, $h < j$,

(3.3) $$a_h = \alpha b_h + (1-\alpha)b_j,\ a_j = (1-\alpha)b_h + \alpha b_j,\ a_i = b_i,\ i \neq h, j .$$

That this is true follows from the fact that if $(a) \prec (b)$, then a can be derived from b by successive applications of a finite number of transformations of the form (3.3) (see [2, p. 47]).

Let $\tilde{b} = (b_1, \cdots, b_{h-1}, b_j, b_{h+1}, \cdots, b_{j-1}, b_h, b_{j+1}, \cdots, b_n)$, so that $\Phi(b_1, \cdots, b_n) = \Phi(\tilde{b}_1, \cdots, \tilde{b}_n)$. By convexity,

$$(\alpha b_i + (1-\alpha)\tilde{b}_i)^{-1} \leqq \alpha b_i^{-1} + (1-\alpha)\tilde{b}_i^{-1} .$$

Then using (2.1) and the convexity of Φ, it follows that

$$\begin{aligned}\Phi(a_1^{-1}, \cdots, a_n^{-1}) &= \Phi[(\alpha b_1 + (1-\alpha)\tilde{b}_1)^{-1}, \cdots, (\alpha b_n + (1-\alpha)\tilde{b}_n)^{-1}] \\ &\leqq \Phi(\alpha b_1^{-1} + (1-\alpha)\tilde{b}_1^{-1}, \cdots, \alpha b_n^{-1} + (1-\alpha)\tilde{b}_n^{-1}) \\ &\leqq \alpha\Phi(b_1^{-1}, \cdots, b_n^{-1}) + (1-\alpha)\Phi(\tilde{b}_1^{-1}, \cdots, \tilde{b}_n^{-1}) . \quad ||\end{aligned}$$

As a consequence of Lemma 3.3., we have that if $(a) \prec (b)$ then

$$\Phi(a_1, \cdots, a_n)\Phi(a_1^{-1}, \cdots, a_n^{-1}) \leqq \Phi(b_1, \cdots, b_n)\Phi(b_1^{-1}, \cdots, b_n^{-1}) .$$

The proof of Theorem 3.2 is completed by the following

LEMMA 3.4. *If* $\alpha_1 \geqq \cdots \geqq \alpha_n > 0$ *and* $a_i = \alpha_i^r / \Sigma\alpha_j^r$, $b_i = \alpha_i^s / \Sigma\alpha_j^s$, $0 < r < s$, *then* $(a) \prec (b)$.

Proof. We must show that for all k,

$$\frac{\sum_1^k \alpha_i^r}{\sum_1^n \alpha_i^r} \leqq \frac{\sum_1^k \alpha_i^s}{\sum_1^n \alpha_i^s}, \qquad r < s ,$$

which is true if and only if

$$\sum_1^k \alpha_i^s \sum_{k+1}^n \alpha_j^r - \sum_1^k \alpha_i^r \sum_{k+1}^n \alpha_j^s = \sum_{i=1}^k \alpha_i^r \sum_{j=k+1}^n \alpha_j^r(\alpha_i^{s-r} - \alpha_j^{s-r}) \geqq 0 .$$

The latter follows from $\alpha_i \geqq \alpha_j$, $i < j$. ||

Observe that by (3.1) and Lemma 3.4, we have

$$\frac{\Phi(\alpha_1^r, \cdots, \alpha_n^r)}{\Phi(\alpha_1^s, \cdots, \alpha_n^s)} \leqq \frac{\Sigma\alpha_i^r}{\Sigma\alpha_i^s} .$$

In view of (2.2), it is perhaps natural to expect that

(3.4) $$\frac{\alpha_1^r}{\alpha_1^s} \leqq \frac{\Phi(\alpha_1^r, \cdots, \alpha_n^r)}{\Phi(\alpha_1^s, \cdots, \alpha_n^s)} \leqq \frac{\Sigma\alpha_i^r}{\Sigma\alpha_i^s},\ 0 < r < s,\ \alpha_1 \geqq \cdots \geqq \alpha_n > 0 ,$$

for any symmetric gauge function Φ. To see this we need only prove the left hand inequality, which may be written in the form

$$(3.5)\qquad \Phi\left(\left[\frac{\alpha_1}{\alpha_1}\right]^s, \cdots, \left[\frac{\alpha_n}{\alpha_1}\right]^s\right) \leqq \Phi\left(\left[\frac{\alpha_1}{\alpha_1}\right]^r, \cdots, \left[\frac{\alpha_n}{\alpha_1}\right]^r\right),$$

and which is a consequence of (2.1).

An interesting counterpart to Theorem 3.2 can be obtained from (3.4).

THEOREM 3.5. *If Φ is a symmetric gauge function satisfying* bIV, *then* $[\Phi(\alpha_1^r, \cdots, \alpha_n^r)]^{1/r}$ *is decreasing in* $r > 0$ *whenever* $\alpha_i > 0$, $i = 1, 2, \cdots, n$. *Thus* $[\Phi(\alpha_1^r, \cdots, \alpha_n^r)\Phi(\alpha_1^{-r}, \cdots, \alpha_n^{-r})]^{1/r}$ *is decreasing in* $r > 0$.

Proof. We have that

$$1 \leqq \Phi\left(\left[\frac{\alpha_1}{\alpha_1}\right]^s, \cdots, \left[\frac{\alpha_n}{\alpha_1}\right]^s\right) \leqq \Phi\left(\left[\frac{\alpha_1}{\alpha_1}\right]^r, \cdots, \left[\frac{\alpha_n}{\alpha_1}\right]^r\right),$$

the first inequality by bIV and (2.1). The second inequality is (3.5). Thus

$$\begin{aligned}\left\{\Phi\left(\left[\frac{\alpha_1}{\alpha_1}\right]^s, \cdots, \left[\frac{\alpha_n}{\alpha_1}\right]^s\right)\right\}^r &\leqq \left\{\Phi\left(\left[\frac{\alpha_1}{\alpha_1}\right]^r, \cdots, \left[\frac{\alpha_n}{\alpha_1}\right]^r\right)\right\}^r \\ &\leqq \left\{\Phi\left(\left[\frac{\alpha_1}{\alpha_1}\right]^r, \cdots, \left[\frac{\alpha_n}{\alpha_1}\right]^r\right)\right\}^s,\end{aligned}$$

so that

$$\left\{\Phi\left(\left[\frac{\alpha_1}{\alpha_1}\right]^s, \cdots, \left[\frac{\alpha_n}{\alpha_1}\right]^s\right)\right\}^{1/s} \leqq \left\{\Phi\left(\left[\frac{\alpha_1}{\alpha_1}\right]^r, \cdots, \left[\frac{\alpha_n}{\alpha_1}\right]^r\right)\right\}^{1/r}.$$

The theorem now follows from bII. ||

Theorem 3.5 can, of course, be specialized to yield a kind of converse to (c).

THEOREM 3.6. *If φ is a unitarily invariant norm satisfying* aIV, *then*

$$(\mathrm{c}^*)\qquad [c_\varphi(AA^*)]^{1/2} \leqq c_\varphi(A).$$

Condition (c*) can also be obtained under somewhat different hypotheses. In particular, if φ satisfies aV, then

$$\begin{aligned}c_\varphi(AA^*) &= \varphi(AA^*)\varphi((AA^*)^{-1}) \\ &\leqq \varphi(A)\varphi(A^{-1})\varphi(A^*)\varphi(A^{*-1}) = c_\varphi(A)c_\varphi(A^*).\end{aligned}$$

If also $\varphi(A) = \varphi(A^*)$, then (c*) follows. Of course, $\varphi(A) = \varphi(A^*)$ if φ is unitarily invariant.

4. Independence of the norm axioms and (c). It is our purpose here to show that the condition number inequality (c) does not follow from the usual norm axioms aI – aV. In fact, aII, aIII, aIV, aV and (c) are independent.

REMARK. It has been shown by Ostrowski [4] that aI is implied by aII, aIII, aV, together with $\varphi(A) \not\equiv 0$, so that aI is not included in the list of independent properties. Rella [5] has shown that aII, aIII, aIV and aV are independent, and we add (c) to this list.

The results which prove the independence of aII – aV and (c) are summarized in the following table, where $+(-)$ indicates that a property is true (false).

$\varphi(A)$	aII	aIII	aIV	aV	(c)
1	$-$	$+$	$+$	$+$	$+$
$(\text{rank } A)(\text{tr } AA^*)^{1/2}$	$+$	$-$	$+$	$+$	$+$
$n \max \lvert a_{ij} \rvert$	$+$	$+$	$-$	$+$	$+$
$\max \lvert a_{ij} \rvert$	$+$	$+$	$+$	$-$	$+$
$\Sigma \lvert a_{ij} \rvert$	$+$	$+$	$+$	$+$	$-$

An example which serves in the last line of the table just as well as $\Sigma \lvert a_{ij} \rvert$ is the norm $\max_i \sum_j \lvert a_{ij} \rvert = \sup_x \Phi(xA)/\Phi(x)$, where $\Phi(x) = \sum_i \lvert x_i \rvert$. Norms of this form are called "subordinate" or "lub" norms, and in this case Φ is a symmetric guage function.

The remainder of this paper is devoted to proving the propositions indicated in the table.

The results for $\varphi(A) \equiv 1$ are obvious, so we begin by considering $\varphi(A) = (\text{rank } A)(\text{tr } AA^*)^{1/2}$. In this case, aII and aIV are obvious, and (c) follows from Theorem 3.1, since $(\text{tr } AA^*)^{1/2}$ is unitarily invariant. As is well known, $(\text{tr } AA^*)^{1/2}$ satisfies aV; this together with rank $AB \leqq (\text{rank } A)(\text{rank } B)$ yields aV for $\varphi(A) = (\text{rank } A)(\text{tr } AA^*)^{1/2}$. That aIII is violated may be seen by taking $A = I$ and B the matrix with a unit in the (1, 1)th place and zeros elsewhere.

For $\varphi(A) = n \max_{i,j} \lvert a_{ij} \rvert$ and $\max_{i,j} \lvert a_{ij} \rvert$ the first four columns of the table are well known, and we need only prove (c). Let e_i be the row vector with one in the ith position and zero elsewhere. Denote $M^{-1} = (m^{ij})$ where $M = (m_{ij})$, and let $U = AA^*$. By Cauchy's inequality,

$$|a_{ij}||a^{\alpha\beta}| = |e_i A e_j^*||e_\alpha A^{-1} e_\beta^*| \leqq [(e_i U e_i^*)(e_j e_j^*)(e_\alpha e_\alpha^*)(e_\beta U^{-1} e_\beta^*)]^{1/2} = (u_{ii} u^{\beta\beta})^{1/2} .$$

Hence,

$$\max_{i\ j} |a_{ij}| \max_{\alpha\ \beta} |a^{\alpha\beta}| \leqq (\max_i |u_{ii}| \max_\alpha |u^{\alpha\alpha}|)^{1/2} ,$$

or

$$c_\varphi(A) \leqq [c_\varphi(AA^*)]^{1/2} .$$

Since $U = AA^*$ is positive semi-definite,

$$u_{ii} u^{ii} = (e_i U e_i^*)(e_i U^{-1} e_i^*) \geqq (e_i e_i^*)^2 = 1 ,$$

and it follows that $c_\varphi(AA^*) \geqq 1$. Thus, we have that

$$c_\varphi(A) \leqq [c_\varphi(AA^*)]^{1/2} \leqq c_\varphi(AA^*) , \tag{4.1}$$

which gives (c).

Note that the left inequality of (4.1) is a reversal of inequality (c*). That (4.1) also holds if $\varphi(A)$ is the maximum of the absolute values of the characteristic values of A was proved by O. Taussky-Todd [6].

Since the first four columns of the table are well known for $\varphi(A) = \Sigma |a_{ij}|$, we again need consider only (c). If $A = \begin{pmatrix} B & 0 \\ 0 & 2I \end{pmatrix}$, where $B = \begin{pmatrix} 2 & 1 \\ -1 & 2 \end{pmatrix}$. Then (c) is violated. This same example shows that (c) is violated for $\varphi(A) = \max_i \sum_j |a_{ij}|$.

Bibliography

1. K. Fan, *Maximum properties and inequalities for the eigenvalues of completely continuous operators*, Proc. Nat. Acad. Sci. U.S.A. **37**, (1951) 760–766.
2. G. H. Hardy, J. E. Littlewood and G. Pólya, *Inequalities* (2nd. Ed.), Cambridge University Press, Cambridge. (1959).
3. A. Ostrowski, *Sur quelques applications des fonctions convexes et concaves au sens de I. Schur*, J. Math. Pure Appl. **31**, (1952), 253–292.
4. ———, *Über Normen von Matrizen*, Math. Z., **63**, (1955), 2–18.
5. T. Rella, *Über den absoluten Betrig von Matrizen*, Proc. of the International Congress of Mathematicians, Oslo, (1936). 29–31.
6. R. Schatten, *A theory of cross-spaces*, Princeton University Press, Princeton (1950).
7. O. Taussky-Todd, *Note on the condition of matrices*, Math. Tables Aids Comput., **4**, (1950), 111–112.
8. Von Neumann, *Some matrix-inequalities and metrization of matric-space*, Tomsk. Univ., Rev. **1**, (1937) 286–300, (in Collected Works, Vol. IV, Pergamon Press, 1962).

BOEING SCIENTIFIC RESEARCH LABORATORIES
AND STANFORD UNIVERSITY

PACIFIC JOURNAL OF MATHEMATICS
Vol. 15, No. 1, 1965

FINITISTIC GLOBAL DIMENSION FOR RINGS

HORACE MOCHIZUKI

The finitistic global dimensions $lfPD(R), lFPD(R)$, and $lFID(R)$ are defined for a ring R. We obtain the following results for R semiprimary with Jacobson radical N. C is a simple left R-module and $l.\dim_R C < \infty$, and suppose that $l.\dim_R N^{i-1}/N^i < \infty$ for $i \geqq 3$. Then $m \leqq lfPD(R) = lFPD(R) \leqq (m+1)$. Theorem 2: Suppose that $l.\text{inj.}\dim_R P \leqq l.\text{inj.}\dim_R R/N^2 < \infty$ for every projective (R/N^2)-module P and that $l.\text{inj.}\dim_R N^{i-1}*N^i < \infty$ for $i \geqq 3$. Then $lFID(R) = l.\text{inj.}\dim_R R < \infty$. The method of proof uses a result of Eilenberg and a result of Bass on direct limits of modules together with the lemma: If M is a left R-module such that $N^{k-1}M \neq 0$ and $N^kM = 0$, then every simple direct summand of N^{k-1} is isomorphic to a direct summand of N^{k-1}/N^k.

1. We begin by discussing some further properties of perfect and left perfect rings. The rest of the paper is devoted to giving sufficient conditions for finiteness and equality of certain finitistic global dimensions for a semi-primary ring.

Let R be a ring (with identity). By an R-module we shall always mean a left unitary module over R. In ([7]) and ([10]), Eilenberg and Nakayama define what they called minimal epimorphisms. Bass ([1]) altered this definition to call minimal epimorphisms projective covers. Eilenberg ([7)] studied the dimension theory for modules having minimal epimorphisms and said that a category of modules is perfect if every member of the category has a projective cover. Thus Bass ([1]) called a ring R for which every R-module has a projective cover a left perfect ring. There are two special types of left perfect rings about which we are particularly interested. One is the semi-primary ring R where the Jacobson radical (J-radical) N is nilpotent and R/N is semi-simple with minimum condition (semi-simple), and the other is a ring with minimum condition on left ideals (left Artinian ring).

We define the following finitistic global dimensions for R, using the definitions and notation of ([1]) and ([3]). $lFPD(R) = \sup l.\dim_R M$ for all R-modules of finite projective dimension, $lfPD(R) = \sup l.\dim_R M$ for all finitely generated (f.g.) R-modules of finite projective dimension, $lFWD(R) = \sup w.l.\dim_R M$ for all R-modules of finite weak dimension, $lFID(R) = \sup l.\text{inj.}\dim_R M$ for all R-modules of finite injective dimension.

In § 2 we discuss some further properties of left perfect and perfect rings.

Received January 17, 1964.

In § 3 we give a partial answer to the following questions of Rosenberg and Zelinsky.

(1) Does $lfPD(R) = lFPD(R)$?

(2) Is $lfPD(R)$ finite?

We prove that if R is a semi-primary ring with J-radical N such that N^i/N^{i+1} has finite projective dimension for $i \geqq 2$, then $m \leqq lfPD(R) = lFPD(R) \leqq (m + 1)$ where $m = \sup\{l.\dim_R C$: C is a simple R-module of finite projective dimension$\}$.

In § 4 we prove in a manner similar to § 3 that if R is a left Artinian ring with J-radical N such that N^i/N^{i+1} has finite injective dimension for $i \geqq 2$ and R has finite self-injective dimension, then $lFID(R) = l.\text{inj.}\dim_R R$. We also show that if R is a ring such that the direct product of projectives is projective, if the J-radical N of R has the property that N^i/N^{i+1} has finite injective dimension for $i \geqq 2$, and if R has finite self-injective dimension, then $lFID(R) = l.\text{inj.}\dim_R R$. We conclude by giving examples for the above theorems.

This paper is essentially the first half of the author's dissertation at the University of Washington, Seattle, and was written during his tenure as a National Science Foundation Cooperative Graduate Fellow. The author wishes to express his gratitude to Professor James P. Jans for his advice and encouragement during the preparation of the dissertation.

2. Left perfect rings. Eilenberg ([7]) and Bass ([1]) introduced the following concepts.

DEFINITION 2.1. A submodule B of an R-module A is called superfluous if $B + C = A$ implies $C = A$ whenever C is a submodule of A. An R-homomorphism $f: A \to D$ is called minimal if Ker f is superfluous in A. If A is projective and f is an epimorphism, then f is called a minimal epimorphism. The ring R is called left perfect if every R-module has a minimal epimorphism.

DEFINITION 2.2. An ideal N of a ring R is called left (right) T-nilpotent if given any sequence $\{a_i\}$ of elements in N, we can find an n such that $a_1 a_2 \cdots a_n = 0$ $(a_n \cdots a_2 a_1 = 0)$.

Bass proved the following theorem.

THEOREM 2.3 ([1, Theorem P, p. 467]): *Let R be a ring and N its J-radical. Then the following are equivalent.*

(1) *N is left T-nilpotent and R/N is semi-simple.*

(2) *R is left perfect.*

(3) *Every R-module has the same weak as projective dimension.*

(4) *A direct limit of R-modules of projective dimension* $\leqq n$ *has projective dimension* $\leqq n$.

(5) *R has no infinite sets of orthogonal idempotents, and every nonzero right R-module has nonzero socle (sum of all simple submodules of the right R-module).*

In [7] Eilenberg generalized the concept of semi-primary ring, and generalized a number of theorems of M. Auslander. We state two of them here in a slightly more restricted situation.

PROPOSITION 2.4 ([7, Theorem 11, p. 333]). Let R be a left perfect ring with J-radical N. If A is an R-module, then the following are equivalent.

(1) $\mathrm{Ext}_R^{n+1}(A, R/N) = 0$ where R/N is considered as an R-module.

(2) $\mathrm{Tor}_{n+1}^R(R/N, A) = 0$ where we consider R/N as a right R-module.

(3) $l.\dim_R A \leqq n$.

PROPOSITION 2.5 ([7, Theorem 12, p. 334]): Let R be a perfect (i.e., left and right perfect) ring with J-radical N. Then the following are equivalent.

(1) $l.\,\mathrm{gl.}\dim R \leqq n$.

(2) $l.\dim_R C = w.l.\dim_R C \leqq n$ for all simple R-modules C.

(3) $l.\dim_R (R/N) = w.l.(R/N) \leqq n$ where we consider R/N as an R-module.

(4) $l.\dim_R N = w.l.\dim_R N \leqq n$.

(5) $l.\,\mathrm{inj.}\dim_R (R/N) \leqq n$ where R/N is considered as an R-module.

(6) $r.\dim_R (R/N) = w.r.\dim_R (R/N) \leqq n$ where we treat R/N as a right R-module.

REMARKS. From Proposition 2.4 it is clear that for a left perfect ring R, $l.\,\mathrm{gl.}\dim R \leqq w.r.\dim_R (R/N) \leqq r.\dim_R (R/N) \leqq r.\,\mathrm{gl.}\dim R$. By interchanging the l and the r in Proposition 2.5, we see that $l.\,\mathrm{gl.}\dim R = r.\,\mathrm{gl.}\dim R = \mathrm{gl.}\dim R$ for a perfect ring R.

The following proposition asserts that the simple modules of a right perfect ring serve as test modules for determining injective dimensions of modules.

PROPOSITION 2.6. Let R be a right perfect ring with J-radical N. If A is an R-module, then the following statements are equivalent.

(a) $\mathrm{Ext}_R^{n+1}(C, A) = 0$ for all simple R-modules C.

(b) $\mathrm{Ext}_R^{n+1}(R/N, A) = 0$.

(c) $l.\,\mathrm{inj.}\dim_R A \leqq n$.

Furthermore if $n \geqq 1$, then (b) becomes $\mathrm{Ext}_R^{n+1}(R/N, A) \cong \mathrm{Ext}_R^n(N, A) = 0$.

Proof. (a) $\Leftrightarrow$ (b) and (c) $\Rightarrow$ (a) are obvious. We shall show that (a) $\Rightarrow$ (c)

It is well known ([6]) that we can embed any R-module in an injective R-module. Thus it is possible to form the exact R-sequence:

$$0 \to A \xrightarrow{d_0} Q_0 \xrightarrow{d_1} Q_1 \xrightarrow{d_2} \cdots \xrightarrow{d_{n-1}} Q_{n-1} \xrightarrow{d_n} Q_n \longrightarrow 0$$

where Q_i, $0 \leqq i \leqq (n-1)$, are injective. $\mathrm{Ext}_R^{n+1}(M, A) \cong \mathrm{Ext}_R^1(M, Q_n)$ for all R-modules M where we use the exact sequences

$$0 \longrightarrow A \longrightarrow Q_0 \longrightarrow Im\, d_1 \longrightarrow 0$$

and

$$0 \longrightarrow Im\, d_i \longrightarrow Q_i \longrightarrow Im\, d_{i+1} \longrightarrow 0\ , \qquad 1 \leqq i \leqq (n-1)\ .$$

If we can show that Q_n is injective, then A would have injective dimension $\leqq n$.

It is well known ([3, Chapter I, Theorem 3.2, p. 8]) that Q_n is injective if and only if for each left ideal L, each R-diagram

$$\begin{array}{ccccc} 0 & \longrightarrow & L & \xrightarrow{j} & R \\ & & f\big\downarrow & & \\ & & Q_n & & \end{array}\ ,$$

with j the embedding map can be embedded in a commutative diagram

$$\begin{array}{ccccc} 0 & \longrightarrow & L & \xrightarrow{j} & R \\ & & f\big\downarrow & \swarrow g & \\ & & Q_n & & \end{array}\ .$$

By using Zorn's Lemma (as in the proof of Theorem 3.2 in [3]), we can show that there exists a left ideal L_0 of A containing L such that

$$\begin{array}{ccccc} 0 & \longrightarrow & L & \xrightarrow{k} & L_0 \\ & & f\big\downarrow & \swarrow g_0 & \\ & & Q_n & & \end{array}$$

(k the embedding map) is commutative and that f cannot be extended to any left ideal of R properly containing L_0. If $L_0 = R$, then we are done. If $L_0 \neq R$, then $R/L_0 \neq 0$. According to Theorem 2.3, R/L_0 has nonzero socle $S(R/L_0) = S$.

$S = \bigoplus \Sigma C_i$ is the direct sum of simple R-modules. It is well-known that $\mathrm{Ext}^1_R(\bigoplus \Sigma C_i, Q_n) \cong \Pi\, \mathrm{Ext}^1_R(C_i, Q_n)$, which is the zero module. There exists a left ideal L_1 of R containing L_0 such that $L_1/L_0 \cong S$. $\mathrm{Ext}^1_R(S, Q_n) = 0$ implies that we can extend f (and g_0) to L_1, contradicting the maximality of L_0. Q_n is therefore injective.

COROLLARY 2.7. *If R is a right perfect ring with J-radical N, then* $l.\,\mathrm{gl.\,dim}\, R = l.\,\mathrm{dim}_R\, R/N \geqq r.\,\mathrm{gl.\,dim}\, R$.

Proof. Since $l.\,\mathrm{gl.\,dim}\, R$ is the supremum of injective dimensions of all the R-modules and since $l.\,\mathrm{gl.\,dim}\, R \geqq l.\,\mathrm{dim}\,_R R$, it follows from Proposition 2.6 that $l.\,\mathrm{gl.\,dim}\, R = l.\,\mathrm{dim}\,_R R$. The second part is essentially contained in a theorem proved by Eilenberg ([7, Theorem 12, p. 334]).

In [4] Chase proved some necessary and sufficient conditions that direct products of projective modules be projective. A module A of a ring R is called finitely related if there exists an exact sequence $0 \to K \to F \to A \to 0$ of R-modules where F is free and both F and K are *f.g.*

PROPOSITION 2.8 ([4, Theorem 3.3, p. 467]). For any ring R the following statements are equivalent.

(1) The direct product of any family of projective R-modules is projective.

(2) R is left perfect and finitely generated right ideals of R are finitely related.

Let R be a ring satisfying (1) and (2) in Proposition 2.8. Let $\bigoplus \Sigma R_a\,(a \in X)$ be a direct sum of copies of R as an R-module. Considering the exact sequence

$$0 \longrightarrow \bigoplus \Sigma R_a \longrightarrow \Pi R_a \longrightarrow (\Pi R_a)/(\bigoplus \Sigma\, R_a) \longrightarrow 0\ ,$$

we note that $(\Pi R_a)/(\bigoplus \Sigma R_a)$ is the direct limit of projective R-modules and is therefore projective. The sequence splits, and $\bigoplus \Sigma R_a$ is embedded as a direct summand of ΠR_a.

PROPOSITION 2.9: Let R be a ring satisfying (1) and (2) of Proposition 2.8. If P is projective, then $l.\,\mathrm{inj.\,dim}_R\, P \leqq l.\,\mathrm{inj.\,dim}_R\, R$.

Proof. If P is projective, then P is a direct summand of a direct product ΠR_a copies of R. It then follows by an exercise in C and E ([3, Chapter VI, Exercise 7, p. 123]) that $l.\,\mathrm{inj.\,dim}_R\, R$.

COROLLARY 2.10. *Let R be a ring as in Proposition 2.8. If $lFID(R)$ and $l.\,\mathrm{inj.\,dim}_R\, R$ are both finite, then they are equal.*

Proof. $l.\,\mathrm{inj.\,dim}_R\, P \leqq l.\,\mathrm{inj.\,dim}_R\, R \leqq lFID(R) = n$ where P is any projective R-module. Let A be an R-module such that $l.\,\mathrm{inj.\,dim}_R A = n$. Considering an exact sequence $0 \to K \to P \to A \to 0$ where P is projective, we note that $l.\,\mathrm{inj.\,dim}_R K \leqq n$. Thus we get part of the exact sequence in Ext as follows:

$$\mathrm{Ext}^n_R(B, K) \longrightarrow \mathrm{Ext}^n_R(B, P) \longrightarrow \mathrm{Ext}^n_R(B, A) \longrightarrow 0 ,$$

where B is an arbitrary R-module. But then $\mathrm{Ext}^n_R(B, P) \neq 0$ for an R-module B such that $\mathrm{Ext}^n_R(B, A) \neq 0$. Hence $n \leqq l.\,\mathrm{inj.\,dim}_R P \leqq l.\,\mathrm{inj.\,dim}_R R \leqq n$.

COROLLARY 2.11. *Let R be as in Proposition 2.8. If $lFID(R) = l.\,\mathrm{inj.\,dim}_R R = 0$, then R is left Noetherian, i.e., R is quasi-Frobenius* ([9, Theorem 18, p. 11]).

Proof. According to theorem of Bass ([2, Theorem 1.1, p. 19]) R is left Noetherian if and only if the direct sum of injective R-modules is injective. Let $\{Q_i : i \in I\}$ be a collection of injective R-modules. For each $i \in I$, we consider an exact sequence $0 \to K_i \to P_i \to Q_i \to 0$ where P_i is projective and thus injective. Since $l.\,\mathrm{inj.\,dim}_R K_i$ is finite, K_i is injective, and the sequence splits. Thus Q_i is also projective, and $\bigoplus \Sigma Q_i (i \in I)$ is a projective R-module and hence an injective R-module.

3. Sufficient conditions that $lfPD(R) = lFPD(R) < \infty$. In this section we attempt to give relatively simple sufficient homological conditions to answer questions (1) and (2) of Rosenberg and Zelinsky (appearing in the introduction) in the affirmative. We have the following theorem.

THEOREM 3.1. *Let R be a semi-primary ring with J-radical N. If $l.\,\dim_R (N^{r-1}/N^r) < \infty$ for $r \geqq 3$, then $m \leqq lfPD(R) = lFPD(R) \leqq (m+1)$ where $m = \max\{l.\,\dim_R C : C$ is a simple R-module of finite projective dimension$\}$.*

Before we begin the proof of 3.1 we need a preliminary lemma.

LEMMA 3.2. *Let R be a semi-primary ring with J-radical N such that $N \neq 0$. If M is an R-module such that $N^{r-1}M \neq 0$ and $N^r M = 0$, then $N^{r-1}M$ is the direct sum of simple R-modules which appear as direct summands of N^{r-1}/N^r. Thus $N^{r-s}M/N^{r-s+1}M$, $r \geqq (s-1)$, is the direct sum of simple R-modules which appear as direct summands of N^{r-s}/N^{r-s+1}.*

Proof. The second part follows easily from the first part by noting that $N^{r-s+1}(M/N^{r-s+1}M) = 0$ and $N^{r-s}(M/N^{r-s+1}M) \neq 0$.

We first observe that $(N^{r-1}/N^r)M = \bigoplus \Sigma C_i \, (i \in I)$ is the direct sum of simple R-modules C_i. Let $x \in M$. Then the map $f_x \colon a \to ax$, $a \in N^{r-1}/N^r$, is an R-homomorphism of N^{r-1}/N^r onto $(N^{r-1}/N^r)x$. Let C_i, $i \in I$, be one of the direct summands of $(N^{r-1}/N^r)M$. If $x_0 \in C_i$, then $x_0 = \Sigma a_j x_j \, (1 \leqq j \leqq n)$ where $a_j \in N^{r-1}/N^r$ and $x_j \in M$. Furthermore x_0 generates C_i.

Let $\varphi \colon L = \bigoplus \Sigma(N^{r-1}/N^r)$ (n copies) $\to T = (N^{r-1}/N^r)x_j \, (1 \leqq j \leqq n)$ be the R-epimorphism given by $\varphi\,(\Sigma a_j \, (1 \leqq j \leqq n)) = \Sigma a_j x_j \, (1 \leqq j \leqq n)$. Since L and T are both modules over the semi-simple ring R/N, $L \cong \operatorname{Ker} \varphi \bigoplus T$. C_i is a direct summand of L. By a well-known theorem ([10, Chapter IV, Theorem 2, p. 64]) C_i is isomorphic to a direct summand of N^{r-1}/N^r.

Proof of Theorem 3.1. Let M be an R-module of finite projective dimension. If $NM = 0$, then M is a direct sum of simple R-modules of finite projective dimension. Thus M has projective dimension $\leqq m$ ([3, Chapter VI, Exercise 7, p. 123)].

Suppose then that $NM \neq 0$. We assert that $l.\dim_R N^i M \leqq m$ for $i \geqq 2$. If $N^2 M = 0$, then there is nothing to prove. Hence assume $N^2 M \neq 0$. Let t be the integer such that $N^t M = 0$ and $N^{t-1} M \neq 0$. We have the submodules $N^{t-j}M$, $1 \leqq j \leqq (t-2)$ to examine. We induce on the integer j. If $j = 1$, the $N^{t-1}M$ is the direct sum of simple R-module of projective dimension $\leqq m$ by Lemma 3.2 and the hypotheses of the theorem. $l.\dim_R (N^{t-1}M) \leqq m$, as above. Assume that $l.\dim_R (N^{t-j+1}M) \leqq m$ where $1 < j \leqq (t-2)$. $(N^{t-j}M)/(N^{t-j+1}M)$ has finite projective dimension $\leqq m$, being the direct sum of simple R-modules of projective dimension $\leqq m$ (Lemma 3.2 and the hypotheses of the theorem). From the exact sequence in Ext in the first variable for the exact sequence $0 \to N^{t-j+1}M \to N^{t-j}M \to (N^{t-j}M)/(N^{t-j+1}M) \to 0$ we see that $l.\dim_R (N^{t-j}M) \leqq m$. We have therefore proved our assertion. Since $l.\dim_R (N^2 M) \leqq m$, from the exact sequence $0 \to N^2 M \to M \to M/N^2 M \to 0$ we conclude that $l.\dim_R M/N^2 M < \infty$.

From the exact sequence $0 \to N^2 \to R \to R/N^2 \to 0$ we notice in particular that $l.\dim_R N^2 \leqq m$ and that $l.\dim_R (R/N^2) \leqq m + 1$. Now, R/N^2 as a ring is semi-primary with J-radical N/N^2, and $M/N^2 M$ is an R/N^2-module. Thus $M/N^2 M$ has a minimal epimorphism as an R/N^2 module. Let $0 \to K \to P \to M/N^2 M \to 0$ be the minimal (R/N^2)-epimorphism for $M/N^2 M$. Then $l.\dim_R P \leqq m + 1$ (since P is a direct summand of a direct sum of copies of R/N^2 as an R-module), and $K \subseteqq (N/N^2)P$ ([7, p. 330]). $K = \bigoplus \Sigma C_a \, (a \in I)$ is a direct sum of simple R-modules C_a, and $l.\dim_R K < \infty$. Again applying an exercise in C and E ([3, Chapter VI, Exercise 7, p. 123]), we see that $l.\dim_R C_a < \infty$ for all $a \in I$. Therefore $l.\dim_R K \leqq m$. Using the exact sequence in Ext for the exact sequences $0 \to K \to P \to M/N^2 M \to 0$ and $0 \to N^2 M \to M \to M/N^2 M \to 0$,ǝʍ conclude that $l.\dim_R M \leqq l.\dim_R M/N^2 M \leqq (m + 1)$.

Lastly we assert that M/N^2M is the direct limit of *f.g.* R-modules of finite projective dimension. Let $P = \bigoplus \Sigma P_\alpha \,(\alpha \in \Gamma)$ be a direct decomposition of P into *f.g.* projective (R/N^2)-modules (each of projective dimension $\leqq m+1$). This fact follows from a result of Eilenberg ([7, Proposition 3, p. 330]). If Δ is an arbitrary finite subset of Γ and if J is any finite subset of I such that $\bigoplus \Sigma C_\alpha(\alpha \in J) \subseteqq \bigoplus \Sigma P_\alpha(\alpha \in \Delta)$, then M/N^2M is the direct limit of *f.g.* R-modules $S(\Delta, J)$ where $0 \to \bigoplus \Sigma C_\alpha(\alpha \in J) \to \bigoplus \Sigma P_\alpha(\alpha \in \Delta) \to S(\Delta, J) \to 0$ is exact. Since the first two both have finite projective dimension, so does $S(\Delta, J)$. From Theorem 2.3 it follows that $l.\dim_R M \leqq l.\dim_R (M/N^2M) \leqq lfPD(R)$. Since M was arbitrary with finite projective dimension, we can write that $m \leqq lfPD(R) \leqq lFPD(R) \leqq (m+1)$.

4. Sufficient conditions that $lFID(R) < \infty$.

We state the main theorem of this section.

THEOREM 4.1. *Let R be a semi-primary ring with radical N such that*

(a) *for any projective (R/N^2)-module P,*

$$l.\,\text{inj.}\dim_R P \leqq l.\,\text{inj.}\dim_R R/N^2 < \infty\ .$$

(b) *if $r \geqq 3$, then* $l.\,\text{inj.}\dim_R (N^{r-1}/N^r) < \infty$.

Then $l.\,\text{inj.}\dim_R R < \infty$, and $lFID(R) = l.\,\text{inj.}\dim_R R$.

Proof. Let M be an R-module of finite injective dimension over R. Then it follows, in a manner similar to the proof of Theorem 3.1, that $l.\,\text{inj.}\dim_R N^2M \leqq m$ where $m = \max\{l.\,\text{inj.}\dim_R C : C$ is a simple R-module of finite injective dimension$\}$. From the exact sequence $0 \to N^2 \to R \to R/N^2 \to 0$ it is evident that $l.\,\text{inj.}\dim_R N^2 \leqq m$ and $l.\,\text{inj.}\dim_R R \leqq \max(m, n) < \infty$.

Obviously, $l.\,\text{inj.}\dim_R (M/N^2M) < \infty$. M/N^2M is an R/N^2-module and therefore has a minimal R/N^2-epimorphism $0 \to K \to P \to M/N^2 \to 0$. As in the proof of Theorem 3.2, K is the direct sum of simple R-modules of finite injective dimension. K is thus a direct summand of a direct product of simple R-modules of finite injective dimension, and we have that $l.\,\text{inj.}\dim_R K \leqq m$ ([3, Chapter VI, Exercise 7, p. 123]). From the exact sequence in Ext for the second variable applied to the exact sequence $0 \to K \to P \to M/N^2M \to 0$ we deduce that $l.\,\text{inj.}\dim_R(M/N^2M) \leqq \max(m, n)$ and hence $l.\,\text{inj.}\dim_R M \leqq \max(m, n)$.

$$lFID(R) \leqq \max(m, n) < \infty\ ,$$

and by Corollary 2.10, $lFID(R) = l.\,\text{inj.}\dim_R R$.

We remark that a semi-primary ring R satisfies condition (a) of Theorem 4.1 if R is a left Artinian ring with $l.\,\text{inj.}\dim_R(R/N^2) = n < \infty$, or if R/N^2 satisfies (1) and (2) of Proposition 2.8 and $l.\,\text{inj.}\dim_R R/N^2 = n < \infty$. In the first case we apply the well-known fact ([2, Theorem 1.1, p. 19]) that the direct sum of modules of injective dimension $\leqq n$ has injective dimension $\leqq n$ for Noetherian rings. Thus P a direct summand of a free R-module implies that the injective dimension of P is less than or equal to the injective dimension of R. In the second case we use the remarks prior to Proposition 2.9 together with an exercise in C and E ([3, Chapter VI, Exercise 7, p. 123]) to find that free R-modules and therefore projective R-modules have injective dimension less than or equal to the injective dimension of R.

5. **Examples.** In this section we give examples which satisfy Theorems 3.1 and 4.1 respectively. The construction is essentially that given by Chase in [5].

Let R' and R'' be rings and X a left R'-, right R''-bimodule. We form the ring R consisting of matrices $(a', x, 0, z'')$ where $a' \in R'$, $x \in X$, and $a'' \in R''$. Addition is componentwise and multiplication is given by the equation

$$(a_1', x_1, 0, a_1'')(a_2', x_2, 0, a_2'') = (a_1'a_2', a_1'x_2 + x_1a_2'', 0, a_1''a_2'')\ .$$

Chase proved the following proposition.

PROPOSITION 3.7 ([5, Lemma 4.1, p. 17]). *Let R be as above, and suppose further that R' is semi-primary (respectively Artinian) with radical N' and R'' is semi-simple (with minimum conditions). Then R is semi-primary (respectively Artinian) with radical N consisting of the matrices $(a', x, 0, 0)$ where $a' \in N'$ and $x \in X$. Moreover* $\text{gl.}\dim R = \max(\text{gl.}\dim R', 1 + l.\dim_R X)$.

If G is the finite group of order 2 and K is a field of characteristic 2, then $K(G)$, the group algebra, is a quasi-Frobenius algebra with nonzero radical $N(G)$ such that $N(G)^2 = 0$ ([9, p. 7]). Eilenberg et al. ([8, Proposition 15, p. 94]) have shown that for each positive integer m, there exists a semi-primary ring R_m' with radical N_m' such that $\text{gl.}\dim R_m' = m$ and $(N_m')^2 \neq 0$. Let $R' = K(G) \oplus R''$ (ring direct sum). The radical of R' is $N' = N(G) + N_m'$. We can suppose that R_m' is a finite dimensional algebra over K. Then R' is a vector space over K.

(i) Let $R' = R'$, $X = N'$ and $R'' = K$ in Proposition 3.7. Then the following facts hold.

(a) $\text{gl.}\dim R = l.\dim_R N = \infty$.

(b) $l.\dim_R N^j = l.\dim_{R'}(N')^j \leqq (m-1)$ if $j \geqq 2$.

(c) $l.\,\text{inj.}\dim_R R = \max\{l.\,\text{inj.}\dim_{R'} R', l.\,\text{inj.}\dim_{R'} N'\}$
$= l.\,\text{inj.}\dim_{R'} N' = \infty$.

Thus we have an example of a semi-primary (Artinian) ring satisfying the hypotheses of Theorem 3.1 but not those of Theorem 4.1.

(ii) Suppose that $(N')^q = (N'_m)^q \neq 0$ and $(N')^{q+1} = (N'_m)^{q+1} = 0$ where $q \geqq 2$. Let $R' = R'$, $X = (N')^q$ and $R'' = K$ in Proposition 3.7. Then the following facts hold.

(a) $\text{gl.}\dim R = l.\dim_R N = \infty$.

(b) $l.\text{inj.}\dim_R R = l.\text{inj.}_{R'} R' = m$.

(c) $l.\text{inj.}\dim N^j = l.\text{inj.}\dim_{R'} (N')^j \leqq m$.

This gives an example of an Artinian ring satisfying the hypotheses of Theorem 4.1.

References

1. H. Bass, *Finitistic dimension and a homological generalization of semi-primary rings*, Trans. Amer. Math. Soc. **90** (1960).
2. ———, *Injective dimension in Noetherian rings*, Trans. Amer. Math. Soc. **102** (1962), 18-29.
3. H. Cartan and S. Eilenberg, *Homological algebra*, Princeton University Press, Princeton, 1956.
4. S. Chase, *Direct products of modules*, Trans. Amer. Math. Soc. **97** (1960), 457-473.
5. ———, *A generalization of the ring of triangular matrices*, Nagoya Math. J. **18** (1961), 13-25.
6. B. Eckmann and A. Schopf, *Über injective moduln*, Arch. Math. **4** (1953), 75-78.
7. S. Eilenberg, *Homological dimensions and syzygies*, Ann. of Math. **64** (1956), 328-336.
8. S. Eilenberg, H. Nagao and T. Nakayama, *On the dimension of modules and algebras, IV: Dimension of residue rings of hereditary rings*, Nagoya Math. J. **10** (1956), 87-95.
9. S. Eilenberg and T. Nakayama, *On the dimension of modules and algebras, II: Frobenius algebras and quasi-Frobenius algebras*, Nagoya Math. J. **9** (1955), 1-16.
10. ———, *On the dimension of modules and algebras, V: Dimension of residue rings*, Nagoya Math. J. **11** (1957) 9-12.
11. N. Jacobson, *Structure of rings*, Amer. Soc. Colloquium Publications, **37**, Providence, 1956.

