Pacific Journal of Mathematics

IN THIS ISSUE—

Vol. 14, No. 3 **July, 1964**

PACIFIC JOURNAL OF MATHEMATICS

Mathematical papers intended for publication in the *Pacific Journal of Mathematics* should by typewritten (double spaced), and on submission, must be accompanied by a separate author's résumé. Manuscripts may be sent to any one of the four editors. All other communications to the editors should be addressed to the managing editor, L. J. Paige at the University of California, Los Angeles 24, California.

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

The *Pacific Journal of Mathematics* is published quarterly, in March, June, September, and December. Effective with Volume 13 the price per volume (4 numbers) is $18.00; single issues, $5.00. Special price for current issues to individual faculty members of supporting institutions and to individual members of the American Mathematical Society: $8.00 per volume; single issues $2.50. Back numbers are available.

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

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

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

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

THE ESSENTIAL SPECTRUM OF A CLASS OF ORDINARY DIFFERENTIAL OPERATORS

E. BALSLEV AND T. W. GAMELIN

Introduction. The purpose of this paper is to give a method of determining the essential spectrum of a class of ordinary differential operators in L^p of an interval with ∞ as a singular endpoint. The method relies on the mapping theorem for the essential spectrum, proved for ordinary differential operators by Rota [9]. A discussion of this type of theorem is presented in § 1. The essential spectrum of the constant coefficient operator and the Euler operator is determined in § 4. It is found that the essential spectrum of the Euler operator is an algebraic curve which varies with the index $p, 1 < p < \infty$.

In §§ 5 and 6 the class of differential operators which are compact with respect to the constant coefficient operator, or Euler operator, is determined. By a fundamental theorem of perturbation theory, these operators may be added to the original operator without altering the essential spectrum.

The results apply to differential equations of Fuchsian type. This includes the Riemann differential equation, whose spectral theory was investigated by Rota [10].

1. Spectral mapping theorems. Let A be a closed, densely-defined operator in a Banach space $\mathfrak{X}$. A is a Fredholm operator if the null space $\mathscr{N}(A)$ of A is finite dimensional and the range $\mathscr{R}(A)$ of A is closed and of finite codimension in $\mathfrak{X}$. The *Fredholm index* of A is the number

$$\kappa(A) = \dim \mathscr{N}(A) - \operatorname{codim} \mathscr{R}(A) .$$

A complex number λ is in the *essential resolvent set* of A, denoted by $\rho_e(A)$, if $\lambda I - A$ is a Fredholm operator. Otherwise λ is in the *essential spectrum* of A, denoted by $\sigma_e(A)$. $\rho(A)$ and $\sigma(A)$ will denote the resolvent set and spectrum of A respectively.

Let $B(\mathfrak{X})$ denote the ring of bounded operators on $\mathfrak{X}$, and let $\mathscr{C}$ denote the ideal of compact operators in $\mathscr{B}(\mathfrak{X})$. $\mathscr{A} = \mathscr{B}(\mathfrak{X})/\mathscr{C}$ is a Banach algebra. The coset $A + \mathscr{C}$ of an element $A \in \mathscr{B}(\mathfrak{X})$ will be denoted by $\widetilde{A}$, and its spectrum will be denoted by $sp(\widetilde{A})$. The invertible elements of $\mathscr{A}$ are the cosets $\widetilde{B} = B + \mathscr{C}$, where $B \in \mathscr{B}(\mathfrak{X})$ is a Fredholm operator (cf [1]). In particular, $sp(\widetilde{A}) = \sigma_e(A)$ for all $A \in \mathscr{B}(\mathfrak{X})$.

Received October 23, 1963.

LEMMA 1. *Let $A \in \mathscr{B}(\mathfrak{X})$, and let f be analytic in a neighborhood of $\sigma(A)$. Then $\sigma_e(f(A)) = f(\sigma_e(A))$. If $\mu \in \rho_e(A)$, then*

$$\kappa(\mu I - f(A)) = \Sigma\{\kappa(\lambda I - A) \colon \lambda \in f^{-1}(\mu)\} ,$$

where λ is counted in the set $f^{-1}(\mu)$ according to its multiplicity as a solution of $f(z) - \mu = 0$.

Proof. The first assertion of the lemma is a trivial consequence of the spectral mapping theorem for Banach algebras:

$$\sigma_e(f(A)) = sp(\widetilde{f(A)}) = sp(f(\tilde{A})) = f(sp(\tilde{A})) = f(\sigma_e(A)) .$$

By replacing f by $\mu - f$ it suffices to establish the formula

$$\kappa(f(A)) = \Sigma\{\kappa(\lambda I - A) \colon \lambda \in f^{-1}(0)\} .$$

We can decompose the spectrum of A into a finite number of spectral (closed and open) subsets F_i, $i = 1, \cdots, n$, such that f is analytic in an open connected neighborhood of each F_i. Corresponding to each spectral set F_i, there is a projection E_i onto a closed invariant subspace $\mathfrak{X}_i$ of $\mathfrak{X}$ such that $I = \sum_{i=1}^n E_i$, $E_iE_j = 0$, $i \neq j$, and $\sigma(A \mid \mathfrak{X}_i) = F_i$ (cf [5], VII. 3).

Since the index κ satisfies the appropriate additivity conditions, it suffices to prove the formula for the restriction operators $A \mid \mathfrak{X}_i$, i.e. we may assume that f is analytic in a connected open neighborhood of $\sigma(A)$.

If f is identically zero, then $f(A) = 0$ is Fredholm, so $\mathfrak{X}$ is finite dimensional, and the result is trivial. If f is not identically zero, it has a finite number of zeros $z_1, \cdots, z_n \in \sigma(A)$, counted according to their multiplicity. Let

$$g(z) = f(z)/(z_1 - z) \cdots (z_n - z) .$$

g is analytic and nonzero in a neighborhood of $\sigma(A)$, so $g(A)$ is invertible and has index zero. Now

$$f(A) = (z_1 I - A) \cdots (z_n I - A)g(A) ,$$

where the $z_i I - A$ are Fredholm. Since the index of a product of Fredholm operators is the sum of their indices, we have

$$\begin{aligned} \kappa(f(A)) &= \sum_{i=1}^n \kappa(z_i I - A) \\ &= \Sigma\{\kappa(\lambda I - A) \colon \lambda \in f^{-1}(0) \cap \sigma(A)\} \\ &= \Sigma\{\kappa(\lambda I - A) \colon \lambda \in f^{-1}(0)\} . \end{aligned}$$

If A and B are unbounded operators with domains $\mathscr{D}(A)$ and

$\mathscr{D}(B)$, then their product is defined by

$$\mathscr{D}(AB) = \{x \in \mathscr{D}(B): Bx \in \mathscr{D}(A)\},\ (AB)x = A(Bx) .$$

A and B commute if $AB = BA$.

If A and B are closed, densely-defined Fredholm operators, then AB is closed and densely-defined, AB is Fredholm, and $\kappa(AB) = \kappa(A) + \kappa(B)$ (cf [6]). Conversely, if $\{A_i\}_{i=1}^n$ is a commuting set of closed operators such that $A = A_1 \cdots A_n$ is closed, densely-defined and Fredholm, then each of the A_i is densely-defined and Fredholm. For $\mathscr{N}(A) \supseteqq \mathscr{N}(A_i)$ and $\mathscr{R}(A) \subseteqq \mathscr{R}(A_i)$ for each i. As a special case of these remarks, we can state a version of Lemma 1 for unbounded operators. For ordinary differential operators, the spectral mapping theorem is due to Rota [9].

LEMMA 2. *Let A be a closed, densely-defined operator in $\mathfrak{X}$, and let p be a polynomial of degree n.*

(a) *If $p(\sigma_e(A))$ is not the entire complex plane, then $p(A)$ is densely defined and closed.*

(b) *If $p(A)$ is densely defined and closed, then $\sigma_e(p(A)) = p(\sigma_e(A))$. If $\mu \in \rho_e(p(A))$,*

$$\kappa(\mu I - p(A)) = \sum_{i=1}^{n} \kappa(\lambda_i I - A) ,$$

where $\lambda_1, \cdots, \lambda_n$ are the solutions of $p(z) - \mu = 0$, counted according to their multiplicity.

Proof. $\mu I - p(A) = (\lambda_1 I - A) \cdots (\lambda_n I - A)$, where the $\lambda_i I - A$ commute. If $\mu \notin p(\sigma_e(A))$, then each λ_i is in $\rho_e(A)$, so $\mu I - p(A)$ is densely-defined and closed. Hence $p(A)$ is densely-defined and closed.

Part (b) of the lemma is a consequence of the preceeding discussion.

2. Some basic facts about linear operators. Let A be a closed densely-defined linear operator in a Banach space $\mathfrak{X}$. The domain $\mathscr{D}(A)$ of A becomes a Banach space when endowed with the A-topology, or graph topology, defined by the norm $\|x\|_A = \|x\| + \|Ax\|$. A linear operator $B: \mathscr{D}(B) \to \mathfrak{X}$ is said to be *A-defined* if $\mathscr{D}(B) \supseteqq \mathscr{D}(A)$. B is *A-bounded* if the restriction of B to $\mathscr{D}(A)$ is a bounded operator from $\mathscr{D}(A)$, with the graph topology, to $\mathfrak{X}$. Its A-norm $\|B\|_A$ is given by

$$\|B\|_A = \sup_{x \in \mathscr{D}(A)} \{\|Bx\|/\|x\|_A\}$$

B is A-compact if it is compact as an operator from $\mathscr{D}(A)$, with the graph topology, to $\mathfrak{X}$.

If A' is a second operator which is closed on $\mathscr{D}(A) = \mathscr{D}(A')$, then the A'-topology for $\mathscr{D}(A)$ coincides with the A-topology for $\mathscr{D}(A)$. The following lemma gives criteria for $A' = A + B$ to be closed on $\mathscr{D}(A)$, and collects certain facts which will be used later.

LEMMA 3. *Let A be a closed densely-defined operator in $\mathfrak{X}$, and let B be an A-defined (not cecessarily closed) linear operator in $\mathfrak{X}$.*

(a) *If there exist $0 \leqq \alpha < 1$ and $0 \leqq \beta$ such that*

$$\| Bx \| \leqq \alpha \| Ax \| + \beta \| x \| \quad \text{for } x \in \mathscr{D}(A) ,$$

then $A + B$ is closed on $\mathscr{D}(A)$.

(b) *If B is A-compact, then $A + B$ is closed on $\mathscr{D}(A)$, and $\sigma_e(A + B) = \sigma_e(A)$,*

$$\kappa(A + B - \lambda I) = \kappa(A - \lambda I) \quad \text{for } \lambda \in \rho_e(A) .$$

(c) *If $\lambda \in \rho_e(A)$, then there is an $\varepsilon(\lambda) > 0$ such that $\| B \|_A < \varepsilon(\lambda)$ implies $\lambda \in \rho_e(A + B)$.*

(d) *If B is closed and A-compact, then for every $\varepsilon > 0$, there is a $K(\varepsilon) > 0$ such that*

$$\| Bx \| \leqq \varepsilon \| Ax \| + K(\varepsilon) \| x \| , \qquad x \in \mathscr{D}(A) .$$

Proof. (a), (b) and (c) are well-known. Suppose that (d) is not true. Then there is an $\varepsilon > 0$ and a sequence $\{x_n\}$ in $\mathscr{D}(A)$ such that

$$\| Bx_n \| \geqq \varepsilon \| Ax_n \| + n \| x_n \| .$$

Since the inequality is homogeneous, we may assume $\| x_n \|_A = 1$. Passing to a subsequence, if necessary, we may assume, that Bx_n converges to y. Since

$$\| Bx_n \| \geqq \varepsilon \| x_n \|_A + (n - \varepsilon) \| x_n \| = \varepsilon + (n - \varepsilon) \| x_n \| ,$$

x_n converges to 0. Since B is closed, $y = 0$. On the other hand, $\| y \| = \lim \| Bx_n \| \geqq \varepsilon$, a contradiction.

The argument establishing part (d) can be found in [4], p. 39. There are operators B which are A-compact but for which no inequality of the form $\| Bx \| \leqq \varepsilon \| Ax \| + K(\varepsilon) \| x \|$ obtains.

3. Differential operators. Let (α, β) be an interval, where $\alpha = -\infty$ and $\beta = +\infty$ are allowed as endpoints. A *formal differential expression* l on the interval (α, β) is an expression of the form

$$(lf)(t) = \sum_{j=0}^{n} a_j(t) f^{(j)}(t) ,$$

where the a_j are complex-valued measurable functions on (α, β).

The *maximal operator* L in $L^p(\alpha, \beta)$, $1 < p < \infty$, associated with l, is defined by

$\mathscr{D}(L) = \{f \in L^p(\alpha, \beta): f^{(j)}$ exist and are loc. $a.c.$, $0 \leqq j \leqq n-1$, $l(f) \in L^p(\alpha, \beta)\}$ and

$$Lf = l(f)\,, \qquad f \in \mathscr{D}(L)\,.$$

The operator L_0' is the restriction of L to C^∞ functions with compact support contained in (α, β).

If L_0' is closable, then the *minimal operator* L_0 associated with l is the closure of L_0'. A *differential operator* associated with l is an operator L_u such that

$$\mathscr{D}(L_0) \subseteqq \mathscr{D}(L_u) \subseteqq \mathscr{D}(L)$$

and

$$L_u f = l(f)\,, \qquad f \in \mathscr{D}(L_u)\,.$$

Under mild restrictions on the coefficients $a_j(t)$, for instance, that $a_j(t)$ be locally integrable, $0 \leqq j \leqq n-1$, and that $1/a_n(t)$ be locally integrable, the maximal operator L is densely defined and closed. In this case, $\mathscr{D}(L_0)$ is of finite codimension in $\mathscr{D}(L)$.

Any finite dimensional extension of a Fredholm operator is again Fredholm (cf [6]). Hence, under the preceeding restrictions on the coefficients a_j, $\rho_e(L_u) = \rho_e(L)$ for all differential operators L_u determined by l. This set is called the *essential* resolvent set of l, and denoted by $\rho_e(l)$. Its complement $\sigma_e(l)$ is the *essential spectrum* of l.

If $\mathscr{D}(L_u)$ is of codimension k in $\mathscr{D}(L)$, and $\mu \in \rho_e(l)$, then $\kappa(\mu I - L_u) = \kappa(\mu I - L) - k$ (cf [6]). To determine the Fredholm index of $\mu I - L_u$, it suffices then to find the index of $\mu I - L$, or of $\mu I - L_0$.

In the following, D_0 and D will denote respectively the minimal and maximal operators in $L^p(\alpha, \beta)$ determined by the differential expression $(lf)(t) = f'(t)$, where (α, β) is the interval under consideration.

4. The basic formulae for the essential spectrum.

THEOREM 1. *Let M be the maximal differential operator in $L^p[0, \infty)$ associated with the expression*

$$(mf)(t) = \sum_{j=0}^{n} a_j f^{(j)}(t)\,,$$

a_j constants.

Let π be the polynomial

$$\pi(z) = \sum_{j=0}^{n} a_j z^j \ .$$

Then

$$\sigma_e(m) = \{\pi(ir) \colon -\infty < r < \infty\} \ .$$

If $\lambda \in \rho_e(m)$, the Fredholm index $\kappa(\lambda I - M)$ is the number of roots of $\pi(z) = \lambda$, counted according to their multiplicity, which lie in the half-plane $\mathscr{Re}(z) < 0$.

Proof. The equation $\lambda g - D_0 g = f$ is satisfied by

$$g(s) = (\lambda I - D_0)^{-1} f(s) = -e^{\lambda s}\int_0^s e^{-\lambda t} f(t) dt \ .$$

If $\mathscr{Re}(\lambda) < 0$, then $(\lambda I - D_0)^{-1} f = k * f$, where $k \in L^1(-\infty, \infty)$. So $(\lambda I - D_0)^{-1}$ is bounded, and $\lambda \in \rho(D_0)$. In particular, $\kappa(\lambda I - D_0) = 0$ for $\mathscr{Re}(\lambda) < 0$.

If $\mathscr{Re}(\lambda) > 0$, the adjoint differential equation of $\lambda f = D_0 f$ has the solution $e^{-\lambda t} \in L^q[0, \infty)$, which must be orthogonal to the range of $\lambda I - D_0$. If $f \in \mathscr{R}(\lambda I - D_0)$,

$$(\lambda I - D_0)^{-1} f(s) = e^{\lambda s}\int_s^\infty e^{-\lambda t} f(t) dt \ .$$

This is again a convolution operator with an L^1-kernel, and so $(\lambda I - D_0)^{-1}$ is bounded on $\mathscr{R}(\lambda I - D_0)$. It follows that $\mathscr{R}(\lambda I - D_0)$ is the subspace of $L^p[0, \infty)$ orthogonal to $e^{-\lambda t}$, and so is closed and of codimension 1 in $L^p[0, \infty)$. Hence $\lambda \in \rho_e(D_0)$ and $\kappa(\lambda I - D_0) = -1$ for $\mathscr{Re}(\lambda) > 0$.

Since the Fredholm index is constant on each component of $\rho_e(D_0)$, the line $\mathscr{Re}(\lambda) = 0$ must be the essential spectrum of D_0. Since D is an extension of D_0 by one dimension, $\kappa(\lambda I - D) = 1$ if $\mathscr{Re}(\lambda) < 0$ and $\kappa(\lambda I - D) = 0$ if $\mathscr{Re}(\lambda) > 0$.

This establishes the theorem for the special case of the operator D. It suffices now to prove that $M = \pi(D)$;[1] then the general result follows from Lemma 2. From the inequality of Lemma 5 we derive the inequality

$$\| D^n f \| \leqq K\{\| Mf \| + \| f \|\} \ , \qquad f \in C_0^\infty(0, \infty) \ .$$

Thus, the M-norm and D^n-norm on $C_0^\infty(0, \infty)$ are equivalent, and it follows, that

$$\mathscr{D}(M_0) = \mathscr{D}(D_0^n) = \mathscr{D}(\pi(D_0)) \ .$$

Since M is an extension of $\pi(D)$, and since $\dim \mathscr{D}(\pi(D))/\mathscr{D}(\pi(D_0)) \geqq n$, it suffices to show, that $\dim \mathscr{D}(M)/\mathscr{D}(M_0) = n$.

[1] Professor S. Goldberg pointed out, that a proof was missing here.

Since $\mathscr{D}(M) = \mathscr{D}(M - \lambda I)$, we may assume, by altering the constant term of π, that M is Fredholm. Then

$$\begin{aligned}\dim \mathscr{D}(M)/\mathscr{D}(M_0) &= \dim \mathscr{N}(M) + \operatorname{codim} \mathscr{R}(M_0)\\ &= \dim \mathscr{N}(M) + \dim \mathscr{N}(L)\end{aligned}$$

where L is the maximal differential operator associated with the adjoint expression (cf. [9]).

We may also assume, that the roots $\lambda_1, \cdots, \lambda_n$ of $\pi(z) = 0$ have distinct real parts. Then $\mathscr{N}(M)$ is spanned by the exponentials $e^{\lambda_i t}$, and $\mathscr{N}(L)$ is spanned by the exponentials $e^{-\lambda_i t}$. From this it is easy to conclude that $\dim \mathscr{N}(M) + \dim \mathscr{N}(L) = n$.

THEOREM 2. *Let the Euler differential expression l on the interval $[1, \infty)$ be defined by*

$$(lf)(t) = \sum_{k=0}^{n} b_k t^k f^{(k)}(t) ,$$

where the b_k are constants. Let L be the associated maximal operator in $L^p[1, \infty)$, $1 < p < \infty$. Let d be the polynomial

$$d(z) = b_0 + \sum_{k=1}^{n} b_k \prod_{j=0}^{k-1} \left(z - \left(\frac{1}{p} + j\right)\right) .$$

Then $\sigma_e(l) = \{d(ir)\colon -\infty < r < \infty\}$. For $\lambda \in \rho_e(l)$, the Fredholm index $\kappa(\lambda I - L)$ is the number of roots of $d(z) - \lambda = 0$, counted according to their multiplicity, which lie in the half-plane $\mathscr{Re}(z) < 0$.

Proof. For $f \in L^p[1, \infty)$, we define

$$(\pi f)(s) = e^{s/p} f(e^s),\ 0 \leqq s < \infty .$$

It is easily verified, that τ is an isometric isomorphism of $L^p[1, \infty)$ and $L^p[0, \infty)$. Its inverse is given by

$$f(t) = (\tau^{-1}g)(t) = t^{-1/p} g(\log t),\ 1 \leqq t < \infty .$$

We have

$$\begin{aligned}\frac{df}{dt} &= t^{-(1/p)-1} g'(\log t) - \frac{1}{p} t^{-(1/p)-1} g(\log t)\\ &= t^{-(1/p)-1}\left[\left(\frac{d}{ds} - \frac{1}{p}\right) g(s)\right]_{s=\log t} .\end{aligned}$$

By induction on k, the following formula obtains

$$\begin{aligned}\frac{d^k f}{dt^k} = t^{-(1/p)-1}\Big[&\left(\frac{d}{ds} - \left(\frac{1}{p} + k - 1\right)\right)\left(\frac{d}{ds} - \left(\frac{1}{p} + k - 2\right)\right) \cdots\\ &\times \left(\frac{d}{ds} - \frac{1}{p}\right) g(s)\Big]_{s=\log t} .\end{aligned}$$

Therefore

$$\left(\tau t^k \frac{d^k f}{dt^k}\right)(s) = \prod_{j=0}^{k-1}\left(\frac{d}{ds} - \left(\frac{1}{p} + j\right)\right)\tau f(s) .$$

Let l_k be the differential expression

$$(l_k f)(t) = t^k f^{(k)}(t), 1 \leqq t < \infty ,$$

and let L_k be the corresponding maximal operator in $L^p[1, \infty)$. Then

$$L_k = \tau^{-1} \prod_{j=0}^{k-1}\left(D - \left(\frac{1}{p} + j\right)\right)\tau, k \geqq 1 .$$

Consequently,

$$L = \tau^{-1}\left[b_0 I + \sum_{k=1}^{n} b_k \prod_{j=0}^{k-1}\left(D - \left(\frac{1}{p} + j\right)\right)\right]\tau .$$

Since the essential spectrum and Fredholm index remain invariant under isometric isomorphisms, the result follows from Theorem 1.

REMARK. The essential spectrum of L could also be computed by writing L as a polynomial in the operator $x(d/dx)$, which has the essential spectrum $\{-(1/p) + ir, -\infty < r < \infty\}$. The Euler operator was originally represented as a polynomial by George Boole.

5. Perturbation of the constant coefficient operator. The inequalities, on which the results of this section are based, are essentially special cases of similar estimates for elliptic partial defferential operators (cf [4]). Similar results for perturbation of partial differential operators are obtained in [3]. For $p = 2$ theorems of this type for elliptic operators, including Lemma 7, are proved by Birman (cf. [11]).

LEMMA 4. *Given $\varepsilon > 0$, there exists a constant K, depending only on p and ε, such that*

$$\int_N^\infty | b(t)f(t) |^p \, dt \leqq \left\{\sup_{N \leqq s < \infty} \int_s^{s+1} | b(t) |^p \, dt\right\}\left\{\varepsilon\int_N^\infty | f'(t) |^p \, dt + K\int_N^\infty | f(t) |^p \, dt\right\}$$

for all $N \geqq 0$, all functions b locally in $L^p[0, \infty)$, and all functions f in the domain of the maximal operator D in $L^p[0, \infty)$.

Proof. Let r be a small positive number Let a be a continuously differentiable function on $[0, r]$ such that

$$0 \leqq a \leqq 1, a(0) = 1 \quad \text{and} \quad a(r) = 0 .$$

If $f \in \mathscr{D}(D)$, then

$$
\begin{aligned}
f(t) &= -\int_0^r \frac{d}{ds}(a(s)f(t+s))ds \\
&= -\int_0^r a(s)f'(t+s)ds - \int_0^r a'(s)f(t+s)ds\ ;
\end{aligned}
$$

$$
\begin{aligned}
|f(t)| &\leqq \int_0^r |f'(t+s)|\,ds + K_0\int_0^r |f(t+s)|\,ds \\
&\leqq r^{1/q}\Big\{\int_0^r |f'(t+s)|^p\,ds\Big\}^{1/p} + K_0 r^{1/q}\Big\{\int_0^r |f(t+s)|^p\,ds\Big\}^{1/p} \\
&\leqq c_p r^{1/q}\Big\{\int_0^r |f'(t+s)|^p\,ds + K_0\int_0^r |f(t+s)|^p\,ds\Big\}^{1/p}\,,
\end{aligned}
$$

where $(1/p) + (1/q) = 1$.

If r is chosen so that $\varepsilon^{1/p} = r^{1/q}c_p$, then

$$
|f(t)|^p \leqq \varepsilon\int_0^r |f'(t+s)|^p\,ds + K\int_0^r |f(t+s)|^p\,ds\ ;
$$

$$
\begin{aligned}
&\int_N^\infty |b(t)f(t)|^p\,dt \\
&\quad\leqq \int_N^\infty\int_t^{t+r} |b(t)|^p\{\varepsilon|f'(s)|^p + K|f(s)|^p\}dsdt \\
&\quad= \int_N^\infty\int_{\max(s-r,N)}^{s} |b(t)|^p\{\varepsilon|f'(s)|^p + K|f(s)|^p\}dtds \\
&\quad\leqq \Big\{\sup_{N\leqq s<\infty}\int_s^{s+r} |b(t)|^p\,dt\Big\}\Big\{\varepsilon\int_N^\infty |f'(s)|^p\,ds + K\int_N^\infty |f(s)|^p\,ds\Big\}\,.
\end{aligned}
$$

LEMMA 5. *Given $\varepsilon > 0$, there exists $K(\varepsilon) > 0$ such that*

$$
\|D^k f\| \leqq \varepsilon\|D^n f\| + K(\varepsilon)\|f\|,\ f\in\mathscr{D}(D^n),\ 0\leqq k<n\,,
$$

where the norms are taken in $L^p[0,\infty)$.

Proof. Let $[0, r]$ be a finite interval. Replacing f by f' and proceeding as in the proof of Lemma 4, we arrive at the inequality

$$
|f'(t)| \leqq C_p r^{1/q}\Big\{\int_0^r |f''(t+s)|^p\,ds + K_0(r)\int_0^r |f'(t+s)|^p\,ds\Big\}^{1/p}\,.
$$

Suppose $\{f_n\}$ is a D^2-bounded sequence in $L^p[0, r]$. It is easy to see that the derivatives f_n' are uniformly bounded and equicontinuous on the interval $[0, r]$. Hence the operator D in $L^p[0, r]$ is compact with respect to the operator D^2 in $L^p[0, r]$.

By Lemma 3(d), there exists a $K_1(r) > 0$ such that

$$
K_0(r)\int_0^r |f'(t+s)|^p\,ds \leqq \int_0^r |f''(t+s)|^p\,ds + K_1(r)\int_0^r |f(t+s)|^p\,ds\,.
$$

If r is chosen so that $0 < r < 1$ and $\varepsilon^{1/p} = 2C_p r^{1/q}$, then the above inequalities yield the pointwise estimate

$$|f'(t)|^p \leqq \varepsilon\int_0^r |f''(t+s)|^p\, ds + K(\varepsilon)\int_0^r |f(t+s)|^p\, ds .$$

Integrating from 0 to ∞ and exchanging the order of integration, we arrive at the following inequality

$$\| Df \|^p \leqq \varepsilon \| D^2 f \|^p + K(\varepsilon) \| f \|^p .$$

This is equivalent to an inequality of the form

$$\| Df \| \leqq \varepsilon \| D^2 f \| + K(\varepsilon) \| f \| .$$

Inequalities of the form

$$\| D^k f \| \leqq \varepsilon \| D^{k+1} f \| + K(\varepsilon) \| f \|$$

follow easily by induction on k. Since D^{k+1} is D^n-bounded, we finally obtain an inequality of the desired form

$$\| D^k f \| \leqq \varepsilon \| D^n f \| + K(\varepsilon) \| f \| .$$

Let b be a measurable function on the interval $[0, \infty)$, and define the linear operator B in $L^p[0, \infty)$ by

$$\mathscr{D}(B) = \{ f \in L^p[0, \infty) : bf \in L^p[0, \infty) \} ,$$
$$Bf = b \cdot f, f \in \mathscr{D}(B) .$$

B is closed and densely-defined.

In the following, $L^p_{\mathrm{loc}}[\alpha, \infty)$ will denote the space of measurable functions on $[\alpha, \infty)$ which are locally in $L^p[\alpha, \infty)$.

LEMMA 6. *B is D-defined if and only if $b \in L^p_{\mathrm{loc}}[0, \infty)$ and*

$$\limsup_{s \to \infty} \int_s^{s+1} | b(t) |^p\, dt < \infty .$$

If B is D-defined, then for every $\varepsilon > 0$, there exists a $K(\varepsilon) > 0$ such that

$$\| Bf \| \leqq \varepsilon \| Df \| + K(\varepsilon) \| f \|, f \in \mathscr{D}(D) .$$

In particular, $D + B$ is closed on $\mathscr{D}(D)$.

Proof. Suppose that B is D-defined. Since B is closed, B is D-bounded. Let f be a C^∞-function on $(-\infty, \infty)$ such that

$$0 \leqq f \leqq 1$$
$$f(s) = 1, 0 \leqq s \leqq 1$$
$$f(s) = 0, -\infty < s \leqq -1, 2 \leqq s < \infty .$$

Let $f_s(t) = f(t - s)$, and let g_s be the restriction of f_s to the interval $[0, \infty)$.

If $s \geqq 0$, then

$$\int_s^{s+1} | b(t)g_s(t) |^p \, dt = \int_s^{s+1} | b(t) |^p \, dt \leqq || bg_s ||^p \leqq || B ||_D^p \, || g_s ||^p$$
$$\leqq || B ||_D^p \, K_p\{|| f ||_{L^p(-\infty,\infty)}^p + || f' ||_{L^p(-\infty,\infty)}^p\} ,$$

Hence $b \in L^p_{\mathrm{loc}}[0, \infty)$, and

$$\limsup_{s\to\infty} \int_s^{s+1} | b(t) |^p \, dt < \infty .$$

Conversely, suppose $b \in L^p_{\mathrm{loc}}[0, \infty)$ and

$$\limsup_{s\to\infty} \int_s^{s+1} | b(t) |^p \, dt < \infty .$$

Then

$$\sup_{0\leqq s<\infty} \int_s^{s+1} | b(t) |^p \, dt < \infty .$$

It follows from Lemma 4, with $N = 0$, that B is D-bounded and

$$|| Bf || \leqq \varepsilon \, || Df || + K(\varepsilon) \, || f ||, f \in \mathscr{D}(D) .$$

By Lemma 3(a), $D + B$ is closed.

LEMMA 7. *B is D-compact if and only if $b \in L^p_{\mathrm{loc}}[0, \infty)$ and*

$$\lim_{s\to\infty} \int_s^{s+1} | b(t) |^p \, dt = 0 .$$

Proof. Suppose that B is D-compact. By Lemma 6, $b \in L^p_{\mathrm{loc}}[0, \infty)$. Suppose that there exists a sequence $s_n \to \infty$ and a $K > 0$ such that

$$\int_{s_n}^{s_n+1} | b(t) |^p \, dt \geqq K , \qquad n = 1, 2, \cdots .$$

Let $\{g_{s_n}\}$ be the functions defined in the proof of Lemma 6; since $\{g_{s_n}\}$ is a D-bounded sequence, and B is D-compact, we can assume that

$$|| Bg_{s_n} || \xrightarrow[n\to\infty]{} 0 ,$$

passing to a subsequence if necessary. On the other hand,

$$|| Bg_{s_n} ||^p \geqq \int_{s_n}^{s_n+1} | b(t) |^p \, dt \geqq K ,$$

a contradiction. Hence $\lim_{s\to\infty} \int_s^{s+1} | b(t) |^p \, dt = 0$.

Conversely, suppose that $b \in L^p_{loc}[0, \infty)$ and that

$$\lim_{s\to\infty} \int_s^{s+1} | b(t) |^p \, dt = 0 \, .$$

Let χ_N denote the characteristic function of $[0, N]$, and define

$$B_N f = \chi_N b f, \, f \in \mathscr{D}(D) \, .$$

By Lemma 4, there is a constant $K > 0$ such that

$$\begin{aligned} \| (B - B_N) f \|^p &= \int_N^\infty | b(t) f(t) |^p \, dt \\ &\leqq K \Big\{ \sup_{N \leqq s < \infty} \int_s^{s+1} | b(t) |^p \, dt \Big\} \{ \| Df \|^p + \| f \|^p \} \, . \end{aligned}$$

Hence $\| B - B_N \|_D \to 0$ as $N \to \infty$, so it suffices to show that each B_N is D-compact.

For this purpose, let $\{f_k\}$ be a D-bounded sequence in $\mathscr{D}(D)$. Since

$$| f_k(s) - f_k(t) | = \left| \int_s^t f_k'(r) dr \right| \leqq | t - s |^{1/q} \Big\{ \int_s^t | f'(r) |^p \, dr \Big\}^{1/p} ,$$

the f_k are equicontinuous on $[0, N]$. If $\{f_{k_j}\}_{j=1}^\infty$ is a subsequence which converges uniformly on $[0, N]$ then $\{B_N f_{k_j}\}_{j=1}^\infty$ converges in $L^p[0, \infty)$. Hence B_N is D-compact.

THEOREM 3. *Let M be the maximal operator in $L^p[0, \infty), 1 < p < \infty$, corresponding to the differential expression*

$$(mf)(t) = \sum_{j=0}^n a_j f^{(j)}(t), \, a_j \text{ constants}, \; a_n \neq 0 \, .$$

Let B be the maximal operator in $L^p[0, \infty)$ corresponding to the differential expression

$$\sum_{j=0}^{n-1} b_j(t) f^{(j)}(t) \, ,$$

where the b_j are measurable.

(a) *B is M-bounded if and only if $b_j \in L^p_{loc}[0, \infty)$ and*

$$\limsup_{s\to\infty} \int_s^{s+1} | b_j(t) |^p \, dt < \infty, \, 0 \leqq j \leqq n - 1 \, .$$

(b) *B is M-compact if and only if $b_j \in L^p_{loc}[0, \infty)$ and*

$$\lim_{s\to\infty} \int_s^{s+1} | b_j(t) |^p \, dt = 0, \, 0 \leqq j \leqq n - 1 \, .$$

(c) *If B is M-bounded, then for every $\varepsilon > 0$ there exists $K(\varepsilon) > 0$ such that*

$$\| Bf \| \leqq \varepsilon \| Mf \| + K(\varepsilon) \| f \|, f \in \mathscr{D}(M) .$$

In particular, $M + B$ is closed on $\mathscr{D}(M)$.

Proof. Suppose that B is M-bounded. If the functions g_s are constructed as in the proof of Lemma 6, we have

$$\begin{aligned} \sup_{s \geqq 0} \int_s^{s+1} | b_0(t) |^p \, dt &= \sup \int_s^{s+1} | Bg_s(t) |^p \, dt \\ &\leqq \sup_{s \geqq 0} \| Bg_s \|^p \leqq \sup_{s \geqq 0} \| B \|_M^p (\| Mg_s \| + \| g_s \|)^p < \infty . \end{aligned}$$

Hence $b_0 \in L^p_{\mathrm{loc}}[0, \infty)$ and

$$\limsup_{s \to \infty} \int_s^{s+1} | b_0(t) |^p \, dt < \infty.$$

Let $1 \leqq k \leqq n - 1$ and assume that $b_j \in L^p_{\mathrm{loc}}[0, \infty)$ and

$$\limsup_{s \to \infty} \int_s^{s+1} | b_j(t) |^p \, dt < \infty, 0 \leqq j \leqq k - 1 .$$

The functions g_s can be altered so that

$$g_s^{(k)}(t) = 1, s \leqq t \leqq s + 1 .$$

The same type of estimate as used in the preceding paragraph yields the results

$$b_k \in L^p_{\mathrm{loc}}[0, \infty)$$

and

$$\limsup_{s \to \infty} \int_s^{s+1} | b_k(t) |^p \, dt < \infty .$$

By induction, this holds for all $k, 0 \leqq k \leqq n - 1$.

Conversely, assume $b_j(t) \in L^p_{\mathrm{loc}}[0, \infty)$ and

$$\limsup_{s \to \infty} \int_s^{s+1} | b_j(t) |^p \, dt < \infty, 0 \leqq j \leqq n - 1 .$$

Let B_j be the maximal operator corresponding to the expression $b_j(t) f^{(j)}(t)$.

By Lemma 6,

$$\| B_j f \| \leqq \varepsilon_j \| D^{j+1} f \| + K_0(\varepsilon_j) \| D^j f \|, f \in \mathscr{D}(D^{j+1}) .$$

From Lemma 5 we can deduce an inequality of the form

$$\| B_j f \| \leqq \varepsilon_j \| D^n f \| + K(\varepsilon_j) \| f \|, f \in \mathscr{D}(D^n) .$$

Summing over j we arrive at an inequality of the form

$$\| Bf \| \leqq \varepsilon \| D^n f \| + K(\varepsilon) \| f \|, f \in \mathscr{D}(D^n) .$$

Since M is a polynomial in D of order n, $\mathscr{D}(M) = \mathscr{D}(D^n)$, and the M-topology is equivalent with the D^n-topology for $\mathscr{D}(M)$. Hence we get an inequality of the desired form,

$$\| Bf \| \leqq \varepsilon \| Mf \| + K(\varepsilon) \| f \| .$$

By Lemma 3(a), $M + B$ is closed on $\mathscr{D}(M)$. This completes the proof of parts (a) and (c) of the theorem.

If

$$\lim_{s \to \infty} \int_s^{s+1} | b_j(t) |^p \, dt = 0, \ 0 \leqq j \leqq n - 1,$$

then each B_j is D^{j+1}-compact, by Lemma 7. And so B_j is D^n-compact, therefore M-compact. Hence B is M-compact.

Conversely, if B is M-compact, then the relations

$$\lim_{s \to \infty} \int_s^{s+1} | b_k(t) |^p \, dt = 0$$

can be proved by induction on k as in the proof of part (a) and of Lemma 7.

THEOREM 4. *Let M and L be the maximal operators in $L^p[0, \infty)$, $1 < p < \infty$, corresponding to the differential expressions*

$$(mf)(t) = \sum_{j=0}^{n} a_j f^{(j)}(t) \ , \quad a_j \text{ constants}, \ a_n \neq 0,$$

$$(lf)(t) = (mf)(t) + \sum_{j=0}^{n} b_j(t) f^{(j)}(t) \ .$$

Suppose b_n is continuous and satisfies

$$b_n(t) \neq -a_n, \ 0 \leqq t < \infty$$
$$\lim_{t \to \infty} b_n(t) = 0 \ .$$

Suppose $b_j \in L^p_{\mathrm{loc}}[0, \infty)$ and satisfies

$$\lim_{s \to \infty} \int_s^{s+1} | b_j(t) |^p \, dt = 0 \ , \qquad 0 \leqq j \leqq n - 1 \ .$$

Then $\mathscr{D}(L) = \mathscr{D}(M)$, and $\sigma_e(l) = \sigma_e(m)$. If $\lambda \in \rho_e(m)$, $\kappa(\lambda I - M) = \kappa(\lambda I - L)$.

Proof. Let B_n be the maximal operator corresponding to the expression $b_n(t)f^{(n)}(t)$.[2] In view of Theorem 3 and Lemma 3 it suffices to prove the theorem in the case

$$(lf)(t) = (mf)(t) + b_n(t)f^{(n)}(t) .$$

So we assume $b_j(t) = 0$, $0 \leqq j \leqq n-1$. Since the essential spectrum and the Fredholm index are localizable to the endpoint ∞, and since the graph topologies of $\mathscr{D}(L)$ and $\mathscr{D}(M)$ are equivalent on compact subsets of $[0, \infty)$, we may assume, by passing to an interval of the form $[N, \infty)$, that $|b_n(t)| \leqq \varepsilon$, $0 \leqq t < \infty$.

We have

$$\begin{aligned} \| B_n f \| &= \left\{ \int_0^\infty | b_n(t) f^{(n)}(t) |^p \, dt \right\}^{1/p} \\ &\leqq \varepsilon \| D^n f \| \leqq \varepsilon \| D^n \|_M (\| Mf \| + \| f \|) . \end{aligned}$$

If ε is sufficiently small, Lemma 3(a) applies, and $\mathscr{D}(L) = \mathscr{D}(M)$. Also, by Lemma 3(c) and suitable choice of ε, we must have $\sigma_e(l) = \sigma_e(m)$.

Now suppose $|b_n(t)| < |a_n|$, $0 \leqq t < \infty$, so that the hypotheses of the theorem are satisfied for

$$(l_\beta f)(t) = (mf)(t) + \beta b_n(t) f^{(n)}(t) ,$$

where $0 \leqq \beta \leqq 1$. We have shown that $\sigma_e(l_\beta) = \sigma_e(m)$, so that the function $\beta \to \kappa(\lambda I - L_\beta)$ is well-defined, $\lambda \in \rho_e(m)$. This function is continuous and integervalued, hence a constant. In particular, $\kappa(\lambda I - M) = \kappa(\lambda I - L)$.

6. Perturbation of the Euler operator.

THEOREM 5. *Let L be the maximal operator in $L^p[1, \infty)$, $1<p<\infty$, corresponding to the Euler differential expression*

$$(lf)(t) = \sum_{j=0}^{n} b_j t^j f^{(j)}(t) , \quad b_j \text{ constants}, \ b_n \neq 0.$$

Let C be the maximal operator in $L^p[1, \infty)$ corresponding to the expression

$$\sum_{j=0}^{n-1} c_j(t) t^j f^{(j)}(t) ,$$

where the c_j are measurable.

(a) *C is L-bounded if and only if $c_j \in L^p_{\mathrm{loc}}[1, \infty)$ and*

$$\limsup_{s\to\infty} \int_s^{\alpha s} \frac{1}{t} | c_j(t) |^p \, dt < \infty \ \textit{for some } \alpha > 1, \ 0 \leqq j \leqq n-1$$

[2] Professor S. Goldberg pointed out, that the proof was incomplete. The remaining part can be found at the end of the paper.

(b) *C is L-compact if and only if $c_j \in L^p_{\mathrm{loc}}[1, \infty)$ and*

$$\lim_{s\to\infty} \int_s^{\alpha s} \frac{1}{t} | c_j(t) |^p \, dt = 0 \text{ for some } \alpha > 1, 0 \leqq j < n - 1 .$$

(c) *If C is L-bounded, then for every $\varepsilon > 0$ there exists $K(\varepsilon) > 0$ such that*

$$|| Cf || \leqq \varepsilon || Lf || + K(\varepsilon) || f ||, f \in \mathscr{D}(L) .$$

In particular, $L + C$ is closed on $\mathscr{D}(L)$.

Proof. Let M be the maximal operator in $L^p[0, \infty)$ corresponding to the differential expression

$$(mf)(t) = b_0 + \sum_{j=1}^{n} b_j \prod_{k=0}^{j-1} \left(\frac{d}{ds} - \left(\frac{1}{p} + k\right)\right) ,$$

and let B be the maximal operator in $L^p[0, \infty)$ corresponding to the expression

$$c_0(e^s) + \sum_{j=1}^{n-1} c_j(e^s) \prod_{h=0}^{j-1} \left(\frac{d}{ds} - \left(\frac{1}{p} + k\right)\right) .$$

Let τ be the isometry of $L^p[1, \infty)$ and $L^p[0, \infty)$ introduced in the proof of Theorem 2. Then

$$L = \tau^{-1} M \tau$$

and

$$C = \tau^{-1} B \tau .$$

Also,

$$\int_s^{\alpha s} \frac{1}{t} | c_j(t) |^p \, dt = \int_{e^s}^{e^s + e^\alpha} | c_j(u) |^p \, du .$$

Combining Theorem 3 and a downward induction argument on the coefficients c_j, we arrive at parts (a) and (b) of Theorem 5. Part (c) also follows from Theorem 3.

THEOREM 6. *Let L and M be the maximal operators in $L^p[1, \infty)$, $1 < p < \infty$, corresponding to the differential expressions*

$$(lf)(t) = \sum_{j=0}^{n} b_j t^j f^{(j)}(t), \ b_j \text{ constants}, \ b_n \neq 0 .$$

$$(mf)(t) = (lf)(t) + \sum_{j=0}^{n} c_j(t) t^j f^{(j)}(t) .$$

Suppose c_n *is continuous and satisfies*

$$c_n(t) \neq -b_n,\ 1 \leqq t < \infty$$
$$\lim_{t\to\infty} c_n(t) = 0 \ .$$

Suppose $c_j \in L^p_{\text{loc}}[0, \infty)$ *and satisfies*

$$\lim_{s\to\infty} \int_s^{\alpha s} \frac{1}{t} |\, c_j(t)\,|^p \, dt = 0 \ \textit{for some}\ \alpha > 1,\ 0 \leqq j \leqq n-1 \ .$$

Then $\mathscr{D}(L) = \mathscr{D}(M)$, *and* $\sigma_e(l) = \sigma_e(m)$. *If* $\lambda \in \rho_e(m)$, $\kappa(\lambda I - M) = \kappa(\lambda I - L)$.

Proof. A straightforward verification, as in the proof of Theorem 5, shows that the transform of Theorem 6 under τ is Theorem 4.

7. Some special cases. The perturbation criterion of Theorem 5 includes all functions $c(t)$ such that $t^{-1/p}c(t) \in L^p[1, \infty)$. It includes all bounded measurable functions with limit zero at ∞. The criterion shows, for instance, that if $\alpha < j < n$, then $t^\alpha f^{(j)}$ is compact with respect to the Euler operator of degree n. If $\alpha < n$, Theorem 6 shows, that $t^\alpha f^{(n)}$ has no effect on the essential spectrum of l. In particular, if

$$(mf)(t) = \sum_{j=0}^{n} a_j(t) f^{(j)}(t)$$

is a Fuchsian differential expression, where $a_n(t) = 0(t^n)$, then m can be written in the form of Theorem 6, and the essential spectrum of m can be determined from the coefficients as in Theorem 2.