PACIFIC JOURNAL OF MATHEMATICS
Vol. 15, No. 1, 1965

THE COLLINEATION GROUPS OF DIVISION RING PLANES II: JORDAN DIVISION RINGS

R. OEHMKE AND R. SANDLER

In this paper the authors continue their study of the collineation groups of division ring planes (The collineation groups of division ring planes I. Jordan division algebras, J. Reine and Angew. Math. vol. 216, 1964). Some of the results obtained for finite dimensional Jordan division algebras are extended to a special class of infinite dimensional algebras.

As is well-known the study of the collineation group of a projective plane π coordinatized by an algebra $\mathscr{R}$ can be reduced to the study of the autotopism group of $\mathscr{R}$ or the group of autotopic collineations of π, $\mathscr{H}(\pi)$. The pair (a, b), $a, b \in \mathscr{R}$, is defined to be admissible if and only if there exists an element α in $\mathscr{H}(\pi)$ with $(1, 1)\alpha = (a, b)$. Modulo the automorphism group of $\mathscr{R}$, the determination of $\mathscr{H}(\pi)$ is equivalent to the determination of all admissible pairs (a, b) and coset representatives $\varphi_{a,b} \in \mathscr{H}(\pi)$ such that $(1, 1)\varphi_{a,b} = (a, b)$. With either the assumption $\mathscr{R}$ algebraic over its center, or the assumptions characteristic of $\mathscr{R}$ not equal to 0 and the centers of $\mathscr{R}$ and $\mathscr{R}'$ (the algebra of all elements of $\mathscr{R}$ algebraic over the center of $\mathscr{R}$) equal, the admissible pairs (a, b) are determined. Use is made of Kleinfeld's result on the middle nucleus of Jordan rings (Middle nucleus = center in a simple Jordan ring, to appear.) We also prove and use the result that the algebra $\mathscr{E}$ consisting of all right multiplications R_f is commutative, where f is in the subalgebra generated by a and a^{-1} over the base field.

Let $\mathfrak{R}$ be any nonalternative division ring (i.e., $(\mathfrak{R} - \{0\}, \cdot)$ is a loop), and let $\pi(\mathfrak{R})$ be the projective plane coordinatized by $\mathfrak{R}$. Then, as is well known, the study of the collineation group of π, $G(\pi)$, can be reduced to the study of the autotopism group of $\mathfrak{R}$, or the group of autotopic collineations of π, $H(\pi)$. If α is a collineation of π, then $\alpha \in H(\pi)$ if and only if $(\infty)\alpha = (\infty)$, $(0)\alpha = (0)$, $(0, 0)\alpha = (0, 0)$. Now, in [3], the pair (a, b) was defined to be admissible if and only if there exists an element $\alpha \in H(\pi)$ with $(1, 1)\alpha = (a, \mathrm{b})$, and it was shown that, modulo the automorphism group of $\mathfrak{R}$, $H_1(\mathfrak{R})$, the determination of $H(\pi)$ is equivalent to the determination of all admissible pairs (a, b) and coset representatives $\varphi_{a,b} \in H(\pi)$:

$$(1) \qquad (1, 1)\varphi_{a,b} = (a, b) .$$

Received January 23, 1964.

The second part of [3] was concerned with planes coordinatized by finite dimensional Jordan division algebras, and it was proved that if $\mathfrak{R}$ is a finite dimensional Jordan division algebra of characteristic $\neq 2, 3$, then (a, b) is admissible if and only if a and b are elements in the center of $\mathfrak{R}$, and coset representatives of $\varphi_{a,b}$ were obtained for a, b in the center of $\mathfrak{R}$. In this paper, we shall prove the following theorem:

THEOREM A. *If $\mathfrak{R}$ is a Jordan division algebra of characteristic $\neq 2, 3$, and if either*

(*a*) *$\mathfrak{R}$ is algebraic over its center, Z; or,*

(*b*) *$\mathfrak{R}$ has characteristic $\neq 0$, and the center of $\mathfrak{R}$ is equal to the center of $\mathfrak{R}'$—the algebra of all elements algebraic over Z; then (a, b) is admissible if and only if a and b are both in Z.*

We shall need a recent result of Kleinfeld in the proof of Theorem A, and quote it here:

THEOREM 1 [2]. *If $\mathfrak{R}$ is a simple Jordan ring of characteristic $\neq 2$, the three nuclei of $\mathfrak{R}$ are equal.*

This generalizes Theorem 15 of [3] and is useful in that with this result we need only show that an element, a, is in any one of the nuclei of $\mathfrak{R}$ in order to prove that a is in the *center* of $\mathfrak{R}$.

Our first step will be to prove some results about Jordan division rings which are analogous to known theorems about finite dimensional Jordan algebras, and which are necessary tools for this paper. Recall that the linearized form of the Jordan identity can be written [1],

$$\begin{aligned} (2) \qquad & R_xR_{zw} + R_wR_{xz} + R_zR_{xw} \\ &= R_{zw}R_x + R_{xz}R_w + R_{xw}R_z \\ &= R_{x(zw)} + R_zR_xR_w + R_wR_xR_z \\ &= R_{z(xw)} + R_xR_zR_w + R_wR_zR_x , \end{aligned}$$

where R_x is the linear transformation corresponding to multiplication in $\mathfrak{R}$ by the element x.

We now prove

THEOREM 2. *Let a be an element of a Jordan division algebra, $\mathfrak{R}$. Then a and a^{-1} generate an associative subalgebra of $\mathfrak{R}$.*

Proof. If $\mathfrak{R}$ is finite dimensional over Z, this result is a trivial consequence of the well known result [1] that any Jordan algebra is power-associative. For the infinite-dimensional case, it suffices to prove

(3) $$a^{-i}a^k = a^{k-i} \qquad \text{for all} \quad i,\ k \geqq 1.$$

For $i = 1$, set $x = a^k$, $z = w = a$ in (2), and apply the last two resulting transformations to a^{-1} to obtain

$$a^{-1}(R_{a^{k+2}} + 2R_aR_{a^k}R_a) = a^{-1}(R_{a^{k+2}} + R_{a^k}R_aR_a + R_aR_aR_{a^k}),$$

or

$$a^{-1}a^{k+2} + 2a^{k+1} = a^{-1}a^{k+2} + [(a^{-1}a^k)a]a + a^{k+1},$$

which implies $a^{k+1} = [(a^{-1}a^k)a]a = a^kR_a = [(a^{-1}a^k)a]R_a$. Since $\mathfrak{R}$ is a division ring, R_a is nonsingular, and the last equation implies $a^k = (a^{-1}a^k)a = (a^{k-1})R_a = (a^{-1}a^k)R_a$, which, in turn, implies $a^{k-1} = a^{-1}a^k$. Thus, (3) is verified for $i = 1$. For $i > 1$, set $x = a^{-1}$, $y = a$, $z = a^k$ in (2) and apply the first two resulting transformations to a^{-i}:

$$\begin{aligned} &a^{-i}(R_{a^{-1}}R_{a^{k+1}} + R_{a^k} + R_aR_{a^{k-1}}) \\ &= a^{-i}(R_{a^{k+1}}R_{a^{-1}} + R_{a^k} + R_{a^{k-1}}R_a)\ , \end{aligned}$$

or,

(4) $$\begin{aligned} &a^{-(i+1)}a^{k+1} + a^{-i}a^k + [(a^{-i})a]a^{k-1} \\ &= (a^{-i}a^{k+1})a^{-1} + a^{-i}a^k + [a^{-i}a^{k-1}]a\ . \end{aligned}$$

If we assume $a^{-j}a^k = a^{k-j}$ for all k and all $j < i + 1$, (4) becomes

$$a^{-(i+1)}a^{k+1} + 2a^{k-i} = 3a^{k-i}\ ,$$

which implies $a^{-(i+1)}a^{k+1} = a^{k-i}$, which together with the truth of (3) for $i = 1$ and all k, completes the inductive proof of the theorem.

Another result which is analogous to a well-known theorem for finite-dimensional Jordan algebras [1] is:

THEOREM 3. *If $\mathfrak{R}$ is a Jordan division algebra over a field $\mathfrak{F}$ of characteristic $\neq 2$, and if a is any element of $\mathfrak{R}$, then the algebra $\mathfrak{S}$ generated by all R_x, for $x \in \mathfrak{F}[a, a^{-1}]$, is commutative.*

Proof. In (2), set $x = a^{-1}$, $w = a$, $z = a^i$, and get

(5) $$R_{a^{-1}}R_{a^{i+1}} + R_aR_{a^{i-1}} + R_{a^i} = R_{a^{i+1}}R_{a^{-1}} + R_{a^{i-1}}R_a + R_{a^i}\ .$$

In [1], it is shown that the Jordan identity implies that R_{x^i}, R_{x^j} commute for any x, and all i, $j \geqq 0$. Thus, for $i \geqq 1$, (5) can be simplified to

(6) $$R_{a^{-1}}R_{a^{i+1}} = R_{a^{i+1}}R_{a^{-1}}\ .$$

Next, let $x = a^{-2}$, $w = a^i$, $z = a$ in (2), and get

$$(7) \qquad \begin{aligned} &R_{a^{-2}}R_{a^{i+1}} + R_{a^i}R_{a^{-1}} + R_aR_{a^{i-2}} \\ &= R_{a^{i+1}}R_{a^{-2}} + R_{a^{-1}}R_{a^i} + R_{a^{i-2}}R_a \ . \end{aligned}$$

If $i \geqq 2$, we can use (6) and the fact that $R_aR_{a^{i-2}} = R_{a^{i-2}}R_a$ to simplify (7) to

$$(8) \qquad R_{a^{-2}}R_{a^{i+1}} = R_{a^{i+1}}R_{a^{-2}} \ .$$

Thus, for all $i \geqq 3$, R_{a^i} commutes with all elements in $\mathfrak{S}$ generated by $R_{a^{-1}}$ and $R_{a^{-2}}$. Since the set of all R_f, $f \in \mathfrak{F}[a^{-1}]$, is generated by $R_{a^{-1}}$ and $R_{a^{-2}}$ [1], we can conclude that for $i \geqq 3$, R_{a^i} is in the center of $\mathfrak{S}$. Similarly, we can show that for $i \geqq 3$, $R_{a^{-i}}$ is in the center of $\mathfrak{S}$. Next, we substitute in (2), $x = z = a^2$, $w = a^{-4}$, and, using the fact that $R_{a^{-4}}$ is in the center of $\mathfrak{S}$, we conclude

$$(9) \qquad R_{a^2}R_{a^{-2}} = R_{a^{-2}}R_{a^2} \ .$$

Finally, substituting $x = z = a$, $w = a^{-2}$, and using (9), we can deduce

$$(10) \qquad R_aR_{a^{-1}} = R_{a^{-1}}R_a \ .$$

But from (6), we know that $R_{a^{-1}}R_{a^2} = R_{a^2}R_{a^{-1}}$. Thus, we see that $R_{a^{-1}}$ commutes with R_a, R_{a^2}, $R_{a^{-1}}$, and $R_{a^{-2}}$, and hence that $R_{a^{-1}}$ is in the center of $\mathfrak{S}$. Similarly, we can conclude that R_a, R_{a^2}, $R_{a^{-2}}$ are also in the center of $\mathfrak{S}$, and thus that $\mathfrak{S}$ is commutative.

We now turn to the proof of Theorem A. Recall that in [3] the admissibility of (a, b) for any nonalternative $\mathfrak{R}$ was seen to be equivalent with the isomorphism of $\mathfrak{R}$ and a certain isotope of $\mathfrak{R}$, $\mathfrak{S}_{a,b}$, obtained by recoordinatizing π with the new coordinate points $(\infty)' = (\infty)$, $(0)' = (0)$, $(0, 0)' = (0, 0)$, and $(1, 1)' = (a, b)$. Now, in [3], (Sec. 9) the following theorem was proved but not stated explicitly.

THEOREM 4. *If $\mathfrak{R}$ is commutative, and if the middle nucleus of $\mathfrak{R}$ is equal to the center of $\mathfrak{R}'$ and if $\mathfrak{F}[R_{x^i}]$ is commutative for all $x \in R$, then if $\mathfrak{S}_{a,b}$ is commutative, $\mathfrak{S}_{a,b}$ can be represented as follows: $\mathfrak{S}_{a,b} \approx (\mathfrak{R}, \oplus, *)$, where*

$$(11) \qquad x \oplus y = x + y \ ,$$

and multiplication in $\mathfrak{S}_{a,b}$ is given in terms of multiplication in $\mathfrak{R}$:

$$(12) \qquad (y * x) = [(yR_{a^{-1}}^{-1})(xR_{a^{-1}}^{-1})]R_a^{-1}R_{a^{-1}} \ .$$

Also, $a^2b^{-1} \in Z$.

Notice that (12) is equivalent to

$$(13) \qquad \tilde{R}_x = R_{a^{-1}}^{-1}R_xR_{a^{-1}}^{-1}R_a^{-1}R_{a^{-1}} \ .$$

Since if $\mathfrak{S}_{a,b}$ is to be isomorphic with $\mathfrak{R}$, $\mathfrak{S}_{a,b}$ must be commutative and using Theorems 1 and 3, we have the validity of Theorem 4 in our present study. From this point on, then we assume that $\mathfrak{S}_{a,b}$ is of the form given by Theorem 4 and wish to determine under what conditions on the element a, $\mathfrak{R}$ and $\mathfrak{S}_{a,b}$ are isomorphic.

We begin with

THEOREM 5. *Let $\mathfrak{R}$ be a Jordan division ring, of characteristic $\neq 2, 3$, and let $\mathfrak{S}_{a,b}$ be as in Theorem 4. Then if $\mathfrak{S}_{a,b}$ satisfies the Jordan identity, we have*

$$R_{a^i} = R_a^i\,[q_i(T - I) + I] \qquad \textit{for any } i \geqq 0\,, \tag{14}$$

where

$$T = R_a R_{a^{-1}}\,, \tag{15}$$

and

$$q_i = \frac{i(i-1)}{2}\,. \tag{16}$$

Proof. This theorem for $\mathfrak{R}$ finite dimensional is contained in [3] (Sec. 10, Lemma 2). The proof for the infinite-dimensional case is exactly the same, keeping in mind that Theorem 3 must be invoked to let us permute elements of the form R_{a^i} and $R_{a^{-j}}$.

Assume now that $\mathfrak{R}$ has characteristic $p \neq 0$. We shall prove that Theorem 5, together with the Jordan identity for $\mathfrak{R}$ imply that a is algebraic over Z if $\mathfrak{S}_{a,b}$ satisfies the Jordan identity. To see this, observe first that for any k, (14) implies the equalities

$$R_{a^{kp}} = R_a^{kp} = R_{(a^p)^k} = (R_a^p)^k\,, \tag{17}$$

since $q_{kp} = 0$.

Thus, if $c = a^p$, we have

$$R_{c^k} = R_c^k, \qquad \text{for all } k \geqq 0\,. \tag{18}$$

Next, recall [3] that the linearized form of the Jordan identity for $\mathfrak{S}_{a,b}$ can be written

$$\begin{aligned} &R_x R_a^{-1} R_{(zw)} R_a^{-1} + R_w R_a^{-1} R_{(zx)} R_a^{-1} + R_z R_a^{-1} R_{(xw)} R_a^{-1} \\ &= R_{(zw)} R_a^{-1} R_x + R_{(zx)} R_a^{-1} R_a^{-1} R_w + R_{(xw)} R_a^{-1} R_a^{-1} R_z \\ &= R_{(zw)} R_a^{-1} R_x R_a^{-1} + R_z R_a^{-1} R_x R_a^{-1} R_w + R_w R_a^{-1} R_x R_a^{-1} R_z\,. \end{aligned} \tag{19}$$

We wish to prove a commutator identity:

THEOREM 6. *If $\mathfrak{S}_{a,b}$ satisfies the Jordan identity, then*

$$(20)\qquad [R_x, R_{c^{i+1}}] = (i+1)[R_{xc^i}, R_c]\ ,$$

for $i \geqq 0$, $c = a^p$, *and for all* $x \in \mathfrak{R}$.

Proof. In (19), let $w = c$, $z = c^i$. Then we have

$$\begin{aligned} &R_x R_{c^{i+1}} + R_c R_{xc^i} + R_{c^i} R_{xc} \\ &= R_{c^{i+1}} R_x + R_{xc^i} R_c + R_{xc} R_{c^i}\ , \end{aligned}$$

or,

$$(21)\qquad [R_x, R_{c^{i+1}}] = [R_{xc^i}, R_c] + [R_{xc}, R_{c^i}]\ .$$

Thus (20), for $i = 1$ is verified. If $i > 1$, we apply our induction hypothesis to the right hand side of (21), and write

$$(22)\qquad [R_{xc}, R_{c^i}] = i[R_{(xc)c^{i-1}}, R_c] = i[R_x R_c R_{c^{i-1}}, R_c]\ .$$

But by (18), we can write $R_c R_{c^{i-1}} = R_c R_c^{i-1} = R_c^i = R_{c^i}$, so (22) becomes

$$(23)\qquad [R_{xc}, R_{c^i}] = i[R_{xc^i}, R_c]\ ,$$

which allows us to write (21) as

$$(24)\qquad [R_x, R_{c^{i+1}}] = (i+1)[R_{xc^i}, R_c]\ ,$$

and complete the inductive proof of Theorem 6.

Now in (20), if we set $i = p - 1$, we obtain

$$(25)\qquad [R_x, R_{c^i}] = 0, \qquad \text{for all } x \in \mathfrak{R},$$

but this is equivalent, in our case, to asserting that $c = a^p$ *is in the center* of $\mathfrak{R}$. Thus, as promised, we demonstrated that if $\mathfrak{S}_{a,b}$ is a Jordan ring, then $a^p = c \in Z$, and a *is algebraic over* Z.

The completion of the proof of Theorem A is quite simple. If $\mathfrak{S}_{a,b}$ satisfies the Jordan identity, and if a is algebraic over Z, then $a \in \mathfrak{R}'$—the algebra of all algebraic elements. Now, let a', a'' be any two elements of $\mathfrak{R}'$, and consider $\mathfrak{R}[a, a', a'']$, the subalgebra of $\mathfrak{R}'$ generated by a, a', a''. Since (19) holds for all $x, y, z \in \mathfrak{R}$, it certainly also holds for all $x, y, z \in \mathfrak{R}[a, a', a'']$. But in [3], it was shown that if (2) and (19) hold for any commutative algebra, then a is in the center of that algebra. Thus, a is in the center of $\mathfrak{R}[a, a', a'']$ for any $a', a'' \in \mathfrak{R}'$, which completes the proof of Theorem A.

As a final remark, we observe that [3] (Sec. 14), a coset representative $\varphi_{a,b}$ for $H(\pi)/H_i(\mathfrak{R})$ was explicitly determined for every admissible pair (a, b) for which both a and b were in the center. Thus, for those algebras satisfying the conditions of Theorem A, all the coset representatives are actually known, and the collineation group for such a plane is thus *completely determined* modulo the automorphism group of the algebra.

REFERENCES

1. A. A. Albert, *A structure theory for Jordan algebras*, Ann. of Math. **48** (1947), 546–567.
2. E. Kleinfeld, *Middle nucleus = center in a simple Jordan ring*, to appear.
3. R. H. Oehmke and R. Sandler, *The collineation groups of division ring planes I: Jordan division algebras*, J. Reine und Angew. Math. **216** (1964), 67–87.

INSTITUTE FOR DEFENSE ANALYSES

PACIFIC JOURNAL OF MATHEMATICS
Vol. 15, No. 1, 1965

ON NON-CONVEX POLYHEDRAL SURFACES IN E^2

GEORGE H. ORLAND

In this note only the simplest type of non-convex polyhedral surface will be examined. These surfaces will be characterized as the only non-convex polyhedral surfaces which have a convex polar.

Proofs will depend upon the possibility of associating convex surfaces to the non-convex ones. Thus, by comparing a non-convex surface with its convex "second polar", it will be possible to discover how bending affects the angles of the non-convex surface. In a similar way the Gauss-Bonnet relationship will be verified.

The tools of vector analysis provide simple and efficient means for keeping track of the sizes of angles on one side of a non-convex polyhedral surface. The triple scalar product, together with the rules for its manipulation and expansion, will be used a great deal. Thus Lemma 1 expresses known facts in terms of these products. In this connection, $\operatorname{sgn}[t_1t_2t_3]$ is to mean the algebraic sign of the triple scalar product $[t_1t_2t_3]$ of the vectors t_1, t_2, and t_3. Furthermore $\operatorname{sgn}[t_1t_2t_3]$ will only be written when $[t_1t_2t_3] \neq 0$.

Polyhedral corners. Let $t_1, \cdots, t_k$ be an ordered set of vectors in E^3 where $k \geqq 3$ and any three consecutive vectors t_{i-1}, t_i, and t_{i+1} are linearly independent. (All indices will always be reduced modulo k.) The set of all vectors which are linear combinations of t_i and t_{i+1} with nonnegative coefficients will be denoted by $\prod_{i,i+1}$. Furthermore let the intersection of $\prod_{i-1,i}$ and $\prod_{j,j+1}$ be no more than the origin unless $i = j$, $i-1 = j+1$, or $i-1 = j$. When these conditions are satisfied the collection of all vectors in the $\prod_{i,i+1}$'s will be called a *polyhedral corner*. The origin is the *vertex*, the t_i's are the *edges*, and the $\prod_{i,i+1}$'s are the *faces* of the polyhedral corner. The angle between t_i and t_{i+1} is the *face angle* $\varphi_{i,i+1}$ of $\prod_{i,i+1}$. The *normal* to the face $\prod_{i,i+1}$ is the vector $n_{i,i+1} = t_i \times t_{i+1}$. The *exterior angle* e_i formed by $\prod_{i-1,i}$ and $\prod_{i,i+1}$ will have the magnitude of the angle between $n_{i-1,i}$ and $n_{i,i+1}$, and the sign of $[t_{i-1}t_it_{i+1}]$. The *dihedral angle* δ_i formed by $\prod_{i-1,i}$ and $\prod_{i,i+1}$ is $180° - e_i$. A polyhedral corner Σ will be called *convex* if for each i, the plane of $\prod_{i,i+1}$ is a plane of support for Σ.

LEMMA 1. *Let $\Sigma = \Sigma(t_i)$ be a polyhedral corner. Then Σ is convex*

Received December 18, 1963.

if and only if $\operatorname{sgn}[t_{i-1}t_it_{i+1}]$ *is constant.*

Proof. Suppose Σ is convex. It is sufficient to show that

$$\operatorname{sgn}[t_{i-1}t_it_{i+1}] = \operatorname{sgn}[t_it_{i+1}t_{i+2}]$$

for any i. However t_{i-1} and t_{i+2} must lie on the same side of the plane of $\prod_{i,i+1}$, so this equality holds.

The "if" part of this lemma will not be used in what follows and is only mentioned for completeness. For this reason the proof, which is not trivial, will be omitted.

LEMMA 2. *An ordered set of vectors* $\{t_1, \cdots, t_k\}$ *determines a convex polyhedral corner if and only if* $\operatorname{sgn}[t_{i-1}t_it_j]$ *is constant for all* i *and all* j *different from* $i-1$ *and* i.

Proof. If $\Sigma = \Sigma(t_i)$ is a convex corner, then $\operatorname{sgn}[t_{i-1}t_it_{i+1}]$ is constant for all i. For any particular i, the plane of $\prod_{i-1,i}$ supports Σ so that $\operatorname{sgn}[t_{i-1}t_it_j] = \operatorname{sgn}[t_{i-1}t_it_{i+1}]$ for all j different from $i-1$ and i.

To prove the converse we must first show that the t_i's determine a polyhedral corner. Since $[t_{i-1}t_it_{i+1}] \neq 0$ for all i, any three consecutive vectors are linearly independent. Now suppose the faces $\prod_{i-1,i}$ and $\prod_{j,j+1}$ are distinct, and $i \neq j$ and $i-1 \neq j+1$. Any vector common to both of them is of the form $\alpha t_{i-1} + \beta t_i = \gamma t_j + \delta t_{j+1}$, where $\alpha, \beta, \gamma, \delta$ are nonnegative. By taking the inner product of both sides of this equation with $t_{i-1} \times t_i$ we get

$$0 = \gamma[t_{i-1}t_it_j] + \delta[t_{i-1}t_it_{j+1}] ,$$

so that $\gamma = \delta = 0$ and $\prod_{i-1,i} \cap \prod_{j,j+1} = 0$. Thus we have a polyhedral corner $\Sigma(t_i)$. The hypothesis now implies that $\Sigma(t_i)$ is convex.

Corners and polars. Starting with some given polyhedral corner $\Sigma = \Sigma(t_i)$ we may ask whether or not the vectors $n_{k,1}, n_{1,2}, \cdots, n_{k-1,k}$ form the edges of a new polyhedral corner. For this to occur we must first of all have every consecutive set of three of them linearly independent. However

$$[n_{i-1,i}n_{i,i+1}n_{i+1,i+2}] = [t_it_{i+1}t_{i+2}][t_{i-1}t_it_{i+1}] \neq 0 ,$$

so that this is automatically satisfied. The second condition, that nonadjacent faces intersect in exactly the origin, need not be satisfied in every instance. When the normals do form a new polyhedral corner we call it the *polar polyhedral corner*, or just the *polar* of Σ, and denote it by Σ^p.

LEMMA 3. *A polyhedral corner $\Sigma = \Sigma(t_i)$ is convex if and only if it has a polar Σ^p which is convex, and has* sgn $[n_{i-1,i}n_{i,i+1}n_{j,j+1}]$ *positive for all i and for all j different from $i-1$ and i.*

Proof. Suppose Σ is convex. Using Lemma 2, the existence and convexity of Σ^p will follow from the positivity of sgn $[n_{i-1,i}n_{i,i+1}n_{j,j+1}]$ for all i and all j different from $i-1$ and i. This in turn follows by expanding

$$\begin{aligned}[n_{i-1,i}n_{i,i+1}n_{j,j+1}] &= [(t_{i-1} \times t_i) \times (t_i \times t_{i+1})]\cdot(t_j \times t_{j+1}) \\ &= [t_{i-1}t_it_{i+1}]\cdot[t_it_jt_{j+1}]\end{aligned}$$

and noting that, by Lemma 2, it is positive.

To prove the converse let sgn $[n_{i-1,i}n_{i,i+1}n_{j,j+1}]$ be positive for all i and all j different from $i-1$ and i. The preceding equation tells us that sgn $[t_{i-1}t_it_{i+1}] =$ sgn $[t_it_jt_{j+1}]$ for those same i's and j's. One more application of Lemma 2 shows that $\Sigma(t_i)$ is convex.

There is still another way for a polyhedral corner to have a convex polar. This would occur if the sign in Lemma 3 were taken to be negative.

LEMMA 4. *Let $\Sigma = \Sigma(t_i)$ be a polyhedral corner with at least four edges. Then* sgn $[t_it_jt_{j+1}]$ *is constant as a function of j, where j is different from $i-1$ and i, and alternates as a function of i if and only if* sgn $[n_{i-1,i}n_{i,i+1}n_{j,j+1}]$ *is negative for all i and all j different from $i-1$ and i. When this is the case Σ has a convex polar. In fact Σ has exactly four edges.*

Proof. The equivalence can be shown, in the following way, to be a consequence of the identity

$$[n_{i-1,i}n_{i,i+1}n_{j,j+1}] = [t_{i-1}t_it_{i+1}]\cdot[t_it_jt_{j+1}] \ .$$

First suppose sgn $[t_it_jt_{j+1}]$ satisfies the hypothesis. Then

$$\begin{aligned}\text{sgn}\,[t_{i-1}t_it_{i+1}] &= \text{sgn}[t_{i-1}t_{i+1}t_{i+2}] \\ &= -\text{sgn}\,[t_it_{i+1}t_{i+2}] = -\text{sgn}\,[t_it_jt_{j+1}]\end{aligned}$$

and sgn $[n_{i-1,i}n_{i,i+1}n_{j,j+1}]$ is negative. Conversely the negativity of sgn $[n_{i-1,i}n_{i,i+1}n_{j,j+1}]$ shows that sgn $[t_it_jt_{j+1}]$ is constant as j varies, and also that sgn $[t_it_jt_{j+1}] = -$sgn $[t_{i-1}t_it_{i+1}] = -$sgn $[t_{i-1}t_jt_{j+1}]$.

When these hypotheses are satisfied, Lemma 2 shows that Σ^p exists and is convex.

The expansion of $(t_1 \times t_2) \times (t_j \times t_{j+1})$ can be performed in two ways to yield the identity

$$[t_1t_2t_{j+1}]t_j - [t_1t_2t_j]t_{j+1} = [t_1t_jt_{j+1}]t_2 - [t_2t_jt_{j+1}]t_1 .$$

By hypothesis $\operatorname{sgn}[t_1t_2t_{j+1}] = \operatorname{sgn}(-[t_1t_2t_j])$ and

$$\operatorname{sgn}[t_1t_jt_{j+1}] = \operatorname{sgn}(-[t_2t_jt_{j+1}]) .$$

Now should $\operatorname{sgn}[t_1t_2t_{j+1}] = \operatorname{sgn}[t_1t_jt_{j+1}]$, then the faces $\Pi_{1,2}$ and $\Pi_{j,j+1}$ would intersect in points besides the origin. We conclude that

$$\operatorname{sgn}[t_1t_2t_{j+1}] = -\operatorname{sgn}[t_1t_jt_{j+1}]$$

when neither j nor $j+1$ is equal to 1 or 2. Now suppose Σ has five or more edges. Then

$$\operatorname{sgn}[t_1t_2t_5] = \operatorname{sgn}[t_2t_3t_5] = \operatorname{sgn}[t_3t_4t_5] = \operatorname{sgn}[t_1t_4t_5]$$

which is impossible. (It is easy to give examples of four-edged polyhedral corners which do satisfy the hypothesis.)

We shall call such non-convex polyhedral corners, which have convex polars, *saddle corners.*

Bending a saddle corner. Starting with a saddle corner $\Sigma = \Sigma(t_i)$ we shall form Σ^{pp}, the polar of the polar of Σ, and see how it is related to Σ. An edge of Σ^p is $n_{i,i+1} = t_i \times t_{i+1}$ and a corresponding edge of Σ^{pp} is

$$m_i = n_{i-1,i} \times n_{i,i+1} = (t_{i-1} \times t_i) \times (t_i \times t_{i+1}) = [t_{i-1}t_it_{i+1}]t_i .$$

Corresponding to the face angle $\varphi_{i,i+1} = \sphericalangle(t_i, t_{i+1})$ of Σ is the face angle

$$\varphi'_{i,i+1} = \sphericalangle(m_i, m_{i+1}) = \sphericalangle([t_{i-1}t_it_{i+1}]t_i, [t_it_{i+1}t_{i+2}]t_{i+1})$$

of Σ^{pp}. Since

$$\operatorname{sgn}[t_{i-1}t_it_{i+1}] = -\operatorname{sgn}[t_it_{i+1}t_{i+2}] ,$$

$\varphi'_{i,i+1} = 180° - \varphi_{i,i+1}$. Corresponding to the dihedral angle

$$\delta_i = 180° - \operatorname{sgn}[t_{i-1}t_it_{i+1}] \sphericalangle (n_{i-1,i}, n_{i,i+1})$$

of Σ, the dihedral angle of Σ^{pp} is

$$\delta'_i = 180° - \operatorname{sgn}[m_{i-1}m_im_{i+1}] \sphericalangle (m_{i-1} \times m_i, m_i \times m_{i+1}) .$$

Now

$$\begin{aligned}
&\operatorname{sgn}[m_{i-1}m_im_{i+1}] \\
&\quad = \operatorname{sgn}[[t_{i-2}t_{i-1}t_i]t_{i-1}, [t_{i-1}t_it_{i+1}]t_i, [t_it_{i+1}t_{i+2}]t_{i+1}] \\
&\quad = \operatorname{sgn}([t_{i-2}t_{i-1}t_i]\cdot[t_{i-1}t_it_{i+1}]^2[t_it_{i+1}t_{i+2}]) = +1 ,
\end{aligned}$$

and

$$\begin{aligned}
&\sphericalangle(m_{i-1} \times m_i, m_i \times m_{i+1}) \\
&\quad = \sphericalangle([t_{i-2}t_{i-1}t_i]t_{i-1} \times [t_{i-1}t_it_{i+1}]t_i, [t_{i-1}t_it_{i+1}]t_i \times [t_it_{i+1}t_{i+2}]t_{i+1}) \\
&\quad = \sphericalangle(-n_{i-1,i}, -n_{i,i+1}) = \sphericalangle(n_{i-1,i}, n_{i,i+1}) .
\end{aligned}$$

Thus

$$\delta'_i = 180° - \sphericalangle(n_{i-1,i}, n_{i,i+1}) .$$

The second relation will be used to obtain the following result. The first will be used later.

BENDING THEOREM. *Let $\Sigma = \Sigma(t_i)$ and $\Gamma = \Gamma(r_i)$ be two saddle corners whose corresponding face angles are equal. Let δ_i be the dihedral angle of Σ at t_i and γ_i be the dihedral angle of Γ at r_i. Then* $\operatorname{sgn}(\delta_i - \gamma_i)$ *alternates as a function of i, in the case where for all i,* $\operatorname{sgn}[r_{i-1}r_ir_{i+1}] = -\operatorname{sgn}[t_{i-1}t_it_{i+1}]$. *In the case where for all i,* $\operatorname{sgn}[r_{i-1}r_ir_{i+1}] = \operatorname{sgn}[t_{i-1}t_it_{i+1}]$, *either all $\delta_i - \gamma_i$ have the same sign or they are all zero.*

Proof. The normal $r_i \times r_{i+1}$ to a face of Γ will be called $s_{i,i+1}$. In the case where $\operatorname{sgn}[r_{i-1}r_ir_{i+1}] = -\operatorname{sgn}[t_{i-1}t_it_{i+1}]$,

$$\begin{aligned}
\delta_i - \gamma_i &= [180° - \operatorname{sgn}[t_{i-1}t_it_{i+1}] \sphericalangle (n_{i-1,i}, n_{i,i+1})] \\
&\qquad - [180° - \operatorname{sgn}[r_{i-1}r_ir_{i+1}] \sphericalangle (s_{i-1,i}, s_{i,i+1})] \\
&= \operatorname{sgn}[r_{i-1}r_ir_{i+1}][\sphericalangle(s_{i-1,i}, s_{1,i+1}) + \sphericalangle(n_{i-1,i}, n_{i,i+1})]
\end{aligned}$$

which alternates as a function of i. In the case where

$$\operatorname{sgn}[r_{i-1}r_ir_{i+1}] = \operatorname{sgn}[t_{i-1}t_it_{i+1}] ,$$

$$\begin{aligned}
\delta_i - \gamma_i &= \operatorname{sgn}[r_{i-1}r_ir_{i+1}][\sphericalangle(s_{i-1,i}, s_{i,i+1}) - \sphericalangle(n_{i-1,i}, n_{i,i+1})] \\
&= \operatorname{sgn}[r_{i-1}r_ir_{i+1}]\{[180° - \sphericalangle(n_{i-1,i}, n_{i,i+1})] \\
&\qquad - [180° - \sphericalangle(s_{i-1,i}, s_{i,i+1})]\} \\
&= \operatorname{sgn}[r_{i-1}r_ir_{i+1}]\cdot(\delta'_i - \gamma'_i) ,
\end{aligned}$$

where δ'_i and γ'_i are dihedral angles of Σ^{pp} and Γ^{pp} respectively, corresponding to δ_i and γ_i respectively. We are now able to apply a well-known "four vertex" theorem [1, Chapt. II, p. 12] to the two convex polyhedral corners Σ^{pp} and Γ^{pp}. For the situation we are considering, this theorem states that either $\delta'_i - \gamma'_i$ is zero for all i, or is zero for no i and alternates in sign as a function of i. Since $[r_{i-1}r_ir_{i+1}]$ also alternates in sign, the assertion is proved.

This theorem has the following interpretation when we think of the saddle corner as having hinged edges. Picture the corner being bent and thereby having its dihedral angles altered. Then all of its dihedral angles will be altered in the same direction provided it remains a saddle corner throughout the process.

The case where the signs alternate arises in the following way. Take a saddle corner, form a mirror image of it, and "bend" this mirror image. Then compare this last corner with the original one.

The Gauss-Bonnet result. In [2] Polya discussed and proved a version of the Gauss-Bonnet theorem for convex polyhedral surfaces. We shall extend his Lemma II to saddle corners. (His Lemma II is just the statement of the $G-B$ theorem for convex corners.) Other methods will probably be needed to extend this lemma to general non-convex polyhedral angles.

Let Σ be a saddle corner so Σ^p is convex. Call $-K$ the (negative) total curvature of Σ which is, in magnitude, equal to the solid angle included by Σ^p. Since $(\Sigma^{pp})^p$ is Σ^p (or its mirror image), the total curvature of Σ^{pp} is just K. The total geodesic curvature can be found by taking a polygonal path around a polyhedral corner and computing the total change in direction along this path. This turns out to be exactly the sum of the face angles for that corner. The $G-B$ theorem is valid for Σ^{pp} so $K+\varphi'_{1,2}+\cdots+\varphi'_{4,1}=360°$. The sum of the face angles of Σ is

$$\begin{aligned}\varphi_{1,2}+\cdots+\varphi_{4,1} &= (180°-\varphi'_{1,2})+\cdots+(180°-\varphi'_{4,1})\\ &= 720°-(\varphi'_{1,2}+\cdots+\varphi'_{4,1})\,.\end{aligned}$$

Therefore

$$\begin{aligned}&(-K)+(\varphi_{1,2}+\cdots+\varphi_{4,1})\\ &\quad=-K+720°-(\varphi'_{1,2}+\cdots+\varphi'_{4,1})\\ &\quad=720°-360°=360°\,,\end{aligned}$$

and the $G-B$ theorem is valid for Σ.

References

1. H. Hopf, *Geometry*, Mimeographed lecture notes presented to the Institute of Mathematics and Mechanics, New York, N.Y. (1952).

2. G. Polya, *An elementary analogue to the Gauss-Bonnet theorem*, Amer. Math. Monthly, **61** (1954), 601-603.

University of Illinois

PACIFIC JOURNAL OF MATHEMATICS
Vol. 15, No. 1, 1965

COLLINEATION GROUPS OF SEMI-TRANSLATION PLANES

T. G. OSTROM

This paper consists in an investigation of the collineations of a class of planes constructed by the author. The construction consists of replacing the lines of a net embedded in a given plane by subplanes of the same net.

For the case in question, the given plane is the dual of a translation plane. The full collineation group of the new plane is isomorphic to a subgroup of the collineation group of the original plane. The main point of the argument is to show that the new planes admit no collineations displacing the line at infinity.

I. In [2], the author introduced a new class of affine planes. These new planes were obtained by a construction which consists of starting with a plane which is the dual of a translation plane and modifying some of the lines. By the very process of construction, a part of the collineation group of the original plane is carried over to the new plane.

However, the full collineation group for these new planes has not been previously determined; in particular, it has not been known whether there are any collineations displacing the line at infinity. In this paper, we show that (with mild restrictions on the nature of the original plane) the full collineation group on each new plane is precisely the group "inherited" from the original plane.

II. Preliminary definitions and summary of previous results. We shall be using Hall's ternary [4] and certain slight modifications of the ternary as coordinate systems for planes. The point at infinity on the line $y = xm$ will be denoted by (m); the point at infinity on $x = 0$ will be denoted by (∞).

In any case where the coordinate system contains a subfield $\mathfrak{F}$ it should be understood that small Greek letters (with the exception of ρ and σ) denote elements of $\mathfrak{F}$.

For any affine plane Π and any set $\mathfrak{S}$ of parallel classes, the system consisting of the points of Π and lines belonging to the parallel classes in $\mathfrak{S}$ will be called a net N embedded in Π. If (m) is the point at infinity corresponding to some parallel class in N, we shall

Received January 28, 1964. This work was supported (in part) by Grant No. GP1623 from the National Science Foundation.

find it convenient to speak of (m) as "belonging to N."
A quasifield (Veblen-Wedderburn system) will be said to be a left quasifield if the left distributive law, $a(b + c) = ab + ac$, holds.

Let $\mathfrak{X}$ be a coordinate system with associative and commutative addition. If $\mathfrak{X}$ contains a subfield $\mathfrak{F}$ such that

$$a(\alpha + \beta) = a\alpha + a\beta \tag{1}$$

$$(a\alpha)\beta = a(\alpha\beta) \tag{2}$$

$$(a + b)\alpha = a\alpha + b\alpha \,, \tag{3}$$

for all a, b in $\mathfrak{X}$ and all α, β in $\mathfrak{F}$, we shall say that $\mathfrak{X}$ is a right vector space over $\mathfrak{F}$.

If lines whose slopes are in $\mathfrak{F}$ can be represented by equations of the type $y = x\alpha + b$, we shall say that $\mathfrak{X}$ is linear with respect to $\mathfrak{F}$.

Now let $\mathfrak{X}$ be a left quasifield of order q^2 $(q > 4)$. Suppose that $\mathfrak{X}$ is a right vector space over a subfield $\mathfrak{F}$ of order q. Let Π be the affine plane coordinatised in the usual sense by $\mathfrak{X}$. (Note: The line of slope m through the origin is written $y = xm$ rather than with m on the left.)

We can then define another plane $\bar{\Pi}$ [2] whose points are identical with the points of Π. The lines of $\bar{\Pi}$ are of two kinds:

(1) Lines of Π which have finite slopes not in $\mathfrak{F}$
(2) Sets of points (x, y) such that $x = a\alpha + c$, $y = a\beta + d$,

where $a \neq 0$, c, d are fixed elements of $\mathfrak{X}$ while α and β vary over $\mathfrak{F}$.

Now the lines of type (2) may be identified with subplanes (of order q) of Π. If a permutation σ on the points of Π induces a collineation of either Π or $\bar{\Pi}$ which carries lines of type (1) into lines of type (1), then σ induces collineations of both planes. If σ is a translation (elation with axis L_∞) of either plane, then σ is a translation of both planes [3].

Now let t be a fixed element of $\mathfrak{X}$ which is not in $\mathfrak{F}$. Each element of $\mathfrak{X}$ can be written uniquely in the form $t\alpha + \beta$. The lines of $\bar{\Pi}$ can be written in a more convenient form if each point is assigned new coordinates as follows:

$$\text{If} \quad (x, y) = (t\xi_1 + \xi_2, t\eta_1 + \eta_2), \text{ let}$$
$$(\bar{x}, \bar{y}) = (t\xi_1 + \eta_1, t\xi_2 + \eta_2).$$

Define a new operation $*$ such that

$$(t\xi_1 + \eta_1)*(t\lambda_1 + \lambda_2) = t\xi_2 + \eta_2 \qquad \text{is equivalent to}$$
$$(t\xi_1 + \xi_2)\,(t\mu_1 + \mu_2) = t\eta_1 + \eta_2; \;\; (t\xi_1 + \xi_2)*\lambda_2 = t\xi_1\lambda_2 + \eta_1\lambda_2 \,,$$

where $\lambda_1 \neq 0$ and $\lambda_1(t\mu_1 + \mu_2) = t + \lambda_2$.
See reference [3].

Then the lines of $\bar{\Pi}$ can be represented by equations of the following forms:

Type (1): $\bar{y} = (\bar{x} - \alpha)*m + \beta, m \notin \mathfrak{F}$

Type (2): $\bar{y} = \bar{x}\delta + b$ or $x = c$.

Let Π_0 denote the affine subplane of Π which is coordinatised by $\mathfrak{F}$ in $\mathfrak{T}$; let $\bar{\Pi}_0$ be the affine subplane of $\bar{\Pi}$ which is coordinatised by $\mathfrak{F}$ in $\bar{T}$. Then Π_0 is the set of points for which $\bar{x} = 0$; $\bar{\Pi}_0$ is the set of points for which $x = 0$.

The plane Π admits all translations of the form $(x, y) \rightarrow (x, y + b)$. The points of $\bar{\Pi}_0 (x = 0)$ are in a single transitive class under this group of translations—which also acts as a group of translations on $\bar{\Pi}$. There will be further translations if and only if there is some element c such that $(x + c)m = xm + cm$ for all x and all m. If there are no further translations, $\bar{\Pi}$ is what we call a strict semi-translation plane; we shall say that T is a strict left quasifield.

III. The collineation group. It is well known that a net can be coordinatised in much the same fashion as a plane. If the net is embedded in a plane, a coordinate system for the plane induces a coordinate system for the net, provided the lines $x = 0$, $y = 0$ and $y = x$ all belong to the net. Conversely, any coordinate system for the net can be extended to form a coordinate system for the whole plane.

LEMMA 1. *Let N be a net with $q + 1$ parallel classes. Let N be coordinatised by a system $\mathfrak{C}$, let F be the subset of $\mathfrak{C}$ such that $x\alpha$ is defined for all x in $\mathfrak{C}$, all α in $\mathfrak{F}$. Suppose that*

(1) *Addition in $\mathfrak{C}$ is associative.*

(2) *F is a field of order q with respect to addition and multiplication in $\mathfrak{C}$.*

(3) *The additive group in $\mathfrak{C}$ is a right vector space over F.*

(4) *$\mathfrak{C}$ is linear.*

Then N can be embedded in a Desarguesian plane.

Proof. The additive group is isomorphic to the additive group of a field $\mathfrak{K}$ which is a quadratic extension of $\mathfrak{F}$. For instance, if q is odd, multiplication in $\mathfrak{K}$ may be defined as follows

$$(t\xi_1 + \xi_2) \circ (t\lambda_1 + \lambda_2) = t(\xi_1\lambda_2 + \xi_2\lambda_1) + (\delta\xi_1\lambda_1 + \xi_2\lambda_2) ,$$

where δ is a fixed nonsquare element of $\mathfrak{F}$ and t is a fixed element not in $\mathfrak{F}$. Then the net N will be embedded in the Desarguesian plane coordinatised by $\mathfrak{K}$.

LEMMA 2. *Let $\mathfrak{T}$ be a left quasifield coordinatising a plane Π of order q^2. Suppose that* (1) *$\mathfrak{T}$ is a right vector space over a subfield $\mathfrak{F}$ of order q and* (2) *$\mathfrak{T}$ is linear with respect to $\mathfrak{F}$. Let $\mathfrak{T}'$ be any other coordinate system for Π subject to the following condition* (a). *The point (∞) is the same for both $\mathfrak{T}$ and $\mathfrak{T}'$,* (b) *$\mathfrak{T}'$ is an extension of a coordinate system for the net N consisting of those parallel classes whose slopes in $\mathfrak{T}$ are infinite or belong to $\mathfrak{F}$.*

Then $\mathfrak{T}'$ is also a left quasifield satisfying conditions (1) *and* (2).

Proof. The plane Π is a dual translation plane with special point (∞). This implies that $\mathfrak{T}'$ is a left quasifield.

It follows from Lemma 1 that any coordinate system for N must have properties (1) and (2). These properties will carry over to $\mathfrak{T}'$.

We now return to the construction discussed in part II. It is to be understood that $\mathfrak{T}$ is a left quasifield of order q^2 which is a right vector space over a subfield of order q, that T is linear with respect to $\mathfrak{F}$, and that $\bar{\Pi}$ is the new plane introduced in part II.

Since we shall ultimately be concerned with collineations which might displace the line at infinity, we shall want to deal with the projective version of $\bar{\Pi}$. We modify our previous notation so that (m) denotes the point at infinity on $y = x*m$.

THEOREM 1. *If $\mathfrak{T}$ is a strict left quasifield, then the affine collineations of $\bar{\Pi}$ are precisely those which it shares with Π.*

Proof. For each α in $\mathfrak{F}$, there are exactly q translations of $\bar{\Pi}$ with center (α). Likewise, there are q translations with center (∞). If $\mathfrak{T}$ is a strict left quasifield, so that Π and $\bar{\Pi}$ admit exactly q^2 translations, we have exhausted the translations in $\bar{\Pi}$.

This implies that no affine collineations of $\bar{\Pi}$ carry a line of type (1) into a line of type (2). Hence every affine collineation of $\bar{\Pi}$ is a collineation of Π.

LEMMA 3. *Suppose that $\bar{\Pi}$ admits a collineation which carries the line at infinity into some line L. Then, without loss of generality, we may take L to be $\bar{x} = 0$.*

Proof. By Lemma 2 of [3], L is some line of type 2, hence L consists of the set of points of an affine subplane of Π. By Lemma 2, we can choose a new coordinate system $\mathfrak{T}'$ for Π such that this subplane is coordinatised by a field of order q and $\mathfrak{T}'$ is a left quasifield satisfying (1) and (2) of Lemma 2. If $\mathfrak{T}$ is initially chosen in this way, L has the equation $\bar{x} = 0$. Since the basic construction consists of replacing lines by subplanes (see [3]), the change of coordinate system for Π does not alter the nature of $\bar{\Pi}$.

LEMMA 4. *If $\bar{\Pi}$ admits a collineation carrying L_∞ into $\bar{x} = 0$, multiplication in $\bar{T}$ takes the form*

$$\begin{aligned}(t\alpha_1 + \beta_1) * (t\alpha_2 + \beta_2) = t[&h(\alpha_1, \alpha_2) - \beta_1\alpha_2 + \alpha_1\beta_2] \\ &+ [\beta_1\alpha_1^{-1} h(\alpha_1, \alpha_2) + k(\alpha_1, \alpha_2) \\ &- \beta_1^2 \alpha_1^{-1} \alpha_2 + \beta_1 \beta_2] \quad \alpha_1 \neq 0\end{aligned}$$

and

$$\beta_1 * (t\alpha_2 + \beta_2) = t\alpha_2 \beta_1 + \beta_1 \beta_2 + R(\beta_1, \alpha_2)$$

where h, k, and R are functions from $\mathfrak{F} \times \mathfrak{F}$ into $\mathfrak{F}$.

Proof. By Lemma 2 of [3], (∞) is the center of q elations with axis $\bar{x} = 0$. These collineations act on Π in such a way as to leave Π_0 pointwise fixed. Since $x = 0$ is fixed in Π, $\bar{\Pi}_0$ is fixed (not pointwise) in $\bar{\Pi}$. Thus we have a group of elations of $\bar{\Pi}$ which is transitive on the q points of $\bar{\Pi}_0 \cap L_\infty - (\infty)$.

There is a similar group of elations in Π which has center (∞), axis $x = 0$, and is transitive on the points at infinity of Π_0 (excluding the point at infinity of $x = 0$). These collineations carry over into $\bar{\Pi}$, appearing as collineations which leave $\bar{\Pi}_0$ pointwise fixed. The collineations leaving $\bar{\Pi}_0$ pointwise fixed impose automorphisms of $\bar{T}$ which fix each element of F. The elations of $\bar{\Pi}$ with center (∞) and axis $x = 0$ impose the "partial distributive law" $a*(b + \alpha) = ab + a\alpha$ $a, b \in \bar{\mathfrak{T}}$, $\alpha \in \mathfrak{F}$, on $\bar{T}$. Lemma 4 then follows from Theorem 2 and 3 of [1].

LEMMA 5. *Under the conditions of the previous Lemmas, $\bar{\mathfrak{T}}$ has the property that if $b * a = -1$, then $b * (a * m) = (-1) * m$ for all m in $\bar{\mathfrak{T}}$.*

Proof. The proof is essentially the same as the proof of Theorem 11 in [1].

LEMMA 6. *Under the conditions of the previous lemmas, there exist functions f and g such that $h(\alpha_1, \alpha_2) = f(\alpha_1)\alpha_2$, $k(\alpha_1, \alpha_2) = g(\alpha_1)\alpha_2$.*

Proof. Given $t\alpha_1$ $(\alpha_1 \neq 0)$, let $t\alpha_2 + \beta_2$ be determined so that $t\alpha_1 * (t\alpha_2 + \beta_2) = -1$. By Lemma 4, $h(\alpha_1, \alpha_2) + \alpha_1\beta_2 = 0$

$$k(\alpha_1, \alpha_2) = -1.$$

By Lemma 5, we have $t\alpha_1 * (t\alpha_2\gamma + \beta_2\gamma) = -\gamma$, which is equivalent to the pair of equations

$$h(\alpha_1, \alpha_2\gamma) + \alpha_1\beta_2\gamma = 0$$
$$k(\alpha_1, \alpha_2\gamma) = -\gamma.$$

Now, by Theorem 11 of [3], $\bar{T}$ is a right vector space over $\mathfrak{F}$. In particular, $(t\alpha_1 + \beta_1) * \beta_2 = t\alpha_1\beta_2 + \beta_1\beta_2$. From this, and our definition of α_2, we know that $\alpha_2 \neq 0$. We easily get $k(\alpha_1, \alpha_2\gamma) = k(\alpha_1, \alpha_2)\gamma$ for each nonzero α_1 and γ in F, where α_2 depends on α_1. Letting $\alpha_2\gamma = \alpha$, we get $k(\alpha_1, \alpha) = k(\alpha_1, \alpha_2)\alpha_2^{-1}\alpha = g(\alpha_1)\alpha$. Moreover, $k(\alpha_1, 0) = 0$. This establishes the part of our Lemma that pertains to k. A similar argument works for h.

THEOREM 2. *Under the hypotheses of Theorem* 1 *and the additional requirement that* $q > 4$, $\bar{\Pi}$ *admits no collineations displacing* L_∞; *the full collineation group of* $\bar{\Pi}$ *is the group of affine collineations which it shares with* Π.

Proof. The relations between the multiplications in T and $\bar{T}$ is reciprocal, i.e.

$$(t\xi_1 + \eta_1)(t\lambda_1 + \lambda_2) = t\xi_2 + \eta_2 \iff$$
$$(t\xi_1 + \xi_2) * (t\mu_1 + \mu_2) = t\eta_1 + \eta_2 \qquad \text{if } \lambda_1 \neq 0\,,$$

where
$$\lambda_1 * (t\mu_1 + \mu_2) = t + \lambda_2\,.$$

Let us assume that $\bar{\Pi}$ does admit a collineation displacing L_∞. We shall show that we must have $q \leqq 4$. Now let $\lambda_2 = 0$, $\xi_1 \neq 0$, $\bar{\lambda}_1 \neq 0$. We have:

$$(t\xi_1 + \eta_1)(t\lambda_1) = t\xi_2 + \eta_2$$

is equivalent to

$$(t\xi_1 + \xi_2) * (t\lambda_1^{-1} - \lambda_1^{-1} R(\lambda_1, \lambda_1^{-1})) = t\eta_1 + \eta_2\,,$$

which is in turn equivalent to the pair of equations (by Lemmas 4 and 6)

$$\eta_1 = f(\xi_1)\lambda_1^{-1} - \xi_2\lambda_1^{-1} - \xi_1\lambda_1^{-1}R(\lambda_1, \lambda_1^{-1})$$
$$\eta_2 = \xi_2\,\xi_1^{-1} f(\xi_1)\lambda_1^{-1} + g(\xi_1)\lambda_1^{-1} - \xi_2^2\,\xi_1^{-1}\,\lambda_1^{-1} - \xi_2\,\lambda_1^{-1}\,R(\lambda_1, \lambda_1^{-1}).$$

Let $R(\lambda_1, \lambda_1^{-1}) = S(\lambda_1)$. Solving for ξ_2 and η_2, we get

$$(t\xi_1 + \eta_1)(t\lambda_1) = t[f(\xi_1) - \xi_1 S(\lambda_1) - \eta_1\lambda_1]$$
$$+ [g(\xi_1)\lambda_1^{-1} + f(\xi_1)\eta_1\xi_1^{-1} - \eta_1 S(\lambda_1) - \eta_2^2\xi_1^{-1}\lambda_1]\,.$$

By hypotheses, $\mathfrak{T}$ is a left quasifield which is a right vector space over $\mathfrak{F}$. Hence

$$(t\xi_1 + \eta_1)[t(\lambda + \mu)] = (t\xi_1 + \eta_1)(t\lambda) + (t\xi_1 + \eta_1)(t\mu) .$$

Carrying out the multiplications in the above equation and separating the components, we get the two equations

$$\begin{aligned} f(\xi_1) - \xi_1 S(\lambda + \mu) - \eta_1(\lambda + \mu) = [f(\xi_1) - \xi_1 S(\lambda) - \eta_1\lambda] \\ + [f(\xi_1) - \xi_1 S(\mu) - \eta_1\mu] , \end{aligned}$$

$$\begin{aligned} & g(\xi_1)(\lambda + \mu)^{-1} + f(\xi_1)\eta_1\xi_1^{-1} - \eta_1 S(\lambda + \mu) - \eta_1^2 \xi_1^{-1}(\lambda + \mu) \\ & = [g(\xi_1)\lambda^{-1} + f(\xi_1)\eta_1\xi_1^{-1} - \eta_1 S(\lambda) - \eta_1^2 \xi_1^{-1}\lambda] \\ & \quad + [g(\xi_1)\mu^{-1} + f(\xi_1)\eta_1\xi_1^{-1} - \eta_1 S(\mu) - \eta_1^2 \xi_1^{-1}\mu]. \end{aligned}$$

Eliminating $f(\xi_1)$, we find that the terms involving S also drop out and we get

$$g(\xi_1)(\lambda + \mu)^{-1} = g(\xi_1)\lambda^{-1} + g(\xi_1)\mu^{-1}.$$

Now if $g(\xi_1) = 0$, then $(t\xi_1)(t\lambda) = t[f(\xi_1) - \xi_1 S(\lambda_1)]$. But the solution of any equation of the type $(t\xi)x = t\beta$ is $x = \xi^{-1}\beta$, which is in $\mathfrak{F}$.

Since $t\lambda \notin \mathfrak{F}$, we have a contradiction. We conclude that $g(\xi_1) \neq 0$. Hence we must have $(\lambda + \mu)^{-1} = \lambda^{-1} + \mu^{-1}$ for all λ, μ in $\mathfrak{F}$ except in the cases that λ, μ, or $\lambda + \mu$ is zero.

With $\mu = 1$, this equation is equivalent to

$$\lambda^2 + \lambda + 1 = 0, \ \lambda \neq 0, -1 .$$

Hence $\mathfrak{F}$ can contain at most 4 elements. Since we assumed $q > 4$, the theorem is proved.

REFERENCES

1. D. L. Morgan and T. G. Ostrom, *Coordinate systems for semi-translation planes*, Trans. Amer. Math. Soc. (to appear).
2. T. G. Ostrom, *A class of non-desarguesian affine planes*, Trans. Amer. Math. Soc. **104** (1962), 483-487.
3. ——, *Semi-translation planes*, Trans. Amer. Math. Soc. (to appear).
4. M. Hall, *Projective planes*, Trans. Amer. Math. Soc. **54** (1943), 229-277.

PACIFIC JOURNAL OF MATHEMATICS
Vol. 15, No. 1, 1965

ON ANTI-COMMUTATIVE ALGEBRAS AND GENERAL LIE TRIPLE SYSTEMS

ARTHUR A. SAGLE

A general Lie triple system as defined by K. Yamaguti, is considered as an anti-commutative algebra A with a trilinear operation $[x, y, z]$ in which (among other things) the mappings $D(x, y) : z \to [x, y z]$ are derivations of A. It is shown that if the trilinear operation is homogeneous, and A is irreducible as a general L. t. s. or irreducible relative to the Lie algebra $I(A)$ generated by the $D(x, y)$'s, then A is a simple algebra. The main result is the following. If A is a simple finite-dimensional anti-commutative algebra over a field of characteristic zero which is a general L. t. s. with a homogeneous trilinear operation $[x, y, z]$, then A is (1) a Lie algebra; or (2) a Malcev algebra; or (3) an algebra satisfying $J(x, y, z)w = J(w, x, yz) + J(w, y, zx) + J(w, z, xy)$ where $J(x, y, z) = xy \cdot z + yz \cdot x + zx \cdot y$. Furthermore in all three cases I(A) is the derivation algebra of A and $I(A)$ is completely reducible in A.

1. *A general Lie triple system* (*general* L. t. s.) has been defined in [6] to be a vector space V over a field F which is closed with respect to a trilinear operation $[x, y, z]$ and a bilinear operation xy so that

$$[x, y, z] = 0, \tag{1.1}$$

$$x^2 = 0, \tag{1.2}$$

$$[x, y, z] + [y, z, x] + [z, x, y] - (xy)z - (yz)x - (zx)y = 0, \tag{1.3}$$

$$[wx, y, z] + [xy, w, z] + [yw, x, z] = 0, \tag{1.4}$$

$$[[u, v, w], x, y] + [[v, u, x], w, y] + [v, u, [w, x, y]] + [w, x, [u, v, y]] = 0, \tag{1.5}$$

$$[w, x, yz] + z[w, x, y] + y[x, w, z] = 0. \tag{1.6}$$

A general L. t. s. is an extension of a Lie triple system used in differential geometry and Jordan algebras. Next we note that if V is a Lie algebra with multiplication xy, then V becomes a general L. t. s. by setting $[x, y, z] = (xy)z$. As an extension of this it was shown in [7] that if V is a Malcev algebra [2] with multiplication xy, then V becomes a general L. t. s. by setting $[x, y, z] = -(xy)z + (yz)x + (zx)y$.

Received January 23, 1964. Sponsored by the U. S. Army Research Office (Durham) and by NSF Grant GP-1453.

In this paper we shall take the point of view that a general L. t. s. is an algebra A with multiplication xy satisfying (1.1) – (1.6); that is, A is an anti-commutative algebra with linear transformations

$$D(x, y)\colon\ A \to A : z \to zD(x, y) \equiv [x, y, z]$$

for all $x, y \in A$ satisfying (1.1)–(1.6). Thus from (1.6) we see that each $D(x, y)$ is a derivation of A satisfying various identities. Motivated by the above examples we shall assume that the trilinear product $[x, y, z]$ is a linear homogeneous expression in the products of x, y and z. Thus using anti-commutativity we assume there exist fixed α, β, $\gamma \in F$ so that

(1.7) $\qquad [x, y, z] = \alpha\, xy\cdot z + \beta\, yz\cdot x + \gamma\, zx\cdot y.$

With (1.7) as the form for the trilinear operation, we next consider irreducibility conditions on A. First suppose A is irreducible as a general L. t. s.; that is, A has no proper general L. t. s. subspace B so that $[B, A, A] \subset B$, then A is a simple algebra. For if B is a proper ideal of A, then from (1.7) B is general L. t. s. subspace so that $[B, A, A] \subset B$. Next let $I(A)$ be the subspace of the derivation Lie algebra $D(A)$ generated by all derivations of the form $D(x, y)$, then from (1.5) $I(A)$ is a Lie subalgebra of $D(A)$ under commutation. If A is $I(A)$-irreducible, then by (1.7) A is a simple algebra. Motivated by these remarks the main result is the following

THEOREM. *If A is a simple finite dimensional anti-commutative algebra over a field of characteristic zero with a nonzero trilinear operation $[x, y, z]$ satisfying* (1.1)–(1.7), *then*

(1) *A is a simple Lie algebra with $\beta = \gamma$, $\alpha - \beta = 1$; or*

(2) *A is a simple Malcev algebra* [4] *with $\alpha = -1, \beta = \gamma = 1$; or*

(3) *A is a simple algebra satisfying $J(x, y, z)w = J(w, x, yz) + J(w, y, zx) + J(w, z, xy)$ with $\alpha = 1/2$, $\beta = \gamma = 1/4$ and $J(x, y, z) = xy\cdot z + yz\cdot x + zx\cdot y$.*

Furthermore in all three cases $I(A)$ is the derivation algebra of A and $I(A)$ is completely reducible in A.

It should be noted that since the trilinear operation $[x, y, z]$ given by (1.7) is homogeneous, any nonzero scalar multiple would also be an admissible trilinear operation. Therefore all superfluous nonzero scalars will be eliminated to obtain the final above normalizations for α, β and γ.

2. Identities. We investigate the identities (1.1)–(1.7) with the assumption that A is a simple finite dimensional anti-commutative algebra over a field F of characteristic zero and with multiplication denoted by xy. From (1.1),

$$\begin{aligned} 0 = [x, x, y] &= \alpha\, xx\cdot y + \beta\, xy\cdot x + \gamma\, yx\cdot x \\ &= (\beta - \gamma)\, xy\cdot x . \end{aligned}$$

Thus $\gamma = \beta$ or $xy\cdot x = 0$ for all $x, y \in A$. Suppose this last equation holds then we must have

$$xy\cdot z + x\cdot yz = 0 \qquad x, y, z \in A .$$

Now let $0 \neq b \in A$ and consider $B = bA$. B is an ideal of A; for if $y, z \in A$, then $by\cdot z = -b\cdot yz$. Thus $B = 0$ or $B = A$. $B = 0$ implies bF is an ideal of A and therefore $A = bF$. This implies $A^2 = 0$, a contradiction to the simplicity of A. But if R_x or $R(x)$ denotes the mapping $a \to ax$, then $B =$ A implies R_b is surjective and since A is finite dimensional, R_b is injective. This contradicts $bR_b = 0$ with $b \neq 0$. Thus we must have $\gamma = \beta$.

From (1.2), $xy = -yx$ which is just the statement that A is anti-commutative.

From (1.3), $\gamma = \beta$ and setting $J(x, y, z) = xy\cdot z + yz\cdot x + zx\cdot y$ we have

$$\begin{aligned} J(x, y, z) &\doteq \alpha(xy\cdot z + yz\cdot x + zx\cdot y) \\ &\quad + \beta(yz\cdot x + zx\cdot y + xy\cdot z) \\ &\quad + \gamma(zx\cdot y + xy\cdot z + yz\cdot x) \\ &= (\alpha + 2\beta)\, J(x, y, z) . \end{aligned}$$

Thus if A is not a Lie algebra we have

$$1 = \alpha + 2\beta . \tag{2.1}$$

In case A is a Lie algebra, then $\beta = \gamma$ and $D(x, y) = (\beta - \alpha)\, R(xy)$. The remaining identities give no more information and therefore the first part of the main theorem is proved by setting $\beta - \alpha = -1$. Henceforth A is assumed to be a non-Lie algebra.

From (1.4) we obtain

$$\begin{aligned} 0 &= \alpha[(wx\cdot y)z + (xy\cdot w)z + (yw\cdot x)z] \\ &\quad + \beta[yz\cdot wx + wz\cdot xy + xz\cdot yw \\ &\quad + (z\cdot wx)y + (z\cdot xy)w + (z\cdot yw)x] \\ &= \alpha\, J(w, x, y)z \\ &\quad + \beta\, [yz\cdot wx + (z\cdot wx)y + (wx\cdot y)z - (wx\cdot y)z \\ &\quad + wz\cdot xy + (z\cdot xy)w + (xy\cdot w)z - (xy\cdot w)z \\ &\quad + xz\cdot yw + (z\cdot yw)x + (yw\cdot x)z - (yw\cdot x)z] \end{aligned}$$

$$= \alpha J(w, x, y)z - \beta J(w, x, y)z + \beta [J(y, z, wx) + J(w, z, xy) + J(x, z, yw)] .$$

Thus

$$(2.2) \quad (\alpha - \beta) J(w, x, y)z = \beta [J(z, w, xy) + J(z, x, yw) + J(z, y, wx)] .$$

From (2.2) we see that $\beta \neq 0$. For suppose $\beta = 0$, then from (2.1), $\alpha = 1$ and from (2.2), $J(w, x, y)z = 0$. Now if $J(A, A, A)$ denotes the subspace spanned by the elements $J(w, x, y)$ for all $w, x, y \in A$ we see that $J(A, A, A)$ is a nonzero ideal of A and so $A = J(A, A, A)$. But $J(w, x, y)z = 0$ implies $A^2 = 0$, a contradiction.

We rewrite (1.5) in terms of the derivations $D(u, v)$ by operating on y in (1.5) to obtain

$$(2.3) \qquad [D(w, x), D(v, u)] = D(wD(v, u), x) + D(w, xD(v, u)) ,$$

where for linear transformations S and T, $[S, T] = ST - TS$. We shall not use this identity since a straightforward computation, as suggested by the referee, shows (1.6) and (1.7) imply (1.5).

Next using $[x, y, z] = (\alpha - \beta)xy \cdot z + \beta J(x, y, z)$ we obtain from (1.6),

$$\begin{aligned} 0 = &(\alpha - \beta) wx \cdot yz + \beta J(w, x, yz) \\ &+ (\alpha - \beta) z(wx \cdot y) + \beta z J(w, x, y) \\ &+ (\alpha - \beta) y(xw \cdot z) + \beta y J(x, w, z) \end{aligned}$$

and therefore

$$(2.4) \qquad \begin{aligned} &(\alpha - \beta) J(wx, y, z) - \beta J(yz, w, x) \\ &= \beta [J(w, x, z)y - J(w, x, y)z] . \end{aligned}$$

3. Proof of main theorem. We shall investigate first the restrictions imposed by (2.2) and (2.4). In (2.4) set $w = y$ and $z = x$ to obtain

$$\begin{aligned} (\alpha - 2\beta) J(xy, x, y) &= \beta [J(y, x, x)y - J(y, x, y)x] \\ &= 0 . \end{aligned}$$

Thus we have

CASE I. $\qquad J(xy, x, y) = 0, \qquad$ or

CASE II. $\qquad \alpha = 2\beta$.

We shall show that in Case I, A must be a non-Lie Malcev algebra (since we are assuming A is not a Lie algebra) and that Case II yields an anti-commutative algebra satisfying $J(x, y, z)w = J(w, x, yz) + J(w, y, zx) + J(w, z, xy)$.

For Case I we linearize $J(xy, x, y) = 0$ to obtain

$$J(wx, y, z) + J(yz, w, x) = J(wy, z, x) + J(zx, w, y)\,, \tag{3.1}$$

$$\begin{aligned} wJ(x, y, z) - xJ(y, z, w) + yJ(z, w, x) - zJ(w, x, y) \\ = 3\,[J(wx, y, z) + J(yz, w, x)]\,. \end{aligned} \tag{3.2}$$

Using (3.2) and (2.2) we have

$$\begin{aligned} &3\,(\alpha - \beta)\,[J(wx, y, z) + J(yz, w, x)] \\ &= (\alpha - \beta)\,[-J(x, y, z)w + J(y, z, w)x - J(z, w, x)y + J(w, x, y)z] \\ &= \beta[-J(w, x, yz) - J(w, y, zx) - J(w, z, xy) \\ &\quad + J(x, y, zw) + J(x, z, wy) + J(x, w, yz) \\ &\quad - J(y, z, wx) - J(y, w, xz) - J(y, x, zw) \\ &\quad + J(z, w, xy) + J(z, x, yw) + J(z, y, wx)] \\ &= \beta[2J(xy, z, w) + 2J(zw, x, y) \\ &\quad - 2J(wx, y, z) - 2J(yz, w, x) \\ &\quad + 2J(wy, x, z) + 2J(xz, w, y)] \\ &= -\,6\beta\,[J(wx, y, z) + J(yz, w, x)]\,, \qquad \text{using (3.1).} \end{aligned}$$

Thus $\alpha - \beta = -\,2\beta \neq 0$ or

$$J(wx, y, z) + J(yz, w, x) = 0. \tag{3.3}$$

Now in case $\alpha - \beta = -\,2\beta$ we have from (2.2),

$$2zJ(w, x, y) = J(z, w, xy) + J(z, x, yw) + J(z, y, wx)$$

and using this identity with (3.1) we have

$$\begin{aligned} 2wJ(w, x, y) &= J(w, x, yw) + J(w, y, wx) \\ &= 2J(w, x, yw)\,. \end{aligned}$$

Thus A is a Malcev algebra. Using the results of [3] we may assume $\alpha = -1$, $\beta = \gamma = 1$. Also from [3] the derivation algebra equals $I(A)$ and $I(A)$ is completely reducible in A.

So we next assume A satisfies (3.3), then using (2.4) we obtain

$$\alpha\, J(wx, y, z) = \beta[J(w, x, z)y - J(w, x, y)z]\,,$$

and therefore

$$\begin{aligned} \alpha(\alpha - \beta)\; J(wx, y, z) &= \beta(\alpha - \beta)\;[J(w, x, z)y - J(w, x, y)z] \\ &= \beta^2[J(y, w, xz) + J(y, x, zw) \\ &\quad + J(y, z, wx) - J(z, w, xy) \\ &\quad - J(z, x, yw) - J(z, y, wx)], \qquad \text{using (2.2)} \\ &= 2\beta^2\; J(wx, y, z), \qquad \text{using (3.3).} \end{aligned}$$

Thus
$$J(wx, y, z) = 0, \qquad \text{or}$$
$$\alpha(\alpha - \beta) = 2\beta^2.$$

In the first case, A is a Lie algebra which is a contradiction and in the second case $\alpha = -\beta$ or 2β. Thus as subcases we have

Case A. $\alpha = -\beta$ and therefore $\alpha = -1$, $\beta = \gamma = 1$;

Case B. $\alpha = 2\beta$ and therefore $\alpha = 1/2$, $B = \gamma = 1/4$.

First consider Case A, then from (2.2) and (3.3) we have

$$\begin{aligned} 2wJ(w, x, y) &= J(w, w, xy) + J(w, x, yw) + J(w, y, wx) \\ &= 2J(w, x, yw) \end{aligned}$$

and therefore A is a Malcev algebra. We shall next show that this Malcev algebra of Case A actually does not exist. First for any anti-commutative algebra A define the linear transformation $\Delta(x, y)$ by

$$z\ \Delta(x, y) = J(x, y, z)$$

and let $\Delta(A, A)$ be the linear space of transformations spanned by these $\Delta(x, y)$'s for all $x, y \in A$. Using (3.3) we have

$$\begin{aligned} 0 &= J(wx, y, z) + J(w, x, yz) \\ &= w(R_x \Delta(y, z) + \Delta(x, yz)) \end{aligned}$$

and therefore

$$R_x \Delta(y, z) = -\Delta(x, yz) \in \Delta(A, A)\ . \tag{3.4}$$

From identities (2.32) and (2.34) of [2] we also have

$$2\ \Delta(y, z)R_x = 2R_x \Delta(y, z) - 2R(J(x, y, z)) - 4\ \Delta(yz, x) \in \Delta(A, A)\ ,$$

using also the preceding identity. Thus we see from these identities that $\Delta(A, A)$ is an ideal in the transformation algebra $T(A)$ which is generated by $R(A) = \{R_x : x \in A\}$. But since A is simple, $T(A)$ is a simple associative algebra and therefore $\Delta(A, A) = 0$ or $T(A) = \Delta(A, A)$. But $\Delta(A, A) = 0$ implies A is a Lie algebra and therefore $\Delta(A, A)$ is a simple associative algebra. But we shall next show that $\Delta(x, y)^2 = 0$ and therefore conclude that $\Delta(A, A)$ have a basis consisting of nilpotent elements. Thus $\Delta(A, A)$ must be a nilpotent associative algebra, a contradiction to the simplicity of $\Delta(A, A)$. Hence Case A does not exist. So to show $\Delta(x, y)^2 = 0$ we have

$$\begin{aligned} -2z\ \Delta(x, y)\ R(xy) &= 2xy \cdot J(x, y, z) \\ &= J(xy, x, yz) + J(xy, y, zx) \\ &\quad + J(xy, z, xy) \end{aligned}$$

$$\begin{aligned} &= -J(x\cdot yz, x, y) - J(y\cdot zx, x, y) \\ &\quad - J(z\cdot xy, x, y), \quad \text{using (3.3)} \\ &= J(J(x, y, z), x, y) \\ &= z\,\Delta(x, y)^2 , \end{aligned}$$

that is, $\Delta(x, y)^2 = -2\,\Delta(x, y)\,R(xy)$. But from identity (2.33) of [2], $\Delta(x, y)^2 = -3\,\Delta(x, y)\,R(xy)$ and therefore $\Delta(x, y)^2 = 0$.

Next we derive more identities for Case B and use methods similar to those used in Case A to show Case B does not exist either. Using the notation from Case A we obtain again from (3.3) the identity (3.4). Also with $\alpha = 1/2$, $\beta = 1/4$ in (2.2) we obtain

$$J(w, x, y)z = J(z, w, xy) + J(z, x, yw) + J(z, y, wx) . \tag{3.5}$$

From (3.5) and (3.4) we have

$$\Delta(x, y)R_z = -\Delta(z, xy) - R_y\,\Delta(z, x) + R_x\,\Delta(z, y) \in \Delta(A, A) .$$

Thus as in Case A we see $\Delta(A, A) = T(A)$ is a simple associative algebra. Next from (3.5) we also have

$$R(J(w, x, y)) = \Delta(wx, y) + \Delta(xy, w) + \Delta(yw, x) , \tag{3.6}$$

and using (3.4) we obtain

$$\begin{aligned} \Delta(x, y)^2 &= [R_x, R_y]\,\Delta(x, y) - R(xy)\,\Delta(x, y) \\ &= R_xR_y\,\Delta(x, y) - R_yR_x\,\Delta(x, y) + \Delta(xy, xy) \\ &= \Delta(x, y\cdot xy) - \Delta(y, x\cdot xy) + \Delta(xy, xy) \\ &= -\Delta(xy\cdot x, y) - \Delta(xy, xy) - \Delta(y\cdot xy, x) \\ &= -R(J(xy, x, y)) \\ &= 0 , \end{aligned}$$

where the last equality follows from (3.3). Thus as in Case A, we conclude that Case B does not exist, this completes Case I.

Next consider Case II where $\alpha = 2\beta \neq 0$. From (2.2) and (2.4) we see that A satisfies (3.5) and

$$\begin{aligned} &J(wx, y, z) - J(yz, w, x) \\ &\quad = J(w, x, z)y - J(w, x, y)z . \end{aligned} \tag{3.7}$$

Next we rewrite (3.5) and (3.7) in terms of right multiplications to obtain (3.6) and

$$\Delta(z, wx) - R_z\,\Delta(w, x) = -R(J(w, x, z)) - \Delta(w, x)R_z ,$$

by operation on y in (3.7). Using this and (3.6) we have

$$\Delta(w, x)R_z - R_z\,\Delta(w, x) = -\Delta(wx, z) - \Delta(xz, w)$$

$$- \Delta(zw, x) - \Delta(z, wx)$$
$$= \Delta(zx, w) - \Delta(zw, x) ;$$

$$[\Delta(w, x), R_z] = \Delta(zx, w) - \Delta(zw, x) . \tag{3.8}$$

Now using (3.8) and the Jacobi identity we have

$$\begin{aligned}[\Delta(w, x), \ \Delta(u, v)] &= [\Delta(w, x), \ [R_u, R_v]] - [\Delta(w, x), R(uv)] \\ &= - [[R_u, \Delta(w, x)], R_v] - [R_u, [R_v, \Delta(w, x)] \\ &\quad - [\Delta(w, x), R(uv)] \in \Delta(A, A) .\end{aligned}$$

Thus defining the *Lie transformation algebra of A*, denoted by $L(A)$, to be the Lie algebra generated by $R(A) = \{R_x : x \in A\}$ [5], the above calculations prove

LEMMA 3.9. *$\Delta(A, A)$ is a Lie algebra and $L(A) = \Delta(A, A)$.*

For clearly $\Delta(A, A) \subset L(A) = \Sigma_i M_i$ where $M_1 = R(A)$ and $M_i = [M_{i-1}, M_1]$ for $i > 1$. But also since $A = J(A, A, A)$ we have from (3.6) that $M_1 \subset \Delta(A, A)$ and since $\Delta(A, A)$ is a Lie algebra, $M_i \subset \Delta(A, A)$; thus $L(A) \subset \Delta(A, A)$.

Next we consider the center C of $L(A) = \Delta(A, A)$. Since A is simple, $\Delta(A, A)$ is an irreducible Lie algebra over a field of characteristic zero and therefore from [1, Th. 1], $\Delta(A, A) = C \oplus \Delta'$ where Δ' is a semi-simple Lie algebra and $C = \{\Delta \in \Delta(A, A) : [\Delta, T] = 0$ for all $T \in \Delta(A, A)\}$.

LEMMA 3.10. *$C = 0$ and therefore $\Delta(A, A)$ is a semi-simple Lie algebra.*

Proof. Let $\Delta = \Sigma_i \Delta(x_i, y_i) \in C$, then for any $u, v \in A$ we have

$$\begin{aligned}(uv)\Delta &= - vR_u \Delta \\ &= - v\Delta R_u, \text{ since } R_u \in \Delta(A, A) \\ &= - v\Delta \cdot u \\ &= u \cdot v\Delta .\end{aligned}$$

But next we have

$$\begin{aligned}(vu)\Delta &= \Sigma_i \, vu\Delta \, (x_i, y_i) \\ &= \Sigma_i \, J(vu, x_i, y_i) \\ &= \Sigma_i \, [J(x_i y_i, v, u) - J(x_i, y_i, u)v + J(x_i, y_i, v)u] , \\ &\qquad \text{using (3.7)} \\ &= J(z, v, u) - \Sigma_i u\Delta(x_i, y_i)v + \Sigma_i v\Delta(x_i, y_i)u, \\ &\qquad \text{where } z = \Sigma x_i y_i\end{aligned}$$

$$\begin{aligned} &= J(z, v, u) - u\Delta \cdot v + v\Delta \cdot u \\ &= J(z, v, u) + v \cdot u\Delta - u \cdot v\Delta \\ &= J(z, v, u) + (vu)\Delta - (uv)\Delta \,. \end{aligned}$$

Therefore $(uv)\Delta = J(z, v, u)$ where $z = \Sigma x_i y_i$.

Case 1. $z = \Sigma x_i y_i = 0$. Then $(uv)\Delta = 0$ and since $A = A^2$ we have $\Delta = 0$.

Case 2. $z = \Sigma\, x_i y_i \neq 0$. Then there exists $w \in A$ with $a = zw \neq 0$; otherwise zF would be a nonzero ideal in A. Therefore

$$a\Delta = (zw)\Delta = J(z, w, z) = 0$$

and $K = \ker \Delta \neq 0$. But K is an ideal of A. For if $x \in K$, $y \in A$, then $(yx)\Delta = y \cdot x\Delta = 0$ and therefore $KA \subset K$. Thus $K = A$ which means $\Delta = 0$. From both of these cases we conclude $C = 0$.

Next as for Malcev algebras we have the following

DEFINITION. The set $N = \{n \in A : J(n, A, A) = 0\}$ is called the *J-nucleus of A*.

LEMMA 3.11. *If $a, b \in A$ are such that $J(a, b, A) = 0$, then $ab \in N$.*

Proof. Suppose $J(a, b, u) = 0$ for all $u \in A$, then from (3.7)

$$\begin{aligned} J(ab, y, z) &= J(yz, a, b) + J(a, b, z)y - J(a, b, y)z \\ &= 0. \end{aligned}$$

COROLLARY 3.12. *N is an ideal of A which is a Lie algebra and therefore $N = 0$.*

COROLLARY 3.13. *R_x is a derivation of A if and only if $x \in N = 0$.*

Now let

$$D(x, y) = \frac{1}{4}[R_x, R_y] - \frac{1}{2}R_{xy}$$

i.e.
$$zD(x, y) = [x, y, z] = \frac{1}{2}xy \cdot z + \frac{1}{4}(yz \cdot x + zx \cdot y)\,.$$

Then $D(x, y)$ is an inner derivation, that is, $D(x, y) \in L(A) = \Delta(A, A)$. If $I(A)$ denotes the linear subspace spanned by all such $D(x, y)$'s, then we have

LEMMA 3.14. *The derivation algebra of A equals $I(A)$* and *$I(A)$ is completely reducible in A.*

Proof. First we shall show that any derivation D of A in $\Delta(A, A)$ is actually in $I(A)$. For let $D = R_z + (1/4)\Sigma_i[R_{x_i}, R_{y_i}] \in \Delta(A, A) = L(A)$. Then write $D = R(z + (1/2)\Sigma_i x_iy_i) + \Sigma_i D(x_i, y_i)$. From this equation $R(z + (1/2)\Sigma_i x_iy_i)$ is a derivation and therefore by Corollary 3.13 it equals zero. Thus $D \in I(A)$.

Next we shall show any derivation of A is in $I(A)$. Since D is a derivation $[R_x, D] = R(xD)$ and therefore using the Jacobi identity we obtain $[\Delta(A, A), D] \subset \Delta(A, A)$. Now the map

$$\Delta(A, A) \to \Delta(A, A): X \to [X, D]$$

is a derivation of $\Delta(A, A)$. But since $\Delta(A, A)$ is a semi-simple Lie algebra all derivations are inner and therefore there exists $T \in \Delta(A, A)$ so that

$$[X, D] = [X, T] \text{ for all } X \in \Delta(A, A). \tag{3.15}$$

Now since $\Delta(x, y) = [R_x, R_y] - R(xy) = 4D(x, y) + R(xy)$ we have, using Corollary 3.13,

$$\Delta(A, A) = R(A) \oplus I(A) \tag{3.16}$$

as a vector space sum. Therefore let $T = R_z + D_1$ where $D_1 \in I(A)$ and $z \in A$, then for any $x \in A$,

$$\begin{aligned}[R_x, R_z] &= [R_x, T - D_1] \\ &= [R_x, T] - [R_x, D_1] \\ &= [R_x, D - D_1], \quad \text{using (3.15)} \\ &= R(x\tilde{D}),\end{aligned}$$

where $\tilde{D} = D - D_1$ is a derivation. Therefore

$$\begin{aligned}R(x(\tilde{D} - 2R_z)) &= R(x\tilde{D}) - 2R(xz) \\ &= [R_x, R_z] - 2R(xz) \\ &= 4D(x, z).\end{aligned}$$

Thus $R(x(\tilde{D} - 2R_z))$ is a derivation and using Corollary 3.13 we have

$$x(\tilde{D} - 2R_z) = 0 \qquad \text{for any } x \in A.$$

However this implies $2R_z = \tilde{D}$ is a derivation and therefore $0 = 2R_z = \tilde{D}$. Thus from the definition of $\tilde{D}$ we have $D = D_1 \in I(A)$ so that every derivation of A is in $I(A)$.