For instance, consider the Riemann differential expression

$$(mf)(t) = t(t+1)f''(t) + (at+b)f'(t) + \frac{ct^2 + dt + e}{t(t+1)} f(t) \ .$$

Except for the change of variable $t \to -t$ this is the equation investigated by Rota [10]. By Theorem 6, $\sigma_e(m) = \sigma_e(l)$, where

$$(lf)(t) = t^2 f''(t) + atf'(t) + cf(t) \ .$$

By Theorem 2,

$$\sigma_e(l) = \{d(ir) \colon -\infty < r < \infty\} \ ,$$

where

$$d(z) = \left(z - \frac{1}{p}\right)\left(z - \frac{1}{p} - 1\right) + a\left(z - \frac{1}{p}\right) + c \ .$$

Hence

$$\sigma_e(m) = \left\{ -r^2 + ir\left(a - 1 - \frac{2}{p}\right) + \frac{1}{p^2} + (1-a)\frac{1}{p} + c \colon -\infty < r < \infty \right\}.$$

This is equivalent to the expression obtained by Rota.

8. Remarks.

(a) The Euler operator in $L^p(0, 1]$.

The mapping τ defined as in the proof of Theorem 2 by

$$\tau f(s) = e^{s/p} f(e^s)$$

also establishes an isometric isomorphism of $L^p(0, 1]$ and $L^p(-\infty, 0]$.

The Euler operator

$$L = \sum_{j=0}^{n} a_j t^j D^j$$

in $L^p(0, 1]$ is isometric isomorphic via τ to the constant coefficient operator

$$M = a_0 + \sum_{j=1}^{n} a_j \prod_{k=0}^{j-1} \left(D - \left(\frac{1}{p} + k \right) \right)$$

in $L^p(-\infty, 0]$.

The operator D in $L^p(-\infty, 0]$ is isometric isomorphic to the operator $(-D)$ in $L^p[0, \infty)$; therefore D in $L^p(-\infty, 0]$ has the essential spectrum $\{it \colon -\infty < t < \infty\}$, and the Fredholm index of $\lambda I - D$ is 0 for $\mathscr{R}e\lambda < 0$ and 1 for $\mathscr{R}e\lambda > 0$.

It follows, that l on the interval $(0, 1]$ has the same essential spectrum as l on the interval $[1, \infty)$ and the Fredholm index of $\lambda I - L$ is the number of roots of the polynomial $d(z) - \lambda$ of Theorem 2, counted with multiplicity, which lie in the half-plane $\mathscr{R}e z > 0$.

The perturbation results also carry over to the interval $(0, 1]$. The Theorems of § 6 are true for the operator L in $L^p(0, 1]$, when 1 is substituted for 0 and 0 for ∞, in particular we now take the limes and lim sup of $\int_s^{\alpha s} (1/u) |c(u)|^p du$ as $s \to 0$.

The Euler operator L in $L^p(0, \infty)$ is isometric isomorphic via τ to the constant coefficient operator considered above in $L^p(-\infty, \infty)$, and the essential spectrum is given by the same formula. The Fredholm index is 0, $\sigma_e(L) = \sigma(L)$ and $L_0 = L$.

(b) The condition

$$\limsup_{s \to \infty} \int_s^{\alpha s} \frac{1}{t} |b(t)|^p dt < \infty \text{ for some } \alpha > 1$$

is equivalent to the condition

$$\limsup_{s\to\infty} s\int_s^\infty \frac{1}{t^2}\,|\,b(t)\,|^p\,dt < \infty$$

Also, the condition

$$\lim_{s\to\infty}\int_s^{\alpha s}\frac{1}{t}\,|\,b(t)\,|^p\,dt = 0 \text{ for some } \alpha > 1$$

is equivalent to the condition

$$\lim_{s\to\infty} s\int_s^\infty \frac{1}{t^2}\,|\,b(t)\,|^p\,dt = 0\ .$$

This second set of conditions could just as well have been used in Theorem 5 and 6.

The proof of these assertions follows from the inequalities

$$\begin{aligned}\sup_{s\geqq N}\int_s^{\alpha s}\frac{1}{t}\,|\,b(t)\,|^p\,dt &\leqq \sup_{s\geqq N}\alpha s\int_s^{\alpha s}\frac{1}{t^2}\,|\,b(t)\,|^p\,dt\\ &\leqq \alpha\sup_{s\geqq N} s\int_s^\infty\frac{1}{t^2}\,|\,b(t)\,|^p\,dt\ ,\end{aligned}$$

and

$$\begin{aligned}\sup_{s\geqq N} s\int_s^\infty\frac{1}{t^2}\,|\,b(t)\,|^p\,dt &= \sup_{s\geqq N} s\sum_{n=0}^\infty\int_{\alpha^n s}^{\alpha^{n+1}s}\frac{1}{t^2}\,|\,b(t)\,|^p\,dt\\ &\leqq \sup_{s\geqq N} s\sum_{n=0}^\infty\frac{1}{\alpha^n s}\int_{\alpha^n s}^{\alpha^{n+1}s}\frac{1}{t}\,|\,b(t)\,|^p\,dt\\ &\leqq \frac{\alpha}{\alpha-1}\sup_{s\geqq N}\int_s^{\alpha s}\frac{1}{t}\,|\,b(t)\,|^p\,dt\ .\end{aligned}$$

(c) A basis of solutions $f_1(\lambda, t), \cdots, f_n(\lambda, t)$ of a differential equation $l(f) = \lambda f$ of order n is said to be a *norm-analytic basis* at λ_0 if there is a neighborhood N of λ_0 such that (i) the functions f_i are analytic in λ for $\lambda \in N$ and (ii) there is an integer k such that for each $\lambda \in N$, $\{f_i\}_{i=1}^k$ span the set of solutions of $l(f) = \lambda f$ which lie in L^p. In [10], Rota proved the following criterion:

LEMMA. *If at λ either the differential operator l in L^p or its adjoint l^* in L^q,* $(1/p) + (1/q) = 1$ (cf [9], for definition of adjoint), *does not have a norm-analytic basis of solutions, then λ belongs to the essential spectrum of l.*

If l is the Euler differential expression of Theorem 2, the equation $l(y) = \lambda y$ has solutions $\varphi_j(t) = t^{\alpha_j}$, where α_j is a root of the algebraic equation

$$b_n z(z-1)\cdots(z-n+1)+\cdots+b_1 z+b_0=\lambda\ .$$

Now $\varphi_j \in L^p[1, \infty)$ if and only if $\mathscr{R}_e(\alpha_j) < -(1/p)$.

Hence l will not have a norm-analytic basis at any point of the curve

$$\lambda = b_n\Big(ir-\frac{1}{p}\Big)\Big(ir-\frac{1}{p}-1\Big)\cdots\Big(ir-\frac{1}{p}-n+1\Big) + \cdots + b_1\Big(ir-\frac{1}{p}\Big)+b_0\ , \qquad -\infty < r < \infty\ .$$

This curve is identical to $\{d(ir): -\infty < r < \infty\}$, where $d(z)$ is the polynomial defined in Theorem 2.

If λ is not on this curve, then it can be shown that the resolvent operator $(\lambda I - l)^{-1}$ is a sum of integral operators whose kernels are of the Hardy-Littlewood-Polya type (cf [7], or [5] pp. 531–532). This yields another proof of Theorem 2, but the details are more complicated.

This method also shows that the essential spectrum of the Euler operator is precisely the set of points at which l or l^* does not have a norm-analytic basis of solutions. That this is not true in general is shown by the following example.

Define

$$(lf)(t) = f'(t) + (\sin t + t\cos t)f(t),\ 0 \leqq t < \infty\ .$$

The equation $lf = \lambda f$ has the solution

$$\varphi_\lambda(t) = \exp\left[t(\lambda - \sin t)\right]\ ,$$

while the adjoint equation $l^*g = \lambda g$ has the solution

$$\psi_\lambda(t) = 1/\varphi_\lambda(t)\ .$$

Now $\varphi_\lambda \in L^p[0, \infty)$ if $\mathscr{R}_e(\lambda) < -1$ and $\varphi_\lambda \notin L^p[0, \infty)$ if $\mathscr{R}_e(\lambda) > -1$, so l does not have a norm-analytic basis on the line $\mathscr{R}_e(\lambda) = -1$.

Similarly, l^* does not have a norm-analytic basis on the line $\mathscr{R}_e(\lambda) = 1$. l and l^* have norm-analytic bases if $\mathscr{R}_e(\lambda) \neq \pm 1$.

Since 0 is a regular endpoint for the differential expression l, a necessary condition that a point λ be in $\rho_e(l)$ is that either $\varphi_\lambda \in L^p[0, \infty)$ or $\psi_\lambda \in L^q[0, \infty)$, $(1/p) + (1/q) = 1$ (cf [9]). Hence the entire strip $\{-1 \leqq \mathscr{R}_e(\lambda) \leqq 1\}$ is contained in the essential spectrum of l. It is easy to see that $\sigma_e(l)$ actually coincides with this vertical strip.

It seems possible that the boundary of the essential spectrum of an arbitrary differential expression consists of points λ at which either $lf = \lambda f$ or $l^*g = \lambda g$ does not have a norm-analytic basis of solutions.

(d) The fact that the isomorphism $(\tau f)(s) = e^{s/p}f(e^s)$ converts a resolvent operator of Hardy-Littlewood-Polya type into a resolvent operator of convolution type is a special case of the following situation.

Let K be a measurable function on $[0, \infty)$, and let

$$(Tf)(x) = \frac{1}{x}\int_0^x K\left(\frac{y}{x}\right)f(y)dy, f \in L^p[0, \infty) .$$

The mapping τ may be regarded as an isometric isomorphism of $L^p[0, \infty)$ and $L^p(-\infty, \infty)$.

The operator $S = \tau T \tau^{-1}$ in $L^p(-\infty, \infty)$ is given by

$$(Sg)(\omega) = \int_{-\infty}^{\infty} K(e^{z-\omega})e^{((1/p)-1)(\omega-z)}g(z)dz .$$

S is a convolution operator with kernel

$$J(r) = K(e^{-r})e^{((1/p)-1)r} .$$

Conversely, a convolution operator in $L^p(-\infty, \infty)$ with kernel J determines a Hardy-Littlewood-Polya operator in $L^p[0, \infty)$ with kernel

$$K(s) = s^{(1/p)-1}J(-\log s) .$$

The norm of S is at most the L^1-norm of J. Hence if

$$\int_{-\infty}^{\infty} |J(r)|\, dr = \int_0^{\infty} |K(s)|\, s^{-1/p}\, ds < \infty ,$$

then T is bounded, and

$$|| T || \leqq \int_0^{\infty} |K(s)|\, s^{-1/p}\, ds .$$

This last statement is just the Hardy-Littlewood-Polya inequality (cf [7]).

Added in Proof. Professor S. Goldberg has pointed out that the proof that $D(L) = D(M)$ in Theorem 4 is incomplete, i.e., it must be shown that $f \in L^p$ and $lf \in L^p$ imply $mf \in L^p$. This follows easily with the aid of a more general form of theorem 3(c), namely, that inequalities of the form

$$|| Bf || \leqq \epsilon \, || mf || + K \, || f ||$$

obtain, where the norm is taken in $L^p[0, N)$ for $1 \leqq N \leqq \infty$, and K depends on ϵ and p but not on N. These inequalities result from modifying and sharpening the proofs of § 5.

Bibliography

1. F. Atkinson, *Normal solvability of linear equations in normed spaces,* Mat. Sb. **28** (1951), 3-14.
2. E. Balslev, *Perturbation of ordinary differential operators,* to appear, Math. Scand.
3. E. Balslev, *Perturbation of differential operators,* Ph. D. thesis (1963), University of California, Berkeley, Calif.
4. F. E. Browder, *On the spectral theory of elliptic differential operators I,* Math., Ann., **142** (1961), 22-130.
5. Dunford and Schwartz, *Linear Operators, Part I,* Interscience Press, 1958.
6. I. C. Gokhberg and M. G. Krein, *Fundamental theorems on deficiency numbers, root numbers and indices of linear operators,* Uspekhi Mat. Nauk, **12** (1957), 43-118, (Am. Math. Soc. Translations, Series 2, vol. 13).
7. Hardy, Littlewood and Polya, *Inequalities,* Cambridge, 1952.
8. T. Kato, *Perturbation theory for nullity, deficiency and other quantities of linear operators,* J. d'Analyse Math., **6** (1958), 261-322.
9. G. C. Rota, *Extension theory of differential operators,* Communs. Pure App. Math. **11** (1958), 23-65.
10. G. C. Rota, *On the spectra of singular boundary value problems,* J. Math. and Mech., **10** (1961), 83-90.
11. M. S. Birman, *On the spectrum of singular boundary value problems,* Mat. Sbornik T. 55 (2) **97** (1961), 125-174.

BOUNDS FOR DERIVATIVES IN ELLIPTIC BOUNDARY VALUE PROBLEMS

J. H. BRAMBLE AND L. E. PAYNE

I. Introduction. In a recent paper [7], Payne and Weinberger gave pointwise bounds for solutions of second order uniformly elliptic partial differential equations. The bounds for the function and its gradiant involved derivatives of the boundary data. Later [2] the present authors gave a method for obtaining bounds in which no derivatives of the boundary data appeared. Pointwise bounds for derivatives were not dealt with. In [4] the authors gave a method for bounding derivatives for Poisson's equation. The method was, however, restricted to the Laplace operator (or the constant coefficient case) and was not generally applicable.

In this paper we consider the operator

$$Lu \equiv (a^{ij}u_{,i})_{,j} \tag{1.1}$$

where u is a sufficiently smooth function defined in some region R (with boundary C) of Euclidean N dimensional space. Here the notation $u_{,i}$ denotes the partial derivative of u with respect to the cartesian coordinate x^i. In (1.1) the summation convention is used, i.e. $(a^{ij}u_{,i})_{,j} \equiv \sum_{i\ j=1}^{N} (a^{ij}u_{,i})_{,j}$. The coefficient matrix a^{ij} may be a function of position and is assumed to be uniformly positive definite and bounded above. That is there exist positive constant a_0 and a_1 such that

$$a_0 \sum_{i=1}^{N} \xi_i^2 \leqq a^{ij}\xi_i\xi_j \leqq a_1 \sum_{i=1}^{N} \xi_i^2 \tag{1.2}$$

for any real vector $\xi = (\xi_1, \cdots, \xi_N)$. We shall give a method involving the use of a parametrix, for obtaining bounds on any derivative of a function u at an arbitrary interior point P of R. These bounds are in terms of Lu and $\max_{S(P)} |u|$, where $S(P)$ is a sphere containing P. Estimates of this type for very general elliptic operators are described by John [6]. His method does not involve the parametrix and hence the expressions which could be derived would turn out to be quite different. Thus the problem is reduced to that of bounding $\max_{S(P)} |u|$ in terms of quantities which are data of some boundary value problem. We assume throughout that Lu and the coefficients a^{ij} are sufficiently smooth so that all subsequent indicated operations are valid.

In this paper we concern ourselves only with the derivation of appropriate a priori inequalities. The manner of applying such ine-

Received September 18, 1963. This research was supported in part by the National Science Foundation under grant-NSF GP-3.

qualitites to obtain bounds has been thoroughly discussed in previous papers (see e.g. [2, 4, 7]).

II. Mean value expressions. To obtain the desired bounds we shall first need a certain expression which is in a sense analogous to the solid mean value theorem for harmonic function. One such expression was given in [2]; however, it is quite complicated. We derive now a simpler expression.

Since a fundamental solution corresponding to the operator L is not in general known we make use of a Levi function (or parametrix) (c.f. Miranda [6]).

Let P and Q be two points in R. One possible definition of a parametrix is

$$(2.1)\qquad \begin{aligned} \Gamma(P, Q) &= -(2\pi)^{-1}[a(Q)a(P)]^{1/4}\log\rho\,, && N=2 \\ \Gamma(P, Q) &= 2^{1/2(N-2)}[(N-2)\omega_N]^{-1}[a(Q)a(P)]^{1/4}\rho^{-(N-2)}\,, && N\geqq 3 \end{aligned}$$

where ω_N denotes the surface of the unit sphere in N dimensions,

$$\rho^2 = [a_{ij}(Q) + a_{ij}(P)](x_P^i - x_Q^i)(x_P^j - x_Q^j)\,,$$

and $a(Q)$ denotes the determinant of the matrix $a_{ij}(Q)$, the inverse of $a^{ij}(Q)$. If the a^{ij} are twice continuously differentiable in the neighborhood of P, this function Γ has the property that

$$(2.2)\qquad L_Q\Gamma = O(r_{PQ}^{-(N-2)}),\ r_{PQ}\to 0$$

where r_{PQ} is the distance from P to Q. An alternate form for a parametrix is

$$(2.3)\qquad \begin{aligned} \bar{\Gamma}(P, Q) &= (2\pi)^{-1}[a(P)]^{1/2}\log\bar{\rho} \\ \bar{\Gamma}(P, Q) &= [(N-2)\omega_N]^{-1}[a(P)]^{1/2}[\bar{\rho}]^{-(N-2)}\,. \end{aligned}$$

Here $\bar{\rho}^2 = a_{ij}(P)(x_P^i - x_Q^i)(x_P^j - x_Q^j)$. The function $\bar{\Gamma}(P, Q)$ is such that if the a^{ij} are continuously differentiable in the neighborhood of P, then

$$(2.4)\qquad L_Q\bar{\Gamma} = O(r_{PQ}^{-(N-1)}),\ r_{PQ}\to 0\,.$$

Comparing (2.2) and (2.4) we see that Γ is a better approximation to the fundamental solution than is $\bar{\Gamma}$ near $Q = P$.

Now let $S_a(P)$ be the interior of a sphere of radius a with center at P, and such that $S_a(P)\subset R$. We define the function $f_n(P, Q)$ as follows (for P fixed)

$$(2.5)\qquad \begin{aligned} &\text{(a)}\quad f_n(P, Q) = \begin{cases} 1, & Q = P \\ 0, & r_{PQ}\geqq a \end{cases} \\ &\text{(b)}\quad f_n^{(i)}(P, P) = 0,\ i = 1, 2, \cdots, N-1 \\ &\text{(c)}\quad f_n(P, Q)\in C^{n-1}(E^N) \end{aligned}$$

(continuous derivatives up to and including those of order $n-1$ at each point of Euclidean N-space.) One such function, for example, is the polynomial with values

$$\left[\int_{r_{PQ}}^{a}\rho^{n-1}(a^2-\rho^2)^{n-1}d\rho\right]\left[\int_{0}^{a}\rho^{n-1}(a^2-\rho^2)^{n-1}d\rho\right]^{-1}, \; r_{PQ}\leqq a \, .$$

Another possible choice is the function

$$\left\{\int_{r_{PQ}}^{a}\exp\left[-\rho^{-2}(a^2-\rho^2)^{-1}\right]d\rho\right\}\left\{\int_{0}^{a}\exp\left[-\rho^{-2}(a^2-\rho^2)^{-1}\right]d\rho\right\}^{-1}, \; r_{PQ}\leqq a$$

which satisfies (2.5) for all n. Clearly

$$\Gamma_n(P, Q) \equiv f_n(P, Q)\Gamma(P, Q) \tag{2.6}$$

also satisfies (2.2). But $\Gamma_n(P, Q)$ has all derivatives up to and including those of order $n-1$ vanishing on $r_{PQ}=a$. Using (2.1) and (2.2) we find from Green's identity that

$$u(P)=\int_{S_a(P)}u(Q)L_Q\Gamma_n(P, Q)dV_Q-\int_{S_a(P)}\Gamma_n(P, Q)Lu(Q)dV_Q \, , \tag{2.7}$$

provided $n\geqq 2$. This expression is analogous to (5.8) of [2]. In addition to being simpler it possesses the advantage that the integration is taken over spheres, rather than ellipsoids which vary from point to point. We could as well have defined

$$\bar{\Gamma}_n(P, Q)=f_n(P, Q)\bar{\Gamma}(P, Q) \tag{2.8}$$

and obtained

$$u(P)=\int_{S_a(P)}u(Q)L_Q\bar{\Gamma}_n(P, Q)dV_Q-\int_{S_a(P)}\bar{\Gamma}_n(P, Q)Lu(Q)dV_Q \, , \tag{2.9}$$

with $n\geqq 2$.

III. Pointwise bounds. Either (2.7) or (2.9) can be used to obtain bounds in the Dirichlet problem. Using the Schwarz inequality we have

$$\left[\int_{S_a(P)}u(Q)L_Q\bar{\Gamma}_n(P, Q)dV_Q\right]^2\leqq\left[\int_R u^2r_{PQ}^{-q}dV\right]^2\left[\int_{S_a(P)}r_{PQ}^{q}(L\bar{\Gamma}_n)^2dV_Q\right]^2 . \tag{3.1}$$

Equation (2.9) together with (3.1) and the bounds given by Theorem I and II of [2], yield pointwise bounds for u in terms of Lu in R and the values of u on C.

In order to bound the first derivatives of u we can use (2.7), with $n\geqq 3$, to obtain

$$
\begin{aligned}
\frac{\partial u(P)}{\partial x_P^i} &= \int_{S_a(P)} u(Q) L_Q \frac{\partial \Gamma_n(P, Q)}{\partial x_P^i} d V_Q \\
&\quad - \frac{\partial}{\partial x_P^i}\left[\int_{S_a(P)} \Gamma_n(P, Q) L u(Q) d V_Q\right] .
\end{aligned}
\tag{3.2}
$$

Hence we have

$$
\begin{aligned}
\left|\frac{\partial u(P)}{\partial x_P^i}\right| &\leqq \max_{Q \in S_a(P)} |u(Q)| \int_{S_a(P)} \left| L_Q \frac{\partial \Gamma_n(P, Q)}{\partial x_P^i}\right| d V_Q \\
&\quad + \left|\frac{\partial}{\partial x_P^i}\left[\int_{S_a(P)} \Gamma_n(P, Q) L u(Q) d V_Q\right]\right| .
\end{aligned}
\tag{3.3}
$$

Now if a is so chosen that we can obtain a bound for $\max_{Q \in S_a(P)} |u(Q)|$ then (3.3) provides a bound for $|\partial u(P)/\partial x_P^i|$. If, for example, the least distance from P to the boundary C is r_0, then we could choose $a = (1/2) r_0$. Thus the closure $\overline{S_a(P)}$ of $S_a(P)$ is a compact subset of R and hence only interior bounds for u are required. Note that we could not replace (3.2) by a similar expression involving $\bar{\Gamma}_n$ since the integrals on the right would not exist.

We note from (3.2) that

$$
\int_{S_a(P)} L_Q \frac{\partial \Gamma_n(P, Q)}{\partial x_P^i} d V_Q = 0 .
\tag{3.4}
$$

Thus if $n \geqq 4$ we have the representation

$$
\begin{aligned}
\frac{\partial^2 u(P)}{\partial x_P^i \partial x_P^j} &= \int_{S_a(P)} [u(Q) - u(P)] L_Q \frac{\partial^2 \Gamma_n(P, Q)}{\partial x_P^i \partial x_P^j} d V_Q \\
&\quad - \frac{\partial^2}{\partial x_P^i \partial x_P^j}\left[\int_{S_a(P)} \Gamma_n(P, Q) L u(Q) d V_Q\right]
\end{aligned}
\tag{3.5}
$$

since

$$
[u(Q) - u(P)] L_Q \frac{\partial^2 \Gamma_n(P, Q)}{\partial x_P^i \partial x_P^j} = O(r_{PQ}^{-(N-1)})
\tag{3.6}
$$

for $r_{PQ} \to 0$. From (3.5) we see that

$$
\begin{aligned}
\left|\frac{\partial^2 u(P)}{\partial x_P^i \partial x_P^j}\right| &\leqq \max_{Q \in S_a(P)} \left|\frac{u(Q) - u(P)}{r_{PQ}}\right| \int_{S_a(P)} r_{PQ} \left| L_Q \frac{\partial^2 \Gamma_n(P, Q)}{\partial x_P^i \partial x_Q^j}\right| d V_Q \\
&\quad + \left|\frac{\partial^2}{\partial x_P^i \partial x_P^j}\left[\int_{S_a(P)} \Gamma_n(P, Q) L u(Q) d V_Q\right]\right| .
\end{aligned}
\tag{3.7}
$$

Now

$$
\max_{Q \in S_a(P)} \left|\frac{u(Q) - u(P)}{r_{PQ}}\right| \leqq \max_{Q \in S_a(P)} |\operatorname{grad} u(Q)| .
\tag{3.8}
$$

Clearly we can use (3.3) with a smaller value of a to bound the right hand side of (3.8). Thus we can bound an arbitrary second derivative of u in terms of Lu in R and the maximum of $|u|$ over a compact subset of R. In order to treat an arbitrary third derivative we note from (3.5) that

$$(3.9)\quad \int_{S_a(P)} (x_Q^\alpha - x_P^\alpha) L_Q \frac{\partial^2 \Gamma_n(P, Q)}{\partial x_P^i \partial x_P^j} d V_Q = \frac{\partial^2}{\partial x_P^i \partial x_P^j}\left[\int_{S_a(P)} \Gamma_n(P, Q) L x_Q^\alpha d V_Q\right]$$

for $\alpha, i, j = 1, \cdots, N$. Combining (3.9) and (3.5) we have

$$\begin{aligned} \frac{\partial^2 u(P)}{\partial x_P^i \partial x_P^j} &= \int_{S_a(P)} [u(Q) - u(P) - (x_Q^\alpha - x_P) u_{,\alpha}(P)] L_Q \frac{\partial^2 \Gamma_n(P, Q)}{\partial x_P^i \partial x_P^j} d V_Q \\ (3.10)\qquad & - \frac{\partial^2}{\partial x_P^i \partial x_P^j}\left[\int_{S_a(P)} \Gamma_n(P, Q) Lu(Q) d V_Q\right] \\ & - u_{,\alpha}(P) \frac{\partial^2}{\partial x_P^i \partial x_P^j}\left[\int_{S_a(P)} \Gamma_n(P, Q) L x_Q^\alpha d V_Q\right] \end{aligned}$$

where we have summed over α from 1 to N. It follows from (3.10) that if $n \geqq 5$

$$\begin{aligned} \frac{\partial^3 u(P)}{\partial x_P^i \partial x_P^j \partial x_P^k} &= \int_{S_a(P)} [u(Q) - u(P) - (x_Q^\alpha - x_P^\alpha) u_{,\alpha}(P)] L_Q \frac{\partial^3 \Gamma_n(P, Q)}{\partial x_P^i \partial x_P^j \partial x_P^k} d V_Q \\ (3.11)\qquad & - \frac{\partial^3}{\partial x_P^i \partial x_P^j \partial x_P^k}\left[\int_{S_a(P)} \Gamma_n(P, Q) Lu(Q) d V_Q\right] \\ & - u_{,\alpha}(P) \frac{\partial^3}{\partial x_P^i \partial x_P^j \partial x_P^k}\left[\int_{S_a(P)} \Gamma_n(P, Q) L x_Q^\alpha d V_Q\right]. \end{aligned}$$

The first integral on the right may be bounded as

$$\begin{aligned} (3.12)\quad & \left|\int_{S_a(P)} [u(Q) - u(P) - (x_Q^\alpha - x_P^\alpha) u_{,\alpha}(P)] L_Q \frac{\partial^3 \Gamma_n(P, Q)}{\partial x_P^i \partial x_P^j \partial x_P^k} d V_Q\right| \\ & \qquad \leqq \max_{\substack{Q \in S_a(P) \\ \alpha\ \beta = 1 \cdots N}} |u_{,\alpha\beta}(Q)| \int_{S_a(P)} r_{PQ}^2 \left|L_Q \frac{\partial^3 \Gamma_n(P, Q)}{\partial x_P^i \partial x_P^j \partial x_P^k}\right| d V_Q . \end{aligned}$$

Now (3.11) and (3.12) can be used to reduce the problem of bounding third derivatives to that of bounding second derivatives. It is clear how to proceed to higher derivatives. In each of the preceding bounds certain differentiability assumptions must be made. These conditions become more and more stringent the more derivatives of u that we wish to bound. Some conditions of this nature are of course required since in general u cannot be expected to be smooth.

Thus for an arbitrary derivative at P the method described above yields a bound in terms of Lu in R and the maximum of $|u|$ on a compact subset (for example $S_a(P)$ for some a) of R. These bounds, together with bounds for $|u|$ in $S_a(P)$ in terms of data in various boundary value problems, yield pointwise bounds for derivatives at interior points in terms of the respective data. For such bounds see [1, 2, 3, 4, 5, 7, 8].

The techniques which we have used here to bound derivatives of solutions to boundary value problems at interior points in terms of the operator and bounds for the solution itself, will carry over quite naturally to higher order equations and to equations of other than elliptic type.

BIBLIOGRAPHY

1. F. J. Bellar, *Pointwise bounds in parabolic and elliptic partial differential equations*, Ph. D. dissertation, University of Maryland (1961).

2. J. H. Bramble, L. E. Payne, *Bounds for solutions of second order elliptic partial differential equations*, Contributions to Differential Equations, Vol. I, (1963), 95-127.

3. ———, *Some integral inequalities for uniformly elliptic operators*, Contributions to Differential Equations, Vol. I, (1963), 129-135.

4. ———, *Bounds for derivatives in the Dirichlet problem for Poisson's equation*, J. Soc. Ind. Appl. Math., **10** (1962), 370-380.

5. ———, *Bounds in the Neumann problem for second order uniformly elliptic operators*, Pacific J. Math., **12** (1962), 823-833.

6. C. Miranda, *Equazioni alle derivate parziali di tipo ellitico*, Springer-Verlag, Berlin (1955).

7. L. E. Payne, H. F. Weinberger, *New bounds for solutions of second order elliptic partial differential equations*, Pacific J. Math., **8** (1958), 551.

8. G. N. Trytten, *Pointwise bounds for solutions in the Cauchy problem for elliptic equations*, Arch. Rat. Mech. Anal., **13** (1963), 222.

UNIVERSITY OF MARYLAND

INTEGRAL INEQUALITIES FOR FUNCTIONS WITH NONDECREASING INCREMENTS

H. D. BRUNK

1. Introduction. One of the fundamental inequalities of analysis is Jensen's inequality,

$$\int f(x)\,dG(x) \geqq f\left(\int x\,dG(x)\right), \tag{1.1}$$

for convex f, with G a probability distribution function. However, G need not be a probability distribution function in order that (1.1) hold for all convex f. Let $X(t)$ be nondecreasing for $\alpha \leqq t \leqq \beta$. It was shown in [1] that under mild regularity conditions on G, if $G(\alpha) = 0$, necessary and sufficient conditions for

$$\int_\alpha^\beta f[X(t)]\,dG(t) \geqq f\left(\int_\alpha^\beta X(t)\,dG(t)\right) \tag{1.2}$$

for all convex f are

$$G(\beta) = 1\,, \tag{1.3}$$

and

$$\int_\alpha^t G(u)\,dX(u) \geqq 0\,, \quad \int_t^\beta [1 - G(u)]\,dX(u) \geqq 0 \quad \text{for } \alpha \leqq t \leqq \beta\,. \tag{1.4}$$

This result was applied to show that:

(i) sufficient conditions in order that (1.2) hold for convex f are $X(\alpha) = 0, f(0) \leqq 0$, and $0 \leqq G(t) \leqq 1$ for $\alpha \leqq t \leqq \beta$; and

(ii) if f is convex on $[0, b]$ with $f(0) \leqq 0$, if $0 \leqq a_1 \leqq \cdots \leqq a_m \leqq b$, if $0 \leqq h_1 \leqq \cdots \leqq h_m \leqq 1$, then

$$\sum_{j=1}^m (-1)^{j-1} h_j f(a_j) \geqq f\left[\sum_{j=1}^m (-1)^{j-1} h_j a_j\right]. \tag{1.5}$$

The latter, (ii), was proved independently by Olkin [5]. Ciesielski [2] obtained results (under unnecessarily stringent hypotheses) related to (i) through change of variable, and obtained also analogous two-dimensional results. These provided part of the motivation for the present study of k-dimensional analogues of (1.2).

In the present paper, $X(\cdot)$ denotes a map from the real interval $[\alpha, \beta)$ into an interval I in k-dimensional Euclidean space R^k such that each component of X is nondecreasing. The function f is a map from

Received November 29, 1963. This research was supported by the United States Air Force Office of Scientific Research.

R^k into the reals. The property of f critical for inequality (1.2) in this context is that of having nondecreasing increments, rather than convexity; for $k=1$ it coincides with convexity. Functions with nondecreasing increments are discussed briefly in § 2. In § 3, conditions (1.3) and (1.4) are shown to be necessary and sufficient for (1.2) ($k \geqq 1$), and k-dimensional analogues are given of (i) and (ii), above. Section 4 is devoted to the k-dimensional analogue of a related theorem of Levin and Stečkin [4], giving conditions on H necessary and sufficient in order that $\int_\alpha^\beta f[X(t)]\,dH(t) \geqq 0$ for all f with nondecreasing increments.

2. Functions with nondecreasing increments. Let R^k denote the k-dimensional vector lattice of points $x=(x_1,\cdots,x_k)$, x_i real for $i=1,2,\cdots,k$, with the partial ordering $x=(x_1,\cdots,x_k)\leqq y=(y_1,\cdots,y_k)$ if and only if $x_i \leqq y_i$ for $i=1,2,\cdots,k$.

DEFINITION 2.1. A real-valued function f on an interval $I \subset R^k$ will be said to have *nondecreasing increments* if

$$f(a+h)-f(a) \leqq f(b+h)-f(b) \tag{2.1}$$

whenever $a \in I$, $b+h \in I$, $0 \leqq h \in R^k$, $a \leqq b$. Even in the one-dimensional case, $k=1$, this does not imply continuity. Indeed, every solution of Cauchy's equation, $f(x+y)=f(x)+f(y)$, has equal increments. (Note that if $f_1, f_2, \cdots, f_k$ are functions of a single real variable satisfying Cauchy's equation, then $f(x) \equiv \sum_{i=1}^k f_i(x_i)$ is a function on R^k satisfying Cauchy's equation.) However, our interest in this paper is solely in *continuous* functions with nondecreasing increments.

It is of interest to note that such a function is convex along positively oriented lines, i.e., lines whose direction cosines are nonnegative, with equations of the form $x=at+b$ where $(0,\cdots,0)\leqq a \in R^k$, $b \in R^k$. If $f(x)$ is continuous with nondecreasing increments for $b \leqq x \leqq a+b$, set $\varphi(t)=f(at+b)$, $0 \leqq t \leqq 1$. In order to prove φ convex, it suffices [3, Theorem 86, page 72] to show that $[\varphi(r)+\varphi(s)]/2 \geqq \varphi[(r+s)/2]$ for $0 \leqq r \leqq s \leqq 1$. Set $c=(s-r)/2$. Then $\varphi(s)-\varphi[(r+s)/2] = \varphi(r+2c)-\varphi(r+c) = f(ar+b+2ca)-f(ar+b+ca) \geqq f(ar+b+ca) - f(ar+b) = \varphi(r+c)-\varphi(r) = \varphi[(r+s)/2]-\varphi(r)$. Thus φ is convex.

It is immediate from the definition that if the partial derivatives $f_i(x) \equiv \partial f/\partial x_i\,(x_1,\cdots,x_k)$ exist for $x \in I$, then f has nondecreasing increments if and only if each of these partial derivatives is nondecreasing in each argument; in other words, if and only if the gradient, $\nabla f \equiv (f_1(x),\cdots,f_k(x))$ is nondecreasing on I. The second partials, if they exist, are then nonnegative. If f is continuous and has nondecreasing increments on I, it may be approximated uniformly on I by polynomials having nondecreasing increments and therefore nonnegative second partial derivatives. To see this, let us set, for convenience, $I = \{x: x \in R^k, (0, \cdots, 0) \leqq x \leqq (1, \cdots, 1)\}$. It is known that the Bernstein polynomials

$$\sum_{i_1=0}^{n_1} \sum_{i_2=0}^{n_2} \cdots \sum_{i_k=0}^{n_k} f(i_1/n_1, i_2/n_2, \cdots, i_k/n_k) \prod_{j=1}^{k} \binom{n_j}{i_j} x_j^{i_j}(1 - x_j)^{n_j - i_j}$$

converge uniformly to f on I as $n_1 \to \infty, \cdots, n_k \to \infty$, if f is continuous. Further, if f has nondecreasing increments these polynomials have nonnegative second partial derivatives, as may be shown by repeated application of the formula

$$(d/dx) \sum_{i=0}^{n} \binom{n}{i} a_i x^i (1 - x)^{n-i} = n \sum_{i=0}^{n-1} \binom{n-1}{i} (a_{i+1} - a_i)\, x^i (1 - x)^{n-1-i} .$$

3. A line integral inequality of Jensen's type. Perhaps the most direct analogue of Jensen's inequality for f defined on an interval $I \subset R^k$ would involve the integral of f over I with respect to a normed measure. The inequality we treat here, however, deals with a line integral over a positively oriented curve. By the term "positively oriented curve" we understand a nondecreasing map $X = (X_1, \cdots, X_k)$ of a real interval $[\alpha, \beta)$ into an interval $I \subset R^k$: $\alpha \leqq t' \leqq t'' < \beta$ implies $X(t') \leqq X(t'')$, i.e., $X_i(t') \leqq X_i(t'')$ for $i = 1, 2, \cdots, k$. Theorem 3.1, below, relates such a map X and a real valued function G of bounded variation on $[\alpha, \beta)$. The integrals $\int_{[\alpha,\beta)} X\, dG$ and $\int_{[\alpha,\beta)} G\, dX$ appearing in the statement of Theorem 3.1 are related through the formula for integration by parts: $\int_J X\, dG + \int_J G\, dX = \int_J d(XG)$ for every interval $I \subset [\alpha, \beta)$ $\left(\text{by } \int_J X\, dG \text{ we understand the vector } \left(\int_J X_1\, dG, \cdots, \int_J X_k\, dG\right),\right.$ $\left.\text{and similarly for } \int_J G\, dx, \int_J d(XG)\right)$. In order for this to hold and also to avoid minor difficulties in the determination of G at common points of discontinuity of X and G, we shall assume henceforth without further reference that X *is nondecreasing and continuous from the right* (*i.e.*, X_i *is nondecreasing and continuous from the right for* $i = 1, \cdots, k$) *and* G *is continuous from the left*. For simplicity of notation, we write $X(\beta)$ for $X(\beta^-)$ and $G(\beta)$ for $G(\beta^-)$. Some further bits of notation will be required: the symbol $[\alpha, t\}$ will refer to either of the left intervals $[\alpha, t]$ or $[\alpha, t)$; and $\{t, \beta)$ to either of the right intervals $[t, \beta)$ or (t, β). Also, if $a = (a_1, \cdots, a_k) \in R^k$, then $a^+ = (a_1^+, \cdots, a_k^+)$, where $a_i^+ = \max(a_i, 0)$, $i = 1, 2, \cdots, k$. Further, we set $\xi_i = \int_{[\alpha,\beta)} X_i\, dG$, $i = 1, 2, \cdots, k$, and $\xi = \int_{[\alpha,\beta)} X\, dG$.

THEOREM 3.1. *If* $G(\alpha) = 0$, *then necessary and sufficient conditions in order that*

$$\int_{[\alpha,\beta)} f[X(t)]\,dG(t) \geqq f\left[\int_{[\alpha,\beta)} X(t)\,dG(t)\right] \tag{3.1}$$

for every continuous function f on I with nondecreasing increments are

$$G(\beta) = 1 \tag{3.2}$$

and

$$\begin{cases} \int_{[\alpha,t\}} G\,dX \geqq 0 \textit{ for every left interval } [\alpha, t\} \subset [\alpha, \beta) \textit{ and} \\ \int_{\{t,\beta)} [1-G]\,dX \geqq 0 \textit{ for every right interval } \{t, \beta) \subset [\alpha, \beta)\,. \end{cases} \tag{3.3}$$

The case $k = 1$ of Theorem 3.1 appears in [1]. We note that for $k = 1$ the class of continuous functions with nondecreasing increments is identical with that of continuous convex functions. If $k > 1$, (3.2) (3.3) do not imply (3.1) for all continuous convex f. For example, set $X(t) = (0, 2t)$ for $0 \leqq t \leqq 1/2$, $X(t) = (2t - 1, 1)$ for $1/2 \leqq t \leqq 2$, $G(0) = 0$, $G(2) = 1$, and let G have saltus 1 at $t = 0$, saltus -1 at $t = 1/2$, and saltus 1 at $t = 1$, being constant on each of the intervening intervals. Set $f(x) = (x_1 - x_2)^2$, where $x = (x_1, x_2)$; then f is convex, but does not have nondecreasing increments. We have $\int_{[0,2)} f[X(t)]\,dG(t) = -1$, while $f\left[\int_{[0,2)} X(x)\,dG(t)\right] = 1$, so that (3.1) fails, although (3.2) and (3.3) are satisfied: indeed, $0 \leqq G \leqq 1$ (cf. Lemma 3.1).

Before proceeding to the proof of Theorem 3.1, we examine relations among the following properties of G, for given X:

$$0 \leqq G(t) \leqq 1 \quad \text{for} \quad t \in [\alpha, \beta)\,; \tag{3.4}$$

$$\begin{cases} \int_{[\alpha,t\}} G\,dX \geqq 0 \text{ for every left interval } [\alpha, t\} \subset [\alpha, \beta)\,, \text{ and} \\ \int_{\{t,\beta)} [1-G]\,dX \geqq 0 \text{ for every right interval } \{t, \beta] \subset [\alpha, \beta)\,; \end{cases} \tag{3.3}$$

$$\begin{cases} \int_{[\alpha,t]} G\,dX \geqq [X(t^+) - \xi]^+ \quad \text{for } t \in [\alpha, \beta)\,, \\ \int_{[\alpha,t)} G\,dX \geqq [X(t^-) - \xi]^+ \quad \text{for } t \in [\alpha, \beta)\,; \end{cases} \tag{3.5}$$

$$\begin{cases} \int_{[t,\beta)} [1-G]\,dX \geqq [\xi - X(t^-)]^+ \quad \text{for } t \in [\alpha, \beta)\,, \\ \int_{(t,\beta(} [1-G]\,dX \geqq [\xi - X(t^+)]^+ \quad \text{for } t \in [\alpha, \beta)\,. \end{cases} \tag{3.6}$$

LEMMA 3.1. *We have* (3.4) $\Rightarrow$ (3.3). *Also, if* $G(\alpha) = 0$ *and* $G(\beta) = 1$, *then* (3.3) $\Leftrightarrow$ (3.5) $\Leftrightarrow$ (3.6).

Proof. That (3.4) implies (3.3) is obvious. Also, if $G(\alpha) = 0$, $G(\beta) = 1$, then

$$\xi = \int_{[\alpha,\beta)} X\,dg = X(\alpha) + \int_{[\alpha,\beta)} (1 - G)\,dX\,,$$

so that

$$\int_{\{t,\beta)} (1 - G)\,dX = \xi - X(\alpha) - \int_{[\alpha,t\}} (1 - G)\,dX\,,$$

where $[\alpha, t\} \cup \{t, \beta)$ is a disjoint partition of $[\alpha, \beta)$; or,

$$\left\{\begin{aligned} \int_{[t,\beta)} (1 - G)\,dX &= \xi - X(t^-) + \int_{[\alpha,t)} G\,dX\,, \\ \int_{(t,\beta)} (1 - G)\,dX &= \xi - X(t^+) + \int_{[\alpha,t]} G\,dX\,. \end{aligned}\right. \tag{3.7}$$

Thus (3.3) implies that

$$\int_{[\alpha,t)} G\,dX \geqq X(t^-) - \xi\,, \quad \int_{[\alpha,t]} G\,dX \geqq X(t^+) - \xi\,.$$

With the first inequality in (3.3), this implies (3.5). Thus (3.3) $\Rightarrow$ (3.5). Also, it is clear from (3.7) that (3.5) and (3.6) are equivalent. Finally, (3.5) and (3.6) clearly imply (3.3), and the proof of Lemma 3.1 is complete.

Lemma 3.2 will be used in the proof of the sufficiency of the conditions in Theorem 3.1.

LEMMA 3.2. *Under the hypotheses of Theorem* 3.1, *and conditions* (3.2) *and* (3.3),

$$\int_{[\alpha,\beta)} \nabla f[X(t)]\cdot d[X(t) - \xi]^+ \leqq f[X(\beta)] - f(\xi)\,.$$

Proof. We observe first that $X(\alpha) \leqq \xi \leqq X(\beta)$. This follows from the inequalities

$$0 \leqq \int_{[\alpha,\beta)} G(u)\,dX(u) = X(\beta) - \int_{[\alpha,\beta)} X(u)\,dG(u) = X(\beta) - \xi\,,$$

and

$$0 \leqq \int_{[\alpha,\beta)} [1 - G(u)]\,dX(u) = -X(\alpha) + \int_{[\alpha,\beta)} X(u)\,dG(u) = \xi - X(\alpha)\,.$$

Since X is nondecreasing, there is, for $i = 1, 2, \cdots, k$, a unique smallest real number τ_i such that $X_i(\tau_i^-) \leqq \xi_i \leqq X_i(\tau_i^+)$. Suppose $\tau_1 \leqq \tau_2 \leqq \cdots \leqq \tau_k$; the proof is similar for other orderings. We have

$$\begin{aligned}
&\int_{[\alpha,\beta)} \nabla f[X(t)]\cdot d[X(t)-\xi]^+ \\
&\quad = \sum_{i=1}^{k} \int_{[\alpha,\beta)} f_i[X(t)]\, d[X_i(t)-\xi_i]^+ \\
&\quad = \sum_{i=1}^{k} \int_{[\tau_i,\beta)} f_i[X(t)]\, dX_i(t) \\
&\quad = \int_{[\tau_k,\beta)} \sum_{i=1}^{k} f_i[X(t)]\, dX_i(t) + \int_{[\tau_{k-1},\tau_k)} \sum_{i=1}^{k-1} f_i[X(t)]\, dX_i(t) \\
&\qquad + \cdots + \int_{[\tau_1,\tau_2)} f_1[X(t)]\, dX_1(t)\,.
\end{aligned}$$

Since $f_i(x) = f_i(x_1, \cdots, x_k)$ is nondecreasing in each argument, $i = 1, 2, \cdots, k$, we have, for $1 \leqq i < j \leqq k$, and for $\tau_{j-i} \leqq t < \tau_j$,

$$\begin{aligned}
&f_i[X_1(t), \cdots, X_k(t)] \\
&\quad \leqq f_i[X_1(t), \cdots, X_{j-i}(t), X_j(\tau_j^-), X_{j+1}(\tau_{j+1}^-), \cdots, X_k(\tau_k^-)] \\
&\quad \leqq f_i[X_1(t), \cdots, X_{j-1}(t), \xi_j, \cdots, \xi_k]\,.
\end{aligned}$$

It follows that

$$\begin{aligned}
&\int_{[\tau_{j-1},\tau_j)} \sum_{i=1}^{j-1} f_i[X(t)]\, dX_i(t) \\
&\quad \leqq \int_{[\tau_{j-1},\tau_j)} \sum_{i=1}^{j-1} f_i[X_1(t), \cdots, X_{j-1}(t), \xi_j, \cdots, \xi_k]\, dX_i(t) \\
&\quad = \int_{[\tau_{j-1},\tau_j)} \nabla f[X_1(t), \cdots, X_{j-1}(t), \xi_j, \cdots, \xi_k] \\
&\qquad \cdot d[X_1(t), \cdots, X_{j-1}(t), \xi_j, \cdots, \xi_k] \\
&\quad = f[X_1(\tau_j^-), \cdots, X_{j-1}(\tau_j^-), \xi_j, \cdots, \xi_k] \\
&\qquad - f[X_i(\tau_{j-1}^-), \cdots, X_{j-1}(\tau_{j-1}^-), \qquad \xi_j, \cdots, \xi_k],\ j = 2, \cdots, k\,.
\end{aligned}$$

Therefore

$$\begin{aligned}
\int_{[\alpha,\beta)} \nabla f[X(t)]\cdot d[X(t)-\xi]^+ &\leqq f[X_1(\beta), \cdots, X_k(\beta)] \\
&\quad - f[X_1(\tau_k^-), \cdots, X_{k-1}(\tau_k^-), \xi_k] \\
&\quad + \sum_{j=2}^{k} \{f[X_1(\tau_j^-), \cdots, X_{j-1}(\tau_j^-), \xi_j, \cdots, \xi_k] \\
&\quad - f[X_1(\tau_{j-1}^-), \cdots, X_{j-2}(\tau_{j-1}^-), \xi_{j-1}, \cdots, \xi_k]\} \\
&= f[X(\beta)] - f(\xi)\,.
\end{aligned}$$

This completes the proof of Lemma 3.2.

Proof of Theorem 3.1; *necessity*. Equation (3.2) follows from (3.1) with $f \equiv 1$ and $f \equiv -1$. For $1 \leqq i \leqq k$, and $\alpha \leqq t < \beta$, set $f(x) = f(x_1, \cdots, x_k) \equiv [x_i - X_i(t^-)]^+$. For this function f, (3.1) yields

$$\int_{[\alpha,\beta)} [X_i(u) - X_i(t^-)]^+ \, dG(u) \geqq [\xi_i - X_i(t^-)]^+ \, .$$

But

$$\begin{aligned}\int_{[\alpha,\beta)} [X_i(u) - X_i(t^-)]^+ \, dG(u) &= \int_{[t,\beta)} [X_i(u) - X_i(t^-)] \, dG(u) \\ &= \int_{[t,\beta)} [1 - G(u)] \, dX_i(u) \, ,\end{aligned}$$

so that

$$\int_{[t,\beta)} [1 - G(u)] \, dX_i(u) \geqq [\xi_i - X_i(t^-)]^+ \, , \qquad i = 1, 2, \cdots, k \, ,$$

verifying the first part of (3.6). The verification of the second part is similar. With Lemma 3.1, this completes the proof of the necessity of (3.2) and (3.3).

Sufficiency. Set $Q(t) = (Q_1(t), \cdots, Q_k(t)) = \int_{[\alpha,t)} G(u) \, dX(u)$ for $\alpha \leqq t < \beta$. Then by (3.5) we have $Q(t^\pm) \geqq [X(t^\pm) - \xi]^+$ for $\alpha \leqq t < \beta$. Since f can be approximated uniformly in I by polynomials with nondecreasing increments, there is no loss in generality in assuming that the partial derivatives $f_i(x)$, $i = 1, 2, \cdots, k$, exist and are nondecreasing in each argument. We then have, for $i = 1, 2, \cdots, k$,

$$\begin{aligned}\int_{[\alpha,\beta)} f_i[X(t)] \, dQ_i(t) &= f_i[X(\beta)]Q_i(\beta) - \int_{[\alpha,\beta)} Q_i(t) \, df_i[X(t)] \\ &\leqq f_i[X(\beta)]Q_i(\beta) - \int_{[\alpha,\beta)} [X_i(t) - \xi]^+ \, df_i[X(t)] \\ &= \int_{[\alpha,\beta)} f_i[X(t)] \, \mathrm{d}[X_i(t) - \xi]^+ \, ,\end{aligned}$$

since

$$Q(\beta) = [X(\beta) - \xi] = [X(\beta) - \xi]^+$$

by (3.7). Therefore

$$\begin{aligned}\int_{[\alpha,\beta)} f[X(t)] \, dG(t) &= f[X(\beta)] - \int_{[\alpha,\beta])} G(t) \, \nabla f[X(t)] \cdot dX(t) \\ &= F[X(\beta)] - \int_{[\alpha,\beta)} \nabla f[X(t)] \cdot dQ(t) \\ &\geqq f[X(\beta)] - \int_{[\alpha,\beta)} \nabla f[X(t)] \cdot d[X(t) - \xi]^+ \geqq f(\xi)\end{aligned}$$

by Lemma 3.2. This completes the proof of the theorem.

In each of the following corollaries, Corollary 3.1 and Corollary 3.2, it is assumed that X is a nondecreasing map, continuous from the right, from $[\alpha, \beta)$ into a k-dimensional interval I containing the origin $0 = (0, \cdots, 0)$; that f is a continuous function from I into the reals which has nondecreasing increments; that G is a real-valued function of bounded variation on $[\alpha, \beta)$, continuous from the left, and that $G(\beta) = 1$.

COROLLARY 3.1. *If* $X(\alpha) = 0 = (0, \cdots, 0)$, *if* $f(0) \leqq 0$, *if* $G(\alpha) \geqq 0$, *and if* (3.3) *holds, then*

$$\int_{[\alpha,\beta)} f[X(t)]\, dG(t) \geqq f\left[\int_{[\alpha,\beta)} X(t)\, dG(t)\right]. \tag{3.1}$$

The case $k = 1$ of this corollary appears in [1].

Proof. Set $G_1(t) = G(t)$ for $t > \alpha$, $G_1(\alpha) = 0$. Then by Theorem 3.1,

$$\int_{[\alpha,\beta)} f[X(t)]\, dG_1(t) \geqq f\left[\int_{[\alpha,\beta)} X(t)\, dG_1(t)\right].$$

But

$$\int_{[\alpha,\beta)} X(t)\, dG_1(t) = \int_{[\alpha,\beta)} X(t)\, dG(t)$$

since $X(\alpha) = 0$. Also

$$\int_{[\alpha,\beta)} f[X(t)]\, dG_1(t) = f(0)G(\alpha) + \int_{[\alpha,\beta)} f[X(t)]\, dG(t)\,,$$

and (3.1) follows.

COROLLARY 3.2. *If either*
(i) $G(\alpha) = 0$ *or*
(ii) $X(\alpha) \geqq 0, f(0) \leqq 0$, *and if*
(iii) $0 \leqq G(t) \leqq 1$ *for* $\alpha \leqq t < \beta$,
then

$$\int_{[\alpha,\beta)} f[X(t)]\, dG(t) \geqq f\left[\int_{[\alpha,\beta)} X(t)\, dG(t)\right]. \tag{3.1}$$

Proof. By Lemma 3.1, (iii) implies (3.3) so that under hypotheses (i) and (iii), (3.1) is immediate from Theorem 3.1. If (ii) and (iii) hold, choose $\alpha^* < \alpha$, set $X^*(\alpha^*) = 0, X^*(t) = X(t)$ for $\alpha \leqq t < \beta$, and let X^* be linear for $\alpha^* \leqq t \leqq \alpha$. Set $G^*(\alpha^*) = 0, G^*(t) = G(\alpha)$ for $\alpha^* \leqq t \leqq \alpha, G^*(t) = G(t)$ for $\alpha \leqq t < \beta$. Then $G^*(\beta) = 1, G^*(\alpha^*) = 0$, and $0 \leqq G^* \leqq 1$. From Lemma 3.1 and Theorem 3.1 it follows that

$$\int_{[\alpha^*,\beta)} f[X^*(t)]\, dG(t) \geqq f\left[\int_{[\alpha^*,\beta)} X^*(t)\, dG^*(t)\right].$$

But

$$\int_{[\alpha^*,\beta)} X^*(t)\, dG^*(t) = \int_{[\alpha,\beta)} X(t)\, dG(t)\ ,$$

and

$$\int_{[\alpha^*,\beta)} f[X^*(t)]\, dG^*(t) = f(0)G(\alpha) + \int_{[\alpha,\beta)} f[X(t)]\, dG(t)\ ,$$

Since $f(0) \leqq 0$ and $G(a) \geqq 0$, conclusion (3.1) follows.

REMARKS ON COROLLARY 3.2. The case $k = 1$ of Corollary 3.2 appears in [1] with the hypothesis $X(\alpha) = 0$. With a change of variable in Corollary 3.2 we obtain the following theorem.

Let Y be a nonincreasing map, continuous from the left, from $(0, 1]$ into $I \subset R^k$, with $Y(1) \geqq 0$. Let H be continuous from the right and of bounded variation on $(0, 1]$, and suppose $H(0) = 0$, $H(t) \geqq 0$ on $(0, 1]$, $\int_{(0,1]} |dH(t)| > 0$. If f is continuous with nondecreasing increments on I, and if $f(0) \leqq 0$, then

$$\int_{(0,1]} f(Y)\, dH \Big/ \int_{(0,1]} |dH| \geqq f\Big(\int_{(0,1]} Y\, dH \Big/ \int_{(0,1]} |dH|\Big)\ .$$

It suffices to set $X(t) = Y(1 - t)$, $G(t) = 1 - \Big[H(1 - t) \Big/ \int_{(0,1]} dH(t)\Big]$ on $[0, 1)$ in Corollary 3.2. Cases $k = 1$ and $k = 2$ of this latter theorem, for discrete and for continuous H, appear in [2], with additional hypotheses: for $k = 1$, that f' is convex; and for $k = 2$, that the first partial derivatives are convex along positively oriented lines.

Ciesielski points out (in the two-dimensional case) that setting $f(x_1, x_2) = x_1 x_2$ yields a generalization of an inequality of Chebyshev [3, page 43]: *if Y_1, Y_2 are nonincreasing, nonnegative and continuous from the left on $(0, 1]$, if H is continuous from the right and of bounded variation on $(0, 1]$, and if $H(0) = 0$, $H(t) \geqq 0$ on $(0, 1]$, then*

$$\int_{(0,1]} Y_1\, Y_2\, dH \int_{(0,1]} |dH| \geqq \int_{(0,11} Y_1\, dH \int_{(0,1]} Y_2\, dH\ .$$

COROLLARY 3.3. *Let f be a continuous map from a k-dimensional interval I containing the origin into the reals, with nondecreasing increments, such that $f(0) \leqq 0$. Let m be a positive integer, and let $1 \geqq h_1 \geqq h_2 \geqq \cdots \geqq h_m \geqq 0$. Let $a_j \in I$, $j = 1, 2, \cdots, m$, with $(1, \cdots, 1) > a_1 \geqq a_2 \geqq \cdots \geqq a_m \geqq (0, \cdots, 0)$. Then*

$$\sum_{j=1}^{m} (-1)^{j-1} h_j f(a_j) \geqq f\Big[\sum_{j=1}^{m} (-1)^{j-1} h_j a_j\Big]\ . \tag{3.8}$$

For inequality (3.1) becomes (3.8) if $\alpha = 0, \beta = 1$, if G has saltus $(-1)^{j-1}h_j$ at $1 - j/m$, $(j = 1, 2, \cdots, m)$ with $G(1) = 1$, and if $X(1 - j/m) =$

$a_j\,(j = 1, 2, \cdots, m)$.

The one-dimensional case appears in [1], and was proved independently by Olkin [5]. For references to earlier special cases by Szegö, Weinberger, and Bellman, cf. [5].

4. An inequality of Levin and Stečkin.

THEOREM 4.1. *Let I denote an interval in R^k; let X be a nondecreasing map from $[\alpha, \beta)$ into I, continuous from the right. Let H be continuous from the left and of bounded variation on $[\alpha, \beta)$, with $H(\alpha) = 0$. Then,*

$$\int_{[\alpha,\beta)} f[X(t)]\, dH(t) \geqq 0 \tag{4.1}$$

for every continuous function f from I into R with nondecreasing increments, if and only if

$$H(\beta) = 0\ , \tag{4.2}$$

$$\int_{[\alpha,\beta)} H(u)\, dX(u) = 0\ , \tag{4.3}$$

and

$$\int_{[\alpha,t\}} H(u)\, dX(u) \geqq 0 \quad for \quad [\alpha, t\} \subset [\alpha, \beta)\ . \tag{4.4}$$

Proof of necessity. The validity of (4.1) for $f \equiv 1$ and for $f \equiv -1$ implies (4.2). Further, (4.1) for $f(x) \equiv x_j$, where $x = (x_1, \cdots, x_k)$, and for $f(x) \equiv -x_j\,(j = 1, 2, \cdots, k)$, implies $\int_{[\alpha,\beta)} X_j(u)\, dH(u) = 0, j = 1, 2, \cdots, k$, or, equivalently, $\int_{[\alpha,\beta)} H(u)\, dX(u) = 0$, which is (4.3). Inequality (4.4) results from (4.1) after integration by parts, on setting, for fixed $j\ (j = 1, 2, \cdots, k)$ and fixed t, $\alpha \leqq t < \beta$, $f(x) = [X_j(t^+) - x_j]^+$ or $[X_j(t^-) - x_j]^+$.

Proof of sufficiency. Since, as remarked in § 2, f may be approximated uniformly on I by functions with continuous nonnegative second partial derivatives, we may assume that the second partials f_{ij} exist and are continuous and nonnegative. We then have

$$\begin{aligned}\int_{[\alpha,\beta)} f[X(t)]\, dH(t) &= -\int_{[\alpha,\beta)} H(t)\, \nabla f[X(t)]\cdot dX(t) \\ &= -\sum_{j=1}^{k} \int_{[\alpha,\beta)} f_j[X(t)]\, H(t)\, dX_j(t) \\ &= \sum_{j=1}^{k}\sum_{i=1}^{k} \int_{[\alpha,\beta)} f_{ij}[X(t)]\, dX_i(t) \int_{[0,t)} H(u)\, dX_j(u)\ ,\end{aligned}$$

by (4.2) and (4.3). But by (4.4) each term in the last sum is nonnegative, so that (4.1) is verified.

The one-dimensional ($k = 1$) version of Theorem 4.1 appears as Theorem D.1 in [4], and indeed the proof of Theorem 4.1 is the natural extension of the proof given in [4].

Sufficiency in the one-dimensional ($k = 1$) version of Theorem 3.1 was proved in [1] as a consequence of Theorem 249 in [3]; it is exhibited below for continuous X as a consequence also of Levin and Stečkin's Theorem D.1 (Theorem 4.1 above, with $k = 1$). Choose τ so that $X(\tau) = \xi = \int_{[\alpha,\beta)} X(t)\, dG(t)$. Set $H(t) = G(t)$ for $\alpha \leqq t < \tau$, $H(t) = G(t) - 1$ for $\tau \leqq t < \beta$. Then $H(\alpha) = 0$, $H(\beta) = 0$. Also

$$\begin{aligned}\int_{[\alpha,\beta)} H(u)\, dX(u) &= \int_{[\alpha,\beta)} G(u)\, dX(u) - \int_{[\tau,\beta)} dX(u) \\ &= X(\beta) - \xi - [X(\beta) - X(\tau)] = 0\ ;\end{aligned}$$

and

$$\int_{[\alpha,t\}} H(u)\, dX(u) = \int_{[\alpha,t\}} G(u)\, dX(u) \geqq 0$$

if $\alpha \leqq t < \tau$, while

$$\begin{aligned}\int_{[\alpha,t\}} H(u)\, dX(u) &= \int_{[\alpha,t\}} G(u)\, dX(u) - \int_{[\tau,t\}} dX(u) \\ &= \int_{[\alpha,t\}} G(u)\, dX(u) - [X(t^{\pm}) - \xi] \geqq 0\end{aligned}$$

for $\tau \leqq t < \beta$. From (4.1) it then follows that

$$\int_{[\alpha,\beta)} f[X(t)]\, dH(t) = \int_{[\alpha,\beta)} f[X(t)]\, dG(t) - f(\xi) \geqq 0\ ,$$

which is (3.1).

References

1. H. D. Brunk, *On an inequality for convex functions*, Proc. Amer. Math. Soc., **7** (1956), 817-824.
2. Z. Ciesielski, *A Note on Some Inequalities of Jensen's Type*, Annales Polonici Mathematici, **4** (1957-58), 269-274.
3. G. H. Hardy, J. E. Littlewood, G. Polya, *Inequalities*, Cambridge University Press, Second Edition, 1952.
4. V. T. Levin, and S. B. Stečkin, *Inequalities*, Amer. Math. Soc. Translations, vol. 14, pp. 1-29 (from appendices to Russian edition of [**3**]).
5. I. Olkin, *On Inequalities of Szegö and Bellman*, Proc. Nat'l. Acad. of Soc., **45** (1959), 230-231.

UNIVERSITY OF CALIFORNIA, RIVERSIDE
AND UNIVERSITY OF MISSOURI

A RESULT CONCERNING INTEGRAL BINARY QUADRATIC FORMS

WILLIAM EDWARD CHRISTILLES

This paper contains an extension of an earlier work by Dickson ([1], p. 95), in which the following theorem was proven:

THEOREM 1. (Dickson's Theorem). *If a number is represented properly by a form $[a, b, c]$ of discriminant $D = 4ac - b^2$, then any divisor of that number is represented by some form of the same discriminant D.*

DEFINITION. ([1], p. 68). A positive form $[a, b, c]$ is called reduced if $-a < b \leqq a, c \geqq a$, with $b \geqq 0$ if $c = a$.

As a consequence of the above definition it follows that $4a^2 \leqq 4ac = D + b^2 \leqq D + a^2, 3a^2 \leqq D$, and finally $a \leqq \sqrt{(1/3)\,D}$

THEOREM 2. *Let M be properly represented by the integral positve definite quadratic form $a\alpha^2 + b\alpha\gamma + c\gamma^2$ of discriminant $D = 4ac - b^2$. If $M \leqq 3D/16$ and $(D, M) = 1$, then in every factorization of M one of the factors is a_i, one of the minimal values of a primitive quadratic form of discriminant D. In other words, $M = M_1M_2$ where M_1 is a unit or a prime and M_2 is the product of no more than two a_i.*

Proof. According to the remark following the definition $a_i \leqq \sqrt{D/3}$, where equality for a primitive reduced form is possible only if $a_i = b_i = c_i = 1$ and hence $D = 3$ so that the inequality $0 < M \leqq 3D/16$ cannot be satisfied. Thus $a_i < \sqrt{D/3}$.

Now assume $M = r_1r_2$. Then according to Theorm 1 it follows that

$$r_1 = a_i\alpha_i^2 + b_i\alpha_i\gamma_i + c_i\gamma_i^2\,, \quad r_2 = a_j\alpha_j^2 + b_j\alpha_j\gamma_j + c_j\gamma_j^2$$

where the two quadratic forms are primitive reduced forms of discriminant D. Hence

$$\begin{aligned}(4a_ir_1)(4a_jr_2) &= [(2a_i\alpha_i + b_i\gamma_i)^2 + D\gamma_i^2][(2a_j\alpha_j + b_j\gamma_j)^2 + D\gamma_j^2] \\ &= (\beta_i^2 + D\gamma_i^2)(\beta_j^2 + D\gamma_j^2) = 16a_ia_jM \\ &< 16(D/3)M \leqq (16D/3)(3D/16) = D^2\,,\end{aligned}$$

where $\beta_i = (2a_i\alpha_i + b_i\gamma_i)$ and $\beta_j = (2a_j\alpha_j + b_j\gamma_j)$. This implies that $\gamma_i\gamma_j = 0$, say $\gamma_i = 0$, and therefore $r_1 = a_i$.

To prove the final statement of the theorem, assume $M \neq a_i$ and

Received November 21, 1963. The author is indebted to the referee for the suggested revision of both the statement and the proof of Theorem 2.

let r_2 be a minimal factor of M so that $r_2 \neq a_j$. If M_1 is any prime factor of r_2, then $M = M_1M_2$ where $M_2 = (M/r_2)(r_2/M_1) = a_ia_j$.

Reference

1. L. E. Dickson, *Introduction to the Theory of Numbers*, Dover Publications, Inc., New York, 1929.

St. Mary's University, San Antonio, Texas

REFINEMENTS FOR INFINITE DIRECT DECOMPOSITIONS OF ALGEBRAIC SYSTEMS

PETER CRAWLEY AND BJARNI JÓNSSON

Introduction. An operator group with a principal series can obviously be written as a direct product of finitely many directly indecomposable admissible subgroups, and the classical Wedderburn-Remak-Krull-Schmidt Theorem asserts that this representation is unique up to isomorphism. Numerous generalizations of this theorem are known in the literature.[1] Thus it follows from results in Baer [1, 2] that if the admissible center of an operator group G satisfies the minimal and the local maximal conditions, then any two direct decompositions of G (with arbitrarily many factors) have isomorphic refinements. In a different direction, it is shown in Crawley [4] that if an operator group G has a direct decomposition each factor of which has a principal series, then any two direct decompositions of G have isomorphic refinements.

The results of this paper yield sufficient conditions for a group (with or without operators) to have the isomorphic refinement property. For operator groups a common generalization of the theorems mentioned above is obtained: *If an operator group G has a direct decomposition such that the admissible center of each factor satisfies the minimal and local maximal conditions, then any two direct decompositions of G have centrally isomorphic refinements.* For groups without operators we obtain the following result which eliminates any assumption of chain conditions: *If a group G (without operators) has a direct decomposition such that the center of each factor is countable and the reduced part of the center of each factor is a torsion group with primary components of bounded order, then any two direct decompositions of G have centrally isomorphic refinements.*

Actually our results hold for a much wider class of algebraic structures, namely for algebras in the sense of Jónsson-Tarski [6], and it is in this more general framework that the theory is developed. The terminology from general algebra used in this preliminary discussion will be explained in §1.

Our techniques are based on an exchange property defined as follows: An algebra B is said to have the exchange property if, for

Received August 27, 1963. This work was supported in part by NSF Grants G-17957 and G-19673. A summary of the results presented here has appeared in Bull. Amer. Math. Soc 69 (1963), 541-547.

[1] For a fairly complete list of references see Baer [1, 2] or Specht [8], p. 449.

any algebras A, C and $D_i\,(i \in I)$, the condition

$$A = B \times C = \prod_{i \in I} D_i$$

implies that there exist subalgebras $E_i \subseteq D_i\,(i \in I)$ such that

$$A = B \times \prod_{i \in I} E_i\,.$$

The principal result relating this notion to the isomorphic refinement problem is Theorem 7.1, which asserts that if an algebra A is a direct product of subalgebras each of which has the exchange property and has a countable generated center, then any two direct decompositions of A have centrally isomorphic refinements. Two related results are obtained where no cardinality conditions are imposed on the centers, but the decompositions involved are of a more special nature. First (4.2), if $A = B_0 \times B_1 \times B_2 \times \cdots = C_0 \times C_1 \times C_2 \cdots$, with countably many factors, and if all the subalgebras B_i and C_j have the exchange property, then these two direct decompositions have centrally isomorphic refinements. Second (5.3), if A is a direct product of subalgebras each having the exchange property, then any two direct decompositions of A into indecomposable factors are centrally isomorphic.

In §§ 8–11 sufficient conditions are given in order for an algebra B to have the exchange property. In § 8 it is shown that if the center B^c of B has the exchange property, then so does B. There it is also shown that in proving the exchange property for an algebra B we may assume that the factors D_i are isomorphic to subalgebras of B. In § 9 we prove that if B^c satisfies the minimal and local maximal conditions, then B has the exchange property and B^c is countably generated. Sections 10 and 11 are devoted to the study of binary algebras (algebras with just one operation, the binary operation $+$). The main result here (11.5) asserts that if the reduced part of the abelian group B^c is a torsion group all of whose primary components are torsion-complete, then B has the exchange property. In the twelfth and final section some counterexamples and open problems are discussed.

1. Fundamental concepts. Our terminology is largely the same as that in Jónsson-Tarski [6], and it will therefore be described very briefly. By an *algebra* we shall mean a system consisting of a set A, a binary operation $+$ called addition, a distinguished element 0 called the zero element of the algebra, and operations $F_t\,(t \in T)$ each of which is of some finite[2] rank $\rho(t)$, subject only to the following conditions:

[2] In Jónsson-Tarski [6] the operations are not required to be of finite rank. The main reason for this restriction is that it insures that the center of an algebra is a central subalgebra.

(i) A is closed under the operation $+$ and the operations $F_t\,(t \in T)$;
(ii) for all $x \in A$, $x + 0 = 0 + x = x$;
(iii) $F_t(0, 0, \cdots, 0) = 0$ for all $t \in T$.

The set T and the function ρ are assumed to be the same for all the algebras under consideration. We shall identify the algebras with the sets of all their elements, and shall in general use the same symbols, $+$, F_t and 0, to denote the operations and the zero elements of all the algebras. If no auxiliary operations F_t are present, i.e. if $T = \varnothing$, then we refer to A as a *binary algebra*.

An obvious example of an algebra is an operator group, i.e. an algebra for which addition is associative, each element has an additive inverse, and each $F_t\,(t \in T)$ is a unary operation which distributes with respect to $+$. Similarly, an ordinary group without operators is a binary algebra.

If A is an algebra, then the sum of finitely many elements $x_0, x_1, \cdots, x_k, \cdots \in A$ is defined recursively by

$$\sum_{k<0} x_k = 0\,; \qquad \sum_{k<n+1} x_k = \sum_{k<n} x_k + x_n \qquad\qquad (n = 0, 1, \cdots)\,.$$

It is convenient to define also the (un-ordered) sum of certain special systems of elements $x_i \in A\,(i \in I)$. This sum is defined if and only if there exist finitely many distinct elements $i_0, i_1, \cdots, i_{n-1} \in I$ such that $x_i = 0$ whenever $i \in I - \{i_0, i_1, \cdots, i_{n-1}\}$ and such that

$$\sum_{k<n} x_k = \sum_{k<n} x_{i_{\varphi(k)}}$$

for every permutation φ of the integers $0, 1, \cdots, n-1$. Under these conditions we let

$$\sum_{i\in I} x_i = \sum_{k<n} x_{i_k}\,.$$

For brevity, a system of elements $x_i \in A\,(i \in I)$ will be said to be *finitely nonzero* if there are only finitely many indices $i \in I$ such that $x_i \neq 0$.

The notions of subalgebra, homomorphism, isomorphism, and congruence relation are assumed to be known. If θ is a congruence relation over an algebra A, then for $x \in A$ we let x/θ be the congruence class to which x belongs, and for $X \subseteq A$ we let $X/\theta = \{x/\theta \mid x \in X\}$. In particular, A/θ is the quotient algebra of A modulo θ. Observe also that if B is a subalgebra of A, then B/θ is a subalgebra of A/θ. It should be noted that if θ' is the restriction of θ to B, then B/θ and B/θ' are in general distinct algebras although they are isomorphic.

A subalgebra B of an algebra A is called a *subtractive subalgebra* of A if it satisfies the following condition: If $a \in A$ and $b \in B$, and if either $a + b \in B$ or $b + a \in B$, then $a \in B$.

By a *central subalgebra* of A we mean a subalgebra C of A

satisfying the following conditions:

(i) for each $c \in C$ there exists $\bar{c} \in C$ such that $c + \bar{c} = 0$;

(ii) if $c \in C$ and $x, y \in A$, then $x+(y+c) = (x+c)+y = (x+y)+c$;

(iii) if $c \in C, t \in T, k < \rho(t)$, and $x_0, x_1, \cdots, x_{\rho(t)-1} \in A$, then

$$\begin{aligned} F_t(x_0, x_1, \cdots, x_{k-1}, x_k + c, x_{k+1}, \cdots, x_{\rho(t)-1}) \\ = F_t(x_0, x_1, \cdots, x_{k-1}, x_k, x_{k+1}, \cdots, x_{\rho(t)-1}) \\ + F_t(0, 0, \cdots, 0, \underset{k\text{th}}{c}, 0, \cdots, 0) . \end{aligned}$$

It is easy to see that the family of all central subalgebras of an algebra A is a complete sublattice of the lattice of all subalgebras of A. In particular, the union of all the central subalgebras of A is a central subalgebra of A. This largest central subalgebra of A is called the *center* of A, and is denoted by A^c. It is clear that if A is an operator group, then A^c is the usual group-theoretic admissible center of A.[3] For a binary algebra A we can alternatively define the center of A as the set of all those elements of A that have an additive inverse and that commute and associate with all the elements of A. If an algebra A is such that $A^c = A$, then we say that A is *abelian*.

Given two subalgebras B and C of an algebra A, a function f is called a *central isomorphism* of B onto C,—in symbols $f\colon B \cong^c C$,—if f is an isomorphism of B onto C and for each $x \in B$ there exists $c \in A^c$ such that $f(x) = x + c$. We say that B and C are *centrally isomorphic*, —in symbols $B \cong^c C$,—if there exists a central isomorphism of B onto C.

By the *outer direct product*[4] of a system of algebras A_i $(i \in I)$,—in symbols

$$\prod_{i \in I}^{o} A_i ,$$

—we mean the algebra consisting of all functions x such that the domain of x is I, $x(i) \in A_i$ for all $i \in I$, and $x(i) = 0$ for all but finitely many $i \in I$. The operations in this algebra are defined componentwise, $(x + y)(i) = x(i) + y(i)$ and

$$F_t(x_0, x_1, \cdots, x_{\rho(t)-1})(i) = F_t(x_0(i), x_1(i), \cdots, x_{\rho(t)-1}(i)) ,$$

and its zero element is the function that associates with each index i

[3] C.f., Specht [**8**], p. 118; here it is called the Ω-center.

[4] Sometimes the outer direct products are referred to as weak outer direct products, and the Cartesian products (which are used only incidentally in this paper) are called strong outer direct products. In other cases, especially in the theory of abelian groups, outer direct products are called direct sums and Cartesian products are called direct products.

the zero element of the corresponding algebra A_i.

The concept of an algebra is designed to make it possible to introduce the notion of an *inner direct product* of subalgebras of an algebra A, and to reduce the study of (isomorphic) representations of subalgebras B of A as outer direct products to considerations involving this new concept. Since the notions of outer and inner direct products are often confused in the literature, and in other cases the connection between the two concepts is not clearly stated, it is perhaps worthwhile to formulate this relationship in some detail. The basic idea is, of course, that given a representation

$$f\colon B \cong \prod_{i \in I}^{o} C_i ,$$

we can associate with each index $i \in I$ a subalgebra B_i of B that is isomorphic to C_i. By definition, this subalgebra consists of all those elements $x \in B$ such that $f(x)(j) = 0$ for all $j \in I - \{i\}$. If a system of subalgebras $B_i\,(i \in I)$ of B corresponds in this manner to a representation of B as an outer direct product, then we say that B is an inner direct product of the subalgebras $B_i\,(i \in I)$. To complete the transition from outer direct products to inner direct products we must find out to what extent the subalgebras determine the representation, and we must formulate intrinsic necessary and sufficient conditions for B to be an inner direct product of a given system of subalgebras.

The solution of the first problem is easy: two representations,

$$f\colon B \cong C = \prod_{i \in I}^{o} C_i \quad \text{and} \quad f'\colon B \cong C' = \prod_{i \in I}^{o} C_i'$$

yield the same system of subalgebras $B_i\,(i \in I)$ if and only if there exist isomorphisms $g_i\colon C_i \cong C_i'$, for all $i \in I$, such that $f' = gf$ where the isomorphism $g\colon C \cong C'$ is induced by the isomorphisms $g_i\,(i \in I)$ in the sense that $g(x)(i) = g_i(x(i))$ for all $x \in C$ and $i \in I$.

Regarding the second problem, we first observe that B is an inner direct product of subalgebras $B_i\,(i \in I)$ of A if and only if, for every element x of the algebra

$$\bar{B} = \prod_{i \in I}^{o} B_i$$

the sum $\sum_{i \in I} x(i)$ exists, and the mapping $x \to \sum_{i \in I} x(i)$ is an isomorphism of $\bar{B}$ onto B.

Consider now a system of subalgebras $B_i\,(i \in I)$ of A, and define $\bar{B}$ as above. In order for the indicated map to be everywhere defined and to be an isomorphism of $\bar{B}$ into A it is obviously necessary and sufficient that the following four conditions be satisfied:

(I) For any finitely nonzero system of elements $a_i \in B_i\,(i \in I)$, the sum $\sum_{i \in I} a_i$ exists.

(II) For any two finitely nonzero systems of elements $a_i, b_i \in B_i\,(i \in I)$, if $\sum_{i \in I} a_i = \sum_{i \in I} b_i$, then $a_i = b_i$ for all $i \in I$.

(III) For any two finitely nonzero systems of elements $a_i, b_i \in B_i\,(i \in I)$,

$$\sum_{i \in I} (a_i + b_i) = \sum_{i \in I} a_i + \sum_{i \in I} b_i \,.$$

(IV) For any $t \in T$, and for any finitely nonzero systems of elements $a_{k,i} \in B_i\,(i \in I)$, $k = 0, 1, \cdots, \rho(t) - 1$,

$$F_t\Big(\sum_{i \in I} a_{0,i}, \cdots, \sum_{i \in I} a_{\rho(t)-1,i}\Big) = \sum_{i \in I} F_t(a_{0,i}, \cdots, a_{\rho(t)-1,i}) \,.$$

Consequently, in order that there exists a subalgebra B of A such that B is an inner direct product of the algebras $B_i\,(i \in I)$, it is necessary and sufficient that (I)–(IV) hold. Furthermore, if such an algebra B exists, then it is unique and can be characterized by either one of the following conditions:

(V) B is the set of all elements $b \in A$ such that $b = \sum_{i \in I} a_i$ for some finitely nonzero system of elements $a_i \in B_i\,(i \in I)$.

(V′) B is the subalgebra of A generated by the union of all the algebras $B_i\,(i \in I)$.

The conditions (I)–(V) or (I)–(IV) and (V′) are often taken as the definition of the phrase "the subalgebra B of A is the inner direct product of the subalgebra $B_i\,(i \in I)$ of A."

Since we shall henceforth be concerned exclusively with inner direct products we will refer to these simply as *direct products*. The direct product of a system of subalgebras $B_i\,(i \in I)$ of an algebra A will be denoted by

$$\prod_{i \in I} B_i \,,$$

and the direct product of finitely many subalgebras $B_0, B_1, \cdots, B_{n-1}$ will also be written

$$B_0 \times B_1 \times \cdots \times B_{n-1} \,.$$

In the finite case our notion obviously coincides with the direct product in Jónsson-Tarski [6], where this notion is defined recursively in terms of the binary operation $\times$.

A subalgebra C of an algebra B is called a *factor* of B if $B = C \times D$

for some algebra D. B is said to be *indecomposable* if it has at least two elements and the only factors of B are B and $\{0\}$. By a *direct decomposition* or, briefly, a *decomposition* of B we mean a representation of B as a direct product of subalgebras. The direct decompositions of B,

$$B = \prod_{i \in I} C_i = \prod_{j \in J} D_j ,$$

are said to be (*centrally*) *isomorphic* if there exists a one-to-one mapping f of I onto J such that, for each $i \in I$, C_i and $D_{f(i)}$ are (centrally) isomorphic. Finally, the second decomposition is said to be a *refinement* of the first if for each $j \in J$ there exists $i \in I$ such that $D_j \subseteqq C_i$.

2. Elementary properties of direct products.

In this section several simple properties of direct products are listed. Since many of these results are already known from the literature (c.f. Jónsson-Tarski [6]), and the derivations of the remaining ones offer no difficulty, all proofs will be omitted.

We assume throughout this section that A is an algebra.

LEMMA 2.1. *If B and C are subalgebras of A such that $B \times C$ exists, then for all $b, b' \in B$ and $c \in C$,*

$$b + c = c + b \quad \textit{and} \quad (b + b') + c = b + (b' + c) = (b + c) + b' .$$

LEMMA 2.2. *Every factor of A is a subtractive subalgebra of A.*

LEMMA 2.3. (*The modular law*) *Suppose B and C are subalgebras of A such that $B \times C$ exists, and suppose D is a subtractive subalgebra of A. If $B \subseteqq D$, then $(B \times C) \cap D = B \times (C \cap D)$. In particular, if $B \subseteqq D \subseteqq B \times C$, then $D = B \times (C \cap D)$.*

LEMMA 2.4. *If, for each $i \in I$, B_i and B'_i are subalgebras of A such that $B'_i \subseteqq B_i$, and if the direct product*

$$B = \prod_{i \in I} B_i$$

exists, then

(i) *the direct product*

$$B' = \prod_{i \in I} B'_i$$

exists and is a subalgebra of B.

(ii) *$B' = B$ if and only if $B'_i = B_i$ for all $i \in I$.*

(iii) *B' is a subtractive subalgebra of B if and only if, for each $i \in I$, B'_i is a subtractive subalgebra of B_i.*

(iv) *B' is a central subalgebra of A if and only if, for each $i \in I$, B'_i is a central subalgebra of B_i.*

LEMMA 2.5. *Suppose $B_i\,(i \in I)$ are subalgebras of A. Then*

$$A = \prod_{i \in I} B_i$$

if and only if there exist homomorphisms f_i of A onto B_i, for all $i \in I$, such that for each $a \in A$

$$a = \sum_{i \in I} f_i(a)\ , \quad \text{and} \quad f_i f_j(a) = 0 \quad \text{whenever } i, j \in I \ \text{ and } \ i \neq j\ .$$

These homomorphisms f_i, if they exist, are unique and have the property that $f_i f_i = f_i$ for all $i \in I$.

DEFINITION 2.6. *Assuming that*

$$A = \prod_{i \in I} B_i\ ,$$

the homomorphisms f_i characterized by the conditions in Lemma 2.5 are called the projections of A onto the algebras B_i induced by the given decomposition of A.

LEMMA 2.7. *Suppose $B_i\,(i \in I)$ are subalgebras of A. Then the direct product*

$$\prod_{i \in I} B_i$$

exists if and only if for each finite subset J of I the direct product

$$\prod_{i \in J} B_i$$

exists.

LEMMA 2.8. *Suppose that $B_i\,(i \in I)$ are subalgebras of A, that $I = \bigcup_{k \in K} J_k$, and that the sets $J_k\,(k \in K)$ are pairwise disjoint. If either the direct product*

$$B = \prod_{i \in I} B_i$$

exists, or if the direct products

$$C_k = \prod_{i \in J_k} B_i (k \in K) \quad \text{and} \quad B' = \prod_{k \in K} C_k$$

exists, then all these direct products exist, and $B = B'$.

LEMMA 2.9. *Given two direct decompositions of A,*

$$A = \prod_{i \in I} B_i \quad \text{and} \quad A = \prod_{j \in J} C_j \, ,$$

the second decomposition of A is a refinement of the first if and only if for each $i \in I$ there exists a subset J_i of J such that

$$B_i = \prod_{j \in J_i} C_j \, .$$

LEMMA 2.10. *If $B_i (i \in I)$ are subalgebras of A, if the direct product*

$$\prod_{i \in I} B_i$$

exists, and if J and K are subsets of I, then

$$\Big(\prod_{i \in J} B_i\Big) \cap \Big(\prod_{i \in K} B_i\Big) = \prod_{i \in J \cap K} B_i \, .$$

LEMMA 2.11. *Suppose $B_i (i \in I)$ are subalgebras of A, and for each $i \in I$ let $\bar{B}_i$ be the subalgebra of A that is generated by the union of all the algebras B_j with $j \in I$ and $i \neq j$. Then the direct product*

$$\prod_{i \in I} B_i$$

exists if and only if $B_i \times \bar{B}_i$ exists for all $i \in I$.

LEMMA 2.12. *If C is a central subalgebra of A, then for all $a, a' \in A$ and $c \in C$,*

$$a + c = c + a \, , \quad \text{and} \quad a + c = a' + c \quad \text{implies } a = a' \, .$$

LEMMA 2.13. *If C is a central subalgebra of A, then C is a subtractive subalgebra of A, and C is an abelian group under the operation $+$.*

LEMMA 2.14. *If B is a subtractive subalgebra of A, and if C is a central subalgebra of A, then*

(i) *$B \cap C$ is a central subalgebra of A.*

(ii) *$B \times C$ exists if and only if $B \cap C = \{0\}$.*

LEMMA 2.15. *Suppose $C_0, C_1, \cdots, C_{n-1}$ are central subalgebras of A, and for $k = 1, 2, \cdots, n-1$ let $\bar{C}_k$ be the subalgebra of A that is generated by the union of the algebras $C_0, C_1, \cdots, C_{k-1}$. Then the* *direct product*

$$\prod_{k<n} C_k$$

exists if and only if $C_k \cap \bar{C}_k = \{0\}$ *for* $k = 1, 2, \cdots, n-1$.

LEMMA 2.16. *If*

$$A = \prod_{i \in I} B_i ,$$

then

$$A^c = \prod_{i \in I} B_i^c .$$

LEMMA 2.17. *Suppose*

$$A = \prod_{i \in I} B_i = \prod_{j \in J} C_j ,$$

and for $i \in I$ *and* $j \in J$ *let* f_i *and* g_j *be the projections of* A *onto* B_i *and onto* C_j *that are induced by these two decompositions. If* $i, i' \in I$, $j \in J$, *and* $i \neq i'$, *then* $f_i g_j f_{i'}$ *maps* A *into the center of* B_i.

LEMMA 2.18. *If*

$$A = B \times C = \prod_{i \in I} D_i ,$$

then

$$B^c \times C = \prod_{i \in I} ((B^c \times C) \cap D_i) .$$

LEMMA 2.19. *If* B, C *and* D *are subalgebras of* A *such that* $B \times C$ *exists, then the conditions*

$$B \times C = B \times D \quad \text{and} \quad B^c \times C = B^c \times D$$

are equivalent.

LEMMA 2.20. *Suppose* $A = B \times C = B \times D$, *and let* f *and* g *be the projections of* A *onto* C *and onto* D *induced by these two decompositions. Then the restriction* g' *of* g *to* C *is a central isomorphism of* C *onto* D, *and the inverse of* g' *is equal to the restriction of* f *to* D.

3. **Exchange properties.** The central concept of this paper, the exchange property, was mentioned in the introduction. We now formulate this notion more precisely.

DEFINITION 3.1. *Given a cardinal* m, *an algebra* B *is said to have the* m-*exchange property if for any algebra* A *containing* B *as a subalgebra, and for any subalgebras* C *and* $D_i (i \in I)$ *of* A, *where the cardinal of* I *does not exceed* m, *the condition*

$$A = B \times C = \prod_{i \in I} D_i$$

implies that there exist subalgebras $E_i \subseteqq D_i (i \in I)$ *such that*

$$A = B \times \prod_{i \in I} E_i .$$

We say that B *has the exchange property if it has the* m-*exchange property for every cardinal* m. *We say that* B *has the finite exchange property if it has the* m-*exchange property for every finite cardinal* m.

It would be of some interest to know whether, for two given cardinals m and n with $1 < m < n$, the m-exchange property implies the n-exchange property. It will be shown later in this section that this is the case whenever n is finite, whence it follows that the 2-exchange property implies the finite exchange property. In all other cases the answer is unknown. However, since every algebra that is known to have the 2-exchange property is also known to have the exchange property, this question is not crucial at the present.

This section will be devoted to a series of lemmas involving or relating to the exchange properties that will be used in the subsequent sections

DEFINITION 3.2. *A congruence relation* θ *over an algebra* A *is said to be consistent with a decomposition*

$$A = \prod_{i \in I} B_i$$

of A *if, for all* $x, y \in A$ *and* $i \in I$,

$$x \theta y \quad \textit{implies} \ f_i(x) \theta f_i(y) ,$$

where f_i *is the projection of* A *onto* B_i *induced by the given decomposition.*

If A is a group, then the congruence relation θ that corresponds to a normal subgroup N of A is consistent with the above decomposition of A if and only if

$$N = \prod_{i \in I} (B_i \cap N) .$$

For an arbitrary algebra A, a congruence relation θ over A is easily seen to be consistent with a given decomposition of A if and only if θ is generated (in an obvious sense that need not be made more precise here) by its restrictions to the factors in the decomposition.