The last part of the main theorem is proved in a manner analogous to the proof of Theorem 9 in [1]. We note from (3.16) that the completely reducible Lie algebra $\Delta(A, A)$ is such that the subalgebra $I(A)$ is splittable and has a complementary subspace, namely $R(A)$, which is $I(A)$-invariant (because $[R_x, D] = R(xD)$). Thus from [**1**, Th. 5], $I(A)$ is completely reducible in A and the proof of the main theorem is complete.

References

1. N. Jacobson, *Completely reducible Lie algebras of linear transformations*, Proc. Amer. Math. Soc. **2** (1951), 105-113.
2. A. Sagle, *Malcev algebras*, Trans. Amer. Math. Soc. **101** (1961), 426-458.
3. ——, *On derivations of semi-simple Malcev algebras*, Portugaliae Math. **21** (1962) 107-109.
4. ——, *Simple Malcev algebras over fields of characteristic zero*, Pacific J. Math. **12** (1962), 1057-1078.
5. R. D. Schafer, *Inner derivations of nonassociative algebras*, Bull. Amer. Math. Soc. **55** (1949), 769-776.
6. K. Yamaguti, *On the Lie triple system and its generalization*, J. Sci. Hiroshima Univ., Ser. A, **21** (1958), 155-160.
7. ——, *Note on Malcev algebras*, Kumamoto J. of Sci., Ser. A, **5** (1962), 203-207.

University of California, Los Angeles

PACIFIC JOURNAL OF MATHEMATICS
Vol. 15, No. 1, 1965

A CHARACTERIZATION OF FREE PROJECTIVE PLANES

L. C. SIEBENMANN

A characterization of free projective planes is given that is more symmetrical than the original definition of M. Hall. *It tends to very simple proofs of two fundamental theorems* due to M. Hall and L. I. Kopejkina-one being the result that every subplane of a free plane is free.

In a fundamental paper Marshall Hall defines a free plane to be a projective plane which either is degenerate or is generated as follows from a 'basis configuration,' π_0, consisting of at least two points on a line together with two isolated points. For each pair of points not already joined in π_0 create a distinct line that joins them and add it to π_0. In the resulting configuration, π_1, consider pairs of lines that do not intersect, and for each create a distinct intersection point and add it to π_1, thus forming π_2. Continuing, construct $\pi_3, \pi_4, \pi_5, \pi_6$, etc. adding alternately lines and points as indicated above. Then $\pi = \bigcup_n \pi_n$ (with the obvious incidence relation) is a projective plane. It is by definition a free plane. Hall proved that a free plane contains no confined configuration, that is, no finite configuration that, like the Desargues configuration, has $\geqq 3$ points on each line and $\geqq 3$ lines through each point. Further, using a complicated argument, he showed that, if a finitely generated plane contains no confined configuration, it is free. It follows that any finitely generated subplane of a free plane is free.

L. I. Kopejkina [2] proved, shortly after, that an arbitrary subplane of a free plane is free. (Of interest is the analogy with free groups.) An exposition of Kopejkina's theorem appears in [3].

Because it suggests a more symmetrical definition of free plane that leads to very direct proofs of the above theorems, we introduce the notion of an extension process.

DEFINITION 1. An *extension process* is a well ordered nested sequence of partial planes $\pi_0 \subset \pi_1 \subset \cdots \subset \pi_n \subset \cdots$ (the subscripts $0, 1, \cdots, n, \cdots$ belonging to a well ordered set) such that if a point p and a line l appear in a term π_n, $n > 0$, and in no earlier term then p is not incident with l—in other words, the *new* elements in π_n may be incident in π_n with elements appearing in earlier terms, but have no incidences among themselves. From this point we adopt as definition the characterization of free plane we aim to justify.

Received February 13, 1964.

DEFINITION 2. A *free plane* is a (possibly degenerate) projective plane, π, for which there exists an extension process $\pi_0 \subset \pi_1 \subset \cdots \subset \pi_n \subset \cdots$ with $\pi_0 = \varnothing$ (the empty plane) and $\pi = \bigcup_n \pi_n$ such that every new point [or line] in a term π_n is incident in π_n with at most two lines [points].

Immediately we can prove

THEOREM I (Kopejkina). *Every subplane of a free plane is free.*

Proof. If π is a free plane and $\pi_0 \subset \pi_1 \subset \cdots \subset \pi_n \subset \cdots$ is an extension process as above for π, then, given any subplane π', the sequence $\pi_0 \cap \pi' \subset \pi_1 \cap \pi' \subset \cdots \subset \pi_n \cap \pi' \subset \cdots$ is visibly such an extension process for π'. So π' is free. ∎

Some definitions and notations are collected in § 2. In § 3 the result of Hall is proved. Our definition of a free plane is apparently broader than M. Hall's. In § 4 we prove that the definitions are equivalent.

2. A set of elements consisting of points and lines, together with a relation of incidence between points and lines is said to form a *partial plane* (or *configuration*) if every two distinct lines [points] are together incident with at most one point [respectively line], (which when it exists we call their *join*). If every two distinct lines [points] are together incident with exactly one point [respectively line] the system becomes a (projective) *plane*. A plane is said to be *non-degenerate* if it contains 4 points no 3 of which are incident with the same line, and otherwise is said to be *degenerate*.

If ρ and σ are subpartial planes of the partial plane π, then $\rho + \sigma$ (or $\rho \cup \sigma$), $\rho \cap \sigma$, and $\rho - \sigma$ are subpartial planes of π defined in the obvious way.

A configuration ρ in a plane π is said to *generate* the least subplane containing ρ. This subplane is denoted by $[\rho]_\pi$ (or $[\rho]$) and is called the *completion* of ρ in π. The plane π is *finitely generated* if, for some finite configuration $\rho \subset \pi$, $[\rho] = \pi$.

An extension process, $\mathscr{E}$, is regularly presented in the form $\mathscr{E} = \{\pi_n; n \in N\}$ where $N = \{0, 1, \cdots, n, \cdots\}$ is a well ordered set. π_{n-} will denote $\bigcup_{m<n} \pi_m$. $\mathscr{E}$ is said to *act on* π_0 and have the *result* $\mathscr{E}(\pi_0) = \bigcup_n \pi_n$. It should be pointed out that the partial plane $\mathscr{E}(\pi_0)$ need not be a full projective plane. The $\mathscr{E}$*-stage* (or simply *stage*) of an element $x \in \mathscr{E}(\pi_0)$ is the least n such that $x \in \pi_n$. If $n > 0$, x is said to *appear at* $\mathscr{E}$*-stage* n; it is incident in π_n with certain elements called its $\mathscr{E}$*-bearers* (or *bearers*), and these must all lie in $\pi_{n-} = \bigcup_{m<n} \pi_m$. Elements in π_0 (by convention) have no bearers. Observe that if $x \notin \pi_0$

is incident with y, and y is *not* a bearer of x, then x must be a bearer of y, for they cannot both appear at the same stage.

If ρ is a subpartial plane of $\pi = \mathscr{E}(\pi_0)$, there is a naturally defined *restriction* of $\mathscr{E}$ to ρ, denoted $\mathscr{E} \cap \rho$, viz. $\{\pi_n \cap \rho; n \in N\}$. This is apparently a process that acts on $\pi_0 \cap \rho$ with result ρ. Also there is a naturally defined *saturation* of $\mathscr{E}$ by ρ, denoted $\mathscr{E} + \rho$, viz. $\{\pi_n + \rho; n \in N\}$. It apparently acts on $\pi_0 + \rho$ with result π. We make two simple but important observations.

(1) The bearers in $\mathscr{E} \cap \rho$ of an element $x \in \rho$ are just those $\mathscr{E}$-bearers that lie in ρ.

(2) The bearers in $\mathscr{E} + \rho$ of an element $x \notin \pi_0 + \rho$ include its $\mathscr{E}$-bearers and in addition any elements of ρ incident with x in π.

If $\mathscr{E}$ and $\mathscr{F}$ are extension processes, $\mathscr{E}$ acting on π_0 and $\mathscr{F}$ acting on $\mathscr{E}(\pi_0)$, then there is a naturally defined composition of $\mathscr{F}$ with $\mathscr{E}$ denoted $\mathscr{F} \circ \mathscr{E}$.

We call a process, $\mathscr{E}$, (a) *bound*; (b) *free*; (c) *hyperfree* if for all $n > 0$ every new element of π_n has (a) $\geqq 2$; (b) 2; (c) $\leqq 2$ bearers. Of course an extension process need not fall into any of these categories. A *free plane* is by definition a plane which is the result of a hyperfree process acting on the empty plane, $\varnothing$.

A bound process $\mathscr{E}$, whose result, $\mathscr{E}(\pi_0)$, is a full projective plane, is called a *completion process* for π_0. If ρ is a configuration in a plane π, there is a canonical completion process $\mathscr{E} = \{\rho_n; n \in J^+\}$, indexed on the integers $\geqq 0$, that acts on ρ with result $[\rho]_\pi$. In fact $\rho_0 = \rho$ and ρ_n is defined inductively as the subpartial plane of π whose elements are those of ρ_{n-1} together with all points [lines] of π that are joins of lines [or points] of ρ_{n-1} resp. as n is even or odd.

3. Now we prove Hall's result. Suppose π is a finite partial plane having P points, L lines, and I incidences between the points and lines. The *rank* of π as defined by Hall is

$$r(\pi) = 2(P + L) - I .$$

LEMMA 1. *Any finite partial plane, π, which contains no confined configurations is the result of a hyperfree process.*

Proof. Set $\pi_m = \pi$ where m is the number of points and lines in π. Since π_m is not confined, there exists some element $x_m \in \pi_m$ which is incident with $\leqq 2$ elements in π_m. Define $\pi_{m-1} \subset \pi_m$ to be π_m less the element x_m. Since π_{m-1} is not confined, the process may be repeated, and after exactly m steps we obtain $\pi_0 = \varnothing$. Then $\mathscr{F} = \{\pi_i; i = 0, 1, \cdots, m\}$ is hyperfree and $\mathscr{F}(\varnothing) = \pi$.

COROLLARY. *The rank of a finite partial plane containing no confined configuration is nonnegative.*

Proof. In fact $r(\pi) \geqq r(\pi_m) \geqq \cdots \geqq r(\pi_0) = 0$. ▮

THEOREM II (Hall). *A free plane contains no confined configuration. A finitely generated plane that contains no confined configuration is free.*

REMARK. Kopejkina [2] constructed a plane (not finitely generated) that contains no confined configuration, but is not free.

Proof. Suppose first that π is a free plane and $\mathscr{F} = \{\pi_n; n \in N\}$ is a hyperfree process such that $\mathscr{F}(\varnothing) = \pi$. If ρ is any finite configuration in π, there is an element $x \in \rho$ having maximal $\mathscr{F}$-stage, m. Since x is incident with at most two elements of π_m, it obviously cannot be incident with at least three elements of $\rho \subset \pi_m$. Thus ρ cannot be a confined configuration.

The proof of the second assertion depends essentially on our definition of free plane. Suppose that the plane π is generated by a finite configuration π_0 and that π contains no confined configuration. Let $\mathscr{C} = \{\pi_n; n \in J^+\}$ be the canonical completion process for π_0. Observe that each partial plane π_n is finite and $\mathscr{C}(\pi_0) = \pi$. Since $\mathscr{C}$ is bound $r(\pi_0) \geqq r(\pi_1) \geqq \cdots \geqq r(\pi_n)$. But by the corollary above $r(\pi_n) \geqq 0$ for all n. Hence for some integer m the minimal rank be attained, and thereafter $r(\pi_n)$ will have this minimal value. But this means that in the bound process $\mathscr{S} = \{\pi_n; n \geqq m\}$ every element has exactly 2 bearers i.e., $\mathscr{S}$ is free. Now, by Lemma 1, $\pi_m = \mathscr{T}(\pi_0)$ where $\mathscr{T}$ is hyperfree. So composing $\mathscr{S}$ with $\mathscr{T}$ we obtain a hyperfree process with $\mathscr{S} \circ \mathscr{T}(\varnothing) = \pi$. ▮

4. This last section is devoted to proving that the adopted definition of free plane is equivalent to Hall's.

LEMMA 2. *Suppose $\mathscr{F}$ is hyperfree and $\pi = \mathscr{F}(\pi_0)$ is a plane. Then*

(1) *$\mathscr{F}_1 = \mathscr{F} \cap [\pi_0]$ is a free completion process for π_0 in π; in fact, for $x \in [\pi_0]$, the $\mathscr{F}$-bearers lie in $[\pi_0]$ and coincide with those in any completion process for π_0 in π.*

(2) *$\mathscr{F}_2 = \mathscr{F} + [\pi_0]$ is still hyperfree; in fact, for $x \notin [\pi_0]$, the $\mathscr{F}_2$-bearers coincide with the $\mathscr{F}$-bearers.*

Proof. (1) Let $\mathscr{C}$ be any completion process for π_0 in π. If the first assertion is false, let x be an element of least $\mathscr{C}$-stage, m, for which the $\mathscr{E}$-bearers do not conincide with the $\mathscr{F}$-bearers. Clearly $m > 0$. Then at least one of the $\geqq 2$ $\mathscr{E}$-bearers of x, say y, must fail to be an $\mathscr{F}$-bearer. Since $x \notin \pi_0$, y must have x as an $\mathscr{F}$-bearer. But y doesn't have x as an $\mathscr{E}$-bearer and y has $\mathscr{E}$-stage $< m$. This is a contradiction.

(2) Supposing the second assertion false, we have some $x \notin [\pi_0]$ with an $\mathscr{F}_2$-bearer y that is not an $\mathscr{F}$-bearer and (hence) lies in $[\pi_0] - \pi_0$. But x is incident with $y \notin \pi_0$; so $x \notin [\pi_0]$ must be an $\mathscr{F}$-bearer of $y \in [\pi_0]$ in contradiction to (1). ∎

REMARK. This lemma has a useful generalization in which π_0 is replaced in (1) and (2) by a partial plane $\rho \subset \pi_0$ that is 'complete' in π_0 (see [1, 4.3]). Then, in the proof of (1), the possibility that $x \in \pi_0$ must be eliminated.

DEFINITION 3. A *free completion* of a partial plane, π_0, is a plane π which is the result of a free process acting on π_0.

Again this seems less restrictive than Hall's definition, but

THEOREM III. *Any two free completions of a partial plane π_0 are related by a unique isomorphism that fixes π_0.*

Proof. Let $\mathscr{F}$ be a free completion process for π_0, and let $\mathscr{E}$ be the canonical completion process for π_0 in $\mathscr{F}(\pi_0)$. Clearly $\mathscr{E}(\pi_0) = [\pi_0] = \mathscr{F}(\pi_0)$; and according to assertion (1) of the above lemma, $\mathscr{E}$ is free. The theorem now follows from the fact that, if $\mathscr{E}$ and $\mathscr{E}'$ are two canonical free completion processes for π_0, there is a unique isomorphism of $\mathscr{E}(\pi_0)$ onto $\mathscr{E}'(\pi_0)$ that fixes π_0. ∎

In a free extension process $\{\pi_0, \pi_1\}$ of just two terms we say that π_1 is derived from π_0 by a *free addition*, and π_0 from π_1 by a *free subtraction*. Two partial planes are *free equivalent* if the one can be derived from the other by a finite sequence of free additions and subtractions. Free equivalent partial planes evidently have isomorphic free completions.

Recall that a basis configuration consists of a number of points on a 'base' line and two isolated points.

THEOREM IV. *Every nondegenerate free plane contains a basis configuration of which it is a free completion.*

Proof. Suppose $\pi = \mathscr{F}(\varnothing)$ is a free plane, where $\mathscr{F} = \{\pi_n; n \in N\}$ is a hyperfree process. We may assume without loss of generality that at each stage one element and no more is added, i.e., $\pi_n = \pi_{n-} + x_n$, where x_n is an element that has 0, 1, or 2 bearers.

Applying Lemma 2 to $\mathscr{F}$ for stages $\geqq n$ we find that $\mathscr{F}' = \mathscr{F} + [\pi_{n-}]$ is hyperfree; applying the same lemma to $\mathscr{F}'$ for stages $>n$ we find that $\mathscr{F}' \cap [\pi_n]$ free completes $[\pi_{n-}] + x_n$ to $[\pi_n]$.

Now let a be the first stage such that $[\pi_a]$ is nondegenerate. Then $[\pi_{a-}]$ must be a degenerate plane, and, as we have observed, $[\pi_a]$ is a free completion of $[\pi_{a-}] + x_a$. By an inspection of the various cases one shows that $[\pi_{a-}] + x_a$ is in every case free equivalent to some basis configuration π_0^α. This implies that $[\pi_a]$ is a free completion of π_0^α.

For $n > a$, one readily shows (cf. [1] or [3]) that $[\pi_{n-}] + x_n$ is free equivalent to $[\pi_{n-}]$ with a set μ_n of points adjoined to the base line, l, of π_0^α. (The set μ_n consists of 2 or 1 points if x_n is incident with 0 or 1 elements in $[\pi_{n-}]$, and of course μ_n is empty when $x_n \in [\pi_{n-}]$.) Thus $[\pi_n]$ is a free completion of $[\pi_{n-}] + \mu_n$.

We will show that π is a free completion of the basis configuration $\pi_0^\beta = \pi_0^\alpha + \bigcup_{n>a} \mu_n$. Let $\{([\pi_{n-}] + \mu_n)^k; k \in J^+\}$ be the canonical process that free completes $[\pi_{n-}] + \mu_n$ to π_n, and form the composition of all these processes to obtain $\mathscr{S} = \{([\pi_{n-}] + \mu_n)^k; (n, k) \in N \times J^+, n \geqq a\}$ where $([\pi_{a-}] + \mu_a)$ is to be read as π_0^α, and $N \times J^+$ is well ordered lexicographically. This is a hyperfree process acting on π_0^α with result π. The saturation $\bar{\mathscr{S}} = \mathscr{S} + \pi_0^\beta$ is the desired free completion process.

Clearly $\bar{\mathscr{S}}(\pi_0^\beta) = \pi$; and $\bar{\mathscr{S}}$ is a bound process since all elements with <2 $\mathscr{S}$-bearers lie in π_0^β. To show that $\bar{\mathscr{S}}$ is actually free observe that every new element, x, of $\bar{\mathscr{S}}$ appeared in $\mathscr{S}$ with exactly 2 bearers. If x has an extra bearer, y, in $\bar{\mathscr{S}}$, then $y \in \mu_n$ where x has $\bar{\mathscr{S}}$-stage $(m, k) \in N \times J^+$ and $n > m$. But y is incident with both x and base line l of π_0^α, i.e., y is the join of x and l. Then $y \in [\pi_m]$ in contradiction to $y \in \mu_n$. So $\bar{\mathscr{S}}$ must be free. ▮

This completes the proof that the definition of free plane we have proposed coincides with M. Hall's definition.

References

1. M. Hall, *Projective planes*, Trans. Amer. Math. Soc. **54** (1943), 229–277.

2. L. I. Kopejkina, *Decomposition of projective planes*, Bull. Acad. Sci. USSR, Ser. Math. Izvestia Akad. Nauk, SSSR), **9** (1945), 495–526.

3. L. A. Skornyakov, *Projective planes*, Trans. Amer. Math. Soc. Series 1, **1** (1962), 51–107.

PACIFIC JOURNAL OF MATHEMATICS
Vol. 15, No. 1, 1965

SIMPLE AREAS

EDWARD SILVERMAN

Let $\lambda \geqq 1, E = E^N$ and g be continuous on $E \times E \times E$ with $g(a, \cdot, \cdot)$ convex, $g(a, kb, kc) = k^2 g(a, b, c)$ for all real k and $(b^2 + c^2)/\lambda \leqq g(a, b, c) \leqq \lambda(b^2 + c^2)$ for all $a, b, c \in E$ where $b^2 = ||b||^2$. If $f(a, d \wedge e) = \min_{b \wedge c = d \wedge e} g(a, b, c)$ then f is a permissible integrand for the two-dimensional parametric variational problem.

Let γ be a simple closed curve in E, B be the closed unit circle in the plane, C be the collection of functions x continuous on B into E for which $x \mid \partial B \in \gamma$ and $D = \{x \in C \mid x$ is a D-map$\}$. Suppose that D is not empty. It was shown in 'A problem of least area', [7], that the problem of minimizing $I(f)$ over D is equivalent to minimizing $I(g)$ over D where $I(f, \mathrm{x}) = \iint f(x, p \wedge q), I(g, x) = \iint g(x, p, q),\ p = x_u,\ q = x_v$ and both integrals are taken over B. The minimizing solution of $I(g)$ is known to have differentiability properties corresponding to g, and this solution also minimizes $I(f)$.

The function f is simple, that is, for each $a \in E$, each supporting linear functional to $f(a, \cdot)$ is simple. If $N = 3$, then, of course, each parametric integrand is simple. In this paper we show that for each simple parametric integrand F there exists G, satisfying the conditions imposed upon g, such that F is obtained from G as f was obtained from g.

In [7] we showed that the two-dimensional parametric problem in the calculus of variations considered by [1, 2, 4, 5, 6] could be reduced to a nonparametric problem provided the parametric integrand f was properly related to a suitable nonparametric integrand $g, f = Ag$. When this occured, not only the existence of the minimizing solution x was given by the nonparametric theory [3] but also its smoothness, if g was smooth. Furthermore, we saw that Ag was simple for each g, that is, each supporting linear functional of Ag was simple. We shall show here that whenever f is simple then there exists g such that $f = Ag$.

Let $E = E^N$. If $a \in E$ or $a \in E^*$ let $a^2 = ||a||^2$. Let $T_1 = E \wedge E$ with norm N_1, thus $N_1(a \wedge b)$ is the area of the parallelogram spanned by a and b, and let $T_2 = E \times E$. We define N_2 on T_2 by $N_2(a, b) = (a^2 + b^2)/2$. Let T^* be the set of all simple linear functionals over T_1 which have norm one. Hence, if $\zeta \in T^*$, there exist ξ and η in E^* such that $\zeta = \xi \wedge \eta$ with $\xi^2 = \eta^2 = 1$ and $\xi \cdot \eta = 0$. We frequently

Received February 13, 1964. (This research was supported in part by National Science Foundation Grant No. G.P. 634).

write ξa for $\xi(a)$.

If φ is defined on $P \times Q$ then φ_p is defined on Q by $\varphi_p(q) = \varphi(p, q)$ for all $p \in P$ and $q \in Q$.

Let $\mathscr{A}$ be the set of all continuous real-valued functions f on $E \times T_1$ for which there exists $\lambda = \lambda(f) \geqq 1$ with $N_1/\lambda \leqq f_a \leqq \lambda N_1$ and such that f_a is convex and positively homogeneous of degree one for each $a \in E$. Let $\mathscr{D}_0$ be the set of all continuous real-valued functions g on $E \times T_2$ for which there exists $\lambda \geqq 1$ with $N_2/\lambda \leqq g_a \leqq \lambda N_2$ and such that g_a is convex and homogeneous of degree two for each $a \in E$. For our purposes, $\mathscr{D}_0$ gives nothing more than $\mathscr{D} = \{h \in \mathscr{D}_0 \mid$ there exists $g \in \mathscr{D}_0$ such that $h(a, b, c) = \max_\theta g(a, b \cos\theta - c \sin\theta, b \sin\theta + c\cos\theta)\}$.

If $g \in \mathscr{D}$ then let $Ag(a, b \wedge c) = \min_{d \wedge e = b \wedge c} g(a, b, c)$ and

$$Ag(a, \alpha) = \inf\left\{\sum_{i=1}^{k} Ag(a, b_i \wedge c_i) \,\middle|\, \sum_{i=1}^{k} b_i \wedge c_i = \alpha\right\}$$

for all $\alpha \in T_1$. We saw in [7] that $Ag \in \mathscr{A}$ and that Ag is simple. Evidently $Ag(a, b \wedge c) = \min_{r \neq 0} g(a, rb, sb + r^{-1}c)$.

If $g \in \mathscr{D}$ then $2g_a^{1/2}$ is convex and positively homogeneous of degree one. Suppose that $\xi, \eta \in E^*$, and so $(\xi, \eta) \in T_2^*$. We say that (ξ, η) *supports* $2g_a^{1/2}$ at (b, c) if $\xi b + \eta c = 2[g(a, b, c)]^{1/2}$ and if $\xi d + \eta e \leqq 2[g(a, d, e)]^{1/2}$ for all (d, e). Furthermore, (ξ, η) supports $2g_a^{1/2}$ *properly* at (b, c) if (ξ, η) supports $2g_a^{1/2}$ at (b, c) and if $\xi b = \eta c$, $\xi c = \eta b = 0$.

The following lemma appears in [7]

LEMMA 1. *If (ξ, η) supports $2g_a^{1/2}$ properly at (b, c) then $g(a, b, c) = Ag(a, b \wedge c) = [b \wedge c, \xi \wedge \eta]$ where $[d \wedge e, \rho \wedge \sigma] = \rho(d)\sigma(e) - \rho(e)\sigma(d)$.*

Proof. If $r \neq 0$ then $4g(a, rb, sb + r^{-1}c) \geqq (r\xi(b) + r^{-1}\eta(c))^2 = (r + r^{-1})^2(\xi b + \eta c)^2/4 \geqq (\xi b + \eta c)^2 = 4g(a, b, c)$ and $g(a, b, c) = [b \wedge c, \xi \wedge \eta]$.

Now suppose that $\xi, \eta \in E^*$, $\xi^2 = \eta^2 = 1$ and $\xi \cdot \eta = 0$. Let $H_{\xi,\eta}(b, c) = [(\xi b + \eta c)^2 + (\xi c - \eta b)^2]/4$. It is easy to see that $H_{\xi,\eta} = H_{\rho,\sigma}$ if $\xi \wedge \eta = \rho \wedge \sigma$, $\rho^2 = \sigma^2 = 1$ and $\rho \cdot \sigma = 0$. Hence we can define $h_{\xi \wedge \eta} = H_{\xi,\eta}$. It quickly follows that $h_\zeta(b\cos\theta - c\sin\theta, b\sin\theta + c\cos\theta) = h_\zeta(b, c)$ for all $\zeta \in T^*$ and all real θ. As the sum of squares of linear functionals, h is continuous, convex and homogeneous of degree two. An easy computation shows that $\rho \wedge \sigma = \zeta$ if (ρ, σ) supports $2h_\zeta^{1/2}$ at (b, c) where $h_\zeta(b, c) \neq 0$.

We define $Ah_\zeta(b \wedge c) = \inf_{d \wedge e = b \wedge c} h_\zeta(d, e)$.

If ϕ is a real number let $\phi^+ = \max\{\phi, 0\}$.

LEMMA 2. $Ah_\zeta(b \wedge c) = [b \wedge c, \zeta]^+$.

Proof. Suppose that $\zeta = \xi \wedge \eta$ where $\xi^2 = \eta^2 = 1$ and $\xi \cdot \eta = 0$. If $[b \wedge c, \xi \wedge \eta] = 1$ then (ξ, η) supports $2h^{1/2} = 2h_\zeta^{1/2}$ properly at $(\eta(c)b - \eta(b)c, -\xi(c)b + \xi(b)c)$. If $[b \wedge c, \xi \wedge \eta] = -1$ then $\xi^2(b) + \eta^2(b) = \delta^2$ for some $\delta > 0$. If $\eta(b) = 0$ let $b' = b/\xi(b)$ and $c' = -\xi(c)b + \xi(b)c$; if $\eta(b) \neq 0$ let $b' = b/\delta$ and $c' = -[\xi(b) + \delta^2\eta(c)]b/[\delta\eta(b)] + \delta c$. In both cases $h(b', c') = 0$ and $b' \wedge c' = b \wedge c$. If $[b \wedge c, \xi \wedge \eta] = 0$ let $\varepsilon > 0$. If $\eta(b) \neq 0$ let $b' = \varepsilon b$ and $c' = [-\eta(c)b + \eta(b)c]/[\varepsilon\eta(b)]$. Then $h(b', c') = \varepsilon^2\delta^2/4$. If $\eta(b) = 0$ and $\xi(b) = 0$ let $b' = b/\varepsilon$ and $c' = \varepsilon c$; now $h(b', c') = \varepsilon^2[\xi^2(c) + \eta^2(c)]/4$. If $\eta(b) = 0$ and $\xi(b) \neq 0$ then let $b' = \varepsilon b$ and $c' = -[\xi(c)b]/[\varepsilon\xi(b)] + c/\varepsilon$ to obtain $h(b', c') = \varepsilon^2\xi^2(b)/4$. The lemma follows by positive homogeneity.

LEMMA 3. *Let* $\lambda \geqq 1$, k *be continuous on* E *into* $[\lambda^{-1}, \lambda]$, $g \in \mathscr{D}$ *and* $f(a, b, c) = \max\{g(a, b, c), k(a)h_\zeta(b, c)\}$. *Then* $f \in \mathscr{D}$ *and* $Af(a, b \wedge c) = \max\{Ag(a, b \wedge c), k(a)Ah_\zeta(b \wedge c)\}$ *for all* $a, b, c \in E$.

Proof. That $f \in \mathscr{D}$ is evident as is the fact that $Af \geqq \max\{Ag, kAh_\zeta\}$. Choose a, b, c with $b \wedge c \neq 0$. Then there exist d and e with $d \wedge e = b \wedge c$ and $Af(a, d \wedge e) = f(a, d, e)$, and there exist (ρ, σ) which supports $2f_a^{1/2}$ properly at (d, e), [7]. Assume, at first, that $f(a, d, e) = g(a, d, e) > k(a)h_\zeta(d, e)$. If (ρ, σ) did not support $2g_a^{1/2}$ at (d, e), then there would exist $(d_n, e_n) \to (d, e)$ such that $k(a)h_\zeta(d_n, e_n) > g(a, d_n, e_n)$ and this is impossible for large n. Hence (ρ, σ) supports $2g_a^{1/2}$ properly at (d, e) and $Ag(a, d \wedge e) = g(a, d, e) = f(a, d, e) = Af(a, d \wedge e)$. If $f(a, d, e) = k(a)h_\zeta(d, e) > g(a, d, e)$, a similar argument, together with the fact that $\rho \wedge \sigma = k(a)(\xi \wedge \eta)$, gives $k(a)Ah_\zeta(d \wedge e) = Af(a, d \wedge e)$. If $g(a, d, e) = k(a)h_\zeta(d, e)$, let $\varepsilon > 0$ and $\phi = \max\{(1 + \varepsilon)^2 g, k \cdot h_\zeta\}$. Obviously $((1 + \varepsilon)\rho, (1 + \varepsilon)\sigma)$ supports $2\phi_a^{1/2}$ properly at (d, e) and $(1 + \varepsilon)^2 g(a, d, e) > k(a)h_\zeta(d, e)$. Hence $Af(a, d \wedge e) \leqq A\phi(a, d \wedge e) = (1 + \varepsilon)^2 Ag(a, d \wedge e)$ and the lemma follows.

Let $f \in \mathscr{A}$ and $\lambda = \lambda(f)$. We define k on $E \times [T_1^* - \{0\}]$ by $1/k(a, \zeta) = \sup_{\alpha \neq 0} [\alpha, \zeta]/f(a, \alpha)$. Then k is continuous, range $k \subset [(\lambda \| \zeta \|)^{-1}, \lambda \| \zeta \|^{-1}]$, k_a^{-1} is convex and

$$f(a, \alpha) = \max_{\zeta \in T_1^*} k(a, \zeta)[\alpha, \zeta] .$$

If $f(a, \alpha) = \max_{\zeta \in T^*} k(a, \zeta)[\alpha, \zeta]$ then f is *simple*.

THEOREM. *Let* k *be as above and* $f(a, \alpha) = \max_{\zeta \in T^*} k(a, \zeta)[\alpha, \zeta]$. *Then* $g(a, b, c) = \max_{\zeta \in T^*} k(a, \zeta)h_\zeta(b, c)$ *is in* $\mathscr{D}$ *and* $f = Ag$.

Proof. Let $\{\zeta_p\}$ be dense in T^* and λ be as above. Let

$$g_1(a, b, c) = \max\{N_2(b, c)/\lambda, k(a, \zeta_1)h_1(b, c)\}$$

and

$$g_{p+1}(a, b, c) = \max\{g_p(a, b, c), k(a, \zeta_{p+1})h_{p+1}(b, c)\}$$

where $h_p = h_{\zeta_p}$.

By the last lemma,

$$Ag_p(a, b \wedge c) = \max\left\{\frac{N_1(b \wedge c)}{\lambda}, \max_{1 \leqq m \leqq p} k(a, \zeta_m)[b \wedge c, \zeta_m]\right\} \leqq f(a, b \wedge c)$$

for each p. Hence $\lim Ag_p \leqq f$. On the other hand, for fixed a, b, c and arbitrary $\varepsilon > 0$ there exists r such that $f(a, b \wedge c) < k(a, \zeta_r)[b \wedge c, \zeta_r] + \varepsilon$ and so $f = \lim Ag_p$.

A little arithmetic shows that

$$| h_p^{1/2}(r, s) - h_p^{1/2}(u, v) | \leqq || (r, s) - (u, v) || .$$

Hence $\{g_p^{1/2}\}$ *is equicontinuous and* $g_0 = \lim g_p$ *is continuous.* It is clear that $g_0 = g$ and that $g \in \mathscr{D}$. Furthermore, if K and L are compact subsets of E^N and T_2, respectively, then, by a theorem of Dini, g_p converges uniformly to g on $K \times L$.

It remains to show that $Ag = \lim Ag_p$. Choose $a, b, c \in E$ and $\varepsilon > 0$. There exist (b_p, c_p) with $N_2(b_p, c_p) \leqq \lambda Ag(a, b \wedge c)$ such that $Ag_p(a, b_p \wedge c_p) = g_p(a, b_p, c_p)$ and $b_p \wedge c_p = b \wedge c$. By passing to a subsequence, if necessary, we can suppose that there exists (b_0, c_0) such that $(b_p, c_p) \to (b_0, c_0)$. Let p be so large that $g_p(a, r, s) > g(a, r, s) - \varepsilon$ for $N_2(r, s) \leqq \lambda Ag(a, b \wedge c)$ and so large that $|| (b_p, c_p) - (b_0, c_0) || < \varepsilon$. Then $Ag(a, b \wedge c) = Ag(a, b_0 \wedge c_0) \leqq g(a, b_0, c_0) < g_p(a, b_0, c_0) + \varepsilon < [g_p^{1/2}(a, b_p, c_p) + \lambda^{1/2}\varepsilon]^2 + \varepsilon = [Ag_p^{1/2}(a, b_p \wedge c_p) + \lambda^{1/2}\varepsilon]^2 + \varepsilon$. Hence $Ag \leqq \lim Ag_p$, and the opposite inequality is evident.

If π is a projection of E onto a plane $P \subset E$, then there exist ξ and η in E^* such that $\xi(\pi e) = \xi(e)$, $\eta(\pi e) = \eta(e)$ and $[b \wedge c, \xi \wedge \eta] \neq 0$ whenever b and c are linearly independent points of P. A computation gives $[b \wedge c, \xi \wedge \eta](\pi e) = [e \wedge c, \xi \wedge \eta]b + [b \wedge e, \xi \wedge \eta]c$ and we can identify π with $\xi \wedge \eta$. Since we can also suppose that $\xi^2 = \eta^2 = 1$, $\xi \cdot \eta = 0$, we can identify the set of projections with the elements of T^*.

THEOREM 2. *Let* $f \in \mathscr{A}$ *and suppose that for each* $a \in E$ *and each* $b \wedge c \neq 0$ *there exists a projection* ζ_0 *(in* T^**) onto the plane determined by* b *and* c *such that* $[b \wedge c, \zeta_0] > 0$ *and such that* $f(a, \zeta_0(d) \wedge \zeta_0(e)) \leqq f(a, d \wedge e)$ *whenever* $[\zeta_0(d) \wedge \zeta_0(e), \zeta_0] > 0$. *Then* f *is simple and* $f(a, b \wedge c) = k(a, \zeta_0)[b \wedge c, \zeta_0]$.

Proof. There exist d and e such that $1/k(a, \zeta_0) = [d \wedge e, \zeta_0]/f(a, d, e)$. Hence

$$\frac{1}{k(a, \zeta_0)} = \frac{[\zeta_0(d) \wedge \zeta_0(e), \zeta_0]}{f(a, d \wedge e)}$$
$$\leqq \frac{[\zeta_0(d) \wedge \zeta_0(e), \zeta_0]}{f(a, \zeta_0(d) \wedge \zeta_0(e))} = \frac{[b \wedge c, \zeta_0]}{f(a, b \wedge c)} \leqq \frac{1}{k(a, \zeta_0)} .$$

It is evident that the converse of this theorem holds.

References

1. Lamberto Cesari, *An existence theorem of calculus of variations for integrals on parametric surfaces*, Amer, J. Math. **74** (1952), 265-295.

2. J. M. Danskin, *On the existence of minimizing surfaces in parametric double integral problems of the calculus of variations*, Riv. Mat. Univ. Parma, **3** (1952), 43-63.

3. C. B. Morrey, Jr., *Multiple integral problems in the calculus of variations and related topics*, University of California, 1943.

4. ———, *The parametric variational problem for double integrals*, Comm. Pure Appl. Math. **14** (1961), 569-575.

5. Ju. G. Rešetnjak, *A new proof of the theorem of existence of an absolute minimum for two-dimensional problems of the calculus of variations in parametric form*, Sibirsk. Mat. Ž. **3** (1962), 744-768.

6. A. G. Sigalov, *Two-dimensional problems of the calculus of variations*, Uspehi Matem. Nauk (N.S.) 6, **42** (1951), 16-101.

7. E. Silverman, *A problem of least area*, Pacific J. Math., **14** (1964), 309-331.

PACIFIC JOURNAL OF MATHEMATICS
Vol. 15, No. 1, 1965

CHEBYSHEV APPROXIMATION TO ZERO

JAMES M. SLOSS

In this paper we shall be concerned with the questions of existence, uniqueness and constructability of those polynomials in $k+1$ variables $(x_1, x_2, \cdots, x_k, y)$ of degree not greater than n_s in x_s and m in y which best approximate zero on $I_1 \times I_2 \times \cdots \times I_{k+1}, I_s = [-1, 1]$, in the Chebyshev sense.

It is a classic result that among all monic polynomials of degree not greater than n there is a unique polynomial whose maximum over the interval $[-1, 1]$ is less than the maximum over $[-1, 1]$ of any other polynomial of the same type and moreover it is given by $\widetilde{T}_n(x) = 2^{1-n} \cos [n \text{ are } \cos x]$, the normalized Chebyshev polynomial.

Our method of attack will be to prove a generalization of an inequality for monic polynomials in one variable concerning the lower bound of the maximum viz. $\max_{-1 \leqq x \leqq 1} | P_n(x) | \geqq 2^{1-n}$ where $P_n(x)$ is a monic polynomial of degree not greater than n. The theorem will show that the only hope for uniqueness is to normalize our class of polynomials. This is done in a very natural way viz. by considering only polynomials, if they exist, of the form:

$$(0.1) \qquad P(x_1, x_2, \cdots, x_k, y) = A_m(x_1, \cdots, x_k)y^m + A_{m-1}(\cdots)y^{m-1} + \cdots + A_0(\cdots)$$

for which $A_m(x_1, x_2, \cdots, x_k)$ is the best polynomial approximation to zero on $I_1 \times I_2 \times \cdots \times I_k$. Thus if $k = 1$, we consider only polynomials of the form:

$$(0.2) \qquad P(x_1, y) = \widetilde{T}_n(x_1)y^m + A_{m-1}(x_1)y^{m-1} + \cdots + A_0(x_1) .$$

We find in the case of (0.2) that there is a unique best polynomial approximation and it is given by $\widetilde{T}_n(x_1)\widetilde{T}_m(y)$. Thus we can consider the question of existence, uniqueness and constructability of a polynomial of the form:

$$(0.3) \qquad P(x_1, x_2, y) = \widetilde{T}_{n_1}(x_1)\widetilde{T}_{n_2}(x_2)y^m + A_{m-1}(x_1, x_2)y^{m-1} + \cdots + A_0(x_1, x_2)$$

that best approximates zero. We find in this case there is a unique best polynomial approximation and it is given by $\widetilde{T}_{n_1}(x_1)\widetilde{T}_{n_2}(x_2)\widetilde{T}_m(y)$. Continuing in this way we shall show that the question is meaning-

Received December 18, 1963.

ful in general and that there is a unique best polynomial approximation to zero of the form (0.1) given by $\tilde{T}_{n_1}(x_1)\tilde{T}_{n_2}(x_2)\cdots\tilde{T}_{n_k}(x_k)\tilde{T}_m(y)$.

The uniqueness and constructability are the most surprising results, since as Buck [1] has shown, $F(x, y) = xy$ has amongst those polynomials of the form

$$p(x, y) = a_0 + a_1(x + y) + a_2(x^2 + y^2)$$

infinitely many polynomials of best approximation which are given by:

$$\alpha f_1 + \beta f_2\,, \quad \alpha \geqq 0\,, \quad \beta \geqq 0\,, \quad \alpha + \beta = 1$$

where

$$f_1(x, y) = \frac{1}{2}(x^2 + y^2) - \frac{1}{4}\,,$$

$$f_2(x, y) = x + y - \frac{1}{2}(x^2 + y^2) - \frac{1}{4}\,.$$

We shall finally normalize the polynomials in a different way and show by construction, the existence of a polynomial, of best approximation in this class. However in this case the question of uniqueness remains open.