LEMMA 3.3. *Suppose the congruence relation θ over the algebra A is consistent with the decomposition*

$$A = \prod_{i \in I} B_i$$

of A. Then

$$A/\theta = \prod_{i \in I} (B_i/\theta) .$$

More generally, for any system of subalgebras $B_i' \subseteq B_i$ $(i \in I)$,

$$\left(\prod_{i \in I} B_i'\right)\Big/\theta = \prod_{i \in I} (B_i'/\theta) .$$

Proof. For each $i \in I$ let f_i be the projection of A onto B_i induced by the given decomposition of A. The consistency of θ is equivalent to the assertion that for each $i \in I$ there exists a map g_i of A/θ onto B_i/θ such that $g_i(x/\theta) = f_i(x)/\theta$ for all $x \in A$. It is obvious that g_i is a homomorphism. For each $x \in A$,

$$x = \sum_{i \in I} f_i(x) ,$$

and therefore

$$x/\theta = \sum_{i \in I} (f_i(x)/\theta) = \sum_{i \in I} g_i(x/\theta) .$$

Finally, if i and j are distinct members of I, then for all $x \in A$, $g_i g_j(x/\theta) = f_i f_j(x)/\theta = 0/\theta$. Hence the first part of the conclusion follows by 2.5. The second part of conclusion follows from the first part together with the observation that the algebra

$$\left(\prod_{i \in I} B_i'\right)\Big/\theta$$

consists of all elements

$$\left(\sum_{i \in I} x_i\right)\Big/\theta = \sum_{i \in I} (x_i/\theta) ,$$

associated with finitely non-zero systems $x_i \in B_i'$ $(i \in I)$.

LEMMA 3.4. *Suppose the congruence relation θ over the algebra A is consistent with the decompositions*

$$A = B \times C = \prod_{i \in I} D_i$$

of A, and suppose the restriction of θ to B is the identity relation. If, for each $i \in I$, $\bar{E}_i$ is a subalgebra of D_i/θ, and if

$$A/\theta = B/\theta \times \prod_{i \in I} \bar{E}_i ,$$

then there exist subalgebras $E_i \subseteqq D_i (i \in I)$ such that $\bar{E}_i = E_i/\theta$ for all $i \in I$ and

$$A = B \times \prod_{i \in I} E_i$$

Proof. For each $i \in I$ let f_i be the projection of A onto D_i induced by the second of the two given decompositions of A. Letting

$$A' = B^c \times C , \tag{1}$$

we infer from 2.18 that

$$A' = \prod_{i \in I} D_i' \quad \text{where } D_i' = A' \cap D_i \, (i \in I) . \tag{2}$$

Obviously $(B/\theta)^c = B^c/\theta$, since the restriction of θ to B is the identity relation. It therefore follows by (1), (2), 3.3 and 2.19 that

$$A'/\theta = B^c/\theta \times C/\theta = \prod_{i \in I} (D_i'/\theta) = (B^c/\theta) \times \prod_{i \in I} \bar{E}_i . \tag{3}$$

Next observe that

$$D_i'/\theta = (A'/\theta) \cap (D_i/\theta) . \tag{4}$$

To prove this we use the fact that

$$A/\theta = (B/\theta) \times (C/\theta) = \prod_{i \in I} (D_i/\theta)$$

and that

$$A'/\theta = (B^c/\theta) \times (C/\theta) ,$$

and we infer by 2.18 that

$$A'/\theta = \prod_{i \in I} ((A'/\theta) \cap (D_i/\theta)) . \tag{5}$$

Since in (4) the left hand side is obviously included in the right hand side, the equality follows from (3) and (5) with the aid of 2.4 (ii).

It follows from (3) and (4), together with the hypothesis $\bar{E}_i \subseteqq D_i/\theta$, that

$$\bar{E}_i \subseteqq D_i'/\theta . \tag{6}$$

Letting

$$E_i = \{x \mid x \in D_i' \text{ and } x/\theta \in \bar{E}_i\}\ ,$$

we see that E_i is a subalgebra of D_i', and we infer from (6) that

$$(7) \qquad \bar{E}_i = E_i/\theta\ .$$

From the fact that D_i' is a subtractive subalgebra of A' and that $\bar{E}_i$ is a subtractive subalgebra of A'/θ it readily follows that E_i is a subtractive subalgebra of A'. Consequently,

$$E = \prod_{i \in I} E_i$$

is also a subtractive subalgebra of A'. Furthermore, if $b \in B^c \cap E$, then

$$b/\theta \in (B^c/\theta) \cap (E/\theta) = \{0/\theta\}\ ,$$

and therefore $b = 0$. Thus $B^c \cap E = \{0\}$, and we infer by 2.14 (ii) that the direct product $B^c \times E$ exists, and is a subalgebra of A'.

To complete the proof it suffices to show that $D_k' \subseteq B^c \times E$ for every $k \in I$. Consider an element $x \in D_k'$. By (3) and (7) there exist an element $b \in B^c$ and a finitely nonzero system of elements $e_i \in E_i$ such that

$$x\theta b + \sum_{i \in I} e_i$$

There exists an element $\bar{b} \in B^c$ such that $b + \bar{b} = 0$. Hence

$$\bar{b} + x\theta \sum_{i \in I} e_i\ .$$

Consequently $f_k(\bar{b}) + x\theta e_k$ and $f_i(\bar{b}) = f_i(\bar{b} + x)\theta e_i$ whenever $k \neq i \in I$. Inasmuch as

$$\bar{b} \in B^c \subseteq \prod_{i \in I} D_i'\ ,$$

we infer that $f_k(\bar{b}) + x \in E_k$ and that $f_i(\bar{b}) \in E_i$ whenever $k \neq i \in I$. Thus

$$f_k(b) = b + \sum_{k \neq i \in I} f_i(\bar{b}) \in B^c \times E\ ,$$

and hence

$$x = f_k(b) + (f_k(\bar{b}) + x) \in B^c \times E\ ,$$

as was to be shown.

Lemma 3.5. *If B is a factor of an algebra A, then there exists* *a unique congruence relation* θ *over* A *with the property that if* C *is any subalgebra of* A *with* $A = B \times C$, *and if* g *is the projection of* A *onto* C *induced by this decomposition, then for all* $x, y \in A$ *the conditions* $x\theta y$ *and* $g(x) = g(y)$ *are equivalent.*

Proof. Since the projection g of A onto C induced by the decomposition $A = B \times C$ is a homomorphism of A onto C, the condition

$$x\theta y \text{ if and only if } g(x) = g(y)$$

defines a congruence relation θ over A. To complete the proof it therefore suffices to show that for any other decomposition $A = B \times C'$, and the induced projection g' of A onto C', the conditions $g(x) = g(y)$ and $g'(x) = g'(y)$ are equivalent. To see that this is true we simply observe that for all $x \in A$, $g'(x) = g'g(x)$ and $g(x) = gg'(x)$. In fact, there exists $b \in B$ such that $x = b + g(x)$; hence

$$g'(x) = g'(b) + g'g(x) = g'g(x) .$$

The second formula is proved similarly

DEFINITION 3.6. *If* B *is a factor of an algebra* A, *then the congruence relation* θ *characterized by the conditions in Lemma* 3.5 *is called the congruence relation over* A *induced by* B.

COROLLARY 3.7. *Suppose* B *and* C *are subalgebras of an algebra* A *such that*

$$\text{(i)} \qquad A = B \times C ,$$

and suppose θ *is the congruence relation over* A *induced by* B. *Then* $0/\theta = B$, *and the restriction of* θ *to* C *is the identity relation over* C. *Furthermore,* θ *is consistent with any decomposition of* A *that is a refinement of the decomposition* (i).

LEMMA 3.8. *If* B, C, D_i $(i \in I)$ *and* E *are subalgebras of an algebra* A *such that*

$$\text{(i)} \qquad A = B \times C \times E = \prod_{i \in I} D_i \times E ,$$

and if θ *is the congruence relation over* A *induced by* E, *then for any subalgebras* $F_i \subseteqq D_i$ $(i \in I)$ *the condition*

$$\text{(ii)} \qquad A/\theta = (B/\theta) \times \prod_{i \in I} (F_i/\theta)$$

implies that

(iii) $$A = B \times \prod_{i \in I} F_i \times E .$$

Proof. Since E/θ is the one-element algebra $\{0/\theta\}$, we have

$$A/\theta = (B/\theta) \times \prod_{i \in I} (F_i/\theta) \times (E/\theta) .$$

Inasmuch as the restriction of θ to B is the identity relation over B, we infer by 3.4 that there exist subalgebras $F_i' \subseteqq D_i \,(i \in I)$ and $E' \subseteqq E$ such that

$$A = B \times \prod_{i \in I} F_i' \times E'$$

and such that $F_i/\theta = F_i'/\theta$ for all $i \in I$. Since the restriction of θ to D_i is the identity relation over D_i, this last condition implies that $F_i' = F_i$, and by the modular law we have

$$E = E' \times E'' \quad \text{where } E'' = E \cap \Big(B \times \prod_{i \in I} F_i\Big) .$$

If $x \in E''$, then $x = y + z$ for some $y \in B$ and $z \in \prod_{i \in I} F_i$. Hence $y/\theta + z/\theta = x/\theta = 0/\theta$, and it follows by (ii) that $y/\theta = z/\theta = 0/\theta$. Recalling that the restrictions of θ to B and to $\prod_{i \in I} F_i$ are the identity relations over these algebras we infer that $y = z = 0$, hence $x = 0$. Thus $E'' = \{0\}$, $E' = E$, and (iii) holds.

COROLLARY 3.9. *If $B, C, D_i \,(i \in I)$ and E are subalgebras of an algebra A with*

$$A = B \times C \times E = \prod_{i \in I} D_i \times E ,$$

and if B has the m-exchange property, where m is the cardinal of I, then there exist subalgebras $F_i \subseteqq D_i \,(i \in I)$ such that

$$A = B \times \prod_{i \in I} F_i \times E .$$

LEMMA 3.10. *Suppose m is a cardinal and n is a positive integer, and suppose $B_0, B_1, \cdots, B_n$ are subalgebras of an algebra B with $B = B_0 \times B_1 \times \cdots \times B_n$. Then B has the m-exchange property if and only if each of the algebras $B_k \,(k = 0, 1, \cdots, n)$ has the m-exchange property.*

Proof. It suffices to consider the case $n = 1$. First suppose B_0 and B_1 have the m-exchange property. If A is an algebra that contains B as a subalgebra, if C and $D_i \,(i \in I)$ are subalgebras of A with

(1) $$A = B_0 \times B_1 \times C = \prod_{i \in I} D_i ,$$

and if the cardinal of I does not exceed m, then there exist subalgebras $E_i \subseteqq D_i \, (i \in I)$ such that

$$A = B_0 \times \prod_{i \in I} E_i \, .$$

From this and the first decomposition in (1) it follows by 3.9 that there exist subalgebras $F_i \subseteqq E_i \, (i \in I)$ such that

$$A = B_0 \times B_1 \times \prod_{i \in I} F_i \, .$$

Thus B has the m-exchange property.

Now suppose B has the m-exchange property. Consider an algebra A containing B_0 as a subalgebra, and subalgebras $C, D_i \, (i \in I)$ with

$$A = B_0 \times C = \prod_{i \in I} D_i \, ,$$

and assume that the cardinal of I does not exceed m. Replacing the given algebras, if necessary, by isomorphic copies, we may assume that there exists an algebra A' such that both A and B_1 are subalgebras of A', and such that $A' = A \times B_1$. Then

$$A' = B \times C = B_1 \times \prod_{i \in I} D_i \, .$$

If m is infinite. then we can apply the m-exchange property to these two decompositions, but in order to accommodate also the finite cases we choose an element $k \in I$, and let $I' = I - \{k\}$ and $E = B_1 \times D_k$. Then

$$A' = B \times C = E \times \prod_{i \in I'} D_i \, .$$

Hence there exist subalgebras $E' \subseteqq E$ and $D_i' \subseteqq D_i \, (i \in I')$ such that

$$A' = B \times E' \times \prod_{i \in I'} D_i' \, . \tag{2}$$

Since $B \times E'$ is a factor of A', and hence a subtractive subalgebra of A', and since $B \subseteqq B \times E' \subseteqq B \times D_k$, it follows from the modular law that $B \times E' = B \times D_k'$ where $D_k' = (B \times E') \cap D_k$. Substituting this into (2) we obtain

$$A' = B \times \prod_{i \in I} D_i' \, .$$

Inasmuch as

$$A' = B_1 \times A = B_1 \times \left(B_0 \times \prod_{i \in I} D_i' \right)$$

and

$$B_0 \times \prod_{i \in I} D_k' \subseteq A ,$$

we conclude by 2.4 that

$$A = B_0 \times \prod_{i \in I} D_k' .$$

Thus B_0 has the m-exchange property.

LEMMA 3.11. *If an algebra B has the 2-exchange property, then B has the finite exchange property.*

Proof. It suffices to show, for an arbitrary integer $m > 1$, that if B has the m-exchange property, then B has the $(m+1)$-exchange property. Assuming that

$$A = B \times C = D_0 \times D_1 \times \cdots \times D_m ,$$

let $E = D_0 \times D_1 \times \cdots \times D_{m-1}$. Then $A = B \times C = E \times D_m$, and since B has the 2-exchange property, there exist algebras $E' \subseteq E$ and $D_m' \subseteq D_m$ such that $A = B \times E' \times D_m'$. Letting

$$E'' = E \cap (B \times D_m') \quad \text{and} \quad D_m'' = D_m \cap (B \times E') ,$$

we infer by the modular law that $E = E' \times E''$ and $D_m = D_m' \times D_m''$. From the decompositions

$$A = B \times (E' \times D_m') = (E'' \times D_m'') \times (E' \times D_m')$$

we see by 2.19 that E'' is isomorphic to a factor of B. Consequently E'' has the m-exchange property by 3.10. Since

$$E = E' \times E'' = D_0 \times D_1 \times \cdots \times D_{m-1} ,$$

it therefore follows that there exist subalgebras $D_i' \subseteq D_i$, $i = 0, 1, \cdots, m-1$, such that

$$E = E'' \times D_0' \times D_1' \times \cdots \times D_{m-1}' .$$

Inasmuch as $E'' \subseteq B \times D_m' \subseteq E'' \times (E' \times D_m)$, and application of the modular law yields

$$B \times D_m' = E'' \times E''' \quad \text{where } E''' = (B \times D_m') \cap (E' \times D_m) ,$$

and we conclude that

$$\begin{aligned} A = E' \times E'' \times E''' = E \times E''' &= D_0' \times D_1' \times \cdots \times D_{m-1}' \times E'' \times E''' \\ &= B \times D_0' \times D_1' \times \cdots \times D_m' . \end{aligned}$$

Thus B has the $(m+1)$-exchange property, as was to be shown.

LEMMA 3.12. *Suppose m is a cardinal greater than 1, and suppose B is an algebra whose center is generated by a set whose cardinal does not exceed m. If B has the m-exchange property, then B has the exchange property.*

Proof. Assuming that

$$A = B \times C = \prod_{i \in I} D_i ,$$

write

$$D_J = \prod_{i \in J} D_i \quad \text{for } J \subseteqq I .$$

Then there exists as set $J \subseteqq I$ such that $B^c \subseteqq D_J$, and such that J is finite if m is finite, and the cardinal of J is at most m if m is infinite. By hypothesis (and by 3.11 in case m is finite), there exist subalgebras $E_i \subseteqq D_i$ for all $i \in J$ and a subalgebra F of D_{I-J} such that

$$A = B \times \prod_{i \in J} E_i \times F .$$

Letting $E_i = F \cap D_i$ for $i \in I - J$, we shall show that

$$F = \prod_{i \in I-J} E_i , \tag{1}$$

whence it follows that

$$A = B \times \prod_{i \in I} E_i .$$

Given $a \in F$, there exists a finitely non-zero system of elements $d_i \in D_i (i \in I - J)$ such that

$$a = \sum_{i \in I-J} d_i .$$

Considering a fixed index $k \in I - J$, we can find elements $b \in B$, $e_i \in E_i$ $(i \in J)$ and $f \in F$ such that

$$d_k = b + \sum_{i \in J} e_i + f . \tag{2}$$

By 2.17, $b \in B^c$, hence $b \in D_J$. Consequently the element

$$x = b + \sum_{i \in J} e_i \tag{3}$$

belongs to D_J. But the elements d_k and f belong to the subtractive subalgebra D_{I-J} of A, and it follows by (2) and (3) that $x \in D_{I-J}$. Thus $x = 0$, $d_k = f$, and $d_k \in F \cap D_k = E_k$. Since this last formula holds for all $k \in I - J$, we conclude that

$$a \in \prod_{i \in I-J} E_i .$$

From this (1) easily follows.

4. Direct decompositions with countably many factors.

The next theorem and its simple proof are included primarily in order to show why a similar argument fails to apply when we drop the assumption that the set I be finite.

THEOREM 4.1. *If the algebra A has the m-exchange property (where m is some cardinal), and if*

$$A = \prod_{i \in I} B_i = \prod_{j \in J} C_j$$

where the set I is finite and the cardinal of J does not exceed m, then these two direct decompositions of A have centrally isomorphic refinements.

Proof. For notational convenience we assume that I consists of the integers $0, 1, \cdots, n$. By 3.10, $B_0, B_1, \cdots, B_n$ have the m-exchange property, and by successive applications of 3.9 we obtain, for each $j \in J$, a sequence of subalgebras

$$C_j \supseteq C'_{0,j} \supseteq C'_{1,j} \supseteq \cdots \supseteq C'_{n-1,j} \supseteq C'_{n,j} = \{0\}$$

such that

$$A = B_0 \times \cdots \times B_i \times \prod_{j \in J} C'_{i,j} \qquad (i = 0, 1, \cdots, n) .$$

Since all the subalgebras $C'_{i,j}$ are factors of A, it follows by the modular law that subalgebras $C_{i,j} (i = 0, \cdots, n, j \in J)$ exist such that for each $j \in J$,

$$C_j = C'_{0,j} \times C_{0,j} , \quad \text{and} \quad C'_{i-1,j} = C'_{i,j} \times C_{i,j} \qquad (i = 1, \cdots, n) .$$

Consequently

$$C_j = \prod_{i \leqq n} C_{i,j} \quad (j \in J)$$

and

$$A = \prod_{i<p} B_i \times \prod_{p \leqq i \leqq n} \prod_{j \in J} C_{i,j} \qquad (p = 0, 1, \cdots, n+1) .$$

comparing the two decompositions obtained from this last formula by taking two successive values of p, $p = k$ and $p = k + 1$, we infer by 2.20 that

$$B_k \cong^c \prod_{j \in J} C_{k,j} \quad (k = 0, 1, \cdots, n) ;$$

and we conclude that B_k has a decomposition

$$B_k = \prod_{j \in J} B_{k,j} \text{ with } B_{k,j} \cong^c C_{k,j} \text{ all } j \in J .$$

Attempting to extend the above argument to the case when both I and J are infinite, one encounters difficulty in connection with the "passage through limits." For instance, in the simplest case, where I is the set of all natural numbers, the above process yields subalgebras $C_{i,j}$, $C'_{i,j}$ $(i = 0, 1 \cdots, j \in J)$ with

$$C_j = C'_{k,i} \times \prod_{i \leq k} C_{k,j} \quad \text{and} \quad B_k \cong^c \prod_{j \in J} C_{k,j} ,$$

but it may happen that the direct product

$$\prod_{i < \infty} C_{i,j}$$

is a proper subalgebra of C_j. It is not known how this difficulty can be overcome in general, but we will show how it can be avoided in certain situations. For the case when I and J are denumerable, this is done below by a simple argument involving a diagonal process.

Observe that in the proof of 4.1 we did not make direct use of the fact that A has the m-exchange property, but applied this property to the factors B_i. Because of the finiteness of I this distinction is immaterial here, but in later results a significant generalization is obtained by assuming the exchange properties for the factors in some decomposition (or decompositions) rather than for the whole algebra. Incidentally, 4.1 could actually be generalized by observing that no use is made of the fact that B_n has the m-exchange property.

THEOREM 4.2. *If an algebra A has two direct decompositions with countably many factors,*

(i) $$A = B_0 \times B_1 \times B_2 \times \cdots = C_0 \times C_1 \times C_2 \times \cdots ,$$

where all the factors B_i and C_j have the $\aleph_0$-exchange property, then these two direct decompositions have centrally isomorphic refinements.

Proof. Since B_0 has the $\aleph_0$-exchange property, there exist subalgebras $C_{0,j}$, $C'_{0,j}$ with $C_j = C_{0,j} \times C'_{0,j}$ for $j = 0, 1, 2, \cdots$ such that

(1) $$A = B_0 \times C'_{0,0} \times C'_{0,1} \times C'_{0,2} \times \cdots ,$$

and from this it follows by 2.20 that

(2) $$B_0 \cong^c C_{0,0} \times C_{0,1} \times C_{0,2} \times \cdots .$$

The factor $C'_{0,0}$ of C_0 has the $\aleph_0$-exchange property by 3.10. Applying

3.9 to (1) and the first decomposition in (i) we obtain subalgebras $B_{i,0}, B'_{i,0}$ with $B_i = B_{i,0} \times B'_{i,0}$ for $i = 0, 1, 2, \cdots$ such that

$$A = B_0 \times C'_{0,0} \times B'_{1,0} \times B'_{2,0} \times B'_{3,0} \times \cdots , \tag{3}$$

and it follows, again by 2.20, that

$$C'_{0,0} \cong^c B_{1,0} \times B_{2,0} \times B_{3,0} \times \cdots .$$

Now, using the fact that $B'_{1,0}$ has the $\aleph_0$-exchange property, we apply 3.9 to (3) and (1). This yields subalgebras $C_{1,j}, C'_{1,j}$ with $C'_{0,j} = C_{1,j} \times C'_{1,j}$ for $j = 1, 2, 3, \cdots$ such that

$$\begin{aligned} A &= B_0 \times C'_{0,0} \times B'_{1,0} \times C'_{1,1} \times C'_{1,2} \times C'_{1,3} \times \cdots , \\ B'_{1,0} &\cong^c C_{1,1} \times C_{1,2} \times C_{1,3} \times \cdots . \end{aligned} \tag{4}$$

Next, from (4) and (3) we obtain subalgebras $B_{i,1}, B'_{i,1}$ with $B'_{i,0} = B_{i,1} \times B'_{i,1}$ for $i = 2, 3, \cdots$ such that

$$\begin{aligned} A &= B_0 \times C'_{0,0} \times B'_{1,0} \times C'_{1,1} \times B'_{2,1} \times B'_{3,1} \times B'_{4,1} \times \cdots , \\ C'_{1,1} &\cong^c B_{2,1} \times B_{3,1} \times B_{4,1} \times \cdots . \end{aligned}$$

Continuing in this manner we obtain subalgebras $B_{i,j}, B'_{i,j}$ for $i > j$ and $C_{i,j}$ for $i \leqq j$ such that the following four conditions hold for $i = 1, 2, 3, \cdots$ and $j = 0, 1, 2, \cdots$:

$$B_i = B_{i,0} \times B_{i,1} \times \cdots \times B_{i,i-1} \times B'_{i,i-1} , \tag{5}$$

$$C_j = C_{0,j} \times C_{1,j} \times \cdots \times C_{j,j} \times C'_{j,j} , \tag{6}$$

$$B'_{i,i-1} \cong^c C_{i,i} \times C_{i,i+1} \times C_{i,i+2} \times \cdots , \tag{7}$$

$$C'_{j,j} \cong^c B_{j+1,j} \times B_{j+2,j} \times B_{j+3,j} \times \cdots . \tag{8}$$

From (2), (7) and (8) we infer that there exist algebras $B_{i,j}$ for $i \leqq j$ and $C_{i,j}$ for $i > j$ such that

$$B_{i,j} \cong^c C_{i,j} \quad \text{for } i, j = 0, 1, 2, \cdots , \tag{9}$$

$$B_0 = B_{0,0} \times B_{0,1} \times B_{0,2} \times \cdots , \tag{10}$$

$$B'_{i,i-1} = B_{i,i} \times B_{i,i+1} \times B_{i,i+2} \times \cdots \quad \text{for } i = 1, 2, 3, \cdots , \tag{11}$$

$$C'_{j,j} = C_{j+1,j} \times C_{j+2,j} \times C_{j+3,j} \times \cdots \quad \text{for } j = 0, 1, 2 \cdots . \tag{12}$$

Together with (5) and (6) the last three formulas yield

$$\begin{aligned} B_i &= B_{i,0} \times B_{i,1} \times B_{i,2} \times \cdots , \\ C_j &= C_{0,j} \times C_{1,j} \times C_{2,j} \times \cdots . \end{aligned}$$

Thus the two original decompositions of A have the refinements

$$A = \prod_{i<\infty} \prod_{j<\infty} B_{i,j} = \prod_{i<\infty} \prod_{j<\infty} C_{i,j} ,$$

and according to (9) these are centrally isomorphic.

5. Decompositions into indecomposable factors. In order to prove the existence of centrally isomorphic refinements for two decompositions with countably many factors we had to assume that all the factors involved had the $\aleph_0$-exchange property. In proving that two decompositions with indecomposable factors are centrally isomorphic we can get by with a much weaker assumption. This is due to the next two lemmas.

LEMMA 5.1. *If an indecomposable algebra B has the 2-exchange property, then B has the exchange property.*

Proof. Suppose

$$A = B \times C = \prod_{i \in I} D_i .$$

Since each element of A is contained in the product of finitely many factors D_i, there exists a finite subset J of I such that

$$(1) \qquad B \cap \prod_{i \in J} D_i \neq \{0\} .$$

Letting

$$E = \prod_{i \in I-J} D_i ,$$

we have

$$A = \prod_{i \in J} D_i \times E .$$

By 3.11 B has the finite exchange property, and there therefore exist subalgebras $D_i' \subseteqq D_i \, (i \in J)$ and $E' \subseteqq E$ such that

$$(2) \qquad A = B \times \prod_{i \in J} D_i' \times E'$$

By the modular law we can find subalgebras D_i'' with $D_i = D_i' \times D_i''$ for $i \in J$, and E'' with $E = E' \times E''$. By 2.20,

$$B \cong \prod_{i \in J} D_i'' \times E'' .$$

But as B is indecomposable, only one of the factors in this last product can be different from $\{0\}$. This cannot be the factor E'', for then we would have $D_i' = D_i$ for all $i \in j$, and the product in (2) could not exist because of (1). Thus $E'' = \{0\}$, $E' = E$, and letting $D_i' = D_i$ for all

$i \in I - J$ we have

$$A = B \times \prod_{i \in I} D_i' .$$

LEMMA 5.2. *If an algebra A is a direct product of subalgebras all of which have the 2-exchange property, then every indecomposable factor of A has the exchange property.*

Proof. Suppose

$$A = B \times C = \prod_{i \in I} D_i$$

where B is indecomposable and the algebras D_i have the 2-exchange property. By 5.1 it suffices to show that B has the 2-exchange property. As in the preceding proof, we choose a finite subset J of I with

$$B \cap \prod_{i \in J} D_i \neq \{0\} . \tag{1}$$

By 3.10 the algebra

$$E = \prod_{i \in J} D_i$$

has the 2-exchange property, and there therefore exist subalgebras $B' \subseteq B$ and $C' \subseteq C$ such that

$$A = E \times B' \times C' .$$

By the modular law, B' is a factor of B, and because $B' \cap E = \{0\} \neq B \cap E$, we have $B' \neq B$. Therefore $B' = \{0\}$. Thus $A = B \times C = E \times C'$. Again by the modular law, $C = C' \times (C \cap E')$, and using 2.20 we infer that $E \cong B \times (C \cap E')$. Thus B is isomorphic to a factor of E, and therefore has the 2-exchange property by 3.10.

THEOREM 5.3. *If an algebra A is a direct product of subalgebras all of which have the 2-exchange property, then any two direct decompositions of A into indecomposable factors are centrally isomorphic.*

Proof. Suppose

$$A = \prod_{i \in I} B_i = \prod_{j \in J} C_j$$

where all the factors B_i and C_j are indecomposable and therefore, by 5.2, have the exchange property. For $I' \subseteq I$ and $J' \subseteq J$ let

$$B(I') = \prod_{i \in I'} B_i \quad \text{and} \quad C(J') = \prod_{j \in J'} C_j ,$$

and recall that, by 3.10, $B(I')$ and $C(J')$ have the exchange property whenever the sets I' and J' are finite. In particular, it follows from this and the indecomposability of the factors C_j that if I' is any finite subset of I, then $A = B(I') \times C(J - J')$ for some subset J' of J. Moreover, since $B(I') \cong^c C(J')$, we see with the aid of 4.1 that J' must also be finite and that, in fact, there must exist a one-to-one map φ of I' onto J' such that $B_i \cong^c C_{\varphi(i)}$ for all $i \in I'$. Similarly, for each finite subset J' of J there exists a one-to-one map ψ of J' into I such that $C_j \cong^c B_{\psi(j)}$ whenever $j \in J'$.

For $k \in I$ let

$$I_k = \{i \mid i \in I \textit{ and } B_i \cong^c B_k\} , \qquad J_k = \{j \mid j \in J \textit{ and } C_j \cong^c B_k\} .$$

From the above considerations we see that each member of J must belong to at least one set J_k, and that if J_k is finite, then I_k must have at least as many elements as J_k. To complete the proof it suffices to show that this last statement also holds when J_k is infinite. To prove this we consider, for each $i \in I$, the set N_i of all elements $j \in J$ such that $A = B_i \times C(J - \{j\})$, and show that

$$(1) \qquad N_i \textit{ is finite for each } i \in I ,$$

$$(2) \qquad \bigcup_{i \in I_k} N_i = J_k .$$

From this our assertion follows, for since J_k is assumed to be infinite, (1) and (2) show that the number of elements in J_k cannot exceed the number of distinct sets N_i with $i \in I_k$, and hence cannot be larger than the number of elements in I_k.

Considering a fixed element $i \in I$, choose a finite subset J' of J such that $B_i \cap C(J') \neq \{0\}$. Then the direct product $B_i \times C(J - \{j\})$ fails to exist whenever $j \in J - J'$, and N_i must therefore be a subset of J'. Thus N_i is finite.

Considering a fixed element $j \in J_k$, choose a finite subset I' of I such that $C_j \cap B(I') \neq \{0\}$. Then there exists a finite subset J' of J such that $A = B(I') \times C(J - J')$. Observing that $j \in J'$, let $J'' = J' - \{j\}$ and apply 3.9 to the direct decompositions

$$A = C(J'') \times C_j \times C(J - J') = \prod_{i \in I'} B_i \times C(J - J') .$$

This yields and element $i \in I'$ such that

$$A = C(J'') \times B_i \times C(J - J') = B_i \times C(J - \{j\}) ,$$

and therefore $j \in N_i$. Since $j \in J_k$ and $C_j \cong^c B_i$, we have $i \in I_k$. Thus (2) holds, and the proof is complete.

6. Factors with countably generated centers: Preliminary lemmas. As a result of Lemma 6.3 below the isomorphic refinement problem for algebras

$$A = \prod_{i \in I} B_i \ ,$$

where the factors B_i have countably generated centers, reduces to the special case where I is countable, and A itself therefore has a countably generated center.

LEMMA 6.1. *If B, C and $D_i (i \in I)$ are subalgebras of an algebra A such that*

$$A = B \times C = \prod_{i \in I} D_i \quad \textit{and} \quad B^c = \prod_{i \in I} (B^c \cap D_i) \ ,$$

then there exist subalgebras $E_i (i \in I)$ such that $B^c \cap D_i \subseteq E_i \subseteq D_i$ and

$$A = C \times \prod_{i \in I} E_i \ .$$

Proof. By 2.16,

$$A^c = B^c \times C^c = \prod_{i \in I} D_i^c \ ,$$

and since each $B^c \cap D_i$ is a factor of A^c and a subalgebra of D_i^c, it follows from the modular law that there exist subalgebras D_i' with $D_i^c = (B^c \cap D_i) \times D_i'$ for all $i \in I$. Thus

$$A^c = B^c \times C^c = B^c \times \prod_{i \in I} D_i' \ ,$$

and it follows from 2.19 and 2.18 that

$$B \times C^c = B \times \prod_{i \in I} D_i' = \prod_{i \in I} D_i''$$

where $D_i'' = (B \times C^c) \cap D_i$ for all $i \in I$. Again using the modular law we infer that, for each $i \in I$, $D_i'' = D_i' \times E_i$ where

$$E_i = D_i'' \cap \Big(B \times \prod_{i \neq j \in I} D_j'\Big) \, .$$

Consequently

$$(1) \qquad B \times C^c = \prod_{i \in I} D_i' \times \prod_{i \in I} E_i \ .$$

Observing that

$$B^c \times C^c = B^c \times \Big(\prod_{i \in I} D_i'\Big)^c = \Big(\prod_{i \in I} E_i\Big)^c \times \Big(\prod_{i \in I} D_i'\Big)^c \ ,$$

and that

$$B^c = \prod_{i\in I}(B^c \cap D_i) \subseteqq \prod_{i\in I} E_i^c = \left(\prod_{i\in I} E_i\right)^c ,$$

we see with the aid of 2.4 that

$$B^c = \left(\prod_{i\in I} E_i\right)^c .$$

Consequently

$$C^c \cap \left(\prod_{i\in I} E_i\right) = \{0\} .$$

According to 2.14 this implies that the direct product

$$A' = C^c \times \prod_{i\in I} E_i \tag{2}$$

exists. Furthermore, A' contains $B^c \times C^c$, and therefore contains all the algebras D_i'. Hence it follows by (1) and (2) that $B \times C^c \subseteqq A'$. The opposite inclusion also holds, since all the algebras E_i are contained in $B \times C^c$. Thus $A' = B \times C^c$. Together with (2) and 2.19 this yields the desired conclusion,

$$A = B \times C = C \times \prod_{i\in I} E_i .$$

LEMMA 6.2. *Suppose $B_i\,(i \in I)$, $C_j\,(j \in J)$ and D are subalgebras of an algebra A such that*

$$A = \prod_{i\in I} B_i \times D = \prod_{j\in J} C_j \times D , \tag{i}$$

and suppose B_i^c is countably generated for each $i \in I$. If $k \in I$, then there exist a countable set $K \subseteqq I$ with $k \in K$ and subalgebras $F_j \subseteqq C_j$ $(j \in J)$ such that

$$A = \prod_{i\in I-K} B_i \times \prod_{j\in J} F_j \times D , \tag{ii}$$

$$\prod_{i\in K} B_i^c \times D^c = \prod_{j\in J} F_j^c \times D^c . \tag{iii}$$

Proof. Since B_k^c is countably generated there exist countably generated subalgebras $E_{j,0} \subseteqq C_j^c\,(j \in J)$ such that $E_{j,0} = \{0\}$ for all but countably many $j \in J$ and such that

$$B_k^c \subseteqq \prod_{j\in J} E_{j,0} \times D^c .$$

Since the algebra

$$F_0 = \prod_{j \in J} E_{j,0}$$

is countably generated, there exists a countable subset I_1 of I such that $k \in I_1$ and

$$F_0 \subseteqq \prod_{i \in I_1} B_i^c \times D^c .$$

Again, since the algebra

$$G_1 = \prod_{j \in I_1} B_i^c$$

is countable generated, there exist countably generated subalgebras $E_{j,1} \subseteqq C_j^c \,(j \in J)$ such that $E_{j,0} \subseteqq E_{j,1} \subseteqq C_j$ for all $j \in J$, $E_{j,1} = \{0\}$ for all but countably many $j \in J$, and

$$G_1 \subseteqq \prod_{j \in J} E_{j,1} \times D^c .$$

Continuing in this manner we obtain an ascending sequence of countable sets $I_0 = \{k\} \subseteqq I_1 \subseteqq I_2 \subseteqq \cdots \subseteqq I$ and, for each $j \in J$, an ascending sequence of subalgebras $E_{j,0} \subseteqq E_{j,1} \subseteqq E_{j,2} \subseteqq \cdots \subseteqq C_j^c$ such that

$$\prod_{i \in I_n} B_i^c \subseteqq \prod_{j \in J} E_{j,n} \times D^c \subseteqq \prod_{i \in I_{n+1}} B_i^c \times D^c$$

for $n = 0, 1, 2, \cdots$. Letting $K = I_0 \cup I_1 \cup \cdots$ and $E_j = E_{j,0} \cup E_{j,1} \cup \cdots$ for all $j \in J$ we therefore have

$$\prod_{i \in K} B_i^c \times D^c = \prod_{j \in J} E_j \times D^c . \tag{1}$$

Letting θ be the congruence relation over A induced by D we have

$$A/\theta = \Big(\Big(\prod_{i \in I-K} B_i\Big)\Big/\theta\Big) \times \Big(\Big(\prod_{i \in K} B_i\Big)\Big/\theta\Big) = \prod_{j \in J} (C_j/\theta)$$

by 3.3 and 3.7. Letting

$$\bar{A} = \Big(\Big(\prod_{i \in K} B_i\Big)\Big/\theta\Big)^c$$

we see by (1) that

$$\bar{A} = \prod_{j \in J} (E_j/\theta) ,$$

and it readily follows that $E_j/\theta = \bar{A} \cap (C_j/\theta)$ for all $j \in J$. We therefore infer by 6.1 that there exist subalgebras $F_j \,(j \in J)$ with $E_j \subseteqq F_j \subseteqq C_j$ such that

$$A/\theta = \Big(\Big(\prod_{i \in I-K} B_i\Big)\Big/\theta\Big) \times \prod_{j \in J} (F_j/\theta) ,$$

and we conclude by 3.8 that (ii) holds. Finally, $\dot{E}_j \subseteqq F_j^c$ for all $j \in J$, so that by (1)

$$\prod_{i \in K} B_i^c \times D^c \subseteqq \prod_{j \in J} F_j^c \times D^c .$$

Since, by (i) and (ii),

$$A^c = \prod_{i \in I-K} B_i^c \times \prod_{i \in K} B_i^c \times D^c = \prod_{i \in I-K} B_i^c \times \prod_{j \in J} F_i^c \times D^c ,$$

we conclude with the aid of 2.4 that (iii) holds.

LEMMA 6.3. *If $B_i (i \in I)$ and $C_j (j \in J)$ are subalgebras of an algebra A such that*

$$A = \prod_{i \in I} B_i = \prod_{j \in J} C_j ,$$

and if B_i^c is countably generated for each $i \in I$, then there exist a (possibly transfinite) sequence of countable pairwise disjoint subsets $I_\alpha (\alpha < \lambda)$ of I and subalgebras $C_{j,\alpha} \subseteqq C_j (j \in J, \alpha < \lambda)$ of A such that $I = \bigcup_{\alpha < \lambda} I_\alpha$ and, for all $\beta \leqq \lambda$,

$$A = \prod_{\beta \leqq \alpha < \lambda} \prod_{i \in I_\alpha} B_i \times \prod_{j \in J} \prod_{\alpha < \beta} C_{j,\alpha} .$$

Proof. Letting $U_\beta = \bigcup_{\alpha < \beta} I_\alpha$, we can write this last formula in the form

$$A = \prod_{i \in I - U_\beta} B_i \times \prod_{j \in J} \prod_{\alpha < \beta} C_{j,\alpha} . \tag{1}$$

Since this condition involves only sets I_α and algebras $C_{j,\alpha}$ with $\alpha < \beta$, it can be used as an induction hypothesis. To secure the convergence of our construction process we impose as a second induction hypothesis the condition

$$\prod_{i \in U_\beta} B_i^c \subseteqq \prod_{j \in J} \prod_{\alpha < \beta} C_{j,\alpha} . \tag{2}$$

First observe that this last condition does in fact permit the passage through the limit ordinals. More precisely, suppose η is a limit ordinal, and suppose the sets I_α and algebras $C_{j,\alpha}$ have been chosen for all $\alpha < \eta$ in such a way that (1) and (2) hold for all $\beta < \eta$. We wish to show that in this case (1) and (2) also hold for $\beta = \eta$. From the fact that the condition (2) holds for $\beta < \eta$ it follows that this condition also holds for $\beta = \eta$. Furthermore, since the direct product

$$\prod_{i \in I - U_\eta} B_i \times \prod_{j \in J} \prod_{\alpha < \beta} C_{j,\alpha}$$

exists for all $\beta < \eta$, we readily see that the direct product

$$A' = \prod_{i \in I - U_\eta} B_i \times \prod_{j \notin J} \prod_{\alpha < \eta} C_{j,\alpha}$$

also exists. In order to prove that $A' = A$, and hence that (1) holds for $\beta = \eta$, it suffices to show that $B_h \subseteqq A'$ whenever $h \in U_\eta$. For each such index h there exists an ordinal $\gamma < \eta$ with $h \in U_\gamma$. Using (1) with $\beta = \gamma$, (2) with $\beta = \eta$, and 2.19 we conclude that

$$\begin{aligned} B_h &\subseteqq \prod_{i \in I - U_\gamma} B_i^c \times \prod_{i \in U_\gamma} B_i = \prod_{i \in I - U_\gamma} B_i^c \times \prod_{j \in J} \prod_{\alpha < \gamma} C_{j,\alpha} \\ &\subseteqq \prod_{i \in I - U_\eta} B_i^c \times \prod_{j \in J} \prod_{\alpha < \eta} C_{j,\alpha} \subseteqq A' . \end{aligned}$$

Now consider an arbitrary ordinal η and suppose the sets I_α and algebras $C_{j,\alpha} \subseteqq C_j \, (j \in J)$ have been defined for all $\alpha < \eta$ in such a way that (1) and (2) hold whenever $\beta \leqq \eta$. If $U_\eta = I$, then we take $\lambda = \eta$. Assuming that $U_\eta \neq I$, let

$$\begin{aligned} D_{j,\eta} &= \prod_{\alpha < \eta} C_{j,\alpha} \quad (j \in J) , \\ D_\eta &= \prod_{j \in J} D_{j,\eta} . \end{aligned}$$

For each $j \in J$, $D_{j,\eta}$ is a factor of A and a subalgebra of C_j, hence $C_j = D_{j,\eta} \times C'_{j,\eta}$ for some subalgebra $C'_{j\,\eta}$. It follows that

$$A = \prod_{i \in I - U_\eta} B_i \times D_\eta = \prod_{j \in J} C'_{j\,\eta} \times D_\eta .$$

Choosing $k \in I - U_\eta$ we infer by 6.2 that there exist a countable set I_η with $k \in I_\eta \subseteqq I - U_\eta$ and subalgebras $C_{j,\eta} \subseteqq C'_{j\,\eta} \, (j \in J)$ such that

$$A = \prod_{i \in I - U_{\eta+1}} B_i \times \prod_{j \in J} C_{j,\eta} \times D_\eta = \prod_{i \in I - U_{\eta+1}} B_i \times \prod_{j \in J} \prod_{\alpha < \eta+1} C_{j,\alpha} , \tag{3}$$

$$\prod_{i \in I_\eta} B_i^c \times D_\eta^c = \prod_{j \in J} C_{j,\eta}^c \times D_\eta^c . \tag{4}$$

Here, in accordance with our earlier notation,

$$U_{\eta+1} = \bigcup_{\alpha < \mu+1} I_\alpha = U_\eta \cup I_\eta .$$

By (3), (1) holds for $\beta = \eta + 1$, and from (4) and the fact that (2) holds for $\beta = \eta$ we infer that (2) holds for $\beta = \eta + 1$.

Since all the sets I_α are nonempty, there must exist an ordinal λ such that $U_\lambda = I$, and the corresponding sets I_α and algebras $C_{j,\alpha}$ $(\alpha < \lambda, j \in J)$ clearly have the properties required by the lemma.

7. Factors with countably generated centers: Fundamental theorem. We are now ready to prove the fundamental theorem relating the exchange property to the isomorphic refinement property.

THEOREM 7.1. *If an algebra A is a direct product of subalgebras each of which has the exchange property and has a countably generated center, then any two direct decompositions of A have centrally* *isomorphic refinements.*

Proof. Suppose

$$A = \prod_{i \in I} B_i \tag{1}$$

where, for each $i \in I$, B_i has the exchange property and B_i^c is countably generated. Since every factor of B_i (and hence every algebra isomorphic to such a factor) has the exchange property and has a countably generated center, it is enough to show that the decomposition (1) and any other decomposition

$$A = \prod_{j \in J} C_j \tag{2}$$

have centrally isomorphic refinements.

Consider first the case when I is countable. For convenience suppose I consists of the integers $0, 1, 2, \cdots$. In this case the center of A is generated by a countable set

$$Z = \{a_0, a_1, a_2, \cdots\} .$$

We shall construct an increasing sequence of finite subsets $I_0, I_1, I_2, \cdots$ of I and, for each $j \in J$, two sequences of subalgebras $D_{j,0}, D_{j,1}, D_{j,2}, \cdots$ and $D'_{j\,0} = C_j, D'_{j,1}, D'_{j,2}, \cdots$ such that the following conditions hold for $k = 0, 1, 2, \cdots$:

$$k \in I_k . \tag{3}$$

$$D'_{j,k} = D_{j,k} \times D'_{j\,k+1} \qquad \text{for all } j \in J . \tag{4}$$

$$A = \prod_{i \in I_k} B_i \times \prod_{i \in J} D'_{j,k+1} . \tag{5}$$

$$a_k \in \prod_{j \in J} \prod_{l \leqq k} D_{j,l} . \tag{6}$$

By (2) there exists a finitely nonzero system of elements $c'_{j,0} \in C_j$ $(j \in J)$ such that

$$a_0 = \sum_{j \in J} c'_{j,0} ,$$

and by (1) there exists a finite subset I_0 of I such that $0 \in I_0$ and such that all the elements $c'_{j\,0}$ belong to the algebra

$$B'_0 = \prod_{i \in I_0} B_i .$$

Since B'_0 has the exchange property, there exist subalgebras $D'_{j,1} \subseteqq C_j$ $(j \in J)$ such that (5) holds for $k = 0$, and letting

$$D_{j,0} = C_j \cap \Big(B_0' \times \prod_{j \neq h \in J} D_{h,1}'\Big)$$

for all $j \in J$, we see that (4) and (6) also hold for $k = 0$. In the case of (6) this is true because $c_{j,0}' \in C_j \cap B_0' \subseteq D_{j,0}$ for all $j \in J$.

Now consider an integer $n > 0$, and assume that the finite subsets $I_0 \subseteq I_1 \subseteq \cdots \subseteq I_{n-1}$ of I and the subalgebras $D_{j,0}, D_{j,1}, \cdots, D_{j,n-1}$, $D_{j,0}' = C_j, D_{j,1}', \cdots, D_{j,n}' (j \in J)$ have been so chosen that (3)–(6) hold for $k = 0, 1, \cdots, n-1$. For each $j \in J$ we have have

$$C_j = \prod_{k<n} D_{j,k} \times C_{j,n}' ,$$

and there exist finitely non-zero systems of elements

$$(7) \qquad c_{j,n} \in \prod_{k<n} D_{j,k} \quad \text{and} \quad c_{j,n}' \in D_{j,n}' \quad (j \in J)$$

such that

$$a_n = \sum_{j \in J} c_{j,n} + \sum_{j \in J} c_{j,n}' .$$

There exists a finite subset I_n of I such that $I_{n-1} \subseteq I_n$ and $n \in I_n$, and such that all the elements $c_{j,n}'$ belong to the algebra

$$B_n' = \prod_{i \in J_n} B_i .$$

Since B_n' has the exchange property, and since

$$A = \prod_{i \in I_n} B_i \times \prod_{i \in I - I_n} B_i = \prod_{i \in I_{n-1}} B_i \times \prod_{j \in J} D_{j,n}' ,$$

there exist subalgebras $D_{j,n+1}' \subseteq D_{j,n}' (j \in J)$ such that (5) holds with $k = n$, and letting

$$D_{j,n} = D_{j,n}' \cap \Big(B_n' \times \prod_{j \neq h \in J} D_{h,n+1}'\Big)$$

for all $j \in J$, we see that (4) and (6) also hold for $k = n$. In the case of (6) this is true because of the first formula in (7) and because of the fact that $c_{j,n}' \in D_{j,n}' \cap B_n' \subseteq D_{j,n}$ for all $j \in J$. Thus we see that the sets I_k and algebras $D_{j,k}$ and $D_{j,k}'$ can be so chosen that (3)–(6) hold for $k = 0, 1, 2, \cdots$.

It follows from (4) that the direct products

$$C_j^* = \prod_{k<\infty} D_{j,k} \ (j \in J) \quad \text{and} \quad A^* = \prod_{j \in J} C_j^*$$

exist, and from (6) we infer that $A^c \subseteq A^*$. Moreover, for any natural number n,

$$A = \prod_{j \in J} \prod_{k \leq n} D_{j,k} \times \prod_{j \in J} D_{j,n}' ,$$

and using this together with (3) and (5) we see by 2.19 that

$$B_n \subseteqq \prod_{i \in I_n} B_i \times \Big(\prod_{j \in J} D'_{j,n}\Big)^c = \prod_{j \in J} \prod_{k \leqq n} D'_{j,k} \times \Big(\prod_{j \in J} D'_{j,n}\Big)^c \subseteqq A^* .$$

Consequently $A^* = A$, and we infer by 2.4 that $C_j^* = C_j$ for all $j \in J$.

From (4) and (5) we see that

$$\begin{aligned} A &= \prod_{i \in I_{n-1}} B_i \times \prod_{j \in J} D_{j,n} \times \prod_{j \in J} D'_{j,n+1} \\ &= \prod_{i \in I_{n-1}} B_i \times \prod_{i \in I_n - I_{n-1}} B_i \times \prod_{j \in J} D'_{j,n+1} , \end{aligned}$$

whence it follows that

$$\prod_{i \in I_n - I_{n-1}} B_i \cong^c \prod_{j \in J} D_{j,n} .$$

Consequently, by 4.1, there exist subalgebras $B_{i,j}$ and $C_{i,j}$, ($i \in I_n - I_{n-1}$, $j \in J$) such that

$$\begin{aligned} B_i &= \prod_{j \in J} B_{i,j} && \text{for all } i \in I_n - I_{n-1} , \\ D_{j,n} &= \prod_{i \in I_n - I_{n-1}} C_{i,j} && \text{for all } j \in J , \\ B_{i,j} &\cong^c C_{i,j} && \text{for all } i \in I_n - I_{n-1} \text{ and } j \in J . \end{aligned}$$

Inasmuch as this holds for every natural number n (with $I_{-1} = \varnothing$), we conclude that

$$A = \prod_{i \in I} \prod_{j \in J} B_{i,j} = \prod_{i \in I} \prod_{j \in J} C_{i,j} ,$$

and that these two decompositions of A are centrally isomorphic and are refinements of the decompositions (1) and (2), respectively.

We now drop the assumption that I is denumerable. By 6.3 there exist a sequence of countable, pairwise disjoint subsets I_α $(\alpha < \lambda)$ of I, and for each $j \in J$ a sequence of subalgebras $D_{j,\alpha}$ $(\alpha < \lambda)$ of C_j such that $I = \bigcup_{\alpha<\lambda} I_\alpha$ and

$$A = \prod_{\beta \leqq \alpha < \lambda} \prod_{i \in I_\alpha} B_i \times \prod_{j \in J} \prod_{\alpha < \beta} D_{j,\alpha} \tag{8}$$

for all $\beta \leqq \lambda$. For $\beta = \lambda$ this yields

$$A = \prod_{j \in J} \prod_{\alpha < \lambda} D_{j,\alpha} ,$$

and using 2.4 we infer that

$$C_j = \prod_{\alpha < \lambda} D_{j,\alpha} \qquad \text{for all } j \in J .$$

Taking in (8) two successive values for β, say $\beta = \gamma$ and $\beta = \gamma + 1$, and comparing the resulting formulas, we see that

$$(9) \qquad \prod_{I \in I_\gamma} B_i \cong^c \prod_{j \in J} D_{j,\gamma} .$$

Since I_γ is countable, it follows from the first part of the proof that the two decompositions in (9) have centrally isomorphic refinements, and inasmuch as this holds for every $\gamma < \lambda$, we conclude that the decompositions (1) and (2) have centrally isomorphic refinements.

The preceding theorem can also be stated in the following, apparently more general, form.

THEOREM 7.2. *If an algebra A has two direct decompositions,*

$$A = \prod_{i \in I} B_i = \prod_{j \in J} C_j ,$$

such that each of the factors B_i ($i \in I$) has a countably generated center and each of the factors C_j ($j \in J$) has the $\aleph_0$-exchange property, then any two direct decompositions of A have centrally isomorphic refinements.

Proof. Choosing the ordinal λ, subsets I_α ($\alpha < \lambda$) of I, and subalgebras $C_{j,\alpha}$ ($j \in J$, $\alpha < \lambda$) according to 6.3, we have

$$(1) \qquad C_j = \prod_{\alpha < \lambda} C_{j,\alpha} \qquad \text{for each } j \in J ,$$

$$(2) \qquad \prod_{i \in I_\alpha} B_i \cong^c \prod_{j \in J} C_{j,\alpha} \qquad \text{for each } \alpha < \lambda .$$

Since, by hypothesis, each of the sets I_α is countable, the first direct product in (2) has a countably generated center, and hence so does the second product. Consequently each of the factors $C_{j,\alpha}$ has a countably generated center. Furthermore, by (1) and 3.10, each of the algebras $C_{j,\alpha}$ has the $\aleph_0$-exchange property. Hence, by 3.12, all the algebras $C_{j,\alpha}$ have the exchange property. Since

$$A = \prod_{j \in J} \prod_{\alpha < \lambda} C_{j,\alpha} ,$$

the conclusion now follows from 7.1.

8. Sufficient conditions for an algebra to have the $\mathfrak{m}$-exchange property. So far we have been primarily concerned with consequences of the exchange property, but in the remainder of this paper we shall investigate conditions that imply that a given algebra has the exchange property. In the present section it will be shown that this problem reduces to considerations that involve only abelian algebras.

THEOREM 8.1. *For any cardinal $\mathfrak{m}$, if the center of an algebra B has the $\mathfrak{m}$-exchange property, then B has the $\mathfrak{m}$-exchange property.*

Proof. Suppose

$$A = B \times C = \prod_{i \in I} D_i ,$$

where the cardinal of I is at most m. Then by 2.18,

$$B^c \times C = \prod_{i \in I} D_i'$$

where $D_i' = (B^c \times C) \cap D_i$ for each $i \in I$. Hence there exist subalgebras $E_i \subseteqq D_i' \, (i \in I)$ such that

$$B^c \times C = B^c \times \prod_{i \in I} E_i ,$$

and we conclude by 2.19 that

$$A = B \times \prod_{i \in I} E_i .$$

THEOREM 8.2. *For any cardinal m, in order for an algebra B to have the m-exchange property it is sufficient (and obviously necessary) that the following condition be satisfied: For any algebra A containing B as a factor, and for any subalgebras C and $D_i \, (i \in I)$ of A, if*

$$A = B \times C = \prod_{i \in I} D_i ,$$

if the cardinal of I does not exceed m, and if each of the algebras $D_i \, (i \in I)$ is isomorphic to a subalgebra of B, then there exist subalgebras $E_i \subseteqq D_i \, (i \in I)$ such that

$$A = B \times \prod_{i \in I} E_i .$$

Proof. Assume that the above condition is satisfied. Suppose

(i) $$A = B \times C = \prod_{i \in I} D_i ,$$

where the cardinal of I does not exceed m. Let f and g be the projections of A onto B and C induced by the first decomposition of A, and for $i \in I$ let h_i be the projection of A onto D_i induced by the second decomposition.

Let θ be the congruence relation over A defined by the condition that, for all $x, y \in A$,

$$x\theta y \text{ if and only if } fh_i(x) = fh_i(y) \text{ whenever } i \in I .$$

We shall show that

(1) θ is consistent with the decompositions (i) of A .

(2) The restriction of θ to B is the identity relation over B.

Suppose $x, y \in A$ and $x\theta y$. Then

$$\begin{aligned} f(x) &= f\Big(\sum_{i\in I} h_i(x)\Big) = \sum_{i\in I} fh_i(x) \\ &= \sum_{i\in I} fh_i(y) = f\Big(\sum_{i\in I} h_i(y)\Big) = f(y)\,. \end{aligned}$$

In particular $f(x)\theta f(y)$. Moreover, this shows that for $x, y \in B$ the condition $x\theta y$ implies that $x = f(x) = f(y) = y$, so that (2) holds. Again assuming that $x\theta y$, if $k \in I$ then

$$\begin{aligned} &fh_ih_k(x) = 0 = fh_ih_k(y) \quad \text{whenever } k \neq i \in I\,, \\ &fh_kh_k(x) = fh_k(x) = fh_k(y) = fh_kh_k(y)\,, \end{aligned}$$

so that $h_k(x)\theta h_k(y)$. From the equations

$$\begin{aligned} fh_kf(x) + fh_kg(x) &= fh_k(f(x) + g(x)) = fh_k(x) = fh_k(y) \\ &= fh_k(f(y) + g(y)) = fh_kf(y) + fh_kg(y) \end{aligned}$$

we infer that

$$(3) \qquad h_ifh_kf(x) + h_ifh_kg(x) = h_ifh_kf(y) + h_ifh_kg(y)$$

for all $i, k \in I$. Since $f(x) = f(y)$, we have

$$h_ifh_kf(x) = h_ifh_kf(y)$$

for all $i, k \in I$, and if $i \neq k$, then this element belongs to A^c. Therefore, by (3),

$$(4) \qquad h_ifh_kg(x) = h_ifh_kg(y) \quad \text{whenever } i, k \in I \text{ and } i \neq k\,.$$

Considering now a fixed index $i \in I$, observe that

$$\sum_{k\in I} h_ifh_kg(x) = h_if\Big(\sum_{k\in I} h_kg(x)\Big) = h_ifg(x) = h_i(0) = 0\,,$$

with the corresponding formula holding with x replaced by y. Hence, in particular,

$$\sum_{k\in I} h_ifh_kg(x) = \sum_{k\in I} h_ifh_kg(y)\,.$$

Furthermore, all the summands in these two sums belong to A^c because $fh_kg(x)$ and $fh_kg(y)$ always belong to A^c. Since, by (4),

$$\sum_{i\neq k\in I} h_ifh_kg(x) = \sum_{i\neq k\in I} h_ifh_kg(y)\,,$$

this implies that

$$h_i f h_i g(x) = h_i f h_i g(y) .$$

Thus in (4) we can omit the condition that $i \neq k$, and we conclude that, for all $k \in I$,

$$f h_k g(x) = \sum_{i \in I} h_i f h_k g(x) = \sum_{i \in I} h_i f h_k g(y) = f h_k g(y) ,$$

so that $g(x)\theta g(y)$. This completes the proof of (1).

From (1) it follows that

$$A/\theta = (B/\theta) \times (C/\theta) = \prod_{i \in I} (D_i/\theta) .$$

Notice that if $k \in I$ and $x, y \in D_k$, then the conditions $x\theta y$ and $f(x) = f(y)$ are equivalent, and therefore the mapping

$$x/\theta \rightarrow f(x) \quad (x \in D_k)$$

is an isomorphism of D_k/θ into B. Since $B \cong B/\theta$, it follows that there exist subalgebras $\bar{E}_i \subseteqq D_i/\theta \, (i \in I)$ such that

$$A/\theta = (B/\theta) \times \prod_{i \in I} \bar{E}_i .$$

Consequently, by 3.4 there exist subalgebras $E_i \subseteqq D_i \, (i \in I)$ such that

$$A = B \times \prod_{i \in I} E_i .$$

Because of 8.1, we may apply the criterion in 8.2 to B^c in place of B, and thus consider decompositions

$$A = B^c \times C = \prod_{i \in I} D_i$$

where the algebras D_i are isomorphic to subalgebras of B^c. However, the algebras D_i need not be central subalgebras of A, and A therefore is not necessarily abelian. We shall now show that it is actually sufficient to consider the case when A is abelian, in which case the factors C and $D_i \, (i \in I)$ of A are of course also abelian.

THEOREM 8.3. *For any cardinal m, in order for an algebra B to have the m-exchange property it is sufficient that the following condition be satisfied: For any abelian algebra A containing B^c as a factor, and for any subalgebras C and $D_i \, (i \in I)$ of A, if*

(i) $$A = B^c \times C = \prod_{i \in I} D_i ,$$

if the cardinal of I does not exceed m, and if each of the algebras $D_i \, (i \in I)$ is isomorphic to a subalgebra of B^c, then there exist subalgebras $E_i \subseteqq D_i \, (i \in I)$ such that

(ii) $$A = B^c \times \prod_{i \in I} E_i \,.$$

Proof. By 8.1 it suffices to show that B^c has the m-exchange property, and by 8.2 it is therefore enough to show that the condition in our theorem implies the property obtained from it by deleting the word "abelian." Assume therefore that (i) holds, that the cardinal of I does not exceed m, and that each of the algebras $D_i \,(i \in I)$ is isomorphic to a subalgebra of B^c. Under the operation $+$ each of the algebras D_i is therefore a commutative cancellation semigroup, and hence so is A. Consequently A can be embedded in an Abelian group $\bar{A}$ in such a way that each element of $\bar{A}$ is the difference of two elements of A. This extension of A is unique up to isomorphism. Furthermore, there is a unique way of extending the operations $F_t \,(t \in T)$ to $\bar{A}$ in such a way that the resulting algebra is abelian: If $a_k = a'_k - a''_k$ with $a'_k, a''_k \in A$ for $k = 0, 1, \cdots, \rho(t) - 1$, then we let

$$F_t(a_0, a_1, \cdots, a_{\rho(t)-1}) = F_t(a'_0, a'_i, \cdots, a'_{\rho(t)-1}) - F_t(a''_0, a''_1, \cdots, a''_{\rho(t)-1}) \,.$$

That this definition is unambiguous and actually does yield an abelian algebra is an easy consequence of the fact that the equation

$$\begin{aligned} &F_t(x_0 + y_0, x_1 + y_1, \cdots, x_{\rho(t)-1} + y_{\rho(t)+1}) \\ &\qquad = F_t(x_0, x_1, \cdots, x_{\rho(t)-1}) + F_t(y_0, y_1, \cdots, y_{\rho(t)-1}) \end{aligned}$$

holds whenever the elements $x_k, y_k \,(k = 0, 1, \cdots, \rho(t) - 1)$ belong to A.

For any subalgebra X of A let $\bar{X}$ be the smallest abelian subalgebra of $\bar{A}$ that contains X. Then $\bar{X}$ consists of all elements of the form $x - x'$ with $x, x' \in X$. It is easy to check the condition

$$A = \prod_{j \in J} X_j$$

implies that

$$\bar{A} = \prod_{j \in J} \bar{X}_j \,.$$

In particular, since $\bar{B}^c = B^c$,

$$\bar{A} = B^c \times \bar{C} = \prod_{i \in I} \bar{D}_i \,.$$

For each $i \in I$, D_i is isomorphic to a subalgebra of B^c, and the same is therefore true of $\bar{D}_i$. Hence, by hypothesis, there exist subalgebras $F_i \subseteqq \bar{D}_i \,(i \in I)$ such that

$$\bar{A} = B^c \times \prod_{i \in I} F_i \,.$$

Given an element $a \in A$, there exist an element $b \in B^c$ and a finitely nonzero system of elements $f_i \in F_i \, (i \in I)$ such that

$$a = b + \sum_{i \in I} f_i \, .$$

Since $-b \in A$, the element

$$a - b = \sum_{i \in I} f_i$$

belongs to A, and there exists a finitely nonzero system of elements $d_i \in D_i \, (i \in I)$ such that

$$a - b = \sum_{i \in I} d_i \, .$$

Inasmuch as $d_i, f_i \in \bar{D}_i$ for all $i \in I$, we infer that $d_i = f_i \in D_i \cap F_i$ for all $i \in I$, and therefore

$$a = b + \sum_{i \in I} d_i \in B^c \times \prod_{i \in I} (D_i \cap F_i) \, .$$

It is now easy to show that (ii) holds with $E_i = D_i \cap F_i$ for all $i \in I$.

9. Factors with central chain conditions. In this section we will show that algebras satisfying certain central chain conditions have the exchange property and have countably generated centers, and these results will be applied to obtain the principal isomorphic refinement theorem for general algebras. The chain conditions involved are made precise in the following two definitions.

DEFINITION 9.1. *An algebra A is said to satisfy the minimal condition if every nonempty family of subtractive subalgebras of A has a minimal member. Similarly, A satisfies the maximal condition if every nonempty family of subtractive subalgebras has a maximal member.*

DEFINITION 9.2. *An algebra A is said to satisfy the local maximal condition if every finitely generated subtractive subalgebra of A satisfies the maximal condition.*

It should be noted that the minimal and (local) maximal conditions as defined above involve only subtractive subalgebras of an algebra A. In particular, since the subtractive subalgebras of an operator group are precisely its admissible subgroups, for groups the minimal and maximal conditions as defined in 9.1 and 9.2 are just the usual group-theoretic chain conditions.

The first theorem of this section makes use of the following lemma which is a consequence of the results of Baer [1].

LEMMA 9.3. ([1]; Theorem D p. 96 and Theorem 3 p. 93)[5] *Let G be an operator group which satisfies the minimal and local maximal conditions. If $G = B \times C = D \times E$ where B is indecomposable, then there exist factors $D' \subseteqq D$ and $E' \subseteqq E$ such that $G = B \times D' \times E'$.*

Suppose now that A is an abelian algebra with auxiliary operations $F_t\,(t \in T)$. For each $t \in T$ and each $k < \rho(t)$ define the unary operation $F_{k,t}$ by

$$F_{k,t}(a) = F_t(0, \cdots, 0, \underset{k\text{th}}{a}, 0, \cdots, 0) \qquad \text{for all } a \in A\ .$$

Since A is abelian, it follows that for each $t \in T$ and elements $a_0, \cdots, a_{\rho(t)-1} \in A$ we have

$$F_t(a_0, \cdots, a_{\rho(t)-1}) = \sum_{k<\rho(t)} F_{k,t}(a_k)\ .$$

Consequently the (subtractive) subalgebras of A and the direct decompositions of A remain unchanged if we replace the operations $F_t\,(t \in T)$ by the operations $F_{k,t}\,(k < \rho(t), t \in T)$. Moreover, this new system so obtained is obviously an abelian operator group. Hence the following lemma is immediate by 9.3.

LEMMA 9.4. *If A is an abelian algebra which satisfies the minimal condition and the local maximal condition, and if $A = B \times C = D \times E$ where B is indecomposable, then there exist factors $D' \subseteqq D$ and $E' \subseteqq E$ such that $A = B \times D' \times E'$.*

THEOREM 9.5. *If the center B^c of an algebra B satisfies the minimal condition and the local maximal condition, then B has the exchange property.*

Proof. By 8.1 we may assume that $B = B^c$. Since B satisfies the minimal condition, it is a direct product of finitely many indecomposable subalgebras, and therefore by 3.10 and 5.1 it is sufficient to show that B has the 2-exchange property.

Consider an abelian algebra A containing B as a subalgebra, and algebras C, D_0 and D_1 such that D_0 and D_1 are isomorphic to subalgebras of B and such that $A = B \times C = D_0 \times D_1$. Then both D_0 and D_1 satisfy the minimal and local maximal conditions, and it readily follows that the same is true of A. Therefore by 9.4 there exist subalgebras $E_0 \subseteqq D_0$ and $E_1 \subseteqq D_1$ such that $A = B \times E_0 \times E_1$, and we conclude by 8.3 that B has the exchange property.

[5] See also Specht [8], pp. 250, 259 and 260.

In order to apply the preceding theorem in conjunction with 7.1, we must further show that under the given hypothesis B^c is countably generated. This observation is based on the following lattice-theoretic lemma. The terminology and simple facts from lattice theory used below can be found in Birkhoff [3].

LEMMA 9.6. *If L is an upper continuous modular lattice, if every decreasing sequence of elements of L is countable, and if every element of L is a join of finite dimensional elements, then every element of L is a join of countably many finite dimensional elements.*

Proof. First consider an element $a \in L$ that is a join of atoms. Then there exists an independence sequence $p_0, p_1, \cdots, p_\xi, \cdots (\xi < \lambda)$ of atoms of L such that

$$a = \sum_{\xi<\lambda} p_\xi .$$

Since the elements

$$\sum_{\eta \leqq \xi < \lambda} p_\xi \quad (\eta < \lambda)$$

form a strictly decreasing sequence, λ must be countable, and therefore a is the join of countably many atoms.

Now consider an arbitrary element $a \in L$. For each $n = 1, 2, \cdots$ let P_n be the set of all the elements $x \in L$ with $x \leqq a$ whose dimension does not exceed n, and let $a_n = \sum P_n$. Then

$$a = \sum_{n<\infty} a_n .$$

By the first part of the proof there is a countable set $Q_1 \subseteqq P_1$ such that $a_1 = \sum Q_1$. Suppose $n > 1$ and $x \in P_n$. Then either $x \leqq a_{n-1}$ or $x + a_{n-1}$ covers a_{n-1}, since each member of $P_n - P_{n-1}$ covers at least one member of P_{n-1}. Consequently a_n is the join of atoms in the quotient sublattice a/a_{n-1}. Since the hypothesis of the lemma is satisfied with L replaced by this sublattice, we again use the first part of the proof to infer that

$$a_n = a_{n-1} + \sum Q_n$$

where Q_n is a countable subset of $P_n - P_{n-1}$. It follows that each a_n is a join of countably many finite dimensional elements, and therefore a is also a join of countably many such element.

COROLLARY 9.7. *If B is an abelian algebra that satisfies the minimal condition and the local maximal condition, then B is countably generated.*

Proof. The lattice L of all subtractive subalgebras of B is modular and upper continuous, and, by hypothesis, every decreasing sequence of elements of L is finite. Also, if C is a finitely generated subtractive subalgebra of B, then the lattice $L(C)$ of all subtractive subalgebras of C satisfies the double chain condition. Consequently $L(C)$ is finite dimensional, i.e., C is a finite dimensional element of L. Since every subtractive subalgebra of B is the lattice join of finitely generated subtractive subalgebras, L satisfies the hypothesis of 9.6. Hence B is the lattice join of countably many finite dimensional elements of L; equivalently, B is generated by the set-union of countably many subtractive subalgebras C such that $L(C)$ is finite dimensional. But if $L(C)$ is finite dimensional, then C is clearly finitely generated. Thus it follows that B is countably generated.

Combining 9.5, 9.7 and 7.1 we obtain our principal isomorphic refinement theorem for algebras with auxiliary operations.

THEOREM 9.8. *If an algebra A has a direct decomposition*

$$A = \prod_{i \in I} B_i$$

such that, for each $i \in I$, B_i^c satisfies the minimal condition and the local maximal condition, then any two direct decompositions of A have centrally isomorphic refinements.

10. Lemmas on abelian groups.

When applied to algebras without auxiliary operations F_t, Theorem 9.8 can be stated in the following equivalent form: *If a binary algebra A is a direct product of subalgebras $B_i\,(i \in I)$ such that, for each $i \in I$, B_i^c is a direct product of finitely many primary cyclic and quasi-cyclic groups, then any two direct decompositions of A have centrally isomorphic refinements.* For every abelian group satisfied the local maximal condition, and the condition imposed on the abelian groups B_i^c above is equivalent to the assertion that they satisfy the minimal condition. In the next section we shall obtain a result that is considerably more general than the one stated above. Here we list a number of known results and prove five lemmas concerning abelian groups that will be used in the proof of this more general theorem.

If G is an abelian group and n is an integer, then the subgroups nG and $G[n]$ are defined by

$$nG = \{nx \mid x \in G\}\,,$$
$$G[n] = \{x \mid x \in G \text{ and } nx = 0\}\,.$$

As usual, we say that an abelian group G is *divisible* if $nG = G$ for every integer $n \neq 0$, and we say that G is *of bounded order* if there exists an integer $n \neq 0$ such that $nG = \{0\}$. An abelian group is said to be *reduced* if it has no nonzero divisible subgroup, and by the *reduced part* of an abelian group G we mean the quotient group G/D where D is the maximal divisible subgroup of G. If X is a subset of a group G, then $[X]$ denotes the subgroup of G generated by X; in particular, if $x \in G$, then the cyclic subgroup of G generated by x is denoted by $[x]$.

Let G be an abelian p-group (p some prime). By the *height* of an element $x \in G$ we mean the largest integer r such that $x \in p^r G$, if a largest such integer r exists, otherwise the height of x is ∞. Thus height $x = \infty$ if $x \in p^n G$ for $n = 1, 2, \cdots$, and height $x = r < \infty$ if $x \in p^r G$ but $x \notin p^{r+1}G$. Obviously the zero element of G has infinite height; if this is the only element in G of infinite height, then we say that G has *no elements of infinite height*. Thus G has no elements of infinite height if and only if $\bigcap_{n<\infty} p^n G = \{0\}$.

If G is an abelian p-group with no elements of infinite height, then a topology can be introduced in G by taking as a neighborhood basis for 0 the subgroup $p^n G\,(n = 1, 2, \cdots)$. This topology is called the *p-adic topology* of G. G can be completed in its p-adic topology, and the torsion subgroup $\widetilde{G}$ of the topological completion of G is also an abelian p-group without elements of infinite height.[6]

An abelian p-group G is said to be *torsion-complete* if G has no elements of infinite height, and G is equal to the torsion subgroup of the topological completion of G, $G = \widetilde{G}$. Alternatively, G is torsion-complete if and only if G has no elements of infinite height, and every Cauchy sequence $\{x_k\}_{k<\infty}$ of G, for which the orders of the elements x_k are bounded, converges to a limit in G.[7] For convenience we will call a Cauchy sequence $\{x_k\}_{k<\infty}$, for which the orders of the x_k are bounded, a *bounded Cauchy sequence*.

An explicit representation of torsion-complete abelian p-groups can be given as follows. Let $U_1, U_2, U_3, \cdots$ be a sequence of p-groups such that U_n is a direct product of cyclic groups of order p^n for each $n = 1, 2, \cdots$. Let Γ be the *Cartesian product* of the groups $U_1, U_2, U_3, \cdots$, that is, Γ is the set of *all* functions f defined on the positive integers such that $f(n) \in U_n$, with addition defined component-wise. Then the torsion subgroup of Γ is torsion-complete. Conversely, if G is a torsion-complete abelian p-group, then there exists a sequence of

[6] This is essentially given by Kaplansky [**7**], p. 50.

[7] Fuchs [**5**], p. 114, calls these groups *closed*. However, we have adopted the terminology of Kaplansky [**7**], p. 54, in order to remain consistent with topological terminology. Fuchs' definition of *Cauchy sequence* also differs somewhat from ours in that he requires a Cauchy sequence to be bounded and converge at a specified rate. Again we have followed Kaplansky [**7**] in using the usual topological concept of Cauchy Sequence.

groups $U_1, U_2, U_3, \cdots$, where U_n is a direct product of cyclic groups of order p^n for each $n = 1, 2, \cdots$, such that G is isomorphic to the torsion subgroup of the Cartesian product of $U_1, U_2, U_3, \cdots$.[8] In particular, every primary abelian group of bounded order is torsion-complete, and every countable torsion-complete primary abelian group is necessarily of bounded order.

By a *pure subgroup* of an abelian p-group G we mean a subgroup S of G such that $S \cap p^nG = p^nS$ for all $n = 1, 2, \cdots$. It is easily seen that the p-adic topology of a pure subgroup of an abelian p-group G with no elements of infinite height is the same as the topology induced by the p-adic topology of G. A subgroup U of a p-group G is called a *basic subgroup* if U has the following properties:

(i) U is a direct product of cyclic groups;

(ii) U is a pure subgroup of G;

(iii) the quotient group G/U is divisible. A subset $X \subseteq G$ is *independent* if the subgroup $[X]$ generated by X is the direct product of the cyclic subgroups $[x]$ generated by the elements $x \in X$. If in addition, $[X]$ is a pure subgroup of G, then X is called a *pure independent* subset.

The following ten lemmas are well known; proofs and references to the original sources can be found in Fuchs [5] as indicated in each case.

LEMMA 10.1. ([5], p. 62) *If a subgroup S of an abelian group G is divisible, then S is a factor of G.*

LEMMA 10.2. ([5], p. 64) *A divisible abelian group is a direct product of subgroups each of which is isomorphic to either the additive group of rationals or a primary quasi-cyclic group.*

LEMMA 10.3. ([5], p. 78) *If S is a subgroup of an abelian p-group G, and if every element of $S[p]$ has the same height in S as it does in G, i.e., if $S[p] \cap p^nG = S[p] \cap p^nS$ $(n = 1, 2, \cdots)$, then S is a pure subgroup of G.*

LEMMA 10.4. ([5], p. 78) *If S is a pure subgroup of an abelian p-group G, and if $S[p] = G[p]$, then $S = G$.*

LEMMA 10.5. ([5], p. 97) *A subgroup U of a primary abelian group G is a basic subgroup if and only if U is generated by a maximal pure independent subset of G.*

LEMMA 10.6. ([5], pp. 98 and 104) *A primary abelian group G*

[8] Fuchs [5], p. 114.

has at least one basic subgroup, and all the basic subgroups of G are isomorphic.

LEMMA 10.7. ([5], p. 104) *If a primary abelian group G is of bounded order, then the only basic subgroup of G is G itself.*

LEMMA 10.8. ([5], pp. 98–99) *Let G be an abelian p-group, and suppose that a subgroup U is a direct product $U = U_1 \times U_2 \times U_3 \times \cdots$, where U_n is a direct product of cyclic groups of order p^n for each $n = 1, 2, \cdots$. Then the following conditions are equivalent:*

(i) *U is a basic subgroup of G;*

(ii) *$G = U_1 \times \cdots \times U_n \times [p^nG \cup \bigcup_{k>n} U_k]$ for each $n = 1, 2, \cdots$;*

(iii) *$U_1 \times \cdots \times U_n$ is a maximal factor of G of bounded order p^n for each $n = 1, 2, \cdots$.*

LEMMA 10.9. ([5], p. 112) *If G is a primary abelian group with no elements of infinite height, then there exists a torsion-complete primary abelian group containing G as a pure subgroup.*

LEMMA 10.10. ([5], p. 117) *If S is a pure subgroup of a primary abelian group G, and if S itself is torsion-complete, then S is a factor of G.*

LEMMA 10.11. *If $U = V \times W$ is a basic subgroup of an abelian p-group G, and if V is of bounded order, then there is a subgroup H of G such that $G = V \times H$ and $W \subseteqq H$.*

Proof. Since U is a direct product of cyclic p-groups, there is an integer m such that

$$V = V_1 \times \cdots \times V_m \quad \text{and} \quad W = W_1 \times \cdots \times W_m \times W'_m ,$$

where V_k and W_k are direct products of cyclic groups of order p^k ($k = 1, \cdots, m$), and W'_m is a direct product of cyclic groups of orders greater than p^m. Then

$$U = (V_1 \times W_1) \times \cdots \times (V_m \times W_m) \times W'_m ,$$

and hence by 10.8,

$$G = V \times W_1 \times \cdots \times W_m \times [p^mG \cup W'_m] .$$

Consequently the subgroup $H = [p^mG \cup W]$ has the required properties.

LEMMA 10.12. *If X is a maximal pure independent subset of an abelian p-group G, and if Y is a pure independent subset of G, then* *there exists a maximal pure independent subset Z of G such that $Y \subseteq Z \subseteq X \cup Y$.*

Proof. By Zorn's Lemma there exists a pure independent subset Z of G which is maximal with respect to the property $Y \subseteq Z \subseteq X \cup Y$. Suppose Z is not a maximal pure independent subset of G. Then there exists a maximal pure independent subset Z' such that $Z \subset Z'$. Choose any $d \in Z' - Z$. If the order of d is p^n, let

$$X_n = \{x \mid x \in X \text{ and } p^n x = 0\} .$$

By 10.5 and 10.11 there exist subgroups H_0 and H_1 of G such that $Z \subseteq H_0$ and

$$G = [d] \times H_0 = H_1 \times \prod_{x \in X_n} [x] .$$

Then there exist an element $e \in H_1$ and a finite subset $\{x_0, \cdots, x_{m-1}\} \subseteq X_n$ such that

$$d \in [e] \times [x_0] \times \cdots \times [x_{m-1}] .$$

Observe that if $u \in G$ is an element of order at most p^n such that $p^{n-1}u \notin H_0$, then u has order exactly p^n, and $[u] \cap H_0 = \{0\}$; therefore, as H_0 has index p^n in G, we must have $G = [u] \times H_0$. Consequently, since H_1 contains no factor of order p^n by 10.8 (iii), it follows that $p^{n-1}e \in H_0$. On the other hand, since $p^{n-1}d \notin H_0$, there exists $k < m$ such that $p^{n-1}x_k \notin H_0$. But then $G = [x_k] \times H_0$, and this implies that $Z \cup \{x_k\}$ is a pure subset of G with $Y \subseteq Z \subset Z \cup \{x_k\} \subseteq X \cup Y$. Since this contradicts the choice of Z, it follows that Z is a maximal pure independent subset of G.

Consider now a torsion-complete primary abelian group G and a pure subgroup S of G. Define $\bar{S}$ to be the subgroup consisting of all those elements $x \in G$ which are limits in G of bounded Cauchy sequences of S. It is easy to see, and is implicit in the proof of the next lemma, that S is just the topological *closure* of S in G. Moreover, if T is a pure torsion-complete subgroup of G containing S, then $T \supseteq \bar{S}$; in particular if S itself is torsion-complete, then $\bar{S} = S$.

LEMMA 10.13. *If S is a pure subgroup of a torsion-complete abelian p-group G, then $\bar{S}$ is a pure torsion-complete subgroup of G.*

Proof. First observe that if $\{s_k\}_{k<\infty}$ is a Cauchy sequence of S converging to an element x, and if $p^m x = 0$, then there is a bounded Cauchy sequence $\{t_k\}_{k<\infty}$ of S which converges to x such that $p^m t_k = 0$ for all k. By picking an appropriate subsequence, if necessary, we may assume that

$$x - s_k \in p^k G \qquad (k = 1, 2, \cdots) .$$

Since $p^m x = 0$, we have $p^m s_k \in p^{k+m}G$. Thus, since S is pure, there is an element $s'_k \in S$ such that $p^m s_k = p^{k+m}s'_k$ for each $k = 1, 2, \cdots$. Let $t_k = s_k - p^k s'_k$ $(k = 1, 2, \cdots)$. Then clearly $p^m t_k = 0$, and

$$x - t_k = (x - s_k) + p^k s'_k \in p^k G \qquad (k = 1, 2, \cdots) ,$$

i.e., $\{t_k\}_{k<\infty}$ is a bounded Cauchy sequence of S, bounded by p^m, which converges to x.

Let $\{x_k\}_{k<\infty}$ be a bounded Cauchy sequence of $\bar{S}$. Since G is torsion-complete, there is an element $x \in G$ which is the limit of $\{x_k\}_{k<\infty}$ in G. By picking an appropriate subsequence, if necessary, we may assume that

$$x - x_k \in p^k G \qquad (k = 1, 2, \cdots) .$$

Since $\{x_k\}_{k<\infty}$ is bounded, there is an integer m such that $p^m x_k = 0$ $(k = 1, 2, \cdots)$. Moreover, since each $x_k \in S$, there are Cauchy sequences $\{s_{k,n}\}_{n<\infty}$ such that $\{s_{k,n}\}_{n<\infty}$ converges to x_k for each $k = 1, 2, \cdots$. And, as observed above, we can choose the $s_{k,n}$ such that

$$p^m s_{k,n} = 0 \quad \text{and} \quad x_k - s_{k,n} \in p^n G$$

for all $n, k = 1, 2, \cdots$. Let $t_k = s_{k,k}$. Then

$$x - t_k = (x - x_k) + (x_k - s_{k,k}) \in p^k G ,$$

and hence $\{t_k\}_{k<\infty}$ is a bounded Cauchy sequence of S which converges to x. Therefore $x \in \bar{S}$, and $\bar{S}$ is torsion-complete.

To see that $\bar{S}$ is pure, let $x \in \bar{S}$, and suppose that $x \in p^r G$. Then there is a bounded Cauchy sequence $\{s_k\}_{k<\infty}$ of S such that

$$x - s_k \in p^k G$$

and hence that

$$s_{k+1} - s_k \in p^k G$$

for all $k = 1, 2, \cdots$. Consequently $s_{r+1} \in p^r G$, and therefore, since S is pure, there exist elements $t_1 \in S$ and $s'_k \in S$ $(k = 1, 2, \cdots)$ such that

$$p^r t_1 = s_{r+1} , \quad \text{and} \quad s_{k+1} - s_k = p^k s'_k \qquad \text{for all } k = 1, 2, \cdots .$$

Define elements $t_k \in S$ $(k = 1, 2, \cdots)$ recursively by $t_{k+1} = t_k + p^k s'_{r+k}$. Then clearly $\{t_k\}_{k<\infty}$ is a bounded Cauchy sequence of S which converges to a limit $t \in \bar{S}$. Moreover, if $p^r t_k = s_{r+k}$, then

$$p^r t_{k+1} = p^r t_k + p^{r+k} s'_{r+k} = s_{r+k} + (s_{r+k+1} - s_{r+k}) = s_{r+k+1} ;$$

hence $p^r t_k = s_{r+k}$ for all $k = 1, 2, \cdots$. It follows that $p^r t = x$, whence

$x \in p^r\bar{S}$. Thus $\bar{S}$ is a pure subgroup of G.

COROLLARY 10.14. *If U is a basic subgroup of a torsion-complete primary abelian group G, then $\bar{U} = G$.*

Proof. By 10.10, 10.8 and 10.13.

LEMMA 10.15. *If $R = S \times T$ is a pure subgroup of a torsion-complete abelian p-group G, then $\bar{R} = \bar{S} \times \bar{T}$.*

Proof. Suppose $x \in \bar{S} \cap \bar{T}$ and $x \neq 0$. Then there are bounded Cauchy sequences $\{s_k\}_{k<\infty}$ and $\{t_k\}_{k<\infty}$ of S and T, respectively, such that

$$x - s_k,\ x - t_k \in p^k G \qquad (k = 1, 2, \cdots)\,.$$

Since $x \neq 0$, x has height r for some integer r. It follows that s_k and t_k must also have height r for each $k > r$. And, as $s_k \in S$, $t_k \in T$, and $R = S \times T$ is a pure subgroup of G, it readily follows that $s_k - t_k$ has height r for each $k > r$. But this is a contradiction since

$$s_k - t_k = (x - t_k) - (x - s_k) \in p^k G \quad (k > r)\,.$$

Consequently $\bar{S} \cap \bar{T} = \{0\}$. On the other hand, if $\{x_k\}_{k<\infty}$ is a bounded Cauchy sequence of R converging to a limit $x \in \bar{R}$, then $\{f(x_k)\}_{k<\infty}$ and $\{g(x_k)\}_{k<\infty}$ are bounded Cauchy sequences of S and T, respectively, where f is the projection of R onto S, and g is the projection of R onto T. Hence there are elements $u \in \bar{S}$ and $v \in \bar{T}$ which are the limits of $\{f(x_k)\}_{k<\infty}$ and $\{g(x_k)\}_{k<\infty}$, respectively. Since

$$x_k = f(x_k) + g(x_k) \qquad \text{for each } k = 1, 2, \cdots\,,$$

it follows that $x = u + v$, and we conclude that $\bar{R} = \bar{S} \times \bar{T}$.

11. Exchange and isomorphic refinement theorems for binary algebras. In the present section conditions are found in order for a binary algebra B to have the exchange property, and these conditions are combined with the results of preceeding sections to obtain uniqueness and isomorphic refinement theorems for binary algebras.

The center B^c of a binary algebra B can be written as a direct product

$$B^c = P \times Q \times R$$

where P is a divisible torsion-free abelian group, Q is a divisible torsion abelian group, and R is a reduced abelian group. The groups Q and

$P \times Q$ are unique, and $P \times Q$ is the maximal divisible subgroup of B^c. Therefore R is isomorphic to the reduced part of B^c. By 3.10, B^c has the exchange property if and only if each of the factors P, Q and R has this property. In the case of P the exchange property readily follows from 8.3 and some elementary properties of vector spaces. Since a torsion abelian group is uniquely a direct product of its primary components, it is clear that a torsion abelian group has the exchange property if and only if each of its primary components has the exchange property. In the case of divisible primary groups, and hence for Q, the exchange property again follows essentially from vector space properties. As for reduced groups, the main lemma of this section asserts that a torsion-complete primary abelian group has the exchange property. Consequently every torsion abelian group with torsion-complete primary components has the exchange property.

LEMMA 11.1. *Every torsion-free divisible abelian group G has the exchange property.*

Proof. Using the criterion of 8.3, suppose

$$A = G \times C = \prod_{i \in I} D_i$$

where each of the factors $D_i\,(i \in I)$ is isomorphic to a subgroup of G. If A', C' and $D_i'\,(i \in I)$ are the maximal divisible subgroups of A, C and $D_i\,(i \in I)$, respectively, then

$$A' = G \times C' = \prod_{i \in I} D_i' .$$

Furthermore, for each $i \in I$ there is a subgroup D_i'' such that $D_i = D_i' \times D_i''$, and thus

$$A = A' \times \prod_{i \in I} D_i'' .$$

Regarding A' as a vector space over the field of rational numbers, we can choose a basis X for G and extend it to a basis Y for A' in such a way that every element of $Y - X$ belongs to one of the factors D_i'. Letting E_i' be the vector space spanned by $D_i' \cap (Y - X)$, we conclude that

$$A' = G \times \prod_{i \in I} E_i' .$$

Therefore

$$A = G \times \prod_{i \in I} E_i$$

where $E_i = E_i' \times D_i''\,(i \in I)$, and hence G has the exchange property.

LEMMA 11.2. *Every primary abelian group G of bounded order has the exchange property.*

Proof. Suppose

$$A = G \times C = \prod_{i \in I} D_i$$

where each of the factors D_i is isomorphic to a subgroup of G. Then A is a primary abelian group of bounded order. Let Y be a maximal pure independent subset of G, and for each $i \in I$ let X_i be a maximal pure independent subset of D_i. Then $X = \bigcup_{i \in I} X_i$ is a maximal pure independent subset of A, and it follows by 10.12 that there exists a maximal pure independent subset Z of A such that $Y \subseteqq Z \subseteqq X \cup Y$. By 10.5 and 10.7, A is generated by Z, and G is generated by Y. Consequently, if E_i is the subgroup generated by the set $D_i \cap (Y - X)$ for each $i \in I$, it follows that

$$A = G \times \prod_{i \in I} E_i \,.$$

Thus G has the exchange property.

LEMMA 11.3. *Every divisible abelian p-group G has the exchange property.*

Proof. Suppose

$$A = G \times C = \prod_{i \in I} D_i$$

where each D_i is isomorphic to a subgroup of G. If A', C' and D_i' $(i \in I)$ are the maximal divisible subgroups of A, C and D_i $(i \in I)$, respectively, then

$$A' = G \times C' \times \prod_{i \in I} D_i' \,.$$

Furthermore, if D_i'' is such that $D_i = D_i' \times D_i''$ for each $i \in I$, then

$$A = A' \times \prod_{i \in I} D_i'' \,.$$

Clearly

$$A'[p] = G[p] \times C'[p] = \prod_{i \in I} D_i'[p] \,,$$

and since $G[p]$ is of bounded order p, there exist subgroups $U_i \subseteqq D_i'[p]$ $(i \in I)$ such that

$$A'[p] = G[p] \times \prod_{i \in I} U_i \,. \tag{1}$$

For each $i \in I$ there exists a divisible subgroup E_i' of D_i' such that $E_i'[p] = U_i$, and it follows from (1) that the direct product

$$A'' = G \times \prod_{i \in I} E_i'$$

exists. Moreover, since A'' is divisible, it is a pure subgroup of A', and using the fact that $A'[p] \subseteqq A''$ we infer by 10.4 that $A' = A''$. Thus

$$A = G \times \prod_{i \in I} E_i$$

where $E_i = E_i' \times D_i''$, and G has the exchange property.