1. NOTATION. Let $n_1, n_2, \cdots, n_k$ be positive fixed integers. Let σ be the finite set of vectors $\{(x_{1j_1}, x_{2j_2}, \cdots, x_{kj_k})\}$, where $j_1, j_2, \cdots, j_k$ are integers with $0 \leqq j_1 \leqq n_1, 0 \leqq j_2 \leqq n_2, \cdots, 0 \leqq j_k \leqq n_k$; and where also $-1 \leqq x_{1j_1} \leqq 1, -1 \leqq x_{2j_2} \leqq 1, \cdots, -1 \leqq x_{kj_k} \leqq 1$ and no two of the x_{1j_1} are the same, no two of the x_{2j_2} are the same, $\cdots$, no two of the x_{kj_k} are the same. Let $Q(x, y) = Q(x_1, x_2, \cdots, x_k, y)$ be any polynomial in $x_1, x_2, \cdots, x_k$ and y of degree $\leqq n_1 + n_2 + \cdots + n_k + m - 1$ where Q is of degree $\leqq n_s$ in x_s, $s = 1, 2, \cdots, k$ and of degree $\leqq m$ in y. Let π be the set of all such polynomials. Thus if $Q(x, y)$ is in π

$$Q(x, y) = p_m(x)y^m + p_{m-1}(x)y^{m-1} + \cdots + p_0(x)$$

where $p_m(x)$ is a polynomial in $x_1, x_2, \cdots, x_k$ of

$$\text{degree} \leqq n_1 + n_2 + \cdots + n_k - 1$$

and $p_s(x), 0 \leqq s \leqq m - 1$, are polynomials of degree $\leqq n_1 + n_2 + \cdots + n_k$ in $x_1, x_2, \cdots, x_k$. Let

$$A[p_m; \pi, \sigma] = \min_{x \text{ in } \sigma} |\, x_1^{n_1}x_1^{n_2} \cdots x_k^{n_k} - p_m(x_1, x_2, \cdots, x_k)\,|$$

which does not depend on the particular Q, but only on the class π and the leading coefficient polynomial of y.

THEOREM 1. *If $Q(x, y)$ is any polynomial in π and if σ is any set of the type described above then*

$$\max_{\substack{-1\leq x_s\leq 1\\ -1\leq y\leq 1}} |x_1^{n_1}x_2^{n_2}\cdots x_k^{n_k}y^m - Q(x_1, x_2, \cdots, x_k, y)| \geq A[p_m; \pi, \sigma]2^{1-m} .$$

Proof. Assume not. Then there exists a $Q^*(x, y)$ in π and a set σ of the type described such that:

$$\max_{\substack{-1\leq x_s\leq 1\\ -1\leq y\leq 1}} |x_1^{n_1}x_2^{n_2}\cdots x_k^{n_k}y^m - Q^*(x, y)| < A[p_m; \pi, \sigma]2^{1-m}$$

consider the polynomial:

$$P(x, y) = x_1^{n_1}x_2^{n_2}\cdots x_k^{n_k}y^m \\ - Q^*(x, y) - [x_1^{n_1}x_2^{n_2}\cdots x_k^{n_k} - p_m(x)]\tilde{T}_m(y)$$

where $p_m(x)$ is the coefficient of y^m in $Q^*(x, y)$ and where

$$(1) \qquad \tilde{T}_m(y) = 2^{1-m}T_m(y) = 2^{1-m}\cos[m \operatorname{arc} \cos y] .$$

Then $P(x, y)$ is a polynomial of degree $\leq m-1$ in y and thus can be written:

$$P(x, y) = q_{m-1}(x)y^{m-1} + q_{m-2}(x)y^{m-2} + \cdots + q_0(x)$$

where $q_s(x)$, $0 \leq s \leq m-1$, are polynomials in $x_1, x_2, \cdots, x_k$ of degree $\leq n_1 + n_2 + \cdots + n_k$.

Let $(x_{1j_1}, x_{2j_2}, \cdots, x_{kj_k})$ belong to σ and y_r be one of the points

$$y_r = \cos\frac{r\pi}{m} , \quad 0 \leq r \leq m , \quad r = \text{integer} .$$

Then $\tilde{T}_m(y_r) = (-1)^r 2^{1-m}$ and we can show that the sign of

$$P[x_{1j_1}, x_{2j_2}, \cdots, x_{kj_k}, y_r]$$

is the same as the sign of $-[x_{1j_1}^{n_1}\cdots x_{kj_k}^{n_k} - p_m(y_{1j_1}, \cdots, x_{kj_k})]. \ \tilde{T}_m(y_r)$. To see this note that:

$$|\tilde{T}_m(y_r)|\,|x_{1j_1}^{n_1}\cdots x_{kj_k}^{n_k} - p_m(x_{1j_1}, \cdots, x_{kj_k})| \\ = |x_{1j_1}^{n_1}\cdots x_{kj_k}^{n_k} - p_m(x_{1j_1}, \cdots, x_{1j_k})|\,2^{1-m} \\ \geq A[p_m; \pi, \sigma]2^{1-m} .$$

But by the assumption

$$\max_{\substack{-1\leq x_s\leq 1\\ -1\leq y\leq 1}} |x_1^{n_1}\cdots x_k^{n_k}y^m - Q^*(x, y)| < A[p_m; \pi, \sigma]2^{1-m}$$

and thus a fortiori

$$| x_{1j_1}^{n_1} \cdots x_{kj_k}^{n_k} y_r^m - Q^*(x_{1j_1}, \cdots, x_{kj_k}, y_r) | < A[p_m; \pi, \sigma] 2^{1-m} .$$

If we fix x in σ then $P(x, y)$ is a polynomial of the one variable y and of degree $\leqq m-1$. And as y takes on the values $y_r = \cos(\pi r/m)$, $P(x, y)$ changes sign $m+1$ times. Thus $P(x, y)$ has m zeros, which means $q_{m-1}(x) = 0, q_{m-2}(x) = 0, \cdots, q_0(x) = 0$ since $P(x, y)$ is only of degree $\leqq m-1$.

Since x was an arbitrary point of σ, then

$$q_s[x_{1j_1}, x_{2j_2}, \cdots, x_{kj_k}] = 0 , \quad 0 \leqq s \leqq m-1$$

where $0 \leqq j_1 \leqq n_1, 0 \leqq j_2 \leqq n_2, \cdots, 0 \leqq j_k \leqq n_k$. But $q_s(x)$ is a polynomial of degree $\leqq n_1$ in x_1, of degree $\leqq n_2$ in $x_2, \cdots$, of degree $\leqq n_k$ in x_k and thus

$$q_s[x_1, x_2, \cdots, x_k] \equiv 0 , \quad 0 \leqq s \leqq m-1 .$$

From which we see $P(x, y) \equiv 0$ and thus:

$$x_1^{n_1} \cdots x_k^{n_k} y^m - Q^*(x, y) \equiv [x_1^{n_1} \cdots x_k^{n_k} - p_m(x)] \tilde{T}_m(y) .$$

But clearly:

$$\max_{\substack{-1 \leqq x_s \leqq 1 \\ -1 \leqq y \leqq 1}} | x_1^{n_1} \cdots x_k^{n_k} - p_m(x) | \, | \tilde{T}(y) | \geqq A[p_m; \pi, \sigma] 2^{1-m}$$

which is a contradiction and thus the theorem is proved.

Let us now consider the subset of polynomials π_0 of π for which $Q(x, y)$ belongs to π and $p_m(x) = 0$. Then by the above theorem, a lower bound for the maximum is

$$A[0; \pi, \sigma] = \min_{x \text{ in } \sigma} | x_1^{n_1} \cdots x_k^{n_k} | < 1$$

which clearly depends on the set σ. We shall next show that for this subset π_0, we get a lower bound for the maximum that is independent of σ and moreover the lower bound is larger than $A[0; \pi, \sigma]$ for all σ, namely it is unity. In the third theorem we shall show that unity is the best possible lower bound i.e. there is a polynomial in π_0 for which the maximum is 2^{1-m}.

THEOREM 2. *Let $Q(x, y)$ be any polynomial in π_0, then*

$$\max_{\substack{-1 \leqq x_s \leqq 1 \\ -1 \leqq y \leqq 1}} | x_1^{n_1} x_2^{n_2} \cdots x_k^{n_k} y^m - Q(x_1, x_2, \cdots, x_k, y | \geqq 2^{1-m} .$$

Proof. By contradiction. Assume there exists a $Q(x_1, \cdots, x_k, y)$ in π_0 such that:

$$\max_{\substack{-1 \leqq x_s \leqq 1 \\ -1 \leqq y \leqq 1}} | x_1^{n_1} x_2^{n_2} \cdots x_k^{n_k} y^m - Q(x_1, \cdots, x_k, y) | < 2^{1-m} .$$

Then there exist δ_s's, $1 \leqq s \leqq k, 1 > \delta_s > 0$ such that:

$$\max_{\substack{-1 \leqq x_s \leqq 1 \\ -1 \leqq y \leqq 1}} | x_1^{n_1} \cdots x_k^{n_k} y^m - Q(x_1, \cdots, x_k, y) | < 2^{1-m} \prod_{s=1}^{k} \delta_s^{n_s} .$$

Let $\widetilde{T}_m(y)$ be given by (1) and consider the polynomial

$$P(x_1, \cdots, x_k, y) \equiv x_1^{n_1} \cdots x_k^{n_k} y^m - Q(x_1, \cdots, x_k, y) - x_1^{n_1} \cdots x_k^{n_k} \widetilde{T}_m(y) .$$

$P(x_1, \cdots, x_k, y)$ is a polynomial of degree $\leqq m - 1$ in y and of degree $\leqq n_s$ in x_s $1 \leqq s \leqq k$.

Let $\sigma^* = \{(x_{1j_1}, x_{2j_2}, \cdots, x_{kj_k})\}$ where $j_1, \cdots, j_k$ are integers with

$$0 \leqq j_1 \leqq n_1 + 1, 0 \leqq j_2 \leqq n_2 + 1, \cdots, 0 \leqq j_k \leqq n_k + 1;$$
$$\delta_1 < x_{1j_1} \leqq 1, \delta_2 < x_{2j_2} \leqq 1, \cdots, \delta_k < x_{kj_k} \leqq 1$$

and the x_{1j_1} are distinct, $\cdots$, the x_{kj_k} are distinct.

Note that for x in σ^*, the sign of $P(x_{1j_1}, \cdots, x_{kj_k}, y)$ is the same as the sign of $-x_{1j_1}^{n_1} \cdots x_{kn_k}^{n_k} \widetilde{T}_m(y_r)$ for $y_r = \cos(r\pi/m)$, $r = 0, 1, \cdots, m$. This follows from the fact that:

$$| x_{1j_1}^{n_1} \cdots x_{kj_k}^{n_k} y_r^m - Q(x_1, \cdots, x_k, y_r) | < 2^{1-m} \prod_{s=1}^{k} \delta_s^{n_s}$$

and the fact that:

$$| x_{1j_1}^{n_1} \cdots x_{kj_k}^{n_k} \widetilde{T}_m(y_r) | = 2^{1-m} \prod_{s=1}^{k} x_{sj_s}^{n_s} > 2^{1-m} \prod_{s=1}^{k} \delta_s^{n_s} .$$

Thus we conclude that $P(x_{1j_1}, \cdots, x_{kj_k}, y)$ has $m + 1$ sign changes for $(x_{1j_1}, \cdots, x_{kj_k})$ in σ^*. Let us write

$$P(x, y) = p_{m-1}(x)y^{m-1} + p_{m-2}(x)y^{m-2} + \cdots + p_0(x)$$

where $p_s(x)$, $0 \leqq s \leqq m - 1$, are polynomials of degree $\leqq n_s$ in x_s, $0 \leqq s \leqq k$. For each x in σ^*, $P(x, y)$ has $m + 1$ sign changes and thus $p_{m-1}(x) = 0$, $p_{m-2}(x) = 0, \cdots, p_0(x) = 0$ for each x in σ^*. If for $(x_{1j_1}, x_{2j_2}, \cdots, x_{kj_k})$ in σ^*, we fix all but the first component, we get $n_1 + 2$ values in σ^* for which $p_s(x) = 0, 0 \leqq s \leqq m - 1$, but these $p_s(x)$ are of degree $\leqq n_1$ in x_1 and thus $p_s(x_1, x_{2j_2}, x_{3j_3}, \cdots, x_{kj_k}) = 0$ for all real x_1. Continuing in this way, we see that $p_s(x_1, x_2, \cdots, x_k) \equiv 0$ for all $(x_1, x_2, \cdots, x_k)$, x_s real. Thus:

$$P(x_1, x_2, \cdots, x_k, y) \equiv 0$$

for all real x_s and real y. Thus

$$x_1^{n_1} \cdots x_k^{n_k} \widetilde{T}_m(y) \equiv x_1^{n_1} \cdots x_k^{n_k} y^m - Q(x_1, \cdots, x_k, y) .$$

But

$$\max_{\substack{-1\leqq x_s\leqq 1\\ -1\leqq y\leqq 1}} |x_1^{n_1}\cdots x_k^{n_k}\widetilde{T}_m(y)| = 2^{1-m}$$

which gives a contradiction and the theorem is proved.

2. Normalization of competing polynomials and construction of the best polynomial. We shall now consider a subset $\pi(\beta)$ of the set of polynomials π. We shall then answer the question of existence, uniqueness and constructability of the best polynomial approximation in the maximum norm to zero within this class $\pi(\beta)$ on the cube

$$-1\leqq x_1\leqq 1, \cdots, -1\leqq x_k\leqq 1, -1\leqq y\leqq 1 .$$

It is apparent from Theorem 1, that if we want uniqueness independent of σ, it is necessary to consider some subset of π.

DEFINITION. A polynomial

$$Q(x, y) = p_m(x_1, x_2, \cdots, x_k)y^m + p_{m-1}(x_1, x_2, \cdots, x_k)y^{m-1} + \cdots + p_0(x_1, x_2, \cdots, x_k)$$

which is in π and for which

$$x_1^{n_1}x_2^{n_2}\cdots x_k^{n_k} - p_m(x_1, x_2\cdots x_k) = \widetilde{T}_{n_1}(x_1)\widetilde{T}_{n_2}(x_2)\cdots\widetilde{T}_{n_k}(x_k)$$

is said to be in $\pi(\beta)$.

LEMMA. *Let $q(y)$ be a polynomial in y, let $y_0 > y_1 > \cdots > y_m$ be any set of real numbers for which*

$$q(y_0)\leqq 0, q(y_1)\geqq 0, q(y_2)\leqq 0, \cdots (-1)^m q(y_m)\leqq 0 .$$

Then $q(y)$ has m zeros including multiplicities on $[y_0, y_m]$.

Proof. (by induction): For $m = 1$ obvious. Assume theorem to be true for $m\leqq k$. Let $y_0 > y_1 > y_2 > \cdots > y_{k+1}$ be any set of real numbers such that

$$q(y_0)\leqq 0, q(y_1)\geqq 0, \cdots (-1)^k q(y_k)\leqq 0, (-1)^{k+1}q(y_{k+1})\leqq 0 .$$

Case 1. $q(y_s)\neq 0$ for some $1\leqq s\leqq k$. Then by the induction hypothesis $q(y)$ has s zeros on $[y_0, y_s]$ and has $k+1-s$ zeros on $[y_s, y_{k+1}]$. But $q(y_s)\neq 0$ thus $q(y)$ has s zeros on $y_0\leqq y\leqq y_s$ and thus $q(y)$ has $s+(k+1-s) = k+1$ zeros on $[y_0, y_{k+1}]$.

Case 2. $q(y_0) < 0$. Then unless $q(y_s) = 0$ for $1\leqq s\leqq k$ we are in Case 1 and we are finished. Therefore, assume $q(y_s) = 0, 1\leqq s\leqq k$.

We may as well assume $q(y) < 0$ on (y_0, y_1) since if not then $q(y)$ has a zero there because $q(y_0) < 0$, and we are finished. Also, we may as well assume $q(y) > 0$ on (y_1, y_2) since if not and $q(y)$ has no zeros on (y_1, y_2) (if does have a zero then we are finished) then since $q(y_0) < 0$ and $q(y_1) = 0$, we must have that $q(y)$ has 2 zeros in (y_0, y_2), continuing in this way we see that we may as well assume that $(-1)^s q(y) < 0$ on (y_s, y_{s+1}) for $0 \leqq s \leqq k$. In particular $(-1)^k q(y) < 0$ for y on (y_k, y_{k+1}). But by assumption $(-1)^{k+1} q(y_{k+1}) \leqq 0$. Thus by the continuity of $q(y)$, we have $q(y_{k+1}) = 0$ and $q(y_s) = 0$ for $1 \leqq s \leqq k+1$ i.e. $q(y)$ has $k+1$ zeros on $[y_0, y_{k+1}]$.

Case 3. $q(y_0) = 0$ proof is obvious making use of Case 1.

THEOREM 3. There exists a unique $Q^*(x, y)$ in $\pi(\beta)$ such that

$$\max_{\substack{-1 \leqq x_s \leqq 1 \\ -1 \leqq y \leqq 1}} | x_1^{n_1} x_2^{n_2} \cdots x_k^{n_k} y^m - Q^*(x, y) |$$

is a minimum. Moreover:

$$Q^*(x, y) = -\tilde{T}_{n_1}(x_1)\tilde{T}_{n_2}(x_2) \cdots \tilde{T}_{n_k}(x_k)\tilde{T}_m(y) + x_1^{n_1} x_2^{n_2} \cdots x_k^{n_k} y^m .$$

Proof. Existence by construction. Let the σ of Theorem 1 be the special set of vectors

$$\sigma(\beta) = \{(x_{1j_1}, x_{2j_2}, \cdots, x_{kj_k})\}$$

where

$$x_{1j_1} = \cos(j_1\pi/n_1), x_{2j_2}, \cdots, x_{kj_k} = \cos(j_k\pi/n_k)$$

$$0 \leqq j_1 \leqq n_1, 0 \leqq j_2 \leqq n_2, \cdots, 0 \leqq j_k \leqq n_k .$$

Then

$$\begin{aligned} A[p_m, \pi(\beta), \sigma(\beta)] &= \min_{x \text{ in } \sigma(\beta)} | x_1^{n_1} x_2^{n_2} \cdots x_k^{n_k} - p_m(x_1, x_2, \cdots, x_k) | \\ &= \min_{x \text{ in } \sigma(\beta)} | \tilde{T}_{n_1}(x_1)\tilde{T}_{n_2}(x_1) \cdots \tilde{T}_{n_k}(x_k) | \\ &= 2^{1-n_1} 2^{1-n_2} \cdots 2^{1-n_k} . \end{aligned}$$

Thus by Theorem 1

$$\max_{\substack{-1 \leqq x_j \leqq 1 \\ -1 \leqq y \leqq 1}} | x_1^{n_1} x_2^{n_2} \cdots x_k^{n_k} y^m - Q(x, y) | \geqq 2^{1-n_1} 2^{1-n_2} \cdots 2^{1-n_k} 2^{1-m} .$$

But the polynomial

$$Q^*(x, y) = x_1^{n_1} x_2^{n_2} \cdots x_k^{n_k} y^m - \tilde{T}_{n_1}(x_1)\tilde{T}_{n_2}(x_2) \cdots \tilde{T}_{n_k}(x_k)\tilde{T}_m(y)$$

clearly belongs to $\pi(\beta)$ and

$$\max_{\substack{-1 \leqq x_s \leqq 1 \\ -1 \leqq y \leqq 1}} |x_1^{n_1} x_2^{n_2} \cdots x_k^{n_k} y^m - Q^*(x, y)| = 2^{1-n_1} 2^{1-n_2} \cdots 2^{1-n_k} 2^{1-m} .$$

Thus $Q^*(x, y)$ is a best approximation from the set $\pi(\beta)$

Uniqueness. Let $Q^*(x, y)$ in $\pi(\beta)$ be a polynomial of best approximation and let

$$\begin{aligned} P(x, y) &= x_1^{n_1} x_2^{n_2} \cdots x_k^{n_k} y^m - Q^*(x, y) - \tilde{T}_{n_1}(x_1) \cdots \tilde{T}_{n_k}(x_k) \tilde{T}_m(y) \\ &= [x_1^{n_1} x_2^{n_1} \cdots x_k^{n_k} - p_m(x)] y^m - p_{m-1}(x) y^{m-1} - \cdots p_0(x) \\ &\quad - \tilde{T}_{n_1}(x_1) \tilde{T}_{n_2}(x_2) \cdots \tilde{T}_{n_k}(x_k) \tilde{T}_m(y) \\ &= q_{m-1}(x) y^{m-1} + q_{m-2}(x) y^{m-2} + \cdots + q_0(x) \end{aligned}$$

where $q_{m-1}(x), \cdots, q_0(x)$ are polynomials of degree $\leqq n_s$ in x_s $0 \leqq s \leqq k$ since $Q^*(x, y)$ is in $\pi(\beta)$.

Let $x^* = (x_1^*, x_2^*, \cdots, x_k^*)$ be a fixed but arbitrary element of $\sigma(\beta)$. Then we claim that $P(x^*, y)$ has m zeros including multiplicities in $[-1, 1]$. To see this let $y_s = \cos(s\pi/m)$, $0 \leqq s \leqq m$, then since

$$|x_1^{*n_1} x_2^{*n_2} \cdots x_k^{*n_k} y^m - Q^*(x^*, y)| \leqq 2^{1-n_1} 2^{1-n_2} \cdots 2^{1-n_k} 2^{1-m} ,$$

$$P(x^*, y_0) \leqq 0, P(x^*, y_1) \geqq 0, \cdots (-1)^m P(x^*, y_m) \leqq 0 .$$

By the lemma $P(x^*, y)$ has m zeros counting multiplicities for $-1 \leqq y \leqq 1$.

Thus $P(x^*, y)$ has m zeros but is only a polynomial of degree $m - 1$, thus $P(x^*, y) \equiv 0$. But this holds for all x^* in $\sigma(\beta)$, thus $P(x, y) \equiv 0$ and the theorem is proved.

We could formulate Theorem 3 in the following way. Let $\pi(k)$, $k \geqq 1$, be the set of polynomials of the form

$$Q(x, y) = p_m(x_1, \cdots, x_k) x_{k+1}^m + p_{m-1}(x) x_{k+1}^{m-1} + \cdots + p_0(x)$$

which is of degree $\leqq n_s$ in x_s, $1 \leqq s \leqq k$ and for which $p_m(x_1 \cdots x_k)$ is a polynomial that best approximates zero, if such exists, on the cube $I_1 \times I_2 \times \cdots \times I_k$, $I_s = [-1, 1]$, $1 \leqq s \leqq k$.

Theorem 3 *alternate.* For $k = 2, 3, 4 \cdots$, the following is true:

Statement k. $\pi(k-1)$ is not empty and there exists a unique $M_k(x_1, x_2, \cdots, x_k, x_{k+1})$ in $\pi(k)$ such that:

$$\max_{\substack{-1 \leqq x_s \leqq 1 \\ -1 \leqq y \leqq 1}} |M_k(x_1, x_2, \cdots, x_k, x_{k+1})|$$

is a minimum. Moreover:

$$M_k(x_1, x_2, \cdots, x_k, x_{k+1}) = \widetilde{T}_{n_1}(x_1)\widetilde{T}_{n_2}(x_2) \cdots \widetilde{T}_{n_k}(x_k)\widetilde{T}_{n_{k+1}}(x_{k_{n+1}}) .$$

Proof. Obvious.

Finally we wish to prove:

THEOREM 4. *There exists a monic polynomial*

$$P(x_1, \cdots, x_k, y) = x_1^{n_1} \cdots x_k^{n_k} y^m - Q(x_1, \cdots, x_k, y)$$

where $Q(x, y)$ *belongs to* π_0 *that best approximates zero on the cube* $I_1 \times I_2 \times \cdots \times I_{k+1}$, $I_s = [-1, 1]$. *The polynomial is*

$$x_1^{n_1} \cdots x_k^{n_k} \widetilde{T}_m(y) .$$

Proof. By Theorem 2

$$\max_{\substack{-1 \leqq x_s \leqq 1 \\ -1 \leqq y \leqq 1}} | P(x_1, \cdots, x_k^{n_k}, y) | \geqq 2^{1-m} .$$

But $x_1^{n_1} \cdots x_k^{n_k} \widetilde{T}_m(y)$ is a monic polynomial of the correct form with

$$\max_{\substack{-1 \leqq x_s \leqq 1 \\ -1 \leqq y \leqq 1}} | x_1^{n_1} \cdots x_k^{n_k} \widetilde{T}_m(y) | = 2^{1-m} .$$

Thus the theorem is correct.

The question of uniqueness in this case is an open one.

REFERENCE

1. R. C. Buck, *Linear spaces and approximation theory, On numerical approximation,* Edited by R. E. Langer, Published by the University of Wisconsin Press, 1959.

UNIVERSITY OF CALIFORNIA, SANTA BARBARA

PACIFIC JOURNAL OF MATHEMATICS
Vol. 15, No. 1, 1965

ISOMETRIC ISOMORPHISMS OF MEASURE ALGEBRAS

ROBERT S. STRICHARTZ

The following theorem is proved:

If G_1 and G_2 are locally compact groups, A_i are algebras of finite regular Borel measures such that $L^1(G_i) \subseteqq A_i \subseteqq \mathscr{M}(G_i)$ for $i = 1, 2$, and T is an isometric algebra isomorphism of A_1 onto A_2, then there exists a homeomorphic isomorphism α of G_1 onto G_2 and a continuous character χ on G_1 such that $T\mu(f) = \mu(\chi(f \circ \alpha))$ for $\mu \in A_1$ and $f \in C_0(G_2)$.

This result was previously known for abelian groups and compact groups (Glicksberg) and when $A_i = L^1(G_i)$ (Wendel) where T is only assumed to be a norm decreasing algebra isomorphism.

A corollary is that a locally compact group is determined by its measure algebra.

If G is a locally compact group with left Haar measure m, then the Banach space $\mathscr{M}(G)$ of finite complex regular Borel measures (the dual of the Banach space $C_0(G)$ of all continuous functions vanishing at infinity on G) can be made into a Banach algebra by defining multiplication of two elements $\mu, \nu \in \mathscr{M}(G)$ to be convolution:

$$\mu * \nu(f) = \iint f(st)d\mu(s)d\nu(t) \qquad \text{for} \qquad f \in C_0(G) .$$

The subspace $L^1(G)$ of all measures absolutely continuous with respect to m is a closed two-sided ideal and hence a subalgebra.

In [1; Theorems 3.1 and 3.2] it is shown that if G_1 and G_2 are either both abelian or both compact, then any algebraic isomorphism T of a subalgebra A_1 of $\mathscr{M}(G_1)$ containing $L^1(G_1)$ onto a subalgebra A_2 of $\mathscr{M}(G_2)$ containing $L^1(G_2)$ which is norm-decreasing on $L^1(G_1)$ has the form

$$(*) \qquad T\mu(f) = \mu(\chi(f \circ \alpha)) \qquad \mu \in A_1 \qquad f \in C_0(G_2)$$

where α is a homeomorphic isomorphism of G_1 onto G_2 and χ is a character on G_1. In this note we shall prove that $(*)$ holds where T is assumed to be an isometry but G_1 and G_2 may be arbitrary locally compact groups. Our starting point will be the theorem of Wendel [2; Theorem 1] that any isometric isomorphism $T: L^1(G_1) \to L^1(G_2)$ is of the form $(*)$.

THEOREM. *If G_1 and G_2 are locally compact groups and T is an isometric isomorphism of a subalgebra A_1 of $\mathscr{M}(G_1)$ containing $L^1(G_1)$*

Received March 8, 1964.

onto a subalgebra A_2 *of* $\mathscr{M}(G_2)$ *containing* $L^1(G_2)$ *then* T *has the form* $(*)$. *Conversely, the equation* $(*)$ *defines an isometric isomorphism of* $\mathscr{M}(G_1)$ *onto* $\mathscr{M}(G_2)$ *for every choice of* α *and* χ.

LEMMA.[1] *Let* $\mu, \nu \in \mathscr{M}(G)$. *Then* $\mu \perp \nu$ *if and only if* $\|\mu + \nu\| = \|\mu - \nu\| = \|\mu\| + \|\nu\|$.

Proof. Suppose $\mu \perp \nu$. Then there exists a disjoint partition of G into sets A, B such that $|\mu|(B) = |\nu|(A) = 0$. Thus

$$\begin{aligned}\|\mu \pm \nu\| &= |\mu \pm \nu|(G) = |\mu \pm \nu|(A) + |\mu \pm \nu|(B) \\ &= |\mu|(A) + |\nu|(B) = \|\mu\| + \|\nu\| .\end{aligned}$$

Conversely, assme $\|\mu + \nu\| = \|\mu - \nu\| = \|\mu\| + \|\nu\|$. Let $\mu = f\nu + \mu_s$ where $f \in L^1(\nu)$ and $\mu_s \perp \nu$ be the Lebesgue decomposition of μ with respect to ν. Then

$$\begin{aligned}\|\mu \pm \nu\| = \|\mu\| + \|\nu\| &= \|f\nu + \mu_s\| + \|\nu\| \\ &= \|f\nu\| + \|\mu_s\| + \|\nu\| .\end{aligned}$$

But $\|\mu \pm \nu\| = \|(1 \pm f)\nu\| + \|\mu_s\|$ so $\|(1 \pm f)\nu\| = \|f\nu\| + \|\nu\|$. Thus $f = 0$ a.e. with respect to ν hence $\mu \perp \nu$.

Proof of theorem. The converse is an easy verification. Let T be an isometric isomorphism of A_1 onto A_2. We shall show first that T maps $L^1(G_1)$ onto $L^1(G_2)$ and hence has the form $(*)$ when restricted to $L^1(G_1)$, and then that $(*)$ extends to all of A_1.

Indeed $L^1(G_i)$ $i = 1, 2$ will be shown to be the intersection of all nontrivial closed left ideals $I \subseteqq A_i$ which satisfy

$(**)$ $\mu \in I$, $\nu \in A_i$ and $\nu \perp \lambda$ whenever $\mu \perp \lambda$ and $\lambda \in A_i$ imply $\nu \in I$.

T and T^{-1} clearly preserve the property of being a closed left ideal and by the lemma they preserve $(**)$. Thus T maps $L^1(G_1)$ onto $L^1(G_2)$.

Now for $\mu \in L^1(G_i)$, the condition $\nu \in A_i$ and $\nu \perp \lambda$ whenever $\lambda \in A_i$ and $\mu \perp \lambda$ is equivalent to $\nu \ll \mu$. Clearly $\nu \ll \mu$ implies it, and conversely any ν satisfying it must be orthogonal to its singular part λ in its Lebesgue decomposition $\nu = f\mu + \lambda$ with respect to μ since $\lambda \in A_i$. So $L^1(G_i)$ is a closed left ideal satisfying $(**)$. Let $I \subseteqq A_i$ be any nontrivial closed left ideal satisfying $(**)$. Then I must contain a nonzero L^1 measure since $\alpha * \mu \in L^1$ and is nonzero for $\mu \neq 0$ in I and α is a suitable element in an L^1 approximate identity. The total variation of this measure is absolutely continuous with respect to it, hence in I. By convolving this with an appropriate L^1 approximation to a point

[1] I am indebted to George Reid for suggesting this lemma.

mass, we get a measure $\nu \in I$ strictly positive in a neighborhood of the identity (the convolution of an L^1 and an L^∞ function is continuous). But there is an L^1 approximate identity absolutely continuous with respect to ν, hence in I. Since I is a closed ideal, $L^1 \subseteqq I$.

Thus we have $(*)$ holding for all $\nu \in L^1(G_1)$. Let $\mu \in A_1$, and $\nu \in L^1(G_1)$. Then $\mu * \nu \in L^1(G_1)$ so

$$\iint f(\alpha(st))\chi(st)d\mu(s)d\nu(t) = T(\mu*\nu)(f) = (T\mu * T\nu)(f)$$
$$= \iint \chi(t) f(r\alpha t)dT\mu(r)d\nu(t)$$

so $(*)$ holds for μ and all functions in $C_0(G_2)$ of the form $\int f(r\alpha t)\chi(t)d\nu(t)$ where $f \in C_0(G_2)$ and $\nu \in L^1(G_1)$. This class of functions is dense in $C_0(G_2)$ since ν may be taken in an L^1 approximate identity. Thus $(*)$ holds for all $C_0(G_2)$ by continuity, which proves the theorem.

COROLLARY. *A locally compact group is determined by its measure algebra.*

This corollary was obtained independently by B. E. Johnson (Proc. Amer. Math. Soc. 1964). His results imply the main theorem under the hypothesis that each A_i contains all point masses.

BIBLIOGRAPHY

1. Irving Glicksberg, *Homomorphisms of certain algebras of measures,* Pacific J. Math. **10** (1960), 167-191.
2. J. G. Wendel, *On isometric isomorphism of groups algebras,* Pacific J. Math. **1** (1951), 305-311.

PRINCETON UNIVERSITY

PACIFIC JOURNAL OF MATHEMATICS
Vol. 15, No.1, 1965

CHARACTER SUMS AND DIFFERENCE SETS

RICHARD J. TURYN

This paper concerns difference sets in finite groups. The approach is as follows: if D is a difference set in a group G, and χ any character of G, $\chi(D) = \sum_D \chi(g)$ is an algebraic integer of absolute value $\sqrt{n}$ in the field of mth roots of 1, where m is the order of χ. Known facts about such integers and the relations which the $\chi(D)$ must satisfy (as χ varies) may yield information about D by the Fourier inversion formula. In particular, if $\chi(D)$ is necessarily divisible by a relatively large integer, the number of elements g of D for which $\chi(g)$ takes on any given value must be large; this yields some non-existence theorems.

Another theorem, which does not depend on a magnitude argument, states that if n and v are both even and a, the power of 2 in v, is at least half of that in n, then G cannot have a character of order 2^a, and thus G cannot be cyclic.

A difference set with $v = 4n$ gives rise to an Hadamard matrix; it has been conjectured that no such cyclic sets exist with $v > 4$. This is proved for n even by the above theorem, and is proved for various odd n by the theorems which depend on magnitude arguments. In the last section, two classes of abelian, but not cyclic, difference sets with $v = 4n$ are exhibited.

A subset D of a finite group G is called a *difference set* if every element $\neq e$ of G can be represented in precisely λ ways as $d_1 d_2^{-1}$, $d_i \in D$. If χ is any nonprincipal character of G, we must then have $|\sum_{d \in D} \chi(d)| = \sqrt{n}$, $n = k - \lambda$, where k is the order of D. We shall write $\chi(D)$ for $\sum_{d \in D} \chi(d)$ (as in [8]). If G is abelian and $|\chi(D)| = \sqrt{n}$ for some subset D and all nonprincipal characters of G, D is a difference set in G.

This work originated in a search for difference sets with G cyclic of order v, and the parameters related by $v = 4n$. Because in this case every divisor of n is a divisor of v, Hall's theorem on multipliers, [5], one of the main tools in the study of difference sets, cannot be applied. The method presented here is particularly suitable for computation of difference sets if v and n have common factors. It is roughly as follows: the numbers $\chi(D)$ are algebraic integers of absolute value $\sqrt{n}$ in the field of mth roots of 1, where m is the order of χ (as an

Received March 17, 1964. Presented to the American Mathematical Society, August 29, 1963. This is essentially a thesis presented to the Department of Mathematics at Harvard University in partial fulfillment of the requirements for the degree of Doctor of Philosophy. The research reported in this paper was partially supported by the Air Force Cambridge Research Laboratories, Office of Aerospace Research under Contract AF19 (628)-2479.

element of the character group of G). We use the known facts about such algebraic integers, together with elementary combinatorial information about these numbers which depends on their being sums of characters taken over the difference set, and the relations which must hold between the various character sums. We may then use the orthogonality of characters (Fourier inversion formula) to obtain information about the characteristic function of D.

The difference sets with $v = 4n$ correspond to (unnormalized) Hadamard matrices. The only known cyclic (i.e., with G cyclic) difference set of this type is the trivial one with $v = 4$. Although we did not succeed in proving that no such cyclic sets exist if $v > 4$, a number of nonexistence theorems are proved; these give bounds on the orders of the cyclic p-subgroups of G, where $p \mid (n, v)$. The proofs depend only on the existence of characters of certain orders.

In his survey of cyclic difference sets with $k \leqq 50$, [5], Hall had left twelve sets of values of (v, k, λ) undecided; it was not known whether a cyclic difference set with these values of (v, k, λ) existed. For all but one of these, $(v, n) > 1$. Nine of these were shown not to correspond to cyclic sets in [14]. Ten have since been shown not to correspond to cyclic sets by Mann ([8]). Of the twelve sets of values, one is left unresolved by [8] or [14], and it is shown here that it cannot correspond to an abelian set.

On the constructive side, we derive two classes of abelian, but not cyclic, difference sets, both with $v = 4n$. One class, for which $v = 36$, contains a set recently found by Menon [10]; the other class, for which $v = 4^t$, was suggested by one of the sets with $v = 36$.

Some of this work appeared in [13] and [14]. However, the use of the full force of Lemma 3 was suggested to me by my reading of Mann's paper [8]. I would like to express my gratitude to Professor Gleason for the large amount of time he spent reading this work; he pointed out a number of errors and is responsible for a great improvement in the quality of the exposition.

We assume throughout that the reader is familiar with cyclotomic fields (see e.g. [15]). We recall in particular the following facts:

(1) The field of mth roots of 1 is of degree $\phi(m)$ over Q (the field of rationals); thus the field of mnth roots of 1 is of degree $\phi(m)$ over the field of nth roots of 1 if $(m, n) = 1$. If $(m, n) = 1$, any $\phi(m)$ consecutive powers of ζ, a primitive mth root of 1, form an integral basis for the field of mnth roots of 1 over the field K of nth roots of 1; the Galois group of $K(\zeta)$ over K is isomorphic to the multiplicative group of integers relatively prime to $m \pmod{m}$. The automorphism σ_j which corresponds to j is defined by $\sigma_j(\zeta) = \zeta^j$ for $(j, m) = 1$. In particular, complex conjugation is σ_{-1}.

(2) If p is prime, the factorization of p in the field $Q(\zeta)$, ζ a primitive mth root of 1, is as follows: if $(p, m) = 1$, and we assume $4 \mid m$ if m is even, let σ_p be the automorphism given by $\sigma_p(\zeta) = \zeta^p$. Then if P is any prime ideal divisor of (p) (where (A) denotes the principal ideal generated by A) σ_p is a generator of the subgroup of automorphisms τ for which $\tau(P) = P$. The prime ideal divisors P_i of (p) are in one-to-one correspondence with the cosets of this subgroup, and $(p) = \pi P_i$. Thus if $(p, m) = 1$, (p) is not divisible by the square of any ideal $\neq(1)$. If $m = p^a n$, $a \geqq 1$, $(p, n) = 1$, and ζ is a primitive p^ath root of 1, then in $Q(\zeta)$ $(p) = (1 - \zeta)^\phi$, $\phi = \phi(p^a)$; ϕ always denotes the Euler function. In the field of mth roots of 1, $1 - \zeta$ factors just as p does in the field of nth roots of 1.

(3) If ζ is a root of 1, $\zeta \neq 1$, $1 - \zeta$ is a unit unless ζ is a primitive p^nth root of 1, p a prime, $n \geqq 1$, and then $1 - \zeta \mid p$. $p \mid 1 - \zeta$ only for $p = 2$, $\zeta = -1$. (A proof follows from $\prod_1^{m-1}(1 - \zeta^i) = m$, ζ a primitive mth root of 1, and the Mobius inversion formula.)

(4) Suppose A and B are algebraic integers in a cyclotomic field, $|A| = |B|$ and $(A) = (B)$. Then $A/B = w$ is a root of 1. This follows from the theorem of Kronecker which asserts that an algebraic integer all of whose conjugates have absolute value 1 are roots of unity. The fact that $|\sigma w| = 1$ for any automorphism σ follows from the lemma below (with $m = 1$).

LEMMA 1. *If* $|w|^w \in Q$ *for some integer* $m \geqq 1$, *and* $\sigma c(w) = c\sigma(w)$, *where c denotes complex conjugation, then* $|w| = |\sigma(w)|$.

For

$$|w|^{2m} = w^m(c(w^m)) \in Q .$$

Therefore

$$\begin{aligned} |w|^{2m} &= \sigma(w^m c(w^m)) \\ &= \sigma(w)^m \sigma(c(w^m)) \\ &= \sigma(w)^m c(\sigma(w)^m) \\ &= |\sigma(w)|^{2m} . \end{aligned}$$

We use the following notations: if G is a group, p a prime, $\sigma_p(G) = a$ if a is the largest integer m such that G has a character of order p^m. If n is an integer, $p^a \| n$ if $p^a \mid n$, $p^{a+1} \nmid n$. Z_n is the cyclic group of order n. w, with or without subscripts, will denote a root of 1. χ_0 always denotes the principal character of G, i.e., $\chi_0(g) = 1$ for all $g \in G$. If a and b are integers, we say that a is semiprimitive mod b if there exists an integer c such that $a^c \equiv -1 \pmod b$. a is self-conjugate mod b if all prime divisors p of a are semiprimitive mod bp^{-e_p}, where $b = \Pi p^{e_p}$.

Difference sets. A (v, k, λ) configuration is a set of v points and b subsets, called blocks, each containing k points, such that the intersection of any two distinct blocks consists of λ points. Defining n to be $k - \lambda$, we have also $k^2 - \lambda v = n$. If M is the incidence matrix of the configuration ($m_{ij} = 1$ if point i is in set j, $m_{ij} = 0$ otherwise), an equivalent definition is that

$$M'M = nI + \lambda J ,$$

where J is the matrix with all entries $= 1$. Since $\{M' - (\lambda J/k)\}M = nI$, $M\{M' - (\lambda J/k)\} = nI$. The entries m_{ij} of M are all 0 or 1, hence $m_{ij}^2 = m_{ij}$, and thus the ii term of the last equation shows that $\sum_j m_{ij} = k$, and therefore $MM' = nI + \lambda J = M'M$.

Assume a (v, k, λ) configuration has a regular transitive group of automorphisms; that is, assume there exists a transitive group G of order v of permutations of the v points, each permutation taking blocks into block; if D is the subset of G of those σ for which $\sigma(P) \in B$, where P is a fixed point and B a fixed block, any element $\alpha \neq e$ of G can be represented in precisely λ ways as $\tau\sigma^{-1}$, with τ, σ in D. We must show $\alpha B \cap B$ contains precisely λ points. This will happen unless $\alpha B = B$, since αB is a clock of the design. So there are at least λ pairs for which $\tau\sigma^{-1} = \alpha$ with $\tau, \sigma \in D$. But since $k(k-1) = \lambda(v-1)$ and there are $k(k-1)$ ordered pairs τ, σ and $v - 1$ elements in G not the identity, we cannot have $\alpha B = B$ for $\alpha \neq e$ (cf. [1]). Replacing P by $\tau_1 P$ and B by $\tau_2 B$ replaces D by $\tau_2 D \tau_1^{-1}$.

Let Y_σ be the characteristic function of D, $y_\sigma = 1$ for $\sigma \in D$, $y_\sigma = 0$ for $\sigma \notin D$.

We then have

$$\sum_{\sigma \in G} y_0 y_{\tau\sigma} = \lambda \qquad \tau \neq e$$

as an equivalent formulation of the condition that $D \cap \tau D$ have precisely λ points for all τ.

A subset D of a group G is called a *difference set* if it satisfies the above conditions; D is cyclic or abelian if G is. The sets σD as σ ranges through G form the blocks of a (v, k, λ) configuration. The complement of a difference set is a difference set, and hence we may assume $k \leqq v/2$.

We shall always assume the difference set is nontrivial, i.e., $1 < k < v - 1$, from which it follows that $v \geqq 7$.

Suppose G is abelian. Let f be a function defined on G, χ a character on G, and let

(1) $$\hat{f}(\chi) = \sum_{g \in G} f(g)\chi(g)$$

The set of equations

(2) $$\sum_{g \in G} f(g) f(hg) = c(h)$$

is equivalent to

(3) $$\hat{f}(\chi)\hat{f}(\bar{\chi}) = \sum_{h \in G} c(h)\chi(h)$$

or if f is real-valued,

$$|\hat{f}(\chi)|^2 = \sum_{h \in G} c(h)\chi(h) .$$

This shows D is a difference set, with parameters $v, k, \lambda, n = k - \lambda$, if and only if

(4) $$\sum_{G} y_g = k$$

$$\left|\sum_{G} y_g \chi(g)\right| = \sqrt{n} \qquad \text{for all } \chi \neq \chi_0 .$$

It also shows that $|\hat{f}(\chi)| = c$ for all χ if and only if $\sum_G f(g)f(gh) = 0$ for $h \neq e$.

Finally, D is a (v, k, λ) difference set if and only if we have, in the group algebra of G,

(5) $$\Big(\sum_{D} g\Big)\Big(\sum_{D} g^{-1}\Big) = ne + \lambda G , \qquad G = \sum_{G} g .$$

The orthogonality relations for characters imply that if f and $\hat{f}$ are related by (1), we must have

(6) $$f(g) = \frac{1}{v}\sum_{\chi} \hat{f}(\chi)\bar{\chi}(g) .$$

Let f be a function on a group G and restrict χ to a subgroup $\hat{H}$ of the character group. If H is the kernel of $\hat{H}$, i.e., all h such that $\chi(h) = 1$ for all χ in $\hat{H}$, we may define a function F on G/H by summing f over the cosets of H and apply the preceding formulae to F.

We note the following special cases:

(1) Let χ be a character of order p^b, p prime, $b \geqq 1$, ζ a primitive p^bth root of 1, f a function on G with values in a field K such that $[K(\zeta) : K] = \phi(p^b) = \phi$. Let $F_i = \Sigma f(g)$, over all g with $\chi(g) = \zeta^i$, and let $S_i = \sum_{j=0}^{p-1} F_{i+qj}$, $q = p^{b-1}$. Then if

$$\sum_1^{p^b} F_i\zeta^i = \sum_1^{\phi} A_i\zeta^i$$

with $A_i \in K$, so that the A_i are uniquely determined, we have

$$F_i = A_i + \frac{1}{p}\Big(S_i - \sum_{j=0}^{p-2} A_{i+qj}\Big) \qquad 1 \leqq i \leqq \phi$$

(7)

$$F_i = \frac{1}{p}\Big(S_i - \sum_{j=0}^{p-2} A_{i+qj}\Big) \qquad \phi < i \leqq p^b$$

the formula whose repeated application is equivalent to the inversion formula (6) for a cyclic group.

(2) Let $f(g)$ be as before, χ_1 and χ_2 two characters of order p which generate a subgroup of order p^2, ζ a fixed primitive pth root of 1.

We let $F_{ij} = \sum f(g)$ over all g with $\chi_1(g) = \zeta^i$, $\chi_2(g) = \zeta^j$. Let $\sum_G f(g) = S$, and let

$$\sum_{m=i+kj} F_{ij} = S_{m,k}$$

for $k = 1, \cdots p, \infty(1 + \infty j = j)$. The $S_{m,k}$ can be determined by (7) from S and the sum

$$\sum_G f(g)(\chi_1\chi_2^k)^i(g) .$$

Then

(8) $$F_{ij} = \frac{1}{p}\Big(\sum_{m=i+kj} S_{m,k} - S\Big).$$

(3) If $f(g)$ is an algebraic integer for all $g \in G$ and χ ranges over a coset of a subgroup $\hat{H}$ of the character group of G, order of $\hat{H} = m$, then

(9) $$m \Big| \sum_{\chi} f(g)\chi(g) .$$

For if χ_1 is a fixed character in the coset, the sum in question is $\sum_{\chi\in\hat{H}} f(g)\chi(g)\chi_1(g) = f(g)\chi_1(g)\sum_{\chi\in\hat{H}}\chi(g)$ and $\sum_{\chi\in\hat{H}}\chi(g)$ is m if $\chi(g) = 1$ for all $\chi \in H$, 0 otherwise.

If H is a subgroup of G, we will always denote by $\hat{H}$ the set of all characters χ such that $\chi(h) = 1$ for all $h \in H$, and vice versa.

If G is abelian, the group algebra of G is a direct sum of fields; in fact the elements $\sum_G g\chi(g^{-1})$ are eigenvectors for all the elements of the regular representation of G. The eigenvalues of the incidence matrix of a (v, k, λ) configuration have absolute value $\sqrt{n}$, except for one which is k; equations (4) are an explicit restatement of this fact for abelian difference sets.

Call a subset D of a group G *nonperiodic* if $D = Da$ implies $a = e$. A difference set is nonperiodic. A *multiplier* of D is an automorphism σ of G such that $\sigma(D) = Da_\sigma$ for some a_σ in G. (a_σ is unique if D is nonperiodic.) When G is cyclic, all the automorphisms of G are of the form $\sigma(g) = g^m$, (with m relatively prime to the order of G) and the integer m is called a multiplier of D if $\sigma(D) = Da$. The above definition is the obvious generalization to noncyclic groups of the notion of multiplier (see [1]).

LEMMA 2. *The multipliers of D are a subgroup M of the automorphism group of G; $a_{\tau\sigma} = a_\tau \tau(a_\sigma)$ for $\sigma, \tau \in M$. σ leaves a translate Db of D fixed if and only if $a_\sigma = b\sigma(b)^{-1}$.*

The lemma is obvious.

COROLLARY. *If G is of prime order p, every set $D \subseteqq G$ has a translate which is left fixed by all the multipliers of D.*

If G is of prime order, written additively, the only periodic subset of G is G. Since the multipliers are a cyclic group, we may pick a generator σ of the multiplier group. If this is given by $\sigma(i) \equiv ki$ (mod p), $1 - k$ has a multiplicative inverse mod p, so if $\sigma D = Da$, $(1 - k)b = a$, then $\tau(Db) = Db$.

The quadratic residues modulo any prime $\equiv -1(4)$ form a difference set. In [7], E. Lehmer considered the existence of other difference sets defined by power residues (mod v) if v is an odd prime. In particular, it was shown in [7] that if $v = ef + 1$, a prime, and if the eth powers, or eth powers and 0, form a difference set (mod v), then the multipliers are precisely the eth powers. The corollary proves a more general statement.

THEOREM 1. *Let D be a subset of Z_p, p a prime, which is a union of m multiplicative cosets of the eth powers, plus possibly 0. If e is the least number for which this is true, and $e > 1$, the eth powers are all the multipliers of D. If D is a difference set and $q \mid (m, e)$ then $q^a \mid e$ implies $q^a \mid m$.*

Proof. Replace D by a translate left fixed by all the multipliers. Lemma 2 shows that if D has a nontrivial multiplier there is a unique translate of D which it leaves fixed. Thus $e > 1$ shows D must be a set of multiplicative cosets of the set of multipliers, plus possibly 0. Since the eth powers are certainly multipliers, the first statement follows from the minimality of e. If D is a difference set, we may assume by taking the complement of D that $0 \notin D$. Then $k = mf$, and $mf(mf-1) = \lambda ef$, $\lambda = m(mf-1)/e$. Since $mf-1$ is prime to m, $q^a \mid e$, $q \mid m$ implies $q^a \mid m$.

In the second part of the theorem, we did not have to assume e minimal. For example, if $e = 4$, $m = 2$ is impossible; in particular the squares cannot form a difference set mod a prime of the form $4k+1$. Hall [5] has constructed a family of difference sets with $m = 3$, $e = 6$.

Character sums. We first prove a well known theorem of a type originally proved [2] for $(v, k, 1)$ configurations (finite projective planes). (See [3], [4].) The proof given is very direct and yields more in the special case of abelian difference sets.

LEMMA 3. *If η is an algebraic integer such that $|\eta|^2 = n$ for some integer n and $(\eta) = \Pi P_i^{a_i}$, P_i prime ideals, then $\Pi(P_i\bar{P}_i)^{a_i} = (n)$. If η belongs to the field of mth roots of 1 and p is a prime divisor of n semiprimitive mod m then p occurs to an even power in n, say $p^{2b} \| n$, and $p^b \mid \eta$.*

Proof. The first statement is obvious since $\eta\bar{\eta} = n$. If p is semiprimitive mod m the prime ideal divisors of (p) in the field of mth roots of 1 are invariant under complex conjugation. $(p, m) = 1$ implies that (p) is not divisible by the square of any prime ideal, which proves the lemma (cf. [8]).

We remark at this point that, with the notations of the lemma, if $d^2 \mid n$ and d is self-conjugate mod m then $d \mid \eta$. For if $p \mid (d, m)$, $p^a \| m$, (p) is a power of a single prime ideal in the field of p^ath roots of 1, and this ideal factors into distinct prime ideals invariant under complex conjugation in the field of mth roots of 1.

THEOREM 2. *Let G be abelian, D a v, k, λ difference set in G. If v is even, n is a square. If p is a prime which divides n to an odd power and $q \neq p$ is a prime divisor of v, p has odd order in the multiplicative group mod q.*

(The conclusions that v even implies n is a square, and that p is a quadratic residue mod q are known for arbitrary v, k, λ configurations.)

Proof. If v is even, G has a character χ of order 2. $|\chi(D)|^2 = n$ implies n is a square, since $\chi(D)$ is rational. To prove the second part, let χ be a character of order q. Since $|\chi(D)|^2 = n$, Lemma 3 shows that p cannot be semiprimitive mod q; the semiprimitive numbers are precisely those which have even order in the multiplicative group mod q.

THEOREM 3. *If v is a prime and D is a difference set mod v, inversion is not a multiplier of D. If $v = ef + 1$ and D is a union of m multiplicative cosets of the set of* eth *powers mod v, plus possibly* 0, *then f is odd.*

REMARK. Inversion is not a multiplier under much more general conditions (see [8]; never if G is cyclic). The conclusion that f is odd is proved for the eth powers, and the eth powers and 0, in [7].

Proof. Replacing D by its complement if $0 \in D$, we may assume $0 \notin D$. Now replace D by a translate left fixed by all the multipliers. Since inversion is a multiplier $\chi(D)$ must be real for all χ, and thus $= \pm\sqrt{n}$ if $\chi \neq \chi_0$. This shows $\chi(D)$, which lies in the subfield of degree 2 over Q is left invariant by the subgroup of index 2 of the Galois group of the field of vth roots of 1, i.e., by all the automorphisms of the form $\sigma(\zeta) = \zeta^r$, ζ a primitive vth root of 1, r a quadratic residue mod v. Therefore $\chi(D) = \sum_1^{v-1} y_i\zeta^i = \sum_1^{v-1} y_i\zeta^{ri}$, and $\sum_1^{v-1}(y_i - y_{ri})\zeta^i = 0$. Therefore $y_i = y_{ri}$ for all i and any quadratic residue r. This shows D consists of the set of quadratic residues or nonresidues (if D is nontrivial). But this can happen only for v of the form $4k - 1$, and then inversion is not a multiplier since it takes the residues into the nonresidues.

If D is a union of multiplicative cosets of the set of eth powers mod v and f is even, -1 is an eth power and hence a multiplier, which we have just shown is impossible. This proves the last part of the theorem.

If σ is a multiplier of D, $\sigma D = Da$, and we have $\chi(\sigma D) = \chi(a)\chi(D)$; in particular the factorization of the ideals $(\chi(D))$ is unchanged if we replace D by σD. The following theorem is a partial converse.

THEOREM 4. *Suppose D is a difference set in G, G abelian, and σ is an automorphism of G such that the ideals $(\chi(D))$ and $(\chi(\sigma D))$ are the same for each character χ. Then if there exists m such that $m \mid n$, $m > \lambda$ and $(m, v) = 1$, σ is a multiplier of D.*

REMARK. We give below an example of a difference set in which every automorphism leaves the principal ideals generated by the character sums invariant, but the multiplier group has order 2 while the automorphism group has order 96.

Proof. The theorem follows from the generalization of Hall's theorem ([5]; see also [8], [9], [12]). We repeat the proof, essentially the one in [8]. In the group algebra of G, we let $H = D^{-1}(\sigma D) - \lambda G$ (where $D^{-1} = \sum_{g \in D} g^{-1}$, $\sigma D = \sum_{g \in D} \sigma(g)$). Each character of G extends to a homomorphism of the group algebra, and $\chi(H)$ is $nw(\chi)$, with $w(\chi)$ a root of 1 for every character χ. If $\chi = \chi_0$, this follows from the formula $k^2 - \lambda v = n$. If $\chi \neq \chi_0$, we have $\chi(H) = \chi(D^{-1})\chi(\sigma D) = \overline{\chi(D)}\chi\sigma(D)$. Since $(\chi(\sigma D))$ has the same factorization as $(\chi(D))$, and $\chi(D)\bar{\chi}(D) = n$, we conclude that $(\chi(H)) = (n)$. Since $|\chi(H)| = n$, $\chi(H) = nw(\chi)$, with $w(\chi)$ a root of 1.

By the inversion formula (6), if $H = \Sigma h_g g$, we have

$$g_h = \frac{1}{v}\sum_{\chi}\chi(H)\bar{\chi}(g) = \frac{n}{v}\sum_{\chi} w(\chi)\bar{\chi}(g) .$$

Since $m \mid n$, $(m, v) = 1$, we conclude $m \mid h_g$. Since $m > \lambda$ and $h_g \geqq -\lambda$ by the definition of H, $h_g \geqq 0$ for all g. We have seen before that $|\chi(H)| = n$ for all characters χ is equivalent to the assertion $\sum_g h_g h_{gs} = 0$ for all $s \neq e$. Therefore only one $h_g \neq 0$, since all $h_g \geqq 0$. Clearly that h_g is n. Now

$$H + \lambda G = D^{-1}\sigma(D) = \lambda G = ng_0 .$$

Multiplying by D, we get

$$(ne + \lambda G)\sigma D = k\lambda G + ng_0 D$$
$$ne(\sigma D) = ng_0 D$$

so $\sigma D = g_0^{-1}D$.

We note here a consequence of (9).

THEOREM 5. *Assume D is a difference set in G and that $(\chi\psi(D)) = (\psi(D))$ for some nonprincipal character ψ and all characters χ in a group H. If the order of H is relatively prime to n and the order of ψ, there exists g in G such that $\chi\psi(Dg) = \psi(Dg)$ (or $\chi(g)\chi\psi(D) = \psi(D)$) for all χ in H.*

Proof. We shall first prove the theorem for a cyclic group of prime power order. Let χ be a generator of H, of order p^r; we may assume r is the least integer for which the theorem is not known. Assume D is translated so that $\chi^{jp}\psi(D) = \psi(D)$ for all j, and let $\chi\psi(D) = w_1\zeta\psi(D)$, where ζ is a p^rth root of 1 and w_1 is a root of 1 of order prime to p. Then $\chi^j\psi(D) = w_1\zeta^j\psi(D)$ if $(j, p) = 1$ because $\chi^j\psi(D)$ is the conjugate of $\chi\psi(D)$ under the automorphism which is the identity on the field of roots of 1 of order prime to p (to which w_1 and $\psi(D)$ belong) and takes p^rth roots of 1 into their jth powers. Therefore $\Sigma\chi^j\psi(D) = w_1\psi(D)tr\zeta$, the sum over all j with $0 < j < p^r$, $(j, p) = 1$. $tr\zeta$ is $\phi(p^r)$ if $\zeta = 1$, $-p^{r-1}$ if ζ is a primitive pth root of 1, and 0 otherwise. Since $\chi^j\psi(D) = \psi(D)$ if $p \mid j$, we get

$$p^r \Big| \sum_1^{p^r} \chi^j \psi(D) = p^{r-1}\psi(D) + w_1\psi(D) tr\zeta$$

$(p, n) = 1$, $\psi(D) \mid n$ implies $p^r \mid p^{r-1} + w_1 tr\zeta$, and therefore $tr\zeta \neq 0$, $\zeta^p = 1$, and $w_1 = 1$ (since $p \mid 1 - w_1$, and $p = 2$ implies $w_1 \neq -1$). If $\zeta = 1$, the theorem is proved; if $\zeta \neq 1$, take any g such that $\chi(g) = \zeta^{-1}$. Then $\chi^{jp}\psi(Dg) = \psi(Dg)$ because $\zeta^p = 1$, and $\chi^j\psi(Dg) = \psi(Dg)$ for $(j, p) = 1$ because $\chi\psi(Dg) = \psi(Dg)$.

An arbitrary group H may be expressed as a direct product of cyclic groups H_i, with generators χ_i. It is clear from the above proof that we can find g_i in G such that $\chi_i^j\psi(Dg_i) = \psi(Dg_i)$ for all j, and $\chi_j(g_i) = 1$ for $i \neq j$, since the construction of g_i involves only the value of $\chi_i(g_i)$. Then if $g = \Pi g_i$, we have $\chi_i^j\psi(Dg) = \psi(Dg)$ for all i, j. Replacing D by Dg for simplicity, we shall now show that $\chi\psi(D) = \psi(D)$ for all χ in H. Let F be the set of all χ with this property. If χ_1, χ_2 are elements of F of order p^r, p^s, respectively, and generate a group of order p^{r+s}, p prime, $r \geqq s \geqq 1$, we show that this group is contained in F. We may assume that r rnd s (and the χ_i) are picked so that $\chi_1^i\chi_2^j \in F$ if $p \mid ij$, i.e., we take a minimal group for which the theorem is not known. The $\phi(p^r)\phi(p^s) = q$ characters $\chi_1^i\chi_2^j$ with $(ij, p) = 1$ fall into $\phi(p^s)$ equivalence classes, each consisting of all χ^m, $(m, p) = 1$ for some χ. The preceding result shows that $\sum_{(m,p)=1} \chi^m\psi(D) = A\psi(D)$, with A one of $\phi(p^r)$, $-p^{r-1}$, or 0. Now

$$p^{r+s} \mid \sum \chi_1^i\chi_2^j\psi(D) = (p^{r+s} - q)\psi(D) + \psi(D) \sum A$$

the first term being the sum over all $\chi_1^i\chi_2^j$ with $p \mid ij$, $\psi(D)\sum A$ being the sum over the $\phi(p^s)$ equivalence classes. Since $\psi(D) \mid n$, $(p, n) = 1$, we have $p^{r+s} \mid \Sigma A - q$. Since $0 \geqq \Sigma A - q \geqq -p^r\phi(s) > -p^{r+s}$ because $-p^{r-1} \leqq A \leqq \phi(p^r)$ for all A, we must have $A = \phi(p^r)$ for all A, which means all $\chi_1^i\chi_2^j \in F$. We now conclude that F contains all characters of prime power order, by induction on the number of components.

An arbitrary character χ in H may be expressed as $\prod_1^r \chi_i$, with χ_i of order q_i, the q_i distinct prime powers. We prove by induction on r that $\chi \in F$. We have seen that if $r = 1, \chi \in F$. If the theorem is true for $r - 1$, we have $\chi\psi(D) = \zeta_1\bar{\chi}_1\chi\psi(D)$ with ζ_1 a q_1 root of 1, by the first part of the theorem, and $\bar{\chi}_1\chi\psi(D) = \psi(D)$ by the inductive assumption. But $\chi\psi(D) = \zeta\chi_1\psi(D)$ by applying the theorem for $r - 1$, with $\chi_1\psi$ playing the role of ψ, and ζ a $\prod_2^r q_i$ root of 1. Since $\chi_1\psi(D) = \psi(D)$ we conclude $\zeta_1 = \zeta$, which implies $\zeta_1 = \zeta = 1$ and $\chi \in F$.

Abelian Hadamard matrices. An Hadamard matrix is a square matrix H of order h with entries ± 1, any two distinct rows of which are orthogonal, i.e., such that $HH' = hI$. An Hadamard matrix may be normalized to have first row and column consisting of just $+1$'s.

The remaining matrix of order $h - 1$ has the property that the dot product of any two rows is -1, and that the sum of the entries in any row is -1.

Let M be the incidence matrix of a (v, k, λ) design, and J the matrix with all entries $=1$. The matrix $2M - J$ has entries 1 where M has entries 1, -1 where M has entries 0 and

$$(2M - J)(2M - J)' = 4MM' - 4kJ + vJ = 4nI + J(v - 4n) .$$

Thus the dot product of two distinct rows of $2M - J$ is $v \quad 4n$. It is clear that the matrix of order $h - 1$ derived from an Hadamard matrix of order $h > 2$ by normalizing the first row and column is equivalent to the incidence matrix of (v, k, λ) configuration with $v + 1 = 4n$, $k = 2n - 1$, $\lambda = n - 1$. Several classes of abelian difference sets with these parameters are known.

However, the question of the existence of difference sets whose incidence matrix generates an Hadamard matrix without the normalization has not been considered extensively in the literature. By the preceding, these are defined by the condition $v = 4n$. In a recent paper [10] Menon constructed two such difference sets (one for the direct product of two dihedral groups of order six, the other one for the abelian group $Z_6 \times Z_6$ and noted the product theorem (Lemma 4 below). In [9] Menon constructed such sets for the direct product of an even number of copies of Z_2. The connection with Hadamard matrices is mentioned in neither paper. The author's interest in the question is partly due to the following theorem ([11]): if $x_i = \pm 1$, $1 \leqq i \leqq v$ and $|\sum_{i=1}^{v-j} x_i x_{i+j}| \leqq 1$ for all $j > 0$, then if v is odd, $v \leqq 13$; if v is even and >2, the i for which $x_i = 1$ (or -1) form a difference set (mod v) with $v = 4n$. (The problem partly answered by [11] arose in radar design.)

By an abelian Hadamard matrix we mean the Hadamard matrix derived from a difference set in an abelian group with $v = 4n$. (Then $n = N^2$, $v = 4N^2$, $k = N(2N - 1)$, $\lambda = N(N - 1)$, if we normalize so that $2k < v$. We shall call such sets H sets for brevity. Note that we have the formula $k = (v - \sqrt{v(v - 4n) + 4n})/2$; v even implies n must be a square, and therefore the choice $v = 4n$ leads to a simple family of values for v, k, λ.

LEMMA 4. *Let D_i be H sets in G_i, $i = 1, 2$. Then $(D_1, \bar{D}_2) \cup (\bar{D}_1, D_2)$ is an Hadamard difference set in $G_1 \times G_2$. Conversely, if D_i is a difference set in G_i $(D_1, \bar{D}_2) \cup (\bar{D}_1, D_2)$ is a difference set in $G_1 \times G_2$ if and only if both D_i are H sets. ($\bar{D}_i$ denotes the complement of D_i.)*

The first statement follows from the fact that the direct product of two Hadamard matrices is an Hadamard matrix; the second from the fact that if $A_i A_i' = v_i I + (v_i - 4n_i)(J - I)$ for $i = 1, 2, n_i \neq 0$, then $(A_1 \times A_2)(A_1' \times A_2') = (v_1 I + (v_1 - 4n_1)(J - I)) \times (v_2 I + (v_2 - 4n_2)(J - I))$ is of the form $vI + cJ$ only if $v_i - 4n_i = 0$ for $i = 1, 2$.

A more involved proof is given in [10].

Nonexistence theorems. In this chapter we shall prove several theorems of the following general nature: if D is a (v, k, λ) difference set in G and $(n, v) > 1$, there are bounds on the orders of characters of G. For example, if $2 \mid (n, v)$, we can prove that under suitable assumptions $\sigma_2(G)$ must be less than the exponent of 2 in v; in particular, G cannot be cyclic. Our main interest is the nonexistence of H sets; we use the previous notations: $v = 4N^2$, $n = N^2$.

We remark that $p \mid (n, v)$ implies that $p \mid k, \lambda$ and that $(k, v)^2 \mid n$, since $n = k - \lambda$ and $k^2 - \lambda v = n$. We also note that if p is odd and $b \geqq 1$, q semiprimitive mod p implies that q is semiprimitive mod p^b.

If χ is a character of G of order s and $D \subseteqq G$, ζ a primitive sth root of 1, then $\chi(D) = \sum_1^s Y_i \zeta^i$, where Y_i is the number of elements g in D such that $\chi(g) = \zeta^i$. Thus $0 \leqq Y_i \leqq v/s$. The proofs of the first two theorems below depend on this statement about the magnitude of the Y_i and would have direct analogues if the y_g were not restricted to be 0 or 1, (i.e., if we allowed multiplicities in D).

THEOREM 6. *Let D be any subset of G such that $m \mid \chi\chi_1(D)$ for all characters χ in a group $\hat{H}$ of order v_2, $(m, v_2) = 1$, with χ_1 a character of order $v_1 > 1$, $\chi_1^j \notin \hat{H}$, for $1 \leqq j < v_1$, and where not all $\chi\chi_1(D) = 0$ for $\chi \in \hat{H}$. Then $2^{r-1}v \geqq mv_1v_2$, where r is the number of distinct prime divisors of v_1. If $v_1 = 1$, $v \geqq mv_2$.*

Proof. The inversion formula (6) shows that each of the v_2 sums $\Sigma y_g \chi_1(g)$ taken over a coset of the kernel H of $\hat{H}$ is divisible by m, since the sum over Hh is

$$\frac{1}{v_2} \sum_{\chi \in \hat{H}} \bar{\chi}(h)\chi\chi_1(D)$$

and $m \mid \chi\chi_1(D)$ for all χ in $\hat{H}$, $(m, v_2) = 1$. Not all these v_2 sums are 0 since then, by another application of (6), all $\chi\chi_1(D)$ would be. Let S_0 be one of these v_2 sums which is not 0. If $v_1 = 1$, S_0 is a sum of the y_g over a coset of H, $S_0 \neq 0$, and $m \mid S_0$. Thus $S_0 \geqq m$, and since $v/v_2 \geqq S_0$, $v \geqq mv_2$.

For the rest of the theorem, we shall require the following lemma:

LEMMA 5. *Let G be a finite cyclic group, f a function on G with integral values, and χ a generator of the character group of G. Assume $m \mid \hat{f}(\chi) = \sum_{g \in G} f(g)\chi(g)$, $\hat{f}(\chi) \neq 0$. Let r be the number of* *distinct prime divisors of the order of G. If $0 \leqq f(g) \leqq b$ for all g, $m \leqq 2^{r-1}b$; if $|f(g)| \leqq b$, $m \leqq 2^r b$.*

Proof. Let $G = G_1 \times H, G_1$ cyclic of order $q = p^S$, order of H prime to q. Let t be a generator of G_1; $\chi(g) = \zeta$ is a primitive qth root of 1. Then $\hat{f}(\chi) = \sum_1^q F(i)\zeta^i$, with $F(i) = \sum_{h \in H} f(ht^i)\chi(h)$, $m \mid \hat{f}(\chi) = \sum_1^\phi (F(i) - F(j(i)))\zeta^i$, where $\phi = \phi(q)$ and $\phi < j(i) \leqq q$, $j(i) \equiv i (\bmod p^{S-1})$. Since the order of H is relatively prime to q, ζ is of degree ϕ over the field generated by the $F(i)$ over Q. Thus $m \mid F(i) - F(j(i))$ for all i, and at least one of the $F(i) - F(j(i))$ is not 0; we pick one such index i. Now if $r = 1$ the lemma follows because $F(i) - F(j(i))$ is an integer divisible by m and bounded by b if $0 \leqq f(g) \leqq b$, by $2b$ if $|f(g)| \leqq b$. If $r > 1$, H is a cyclic group whose order has $r - 1$ distinct prime factors, and χ restricted to H is a generator of the character group of H. But $m \mid F(i) - F(j(i)) = \sum_{h \in H} (f(ht^i) - f(ht^{j(i)}))\chi(h)$, and the lemma follows by induction on r since now $|f(ht^i) - f(ht^{j(i)})| \leqq b$ if $0 \leqq f(g) \leqq b$, $\leqq 2b$ if $|f(g)| \leqq b$.

Returning to the proof of the theorem with $v_1 > 1$, we pick S_0 as before. We may write $S_0 = \sum_1^{v_1} Y_i\zeta^i$, with ζ a primitive v_1th root of 1 and Y_i the number of elements g of D in the chosen coset of the kernel of the group generated by $\hat{H}$ for which $\chi_1(g) = \zeta^i$. Since χ_1 and $\hat{H}$ generate a group of order $v_1 v_2$, there are $v/v_1 v_2$ such elements in G, and therefore at most that many in D. Thus $0 \leqq Y_i \leqq v/v_1 v_2$, and Lemma 5 implies that $m \leqq 2^{r-1}(v/v_1 v_2)$, which proves the theorem.

This proof depends on the simple structure of the irreducible polynomial for a p^mth root of 1.

COROLLARY 1. *Let D be a difference set in a group G which has a character of order v_1. If $m^2 \mid n$ and m is self conjugate* mod v_1, *then $v_1 m \leqq 2^{r-1}v$, where r is the number of distinct prime divisors of (m, v_1).*

If D is a difference set $\chi(D) \neq 0$ for all χ. Let χ be a fixed character of order v_1, and let $\chi = \chi_1\chi_2$, where the order of χ_1 is the product of the distinct prime power divisors q_i of v_1 for which $(q_i, m) > 1$, and the order of χ_2 is relatively prime to the order of χ_1. If follows from the remarks after Lemma 3 that $m \mid \chi(D)$ for any character χ of order dividing v_1, and thus in particular $m \mid \chi_2^j\chi_1(D)$ for all j. Theorem 6 shows that $2^{r-1}v \geqq mv_1$.

In [5], Hall listed twelve sets of (v, k, λ) with $k \leqq 50$ for which the existence of a cyclic difference set had not been decided. The theorems in [8], [13], [14] showed there were no cyclic difference sets for all the sets of (v, k, λ) with the exception of (120, 35, 10).

As an example of the above corollary, we see that there is no abelian difference set with the parameters (120, 35, 10). For an abelian group of order $120 = 2^3 \cdot 3 \cdot 5$ must have a character of order 30. Since $n = 25$ and $5 \equiv -1 \pmod 6$, the existence of such a difference set would imply $30 \cdot 5 \leqq 120$ by the corollary ($m = 25$, $v_1 = 30$).

COROLLARY 2. *There is no cyclic H set if N is a prime power. If an H set exists with $N = 2^a (v = 2^{2a+2})$ we must have $\sigma_2(G) \leqq a + 2$. If and H set exists with $N = 3^a$, $\sigma_3(G) \leqq a + 1$; if we assume also that $\sigma_2(G) \geqq 1$, we can conclude $\sigma_3(G) \leqq a$. If an H set exists with $N = p^a$, p a prime $\geqq 5$, $\sigma_p(G) \leqq a$.*

Proof. If there is an H set with $N = 2^a$, $\sigma_2(G) = b$, put $v_1 = 2^b$, $m = N$ in Theorem 6: we conclude $2^{2a+2} \geqq 2^{a+b}$, or $b \leqq a + 2$. If $N = p^a$, p an odd prime, and $\sigma_p(G) = b$ we conclude similarly $4p^{2a} \geqq p^{a+b}$, or $p^b \leqq 4p^a$. Thus $p^{b-a} \leqq 4$, so $b \leqq a$ if $p \geqq 5$, $b \leqq a + 1$ if $p = 3$. If also $\sigma_2(G) \geqq 1$ (as when G is cyclic) there is a character of order 2 and we can put $v_1 = p^b$, $m = N$, $v_2 = 2$ in Theorem 6: we get $4p^{2a} \geqq 2p^{a+b}$, or $p^b \leqq 2p^a$, and $b \leqq a$ for $p > 2$.

COROLLARY 3. *If there is an H set with $N = p^a M_1 M_2$, with M_1 self conjugate* mod p^b *and $p^{b-a} > 4M_1 M_2^2$, then $\sigma_p(G) < b$.*

Proof. Apply Theorem 6 with $v_1 = p^b$, $m = p^a M_1$, $v_2 = 1$; we conclude $4p^{2a} M_1^2 M_2^2 \geqq p^{a+b} M_1$.

COROLLARY 4. *There is no cyclic H set if $N = p^a M_1 M_2$, p odd, M_1 self conjugate* mod p, *and $p^a > 2M_1 M_2^2$. If p and all prime divisors of M_1 are of the form $4k - 1$ there is no cyclic H set if $p^a > M_1 M_2^2$.*

Proof. The first statement follows from Theorem 6 with $v_2 = 2$, $v_1 = p^{2a}$, $m = p^a M_1$. For the second statement we first make the following observation: if $t^r \equiv -1 \pmod A$, $t^s \equiv -1 \pmod B$, then t is semiprimitive mod AB if and only if the same power of 2 divides both r and s. If $p = 4k - 1$, $(1/2)\phi(p^{2a}) = r$ is odd. Now if $q \equiv -1 \pmod 4$, $q \equiv -1 \pmod p$, we conclude $q^r \equiv -1 \pmod{4p^{2a}}$. Now put $v_2 = 4$, $v_1 = p^{2a}$, $m = p^a M_1$ in Theorem 6. We conclude $M_1 M_2^2 \geqq p^a$, contradicting the hypothesis.

Theorem 6 also gives a simple proof of Theorem 7 of [8]:

COROLLARY 5. *If $p \mid (n, v)$, $v = p^a v_1$ with p semiprimitive* mod v_1, *then there exists no cyclic difference set with parameters (v, k, λ).*

Putting p^a, v_1, p^b of the corollary equal to v_1, v_2, m, respectively in Theorem 6, we conclude $p^a v_1 \geqq p^{a+b} v_1$, or $1 \geqq p^b$.

Other examples of Theorem 6 can easily be given. For example (since $3^2 \equiv -1 \pmod 5$, $5 \equiv -1 \pmod 3$), there is no cyclic set with $N = 3^a 5^b M$ if $3^a 5^b > 2M^2$.

We note an analogous theorem which can be proved in the same fashion.

THEOREM 7. *If* χ *is a character of* G *of order* $v_1 > 1$, $v_1 = \prod_1^r q_i$ *with* q_i *powers of distinct primes and* D *is a subset of* G *such that* $\chi(D) = w \prod_{j=1}^r (\sum_{i=1}^{\phi(q_j)} A_{i,j}\zeta_j^i) A_{i,j}$ *rational integers,* ζ_j *a primitive* q_jth *root of* 1, *then* $2^{r-1}v \geqq v_1 \prod_{j=1}^r \max_i A_{i,j}$.

The preceding results depended on elementary considerations about the magnitude of the characteristic function of a difference set summed over a subset of the group. We shall now prove a result which depends only on the fact that the characteristic function has integer values, with no restrictions on the size of the integers.

It is easy to see that the only algebraic integers in the field of 2^mth roots of $1, m \geqq 3$, of absolute value 3^a are of the form $w3^{a-b}(1 \pm 2\sqrt{-2})^b$, $0 \leqq b \leqq a$. (Note that $\sqrt{-2} = \zeta + \zeta^3$, ζ a primitive 8th root of 1.) Thus, since $|A + B\sqrt{-2}| = 3^m$ implies $A^2 + 2B^2 = 3^{2m}$, $\max A^2, B^2 \geqq 3^{2m-1}$, and theorem 7 implies there are no H sets with $\sigma_2(G) = 2t + 2$ if $N = 2^t 3^S$ and $2^{2t} > 3^{2S+1}$. However, we shall remove any magnitude restrictions and prove that there are no H sets with $\sigma_2(G) = 2t + 2$, $t \geqq 1$ if $2^t \| N$.

We first make the following remarks: if p is a prime and k, λ, v are integers such that $k^2 - \lambda v = n$, $k - \lambda = n$, and $p \mid (n, v)$ then $p \mid (k, \lambda)$; since $k(k-1) = \lambda(v-1)$, and p does not divide $k-1, v-1, p$ divides k and λ to the same power, say $p^r \| k, \lambda$. Thus $p^r \mid n$. Assume that $p^S \| n$; then

(a) $2r > S$ implies $p^{S-r} \| v$

(b) $2r < S$ implies $p^r \| v$

and in either case the power of p which divides v is less than $S/2$. Finally

(c) $2r = S$ implies $p^r \mid v$.

In case (c), assume further that $p = 2$. Then n is a square (by Theorem 2); if $k = 2^r k_1$, $n = 2^{2r} n_1^2$ with k_1, n_1 odd we get $2^{2r}(k_1^2 - n_1^2) = \lambda v$. Since $2^r \| \lambda$, we conclude that $2^{r+3} \mid v$, as $k_1^2 - n_1^2 \equiv 0 \pmod 8$.

We shall now show that if D is a difference set such that $2 \mid (n, v)$, and 2 divides k^2 and n to the same power, then the group G cannot have a character of order 2^a, where $2^a \| v$; in particular, G cannot be cyclic.

Let D be a difference set, and assume $p^t \| k, \sqrt{n}, p^{t+S} \| v$ with $t \geqq 1$. Then we know that for $p = 2, S \geqq 3$. Let χ be a character of G of order p^{t+S}, and let Y_m^j be the number of elements g of D such that

$\chi_j(g) = \chi^{2^{t+S-j}} = \zeta_j^m$, where $\zeta_h = \exp(2\pi i/2^h)$, for $1 \leqq h \leqq t+S$. Then $\chi_j(D) = \sum_1^T Y_m^j \zeta_j^m$, $T = 2^j$.

LEMMA 6. *If* $2^{t-a} \| Y_m^h$ *for some* $m, a > 0$ *then* $2^{t-a-j} \| Y_m^{h+j}$ *for* $h + j \leqq t + S$.

Proof. We note the formula

$$Y_m^{h+1} = \frac{Y_m^h + Y_m^{h+1} - Y_{m+\phi}^{h+1}}{2} \qquad \phi = 2^h \tag{10}$$

which follows from $Y_m^h = Y_m^{h+1} + Y_{m+\phi}^{h+1}$. But $\chi_{h+1}(D) = \sum_1^\phi (Y_m^{h+1} - Y_{m+\phi}^{h+1})\zeta_{h+1}^m$. Since $2^t \| \chi_{h+1}(D) |$ and there is only one prime ideal which divides 2 in the field $Q(\zeta_{h+1})$, we conclude that $2^t | \chi_{h+1}(D)$ and therefore $2^t | Y_m^{h+1} - Y_{m+\phi}^{h+1}$, $1 \leqq m \leqq \phi$, since the ζ_{h+1}^i, $1 \leqq i \leqq \phi$, form an integral basis for $Q(\zeta_{h+1})$. Thus $2^{t-a} \| Y_m^h$, $a > 0$ implies $2^{t-a-1} \| Y_m^{h+1}$ and by induction $2^{t-a-j} \| Y_m^{h+j}$ for $h + j \leqq t + S$.

COROLLARY. $2^t | Y_m^S$ *for all* m.

For if not, put $j = t$ in the lemma; this would imply Y_m^{t+S} is not an integer.

LEMMA 7. *Assume* D *is a difference set such that* $2 | n, v$, $2^{2t} \| n, k^2$, $2^{t+S} \| v$, $S \geqq 0$. *Then* $S \geqq 3$, *and there exist integers* Z_m *such that*

$$\left|\sum_1^T Z_m \zeta_h^m\right| = M \qquad 1 \leqq h \leqq S, \qquad T = 2^S$$

$$\sum_1^T Z_i = k_1 \tag{11}$$

k_1, M *odd integers*, $2^S \| k_1^2 - M^2$.

Proof. We have seen that $2^t | Y_m^S$. Let $Z_m = 2^{-t} Y_m^S$, $k_1 = 2^{-t}k$, $M^2 = 2^{-2t}n$. Then equations (11) are a summary of the known properties of D; the fact that $2^S \| k_1^2 - M^2$ follows from $k^2 - n = \lambda v$, and $2^t \| k, \lambda, \sqrt{n}$.

THEOREM 8. *There are no sets of integers* Z_i, k_1, M *which satisfy* (11) *for* $S \geqq 3$.

Proof. Let $\hat{Z}_i = \sum_1^T Z_m \zeta_S^{im}$, $0 \leqq i \leqq T - 1$. Then $Z_0 = k_1$, $|\hat{Z}_i| = M$ for $i > 0$. The latter equations imply (e.g., by (3)) that

$$\sum_1^T Z_m^2 = M^2 + \frac{k_1^2 - M^2}{2^S}.$$

The assumption $2^S \| k_1^2 - M^2$ implies therefore that ΣZ_m^2 is even. Since the Z_m are integers and ΣZ_m is odd, this is impossible.

COROLLARY. *If D is a difference set and $2^t \| k, \sqrt{n}, 2^{t+S} \| v, t \geqq 1$ then $\sigma_2(G) < t + S$. In particular, there are no cyclic difference sets for such values of $v, k\, n$.*

This follows from the theorem and Lemma 7.

We remark that Lemma 6 holds for arbitrary primes.

Existence theorems. We shall now give some existence theorems for H sets. We denote $Z_2 \times Z_2$ by K_4.

THEOREM 9. *The following are abelian H sets for $N = 2^{h-1}$, $h > 1$:*

(1) *All h-tuples with an odd number of zero components, $G = \prod_1^r k_4 \prod_1^{h-r} Z_4$.*

(2) *The subset of $GF(2^h) \times GF(2^h)$ consisting of all pairs $(m_1 + m_2, m_1 m_2)$, $m_i \in GF(2^h)$.*

Proof. The set $\{0\}$ is an H set in K_4 or Z_4; by Lemma 4 (taking Kronecker products of the Hadamard matrices) we get the first statement.

To prove the second statement, let $q = 2^h$. The set $D =$ all $(m_1 + m_2, m_1 m_2)$ is easily seen to be the set of points which lie on one of the $q + 1$ lines in the affine plane $GF(q) \times GF(q)$

$$L_\infty: \quad X = 0$$
$$L_m: \quad Y = mX + m^2 \qquad m \in GF(q) \, .$$

All these lines have distinct slopes, and it is easily verified that each point in D lies on precisely two of these lines. The number of points in D is $q(q + 1)/2$, since there are $q + 1$ lines of q points each, and each point lies on two lines. (Note that here $2k > v$.) We now consider $D \cap D + a$ for $a \in G$, a not the identity. If the vector a is parallel to L_α (where $\alpha \in GF(q)$ or $\alpha = \infty$), then $P \in L_\alpha$ implies $P + a \in L_\alpha$. If $\beta \neq \alpha, L_\beta \cap L_\beta + a$ will be empty, since $L_\beta + a$ has the same slope as L_β, but $L_\beta + a = L_\beta$ only if the slope of $a =$ the slope of L_β. Any line not one of the L_γ contains $q/2$ points of D; for it intersects q of the lines L_α, and each point of intersection lies on precisely two of the L_γ. Count all the points of $D \cap D + a$ twice: there are q points on L_α, and $q/2$ on each of the other q lines. Therefore $D \cap D + a$ contains $q + q(q/2)$ points each counted twice, and the order of $D \cap D + a$ is $(q^2 + 2q)/4$ for $a \neq 0$, independent of a; clearly $n = q^2/4$, $v = 4n$.