LEMMA 11.4. *Every torsion-complete abelian p-group G has the exchange property.*

Proof. We first prove that G has the 2-exchange property and hence the finite exchange property. Thus suppose

$$A = G \times C = D_0 \times D_1$$

where D_0 and D_1 are isomorphic to subgroups of G. Then A is an abelian p-group without elements of infinite height, and hence by 10.9 there is a torsion-complete abelian p-group A' containing A as a pure subgroup. By 10.13 we may assume that A' is the closure of A, $A' = \bar{A}$, and in this case it follows by 10.15 that

$$\bar{A} = G \times \bar{C} = \bar{D}_0 \times \bar{D}_1 .$$

Choose maximal pure independent subsets X_0, X_1 and Y of $\bar{D}_0$, $\bar{D}_1$ and G respectively. Then $X = X_0 \cup X_1$ is a maximal pure independent subset of $\bar{A}$, and by 10.12 there is a maximal pure independent subset Z of A such that $Y \subseteqq Z \subseteqq X \cup Y$. Since every subset of Z generates a factor of $[Z]$, the subgroups generated by $\bar{D}_0 \cap Z$ and $\bar{D}_1 \cap Z$ are pure in $\bar{A}$. Let E_0 and E_1 be the closures of the subgroups generated by $\bar{D}_0 \cap Z$ and $\bar{D}_1 \cap Z$, respectively. Then by 10.14 and 10.15,

$$\bar{A} = G \times E_0 \times E_1 .$$

Since $E_0 \times (G \times E_1) \supseteqq D_0 \times \bar{D}_1 \supseteqq G \times E_1$, we infer from the modular law that

$$D_0 \times \bar{D}_1 = (E_0 \cap (D_0 \times \bar{D}_1)) \times G \times E_1 = G \times (D_0 \cap E_0) \times E_1 .$$

Therefore $E_1 \times (G \times (D_0 \cap E_0)) \supseteqq D_0 \times D_1 \supseteqq G \times (D_0 \cap E_0)$, and a second application of the modular law yields

$$\begin{aligned} A = D_0 \times D_1 &= (E_1 \cap (D_0 \times D_1)) \times G \times (D_0 \cap E_0) \\ &= G \times (D_0 \cap E_0) \times (D_1 \cap E_1) \,. \end{aligned}$$

Consequently G has the 2-exchange property.

Now suppose

$$A = G \times C = \prod_{i \in I} D_i \tag{1}$$

where each of the factors D_i is isomorphic to a subgroup of G. A is therefore an abelian p-group with no elements of infinite height. For each $i \in I$ let f_i be the project of A onto D_i induced by the second decomposition in (1). We begin by proving the following statement:

(S) *There exist a finite set $J \subseteqq I$ and subgroups G_0 and G_1 such that $G = G_0 \times G_1$, G_0 is of bounded order, and*

$$G_1[p] \subseteqq \prod_{i \in J} D_i \,. \tag{2}$$

Assume that (S) is false. Then for every finite subset $J \subseteqq I$ and every decomposition $G = G_0 \times G_1$ where G_0 is of bounded order, there is an element $x \in G_1[p]$ and an index $i \in I - J$ such that $f_i(x) \neq 0$. Using this we shall construct a sequence of elements $x_0, x_1, x_2, \cdots \in G[p]$ and a sequence of indices $i_1, i_2, i_3, \cdots \in I$ such that the following conditions hold for every positive integer n:

$$\text{height } x_n > \text{height } f_i(x_{n-1}) \text{ whenever } i \in I \text{ and } f_i(x_{n-1}) \neq 0\,; \tag{3}$$

$$f_{i_n}(x_0) = f_{i_n}(x_1) = \cdots = f_{i_n}(x_{n-1}) = 0 \neq f_{i_n}(x_n)\,. \tag{4}$$

Pick any element $x_0 \in G[p]$. Suppose the elements $x_1, \cdots, x_m \in G[p]$ and the indices $i_1, \cdots, i_m \in I$ have been so chosen that (3) and (4) hold for $n = 1, \cdots, m$. Then the set

$$J_m = \{i \mid i \in I \text{ and } f_i(x_n) \neq 0 \text{ for some } n \leqq m\}$$

is finite, and we can choose a positive integer r such that

$$r \geqq \text{height } f_i(x_n) \text{ whenever } i \in J_m,\, n \leqq m \text{ and } f_i(x_n) \neq 0\,.$$

By 10.8, G has a decomposition $G = G_0 \times G_1$ such that $p^{r+1}G_0 = \{0\}$ and such that G_1 has no factor of order less that p^{r+2}. Therefore there exists and element $x_{m+1} \in G_1[p]$ and an index $i_{m+1} \in I - J_m$ such that $f_{i_{m+1}}(x_{m+1}) \neq 0$. Since the height of x_{m+1} is necessarily larger than r, we infer from the choice of r that (3) holds for $n = m + 1$. Also, since $i_{m+1} \notin J_m$, it follows that (4) also holds with $n = m + 1$. Thus the existence of the sequences of elements $x_n \in G[p]$ and of indices $i_n \in I$ satisfying (3) and (4) follows by induction.

For each $m = 0, 1, 2, \cdots$ let

$$y_m = x_0 + \cdots + x_m .$$

If $m > n$, then it follows from (4) that

$$f_{i_n}(y_m) = f_{i_n}(x_n) + \cdots + f_{i_n}(x_m) .$$

From (3) we infer that the height of $f_{i_n}(x_n)$ is less than the height of $f_{i_n}(x_k)$ for $k = n + 1, \cdots, m$. Consequently

$$\text{(5)} \qquad \text{height } f_{i_n}(y_m) = \text{height } f_{i_n}(x_n) \text{ whenever } m > n .$$

Notice that (3) also implies that the height of x_m is at least m. Therefore

$$y_{m+1} - y_m = x_{m+1} \in p^m G \qquad (m = 0, 1, 2, \cdots) ,$$

and since each y_m has order p, the sequence $\{y_m\}_{m<\infty}$ is a bounded Cauchy sequence of G which must converge to a limit $y \in G$. Furthermore, for each $i \in I$, the sequence $\{f_i(y_m)\}_{m<\infty}$ is a bounded Cauchy sequence of D_i which converges to $f_i(y)$. Now $f_i(y) = 0$ for all but finitely many $i \in I$, and therefore there is a positive integer n such that $f_{i_n}(y) = 0$. But the sequence $\{f_{i_n}(y_m)\}_{m<\infty}$ cannot converge to 0, since according to (5) the heights of the elements $f_{i_n}(y_0), f_{i_n}(y_1), f_{i_n}(y_2), \cdots$ are bounded. Thus we have a contradiction, and hence (S) must be true.

Choose J, G_0 and G_1 according to (S). Considering the decomposition

$$A = \prod_{i \in J} D_i \times \prod_{i \in I-J} D_i ,$$

let f be the projection of A onto the factor $\prod_{i \in J} D_i$, and let G^* be the image of G_1 under f. It follows from (2) that f maps G_1 isomorphically onto G^*, and that

$$\text{(6)} \qquad G^*[p] = G_1[p] .$$

In particular, G^* is torsion-complete. Furthermore, if $x \in G^*[p]$, then $x = f(x)$, and the height of x in G^* is at least as large as the height of x in G_1. Since G_1 is a pure subgroup of A, it follows by 10.3 that G^* is a pure subgroup of A. Thus, by 10.10, G^* is a factor of A, and consequently

$$\prod_{i \in J} D_i = G^* \times H$$

for some subgroup H. By the first part of the proof, G^* has the finite exchange property, and thus there exist subgroups $E_i \subseteq D_i \, (i \in J)$ such that

$$\prod_{i \in J} D_i = G^* \times \prod_{i \in J} E_i \ .$$

Therefore

$$A = G_0 \times G_1 \times C = G^* \times \prod_{i \in I} E_i$$

where $E_i = D_i$ if $i \in I - J$. From (6) and this last decomposition we see that the direct product

$$A' = G_1 \times \prod_{i \in I} E_i \ ,$$

exists. Moreover, if $y \in G^*$ then there is an element $x \in G_1$ such that $y = f(x)$. Hence $x = y + z$ for some element $z \in \prod_{i \in I-J} D_i = \prod_{i \in I-J} E_i$, and we conclude that $y = x - z \in A'$. This shows that $G^* \subseteqq A'$, and therefore $A' = A$. Finally, G_0 is of bounded order and thus has the exchange property by 11.2. According to 3.9 we can therefore find subalgebras $F_i \subseteqq E_i \ (i \in I)$ such that

$$A = G_0 \times G_1 \times \prod_{i \in I} F_i = G \times \prod_{i \in I} F_i \ .$$

Hence G has the exchange property, and the proof of 11.4 is complete.

THEOREM 11.5. *If B is a binary algebra such that the reduced part of B^c is a torsion group each primary component of which is torsion-complete, then B has the exchange property.*

Proof. This is an immediate consequence of 8.1, 11.1, 11.3 and 11.4, together with the introductory remarks of this section.

Combining 11.5 with 4.2, 5.3 and 7.1, respectively, we obtain the following principal uniqueness and isomorphic refinement theorems for binary algebras.

THEOREM 11.6. *If a binary algebra A has two direct decompositions with countably many factors,*

$$A = B_0 \times B_1 \times B_2 \times \cdots = C_0 \times C_1 \times C_2 \times \cdots \ ,$$

where the reduced parts of all the groups B_i^c and C_j^c are torsion groups with torsion-complete primary components, then these two direct decompositions of A have centrally isomorphic refinements.

COROLLARY 11.7. *If A is a binary algebra such that the reduced part of A^c is a torsion group with torsion-complete primary components, then any two countable direct decompositions of A have centrally isomorphic refinements.*

THEOREM 11.8. *If a binary algebra A has a direct decomposition*

$$A = \prod_{i \in I} B_i$$

where, for each $i \in I$, the reduced part of B_i^c is a torsion group with torsion-complete primary components, then any two direct decompositions of A into indecomposable factors are centrally isomorphic.

THEOREM 11.9. *If a binary algebra A has a direct decomposition*

$$A = \prod_{i \in I} B_i$$

where, for each $i \in I$, B_i^c is countable and the reduced part of B_i^c is a torsion group each primary component of which is of bounded order, then any two direct decompositions of A have centrally isomorphic refinements.

A final theorem describes a class of binary algebras with uncountable centers having the isomorphic refinement property.

THEOREM 11.10. *If A is a binary algebra such that the maximal divisible subgroup of A^c is countable and the reduced part of A^c is a torsion group each primary component of which is a torsion-complete group with countable basic subgroups, then any two direct decompositions of A have centrally isomorphic refinements.*

Proof. Suppose

$$A = \prod_{i \in I} B_i = \prod_{j \in J} C_j \, . \tag{1}$$

Since the maximal divisible subgroup of A^c is countable and the basic subgroups of each primary component of the reduced part of A^c are countable, it follows that there exists a countable subset I' of I such that $B_i^c = \{0\}$ for each $i \in I - I'$. The factor $\prod_{i \in I - I'} B_i$ has the exchange property, and hence there are subalgebras D_j, D_j' $(j \in J)$ such that $C_j = D_j \times D_j'$ and

$$A = \prod_{i \in I - I'} B_i \times \prod_{j \in J} D_j \, .$$

Consequently

$$\prod_{i \in I - I'} B_i \cong^c \prod_{j \in J} D_j' \, , \tag{2}$$

and, as $\prod_{i \in I - I'} B_i^c = \{0\}$, we infer by 2.19 that

$$\prod_{i \in I'} B_i = \prod_{j \in J} D_j \, . \tag{3}$$

Repeating the argument above for the factor $\prod_{j\in J} D_j$, there is a countable subset J' of J such that $D_j^c = \{0\}$ for each $j \in J - J'$, and there are subalgebras E_i, E_i' $(i \in I')$ such that $B_i = E_i \times E_i'$ and

$$(4) \qquad \prod_{j\in J-J'} D_j \cong^c \prod_{i\in I'} E_i' \,,$$

$$(5) \qquad \prod_{j\in J'} D_j = \prod_{i\in I'} E_i \,.$$

The pairs of decompositions (2) and (4) each have centrally isomorphic refinements by 11.9, and the decompositions (5) have centrally isomorphic refinements by 11.7. Therefore the original decompositions (1) have centrally isomorphic refinements, and the proof is complete.

12. Counterexamples and open problems. This final section contains two examples that yield negative answers to some questions related to the results in this paper. A number of unsolved problems suggested by our investigations are also mentioned.

In 3.10 it was shown that if an algebra B is a direct product of finitely many subalgebras each of which has the exchange property, then B has the exchange property. The first example shows that this result cannot be extended to products of infinitely many subalgebras. In fact, the example shows that if B is an abelian p-group such that

$$B = B_1 \times B_2 \times B_3 \times \cdots$$

where, for $k = 1, 2, 3 \cdots$, B_k is a cyclic group of order p^k, then B does not have the 2-exchange property. Thus the simplest unbounded abelian p-group fails to have the exchange property.

Let

$$A = \prod_{k=1}^{\infty} [u_k] \times \prod_{k=1}^{\infty} [v_k]$$

where, for $k = 1, 2, 3, \cdots$, $[u_k]$ and $[v_k]$ are cyclic groups of order p^k. Also, let

$$B = \prod_{k=1}^{\infty} [u_k + pv_{k+1}] \,, \qquad C = \prod_{k=1}^{\infty} [v_k] \,,$$

$$D_1 = \prod_{k=1}^{\infty} [v_k + pu_{k+1}] \,, \qquad D_2 = \prod_{k=1}^{\infty} [u_k] \,.$$

It is easy to check that

$$A = B \times C = D_1 \times D_2 \,,$$

and in order to prove that B does not have the 2-exchange property it is sufficient to show that the assumption that

$$(1) \qquad A = B \times E_1 \times E_2 , \qquad E_1 \subseteqq D_1 , \qquad E_2 \subseteqq D_2$$

leads to a contradiction.

Assume that (1) holds. Since A is a direct product of finite groups, it and all its direct factors have the unique factorization property. Inasmuch as $C \cong E_1 \times E_2$, this implies that for each positive integer k only one of the groups E_1 and E_2 has a cyclic factor of order p^k. Observing that

$$B \times E_2 \subseteqq B \times D_2 = B \times pC ,$$

we have $v_k = b + pc + e$ where $b \in B$, $c \in C$, $b + pc \in B \times E_2$, and $e \in E_1$. Using the fact that $B \times C$ exists we see that, for $r = 1, 2, 3, \cdots, k-1$, the element $p^r(v_k - b) = p^{r+1}c + p^r e$ has height r, and hence the height of $p^r e$ is also r. Since $p^k e = 0$, this shows that $[e]$ is a pure subgroup of E_1, and hence a factor of E_1, of order p^k. Consequently E_2 cannot have a direct factor of order p^k, and since this is true for every positive integer k, we infer that $E_2 = \{0\}$, and hence $A = B \times E_1 \subseteqq B \times D_1$. But it is easy to see that neither u_1 nor v_1 belongs to $B \times D_1$, and we have thus arrived at a contradiction.

In 8.1 it was shown that if the center of an algebra B has the exchange property, then B has the exchange property. Our second example shows that the converse of this result is false. For this purpose we construct a group B such that

(i) B^c is an infinite cyclic group.

(ii) The commutator subgroup of B equals B, $[B, B] = B$.

First observe that this does in fact imply that B has the required properties. In fact, suppose $B^c = [u]$ and let $A = B^c \times C$ where $C = [v]$ is also an infinite cyclic group. Also let $D_1 = [2u + 3v]$ and $D_2 = [3u + 5v]$. Then $A = D_1 \times D_2$. Since $B^c \times D_1 = B^c \times [3v] \neq A$ and $B^c \times D_2 = B^c \times [5v] \neq A$, we see that B^c does not have the 2-exchange property. On the other hand, suppose A is any algebra containing B as a subalgebra, and suppose C and D_i $(i \in I)$ are subalgebras of A such that

$$A = B \times C = \prod_{i \in I} D_i .$$

Let g and h_i be the projections of A onto C and D_i induced by these two direct decompositions of A. Then gh_i maps B homomorphically into the center of C, whence it follows by (ii) that, for each $b \in B$, $gh_i(b) = 0$ or, equivalently, $h_i(b) \in B$. Thus, for each $i \in I$, h_i maps B into $B \cap D_i$, and we infer that

$$B = \prod_{i \in I} (B \cap D_i) .$$

It follows by the modular law that for each $i \in I$ there exists a subalgebra $E_i \subseteqq D_i$ such that $D_i = (B \cap D_i) \times E_i$, and we conclude that

$$A = B \times \prod_{i \in I} E_i .$$

Hence B has the exchange property.

In order to construct a group having the properties (i) and (ii) we proceed as follows. For $n = 2, 3, 4, \cdots$ let H_n be the group of all n by n matrices of determinant 1 over a field of characteristic 0 that contains a primitive nth root of unity. Then the center of H_n contains a cyclic group of order n, and the commutator subgroup of H_n equals H_n. The Cartesian product H of $H_2, H_3, H_4, \cdots$ therefore has the properties that its center contains an infinite cyclic group and that the commutator subgroup of H is equal to H. We now take for B a free amalgamated product of two isomorphic copies B_1 and B_2 of H, with amalgamated subgroup $Z = B_1 \cap B_2$ an infinite cyclic group contained in the centers of both B_1 and B_2. It is known that $B^c = B_1^c \cap B_2^c$, so that B^c is in this case the infinite cyclic group Z. Thus (i) holds, and it is obvious that (ii) is also satisfied.

The most interesting unsolved problem suggested by the results in this paper is whether in Theorem 7.1 the assumption of countably generated centers is needed. Specifically, *is it true that if an algebra A is a direct product of subalgebras each of which has the exchange property, then any two direct decompositions of A have isomorphic refinements*? Even if the answer is negative, one might hope for an affirmative answer in special cases, such as for groups whose centers are of bounded order. Of course, if the answer should turn out to be affirmative, then this would include Theorems 4.2, 5.3 and 7.1 as special cases.

Another problem concerns the relation of the finite exchange property and the exchange property: *Is the exchange property always implied by the finite exchange property*? In connection with Theorem 7.1 it would be particularly interesting to know whether for an algebra B with a countable generated center the finite exchange property implies the $\aleph_0$-exchange property (and therefore the exchange property). It is not hard to show that for such an algebra B the condition

$$A = B \times C = D_0 \times D_1 \times D_2 \times \cdots$$

implies that

$$A = B \times E_0 \times E_1 \times E_2 \times \cdots$$

where each of the factors E_k is a subalgebra of the finite product $D_0 \times D_1 \times \cdots \times D_k$, but we do not know whether the factors E_k can be replaced by subalgebras of the factors D_k.

Theorem 8.3 raises the problem of determining those abelian operator groups that have the exchange property. In this regard the following question seems particularly relevant: *Is it true that if an abelian operator group satisfies the minimal condition, then it has the exchange property*? For ordinary reduced abelian groups the results in § 11 apply only to groups with no elements of infinite height. It would be of interest to know whether the class of all reduced primary abelian groups having the exchange property contains any groups with (nonzero) elements of infinite height.

BIBLIOGRAPHY

1. R. Baer, *Direct decompositions*, Trans. Amer. Math. Soc., **62** (1947), 62–98.
2. ———, *Direct decompositions into infinitely many summands*, Trans. Amer. Math. Soc., **64** (1948), 519–551.
3. G. Birkhoff, *Lattice theory*, Amer. Math., Soc. Colloquium Publications, vol. 25, rev. ed., New York, 1948.
4. P. Crawley, *Direct decompositions with finite dimensional factors*, Pacific J. Math., **12** (1962), 457–468.
5. L. Fuchs, *Abelian groups*, Budapest, 1958.
6. B. Jónsson and A. Tarski, *Direct decompositions of finite algebraic systems*, Notre Dame Mathematical Lectures, no. 5, 1947.
7. I. Kaplansky, *Infinite abelian groups*, Ann Arbor, 1956.
8. W. Specht, *Gruppentheorie*, Berlin, Gottingen and Heidelberg, 1956.

ON CONTINUOUS MATRIX-VALUED FUNCTIONS ON A STONIAN SPACE

DON DECKARD AND CARL PEARCY

1. Introduction. In this paper the authors continue the study (begun in [9] and carried on in [3] and [10]) of matrices with entries from the algebra $C(\mathfrak{X})$ of all continuous complex-valued functions on an extremely disconnected, compact Hausdorff space $\mathfrak{X}$. (Such spaces are sometimes called Stonian after M. H. Stone, who considered them in [14].) One of the authors has shown ([10], Theorem 3) that if A and B are $n \times n$ matrices over $C(\mathfrak{X})$ such that $A(x)$ is unitarily equivalent to $B(x)$ for each $x \in \mathfrak{X}$, then A and B are unitarily equivalent in the algebra $M_n(\mathfrak{X})$ of all $n \times n$ matrices over $C(\mathfrak{X})$. It is thus natural to ask whether the similarity of $A(x)$ and $B(x)$ for each $x \in \mathfrak{X}$ is sufficient to guarantee the similarity of A and B in $M_n(\mathfrak{X})$. We show by example in § 2 that the answer is no; however, we also show that if the hypothesis is strengthened by the addition of a uniform boundedness requirement, then the similarity of A and B in $M_n(\mathfrak{X})$ does indeed follow. As a by-product of the technique introduced to give this result, we obtain a new short proof of Theorem 3 of [10].

In § 3 we show that a certain class of entire functions maps $M_n(\mathfrak{X})$ onto itself; this is a generalization (with a different proof) of a result of Kurepa [8] for $n \times n$ matrices, and adds to the information obtained by Brown [1] on the question of which entire functions map which Banach algebras onto themselves. As a corollary, we learn that every invertible element of $M_n(\mathfrak{X})$ has a logarithm. Section 4 is devoted to proving that an element of $M_n(\mathfrak{X})$ has an identically vanishing trace if and only if it is a commutator in $M_n(\mathfrak{X})$. (See Remark 2, § 4, for a paraphrase of this result cast in the terminology of operator theory on Hilbert space.) Finally, in § 5 the authors give two examples which indicate that it is probably fruitless to pursue the structure theory of matrices over $C(\mathfrak{X})$ where $\mathfrak{X}$ is a more general topological space than a Stonian space.

2. Similarity in $M_n(\mathfrak{X})$. The most convenient definition of $M_n(\mathfrak{X})$ is as follows. Let M_n denote the full ring of $n \times n$ complex matrices under the operator norm, and let $\mathfrak{X}$ be any Stonian space. Denote by $M_n(\mathfrak{X})$ the *-algebra of continuous functions from $\mathfrak{X}$ to M_n, where the algebraic operations in $M_n(\mathfrak{X})$ are defined pointwise. Under the norm $\|A\| = \sup_{x \in \mathfrak{X}} \|A(x)\|$, $M_n(\mathfrak{X})$ is a C^*-algebra identifiable with the C^*-algebra of all $n \times n$ matrices over $C(\mathfrak{X})$. Moreover, $M_n(\mathfrak{X})$ is an

Received July 12, 1963.

AW^*-algebra [7], and this fact is used briefly in this section.

We first show that pointwise similarity of $A(x)$ and $B(x)$ on $\mathfrak{X}$ is not sufficient to ensure that A and B be similar in $M_n(\mathfrak{X})$. For this purpose, let $\mathscr{S}$ be the Stone-Czech compactification of the natural numbers $\mathscr{N}$. Then $\mathscr{S}$ is a Stonian space. (See, for example, the discussion on page 295 of [12].) Consider elements A and B of $M_2(\mathscr{S})$ defined by:

$$A(x) = \begin{pmatrix} 0 & 1/x^2 \\ 0 & 0 \end{pmatrix}, \qquad B(x) = \begin{pmatrix} 0 & 1/x \\ 0 & 0 \end{pmatrix}$$

for each natural number $x \in \mathscr{N}$. Then $A(x) = B(x) = 0$ for $x \in \mathscr{S} - \mathscr{N}$, and it is obvious that $A(x)$ and $B(x)$ are similar for each $x \in \mathscr{S}$. Suppose that $S = (s_{ij})$ is an invertible element in $M_2(\mathscr{S})$ satisfying $SA = BS$. Calculation yields $s_{21}(x) = 0$ for $x \in \mathscr{N}$ so that $s_{21} \equiv 0$. Furthermore, $s_{11}(x) = xs_{22}(x)$ for $x \in \mathscr{N}$, and the invertibility of S guarantees that s_{22} never vanishes. Thus s_{11} is unbounded, contradicting $s_{11} \in C(\mathscr{S})$, and it follows that A and B are not similar in $M_2(\mathscr{S})$.

The following theorem gives necessary and sufficient conditions for A and B to be similar in $M_n(\mathfrak{X})$.

THEOREM 1. *Let $\mathfrak{X}$ be any Stonian space, and let $A, B \in M_n(\mathfrak{X})$. Suppose that there is a dense subset $\mathscr{D} \subset \mathfrak{X}$ and a positive number M such that for $x \in \mathscr{D}$, there is an invertible matrix $S(x)$ satisfying $S(x)A(x)S^{-1}(x) = B(x)$, $\| S(x) \| < M$, and $\| S^{-1}(x) \| < M$. Then there is an invertible element $T \in M_n(\mathfrak{X})$ satisfying $TAT^{-1} = B$, $\| T \| \leqq M$, and $\| T^{-1} \| \leqq M$.*

Proof. We consider collections $\{\mathscr{U}_i\}$ of nonempty, disjoint, compact open sets $\mathscr{U}_i \subset \mathfrak{X}$ with the property that if $\mathscr{U}_i \in \{\mathscr{U}_i\}$, then there is an invertible element $T_i \in M_n(\mathscr{U}_i)$ satisfying $T_i(x)A(x)T_i^{-1}(x) = B(x)$, $\| T_i(x) \| < M$, and $\| T_i^{-1}(x) \| < M$ for each $x \in \mathscr{U}_i$. Let $\{\mathscr{U}_i\}_{i \in I}$ be a maximal such collection, and denote $\mathscr{U} = \overline{\bigcup_{i \in I} \mathscr{U}_i}$. Then $\mathscr{U}$ is compact open, and it follows from Lemma 2.1 of [3] that the function $\tilde{T}$ defined on $\bigcup_{i \in I} \mathscr{U}_i$ so as to extend each of the T_i can be extended to an element $T \in M_n(\mathscr{U})$. Similarly, there is a function $Z \in M_n(\mathscr{U})$ which extends each of the T_i^{-1}. It is clear from continuity considerations that $Z = T^{-1}$, and that T has all the desired properties on $\mathscr{U}$, so that it suffices to prove $\mathscr{U} = \mathfrak{X}$. Suppose, to the contrary, that $\mathfrak{X} - \mathscr{U} \neq \phi$. To obtain a contradiction, it suffices to find a compact open set $\mathscr{V} \subset \mathfrak{X} - \mathscr{U}$ and an invertible element $V \in M_n(\mathscr{V})$ such that for $x \in \mathscr{V}$, $V(x)A(x) = B(x)V(x)$, $\| V(x) \| < M$, and $\| V^{-1}(x) \| < M$. To do this, we regard the equation $VA = BV$ as a system of linear equations

$$\text{(L)}\qquad \begin{array}{l} c_{11}v_1 + c_{12}v_2 + \cdots + c_{1m}v_m = 0 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ c_{m1}v_1 + c_{m2}v_2 + \cdots + c_{mm}v_m = 0 \end{array}$$

where

(1) the unknown functions v_i are the entries, in some prescribed order, of the matrix V

(2) the coefficients $c_{ij} \in C(\mathfrak{X} - \mathscr{U})$ are the appropriate combinations of the entries of the matrices A and B

(3) $m = n^2$.

For $x \in \mathfrak{X} - \mathscr{U}$, consider the corresponding system $(L(x))$ of linear equations, and let $x_0 \in \mathfrak{X} - \mathscr{U}$ be a point such that the rank $r(x)$ of the system $(L(x))$ assumes its maximum r_0 at x_0. (The case $r_0 = 0$ leads trivially to a contradiction of $\mathfrak{X} - \mathscr{U} \neq \phi$, and we ignore it. The case $r_0 = m$ cannot occur.) Then there is some $r_0 \times r_0$ minor N of the coefficient determinant of the system $(L(x_0))$ which is nonzero, and by continuity there exists a compact open neighborhood $\mathscr{V}_1 \subset \mathfrak{X} - \mathscr{U}$ of x_0 such that for $x \in \mathscr{V}_1$, the same minor N remains a nonzero minor of maximum size. According to the hypothesis, there is a point $x_1 \in \mathscr{V}_1$ and an invertible matrix $S(x_1)$ such that $S(x_1)A(x_1) = B(x_1)S(x_1)$, $\|S(x_1)\| < M$, and $\|S^{-1}(x_1)\| < M$. Let the corresponding nontrivial solution of the system $(L(x_1))$ be denoted by $(\mu_1, \mu_2, \cdots, \mu_m)$ (i.e., the μ_i are the entries of the matrix $S(x_1)$). We wish to define an m-tuple $(v_1(x), v_2(x), \cdots, v_m(x))$ at each point of $\mathscr{V}_1$ in such a way that

(1) the m-tuple is a solution of $(L(x))$ for each $x \in \mathscr{V}_1$,

(2) $v_i \in C(\mathscr{V}_1)$ for $1 \leqq i \leqq m$, and

(3) $v_i(x_1) = \mu_i$ for $1 \leqq i \leqq m$. This is accomplished as follows. Since for $x \in \mathscr{V}_1$, N is a nonzero minor of maximum size, it suffices to solve (continuously on $\mathscr{V}_1$) the r_0 equations affiliated with N. Thus for the appropriate $m - r_0$ values of i (the values not affiliated with N), define $v_i(x) \equiv \mu_i$ on $\mathscr{V}_1$; then for $x \in \mathscr{V}_1$ the other r_0 numbers $v_i(x)$ are determined by Cramer's rule, and since the functions c_{ij} are continuous it follows that (1), (2), and (3) above are satisfied. Next place the resulting functions $v_i \in C(\mathscr{V}_1)$ in their appropriate positions in the matrix V, and shrink the neighborhood $\mathscr{V}_1$ of x_1 to a compact open neighborhood $\mathscr{V} \subset \mathscr{V}_1$ of x_1 such that for $x \in \mathscr{V}$, the matrix $V(x)$ is invertible and the inequalities $\|V(x)\| < M$ and $\|V^{-1}(x)\| < M$ remain valid. The existence of the compact open set $\mathscr{V}$ contradicts the maximality of the collection $\{\mathscr{U}_i\}_{i \in I}$, and thus the proof is complete.

We can prove Theorem 3 of [10] in a similar fashion.

THEOREM 2. *If $\mathfrak{X}$ is Stonian and $A, B \in M_n(\mathfrak{X})$ are such that $A(x)$ and $B(x)$ are unitarily equivalent at each point of a dense subset of $\mathfrak{X}$, then A and B are unitarily equivalent in $M_n(\mathfrak{X})$.*

Proof. We consider collections $\{\mathscr{U}_i\}$ of nonempty, disjoint, compact open subsets $\mathscr{U}_i \subset \mathfrak{X}$ with the property that if $\mathscr{U}_i \in \{\mathscr{U}_i\}$, then there is a unitary element $U_i \in M_n(\mathscr{U}_i)$ satisfying $U_i(x)A(x)U_i^*(x) = B(x)$ for each $x \in \mathscr{U}_i$. As before, we choose a maximal collection $\{\mathscr{U}_i\}_{i \in I}$, and define $\mathscr{U} = \overline{\bigcup_{i \in I} \mathscr{U}_i}$. Again it suffices to prove $\mathscr{U} = \mathfrak{X}$. The argument then proceeds exactly as above, except that the system of linear equations to be considered is the system equivalent to the pair of equations $VA = BV$ and $VA^* = B^*V$. (Thus the system consists of $2n^2$ equations in n^2 unknowns, but it is clear that this has no effect on the argument.) Then, proceeding essentially as above, we obtain a compact open subset $\mathscr{V} \subset \mathfrak{X} - \mathscr{U}$ and an invertible (not necessarily unitary) element $V \in M_n(\mathscr{V})$ such that for $x \in \mathscr{V}$, $V(x)A(x) = B(x)V(x)$ and $V(x)A^*(x) = B^*(x)V(x)$. One knows from ([14], Lemma 2.1) that we can write V in polar form $V = UP$ where U is a unitary element of $M_n(\mathscr{V})$. A standard calculation shows that for $x \in \mathscr{V}$, $U(x)A(x)U^*(x) = B(x)$; thus the existence of $\mathscr{V}$ contradicts the maximality of the collection $\{\mathscr{U}_i\}_{i \in I}$, and the proof is complete.

REMARK. One would naturally like to have a collections of global objects to attach to an element $A \in M_n(\mathfrak{X})$ which would serve as a complete set of similarity invariants for A. In this connection, it is easy to see that one cannot always obtain an element $J \in M_n(\mathfrak{X})$ such that A is similar to J in $M_n(\mathfrak{X})$ and such that $J(x)$ is in Jordan form for each $x \in \mathfrak{X}$.

3. Entire functions on $M_n(\mathfrak{X})$. We say that an entire function f has property (K) if, for every complex number ζ, there is a complex number z satisfying $f(z) = \zeta$ and $f'(z) \neq 0$. In [8] Kurepa showed that an entire function f maps M_n onto itself if and only if f has property (K). The study was then taken up by Brown [1] who characterized the class of entire functions f which map the algebra $\mathscr{L}(\mathscr{H})$ of all bounded operators on an infinite dimensional Hilbert space $\mathscr{H}$ onto itself. Brown showed that such an f maps every Banach algebra onto itself, and we say that such an f has property (B). Since certain W^*-algebras of operators on Hilbert space have faithful C^*-representations as an $M_n(\mathfrak{X})$ (see [9]), one has, in a sense, $\mathscr{L}(\mathscr{H}) \supset M_n(\mathfrak{X}) \supset M_n$. Thus it is of interest to discover which entire functions map $M_n(\mathfrak{X})$ onto itself, and the answer is given by

THEOREM 3. *If f is an entire function and $\mathfrak{X}$ is a Stonian space, then f maps $M_n(\mathfrak{X})$ onto itself if and only if f has property (K).*

Proof. Since for each $x \in \mathfrak{X}$, $[p(A)](x) = p(A(x))$ for every polynomial $p(z)$, and since f is the uniform limit of polynomials on compact sets of the z-plane, $[f(A)](x) = f(A(x))$ for each $x \in \mathfrak{X}$. Thus, if f maps $M_n(\mathfrak{X})$ onto itself, then f must map M_n onto itself, so that by Kurepa's theorem [8], f has property (K). Now suppose that f has property (K), and let $A \in M_n(\mathfrak{X})$. We look for $B \in M_n(\mathfrak{X})$ such that $f(B) \quad A$. Let x_0 be an arbitrary point of $\mathfrak{X}$ and let $\zeta_1, \cdots, \zeta_p$ be the distinct eigenvalues of $A(x_0)$. Choose $z_1, \cdots, z_p$ to be complex numbers with the properties that $f(z_i) = \zeta_i$ and $f'(z_i) \neq 0$. For $i = 1, \cdots, p$, let $\mathscr{D}_i$ be a (non-degenerate) closed disc about z_i such that f is Schlicht on $\mathscr{D}_i$, and arrange it so that the sets $f(\mathscr{D}_i)$ are mutually disjoint. Let g denote the inverse of the restriction of f to $\bigcup_{i=1}^{p} \mathscr{D}_i$. Then g is defined and continuous on $\mathscr{D} = \bigcup_{i=1}^{p} f(\mathscr{D}_i)$ and is analytic at each interior point of $\mathscr{D}$. It follows from Lemma 2.2 of [3] that there exists a compact open neighborhood $\mathscr{N}_0 = \mathscr{N}(x_0)$ of x_0 such that for $x \in \mathscr{N}_0$, the spectrum of $A(x)$ (denoted hereafter $\varLambda[A(x)]$) is a subset of the interior of $\mathscr{D}$. If A_0 denotes the restriction of A to $\mathscr{N}_0$, then A_0 is an element of the C^*-algebra $M_n(\mathscr{N}_0)$, and it is clear that the spectrum of A_0 is $\bigcup_{x \in \mathscr{N}_0} \varLambda[A(x)]$. As usual, following Dunford [5], $g(A_0) \in M_n(\mathscr{N}_0)$ can be defined as the sum of the p integrals $1/2\pi i \int_{\Gamma_i} g(\lambda)(A_0 - \lambda I)^{-1} d\lambda$, where Γ_i is the boundary of the set $f(\mathscr{D}_i)$. If we denote $B_0 = g(A_0)$, it follows from Theorem 2.10 of [5] that $f(B_0) = A_0$. Since this construction was carried out about an arbitrary point $x_0 \in \mathfrak{X}$, we can apply the compactness of $\mathfrak{X}$ to obtain points $x_1, \cdots, x_r \in \mathfrak{X}$ and compact open neighborhoods $\mathscr{N}_i$ of the x_i such that $\bigcup_{i=1}^{r} \mathscr{N}_i = \mathfrak{X}$ and such that the above construction has been carried out to yield a corresponding B_i on each $\mathscr{N}_i$. Furthermore, we can assume that the $\mathscr{N}_i$ are pairwise disjoint. The element $B \in M_n(\mathfrak{X})$ defined by $B(x) = B_i(x)$ for $x \in \mathscr{N}_i$ is such that $f(B) = A$, and the proof is complete.

COROLLARY 3.1. *If $\mathfrak{X}$ is a totally disconnected, compact Hausdorff space, then each invertible element of $M_n(\mathfrak{X})$ has a logarithm in $M_n(\mathfrak{X})$, and thus has roots of all orders in $M_n(\mathfrak{X})$.*

Proof. Observe first that the proof of Theorem 3 above goes through word for word in the case that $\mathfrak{X}$ is only compact Hausdorff and totally disconnected. Then observe that if $A \in M_n(\mathfrak{X})$ and an entire function f are given, in order to carry out the construction in the above proof to obtain a B such that $f(B) = A$, it suffices to know that for each ζ in the spectrum of A, there is a complex number z such that $f(z) = \zeta$ and $f'(z) \neq 0$. These observations complete the proof.

It results easily from Theorem 3 that if

$$\mathfrak{A} = \sum_{k=0}^{k_0} \oplus M_{n_k}(\mathfrak{X}_k)$$

is any finite C^*-sum of algebras $M_{n_k}(\mathfrak{X}_k)$ where the $\mathfrak{X}_k$ are Stonian spaces, then the entire functions which map $\mathfrak{A}$ onto itself are exactly those with property (K). However, if one considers algebras

$$\mathfrak{B} = \sum_{k=1}^{\infty} \oplus M_{n_k}(\mathfrak{X}_k)$$

which are C^*-sums of infinitely many $M_{n_k}(\mathfrak{X}_k)$ where $n_k \to \infty$ and the $\mathfrak{X}_k$ are only assumed to be compact Hausdorff spaces, then the situation is different, as is demonstrated by the following theorem.

THEOREM 4. *If $\mathfrak{B}$ is any algebra of the form*

$$\mathfrak{B} = \sum_{k=1}^{\infty} \oplus M_{n_k}(\mathfrak{X}_k)$$

where $n_k \to \infty$ and each $\mathfrak{X}_k$ is a compact Hausdorff space, then the entire functions which map $\mathfrak{B}$ onto itself are exactly those with property (B)

The proof of this theorem is patterned after an argument of Brown [1], and depends on the following lemma.

LEMMA 3.2. *Let f be any entire function, let $g(z)$ be the polynomial*

$$g(z) = \sum_{i=0}^{n-1} a_i z^i ,$$

and let $A \in M_n$ be the "analytic Toeplitz" matrix

$$A = \begin{bmatrix} a_0 & & & & & \\ a_1 & a_0 & & & & \\ a_2 & a_1 & a_0 & & & \\ \cdot & \cdot & \cdot & \cdot & & \\ \cdot & \cdot & \cdot & \cdot & \cdot & \\ a_{n-1} & \cdot & \cdot & \cdot & a_1 & a_0 \end{bmatrix}$$

Then $f(A)$ is an "analytic Toeplitz" matrix

$$f(A) = \begin{bmatrix} b_0 & & & & & \\ b_1 & b_0 & & & & \\ b_2 & b_1 & b_0 & & & \\ \cdot & \cdot & \cdot & \cdot & & \\ \cdot & \cdot & \cdot & \cdot & \cdot & \\ b_{n-1} & \cdot & \cdot & \cdot & b_1 & b_0 \end{bmatrix}$$

and the entire function $h(z) = f(g(z))$ *has a power series expansion*

$$h(z) = \sum_{i=0}^{\infty} \beta_i z^i$$

where $\beta_i = b_i$ *for* $0 \leqq i \leqq n-1$.

Proof. If f is any positive integral power of z, or more generally any polynomial, an inductive computation shows that the result is valid. For an arbitrary entire function f, let $p_n(z)$ be a sequence of polynomials which converges uniformly to f on every compact subset of the z-plane. Then, since $p_n(g(z))$ converges uniformly to $h(z)$ on compact subsets of the plane, the coefficients in the power series expansions of the $p_n(g(z))$ must converge to the corresponding coefficients in the power series expansion of $h(z)$. (See, for example, ([2], § 211)) Furthermore, since $p_n(A)$ converges to $f(A)$ in the norm topology of M_n, the entries of $p_n(A)$ must converge to the corresponding entires of $f(A)$, and the result follows.

Proof of Theorem 4. For convenience we take $n_k = n$. It will be clear that this does not affect the argument. Let

$$B = \left(\sum_{n=1}^{\infty} \oplus B_n\right) \in \mathfrak{B}$$

be defined by setting

$$B_n \equiv \begin{bmatrix} 0 & & & & & \\ 1 & 0 & & & & \\ & 1 & 0 & & & \\ & & \cdot & \cdot & & \\ & & & \cdot & \cdot & \\ & & & & 1 & 0 \end{bmatrix}$$

for each positive integer n. Let f be an entire function which maps onto $\mathfrak{B}$, and suppose that

$$A = \sum_n \oplus A_n$$

satisfies $f(A) = rB$ where r is some fixed positive real number. Since for any central projection $E \in \mathfrak{B}, f(EA) = Ef(A)$, it is clear that for each positive integer $n, f(A_n) = rB_n$. Now choose an arbitrary $x_n \in \mathfrak{X}_n$ for each integer n. The fact that $f[A_n(x_n)] = rB_n(x_n)$ follows just as in the proof of Theorem 3. Since $A_n(x_n)$ commutes with $B_n(x_n) = 1/r\, f[A_n(x_n)]$ and B_n is identically constant on $\mathfrak{X}_n$, a matrix calculation shows that for each positive integer n, the matrix $A_n(x_n)$ has the form

$$A_n(x_n) = \begin{bmatrix} a_0^n & & & & & \\ a_1^n & a_0^n & & & & \\ a_2^n & a_1^n & a_0^n & & & \\ \cdot & \cdot & \cdot & \cdot & & \\ \cdot & \cdot & \cdot & \cdot & \cdot & \\ a_{n-1}^n & \cdot & \cdot & \cdot & a_1^n & a_0^n \end{bmatrix}$$

where the a_i^n are of course complex numbers. Define the sequence $g_n(z)$ of polynomials by

$$g_n(z) = \sum_{i=0}^{n-1} a_i^n z^i ,$$

and let $h_n(z) = f(g_n(z))$. Since $f[A_n(x_n)] = rB_n(x_n)$, it follows from Lemma 3.2 that for each positive integer n, $h_n(z)$ is an entire function having a power series expansion

$$h_n(z) = rz + \sum_{k=n}^{\infty} \beta_k^n z^k .$$

Since $A = \sum_n \oplus A_n$ is a bounded operator, it follows that there exists a positive number M such that

$$\sum_{i=0}^{n-1} | a_i^n |^2 < M$$

for each n. Let $\mathscr{D}$ denote the disc $\mathscr{D} = \{z: |z| \leqq 1/2\}$ and observe that it follows from the above inequality that the sequence $g_n(z)$ is uniformly bounded on $\mathscr{D}$ by the number $2\sqrt{M}$. It follows from Montel's theorem ([2], § 416) that one can extract a subsequence $g_{n_k}(z)$ which converges uniformly on $\mathscr{D}$ to a function $g(z)$ which is analytic on $\mathscr{D}$. It follows that $h_{n_k}(z) = f(g_{n_k}(z))$ converges uniformly to $f(g(z))$ on $\mathscr{D}$, and by virtue of the form of the power series expansion of each $h_{n_k}(z)$, we must have $f(g(z)) = rz$ on $\mathscr{D}$. It is now clear that $g(z)$ is a Schlicht mapping of the interior $\mathscr{D}^\circ$ of $\mathscr{D}$ onto some bounded domain $g(\mathscr{D}^\circ)$ and that f is a Schlicht mapping of $g(\mathscr{D}^\circ)$ onto the open disc $\{z: |z| < r/2\}$. Since r was arbitrary, it follows from ([1], Theorem 2) that f has property (B), and the proof is complete.

4. Commutators in $M_n(\mathfrak{X})$. We introduce the notation $\sigma(B)$ for the trace in the usual sense of an $n \times n$ complex matrix B. In this section, we generalize another result known for M_n, and thereby set forth a class of operators on Hilbert space which are commutators. (See Remark 2 at the end of this section.) More precisely, we establish

THEOREM 5. *If $\mathfrak{X}$ is a Stonian space and $A \in M_n(\mathfrak{X})$, then A*

satisfies $\sigma[A(x)] \equiv 0$ *if and only if there are elements* B *and* C *in* $M_n(\mathfrak{X})$ *such that* $A = BC - CB$.

One half of the theorem is trivial; to prove the other half we use an idea suggested by Halmos in [6]. The crucial lemma is the following.