If D is a difference set with v even, n is a square, and an obvious possible value for the $\chi(D)$ is $w(\chi)\sqrt{n}$ for $\chi \neq \chi_0$, with $w(\chi)$ an appropriate root of 1 for each χ. ($w(\chi)$ must have order dividing the order of χ, or twice it if the order of χ is odd; $(m, v) = 1$ implies $w(\chi^m) = w(\chi)^m$.) If $v = 4N^2, k = N(2N - 1)$ we must have

$$y_g = \frac{1}{4N}\Big(2N - 1 + \sum_{\chi \neq \chi_0} w(\chi)\bar{\chi}(g)\Big)$$

if $\chi(D) = w(\chi)N$, where $|w(\chi)| = 1$ for any H set, and $w(\chi)$ is a root of 1 if we assume $(\chi(D)) = (N)$ for $\chi \neq \chi_0$.

We now note a simple lemma.

LEMMA 8. *Let D be an H set in G, G_1 normal in G of index 4. If $(\chi(D)) = (N)$ for $\chi \neq \chi_0$ and G_1 in the kernel of χ, then the numbers of elements in the cosets of G_1 are*

$$\left(\frac{N^2}{2}, \frac{N^2}{2}, \frac{N^2}{2}, \frac{N(N-2)}{2}\right) \tag{12}$$

or

$$\left(\frac{N(N-1)}{2}, \frac{N(N-1)}{2}, \frac{N(N-1)}{2}, \frac{N(N+1)}{2}\right). \tag{13}$$

Only the second case can arise if N is odd.

This is a trivial application of, for example, formulae (7) and (8). The first case arises if $G/G_1 = Z_4$ and $\chi(D) = \pm iN$ or if $G/G_1 = Z_2 \times Z_2$ and the three characters of order 2 on G/G_1 do not give equal character sums. The formula (12) does not yield integers if N is odd.

This lemma is proved incorrectly in [10]; the assumption on $\chi(D)$, if $G/G_1 = Z_4$ is not explicitly stated.

We shall now describe certain difference sets in the abelian groups of order 36 which have no elements of order 9. It will be convenient to consider $Z_3 \times Z_3$ as an affine plane (over the field Z_3); we denote it by A_3. We shall refer to these sets as Q sets.

Let G_4 be K_4 or Z_4, and let $0, 1, 2, 3$ be the elements of G_4. In the affine plane A_3 take four lines $L_i, 0 \leqq i \leqq 3$, one of each slope (i.e., four distinct mutually intersecting lines). Let S_0 be the complement of L_0 in $A_3, S_i = L_i$ for $i = 1, 2, 3$. We let D be the subset of $G_4 \times A_3$ consisting of all pairs (i, x) with $x \in S_i, 0 \leqq i \leqq 3$. It is not hard to verify that D is an H set; this will be shown in the course of Theorem 12.

We now enumerate Z_4 in the usual manner by $i = 0, 1, 2, 3$, and let $0 = (0, 0), 1 = (0, 1), 2 = (1, 0), 3 = (1, 1)$ in K_4. We let Q_1 be the

Q set in $Z_4 \times A_3$ for which L_0 is $X = 0, L_1$ is $Y = X, L_2$ is $Y = 2X$ and L_3 is $Y = 0$. Q_2 is like Q_1 except that L_0 is $X = 1$. Q_3 is like Q_1 except that L_3 is $Y = 1$. Q'_1, Q'_2, Q'_3 are the Q sets in $K_4 \times A_3$ defined like Q_1, Q_2, Q_3.

We call two subsets D_1, D_2 of a group G equivalent if $D_2 = (\sigma D_1)\alpha$, where σ is an automorphism of G and $a \in G$.

THEOREM 10. *Any Q set is equivalent to one of Q_i or Q'_i, $i =$ 1, 2, 3; these are inequivalent.*

We first prove a simple lemma.

LEMMA 9. *Assume there are $N + 1$ distinct mutually intersecting lines L_i (i.e., one of each slope) in the affine plane $GF(N) \times GF(N)$, (N any prime power), such that any point in the plane lies on not more than two lines; then N is even.*

To prove the lemma, fix one of the $N + 1$ lines, say L_0. It contains N points of the plane and intersects N of the lines L_i. Since a point lies on at most two of the L_i, each point of L_0 lies on precisely two of the lines L_i. This proves any point of the plane lies on none or two of the lines L_i. Now take a line parallel to L_0, but $\neq L_0$. It must intersect all the L_i except L_0; each point of intersection is on two of the L_i, and there are N intersections; thus N is even.

We now return to the proof of the theorem.

Every automorphism σ of $G_4 \times A_3$ induces automorphisms σ_4, σ_3 on G_4 and A_3, respectively. In an arbitrary Q set, the element of G_4 which corresponds to S_0 is determined (it is the only element x of G_4 for which there are six elements (x, y) in the set). Lemma 9 shows that the four lines L_i have a point P of triple or quadruple intersection, necessarily unique, and it is also uniquely determined by the Q set. Since L_0 is uniquely determined by the set, we see that the sets Q_i, Q'_i are indeed inequivalent. To show that any Q set is equivalent to one of Q_i, Q'_i, we first translate the Q set so that the identity element of G_4 corresponds to L_0 and the point P of A_3 corresponds to the origin of A_3. We now observe that the automorphism group of A_3 is transitive on quadruples of distinct slopes: given four distinct lines through the origin, we may clearly transform the first and second into $X = 0$ and $Y = 0$, respectively, by an automorphism (since A_3 is a vector space). If it is necessary to interchange the other two slopes, the linear transformation S which takes (x, y) into $(x, -y)$ ($(x, y) \in A_3$) will leave the X and Y axes invariant but will interchange the other two lines through the origin. If four of the lines L_i go through P, we have shown the Q set is equivalent to Q_1 or Q'_1. If one of the L_i does not contain P, we first apply an automorphism to A_3 which will transform the slopes to correspond to the slopes of Q_i or Q'_i, $i = 1$ or 2. If the line L_i which does not contain P now coincides with the corresponding line in Q_i or Q'_i we have shown the desired equivalence; if it does not, we apply the inversion automorphism to A_3. This will leave invariant the lines through the origin, and take a line not through the origin into the other line parallel to itself and not through the origin.

THEOREM 11. *The multiplier groups of* $Q_1, Q_2, Q_3; Q'_1, Q'_2, Q'_3$ *are of orders* $4, 2, 2; 12, 6, 6,$ *respectively.*

It is clear that a multiplier of any of the Q_i, Q'_i must leave the sets fixed, since we have seen that identity elements of G_4 and A_3 are special elements of the sets. We have also seen that the automorphism group of A_3 is transitive on quadruples of slopes, and only the transformations $\pm I$ of A_3 leave all the slopes invariant. A multiplier of one of the Q_i or Q'_i restricted to G_4 is a permutation τ of $0, 1, 2, 3$ (which leaves 0 fixed); the permutation of the slopes of the L_i in A_3 must induce the same permutation of $0, 1, 2, 3$. We can always find precisely two automorphisms of A_3, σ and $(-I)\sigma$, which leave the Y axis fixed and take the slope of L_i into that of $L_{\sigma(i)}$, $i = 1, 2, 3$. If the L_i all go through the origin both (τ, σ) and $(\tau, (-I)\sigma)$ will be multipliers. However, if one of the L_i does not go through the origin, only one of these two automorphisms will take Q_i or Q'_i into itself (the other will take the L_i not through the origin into the line $(-I)L_i$). The theorem now follows because Z_4 and K_4 have 2 and 6 automorphisms, respectively.

THEOREM 12. *The only H sets for which N is an odd prime, satisfying the condition* $(\chi(D)) = (N)$ *for all* $\chi \neq \chi_0$ *are the Q sets described above and their complements, if G is abelian.*

Proof. We have seen that if N is an odd prime and an H set exists with $n = N^2$ then $\sigma_N(G) < 2$, (Corollary 1 of Theorem 6). Thus G must be $K_4 \times Z_N \times Z_N$ or $Z_4 \times Z_N \times Z_N$.

We shall assume that $k = N(2N - 1)$ (by taking the complement of the H set D if necessary). We consider first $G = K_4 \times Z_N \times Z_N$, and write $abcd$ for $y_{(a,b,c,d)}$ with $a, b \in Z_2$, $c, d \in Z_N$. We let χ_α, χ_β be the characters of order 2 defined by $\chi_\alpha((1, b, c, d)) = \chi_\beta((a, 1, c, d)) = -1$, for all a, b, c, d. We may assume D is translated so that $\chi_\alpha(D) = \chi_\beta(D) = N$ (since they are both $\pm N$, being rational integers of absolute value N). Lemma 8 then shows that $\chi_\alpha\chi_\beta(D) = N$, since N is odd.

Let ζ be a fixed primitive Nth root of 1, and define χ_N, χ_∞ by

$$\chi_N((a, b, c, d)) = \zeta^c$$
$$\chi_\infty((a, b, c, d)) = \zeta^d$$

c, d being integers $(\bmod N)$. For $0 < k < N$, let

$$\chi_k = \chi_N \chi_\infty^k .$$

For any nonprincipal character χ, $(\chi(D)) = (N)$ and $|\chi(D)| = N$, so that $\chi(D) = wN$ for some root of 1, w. We may thus write

$$\chi_k(D) = u_k \zeta^{e_k} N$$

for $k = 1, \cdots, N, \infty, u_k = \pm 1$. By Theorem 5, if χ_γ is any character of order 2 we have $\chi_\gamma \chi_k(D) = \pm \chi_k(D)$ (put $\chi_k = \psi$ in Theorem 5). We may therefore write

$$\chi_\alpha \chi_\beta \chi_k(D) = t_k \zeta^{e_k} N$$
$$\chi_\alpha \chi_k(D) = w_k \zeta^{e_k} N$$
$$\chi_\beta \chi_k(D) = v_k \zeta^{e_k} N$$

with $v_k, w_k, t_k = \pm 1$.

We first prove that $\sum_k u_k = -1 \pm N (k = 1, \cdots, N, \infty)$. For $N^2 \mid \sum_{i,j} \chi_N^i \chi_\infty^j(D) = N(2N - 1) + \sum_k \sum_{i \neq 0} \chi_k^i(D), i, j \pmod N$. But if $\chi_k(D) = u_k \zeta^{e_k} N$ then $\chi_k^i(D) = u_k \zeta^{i e_k} N$ for $i \not\equiv 0 \pmod N$ (as in the proof of Theorem 5) and therefore $\sum_{i \neq 0} \chi_k^i(D) = u_k N(-1 + \delta_{0,e_k} N)$. Thus $N^2 \mid -N + \sum_k - u_k N$, and $N \mid 1 + \sum_k u_k$. Therefore $\Sigma u_k \equiv -1 \pmod N$, and Σu_k is not more than $N + 1$ in absolute value, since $u_k = \pm 1$ for all k. Since N is odd, Σu_k is even, and therefore $\Sigma u_k = -1 - N$ or $-1 + N$. Thus all the u_k are -1 or all but one are $+1$.

Similarly each of $\Sigma v_k, \Sigma w_k, \Sigma t_k$ is $1 \pm N$; the argument is the same, but the term which corresponds to $i = j = 0$ in e.g., the sum $\sum_{i,j} \chi_\alpha \chi_N^i \chi_\infty^j(D)$ is now N instead of $N(2N - 1)$. The v_k are all $+1$ or all but one are -1; the same is true, independently for the w_k and for the t_k.

We shall write δ_k for δ_{c+kd, e_k}, and Δ for $\sum_k \delta_k$. Δ and the δ_k depend on c, d. We shall refer to the set of c, d such that $c + kd = e_k$ as line k; these are the points of A_3 for which $\delta_k = 1$. Δ is the number of the lines k on which the point c, d lies.

Now the inversion formula gives

$$00cd = \frac{N+1}{2N} + \frac{1}{4N} \sum_k (N\delta_k - 1)(u_k + v_k + w_k + t_k)$$

$$10cd = \frac{N-1}{2N} + \frac{1}{4N} \sum_k (N\delta_k - 1)(u_k + v_k - w_k - t_k)$$

$$01cd = \frac{N-1}{2N} + \frac{1}{4N} \sum_k (N\delta_k - 1)(u_k + w_k - v_k - t_k)$$

$$11cd = \frac{N-1}{2N} + \frac{1}{4N} \sum_k (N\delta_k - 1)(u_k + t_k - v_k - w_k) .$$

The first of the above formulae, for example, is

$$00cd = \frac{1}{4N^2}\sum_{\chi}\chi(D)\overline{\chi}(00cd)$$
$$= \frac{1}{4N^2}\sum_{k}\sum_{i\neq 0}\Big(\chi_k^i(D) + \chi_\alpha\chi_k^i(D) + \chi_\beta\chi_k^i(D) + \chi_\alpha\chi_\beta\chi_k^i(D)\Big)$$
$$\cdot\,\overline{\chi}_k^i((0,0,c,d)) + \frac{1}{4N^2}(\chi_0(D) + \chi_\alpha(D) + \chi_\beta(D) + \chi_\alpha\chi_\beta(D))$$

since $\chi_k(0, 0, c, d) = \chi_\gamma\chi_k(0, 0, c, d)$ for any χ_γ of order 2. The last term is $(N(2N-1)+3N)/4N^2$, the first term in the formula for $00cd$; the sum is clearly $(1/4N^2)\sum_k\sum_{i\neq 0}(u_k + v_k + w_k + t_k)\zeta^{ie_k}N\cdot\zeta^{-c-kd}$ (with the convention $\zeta^{-c-\infty d} = \zeta^{-d}$) which reduces to the first formula. The above follow similarly, except that e.g., $\chi_\alpha\chi_k((1,0,c,d)) = \chi_\alpha\chi_\beta\chi_k((1,0,c,d)) = -\chi_k((1,0,c,d))$.

It is clear that finding an H set D of the required type is precisely equivalent to finding u_k, v_k, w_k, t_k all ± 1, and $e_k \bmod N$ which yield 0 or 1 in all the above equations. We shall now consider all the possible types of solution (using the symmetry of the v_k, w_k, t_k in the problem).

I. $\qquad u_k = -1,\ v_k = w_k = t_k = 1 \qquad$ for all k.

Then

$$00cd = \frac{\Delta}{2}$$
$$10cd = 01cd = 11cd = 1 - \frac{\Delta}{2}\,.$$

Since $00cd$ must be 0 or 1, any point c, d must lie on none or two of the lines k; but Lemma 9 then shows N is even. For another proof, note that these formulae show the resulting Hadamard matrix would be equivalent to a direct product of an $N\times N$ matrix $((2(00cd)-1))$ by the 4×4 matrix $2I - J$, which requires the $N\times N$ matrix to be Hadamard, i.e., $N = 1$ or N even. This case suggested the construction in Theorem 8.

II. $\qquad u_k = -1,\ v_k = w_k = 1,\ t_k = -1 \qquad$ for all k, except $t_m = 1$.

Then

$$00cd = \frac{1}{2}(1 + \delta_m)$$

and $00cd = 1/2$ for cd not on line m, which is impossible.

III. $\quad u_k = -1,\ v_k = 1,\ w_k = t_k = -1\quad$ for all k, except $w_j = t_m = 1$.

Then

$$10cd = 1 - \frac{1}{2}(\Delta - \delta_j + \delta_m) .$$

If $j = m$, $10cd$ is fractional unless each point which lies on one of the lines k lies on at least one other. But this would mean each point of one of the lines k would lie on precisely two, and by Lemma 9 this would mean N is even. If $j \neq m$, the formula

$$01cd = \frac{1}{2}(\Delta - \delta_j - \delta_m) = \frac{1}{2} \sum_{k \neq j, m} \delta_k$$

shows that any point on one of the $N - 1$ lines k, but $\neq j, m$ must lie on another. Since any one of these $N - 1$ lines intersects the others in at most $N - 2$ points, this is impossible.

IV. $u_k = v_k = w_k = t_k = -1$ for all k, except $t_m = w_j = v_n = 1$.

Then

$$11cd = \frac{1}{2}(1 + \delta_m - \delta_n - \delta_j) .$$

If cd is on none of the lines j, m, n we have $11cd = 1/2$, which is impossible. But these three lines contain at most $3(N - 1) + 1$ points, and for $N > 2$, $N^2 > 3N - 2$, so such a point exists.

V. $u_k = v_k = w_k = t_k = 1$ for all k, except $u_h = -1$.

$$00cd = -\frac{1}{2} + \Delta - \frac{\delta_h}{2}$$

and $00cd$ is not an integer for cd not on line h.

VI. $u_k = v_k = w_k = 1, t_k = -1$ for all k, except $u_h = -1, t_m = 1$

$$00cd = \frac{1}{2}(\Delta - \delta_h + \delta_m) .$$

If $m = h$, N is even by Lemma 9. If $m \neq h$, $\sum_{k \neq m, h} \delta_k$ must be an integer for all cd which was shown impossible in III.

VII. $u_k = v_k = 1, t_k = w_k = -1$
for all k, except $u_h = -1, t_m = w_j = 1$.

$$00cd = \frac{1}{2}(1 + \delta_m + \delta_j - \delta_h)$$

so for a point not on lines h, j, m we would have $00cd = 1/2$, as in IV.

VIII. $u_k = 1, v_k = w_k = t_k = -1$
for all k, except $u_h = -1, t_m = w_j = v_n = 1$.

We have

$$00cd = 1 - \frac{1}{2}(\Delta + \delta_h) + \frac{1}{2}(\delta_m + \delta_j + \delta_n)$$

$$01cd = \frac{1}{2}(\Delta - \delta_h + \delta_j - \delta_m - \delta_n)$$

and the two formulae analogous to $01cd$. First, we note that h is not equal to any of j, m, n: for if say $j = h$ (by symmetry), we would have $01cd = (1/2)(\Delta - \delta_m - \delta_n) = (1/2)\sum_{k \neq m,n} \delta_k$ and this is shown impossible in III. Second, we note that j, m, n are distinct: again, by symmetry, if say $j = m$ we would have $01cd = (1/2)(\Delta - \delta_h - \delta_n)$ as before. Therefore j, m, n, h are all distinct. But then $01cd = (1/2)(\sum_{k \neq m,n,h,j} \delta_k) + \delta_j$ and the sum in parenthesis must be an even integer for all cd. This is impossible (as in III) if there are more than 4 lines in the plane; but if $N = 3$, the sum is zero, and the formulae reduce to

$$00cd = 1 - \delta_h$$
$$01cd = \delta_j$$
$$10cd = \delta_n$$
$$11cd = \delta_m$$

which clearly give 0, 1 values for any choice of the lines h, j, m, n (one of each slope) for all c, d.

We now turn to the group $Z_4 \times Z_N \times Z_N = Z_4 \times A_N$. We write abc for $y_{(a,b,c)}$, $a \in Z_4$, $b, c \in Z_N$. We define the characters χ_k $k = 1, \cdots, N, \infty$ of $Z_N \times Z_N$ as before. We let ψ be a fixed character of order 4. We have

$$\chi_k(D) = u_k N \zeta^{e_k}$$
$$\psi^2 \chi_k(D) = v_k N \zeta^{e_k}$$

with $u_k, v_k = \pm 1$. Again we get $\sum_k u_k = -1 \pm N$, $\sum_k v_k = 1 \pm N$. By Theorem 5, (with $\chi_k = \psi$) we conclude that $\chi_k(D) = w\chi_k(D)$ with w a fourth root of 1; write $\psi\chi_k(D) = w_k N i^{a_k} \zeta^{e_k}$, with $a_k = 0$ or 1, and $w_k = \pm 1$. Lemma 8 shows we may normalize the set so that $\psi(D) = \psi^2(D) = N$.

We use the $\sum_I$, $\sum_{II}$ to denote the sum over those values of k for which $a_k = 0$ or $a_k = 1$, respectively. As before, the inversion formula gives

$$0bc = \frac{N+1}{2N} + \frac{1}{2N} \sum_{I} (u_k + w_k)(N\delta_k - 1)$$

$$2bc = \frac{N-1}{2N} + \frac{1}{2N} \sum_{I} (u_k - w_k)(N\delta_k - 1)$$

$$1bc = \frac{N-1}{2N} + \frac{1}{2N} \sum_{II} (u_k + w_k)(N\delta_k - 1)$$

$$3bc = \frac{N-1}{2N} + \frac{1}{2N} \sum_{II} (u_k - w_k)(N\delta_k - 1)$$

since for example, in the first formula, $\psi\chi_k(D) + \psi^3\chi_k(D) = 0$ if $\psi\chi_k(D) = \pm i\zeta^{e_k}N$. (As in Theorem 5, $\psi^3\chi_k(D)$ is the conjugate of $\psi\chi_k(D)$ under the automorphism σ defined by $\sigma(i) = -i, \sigma(\zeta) = \zeta$). But then, since $4 \mid \chi_k(D) + \psi^2\chi_k(D) + \psi\chi_k(D) + \psi^3\chi_k(D)$, $4 \mid (u_k + v_k)N\zeta^{e_k}$, so $u_k = -v_k$ if $\psi\chi_k(D) = \pm i\chi_k(D)$. If $\psi\chi_k(D) = \pm\chi_k(D)$, we conclude $u_k = v_k$.

By considering $0bc \pm 2bc$ for any bc we conclude that $\sum_I u_k \equiv 0 (\bmod N)$, and $\sum_I w_k \equiv 1 \pmod N$. The second of these shows there exist values of k which occur in $\sum_I$, i.e., for which $\psi\chi_k(D) = \pm i\chi_k(D)$. Since $\sum_I u_k + \sum_{II} u_k = \sum u_k = -1 \pm N$, we conclude $\sum_{II} u_k \equiv -1 (\bmod N)$, so that there exist values of k which occur in $\sum_{II}$. The formula for $1bc - 3bc$ shows $\sum_{II} w_k \equiv 0 (\bmod N)$.

$\sum u_k = -1 \pm N$; if $\sum u_k = -1 - N, u_k = -1$ for all k, and $\sum_I u_k \equiv 0 \pmod N$ implies there are N values of k which occur in $\sum_I$ (since there is at least one). We would then have precisely one value of k in $\sum_{II}$, which would imply $\sum_{II} w_k = \pm 1$; but $\sum_{II} w_k \equiv 0 (\bmod N)$. Therefore we must have $\sum u_k = -1 + N$. Now $\sum_I u_k \equiv 0 (\bmod N)$ would imply $\sum_I u_k = N$ or 0. The first of these would imply there are N values of k in $\sum_I$, therefore $\sum_{II} w_k = \pm 1$, which would contradict $\sum_{II} w_k \equiv 0 (\bmod N)$. Therefore $\sum_I u_k = 0$, and there are two values of k in $\sum_I$, $N - 1$ in $\sum_{II}$. But since $\sum w_k \equiv 1 (\bmod N)$, $\sum w_k = 1 \pm N$, to get $\sum_{II} w_k \equiv 0 (\bmod N)$ we must again have $\sum_{II} w_k = 0$, and there are two values of k in $\sum_{II}$. Therefore $N = 3$.

There are two values of k in $\sum_I$, and $\sum_I u_k = 0$. Pick the two generators of $Z_N \times Z_N$ so that $u_3 = -1, u_\infty = +1$, with $3, \infty$ the values of k in $\sum_I$, i.e., $\chi_k(D) \pm \chi_4\chi_k(D)$ for $k = 3, \infty$. $w_3 + w_\infty = 1 - 3 = -2$, so $w_3 = w_\infty = -1$. Therefore $w_1 + w_2 = 0$, and by applying the automorphism $(x, y, z) \to (x, y, -z)$ we may assume $w_1 = 1$, $w_2 = -1$.

The formulae now reduce to (with notations as in the first part of the theorem)

$$0bc = 1 - \delta_3$$

$$2bc = \delta_\infty$$

$$1bc = \delta_1$$
$$3bc = \delta_2 .$$

Clearly, any choice of the lines gives a difference set.

COROLLARY. *The Q sets are the only abelian H sets with N a prime of the form $4k-1$.*

Proof. By Corollary 1 of Theorem 6, we must have $\sigma_N < 2$. The characters of G must all have order dividing $4N$; if N is a prime of the form $4k-1, N$ remains prime in $Q(i)$, and the only integers of absolute value N in the field of $4N$th roots of 1 are wN, w a root of 1. Thus $(\chi(D)) = (N)$ for all $\chi \neq \chi_0$, and the corollary follows from Theorem 12.

We remark that given a set of values of v, k, λ and an abelian group G of order v, one often very useful way of constructing difference sets in G with the given parameters is to construct first all the sets of algebraic integers which might be the $\chi(D)$, and then to construct D from these. Theorem 12 is an example of this procedure.

THEOREM 13. *Let $G = \prod_1^r Z_2 \prod_1^s Z_4 \prod_1^t Z_8 \prod_1^{2q} Z_3$, with $r \geqq t, r-t$ even, $r - t + 2s \geqq 2q$. Then there is an H set in G.*

Proof. The following two subsets of $Z_8 \times Z_2$ are inequivalent H sets:

$$(00, 10, 20, 50, 01, 61)$$
$$(00, 10, 21, 51, 01, 61) .$$

The theorem now follows from the previous theorem by Lemma 4.

It is easy to check that all the H sets of Theorem 13 satisfy the condition $(\chi(D)) = (N)$ for all $\chi \neq \chi_0$.

Addendum. "The case $r=1$ of Theorem 6 has been obtained independently by methods similar to those of this paper: K. Yamamoto, Decomposition fields of difference sets, Pacific J. Math., **13** (1963), 337–352, and R. A. Rankin, Difference sets, Acta Arithmetica, **9** (1964), 161–168. The second paper also contains a special case of Theorem 5."

REFERENCES

1. R. H. Bruck, *Difference sets in a finite group,* Trans. Amer. Math. Soc., **78** (1955), 464-481.

2. R. H. Bruck, and H. J. Ryser, *The nonexistence of certain finite projective planes,* Canad. J. Math. **1** (1949), 88-93.

3. S. Chowla, and H. J. Ryser, *Combinatorial problems*, Canad. J. Math. **57** (1935), 391-424.

4. Marshall Hall, and H. J. Ryser, *Cyclic incidence matrices*, Canad. J. Math. **3** (1951), 495-502.

5. Marshall, Jr. Hall, *A survey of difference sets*, Proc. Amer. Soc. **7** (1956), 975-986.

6. J. B. Kelly, *A characteristic property of quadratic residues*, Proc. Amer. Soc. **5** (1954), 38-46.

7. Emma Lehmer, *On residue difference sets*, Canad. J. Math. **5** (1953), 425-432.

8. H. B. Mann, *Balanced incomplete block designs and abelian difference sets*, Illinois J. Math. **8** (1964), 252-261.

9. P. Kesava Menon, *Difference sets in abelian groups*, Proc. Amer. Math. Soc. **11** (1960), 368-376.

10. ———, *On difference sets whose parameters satisfy a certain relation*, Proc. Amer. Math. Soc. **13** (1962), 739-745.

11. R. Turyn, and J. Storer, *On binary sequences*, Proc. Amer. Math. Soc. **12** (1961), 394-399.

12. R. Turyn, *The multiplier theorem for difference sets*, Canad. J. Math. **16** (1964), 386-388.

13. ———, *Optimum finite codes*, final report, Sylvania Electric Products, Inc., 1960.

14. ———, *Finite binary sequences*, final report, (Chapter VI) Sylvania Electric Products, Inc., 1961.

15. Hermann Weyl, *Algebraic theory of numbers*, Princeton, 1940.

PACIFIC JOURNAL OF MATHEMATICS
Vol. 15, No. 1, 1965

CONCERNING KOCH'S THEOREM ON THE EXISTENCE OF ARCS

L. E. WARD, JR.

A theorem of R. J. Koch asserts that if X is a compact space endowed with a partial order Γ such that

(i) Γ is a closed subset of $X \times X$,

(ii) there exists $0 \in X$ such that $(0, x) \in \Gamma$ for each $x \in X$, and

(iii) for each $x \in X$ the set $L(x) = \{y : y \leqq x\}$ is connected, then each point of X lies in a connected chain containing 0. In particular, X is arcwise connected. This is a corollary of the theorem: if X is a compact space and Γ is a partial order satisfying (i), and if W is an open subset of X such that each neighborhood of each point x of W contains a point $y \neq x$ with $(y, x) \in \Gamma$, then each point of W is the supremum of a connected chain which meets $X - W$. A new proof of these results is presented.

The first of these theorems is generalized in several ways. The compactness is relaxed to local compactness and the assumption that each closed chain has a zero. Moreover, the existence of a zero need not be assumed. If the set E of minimal elements is closed, then E is joined by connected chains to all other points of X. If the set function L is continuous, then E is necessarily closed.

1. A classical problem of topology is to determine when a space is arcwise connected. Here it will be convenient to adopt the terminology of A. D. Wallace [6] and call a subset A of a space an *arc* if A is a continuum with exactly two noncutpoints. If A is also separable then it is a *real arc*.

A few years ago R. J. Koch [4] proved a remarkable theorem of this type. He showed that a compact partially ordered space is arcwise connected if certain natural conditions are imposed on the partial order. It is the purpose of this paper to study Koch's result in detail. His proof, although ingenious, is long and very complicated. Since the theorem is fundamental to the structure theory of partially ordered spaces, and since it has been applied [3, 4, 6] to a variety of problems in topological algebra, it is of some interest to exhibit a shorter and simpler proof. This is done in § 2. In the later sections, some generalizations of Koch's theorem are obtained.

Received February 13, 1964. Presented to the American Mathematical Society, January 24, 1964. The author gratefully acknowledges the support of the National Science Foundation.

Perhaps the most celebrated theorem on arcwise connectivity is the assertion that every locally connected and metrizable continuum is real arcwise connected. I suspect the existence of an intimate relationship between Koch's theorem and this result. In the final section of the present paper, this possible relationship is discussed, but I have not been able to resolve the question satisfactorily.

2. A short proof of Koch's theorem. If Γ is a partial order on a set X, we identify Γ with its graph and treat the symbols $x \leqq y$, $x\Gamma y$ and $(x, y) \in \Gamma$ as synonyms. Recall that a *chain* of a partially ordered set (X, Γ) is a subset C of X such that $a\Gamma b$ or $b\Gamma a$ obtains for each a and b in C. We also define

$$L(a, \Gamma) = \{x \in X : x\Gamma a\} ,$$
$$M(a, \Gamma) = \{x \in X : a\Gamma x\} ,$$

for each $a \in X$. Where no ambiguity may occur we shall write $L(a)$ for $L(a, \Gamma)$ and $M(a)$ for $M(a, \Gamma)$. Moreover, if $A \subset X$ we define

$$L(A) = \bigcup \{L(x) : x \in A\} ,$$

and it is convenient to adopt the notation

$$[x, y]_\Gamma = M(x, \Gamma) \cap L(x, \Gamma) .$$

In case X is a topological space, the partial order Γ is *continuous* provided Γ is a closed subset of $X \times X$. When this occurs, $X = (X, \Gamma)$ is called a *continuously partially ordered space*. It is well-known [7] that if X is a continuously partially ordered space then the sets $L(x)$ and $M(x)$ are closed, for each $x \in X$, X is a Hausdorff space, and, if X is compact, it admits a minimal element, i.e., an element having no proper predecessors. A *zero* of a continuously partially ordered space is an element which precedes every other element. In the compact case, a unique minimal element is necessarily a zero. Finally, we remark that in a compact, continuously partially ordered space, a connected chain joining two distinct points is an arc. An arc which is also a chain will be termed an *order arc* or a *Γ-arc*.

(2.1) THEOREM (*Koch*). *Let W be an open subset of the compact, continuously partially ordered space X, and suppose, for each $x \in W$, that each neighborhood of x contains an element y with $y > x$. Then each $x \in W$ is the supremum of an order arc C such that $C - W$ is nonempty.*

(2.2) COROLLARY. *If X is a compact, continuously partially ordered space with zero, 0, such that $L(x)$ is connected for each $x \in X$, then each $x \in X - \{0\}$ is joined to 0 by an order arc.*

The corollary follows easily from Theorem 2.1 by taking $W = X - \{0\}$ (see [4]). Our proof of Theorem 2.1 is embodied in two main lemmas.

(2.3) LEMMA. *Let W be an open subset of the compact space X. If X admits a partial order satisfying the hypotheses of* (2.1), *then X admits such a partial order which is minimal.*

Proof. Let $\{\Gamma_\alpha\}$ be a maximal nest of partial orders satisfying the hypotheses of (2.1), and let $\Gamma_0 = \bigcap \{\Gamma_\alpha\}$. It is readily verified that Γ_0 is a continuous partial order on X. Let $x \in W$ and let U be a neighborhood of x; clearly we may assume that $U \subset W$. Since X is regular, there exists an open set V with $x \in V \subset \bar{V} \subset U$, and since X is normal, there exists an open set R with $X - U \subset R \subset \bar{R} \subset X - \bar{V}$. For each α, let x_α be a Γ_α-minimal element of $L(x, \Gamma_\alpha) \cap \bar{V}$; then there must exist $y_\alpha \neq x_\alpha$ such that

$$y_\alpha \in L(x_\alpha, \Gamma_\alpha) - \bar{R} \subset L(x, \Gamma_\alpha) - (\bar{R} \cup \bar{V}) .$$

Since the closed sets $L(x, \Gamma_\alpha) - (R \cup V)$ are nested and nonempty, there exists $y \in L(x, \Gamma_0) - R \cup V$. Thus $(y, x) \in \Gamma_0$, $y \neq x$, and $y \in U$. Therefore Γ_0 satisfied the hypotheses of (2.1) and is minimal.

(2.4) LEMMA. *Let W be an open subset of the compact space X, and suppose Γ is a partial order on X which is minimal with respect to satisfying the hypotheses of* (2.1). *Then every maximal chain of (X, Γ) is connected.*

Proof. If not then the compactness of X guarantees [7] the existence of elements a and b of X with $(a, b) \in \Gamma$, $a \neq b$, and

$$[a, b]_\Gamma = \{a\} \cup \{b\} .$$

Since X is a Hausdorff space, there are disjoint open sets U and V with $a \in U$ and $b \in V$. Let

$$F = \{(x, y) \in X \times X : [x, y]_\Gamma - (U \cup V) \neq 0\} .$$

A routine argument involving the continuity of Γ shows that F is closed and hence

$$\Delta = \Gamma - (U \times V - F)$$

is also closed. Since Γ is reflexive and $U \cap V = 0$, one sees that Δ is reflexive, and the anti-symmetry of Γ implies that Δ has the same property. To see that Δ is transitive, suppose that $p\Delta q$ and $q\Delta r$ but $(p, r) \in X \times X - \Delta$. Since $p\Gamma r$, it is clear that

$$(p, r) \in U \times V - F$$

and thus $[p, r]_\Gamma \subset U \cup V$, so that $q \in U$ or $q \in V$. If $q \in U$ then, since $r \in V$ and $(q, r) \in \Delta$, we infer that $(q, r) \in F$ and consequently

$$[q, r]_\Gamma - (U \cup V) \neq 0 .$$

But then

$$[p, r]_\Gamma - (U \cup V) \neq 0 ,$$

i.e., $(p, r) \in F$, a contradiction. A similar contradiction ensues if $q \in V$ and therefore Δ is transitive.

Now let $x \in W$ and let N be a neighborhood of x. If $x \in X - V$ then $L(x, \Delta) = L(x, \Gamma)$ and hence there exists $y \in N$, $y \neq x$, with $y \Delta x$. If $x \in V$ then

$$L(x, \Delta) \cap V = L(x, \Gamma) \cap V$$

and hence the desired y exists in $N \cap V$. Therefore Δ satisfies the hypotheses of (2.1), contrary to the minimality of Γ.

Proof of Theorem (2.1). In view of Lemma 2.3 we may assume that Γ is minimal, for any Γ-arc will be an order arc with respect to a partial order which contains Γ. Let $x \in W$ and let D be a maximal chain of X such that $x \in D$. By Lemma 2.4, D is an order arc, and by hypothesis, $C = D \cap L(x)$ is nondegenerate and hence C is also an order arc. Since X is compact, C has a least element which cannot lie in W.

It should be noted that the chief applications to topological algebra arise from Theorem 2.1. From a purely topological point of view, hewever, Corollary 2.2 is the more interesting, and it is this result which we shall generalize in several ways.

3. A lemma on quotient spaces. If X is a space and F is a closed subset of X, we denote by X/F the quotient space which is obtained when F is identified with a point.

(3.1) LEMMA. *Let (X, Γ) be a continuously partially ordered space and let F be a compact subset of X such that $F = L(F)$. Then X/F is a continuously partially ordered space. If, for each $x \in X$, it follows that $L(x, \Gamma)$ meets F, then F is a zero for X/F. Finally, if X is compact and, for each $x \in X$, each component of $L(x, \Gamma)$ meets F, then X/F satisfies the hypotheses of Corollary 2.2 and hence each point of $X/F - \{F\}$ is joined to F by an order arc of X/F.*

Proof. Define the relation Δ on X/F by $p\Delta q$ provided $p, q \in X-F$ and $p\Gamma q$, or $p = F$ and $L(q, \Gamma)$ meets F. It is clear that Δ is a partial order, and the proof that Δ is continuous is routine except for the verification of the fact that if $(F, q) \notin \Delta$, then there are open sets U and V such that $q \in U$, $F \subset V$ and $L(U, \Gamma)$ and V are disjoint. To see this we note that since Γ is continuous and $L(q, \Gamma)$ and F are disjoint there exist, for each $t \in F$, open sets U_t and V_t such that $q \in U_t$, $t \in V_t$ and $L(U_t, \Gamma)$ and V_t are disjoint. Since F is compact, a familiar argument shows that the desired sets U and V exist. That F is a zero if each $L(x, \Gamma)$ meets F is obvious. If X is compact then so is X/F, and if each component K_α of $L(x, \Gamma)$ meets F, then

$$L(x, \Delta) = \{F\} \cup \bigcup \{K_\alpha - F\}$$

is also connected.

(3.2) COROLLARY. *If X is a compact and continuously partially ordered space, if F is a closed subset of X such that $F = L(F)$ and if, for each $x \in X$, each component of $L(x)$ meets F, then, for each $x \in X - F$, there exists $y < x$ such that y and x are joined by an order arc in X.*

Proof. If $x \in X - F$, then, in X/F, there exists an order arc A_x joining F and x. Let $y \in A_x - \{x\} \cup \{F\}$; then $y < x$ and an order arc joins y and x in X/F. Since this arc is disjoint from F, it remains an order arc in X.

In the following sections we shall also require a simple lemma about compact partially ordered spaces.

(3.3) LEMMA. *If A is a closed subset of a compact, continuously partially ordered space, then $L(A)$ is a closed set.*

Proof. Let Γ denote the graph of the partial order. Choquet [2] first observed[1] that in a continuously partially ordered space the set functions L and M are upper semi-continuous. Therefore, if $x \notin L(A)$, there is an open set U with $x \in U$ such that $M(t) \cap A = 0$ for each $t \in U$. Therefore $U \cap L(A) = 0$, so that $L(A)$ is closed.

4. The locally compact case. Very simple examples exist to show that Koch's theorem fails if X is assumed only to be locally compact. For later reference we describe one of these.

(4.1) EXAMPLE. *There exists a locally compact and continuously partially ordered space Y with zero, 0, such that $L(x)$ is connected,*

[1] I am indebted to the referee for this reference.

for each $x \in Y$, *but certain elements of* $Y - \{0\}$ *are not joined to* 0 *by an arc.*

In the Cartesian plane let A_{-1} denote the closed line segment whose endpoints are (0, 0) and (1, 0), A_0 is the closed line segment whose endpoints are (1, 0) and (1, 1), and, for each $n = 1, 2, \cdots, A_n$ is the closed line segment whose endpoints are $(1 - 2^n, 0)$ and $(1 - 2^n, 1)$. Let

$$X = \bigcup_{n=-1}^{\infty} \{A_n\} .$$

In the relative topology X is a compact space. Give X the coordinatewise partial order, i.e., $(a, b) \leqq (c, d)$ if and only if $a \leqq c$ and $b \leqq d$. Then it is easy to see that X satisfies the hypotheses of Theorem 2.1, with the origin for zero.

Now let S be a closed segment of A_0 which does not contain (1, 1), and let $Y = X - S$. Then Y is a locally compact space which otherwise satisfies all the hypotheses of Theorem 2.1, but no arc joins 0 to (1, 1).

The space Y is even a topological semi-lattice. The author and L. W. Anderson [1] have shown that if a connected and locally compact topological lattice has a zero, then each point is connected by an order arc to zero, and, under suitable auxiliary hypotheses, the same is true of locally compact semi-lattices, but our methods depended very strongly on the lattice structure.

With no additional hypotheses at all, however, some results can be obtained in the locally compact case, using Lemma 3.1 and Corollary 3.2.

(4.2) THEOREM. *Let* X *be a continuously partially ordered space, let* $p \in X$, *and suppose* p *admits a compact neighborhood* N *which contains no minimal elements of* X. *If* $L(x)$ *is connected, for each* $x \in N$, *then there exists* $q \in L(p) - \{p\}$ *such that* q *and* p *lie in an order arc.*

Proof. Let B denote the boundary of N and define

$$F = L(L(p) \cap B) \cap N .$$

We assert that $L(p) \cap B$ is not empty, for otherwise the connectivity of $L(p)$ insures that $L(p) \subset N$; but then $L(p)$ is compact and hence contains a minimal element of X. But, by hypothesis, N contains no minimal elements of X. Moreover, since $p \in L(p) - B$, it follows that $p \in L(p) \cap (N - F)$. By Lemma 3.3, F is a closed subset of $L(p) \cap N$. If $x \in L(p) \cap (N - F)$ then the connectivity of $L(x)$ guarantees that each component of $L(x) \cap N$ meets F. Therefore, the space $L(p) \cap N$ satisfies the hypotheses of Corollary 3.2, and the theorem follows.

Referring to the space Y of Example 4.1, the point $(1, 1)$ can certainly be joined by an order arc to a point $(1, 1 - \varepsilon) < (1, 1)$. In order to continue this arc on to 0 it is necessary to add some further hypothesis such as is contained in our next result.

(4.3) THEOREM. *Let X be a locally compact, continuously partially ordered space with zero, 0, and suppose $L(x)$ is connected, for each $x \in X$. If each closed chain of X has a zero, then each $x \in X - \{0\}$ is joined to 0 by an order arc.*

Proof. If $x \in X - \{0\}$, then Theorem 4.2 assures us that x is the supremum of a nontrivial connected chain. Let C be a maximal such chain; by hypothesis, $z(C)$, the zero of C, exists. If $z(C) \neq 0$, then another application of Theorem 4.2 produces a nontrivial connected chain D, of which $z(C)$ is the supremum. But the chain $C \cup D$ is connected and thus contradicts the maximality of C. Thus C is an order arc joining x to 0.