LEMMA 4.1. *If* $\mathfrak{X}$ *is any Stonian space and* $A \in M_n(\mathfrak{X})$ *is such that* $\sigma[A(x)] \equiv 0$, *then there is an invertible* $S \in M_n(\mathfrak{X})$ *such that* $SAS^{-1} = D = (d_{ij})$ *satisfies* $d_{11} \equiv 0$.

Proof. We consider collections $\{\mathscr{U}_i\}$ of disjoint, nonempty, compact open sets $\mathscr{U}_i \in \mathfrak{X}$ with the property that if $\mathscr{U}_i \in \{\mathscr{U}_i\}$, then there is an invertible $S_i \in M_n(\mathscr{U}_i)$ such that $\| S_i \|, \| S_i^{-1} \| \leqq 6$ and such that for each $x \in \mathscr{U}_i$, the matrix $S_iAS_i^{-1}(x)$ has a zero in the upper left hand corner. Let $\{\mathscr{U}_i\}_{i \in I}$ be a maximal such collection, and define $\mathscr{U} = \overline{\bigcup_{i \in I} \mathscr{U}_i}$. It follows from Lemma 2.1 of [3] that to complete the proof, it suffices to establish $\mathscr{U} = \mathfrak{X}$. Thus, suppose to the contrary that $\mathfrak{X} - \mathscr{U} \neq \phi$. According to Theorem 1 of [3] there exist functions $\lambda_1, \cdots, \lambda_n \in C(\mathfrak{X} - \mathscr{U})$ such that for $x \in \mathfrak{X} - \mathscr{U}$, the numbers $\lambda_1(x), \cdots, \lambda_n(x)$ are exactly the eigenvalues of $A(x)$. Furthermore, there must be at least one point $x_0 \in \mathfrak{X} - \mathscr{U}$ such that some $\lambda_i(x_0) \neq 0$. (Otherwise, we could apply Theorem 2 of [3] to obtain a unitary $U \in M_n(\mathfrak{X} - \mathscr{U})$ such that $UAU^*(x)$ is in upper triangular form for each $x \in \mathfrak{X} - \mathscr{U}$. Then the diagonal entries of $UAU^*(x)$ would be identically zero, and the maximality of the collection $\{\mathscr{U}_i\}_{i \in I}$ would be contradicted.) Since we know from the hypothesis that

$$\sum_{i=1}^{n} \lambda_i \equiv 0 ,$$

there must be at least two distinct i such that $\lambda_i(x_0) \neq 0$. In fact, a little thought convinces one that there exist λ_j and λ_k ($j \neq k$) such that

$$0 < | \lambda_j(x_0) | \leqq | \lambda_k(x_0) | < | \lambda_k(x_0) - \lambda_j(x_0) | .$$

It follows from the circle of ideas connected with the proof of Theorem 2 of [3] that there is a unitary element $U \in M_n(\mathfrak{X}) - \mathscr{U})$ such that $UAU^*(x) = (a_{ij}(x))$ is in upper triangular form for each $x \in \mathfrak{X} - \mathscr{U}$ and such that $a_{11} \equiv \lambda_k$ and $a_{22} \equiv \lambda_j$ on $\mathfrak{X} - \mathscr{U}$. Thus $0 < | a_{22}(x_0) | \leqq | a_{11}(x_0) | < | a_{11}(x_0) - a_{22}(x_0) |$, and by clever choice of U (i.e., by applying an additional rotation, and then changing notation) one can arrange things so that $| a_{11}(x_0) - a_{22}(x_0) | < | a_{12}(x_0) - [a_{11}(x_0) - a_{22}(x_0)] |$. It follows that for some $\delta, 0 < \delta < 1$, there is a compact open neighborhood $\mathscr{V} \subset \mathfrak{X} - \mathscr{U}$ of x_0 such that for $x \in \mathscr{V}, 0 < | a_{22}(x) | \leqq (1 + \delta) | a_{11}(x) | < | a_{12}(x) - [a_{11}(x) - a_{22}(x)] |$. The argument now splits into two cases.

Case I. For every $x \in \mathscr{V}$, $|a_{12}(x)| \geqq |a_{11}(x)|$. In this case we define an invertible $S = (s_{ij}) \in M_n(\mathscr{V})$ to be the direct sum of the 2×2 matrix $(s_{ij}: i, j \leqq 2)$ and the identity element of $M_{n-2}(\mathscr{V})$, where for $x \in \mathscr{V}$, $s_{11}(x) = s_{22}(x) = 1$, $s_{12}(x) = 0$, and $s_{21}(x) = a_{11}(x)/a_{12}(x)$. An easy calculation shows that $\|S\|, \|S^{-1}\| \leqq 4$, and another calculation shows that for $x \in \mathscr{V}$, the matrix $SUAU^*S^{-1}(x)$ has a zero in the upper left hand corner. The existence of $\mathscr{V}$ thus contradicts the maximality of the collection $\{\mathscr{V}_i\}_{i \in I}$, and we proceed to

Case II. There is a compact open subset $\mathscr{W} \subset \mathscr{V}$ such that for $x \in \mathscr{W}$, $|a_{12}(x)| < |a_{11}(x)|$. As before we define an invertible $S = (s_{ij}) \in M_n(\mathscr{W})$ to be the direct sum of the 2×2 matrix $(s_{ij}: i, j \leqq 2)$ and the identity element of $M_{n-2}(\mathscr{W})$. This time for $x \in \mathscr{W}$ we take $s_{11}(x) = s_{12}(x) = s_{21}(x) = [a_{11}(x)/\{a_{12}(x) - [a_{11}(x) - a_{22}(x)]\}]^{1/2}$ and $s_{22}(x) = s_{11}(x)[\{a_{12}(x) + a_{22}(x)\}/a_{11}(x)]$, where the exponent $1/2$ denotes any square root taken in such a way that $s_{11} \in C(\mathscr{W})$. (Theorem 1 of [3] enables us to take continuous square roots.) As a result of the inequalities which are valid on $\mathscr{W}$, one has $|s_{11}(x)| < 1$ and $|s_{22}(x)| \leqq 2 + \delta$ for each $x \in \mathscr{W}$; furthermore, $s_{11}s_{22} - s_{12}s_{21} \equiv 1$ on $\mathscr{W}$, and it follows that $\|S\|, \|S^{-1}\| \leqq 6$. Calculation shows that for $x \in \mathscr{W}$, $SUAU^*S^{-1}(x)$ has a zero for its upper left hand entry, and thus the proof is complete.

The following corollary follows easily by induction on n, and we omit its proof.

COROLLARY 4.2. *If $A \in M_n(\mathfrak{X})$ is such that $\sigma[A(x)] \equiv 0$, then there is an invertible $S \in M_n(\mathfrak{X})$ such that $SAS^{-1} = (a_{ij})$ satisfies $a_{ii} \equiv 0$ for $1 \leqq i \leqq n$.*

Proof of Theorem 5. We are given that $\sigma[A(x)] \equiv 0$. Choose $S \in M_n(\mathfrak{X})$ according to Corollary 4.2 so that $SAS^{-1} = (a_{ij})$ satisfies $a_{ii} \equiv 0$ for $1 \leqq i \leqq n$. Define $B_1 = (b_{ij}) \in M_n(\mathfrak{X})$ by $b_{ii} \equiv i$ for $1 \leqq i \leqq n$ and $b_{ij} \equiv 0$ for $i \neq j$. Also define $C_1 = (c_{ij}) \in M_n(\mathfrak{X})$ by $c_{ij} \equiv a_{ij}/(b_{ii} - b_{jj})$ for $i \neq j$ and $c_{ij} \equiv 0$ for $i = j$. If B and C are defined by $B = S^{-1}C_1S$, then it is easy to see that $B_1C_1 - C_1B_1 = SAS^{-1}$, or, what is the same thing, $BC - CB = A$.

REMARKS.

(1) A stronger version of Lemma 4.1, obtained from the present version by requiring S to be unitary, actually holds. The proof, however, uses a completely different idea and is much longer than the above proof.

(2) A bounded operator B on Hilbert space is called n-normal [9] if the W^*-algebra which B generates satisfies a polynomial identity

of the form

$$\sum (sgn\, \pi) X_{\pi(1)} X_{\pi(2)} \cdots X_{\pi(2n)} = 0 \,,$$

where the sum is taken over all permutations π on $2n$ objects. It is known that such a W^*-algebra is a finite direct sum of algebras each of which has a faithful C^*-representation as some $M_k(\mathfrak{X}_k)$ with $\mathfrak{X}_k$ Stonian and $k \leqq n$. Furthermore such a W^*-algebra has a well-behaved center-valued trace function, so that Theorem 5 can be paraphrased: Any n-normal operator with trace zero is the commutator of a pair of n-normal operators.

(3) There are at least two classes of operators on Hilbert space which possess well-behaved numerical traces. These are operators in the trace-class [13], and operators in W^*-algebras which are factors of type II_1. Is it true that every operator with trace zero in one of these classes is a commutator?

5. **Two examples.** In this section we set forth two examples which show that Theorem 2 of [3] and Theorems 1 and 2 of the present paper cannot be extended to the setting in which $\mathfrak{X}$ is assumed only to be a compact Hausdorff, totally disconnected space. In these examples we take $\mathscr{T}$ to be the compact Hausdorff, totally disconnected space consisting of the set $\{a_1, a_2, \cdots, a_n, \cdots, 0\}$ with the relative topology, where the real sequence $\{a_n\}$ is strictly decreasing to zero and satisfies $\cos(1/a_n) = \sin(1/a_n) = 1/\sqrt{2}$ for n odd and $\cos(1/a_n) = 1$, $\sin(1/a_n) = 0$ for n even.

EXAMPLE 1. (This example is essentially due to Rellich [11].) Define $A \in M_2(\mathscr{T})$ by

$$A(a_n) = \begin{pmatrix} 1 - a_n \cos(2/a_n) & -a_n \sin(2/a_n) \\ -a_n \sin(2/a_n) & 1 + a_n \cos(2/a_n) \end{pmatrix};$$
$$A(0) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Then, even though A is Hermitian, *there exists no unitary* $U \in M_2(\mathscr{T})$ *such that* $UAU^*(t)$ *is in upper triangular form for each* $t \in \mathscr{T}$.

Proof. Assume that such a $U = (u_{ij})$ exists, and let $UAU^*(t) = (b_{ij}(t))$. Then the $b_{ij} \in C(\mathscr{T})$, and the vector $(\bar{u}_{11}(t), \bar{u}_{12}(t)) = V(t)$ has length one at each $t \in \mathscr{T}$ and has entries which are elements of $C(\mathscr{T})$. Futhermore, it is easy to see that $[\mathrm{A}(t) - b_{11}(t)I]V(t) \equiv 0$. In other words, the vector $V(t)$ is a continuous eigenvector for $A(t)$ corresponding to the eigenvalue $b_{11}(t)$. An easy calculation shows that the eigenvalues of $A(a_n)$ are $1 - a_n$ and $1 + a_n$, so that for each n, $b_{11}(a_n) = 1 - a_n$ or $b_{11}(a_n) = 1 + a_n$. Furthermore, it is easy to see that the vector $(\cos(1/a_n), \sin(1/a_n))$ is an eigenvector for $A(a_n)$ corresponding to the eigenvalue $1 - a_n$, and the vector $(\sin(1/a_n), -\cos(1/a_n))$ is an eigenvector for $A(a_n)$ corresponding to the eigenvalue $1 + a_n$. It follows that for n odd, we must have $|\bar{u}_{11}(a_n)| = 1/\sqrt{2}$, and for n even, we must have $|\bar{u}_{11}(a_n)| = 0$ or 1. This contradicts $u_{11} \in C(\mathscr{T})$, and completes the proof.

EXAMPLE 2. Define $A, B \in M_2(\mathscr{T})$ by $A(0) = B(0) = 0$ and

$$A(a_n) = \begin{pmatrix} 0 & a_n \\ 0 & 0 \end{pmatrix}, \qquad B(a_n) = \begin{pmatrix} 0 & (-1)^n a_n \\ 0 & 0 \end{pmatrix}.$$

Then $A(t)$ is unitarily equivalent to $B(t)$ for each $t \in \mathscr{T}$, but *there exists no invertible* $S \in M_2(\mathscr{T})$ *such that* $SAS^{-1} = B$.

Proof. Suppose such an invertible $S = (s_{ij}) \in M_2(\mathscr{T})$ does exist. Then $SA = BS$, and calculation shows that $s_{21} \equiv 0$. Furthermore, $s_{11}(a_n) = (-1)^n s_{22}(a_n)$ for each n, and since S is invertible and $s_{21} \equiv 0$, s_{11} and s_{22} are bounded away from zero. It follows that s_{11} and s_{22} cannot both be continuous at zero, a contradiction.

REMARK. While the theory of elements $A \in M_n(\mathfrak{X})$ is not very satisfactory for $\mathfrak{X}$ only totally disconnected, it is nevertheless true that A has continuous eigenvalues [4].

BIBLIOGRAPHY

1. A. Brown, *Entire functions on Banach algebras*, Michigan Math. J., **10** (1963), 91-96.
2. C. Caratheodory, *Theory of Functions vols. I and II*, Chelsea, New York (1954).
3. D. Deckard and C. Pearcy, *On matrices over the ring of continuous complex-valued functions on a Stonian space*, Proc. Amer. Math. Soc., **14** (1963), 322-328.
4. ———. *On algebraic closure in function algebras*, Proc. Amer. Math. Soc., **15** (1964), 259-263.
5. N. Dunford, *Spectral theory I. Convergence to projections*, Trans. Amer. Math. Soc., **54** (1943), 185-217.
6. P. R. Halmos, *Finite Dimensional Vector Spaces*, Princeton, Van Nostrand (1958).
7. I. Kaplansky, *Algebras of type I*, Ann. of Math., **56** (1952), 460-472.
8. S. Kurepa, , to appear in Publications de 1'Inst. Math., Srpska Akad., Nauk, Belgrad.
9. C. Pearcy, *A complete set of unitary invariants for operators generating finite W*-algebras of type I*, Pacific J. Math., **12** (1962), 1405-1416.
10. ———, *On unitary equivalence of matrices over the ring of continuous complex-valued functions on a Stonian space*, Canad. J. Math., **15** (1963), 323-331.
11. F. Rellich, *Perturbation theory of eigenvalue problems*, lecture notes, New York University (1953).

12. C. Rickart, *General theory of Banach algebras*, Princeton, Van Nostrand (1960).
13. R. Schatten, *Norm ideals of completely continuous operators*, Berlin, Springer-Verlag (1960).
14. M. H. Stone, *Boundedness properties in function-lattices*, Canad. J. Math., **1** (1949), 176-186.
15. T. Yen, *Trace on AW^*-algebras*, Duke J. Math., **22** (1955), 207-222.

HUMBLE OIL & REFINING COMPANY
HOUSTON, TEXAS

ANOTHER CHARACTERIZATION OF THE n-SPHERE AND RELATED RESULTS

R. F. DICKMAN, L. R. RUBIN AND P. M. SWINGLE

In [5] we defined an irreducible $B(J)$-cartesian membrane and an excluded middle membrane property EM, and used these to characterize the n-sphere. There the class $B(J)$ was of $(n-1)$-spheres contained in a compact metric space S. Since part of the proof does not depend upon the fact that elements of $B(J)$ are $(n-1)$-spheres, we consider the possibility of other entries in the class $B(J)$. Recent developments in this direction have been made by Bing in [2] and by Andrews and Curtis in [1]. In [3] and [4] Bing constructed a space B not homeomorphic with E^3, which has been called the dogbone space. By Theorem 6 of [2], the sum of two cones over the one point compactification $\bar{B}$ of B is homeomorphic with S^4. This sum of two cones over a common base X is called the suspension of X.

In [1] Andrews and Curtis showed that if α is a wild arc in S^n that the decomposition space S^n/α is not homeomorphic with S^n. They proved, however, that the suspension of S^n/α is always homeomorphic with S^{n+1} for any arc $\alpha \subset S^n$. The reader will easily see that a class $\bar{B}$ or of S^n/α as described will satisfy the conditions for a class $B(J)$ for which an n-sphere will have property EM.

The results below were obtained in considering such spaces, and Theorem 1 below is a weaker characterization of the n-sphere than is Theorem 2 of [5]. We find it difficult to determine the properties $J \in B(J)$ must have for S to have Property EM, as is shown by our Theorem 4 below.

I. Definition and basic properties. Let S always be a compact metric space and let $B(J)$ be a class of mutually homeomorphic subcontinua of S. We put conditions on this general class $B(J)$ in our theorems below.

We define a $B(J)$-cartesian membrane as we did in [5] and [6]. Let F be a compact subset of S containing $J \in B(J)$. Let M be a subcontinuum of $F, b \in M$ and C be homeomorphic to J. Denote by $(C \times M, b)$ the decomposition space [10: pp 273–274] of the upper semi-continuous decomposition of the cartesian product $C \times M$, where the only nondegenerate element is taken to be $C \times b$ (intuitively the decomposition space is a sort of generalized cone with vertex at the point $C \times b$). With this notation we give:

Received September 18, 1963. This work was done under National Science Foundation Grant G 19672.

DEFINITION 1. We say that F is a $B(J)$-*cartesian membrane* from b to J (or for brevity with base J) if and only if there is a homeomorphism h from $(C \times M, b)$ onto F for some M such that:

(i) for some $a \in M - b$, $J = h(C \times a)$,

(ii) for all $q \in M - b$, $h(C \times q) \in B(J)$, and

(iii) $h(C \times b) = b$.

If M is irreducible from a to b, then we prefix the above definition by *irreducible*. Whenever F is a $B(J)$-cartesian membrane and $F = h(C \times m, b)$, h is assumed to be a homeomorphism from $(C \times M, b)$ onto F with properties (i), (ii) and (iii). We say b is the *vertex* of F and J is the *base* of F.

The definition of $B(J)$-cartesian membrane is rather general; for example, a point or any continuum can be taken as a $B(J)$-cartesian membrane. We shall place restrictions on the space S to limit possibilities such as these when the need arises. The excluded middle membrane property of Theorem 2 in [5] is the following:

Property EM. We say that the space S has *Property EM* with respect to the class $B(J)$ if the following hold:

(1) The class $B(J)$ is not empty;

(2) For each $J \in B(J)$, $S = F_1 + F_2$ where F_1 and F_2 are irreducible $B(J)$-cartesian membranes with base J, such that $F_1 \not\subset F_2$ and $F_2 \not\subset F_1$ and whenever S is such a union and F_3 is any other $B(J)$-cartesian membrane containing J, then F_3 contains F_1 or F_2 but not both; and

(3) If $J \in B(J)$ and $p \in S - J$, then there exists a $B(J)$-cartesian membrane from p to J.

Below F, F', F_1 and F_2 are always irreducible $B(J)$-cartesian membranes.

We proved in [5] that when $B(J)$ is a class of $(n-1)$-spheres and $n > 1$ that:

(A) A necessary and sufficient condition that S be an n-sphere is that S have Property EM.

We observed in our proof of (A) that if S had Property EM with respect to a class of mutually homeomorphic continua, we were able to prove:

(B) That whenever $S = F_1 + F_2$ where F_1 and F_2 have base J, $F_1 \cdot F_2 = J$;

(C) If $F = h(C \times M, b)$ was an irreducible $B(J)$-cartesian membrane, then M was always a simple continuous arc with b as endpoint; and

(D) If $S = F_1 + F_2$ where F_1 and F_2 have base J and F_3 is any other irreducible $B(J)$-cartesian membrane with base J, then $F_1 = F_3$ or $F_2 = F_3$.

In the first paragraph of the proof of Theorem 2 of [5], (D) appeared easily as result (R_1); then by a long proof we showed that $F_1 \cap F_2 = J$, which is (B) above, and we note this long proof only depends upon J being a continuum, not on J being an $(n-1)$-sphere. Finally, the following argument show that (C) holds. Let $S = F_1 + F_2$, where F_1 and F_2 are irreducible $B(J)$-cartesian membranes with base J. By (B) $F_1 \cdot F_2 = J$, and so every element of $B(J)$ separates S. Then if $F_1 = h(C \times M, b)$ where M is irreducible from a to b, and if $z \in M - a - b$, $h(C \times z) \in B(J)$ by (ii) of Definition 1 above. Hence $h(C \times z)$ separates S, and therefore separates F_1. This implies z separates M, and so M is a simple continuous arc, as desired in (C).

II. Characterization of the n-sphere, for $n > 1$. We give now several lemmas that will enable us to characterize the n-sphere.

NOTATION. For a subset K of S, we will use $cl(K)$ to denote the closure of K in S, and for an open subset U of S, we will use $Fr(U)$ to denote the set $cl(U) - U$.

LEMMA 1. *If S has Property EM, then S is homogeneous.*

Proof. Let $x, y \in S$, $x \neq y$, and let J be an element of $B(J)$ such that $J \subset S - x - y$. By (3) of Property EM there exists an irreducible $B(J)$-cartesian membrane $F = h(C \times M, x)$ from x to J and by (D) and (2) of Property EM, $S = F + F'$, where F' has base J. Now by (B) each $J' \in B(J)$ separates S, hence by (ii) of Definition 1, some $J_0 = h(C \times q)$ separates x from y. Then by (2) of Property EM, $S = F_1 + F_2$ where F_1 and F_2 have base J_0. From (D) and (3) of Property EM there exists h_1 and h_2 such that $F_1 = h_1(C \times M_1, x)$ and $F_2 = h_2(C \times M_2, y)$. From (C) M_1 and M_2 are simple continuous arcs and x y are endpoints of M_1 and M_2 respectively. Hence from (B) there exists a homeomorphism from S onto S that carries x onto y; therefore S is homogeneous [10: p 378].

A topological space X is *invertible* [7] if for each nonempty open set U in X there is a homeomorphism h of X onto itself such that $h(X - U)$ lies in U.

LEMMA 2. *If S has Property EM then S is invertible.*

Proof. For any open set U in S and any point $x \in U$, some $J \in B(J)$ separates x from $Fr(U)$; then if $S = F_1 + F_2$ where F_1 and F_2 have base J, we can find a homeomorphism as in Lemma 1, that maps S onto S such that F_1 maps onto F_2 and F_2 maps onto F_1, hence $(S - U)$ into U.

THEOREM 1. *Let $n > 1$ and let each element of $B(J)$ contain a point at which it is locally euclidean of dimension $(n-1)$. Then S is an n-sphere if and only if S has Property EM.*

Proof of the sufficiency. Let $J \in B(J)$ and let x be an element of J at which J is locally euclidean of dimension $(n-1)$. Let U be an open $(n-1)$-cell neighborhood of x in J. Let $F = h(C \times M, b)$ have base J. By (C), M is an arc, and if V is an open subinterval of M containing a point y, $h(U \times V)$ is an open n-cell neighborhood of $h(x, y)$ in F. Since $h(U \times V)$ misses J, $h(U \times V)$ is open in $F - J$, and hence in S. By Lemma 1, S is homogeneous; hence every element of S has an open n-cell neighborhood, and so S is n-manifold. Doyle and Hocking in Theorem 1 of [7], have shown that if S is an invertible, n-manifold, then S is an n-sphere; hence by Lemma 2, S is an n-sphere.

The proof of the necessity is identical to that of Theorem 2 in [5].

Because 0-spheres are not connected the above proof does not hold for $n = 1$. We refer the reader to Theorem 1 of [5] for a characterization of the 1-sphere by an excluded middle membrane principle.

III. Related results.

LEMMA 3. *If S has Property EM then S is locally connected.*

Proof. We note that if F is an irreducible $B(J)$-cartesian membrane with base J, then $F - J$ is an open connected set in S, and proceed as in the proof of Lemma 2.

LEMMA 4. *If S has Property EM and $J \in B(J)$ then J is locally connected.*

Proof. Let $S = F_1 + F$ where F_1 and F have base J and $F = h(C \times M, b)$, where M is an arc from a to b; and $h(C \times a) = J$ as in (1) of Definition 1. Since S is locally connected, the open set $F - J - b$ is locally connected. We define $f(h(c, m)) = h(c, a)$, where $h(c, m)$ is a point in $F - J - b$; then f is a projection onto J and can easily be proved to be continuous and open. Since $F - J - b$ is locally connected and local connectedness is preserved under open, continuous mappings, J is locally connected.

THEOREM 2. *If S has Property EM and $J \in B(J)$, then J contains a 1-sphere.*

Proof. Let $J \in B(J)$, and $F = h(C \times M, b)$ have vertex $b = h(C \times b)$ and base J. Since J is locally connected, C must contain an arc I;

and by (C), M is an arc. Then the set $E' = h(I \times M, b)$ is a closed 2-cell contained in F. Let E be any subset of E' that is homeomorphic to euclidean 2-space E^2.

Let b_i $(i = 1, 2, \cdots)$ be a sequence converging to b in M. Then the half open intervals $M_i = bb_i - b_i$ form a basis of open sets in M at b, and the sets $U_i(b) = h(C \times M_i, b)$ form a basis of open sets in F at b. These open sets have the property that $Fr(U_i(b))$ is homeomorphic to J.

Choose $x \in E$, then $x \notin J$. By the homogeneity of S there exists a basis of open sets $U_i(x)$ which have the property that their boundaries are homeomorphic to J. Now fix i such that $U = U_i(x) \cdot E$ has a compact closure in E. Let V be the component of U that contains x. Since E is locally connected, V is open in E. Also $Fr(V) \subset Fr(U_i(x))$; therefore without loss of generality we can think of $Fr(V)$ as being a subset of J. Let V' be a component of $E - cl(V)$. Then V' is an open connected subset of E and $Fr(V') \subset Fr(V)$. Since $Fr(V')$ is closed and $Fr(V)$ compact, $Fr(V')$ is compact. By Theorem 25 of [10: p 176], $Fr(V')$ is a continuum. Then by Theorem 28 of [10: p 178], $Fr(V')$ is not disconnected by the omission of any point.

Let $r, s \in Fr(V')$, and let Y be an arc from r to s in J. Let $q \in Y - r - s$; now q does not separate r from s in $Fr(V')$; hence q does not separate r from s in J; then there exists an arc Y' from r to s in J that does not contain q, and $Y + Y'$ must contain a 1-sphere.

REMARK. Since J is locally connected, J is arcwise connected and as such cannot be an indecomposable continuum; by Theorem 2, J cannot be hereditarily unicoherent. A simple proof using the Brouwer Invariance of Domain Theorem [9: p. 95] will show that J cannot be a closed n-cell.

LEMMA 5. *Let S be an n-sphere having Property EM with respect to some $B(J)$.* (1) *If G is an $(n-2)$-sphere in $J \in B(J)$, then $J - G$ is not connected*; (2) *if E is a closed $(n-2)$-cell in J, then $J - E$ is connected.*

Proof. (1) Suppose $J - G$ is connected. Let $S = F_1 + F_2$ where F_1 and F_2 have base J; by (B) and (C) we can find h_1 and h_2 such that $F_1 = h_1(J \times M_1, b_1)$, $F_2 = h_2(J \times M_2, b_2)$ and $h_1 \mid (J \times a) = h_2 \mid (J \times a)$ where M_1 and M_2 are arcs from a to b_1 and a to b_2 respectively. Then $K = h_1((J - G) \times (M_1 - b_1)) + h_2((J - G) \times (M_2 - b_2))$ is connected. But $S - K = h_1(G \times M_1, b_1) + h_2(G \times M_2, b_2)$ is an $(n-1)$-sphere is S and must disconnect S by the Jordan Separation Theorem [9: p. 101].

The proof of (2) is similar to that of (1).

THEOREM 3. *A necessary and sufficient condition that S be a 3-sphere is that S have Property EM if and only if $B(J)$ is a collection of 2-spheres.*

Proof. The sufficiency follows from Theorem 2 of [5].

By Theorem 2, every $J \in B(J)$ contains a 1-sphere, and by (1) of Lemma 5 every 1-sphere in J separates J. By (2) of Lemma 5 no proper subcontinuum of a 1-sphere in J separates J; and by Lemma 4, J is locally connected; therefore by Zippin's Characterization in [11: p. 88] J is a 2-sphere. The rest follows from Theorem 2 of [5].

We need Hypothesis:

(H 1) If F_c, F_b and F'' are irreducible $B(J_0)$-cartesian membranes with base J_0 then $F_c + F_b + F''$ is contained in some E^3;

(H 2) If $S_x = F_x + F''$ is a 2-sphere in E^3, x is vertex of $B(J_0)$-cartesian membrane F_x and $t'_\alpha = h_c(c_\alpha \times M'', x)$ $(c_\alpha \in C)$ is a projecting arc from x to J through a point $y \in \operatorname{int}(S_x, E^3)$, (the interior of S_x in E^3), then $t'_\alpha - x \subset \operatorname{int}(S_x, E^3)$; if $q \in \operatorname{int}(S_x, E^3) \cdot J = J'$, then $q \notin \operatorname{cl}(J - J')$.

THEOREM 4. *Let S have Property EM, let* (H 1) *and* (H 2) *hold and let there exist a region R in S such that $J \cdot R$ contains a 1-sphere J_0 and $R \cdot J$ is embedded in the euclidean E^2; let there exist $q \in J - R$. Then J contains a closed 2-cell with J_0 as boundary.*

Proof. By (2) of Property *EM* there exist irreducible $B(J)$-cartesian membranes such that $S = h(C \times M, b) + h'(C \times M', b')$ where $h \mid (C \times a) = h' \mid (C \times a)$ and M, M' are arcs from a to b and a to b' respectively; since $J \supset J_0$, there exists $C_0 \subset C$ homeomorphic to J_0; let $h(C_0 \times M, b) = F_b$ and $h'(C_0 \times M', b') = F''$, where then F_b and F'' are irreducible $B(J_0)$-cartesian membranes from J_0 to b and b' respectively. Let $S_b = F_b + F''$; by Theorem 2 of [5], S_b is a 2-sphere.

By hypothesis there exists $q \in J - R$; thus $q \notin S_b$, and so by (H 2) the projecting arc from b to q does not contain a point of $\operatorname{int}(S_b, E^3)$; let c be an element of this projecting arc. By (3) of Property *EM*, there exists an irreducible $B(J_0)$-cartesian membrane $F_c = h_c(C_0 \times M_c, c)$ with base J_0, a subset of an irreducible $B(J)$-cartesian membrane $h_c(C \times M_c, c)$ from c to J; by the choice of c, $h_c(C \times M_c, c) = h(C \times M, b)$ and thus $S_c = F_c + F''$ is a 2-sphere.

Since $c \notin \operatorname{int}(S_b, E^3)$, there exists a region R' about c such that $cl(R') \cdot S_b = \phi$; then by Lemma 3 of [6] there exists an irreducible $B(J)$-cartesian membrane $F_{0c} = h_c(C \times M'_c, c)$, for $M'_c \subset M_c$, such that $F_c \cdot R' \supset F_{0c}$.

Let $\{t_{\alpha c}\}$ be the class of all projecting subarcs from c to J which are contained in $(S_c - (F_{0c} - J_c')) + \text{int}(S_c, E^3) - (F_{0c} - J_c')$, where J_c' is the base of F_{0c}; that is $t_{\alpha c}$ is an arc from J to F_{0c} in and on S_c.

Let $Z' = \cup t_{\alpha c}$ and let $Z = Z' \cdot J$. Suppose $Z' = Z_1' + Z_2'$ separate [11: p. 8]. Since each $t_{\alpha c}$ is connected, each is contained wholly in Z_1' or in Z_2'; this is also true of J_0 and so of $F_c - F_{0c}$; so let $Z_1' \supset F_c - F_{0c} \supset J_0$.

By Theorem 5.37 of [11: p. 66] S_c is arcwise accessible from the embedding E^3; hence there exists an arc cb' such that $cb' - c - b' \subset \text{int}(S_c, E^3)$. But cb' contains a point of $\text{int}(S_b, E^3)$ and a point c of $S - \text{int}(S_b, E^3) - S_b$; hence cb' contains some $v \in S_b$, because by the Jordan-Brouwer Separation Theorem [11: Theorem 5.23, p. 63] S_b separates E^3 into two domains. Hence by (2) of Property EM there exists a projecting arc from c to J through v, and so some $t_{\alpha c} \supset v$ and $Z' \supset t_{\alpha c}$. Let $Z_i = Z_i' \cdot Z (i = 1, 2)$, where by agreement $Z_1 \supset J_0$. By hypothesis $J \cdot R$ is contained in some euclidean E^2, and so let E be the 2-cell bounded by J_0 in this E^2. Thus $J_0 + E \supset Z$, and because of v above $E \cdot Z \neq \phi$. If $j \in J \cdot E$, by (H 2) the projecting arc cj is such that $cj - c \subset \text{int}(S_c, E^3)$. Thus $j \in Z$, and so $Z = J_0 + J \cdot E = Z_1 + Z_2$ separate. Hence $J = (Z_1 + (J - E)) + Z_2$ separate, which is a contradiction, since J is a continuum. Therefore Z and Z' are connected. By Lemma 4 J is locally connected, and so by (H2) Z is also.

Since Z is closed, Z contains all of its boundary points in the space J. By the Torhorst Theorem [10: p. 191, Theorem 42], the boundary of any complementary domain of Z in E must be a 1-sphere J_0'. Using J_0' in place of J_0, one obtains a 2-sphere S_c' with poles c and b' and with J_0' as a base in S_c'. Thus an arc bc' above exists such that $cb' - c - b' \subset \text{int}(S_c', E^3)$ and there exists a point $v \in S_b \cdot cb'$; also there exists $t_{\alpha c}$ as above, now contained in the $\text{int}(S_c', E^3)$; hence an endpoint of $t_{\alpha c}$ is an element of $\text{int}(J_0', E^2)$; thus a point of Z is in the complementary domain above of Z in E, which is a contradiction. Therefore $Z = E$, and so J contains a closed 2-cell.

If (H 1) and (H 2) hold, J cannot be a plane universal curve.

References

1. J. J. Andrews and M. L. Curtis, *n-Space modulo an arc*, Annals of Math., **75** (1962), 1-7.

2. R. H. Bing, *The cartesian product of certain nonmanifold and a line is E^4*, Annals of Math., **70** (1959), 399-412.

3. ———, *A decomposition of E^3 into points and tame arcs such that the decomposition space is topologically different from E^3*, Annals of Math., **65** (1957), 484-500.

4. ———, *The cartesian product of a certain nonmanifold and a line is E^4*, Bull. Amer. Math. Soc., **64** (1958), 82-84.

5. R. F. Dickman, L. R. Rubin and P. M. Swingle, *Characterization of n-spheres by an excluded middle membrane principle*, Mich. Math. Jour., **11** (1964), 53-59.

6. ———, *Irreducible continua and generalization of hereditarily unicoherent continua*

by means of membranes, to appear.
7. P. H. Doyle and J. G. Hocking, *A characterization of euclidean n-space*, Mich. Math. Jour., **7** (1960), 199–200.
8. ———, *Invertible spaces*, Amer. Math. Monthly **68** (1961), 959–965.
9. W. Hurewicz and H. Wallman, *Dimension theory*, Princeton University Press (1948).
10. R. L. Moore, *Foundations of point set theory*, Amer. Math. Soc. Colloquium Publications **13** (1962).
11. R. L. Wilder, *Topology of Manifolds*, Amer. Math. Soc. Colloquium Publications **32** (1948).

MIAMI, FLORIDA

A NOTE ON REFLEXIVE MODULES

E. E. ENOCHS

For any ring A and left (resp. right) A-module E we let E^* denote the right (resp. left) A-module $\mathrm{Hom}_A(E, A_s)$ (resp. $\mathrm{Hom}_A(E, A_d)$) where A_s (resp. A_d) denotes A considered as a left (resp. right) A-module. Then the mapping $E \to E^{**}$ such that $x \in E$ is mapped onto the mapping $\varphi \to \varphi(x)$ is linear.

Specker [3] has shown that if E is a free Z-module with a denumerable base (where Z denotes the ring of integers) then E is reflexive, i.e. the canonical homomorphism $E \to E^{**}$ is a bijection. In this paper it is shown that a free module E with a denumerable base over a discrete valuation ring A is reflexive if and only if A is not complete and if and only if E is complete when given the topology having finite intersections of the kernels of the linear forms as a fundamental system of neighborhoods of 0. Specker's result can be deduced from these results. We note that this topology has been used and studied by Nunke [2] and Chase [1].

THEOREM 1. *Let A be a discrete valuation ring with prime Π and let E be a free A-module with a denumerable base. Then E is reflexive if and only if A is not complete.*

Proof. Let $(a_i)_{i \in N}$ (N the set of natural numbers) be a base of E and let $E_j = \{\varphi \mid \varphi \in E^*, \ \varphi(a_i) = 0, \ i = 0, 1, 2, \cdots, j-1$. Let $a'_j \in E^*$ be such that $a'_j(a_j) = 1$, $a'_j(a_k) = 0$ if $j \neq k$. Then clearly a'_0, $a'_1, \cdots a'_{j-1}$ generate a supplement of E_j in E^*. For each $x \in E$ the canonical image of x in E^{**} annihilates some E_j and conversely if $\psi \in E^{**}$ annihilates E_j then ψ is the canonical image of $\sum_{i=0,1,\cdots,j-1} \psi(a'_i) a_i$. Hence $E \to E^{**}$ is a surjection if and only if each $\psi \in E^{**}$ annihilates some E_j. If $E \to E^{**}$ is not a surjection let $\psi \in E^{**}$ be such that $\psi(E_j) \neq 0$ for each $j \in N$ and let $\varphi_j \in E_j$ be such that $\psi(\varphi_j) \neq 0$. We can suppose that $\varphi_j \in \Pi^j E_j$ and that $\psi(\varphi_j) \in \Pi^{m}_{j} A$ but $\psi(\varphi_j) \notin \Pi^{m_j+1} A$ where $m_{i+1} > m_i$ for all $i \in N$. To show A complete it suffices to show that every series $\sum_{j \in N} \beta_j \Pi^{m_j}$, $\beta_j \in A$ converges. We can find a scalar multiple of φ_j say φ'_j such that $\psi(\varphi'_j) = \beta_j \Pi^{m}_{j}$. Then let $\varphi \in E^*$ be such that $\varphi(x) = \sum_{j \in N} \varphi'_j(x)$ for all $x \in E$. This sum is defined since for a fixed $x \in E$ and M sufficiently large positive integer we have $\varphi_{M+i}(x) = 0$ for all $i \in N$. Furthermore, since $\varphi'_j \in \Pi^j E_j$ it is clear that the series $\sum \varphi'_j$ converges to φ when E^* is given the topology having

Received December 6, 1963.

the submodules $\Pi^n E^*$, $n \in N$ as a fundamental system of neighborhoods of 0. Under this topology $\psi: E^* \to A$ is continuous. Hence

$$\sum_{j \in N} \psi(\varphi_j') = \sum_{j \in N} \beta_j \Pi^{m_j}$$

converges to $\psi(\varphi)$. Thus A is complete.

Conversely if A is complete let $(a_i')_{i \in N}$ as defined above be a subfamily of the family $(a_i')_{i \in N_1}$, $N_1 \supset N$ where $(a_i' + \Pi E^*)_{i \in N_1}$ is a base of the $A/\Pi\ A$ module $E^*/\Pi\ E^*$. Then if E' is the submodule of E^* generated by the family $(a_i')_{i \in N_1}$ it is easy to see that E' is free with base $(a_i')_{i \in N_1}$ and that E' is a dense pure submodule of E^*, i.e. E^*/E' is divisible and torsion free. Then, since A is complete the map $E^{**} \to E'^*$ which maps an element of E^{**} onto its restriction to E' is a bijection. But this clearly implies the existence of a $\psi \in E^{**}$ such that $\psi(a_i') \neq 0$ for all $i \in N_1$ and hence for all $i \in N$. Thus $E \to E^{**}$ is not a surjection.

COROLLARY. *If A is an integral domain with a prime Π such that the discrete valuation ring A_π is not complete then free A-modules with denumerable bases are reflexive.*

Proof. There exist canonical injections of E, E^* and E^{**} in E_π, E_π^*, and E_π^{**} and furthermore if for $x \in E$, $\varphi \in E^*$, and $\psi \in E^{**}$ we let $\bar{x}$, $\bar{\varphi}$, and $\bar{\psi}$ denote the image of x, φ, and ψ in E_π, E_π^*, and E_π^{**} then $\varphi(x) = \bar{\varphi}(\bar{x})$ and $\psi(\varphi) = \bar{\psi}(\bar{\varphi})$. Then if $(a_i)_{i \in N}$ is a base of E, $(\bar{a}_i)_{i \in N}$ is a base of E_π and if $(a_i')_{i \in N}$ is defined as above we get $\bar{a}_i'(\bar{a}_i) = 1$, $\bar{a}_i'(\bar{a}_j) = 0$ if $i \neq j$. Then if $\psi \in E^{**}$ is such that $\psi(E_j) = 0$ for each j then $\bar{\psi}$ is not in the image E_π under the canonical homomorphism since $\bar{\psi}((E_\pi)_j) \neq 0$ where E_j and $(E_\pi)_j$ are defined as above.

THEOREM 2. *If A is a left Noethrian hereditary ring, then a left A module E is reflexive if and only if E is complete when endowed with the topology having the finite intersections of the kernels of the linear forms as a fundamental system neighborhoods of 0.*

Proof. Clearly E is separated with the topology described in the theorem if and only if the map $E \to E^{**}$ is an injection hence we suppose that E is separated. For each finite subset X of E^* consider the subset X° of E^{**} consisting of all $\psi \in E^{**}$ such that $\psi(X) = 0$. Let E^{**} be endowed with the topology having the submodules X° as a fundamental system of neighborhoods of 0 where X ranges through all finite subsets of E^*. Then it is immediate that E^{**} is complete with this topology. If we can establish that the canonical map $E \to E^{**}$ maps E isomorphically onto a dense subset of E^{**} then it will follow immediately that E is complete if and only if E is reflexive.

Let X be a finite subset of E^*. Then clearly the intersection of the kernels of the elements in X is mapped onto the intersection of X° with the canonical image of E in E^{**} hence E is mapped isomorphically onto a subset of E^{**}. Thus it only remains to prove that the image of E in E^{**} is dense in E^{**}. If $\psi \in E^{**}$ and $X = \{\varphi_1, \varphi_2, \cdots \varphi_n\}$ is a finite set of elements of E^* consider the map $E \rightarrow \prod_{i=1,\cdots,n} A_i$ such that $x \rightarrow (\varphi_i(x))_{i=1,\cdots,n}$ where $A_i = A_s$. Since A is left hereditary the kernel of this map $E_1 = \bigcap_{i=1,\cdots,n} \varphi_i^{-1}(0)$ is a direct summand of E so let $E = E_1 + E_2$ (direct). Then since A is left Noetherian E_2 is a finitely generated projective module so it is relfexive. Now $E^* = E_i^\circ + E_2^\circ$ (direct) and $E^{**} = E_1^{\circ\circ} + E_2^{\circ\circ}$ (direct). Clearly $E_2^{\circ\circ}$ is isomorphic to E_2^{**} and the restriction of the canonical homomorphism $E \rightarrow E^{**}$ maps E_2 isomorphically onto $E_2^{\circ\circ}$. If $\psi = \psi_1 + \psi_2$ where $\psi_1 \in E_1^{\circ\circ}$ let $x \in E_2$ be such that $x \rightarrow \psi_2$ under the map $E \rightarrow E^{**}$. Then since $\psi - \psi_2 \in E_1^{\circ\circ}$ and since $X = \{\varphi_1, \varphi_2, \cdots, \varphi_n\} \subset E_1^\circ$ we get $\psi - \psi_2 \in X^\circ$. This completes the proof.

REFERENCE

1. S. Chase, *Function topologies on Abelian groups*, Illionois J. of Math., **7** (1963), 593-608.

2. R. J. Nunke, *On direct products of infinite cyclic groups*, Proc. Amer. Math. Soc., **13** (1962), 66-71.

3. E. Specker, *Additive gruppen von Folgen ganzer zahlen*, Protugaliae Math., (1950), 131-140.

UNIVERSITY OF SOUTH CAROLINA

ON THE REFLECTION OF HARMONIC FUNCTIONS AND OF SOLUTIONS OF THE WAVE EQUATION

VLADIMIR FILIPPENKO

Introduction. While the analytic extension of a harmonic function across analytic differential boundary conditions is always possible for the case of two independent variables [3], no comparable global theorem exists for harmonic functions in $N > 2$ variables.

This work is concerned with the problem of global extension of a harmonic function $U(x, y, z)$ across a plane on which U satisfies a linear differential boundary condition of the form

$$B(U) \equiv \frac{\partial U}{\delta z} + P_n(x, y)U = 0 \qquad \text{on } \sigma(z = 0) ,$$

where $P_n(x, y)$ is a polynomial of degree n. It is assumed here that the given function U is C^1 in the closure of a cylindrical domain $R: \{x^2 + y^2 < \rho^2, \ -l < z < 0\}$.

The possibility of harmonic reflection is obvious for $n = 0$, $P_n =$ const. as $B(U)$ itself is harmonic. Since it vanishes on $z = 0$, it can be extended harmonically, and the harmonic extension of U can then be found by integrating with respect to z. But such procedure is no longer available in our case. We shall show, how our problem can be reduced to that of solving an initial value problem of a certain hyperbolic differential equation (1.22) of order $2n$ with distinct characteristic surfaces (of normal type).

Classical considerations yield the analyticity of U on σ and, therefore, its harmonic extensibility across σ into a neighborhood of σ. Our result asserts that this neighborhood is the whole of the mirror image of R, denoted by $\bar{R}$.