We note that Theorem 4.3 truly generalizes Corollary 2.2 because, in the compact case, every closed chain has a zero.

Problem. Does Theorem 4.3 remain true if the hypothesis that each closed chain has a zero is weakened to "each chain has an infimum"?

5. Partially ordered spaces without zero. Let K be a continuum which contains no arc. Select $x_1 \in K$ and define $x \leqq y$ if and only if $y = x_1$ or $y = x$. With respect to this relation K is a compact continuously partially ordered space in which each set $L(x)$ is connected but in which there are no arcs. Thus we cannot infer the existence of order arcs without some restrictions on the set of minimal elements, but the hypothesis of Corollary 2.2 that there is only one minimal element is unduly restrictive.

(5.1) THEOREM. *Let X be a compact, continuously partially ordered space in which $L(x)$ is connected, for each $x \in X$. Let E denote the set of minimal elements of X, and suppose, for each $x \in X - E$, that $x \in X - Cl(L(x) \cap E)$. Then each $x \in X - E$ is joined by an order arc to some element of E.*

Proof. Let $x \in X - E$; since $L(x)$ is also a compact, continuously partially ordered space, $L(x) \cap E$ is not empty. Let

$$E_x = L(Cl(L(x) \cap E))$$

and note that $x \in L(x) - E_x$ and, by Lemma 3.3, that E_x is closed. By Corollary 3.2, x is the supremum of a nondegenerate connected chain. The proof now follows that of Theorem 4.3. If C is a maximal connected chain such that $x = \sup C$, then by compactness C has a zero which, by maximality, is a member of E.

(5.2) COROLLARY. *Let X be a locally compact, continuously partially ordered space in which each closed chain has a zero, and in which, for each $x \in X - E$, it follows that $x \in X - Cl(L(x) \cap E)$, where E denotes the set of minimal elements of X. If $L(x)$ is connected, for each $x \in X$, then each $x \in X - E$ is joined by an order arc to some element of E.*

Proof. If $x \in X - E$ then by Theorem 4.2, x is the supremum of some nondegenerate connected chain. If C is a maximal chain with this property, then C is closed and, by maximality, its zero is an element of E.

(5.3) COROLLARY. *Let X be a locally compact, continuously partially ordered space in which each closed chain has a zero, and in which the set E of minimal elements is closed. If $L(x)$ is connected, for each $x \in X$, then each $x \in X - E$ is joined by an order arc to some element of E.*

Some authors have called a partial order on a space "continuous" if the set-valued mapping L is continuous in the following sense: that each set $L(x)$ is closed and, if U and V are open sets such that $L(x) \subset U$ and $L(x)$ meets V, then there exists an open set W containing x such that, if $y \in W$, then $L(y) \subset U$ and $L(y)$ meets V. If a partial order satisfies this condition, let us say that the space is an *L-continuous partially ordered space.* It is a simple exercise to verify that L-continuity of a partial order implies continuity. (See Choquet [2].)

(5.4) THEOREM. *If X is an L-continuous partially ordered space, then the set E of minimal elements of X is closed.*

Proof. If $x \in X - E$ then there exists $p < x$ and hence, if U is a neighborhood of p, $L(x) \cap U$ is not empty. We may select U such that $x \in X - \bar{U}$. By L-continuity, there exists an open set W such that $x \in W \subset X - \bar{U}$ and, for each $t \in W$, $L(t) \cap U$ is not empty. In particular, $L(t)$ is nondegenerate and hence $W \cap E$ is empty.

(5.5) COROLLARY. *If X is a locally compact, L-continuous partially ordered space in which each closed chain has a zero, and if*

$L(x)$ is connected, for each $x \in X$, then each non-minimal element of X is joined by an order arc to some minimal element of X.

6. Concluding remarks. We return to consideration of the theorem that a locally connected, metrizable continuum is real arcwise connected. The problem we wish to raise may be put in this way: *Does Koch's theorem imply the arcwise connectivity of such continua?* Since Mardešić has shown [5] that the natural analog of this result fails in the nonmetrizable case, metrizability (or some slightly weaker condition) must certainly be assumed. Now it can be shown that any locally connected continuum admits a nontrivial quasi-order (i.e., a reflexive, transitive relation) which is continuous, has a zero, and is such that each set $L(x)$ is connected. By an argument similar to that of Lemma 2.3 one can find a minimal quasi-order with the same properties. If, under suitable conditions, this minimal quasi-order is found to be a partial order, then arcwise connectivity would follow from Corollary 2.2.

REFERENCES

1. L. W. Anderson and L. E. Ward, Jr., *A structure theorem for topological lattices,* Proc. Glasgow Math. Assoc. **5** (1961), 1-3.
2. G. Choquet, *Convergences,* Annales Grenoble Univ., Sec. des Sci., Math. et Phys. **23** (1947), 58-112.
3. R. P. Hunter, *Note on arcs in semigroups,* Fund. Math. **49** (1961), 233-245.
4. R. J. Koch, *Arcs in partially ordered spaces,* Pacific J. Math. **9** (1959), 723-728.
5. S. Mardešić, *On the Hahn-Mazurkiewicz theorem in non-metric spaces,* Proc. Amer. Math. Soc. **11** (1960), 929-937.
6. A. D. Wallace, *Relations on topological spaces,* Proceedings of the Symposium on General Topology, Prague, (1961), 356-360.
7. L. E. Ward, Jr. *Partially ordered topological spaces,* Proc. Amer. Math. Soc. **5** (1954), 144-161.

UNIVERSITY OF OREGON

PACIFIC JOURNAL OF MATHEMATICS
Vol. 15, No. 1, 1965

A NEW MEASURE OF A PARTIAL DIFFERENTIAL FIELD EXTENSION

ISRAEL ZUCKERMAN

Let G be a differential field of characteristic zero with the commuting derivations $d_1, \cdots, d_m$. If F is a differential subfield of G, the algebraic and differential degrees of transcendence of G over F, denoted respectively by $d(G/F)$ and $d.d(G/F)$ are numerical invariants of the extension. Unlike the ordinary differential case $(m=1)$ $d.d.$ $(G/F)=0$ does not imply that $d(G/F)$ is finite. In this paper an intermediate measure of the extension is constructed, called the limit vector. The first and last components of this vector correspond to $d.d(G/F)$ and $d(G/F)$ respectively, and the limit vector is additive.

Similar concepts have been developed independently by Kolchin in a work not yet published.

Characteristic sets of prime ideals as defined in [6] play a prominent role in the development of the limit vector, as well as in the development of other results of this paper which do not depend on the limit vector. Further, it is shown that an intermediate field of a finitely generated extension is finitely generated. Kolchin has a prior, but different proof of this. Kolchin's analog of Lüroth's Theorem [2, 3] is extended and some results on characteristic sets of length one are obtained.

Raudenbush [5] shows that the dependence axioms of Van der Waerden [7] are satisfied by differential dependence. It is indicated below that these axioms are more readily established by use of the limit vector. A further result is that a proper specialization of $F[a]$ must reduce the limit vector only if a has a characteristic set of length one over F. A short proof of a theorem of Delsarte [1] on partial linear homogeneous differential equations concludes the paper.

2. Ordering the derivatives. The main source of this subject is Ritt [6], especially § 8 and § 2 of Chapter VIII and §§1–16 of Chapter IX. In general, the terminology and notation are as in [6]. Consider the differential ring $F\{y_1, \cdots, y_r\}$, where the y_i are differential indeterminates. Then $D = d_1^{i_1} \cdots d_m^{i_m}$ will denote a derivative, i.e., the composite of derivations. We associate with D the vector $(i_1, \cdots, i_m)$. The sum of the i_k is called the order of D, or of the associated vector,

Received January 23, 1964. This paper is the author's doctoral dissertation submitted to Rutgers University in March 1963 under the direction of Professor Richard Cohn. The work was partly supported by the National Science Foundation Grant G-19636.

or of Dy_j. y_j is to be considered a derivative of itself of zero order. Let w denote an arbitrary y-derivative. A set of marks will be assigned to the d_i and y_i to achieve a complete ordering of the w. In doing so, the primary objective will be to have w_1 precede w_2 if the order of w_1 is less than the order of w_2. Associate with each d_i, the marks $u_{i1}, \cdots, u_{i,m+2}$; where $u_{i1} = 1$, $u_{i2} = 0$, and for $k > 2$: $u_{ik} = 0$ for $k \neq i+2$, $u_{ik} = 1$ for $k = i + 2$. Associate with each y_i the marks $v_{i1}, \cdots, v_{i,m+2}$, where $v_{i2} = i$ and $v_{ik} = 0$ for $k \neq 2$. Let $w = d_1^{i_1} \cdots d_m^{i_m} y_i$. Then w is assigned $m+2$ marks as follows: The jth mark of w is $v_{ij} + i_1 u_{1j} + \cdots + i_m u_{mj}$. Let w_1 and w_2 be two y-derivatives with marks a_i and b_i respectively. Then we shall say that w_1 precedes w_2 or succeeds w_2 according as the first nonzero difference $b_i - a_i$ is positive or is negative. As one can easily verify the system of marks introduced here achieves the desired complete ordering of the y-derivatives. Such ordering will prevail throughout, with the exception of § 12.

3. A transcendence basis for $a_1, \cdots, a_r$ over F. Let $a_1, \cdots, a_r$ be elements in G. If the a_i are differentially dependent over F, let P be the prime ideal in $F\{y_1, \cdots, y_r\}$ with $(a_1, \cdots, a_r)$ as a generic zero, and let $C = (A_1, \cdots, A_k)$ be a characteristic set of P. (C is said to be a characteristic set of the a_i over F.) Denote the leader and separant of A_i by p_i and S_i respectively. If w is a derivative of some p_i, it will be called principal; otherwise, it will be called parametric. We will call an a-derivative principal if it admits a representation $w(a)$ where w is principal; otherwise, it will be called parametric. If the a_i are differentially independent over F, all a-derivatives will be considered to be parametric. The ordering of the y-derivatives carries over to the ordering of the symbols for the a-derivatives.

THEOREM 1. *The parametric derivatives of $a_1, \cdots, a_r$ are distinct and constitute a transcendence basis for $F(a_1, \cdots, a_r)$ over F. Moreover, a principal a-derivative depends algebraically on parametric derivatives of equal or lower order.*

Proof. If the a_i are differentially independent over F, the proof is immediate. Now assume that the a_i are differentially dependent over F, with the associated prime ideal P and the characteristic set C. Since C is a characteristic set of P, every differential polynomial, (d.p.), in P involves a derivative of some p_i. Hence the parametric a-derivatives are distinct and algebraically independent over F. This proves the first statement of the theorem.

Now, the principal a-derivatives are completely ordered with $p_1(a)$ as the first derivative. Since p_1 is the leader of A_1, $p_1(a)$ depends solely on parametric derivatives. Let $w(a_j)$ be a principal a-derivative succeeding $p_1(a)$. If w is a proper derivative of some leader, say p_i, then an appropriate derivative D of A_i yields

$0 = DA_i(a_1, \cdots, a_r) = S_i(a_1, \cdots, a_r)w(a_j) + B(a_1, \cdots, a_r)$, where the a-derivatives in $S_i(a)$ and $B(a)$ precede $w(a_j)$. w may also be some p_i for $i > 1$. In either case, $w(a_j)$ depends on principal and parametric derivatives whose symbols precede the symbol for $w(a_j)$. By induction, each of these principal derivatives depends on preceding parametric derivatives; hence on parametric derivatives which precede $w(a_j)$. Thus, $w(a_j)$ depends on preceding parametric derivatives, concluding the proof.

4. Restriction to a-derivatives with orders not exceeding n. Let A be a subset of G, and n a positive integer. Then the set of derivatives of elements of A of order not exceeding n will be denoted by $(n; A)$. Let F' be an ordinary field contained in G. Then, as is customary, $F'(n; A)$ will denote the ordinary field extension obtained by adjoining $(n; A)$ to F'. $d(F'(n; A)/F')$ will be denoted by $h(F', n; A)$ or simply by $h(n)$ if no ambiguity arises. (This type of abbreviation is repeated throughout.)

With this notation the following corollary is immediate, noting that the ordering of the a-derivatives is such that a principal derivative depends on parametric derivatives of equal or lower order.

COROLLARY 1. *$h(F, n; a_1, \cdots, a_r)$ is the number of parametric derivatives of the a_i of order not exceeding n.*

THEOREM 2. *For n sufficiently large,*

$$h(F, n; a_1, \cdots, a_r) = H(F, n; a_1, \cdots, a_r)/m!\ ,$$

where $H(n)$ is a polynomial in n with integral coefficients. $H(n) = 0$, or has degree $t \leqq m$ and leading coefficient $c_t > 0$.

Proof. Let $p(n)$ denote the number of derivatives of some y of order not exceeding n. Then

$$p(n) = C(n + m, m) = (n + m)(n + m - 1) \cdots (n + 1)/m!$$

Hence, if the a_i are differentially independent over F, $h(n) = rp(n)$ and the theorem is true in such case with $t = m$ and $c_t = r$.

Now assume that the a_i are differentially dependent over F with leaders p_i of a characteristic set for the a_i over F. Partition the p_i into subsets R_j, each consisting of derivatives of the same y_j. Let $q_1, \cdots, q_s$ be one such subset. For each nonempty subset T of $\{q_1, \cdots, q_s\}$, let $u(T)$ be the m-vector with kth component, $k(T)$, equal to the

maximum of the kth components of the vectors associated with the q_i in T, and let v_T be the sum of the $k(T)$. Let n be a positive integer greater than the maximum of the v over all R_j. Then an m-vector will be called a multiple of $u(T)$ if each of its components is not less than the corresponding component of $u(T)$, and if the sum of its components does not exceed n. Let $S(T)$ denote the set of multiples $u(T)$, and let $N(T)$ denote the number of elements in $S(T)$. Let $B(T)$ denote the number of elements in T.
Then,

$$N(T) = C(n + m - v_T, m) . \tag{1}$$

Note that $p(n)$ is of the form (1) with $v = 0$. Therefore we extend the set of subsets T to include the empty set φ and define $v_\varphi = 0$, $N(\varphi) = p(n)$. The number of parametric derivatives of y_j of order not exceeding n is obtained by subtracting from $p(n)$ the number of elements in the union of the $S(T)$. Thus, it is equal to

$$\sum_T (-1)^{B(T)} N(T) . \tag{2}$$

(2) is the sum of 2^s expressions, each of the form (1). The term of highest degree in n in each expression is the same, $n^m/m!$ Since there are as many positive as negative expressions in (2), the sum of the terms in n^m is zero. Furthermore, by consideration of large n we see that the effective leading coefficient of (2) is positive.

The sum of the number of parametric derivatives of the R_j is then a polynomial with the desired properties. (An empty R_j contributes $p(n)$ elements.)

The following corollary is immediate.

COROLLARY 1. *Let s be an integer with $0 \leqq s \leqq m$. If $a_1, \cdots, a_r$ are differentially dependent over F, then*

$$\lim_{n\to\infty} n^s h(n)/p(n) = \begin{cases} 0 , & \text{if } s < m - t ; \\ c_t , & \text{if } s = m - t ; \\ \infty , & \text{if } s > m - t . \end{cases}$$

If $a_1, \cdots, a_r$ are differentially independent over F, then

$$\lim_{n\to\infty} n^s h(n)/p(n) = \begin{cases} r , & \text{for } s = 0 ; \\ \infty , & \text{for } 0 < s < m . \end{cases}$$

5. Introduction of the limit vector.

We may now define

$$L_s(F, a_1, \cdots, a_r) = \lim_{n\to\infty} n^s h(n)/p(n) \qquad (s = 0, \cdots, m) ,$$

and

$$L(F, a_1, \cdots, a_r) = (L_0, \cdots, L_m) .$$

The latter will be called the limit vector of $a_1, \cdots, a_r$ over F.

The following remarks are evident.

(1) $$d.d(F\langle a_1, \cdots, a_r\rangle/F) = r \Leftrightarrow L_0 = r.$$

(2) For a single element a, $d.d(F\langle a\rangle/F) = 0 \Leftrightarrow L_0 = 0$.

We will later show that $d.d(F\langle a_1, \cdots, a_r\rangle/F) = L_0$, subsuming (1) and (2).

(3) $$d(F\langle a_1, \cdots, a_r\rangle/F) = k < \infty \Leftrightarrow L_m = km!$$

In particular, if each a_i is algebraic over F, $L_s = 0$ for all s.

6. **Results on L_1.** The following corollaries follow from the proof of Theorem 2. In this section, a will be differentially algebraic over F with characteristic set $C = A_1, \cdots, A_k$, $0 < k < \infty$. W_i will denote the vector associated with the leader of A_i.

COROLLARY 2.

$$L_1(F; a) = m \sum_T (-1)^{B(T)+1} v_T .$$

In particular, this shows that $L_1(F; a)$ is divisible by m.

Proof. The coefficient of n^{m-1} in $m!\, C(n + m - v_T, m)$ is

(3) $$(m - v_T) + \cdots + (1 - v_T) = m(m + 1)/2 - mv_T .$$

The first term in (3) is the same for each v_T. Since there are as many positive as negative expressions in (2), these terms cancel in computing C_{m-1}, and the desired result follows.

As a special case of Corollary 2, we have the following.

COROLLARY 3. *If a has a characteristic set of length one over F, then $L_1(F; a) = mg$ where g is the order of such characteristic set.*

The result of Corollary 2 may be carried further so as to depend more directly on the leaders of the characteristic set. We need the following lemma.

LEMMA. *Let $u(1), \cdots, u(k)$, $1 \leqq k < \infty$, be a sequence S of real numbers with $u(1) = min\,(u(i))$. For each subsequence T of S, let $B(T)$ and $M(T)$ denote respectively the number and maximum of its elements. Then,*

$$\sum_T (-1)^{B(T)+1} M(T) = u(1) .$$

Proof. Let T' denote a T which does not have $u(1)$ as its first element; and let T'' be the sequence obtained by adjoining $u(1)$ at the beginning of T'. Then, with the exception of $T = u(1)$, the T can be partitioned into pairs of T' and T'' such that

$$(-1)^{B(T')+1} M(T') + (-1)^{B(T'')+1} M(T'') = 0 .$$

Thus, the desired result is obtained.

COROLLARY 4. *Let w_j denote the minimum of the* jth *components of the W_i. Then*

$$L_1(F; a) = m \sum_{j=1}^{m} w_j .$$

Proof. By the lemma, the sum of the jth components of the vectors associated with the T, with the appropriate signs affixed, is w_j.

Hence, if we sum the $(-1)^{B(T)+1} v_T$ component-wise, the result of Corollary 2 yields $L_1(F; a) = m \sum_{j=1}^{m} w_j$, as desired.

Note that Corollary 4 implies that $L_1(F; a) = 0$ if and only if $w_j = 0$ for all j.

COROLLARY 5. *Let A_t of order v belong to C, and let $k > 1$. Then $L_1(F; a) < mv$.*

Proof. If for $j = 1, \cdots, m$, the jth component of W_t is the minimum of the jth components of the W_i, then for $i \neq t$, A_i would not be reduced with respect to A_t.

We are now in a position to prove a converse of Corollary 3.

COROLLARY 6. *If $L_1(F; a) = mg$, $0 \leqq g < \infty$, and if a satisfies an irreducible d.p. A over F of order g, then A is a characteristic set for a over F.*

Proof. Since a satisfies A, there exists a *d.p.* A' in C with order $g' \leqq g$. By Corollary 5, if $k > 1$, $L_1(F; a) < mg' \leqq mg$. Hence, $C = A'$. By Corollary 3, $g' = g$ and A' has the same order as A. Since A has zero remainder with respect to A' and is irreducible, $A = cA'$ where c is in F; hence, A is a characteristic set for a over F.

7, Additivity of the limit vector.

LEMMA 1. *Let A be a finite set of elements contained in an ex-*

tension of F. If k, q, and r are nonnegative integers and s an integer such that $0 \leqq s \leqq m$, then

$$\lim_{n\to\infty} [(n+k)^s h(n+q, A)/p(n+r)] = \lim_{n\to\infty} [n^s h(n, A)/p(n)] .$$

Proof.

$$\frac{(n+k)^s h(n+q, A)}{p(n+r)} = \frac{(n+q)^s h(n+q, A)}{p(n+q)} \frac{(n+k)^s}{(n+q)^s} \frac{p(n+q)}{p(n+r)} .$$

Clearly,

$$\lim_{n\to\infty} \frac{p(n+q)}{p(n+r)} = 1 , \quad \text{and} \quad \lim_{n\to\infty} \frac{(n+k)^s}{(n+q)^s} = 1 .$$

Furthermore,

$$\lim_{n\to\infty} \frac{(n+q)^s h(n+q, A)}{p(n+q)} = \lim_{n\to\infty} \frac{n^s h(n, A)}{p(n)} .$$

LEMMA 2. *Let A be a set and b an element contained in an extension of F with b differentially algebraic over $F\langle A\rangle$. Let t be the maximal order of derivatives of A appearing in a characteristic set C for b over $F\langle A\rangle$. Let A' be a set of A-derivatives containing $(n+t; A)$. Then,*

$$h(F\langle A\rangle, n; b) = h(F(A'), n; b) .$$

Proof. Let $S(n)$ denote the set of parametric derivatives of b of order not exceeding n with respect to C. Then $S(n)$ is algebraically independent over $F(A')$. Furthermore, $S(n)$ is an algebraic spanning set for $(n; b)$ over $F(A')$ since in the proof of Theorem 1 only derivatives with orders not exceeding n of the *d.p.* in C are present in the algebraic relations obtained for the principal derivatives.

Note that the result holds more readily if b is differentially transcendental over $F\langle A\rangle$.

THEOREM 3.

$$L(F; a_1, \cdots, a_r) = \sum_{i=1}^{r} L(F\langle a_1 , \cdots, a_{i-1}\rangle; a_i) .$$

Proof. For $i = 1, \cdots, r-1$, if a_{i+1} is differentially transcendental over $F\langle a_1, \cdots, a_i\rangle$, let $t_i = 0$. Otherwise, t_i be the maximal order of derivatives of $a_1, \cdots, a_i$ in a characteristic set for a_{i+1} over $F\langle a_1, \cdots, a_i\rangle$. Then,

$$\begin{aligned}&F(n+t_1+\cdots+t_{r-1}; a_1, \cdots, a_r)\\ &\qquad\supset F(n+t_1+\cdots+t_{r-1}; a_1)(n+t_2+\cdots+t_{r-1}; a_2)\\ &\qquad\cdots(n+t_{r-1}; a_{r-1})(n; a_r)\supset F(n; a_1, \cdots, a_r)\,.\end{aligned}$$

Then, by additivity of the transcendence degree and Lemma 2,

$$\begin{aligned}&h(F, n+t_1+\cdots+t_{r-1}; a_1, \cdots, a_r)\\ &\qquad\geqq \sum_{i=1}^{r} h(F\langle a_1, \cdots, a_{i-1}\rangle, n+t_i+\cdots+t_{r-1}; a_i)\\ &\qquad\geqq h(F, n; a_1, \cdots, a_r).\end{aligned}$$

Multiplying by $n^s/p(n)$, then taking the limit of the resulting expressions as n approaches ∞, by application of Lemma 1, the desired result is obtained.

8. Extension of the limit vector to a measure of an arbitrary differential field extension. Let $f(G)$ denote the set of finite subsets of an extension G of F, including the null set. For S and T in f, suppose that k is the first component for which $L(F; S)$ and $L(F; T)$ differ, and that $L_k(F; S) < L_k(F; T)$. Then we will write $L(F; S) < L(F; T)$ and we may define $L(G/F) = \sup_{s\in f} L(F; S)$.

COROLLARY 1. *If G is finitely generated over F, i.e., $G = F\langle a_1, \cdots, a_r\rangle$, then $L(G/F) = L(F; a_1, \cdots, a_r)$.*

Proof. Let $A = \{a_1, \cdots, a_r\}$ and B belong to f. Then $G = F\langle A, B\rangle$. By Theorem 3, $L(F; B) \leqq L(F; A, B) = L(F; A) + L(F\langle A\rangle, B)$. Since $B \subset F\langle A\rangle$, $L(F\langle A\rangle; B) = 0$. Hence, $L(F; B) \leqq L(F; A)$ for all B in f.

Immediate consequences are:

COROLLARY 2. *Given $G = F\langle A\rangle = F\langle B\rangle$, where both A and B belong to f. Then $L(F; A) = L(F; B)$.*

COROLLARY 3. *If H is finitely generated over G, and G is finitely generated over F, then*

$$L(H/F) = L(H/G) + L(G/F)\,.$$

9. A general additivity theorem.

THEOREM 4. *Given $F \subset G \subset H$. Then, $L(H/F) = L(G/F) + L(H/G)$.*

Proof. (Part I): To prove $L(H/F) \geqq L(G/F) + L(H/G)$. Let $g(H)$ denote the set of finite subsets of H, including the null set, which contain no elements of G. Then every V in $f(H)$ is the unique union of an S in $f(G)$ and a T in $g(H)$, and conversely. For a particular $V = S \cup T$,

$$L(F;\, V) = L(F;\, S) + L(F\langle S\rangle;\, T \geqq L(F;\, S) + L(G;\, T)\ .$$

Hence,

$$\begin{aligned}\sup_{V \in f(H)} L(F;\, V) &\geqq \sup_{V} (L(F;\, S) + L(G;\, T)) \\ &= \sup_{S \in F(G)} L(F;\, S) + \sup_{T \in g(H)} L(G;\, T) = L(G;\, F) + L(H;\, G)\ .\end{aligned}$$

Proof. (Part II): To prove $L(H/F) \leqq L(G/F) + L(H/G)$. It suffices to show that $L(F\langle h_1, \cdots, h_r\rangle/F) \leqq L(H/G) + L(G/F)$ for any finite subset $h_1, \cdots, h_r$ of H. Let $H_1 = G\langle h_1, \cdots, h_r\rangle$, and let P be the prime ideal in $G\{y_1, \cdots, y_r\}$ with $h_1, \cdots, h_r$ as a generic zero. If $P = 0$, let $G_1 = F$. Otherwise, let $G_1 = F\langle A\rangle$, where A is the set of coefficients of a characteristic set C of P. Let P_1 be the prime ideal in $G_1\{y_1, \cdots, y_r\}$ with $h_1, \cdots, h_r$ as a generic zero. Then C is also a characteristic set of P_1. Hence, $L(H_1/G) = L(G_1\langle h_1, \cdots, h_r\rangle/G_1)$.

By additivity for the finitely generated case,

$$\begin{aligned}L(F\langle h_1, \cdots, h_r\rangle/F &\leqq L(G_1\langle h_1, \cdots, h_r\rangle/F) \\ &= L(G_1\langle h_1, \cdots, h_r\rangle/G_1) + L(G_1/F) \\ &= L(H_1/G) + L(G_1/F) \\ &\leqq L(H/G) + L(G/F)\ .\end{aligned}$$

10. Remarks on L_0. (1) Almost all of the previous results on L_0 could be obtained more readily: If a is differentially algebraic over F, it is sufficient to consider a single d.p. $B(y)$ in $F\{y\}$ which has a as a nonsingular solution. (i.e., a does not annul the separant of $B(y)$.)

(2) In order to establish the theory of differential dependence for arbitrary m, Raudenbush showed in [3] that differential dependence satisfied the dependence axioms of Van der Waerden [5]. However this theory follows immediately from the results on L_0. Furthermore, since it suffices to consider differentially independent subsets of G in determining $L_0(G/F)$, it is clear that d.d $(G/F) = L_0(G/F)$. Hence the additivity of d.d. follows from the additivity of L_0.

11. A result on finitely generated extensions.

THEOREM 5. *Given $F \subset G \subset H = F\langle a_1, \cdots, a_r\rangle$. Then G is finitely generated over F.*

Proof. If the a_i are differentially independent over G, then it follows readily that $F = G$. Hence the theorem is true in such case.

If the a_i are differentially dependent over G let Q be the set of coefficients of a characteristic set $A_1, \cdots, A_k$ of the a_i over G. Then we assert that $G = F\langle Q\rangle$. For let p_i be the leader of A_i and R the set of parametric derivatives of the a_i. Construct

$$G' = F\langle Q\rangle , \qquad G'^* = G'(R) , \qquad G^* = G(R) .$$

Then

$$H = G'^*(p_1, \cdots, p_k) = G^*(p_1, \cdots, p_k) ,$$

and

$$[H : G'^*] = [H : G^*] = \prod_{i=1}^{k} (\deg A_i \text{ in } p_i) < \infty .$$

Thus

$$G'^* = G^* .$$

Moreover, since the parametric derivatives of the a_i are transcendental over both G and G', $[G : G'] = [G^* : G'^*] = 1$. Hence, $G = G'$ and the theorem is proved.

12. Characteristic sets of length one and Lüroth's theorem. Kolchin [1] and [2] proves the differential analog of Lüroth's theorem for ordinary differential fields of characteristic zero; explicitly, if $F \subset G \subset F\langle y\rangle$, then $G = F\langle a\rangle$, for some a in G. With minor changes Kolchin's proof goes through for the partial case, provided that y has a characteristic set of length one over G.[1]

Using the ideas in [1] and [2], we establish the following converse.

THEOREM 6. *If* $F \subset F\langle a\rangle \subset F\langle y\rangle$, *where* $a \notin F$, *then* y *has a characteristic set of length one over* $F\langle a\rangle$.

Proof. Since $a \in F\langle y\rangle$, $a = P(y)/Q(y)$, where P and Q may be taken to be relatively prime d.p. over F. We assert that $A(z) = aQ(z) - P(z)$ is a characteristic set of y over $F\langle a\rangle$. To begin with, A is clearly irreducible over $F\langle a\rangle$.

Let Y be the prime ideal over $F\langle a\rangle$ with y as a generic zero. Let X be the prime ideal over $F\langle a\rangle$ with A as a characteristic set, and let x be a generic zero of X. x is differentially transcendental over F; for otherwise a, and hence y would be differentially algebraic over F. Therefore, mapping y onto x determines an F-isomorphism

1. In Kolchin's proof, [2] on the 9th and 10th lines from the top on page 400, the result follows by considering the derivatives ordered so that every z derivative is higher than every y derivative. On the 16th line from the top on the same page, the y derivatives need to be considered higher. These orderings differ from those used in this paper.

of $F\langle y\rangle$ onto $F\langle x\rangle$. Under this isomorphism, a remains fixed, for $a = P(y)/Q(y) = P(x)/Q(x)$. Hence the isormorphism leaves $F\langle a\rangle$ fixed. Thus y and x satisfy the same d.p.'s over $F\langle a\rangle$, and y is a generic zero of Q. Hence, $X = Y$ and A is a characteristic set of X.

13. The length of a characteristic set is not a property of the extension. If $F\langle a\rangle = F\langle b\rangle$, and if a has a characteristic set of length one over F, must this also be true of b? The answer is in the negative as the following example will indicate.

EXAMPLE. Let u denote a differential indeterminate and let $m = 2$ with the derivations denoted by subscripts x and y as in u_x and u_y.

Let P be the set of d.p. with zero remainder with respect to u_x. Then P is a prime differential ideal. Since the initial and separant of u_x are both one, P is also generated by u_x over F. Let a be a generic zero of P. Define

(1) $$b = a + pa_y ,$$

where p is in F and is differentially transcendental over the field of rationals contained in F. Then

(2) $$b_x = p_x a_y ,$$

(3) $$b_y = a_y + p_y a_y + pa_{yy} = (1 + p_y)a_y + pa_{yy} ,$$

(4) $$b_{xx} = p_{xx} a_y .$$

Then (1) and (2) imply

(5) $$a = b - pa_y = b - p/p_x(b_x) ,$$

showing that a is in $F\langle b\rangle$. (2) and (4) imply

(6) $$b_x - p_x b_{xx}/p_{xx} = 0 ,$$

showing that b satisfies a second order d.p. over F.

Equations (1), (2), and (3) may be solved for a, a_y, a_{yy} in terms of b, b_x and b_y. Hence, since a, a_y, a_{yy} are algebraically independent over F, b cannot satisfy a first order d.p. over F. Equation (6) yields the irreducible d.p. satisfied by b,

(6′) $$A = u_x - p_x u_{xx}/p_{xx} ,$$

which may be chosen as a d.p. in a characteristic set for b.

By Corollary 3 to Theorem 2, $L_1(F; a) = 2$. Since $F\langle a\rangle = F\langle b\rangle$, $L_1(F; b) = 2$. By the same corollary, A cannot be the sole d.p. in a characteristic set for b.

14. A simply generated extension with no generator having a characteristic set of length one. *Example*: Let P be the prime differential ideal generated by u_x and u_y over F. Then u_x and u_y also constitute a characteristic set of P. Let a be a generic zero of P, and let $G = F\langle a \rangle$.

Then $h(F, n; a) = 1$, for all positive integral n. Hence $L_1(G/F) = 0$. By Corollary 3 to Theorem 2, if G has a generator b with a characteristic set of length one, then $L_1(G/F) = mg$ where g is the order of such characteristic set. Hence, $g = 0$ and b is algebraic over F. This implies that a is algebraic over F which is a contradiction of the fact that a is a generic zero of P.

15. Specializations. a' is called a specialization of a if there exists a differential homomorphism of $F\{a\}$ onto $F\{a'\}$, taking a into a', and leaving F elementwise fixed. Since

$$h(F, n; a') \leqq h(F, n; a), \; L(F; a') \leqq L(F; a) .$$

We investigate when equality holds.

PROPOSITION 1. If a has a characteristic set of length one over F and if a' is a proper specialization of a, (i.e., $F\{a\}$ and $F\{a'\}$ are not isomorphic), then $L(F; a') < L(F; a)$.

Proof. Assume $L(F; a') = L(F; a)$, and let A be a characteristic set for a over F. Then by Corollaries 3 and 6 of Theorem 2, A is also a characteristic set for a' over F. Hence, $F\{a\}$ and $F\{a'\}$ are isomorphic. Thus the proposition is proved.

PROPOSITION 2. If a has a characteristic set of length exceeding one, a proper specialization need not reduce the limit vector.

Proof. The following example will prove the point. As in the example of § 13, we consider $F\{u\}$ with u a differential indeterminate and two derivations denoted by subscripts x and y. Let t be a generic zero of the prime differential ideal P in $F\{u\}$ with characteristic set and generator u_x. Also, let a be a generic zero of the prime differential ideal Q in $F\{u\}$ generated by and with a characteristic set, u_{xx} and u_{xy}. Since Q is properly contained in P, t is a proper specialization of a. But, by direct computation, or by Corollary 4 of Theorem 2, $L(F; a) = L(F; t) = (0, 2, \infty)$, proving the proposition.

16. Order of a prime ideal and systems of linear homogeneous d.p. Let P be a prime differential ideal in $F\{y\}$ with generic zero a.

The algebraic degree of transcendence of $F\langle a\rangle$ over F is called the order of P, (ord P). By Theorem 1, this is the number of parametric derivatives of a. We will show that this use of order agrees with the "order" of a system of linear homogeneous d.p. as used in the study of differential equations.

The following lemmas will lead to this result. The first two are nondifferential and are stated without proof. Lemma 3 is Kolchin's Lemma 1 in [4], where it is proved. Lemmas 4 and 5 are the differential analogues of Lemmas 1 and 2, respectively.

LEMMA 1. *Let S be a system of linear homogeneous polynomials in $K[x_1, \cdots, x_n]$ which is a vector space over K, where K is a nondifferential field. Then*

(a) *(S) is prime and contains no linear homogeneous polynomials which are not in S.*

(b) $$S \cap K[x_1, \cdots, x_r] = 0 = (S) \cap K[x_1, \cdots, x_r] = 0 \,.$$

LEMMA 2. *Let a set T of linear homogeneous polynomials with coefficients in a field K generate a vector space over K and over an extension L of K. Then elements of T linearly independent over K remain linearly independent over L, and the number of such elements in a maximal set is the dimension of both vector spaces.*

LEMMA 3. *Let C be the field of constants of the partial differential field F, and let $a_1, \cdots, a_n$ belong to F. If $a_1, \cdots, a_n$ are linearly dependent over C, then $W_{D_1,\cdots,D_n}(a_1, \cdots, a_n) = 0$ for every choice of derivatives $D_1, \cdots, D_n$ where $W_{D_1,\cdots,D_n}(a_1, \cdots, a_n) = \det(D_i a_j)$. Conversely, if $W_{D_1,\cdots,D_n}(a_1, \cdots, a_n) = 0$ for every choice of $D_1, \cdots, D_n$ of orders $\leqq n-1$, then $a_1, \cdots, a_n$ are linearly dependent over C.*

LEMMA 4. *Let S be a system of homogeneous linear partial d.p. in $F\{y\}$ which is a vector space over F and is closed under the derivations $d_1, \cdots, d_m$. Let T denote a set of y-derivatives. Then $[S]$ is prime and*

$$S \cap F[T] = 0 \Rightarrow [S] \cap F[T] = 0 \,.$$

Proof. Let V denote the set of finite sets of y-derivatives. Then by Part (a) of Lemma 1, if $U \in V$, $S \cap F[U]$ generates a prime ideal S_U in $F[U]$. Since the union of the S_U over all U in V is $[S]$, $[S]$ is prime.

Let T' be a finite subset of T. By Part (b) of Lemma 1,

$$S \cap F[T'] = 0 \Rightarrow S_{T'} \cap F[T'] = 0 \Rightarrow [S] \cap F[T] = 0 .$$

LEMMA 5. *Let S denote a set of homogeneous linear partial d.p. in $F\{y\}$, and let G be an extension of F. Let T denote a set of y-derivatives. If P and Q denote the prime differential ideals generated by S in*

$$F\{y\} \quad \text{and} \quad G\{y\} \quad \text{respectively,}$$

then

$$P \cap F[T] = 0 \Leftrightarrow Q \cap G[T] = 0 .$$

This implies that P and Q have the same orders.

Proof. Let S^* denote the system consisting of the d.p. in S and of their derivatives. Let B be a maximal linearly independent subset of S^* over F, therefore, by Lemma 2, over G. By Lemma 4, P contains a nonzero polynomial in members of T if and only if it contains such a polynomial which is linear homogeneous; that is, if and only if there is a linear dependence among members of T and B over F. In the same way Q contains a nonzero polynomial in the members of T if and only if the members of T and B are linearly dependent over G. But Lemma 2 shows that these conditions are equivalent. Thus the lemma is proved.

THEOREM 7. *Let S be a system of homogeneous linear partial d.p. in $F\{y\}$. If the set of solutions of S is of linear dimension k over constants, then $\operatorname{ord}[S] = k$.*

Proof. (a) $\operatorname{ord}[S] \geqq k$:

Let $u_1, \cdots, u_k$ be a linear basis for the solutions over constants. Let $c_1, \cdots, c_k$ be new constants algebraically independent over $F\langle u_1, \cdots, u_k\rangle$. Let $v = \sum c_i u_i$. Since the u_i are linearly independent over constants, by Lemma 3, for some set of derivatives $D_1, \cdots, D_k$ of orders $\leqq k-1$, $W_{D_1, \cdots, D_k}(u_1, \cdots, u_k) \neq 0$. Hence, $D_1 v, \cdots, D_k v$ are linearly independent over F. Since v belongs to the manifold of $[S]$, $[S]$ contains no linear homogeneous polynomial, and therefore by Lemma 4, contains no polynomial in $D_1 y, \cdots, D_k y$. Hence, $\operatorname{ord}[S] \geqq k$.

(b) $k \geqq \operatorname{ord}[S]$:

Let $\operatorname{ord}[S] = k$. Let v_1 be a generic zero of $[S]$. Then k derivatives, $D_1, \cdots, D_k$, of v_1 are algebraically independent over F. We define inductively a sequence of elements v_j which are solutions of S by the requirement that v_j be a generic zero of the differential ideal generated by S over $F\langle v_1, \cdots, v_{j-1}\rangle\{y\}$.

Then, by Lemma 5, $D_1 v_j, \cdots, D_k v_j$ are algebraically independent over $F\langle v_1, \cdots, v_{j-1}\rangle$. In particular, the $D_i v_j$, $i, j = 1, \cdots, k$, are algebraically independent over F. Hence, $W_{D_1,\cdots,D_k}(v_1, \cdots, v_k) \neq 0$. Then Lemma 3 implies that the v_j are linearly independent over constants. Thus $k \geqq \operatorname{ord}[S]$.

THEOREM 8. *Let P be a prime differential ideal in $F\{y\}$. If P contains a set of nonzero d.p. A_i, $i = 1, \cdots, m$, such that A_i involves only y-derivatives of the form $d_i^t y$ where t is a nonnegative integer, then ord P is finite.*

Proof. If P contains a d.p. free of proper y-derivatives, then ord $P = 0$, and the theorem holds. Hence, we may assume that the A_i involve proper y-derivatives. Then A_i involves a derivative of the leader of some member of a characteristic set C of P. Hence, there exist leaders of members of C of the form $d_i^{t_i} y$, $t_i > 0$, $i = 1, \cdots, m$. Thus, by Theorem 1, $\operatorname{ord} P \leqq t_1 \cdots t_m$. proving the theorem.

If we specify in Theorem 8 that $P = [S]$, where S is a system of linear partial homogeneous d.p. in $F\{y\}$, then we have the hypothesis of a theorem of Delsarte [1], Proposition A, page 37. Then by Theorems 7 and 8, the linear dimension of the solution space of S is finite, which is Delsarte's conclusion.

REFERENCES

1. Jean Delsarte, *Sur certains systèmes d'équations aux dérivées partielles à une seule fonction inconnue, et sur une généralization de la théorie des fonctions de Bessel et des fonctions hypergéométriques,* Premier Colloque Sur Les Équations aux Dérivées Partielles held at Louvain, December 17-19, 1962, pp. 35-62.

2. E.R. Kolchin, *Extensions of differential fields, II,* Ann. of Math. **45** (1945), 358-361.

3. ———, *Extensions of differential fields, III,* Bull. Amer. Math. Soc. **53** (1947), 399-401.

4. ———, *Picard-Vessiot theory of partial differential fields,* Proc. Amer. Math. Soc. **3** (1952), 596-603.

5. H.W. Raudenbush, *Hypertranscendental adjunctions to partial differential fields,* Bull. Amer. Math. Soc. **40** (1934), 714-720.

6. J.F. Ritt, *Differential algebra,* Amer. Math. Soc. Colloquium Publications, vol. 33, New York, 1950.

7. B.L. Van Der Waerden, *Modern algebra,* Volume 1, Revised English Edition, F. Ungar Publishing Co., New York, 1953, Section 33, pp. 99-104.

VASSAR COLLEGE

Pacific Journal of Mathematics

Vol. 15, No. 1 September, 1965