Our method consists of constructing a new function $V(x, y, z)$ from U and a differential expression in V (see (1.6) and (1.18)), which is harmonic in R and vanishes on $z = 0$. Thus, this expression in V can be first extended into $R \cup \sigma \cup \bar{R}$ as a harmonic function $\varphi(x, y, z)$. The solution of the differential equation thus obtained for V in $\bar{R}$ is impeded by its degeneracy. To remove this degeneracy we add to the differential equation the Laplacian of V and its higher derivatives in such a way as to obtain a normal hyperbolic problem (1.22), whose solution is guaranteed by a result of I. G. Petrovsky. This modification of the differential equation can be done in infinitely many ways, in particular, so as to make the characteristic surfaces close down on

Received February 20, 1964. This work was supported by the Office of Naval Research, #222 (62)

parallels to the z-axis. Local extensibility of U, together with the solution of the modified equation, then yields the global extension of U. We note, that this method works equally well for $N > 3$ independent variables.

The above described method, however, seems to fail in the case of the wave equation when σ is part of the tilmelike plane $z = 0$, and the boundary condition on it is as simple as $U_z + xU = 0$.

On the other hand, the oblique derivative problem for the wave equation $U_{xx} + U_{yy} - U_{tt} = 0$, whose solution satisfies the boundary condition

$$B'(U) \equiv U_x + \alpha U_y + (Ay + B)U = 0 \qquad \text{on } x = 0 ,$$

yields to a similarly motivated, yet formally different attack. The domain of extension in this case depends on $\alpha \neq 0$.

I would like to take this opportunity to express my gratitude to professor H. Lewy who suggested this problem and offered advice during its investigation.

1. Analytic extension of harmonic functions.

We consider an open cylindrical domain $R : \{x^2 + y^2 < \rho^2, -l < z < 0\}$ and the plane region $\sigma : \{x^2 + y^2 < \rho^2, z = 0\}$. Denote by $\bar{R}$ the mirror image of R with respect to the $z = 0$ plane.

Let there be given a real function $U(x, y, z)$, $U \in C^1$ in the closure of R, such that:

$$U_{xx} + U_{yy} + U_{zz} \equiv \varDelta U = 0 \qquad \text{in } R \tag{1.1}$$

$$\frac{\partial U}{\partial z} + P_n(x, y) U = 0 \qquad \text{on } \sigma \tag{1.2}$$

where $P_n(x, y)$ is a polynomial in x, y of degree n, given in the form

$$P_n(x, y) = \sum_{k=0}^{n} \sum_{m=0}^{k} A_{km} x^{k-m} y^m , \tag{1.3}$$

the coefficients A_{km} being real.

LEMMA 1. *If $U(x, y, z)$ is harmonic in R, $U \in C^1$ in $R \cup \partial R$, and satisfies condition* (1.2) *on σ, then U can be harmonically extended into $R \cup \sigma \cup G$, where G is the portion $z > 0$ of some neighborhood of σ.*

Proof. Since U is C^1 in $R \cup \partial R$, we have by Green's formula

$$4\pi U(X) = \iint_{\partial R} \left\{ \frac{1}{|X - \tau|} \frac{\partial U(\tau)}{\partial n} - U(\tau) \frac{\partial}{\partial n} \frac{1}{|X - \tau|} \right\} d\tau \tag{1.4}$$

where $X = (x, y, z)$, $\tau = (\xi, \eta, \zeta)$, n is the outer normal, and integration is over the surface of the cylinder $\xi^2 + \eta^2 = \rho^2$, $\zeta = -l$, $\zeta = 0$. By (1.2) this becomes

$$4\pi U(X) = A(X) - \iint_\sigma \left\{ \frac{P_n(\tau)U(\tau)}{|X-\tau|} + U(\tau)\frac{\partial}{\partial\zeta}\frac{1}{|X-\tau|} \right\} d\tau$$

where $A(X)$ stands for the integral in (1.4) taken over the lateral surface and the lower base of the cylinder. By passage to the limit as X tends to $X' \in \sigma$, one obtains in a manner familiar in potential theory,

$$2\pi U(X') = A(X') - \iint_\sigma \frac{P_n(\tau')U(\tau')}{|X'-\tau'|} d\tau'$$

where $A(X')$ is an analytic function on σ. This integral equation is an especially simple case of E. Hopf's equation (6.1) ([2], page 220), and his method yields immediately the result, that $U(x, y, 0)$ is analytic on the open disc σ.

Since, due to condition (1.2), $U_z(x, y, 0)$ is also analytic, we obtain from the Cauchy-Kowalewski theorem, that there exists an analytic solution $\tilde{U}$ of Cauchy's problem with $\tilde{U} = U$, $\tilde{U}_z = U_z$ on σ for $\Delta\tilde{U} = 0$ in some neighborhood G of σ.

If we continue U, given in $R \cup \sigma$, as $\tilde{U}$ in $G - R - \sigma$, this new function is, according to well known arguments, harmonic in $R \cup \sigma \cup G$.

We now introduce the symbolic notation

$$D_z^{-1} f(x, y, z) = \int_0^z f(x, y, \zeta) d\zeta \,, \tag{1.5}$$

and define an analytic function $V(x, y, z)$ for $(x, y, z) \in R \cup \sigma$:

$$V(x, y, z) \equiv D_z^{-(2n-1)} U(x, y, z) + \sum_{k=0}^{2n-2} \frac{z^k}{k!} F_k(x, y) \,, \tag{1.6}$$

where the functions $F_k(x, y)$ $(0 \leqq k \leqq 2n-2)$ are solutions of the following equations on σ:

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right) F_{2n-2} + U_z(x, y, 0) = 0 \tag{1.7}$$

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right) F_{2n-3} + U(x, y, 0) = 0 \tag{1.8}$$

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right) F_r + F_{r+2} = 0 \qquad (0 \leqq r \leqq 2n-4) \tag{1.9}$$

with, say, boundary values zero on $x^2 + y^2 = \rho^2$.

The choice of these functions is motivated by the requirements

$$\Delta V = 0 \quad \text{in } R \tag{1.10}$$

$$V_{z^{2n}} + P_n(x, y) V_{z^{2n-1}} = 0 \quad \text{on } \sigma , \tag{1.11}$$

which are easily verified.

Let s stand for either x or for y, and denote

$$H_{s,z} = s\frac{\partial}{\partial z} - z\frac{\partial}{\partial s}, \quad H^m_{s,0} = (H_{s,z})^m|_{z=0} .$$

We then have the identities:

$$H^{2m+1}_{s,0} = \sum_{k=0}^{m}\sum_{j=0}^{k} a^m_{jk} s^{2k-j+1}\frac{\partial^{2k-j+1}}{\partial s^j \partial z^{2k-2j+1}} \qquad (m = 0, 1, 2, \cdots) \tag{1.12}$$

$$H^{2m}_{s,0} = \sum_{k=0}^{m}\sum_{j=0}^{k} b^m_{jk} s^{2k-j}\frac{\partial^{2k-j}}{\partial s^j \partial z^{2k-2j}} \qquad (m = 1, 2, \cdots) \tag{1.13}$$

where the coefficients a^m_{jk} and b^m_{jk} are real numbers, and $a^m_{0m} = b^m_{0m} = 1$.

Proof. Introducing new variables $t = s + iz$, $\tau = s - iz$, we may write, with $\partial/\partial t = 1/2[(\partial/\partial s) - i(\partial/\partial z)]$ and $\partial/\partial\tau = 1/2[(\partial/\partial s) + i(\partial/\partial z)]$

$$H_{s,z} = i\Big(t\frac{\partial}{\partial t} - \tau\frac{\partial}{\partial \tau}\Big) .$$

Hence,

$$H^n_{s,0} = i^n \sum_{p=0}^{n} (-1)^p \binom{n}{p}\Big(t\frac{\partial}{\partial t}\Big)^{n-p}\Big(\tau\frac{\partial}{\partial\tau}\Big)^p\Big|_{z=0} . \tag{1.14}$$

Now, for any variable ξ (real or complex)

$$\Big(\xi\frac{\partial}{\partial\xi}\Big)^r = \sum_{h=0}^{r} B^r_h \xi^h \frac{\partial^h}{\partial\xi^h}$$

where the coefficients B^r_h are nonnegative integers. Since $\partial\tau/\partial t = \partial t/\partial\tau = 0$, and for $z = 0$ we have $t = \tau = s$, each term in (1.14) is, but for a constant coefficient, of the form

$$t^\alpha\tau^\beta\frac{\partial^{\alpha+\beta}}{\partial t^\alpha \partial\tau^\beta}\Big|_{z=0} = s^{\alpha+\beta}\frac{\partial^{\alpha+\beta}}{\partial t^\alpha\partial\tau^\beta} \qquad (1 \leqq \alpha + \beta \leqq n) .$$

Since $\partial^2/\partial t\partial\tau = 1/4[(\partial^2/\partial s^2) + (\partial^2/\partial z^2)]$, each term in (1.14) is, but for a constant coefficient, either of the form

$$s^{\alpha+\beta}\Big[\frac{\partial^2}{\partial s^2} + \frac{\partial^2}{\partial z^2}\Big]^{\min\alpha,\beta}\Big(\frac{\partial}{\partial t}\Big)^{|\alpha-\beta|},$$

or of the form

$$s^{\alpha+\beta}\Big[\frac{\partial^2}{\partial s^2} + \frac{\partial^2}{\partial z^2}\Big]^{\min\alpha,\beta}\Big(\frac{\partial}{\partial \tau}\Big)^{|\alpha-\beta|} .$$

Now, for any positive integer q, $[(\partial/\partial s) \pm i(\partial/\partial z)]^q$ has terms with imaginary coefficients only of the form $\partial^{\lambda+\mu}/\partial s^\lambda \partial z^\mu$, where μ is odd, and terms with real coefficients only of the form $\partial^{\lambda+\mu}/\partial s^\lambda \partial z^\mu$, where μ is even ($\lambda + \mu = q$). Consequently, as $H^n_{s\,0}$ must have real coefficients, it will consist of terms $s^{\lambda+\mu}(\partial^{\lambda+\mu}/\partial s^\lambda \partial z^\mu)$, where μ is odd when n is odd, and μ is even when n is even, which implies identities (1.12) and (1.13).

LEMMA 2. *There exist differential operators*

$$D_i^r \equiv \sum_{h=0}^{r} C_{irh} \frac{\partial^r}{\partial s^h \partial z^{r-h}} ,$$

where C_{irh} are real constants, such that

$$s^p \frac{\partial^{2p-1}}{\partial z^{2p-1}} = \sum_{i=1}^{P} H^i_{s,0} D_i^{p-1} \qquad \textit{for } z = 0 \tag{1.15}$$

Proof. Starting from the definition of $H^n_{s\,0}$ we see, that the above statement holds for $p = 1$ and $p = 2$, with $D_1^0 = 1$, $D_1^1 = \partial/\partial s$ and $D_2^1 = \partial/\partial z$. Assuming, that the statement holds for $p \leqq 2n$, we prove by induction, that it also holds for $p = 2n + 1$ and $p = 2n + 2$.

Since, by assumption, the lemma holds for $p \leqq 2n$, we have for any nonnegative integers α and β, and any positive integer $q \leqq 2n$

$$s^q \frac{\partial^{2q-1+\alpha+\beta}}{\partial s^\alpha \partial z^{2q-1+\beta}} = \sum_{i=1}^{q} H^i_{s,0} D_i^{q-1+\alpha+\beta} . \tag{1.16}$$

But identity (1.12) yields

$$\begin{aligned} s^{2n+1} \frac{\partial^{4n+1}}{\partial z^{4n+1}} = H^{2n+1}_{s,0} \frac{\partial^{2n}}{\partial z^{2n}} - \sum_{j=1}^{n} a^n_{jn} s^{2n-j+1} \frac{\partial^{4n-j+1}}{\partial s^j \partial z^{4n-2j+1}} \\ - \sum_{k=0}^{n-1} \sum_{j=0}^{k} a^n_{jk} s^{2k-j+1} \frac{\partial^{2n+2k-j+1}}{\partial s^j \partial z^{2n+2k-2m+1}} . \end{aligned}$$

We now observe, that all terms on the right hand side of the above expression are of the form (1.16), where $q = 2n - j + 1$ $(1 \leqq j \leqq n$, i.e. $q \leqq 2n)$, $\alpha = j$, $\beta = 0$, for terms contained in the simple sum, and $q = 2k - j + 1$ $(0 \leqq j \leqq k,\ 0 \leqq k \leqq n - 1$, i.e. $q \leqq 2n - 1)$, $\alpha = j$, $\beta = 2n - 2k$, for terms contained in the double sum. Hence, the above lemma holds for $p = 2n + 1$.

A similar argument, which utilizes identity (1.13) instead of (1.12), shows that this lemma holds also for $p = 2n + 2$, and thus completes the proof.

We now introduce the differential operator of order $2p - 1$

$$Q^p_{s,z} \equiv \sum_{i=1}^{p} H^i_{s,z} D_i^{p-1} \qquad (p \geqq 1) \tag{1.17}$$

where the D_i^{p-1} are those of (1.15). Note that, for $z=0$, $Q_{s,0}^p = s^p(\partial^{2p-1}/\partial z^{2p-1})$.

Define an analytic function $\varphi(x, y, z)$ for $(x, y, z) \in R \cup \sigma$:

$$\varphi(x, y, z) \equiv V_{z^{2n}}(x, y, z) + NV(x, y, z) . \tag{1.18}$$

Here $V(x, y, z)$ is the function defined in (1.6), and $N = N(x, y, z)$ is a differential operator of order $2n-1$ defined by:

$$\begin{aligned} N(x, y, z) = {} & A_{00}\frac{\partial^{2n-1}}{\partial z^{2n-1}} + \sum_{k=1}^{n}(A_{k0}Q_{x,z}^k + A_{kk}Q_{y,z}^k)\frac{\partial^{2n-2k}}{\partial z^{2n-2k}} \\ & + \sum_{k=2}^{n}\sum_{m=1}^{k-1} A_{km}Q_{x,z}^{k-m}Q_{y,z}^m\frac{\partial^{2n-2k+1}}{\partial z^{2n-2k+1}} \end{aligned} \tag{1.19}$$

where the coefficients A_{km} are the coefficients of the polynomial $P_n(x, y)$ defined in (1.3).

LEMMA 3. *$\Delta\varphi = 0$ in R, and $\varphi(x, y, 0) = 0$.*

Proof. Note, that $\Delta H_{x,z} = H_{x,z}\Delta$ and $\Delta H_{y,z} = H_{y,z}\Delta$. Thus, by (1.17) and (1.19), the operators Δ and N commute. Therefore, operating on both sides of (1.18) by Δ, and making use of (1.10), we obtain

$$\Delta\varphi = \left(\frac{\partial^{2n}}{\partial z^{2n}} + N\right)\Delta V = 0 \qquad \text{in } R .$$

Making use of (1.17) and (1.15) we may write, for $z = 0$,

$$\begin{aligned} & N(x, y, z)V(x, y, z)|_{z=0} \\ & = \left\{A_{00}\frac{\partial^{2n-1}}{\partial z^{2n-1}} + \sum_{k=1}^{n}\left(A_{k0}x^k\frac{\partial^{2k-1}}{\partial z^{2k-1}} + A_{kk}y^k\frac{\partial^{2k-1}}{\partial z^{2k-1}}\right)\frac{\partial^{2n-2k}}{\partial z^{2n-2k}}\right. \\ & \quad \left. + \sum_{k=2}^{n}\sum_{m=1}^{k-1}A_{km}x^{k-m}\frac{\partial^{2k-2m-1}}{\partial z^{2k-2m-1}}y^m\frac{\partial^{2m-1}}{\partial z^{2m-1}}\frac{\partial^{2n-2k+1}}{\partial z^{2n-2k+1}}\right\}V(x, y, z)\Big|_{z=0} , \end{aligned}$$

which becomes

$$N(x, y, z)V(x, y, z)|_{z=0} = \sum_{k=0}^{n}\sum_{m=0}^{k}A_{km}x^{k-m}y^m V_{z^{2n-1}}(x, y, 0) . \tag{1.20}$$

Thus, setting $z = 0$ in (1.18) and making use of (1.20) and (1.11) we obtain $\varphi\ (x, y, 0) = 0$.

Hence, if we set for $(x, y, z) \in \bar{R} \cup \sigma$

$$\varphi(x, y, z) = -\varphi(x, y, -z) \equiv -\left[\frac{\partial^{2n}}{\partial \zeta^{2n}} + N(x, y, \zeta)\right]V(x, y, \zeta)\Big|_{\zeta=-z} , \tag{1.21}$$

then φ is harmonic in $R \cup \sigma \cup \bar{R}$.

Since $\varphi(x, y, -z)$ is known for $(x, y, z) \in \bar{R} \cup \sigma$, we shall seek a function $\bar{V}(x, y, z)$ for $(x, y, z) \in \bar{R} \cup \sigma$, which satisfies the following overdetermined system (S) for $\bar{V}$ on $z > 0$:

$$\left.\begin{array}{l} \bar{V}_{z^{2n}}(x, y, z) + N(x, y, z)\bar{V}(x, y, z) = -\varphi(x, y, -z) \\ \Delta \bar{V}(x, y, z) = 0 \\ \left.\dfrac{\partial^r \bar{V}}{\partial z^r}\right|_{z=0} = F_r(x, y) \quad 0 \leqq r \leqq 2n-2 \quad \bar{V}_{z^{2n-1}}(x, y, 0) = U(x, y, 0) \end{array}\right\} \text{(S)}$$

where the functions $F_r(x, y)$ are defined by the equations (1.7), (1.8) and (1.9).

Since, by Lemma 1, U can be continued into $R \cup \sigma \cup G$ as an analytic function, the formula (1.6) can be used to define a function $V^*(x, y, z)$ as an analytic function in $R \cup \sigma \cup G'$, where G' consists of all those points of G, which can be joined in G to points of σ by parallels to the z-axis. This, so defined function V^* is harmonic in $R \cup \sigma \cup G'$, satisfies the initial conditions of (S), and

$$\begin{aligned} V^*_{z^{2n}}(x, y, z) + NV^* &= -\left[\frac{\partial^{2n}}{\partial \zeta^{2n}} + N(x, y, \zeta)\right] V^*(x, y, \zeta)\,|_{\zeta=-z} \\ &= -\varphi(x, y, -z) \qquad \text{in } G' . \end{aligned}$$

Thus, a solution $V^*(x, y, z)$ of system (S) *exists for $(x, y, z) \in G' \cup \sigma$.*

To investigate the size of the domain into which $V(x, y, z)$ can be continued, consider the solution of the following Cauchy problem:

$$\text{(1.22)} \qquad \begin{aligned} M\bar{V}(x, y, z) \equiv \prod_{i=1}^{n}\left[\frac{\partial^2}{\partial z^2} - \alpha_i\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right)\right]\bar{V} + \beta N(x, y, z)\bar{V} \\ = -\beta\varphi(x, y, -z) \end{aligned}$$

$$\text{(1.23)} \qquad \left.\frac{\partial^r \bar{V}}{\partial z^r}\right|_{z=0} = F_r(x, y) \ (0 \leqq r \leqq 2n-2)\ , \ \ \bar{V}_{z^{2n-1}}(x, y, 0) = U(x, y, 0)$$

where $\alpha_i (i = 1, 2, \cdots, n)$ are distinct positive real numbers, and $\beta = \prod_{i=1}^{n}(1 + \alpha_i)$.

Now, for distinct positive α_i, M is a normal hyperbolic operator with the distinct characteristic sheets through a point (x^0, y^0, z^0) of the form $(x - x^0)^2 + (y - y^0)^2 = \alpha_i(z - z^0)^2$. It is a result of I. G. Petrovsky (see [1]), that the Cauchy problem (1.22), (1.23) has the unique C^∞ *solution $\bar{V}(x, y, z)$ in that part R^*_α $(\alpha = (\alpha_1, \alpha_2, \cdots, \alpha_n))$* of the domain of influence of the initial surface σ for the equation $M\bar{V}(x, y, z) = -\beta\varphi(x, y, -z)$, which lies in $\bar{R}$, so that $\varphi(x, y, -z)$ is defined.

In view of the identity

$$\prod_{i=1}^{n}(1 + \alpha_i)\frac{\partial^{2n}}{\partial z^{2n}} - \prod_{i=1}^{n}\left[\frac{\partial^2}{\partial z^2} - \alpha_i\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right)\right] = P(\Delta)$$

where P is a polynomial in $\partial/\partial x$, $\partial/\partial y$, $\partial/\partial z$ the function $V^*(x, y, z)$, which solves system (S) in G' satisfies the above Cauchy problem (1.22), (1.23) in the neighborhood of the initial surface σ, and by uniqueness, the solution $\bar{V}(x, y, z) \in R_{\alpha}^{*}$ must coincide with $V^{*}(x, y, z)$ in that neighborhood. Consequently, $\Delta \bar{V}$ and all its derivatives vanish on σ.

Since the operators M and Δ commute, operating on equation (1.22) by Δ we obtain $M(\Delta \bar{V}) = 0$. Therefore, by uniqueness of the solution of Cauchy's problem for $M(\Delta \bar{V}) = 0$ with homogeneous initial conditions, we conclude that $\bar{V}(x, y, z)$, which solves (1.22), (1.23), is harmonic in R_{α}^{*} and solves system (S) in this domain.

Putting $U(x, y, z) = (\partial^{2n-1}/\partial z^{2n-1}) \bar{V}(x, y, z)$ for $(x, y, z) \in R_{\alpha}^{*}$ we have constructed the harmonic extension of U into $R \cup \sigma \cup R_{\alpha}^{*}$. We now observe, that as $\alpha_i \to 0$ $(i = 1, 2, \cdots, n)$ the characteristic surfaces of M close down on parallels to the z-axis. It follows, that every point of $\bar{R}$ is in some R_{α}^{*} for α_i sufficiently small. In view of the simple connectedness of $R \cup \sigma \cup \bar{R}$, the harmonic extension of U at any point of $\bar{R}$ cannot depend on α, and it follows that U can be harmonically extended into all of $R \cup \sigma \cup \bar{R}$. Thus,

THEOREM 1. *If $U(x, y, z)$ is harmonic in R, $U \in C^1$ in $R \cup \partial R$, and satisfies condition* (1.2) *on σ, then U can be harmonically extended into $R \cup \sigma \cup \bar{R}$.*

REMARK The construction of the extension of U depended on the solution of a hyperbolic problem whose order is twice the degree of the polynomial $P_n(x, y)$, the coefficient in the first order boundary condition. This illustrates the difficulty of extending our result to the case of, say, a coefficient $f(x, y)$, which is an entire function.

2. Extension of solutions of the wave equation. We consider an open domain $D : \{-m < x < 0,\ -l < y < l,\ -l < t < l\}$ and the plane region $\sigma : \{x = 0,\ -l < y < l,\ -l < t < l\}$. Denote, for any domain $\mathscr{D}$, the mirror image of $\mathscr{D}$ with respect to the $x = 0$ plane by $\bar{\mathscr{D}}$.

Let there be given a real function $U(x, y, t)$, $U \in C^4$ in the closure of D, such that:

$$LU \equiv U_{xx} + U_{yy} - U_{tt} = 0 \quad \text{in } D \tag{2.1}$$

$$U_x + \alpha U_y + (Ay + B)U = 0 \quad \text{on } \sigma \tag{2.2}$$

where α, A, B are real constants; $\alpha \neq 0$.

Define a function $V(x, y, t)$ for $(x, y, t) \in D \cup \sigma$:

$$V(x, y, t) \equiv \int_0^x U(\xi, y, t) d\xi + G(y, t) \tag{2.3}$$

where $G(y, t)$ is the C^4 solution of the Cauchy problem:

$$\left.\begin{aligned} &G_{yy} - G_{tt} + U_x(0, y, t) = 0 \\ &G(y, 0) = G_t(y, 0) = 0 \end{aligned}\right\} \tag{2.4}$$

Let P be the parallelepiped bounded by the planes $t \pm y = \pm l$, $x = 0$, $x = -m$. Then, $V(x, y, t) \in C^4(V_x \in C^4)$ is defined in $P \cap D \cup \sigma$, and we have the relations:

$$LV = 0 \quad \text{in} \quad P \cap D \cup \sigma \,, \tag{2.5}$$

$$V_{xx} + \alpha V_{xy} + (Ay + B)V_x = 0 \quad \text{on} \quad P \cap \sigma \,, \tag{2.6}$$

which are easily verified.

We now define for $(x, y, t) \in P \cap D \cup \sigma$ the function:

$$\varphi(x, y, t) \equiv V_{xx} + \alpha V_{xy} + A\Big(y\frac{\partial}{\partial x} - x\frac{\partial}{0y}\Big)V + BV_x \,. \tag{2.7}$$

Since the operators L and $\{y(\partial/\partial x) - x(\partial/\partial y)\}$ commute, operating on both sides of (2.7) by L, and making use of (2.5), we obtain:

$$L\varphi = \Big\{\frac{\partial^2}{\partial x^2} + \alpha\frac{\partial^2}{\partial x \partial y} + (Ay + B)\frac{\partial}{\partial x} - Ax\frac{\partial}{\partial y}\Big\}(LV) = 0 \,.$$

Setting $x = 0$ in (2.7), and making use of (2.6) we have $\varphi(0, y, t) = 0$.

If we now set for $(x, y, t) \in \bar{P} \cap \bar{D} \cup \sigma$

$$\varphi(x, y, t) = -\varphi(-x, y, t)$$

it follows, that $L\varphi = 0$ in $P \cap D \cup \sigma \cup \bar{P} \cap \bar{D}$, and $\varphi \in C^3$.

Since $\varphi(-x, y, t)$ is known for $(x, y, t) \in \bar{P} \cap \bar{D} \cup \sigma$, we now seek a function $\bar{V}(x, y, t)$ for $(x, y, t) \in \bar{P} \cap \bar{D} \cup \sigma$, which solves the following Cauchy problem:

$$M\bar{V}(x, y, t) \equiv \bar{V}_{xx} + \alpha\bar{V}_{xy} + (Ay + B)\bar{V}_x - Ax\bar{V}_y = -\varphi(-x, y, t) \tag{2.8}$$

$$\bar{V}(0, y, t) = G(y, t) \,, \quad \bar{V}_x(0, y, t) = U(0, y, t) \quad \text{on } \bar{P} \cap \sigma \,. \tag{2.9}$$

It is well known, that the function $\bar{V}(x, y, t) \in C^4$, which satisfies (2.8), (2.9), exists in a domain Q. Here Q is that domain, each of whose sections by a plane $t = K(-l < K < l)$ is a right triangle bounded by $x = 0$, $y = l - |K|$ and $y - \alpha x = |K| - l$ if $\alpha > 0$, or by $x = 0$, $y = |K| - l$ and $y - \alpha x = l - |K|$ if $\alpha < 0$. Note that Q does not depend on U, and is a subdomain of $\bar{P} \cap \bar{D} \cup \sigma$.

LEMMA 4. *If $\bar{V}(x, y, t) \in C^4$ in Q is the solution of the Cauchy problem* (2.8), (2.9), *then $L\bar{V} = 0$ in Q.*

Proof. We operate on both sides of (2.8) by L. Since the operators L and $\{y(\partial/\partial x) - x(\partial/\partial y)\}$ commute, and $L\varphi(-x, y, t) = 0$, we obtain:

$$M(L\bar{V}) = 0 \,.$$

setting $x = 0$ in (2.8) we have,

$$\bar{V}_{xx}(0, y, t) = -\alpha \bar{V}_{xy}(0, y, t) - (Ay + B)\bar{V}_{x}(0, y, t)$$

and hence, making use of (2.9) and (2.2), we obtain:

$$\bar{V}_{xx}(0, y, t) = U_x(0, y, t) \; . \tag{2.10}$$

Thus, due to equations (2.9) and (2.4)

$$L\bar{V}|_{x=0} = 0 \; .$$

From (2.3) and (2.7) we have:

$$\begin{aligned}\varphi(-x, y, t) \equiv \varphi(\xi, y, t)|_{\xi=-x} &= U_\xi(\xi, y, t)|_{\xi=-x} + \alpha U_y(-x, y, t) \\ &+ (Ay + B)U(-x, y, t) + AxG_y(y, t) + Ax\int_0^{-x} U_y(\xi, y, t)d\xi\end{aligned}$$

and therefore,

$$\begin{aligned}\frac{\partial}{\partial x}\varphi(-x, y, t)|_{x=0} = &- U_{xx}(0, y, t) - \alpha U_{xy}(0, y, t) \\ &-(Ay + B)U_x(0, y, t) + AG_y(y, t) \; .\end{aligned} \tag{2.11}$$

Differentiating (2.8) with respect to x, and setting $x = 0$ we obtain

$$\bar{V}_{xxx} + \alpha\bar{V}_{xxy} + (Ay + B)\bar{V}_{xx} - A\bar{V}_y = -\frac{\partial}{\partial x}\varphi(-x, y, t)|_{x=0} \quad \text{on } x = 0 \; ,$$

which after substituting (2.9), (2.10) and (2.11) becomes:

$$\bar{V}_{xxx}(0, y, t) = U_{xx}(0, y, t) \; .$$

Hence, by (2.9) and (2.1),

$$\frac{\partial}{\partial x}L\bar{V}|_{x=0} = 0 \; .$$

Consequently, by uniqueness of the solution of Cauchy's problem for $M(L\bar{V}) = 0$ with homogeneous initial conditions, we have that $L\bar{V} = 0$ in Q.

We thus have:

THEOREM 2. *If $U(x, y, t) \in C^4$ in the closure of D solves the wave equation* (2.1) *and satisfies the boundary condition* (2.2) *on σ, then there exists a function $U = \bar{V}_x \in C^3$ in the subdomain Q of $\bar{D}$, which extends U across σ as C^3 solution of the wave equation.*

References

1. L. Gårding, *Solution directe du problème de Cauchy pours les équations hyperboliques*, Colloques internationaux du Centre National de la Recherche Scientifique, LXXI, Nancy, 1956.

2. E. Hopf, *Über den funktionalen, insbesondere den analytischen Charakter der Lösungen elliptischer Differential-gleichungen zweiter Ordnung*, Math. Zeitschrift **34** (1932), 194–233.

3. H. Lewy, *A note on harmonic functions and a hydrodynamical application*, Proc. Amer. Math. Soc., **3**, no. 1, (1952).

University of California, Berkeley

MAPPINGS OF BOUNDED CHARACTERISTIC INTO ARBITRARY RIEMANN SURFACES

D. J. H. FULLER

Introduction. Throughout this paper we consider analytic mappings $f(z)$ of an arbitrary open Riemann surface R into an arbitrary Riemann surface S. Heins [3] introduced the class of Lindelöfian maps when R is hyperbolic, and defined them in terms of Green's functions; further contributions have been made by Rao [4], [5]. In the case of planar regions these maps are the classical functions of bounded characteristic.

Sario [6], [7], has utilized principal functions [1] on the range surface to obtain generalizations of the main theorems for mappings of R into S. In this paper a different first main theorem is obtained in which the proximity function is a generalization of Nevanlinna's proximity function by means of the substitution of a principal function for the logarithmic function. It is shown that the resulting class of functions of bounded characteristic are the Lindelöfian maps, and that an extremal decomposition characterization of these functions can be obtained as in the classical case.

1. An auxiliary family of functions. Analytic mappings from an arbitrary open surface R into an arbitrary surface S can be considered in terms of families $\mathscr{T}$ of LH functions, i.e., harmonic functions, with isolated logarithmic singularities having integral coefficients. For the purposes of this paper we slightly generalize the term, parametric disk: $\varDelta = (Q, \mu)$ is a parametric disk if Q is a classical parametric disk, and there is defined on it a metric μ that is a real scalar multiple of the induced metric.

We let ζ be the local variable on S, and fix $\sigma \in S$ and a parametric disk at σ. If S is closed we define $t(\zeta, \sigma, \alpha)$ for $\alpha \in S\backslash\sigma$ (set difference) as the LH function on S which has singularities $\log|\zeta - \alpha|$ and $-\log|\zeta - \sigma|$ and is normalised by

$$\lim_{\zeta\to\sigma} (t(\zeta, \sigma, \alpha) + \log|\zeta - \sigma|) = 0$$

in terms of the fixed parametric disk. At α a parametric disk is fixed such that

$$\lim_{\zeta\to\alpha} (t(\zeta, \sigma, \alpha) - \log|\zeta - \alpha|) = 0$$

Received September 19, 1963. This paper represents part of a thesis submitted to the faculty of the University of California, Los Angeles, in partial fulfillment of the requirements for the Ph. D. degree. The author is indebted to Professor L. Sario for his guidance and to Dr. K. V. R. Rao for his advice and help.

in terms of it. We define successively

$$t(\zeta, \alpha, \sigma) = -t(\zeta, \sigma, \alpha)\ , \qquad \alpha \in S\backslash\sigma\ ,$$
$$t(\zeta, \alpha, \delta) = t(\zeta, \alpha, \sigma) + t(\zeta, \sigma, \delta)\ , \qquad \alpha, \delta \in S\backslash\sigma\ .$$

These functions form the family $\mathscr{T}$.

If S is open, more than one family can usually be formed. We consider an exhaustion of S by regular regions Ω that contain σ and α, and define on Ω the function $t_\Omega(\zeta, \sigma, \alpha)$ which satisfies the above conditions for $t(\zeta, \sigma, \alpha)$ as well as one of the following:

(a) the normal derivative of $t_\Omega(\zeta, \sigma, \alpha)$ vanishes on the boundary $\partial\Omega$ of Ω,

(b) a consistent partition of the boundaries of the regions Ω is given, and $t_\Omega(\zeta, \sigma, \alpha)$ has constant value and vanishing flux over each part of $\partial\Omega$ ([1] pp. 87–90).

By the theory of normal operators ([1] pp. 152 ff.) $t(\zeta, \sigma, \alpha)$ is defined as the directed limit of $t_\Omega(\zeta, \sigma, \alpha)$ as S is exhausted by the regions Ω. $t(\zeta, \alpha, \sigma)$ and $t(\zeta, \alpha, \delta)$ are then defined as in the case of closed surfaces S. Each condition in (a) and (b) determines a family $\mathscr{T}$. It will be represented by $\mathscr{T}_0$ if (a) is satisfied and by $\mathscr{T}_1(P)$ if (b) is satisfied for a partition P; if P is the identity partition I, we write $\mathscr{T}_1(I)$.

Since each function t is a principal function ([1] p. 169), a family $\mathscr{T}$ will be called a principal family. We note that a change in the fixed parametric disk at σ changes every function $t(\zeta, \sigma, \alpha)$ by the same constant but leaves $t(\zeta, \alpha, \delta)$ unaltered. Further, in view of our definition of parametric disk, for any given $\mathscr{T}$ and constant k there exists a family $\mathscr{T}'$ such that for all m,

$$\{\zeta \mid t(\zeta, \sigma, \alpha) = m\} = \{\zeta \mid t'(\zeta, \sigma, \alpha) = m + k\},\ t \in \mathscr{T},\ t' \in \mathscr{T}'\ .$$

We consider functions belonging to any principal family. If $\alpha, \delta \in S\backslash\sigma$, these functions have the following four obvious properties.

$$(1)\qquad \begin{aligned} t(\zeta, \alpha, \alpha) &= 0\ , \\ t(\zeta, \alpha, \delta) + t(\zeta, \delta, \alpha) &= 0\ , \\ \lim_{\zeta\to\delta} (t(\zeta, \delta, \alpha) + \log|\zeta - \delta| &= t(\delta, \sigma, \alpha)\ ,, \\ t(\sigma, \alpha, \delta) &= 0\ . \end{aligned}$$

LEMMA 1.1. *$t(\alpha, \gamma, \delta) + t(\gamma, \delta, \alpha) + t(\delta, \alpha, \gamma) = 0$ when α, γ, δ are distinct points in S.*

Proof. If S is open we let $\Omega \subseteqq S$ be a regular region containing α, γ, and δ, and consider functions t_Ω defined on Ω. We remove small closed disks in Ω that contain α, γ, δ and apply Green's formula to

$t_\Omega(\zeta, \alpha, \gamma)$ and $t_\Omega(\zeta, \delta, \gamma)$ over the remaining region. On letting the disks shrink to points we obtain

$$-t_\Omega(\alpha, \delta, \gamma) + t_\Omega(\delta, \alpha, \gamma) + R(\gamma) = 0$$

where

$$R(\gamma) = \lim_{\zeta \to \gamma} (t_\Omega(\zeta, \delta, \gamma) - t_\Omega(\zeta, \alpha, \gamma) = t_\Omega(\gamma, \delta, \alpha))$$

and the lemma follows by letting $\Omega \to S$.

If S is closed the same method is applied to S instead of to Ω.

COROLLARY 1.2. $t(\alpha, \delta, \sigma) = t(\delta, \alpha, \sigma), \alpha, \delta \in S \backslash \sigma$.

Proof. This is obvious when α and δ are identical; if they are distinct it follows from replacing γ by σ in the lemma and applying (1).

COROLLARY 1.3. *If α is distinct from σ and γ, then $t(\alpha, \zeta, \gamma)$ is of class LH on $S \backslash \sigma$ with singularities at α and σ.*

Proof. If $\gamma = \sigma$ this is implied by Corollary 1.2. Otherwise

$$t(\alpha, \zeta, \gamma) + t(\zeta, \sigma, \alpha) = t(\alpha, \sigma, \gamma) ,$$

which is constant.

LEMMA 1.4. $\psi: S\backslash\sigma \times S\backslash\sigma \to [-\infty, \infty] \mid \psi(\gamma, \alpha) = t(\delta, \gamma, \alpha)$ *is continuous for every fixed δ.*

Proof. If $\delta = \sigma$ then ψ is identically zero; if not,

$$\psi(\gamma, \alpha) = t(\delta, \gamma, \sigma) + t(\delta, \sigma, \alpha)$$

and each term is continuous by Corollary 1.2.

Sario [8] proves that if $E \subseteq S\backslash\sigma$ is compact and Q is an open set containing E and σ, then $t(\gamma, \alpha, \sigma) \in \mathscr{T}_0$ is uniformly bounded for $\alpha \in E, \gamma \in S\backslash Q$. The same proof holds for $t(\gamma, \alpha, \sigma) \in \mathscr{T}_1(P)$. From the harmonicity of $t(\gamma, \alpha, \sigma)$ in γ and in α, and from its uniform boundedness, it follows by a lemma of Heins ([2] p. 445) that

LEMMA 1.5. *If $S' = S \times S\backslash((\sigma, \sigma) \cup \{(\sigma, \zeta)\} \cup \{(\zeta, \sigma)\} \cup \{(\zeta, \zeta)\})$, then $\phi: S' \to (-\infty, \infty) \mid \phi(\gamma, \alpha) = t(\gamma, \alpha, \sigma)$ is continuous.*

LEMMA 1.6. *If $S' = S \times S\backslash(\sigma, \sigma)$ then $\phi: S' \to [-\infty, \infty] \mid \phi(\gamma, \alpha) = t(\gamma, \alpha, \sigma)$ is continuous.*

Proof. It suffices to consider the continuity at points $(\gamma_0, \alpha_0), \gamma_0 =$

$\alpha_0 \neq \sigma$, and $(\sigma, \alpha), \alpha_0 \neq \sigma$. For the first we let Δ be a parametric disk at α_0 such that $\sigma \notin \bar{\Delta}$, and F be a closed connected neighborhood of α_0 that does not intersect $\partial\Delta$. For every $(\gamma, \alpha) \in F \times F$ there exists $\eta \in \partial\Delta$ such that $t(\gamma, \alpha, \sigma) > t(\eta, \alpha, \sigma)$, which is bounded by Sario's lemma for all $\eta \in \partial\Delta, \alpha \in F$. No generality is lost by taking $t(\eta, \alpha, \sigma) > 0$ for all $\eta \in \partial\Delta, \alpha \in F$.

Let ψ be a homeomorphism of Δ onto a closed disk in the plane, and g the Green's function on this disk. By its extremal property

$$t(\zeta, \alpha, \sigma) - g(\psi(\zeta), \psi(\alpha)) \geqq 0$$

for $\zeta, \alpha \in F$. Since for any n, there exists a neighborhood E_n of the origin of the disk such that $g(z, a) \geqq n$ for $z, a \in E_n$, we have $t(\gamma, \alpha, \sigma) > n$ for $\gamma, \alpha \in \psi^{-1}(E_n) \cap F$, and ϕ is continuous at (γ_0, α_0).

For the second case we let $\partial\Delta$ be the boundary of a parametric disk at α_0, and F and G be compact connected neighborhoods of σ and α_0 that do not intersect $\partial\Delta$. For $\zeta \in F, \alpha \in G$, there exists $\eta \in \partial\Delta$ such that

$$t(\zeta, \alpha, \alpha_0) < t(\eta, \alpha, \alpha_0) = t(\eta, \alpha, \sigma) + t(\eta, \sigma, \alpha_0) ,$$

which, by Sario's lemma, is bounded above, say by M. Hence

$$t(\zeta, \alpha, \sigma) < M + t(\zeta, \alpha_0, \sigma) ,$$

and the lemma follows, since for any n, $t(\zeta, \alpha_0, \sigma) < n$ in some neighborhood of σ.

We conclude this section by noting that the limits of $t(\zeta, \gamma, \sigma)$, $t(\zeta, \sigma, \gamma)$ and $t(\zeta, \gamma, \gamma)$ as $\gamma \to \sigma$ are ∞, $-\infty$ and 0 respectively, and that $t(\zeta, \sigma, \sigma)$ is not defined.

2. Jensen's formula. The main tool used in this paper is Jensen's formula generalized for Riemann surfaces. We let Ω be a regularly imbedded relatively compact region on the surface R and let $v(z)$ be an LH function on $\bar{\Omega}$. The positive singularities of $v(z)$ in Ω will be designated by $a_i, i = 1, \cdots, m$, and the negative singularities by $b_j, j = 1, \cdots, n$; their multiplicities will be given by μ_i and ν_j respectively.

We obtain the formula from the following proposition:

LEMMA 2.1. *If r is not a singularity of $v(z)$, then*

$$v(r) = \frac{1}{2\pi}\int_{\partial\Omega} v(z)d^*p(z, r) + \sum_i \mu_i g(a_i, r) - \sum_j \nu_j g(b_j, r) \tag{2}$$

where $p(z, r)$ and $g(z, r)$ are the capacity and Green's functions defined on Ω with singularities at the point r, and $\partial\Omega$ is oriented counterclockwise about r.

Proof. We first take the case when $v(z)$ has no singularities on $\bar{\Omega}$. Let $\Delta \subseteqq \Omega$ be a small closed disk that contains r. On applying Green's formula to $p(z, r)$ and $v(z)$ over $\Omega \backslash \Delta$, and letting Δ shrink to r we obtain

$$v(r) = \frac{1}{2\pi} \int_{\partial \Omega} v(z) d^* p(z, r) .$$

We next take the case when $v(z)$ has a singularity $\nu \log |z - a|$, $a \in \partial\Omega$, but has none in Ω. Let $p(z, r)$ have the value k on $\partial\Omega$; there exists an $\varepsilon > 0$ such that the boundary components of $\Omega_\varepsilon = \{z \mid p(z, r) < k - \varepsilon\}$, have a natural one-to-one mapping on those of $\partial\Omega$.

Let the components of $\partial\Omega$ be $\{\gamma_i\}$, $i = 1, \cdots, n$, with $a \in \gamma_1$, and the corresponding components of $\partial\Omega_\varepsilon$ be $\{\gamma_{i\varepsilon}\}$. For $i \neq 1$, we apply Green's formula to $v(z)$ and $p(z, r)$ over each component of $\Omega \backslash \Omega_\varepsilon$ and obtain

$$\int_{\gamma_i - \gamma_{i\varepsilon}} v(z) d^* p(z, r) = \varepsilon \int_{\gamma_{i\varepsilon}} d^* v(z) .$$

For $i = 1$, we let $\hat{\Omega}$ be the double of $\bar{\Omega} \backslash \Omega_\varepsilon$. If q is the total flux of $p(z, r)$ along γ_1, the function

$$h(z) = \exp\left[\frac{2\pi}{q} (p(z, r) - k + ip^*(z, r)\right]$$

maps the first component of $\hat{\Omega}$ conformally onto an annulus, such that γ_1 is mapped onto the unit circle B_1 and $\gamma_{1\varepsilon}$ onto $B_{1\varepsilon} = \{w \mid |w| = \exp[-(2\pi/q)\varepsilon]\}$. We may assume that the point a is mapped on $w = 1$. Consequently $d\theta = 2\pi/q \, d^* p(z, r)$.

Since $\int_0^{2\pi} \log |e^{i\theta} - 1| = 0$, it follows that

$$\int_{\gamma_1 - \gamma_{1\varepsilon}} v(z) d^* p(z, r) = \frac{q}{2\pi} \int_{B_1 - B_{1\varepsilon}} (v \cdot h^{-1}(re^{i\theta}) - \nu \log |re^{i\theta} - 1|) \, d\theta .$$

By applying Green's formula to the last integrand and to $\log |re^{i\theta}|$ over the annulus between the circles, we find

$$\int_{\gamma_1 - \gamma_{1\varepsilon}} v(z) d^* p(z, r) = \varepsilon \int_{\gamma_{1\varepsilon}} d^* v(z) .$$

Summing over all the components of $\partial\Omega$ we obtain

$$(3) \qquad \frac{1}{2\pi} \int_{\partial\Omega} v(z) d^* p(z, r) = \frac{1}{2\pi} \int_{\partial\Omega_\varepsilon} v(z) d^* p((z, r) = v(r) .$$

For the general case we note that

$$v(z) - \sum_i \mu_i g(a_i, z) + \sum_j \nu_j g(b_j, z)$$

is a harmonic function on Ω. The application of (3) to this function yields (2). We immediately obtain

COROLLARY 2.2. (Generalized Jensen's formula). *If f is an analytic mapping of R into a Riemann surface S on which is defined a function $t(\zeta, \alpha, \delta)$ belonging to a principal family, and if $f(r)$, α and δ are distinct, then,*

$$t(f(r), \alpha, \delta) = \frac{1}{2\pi}\int_{\partial\Omega} t(f(z), \alpha, \delta) d^*p(z, r) + \sum_i \mu_i g(a_i, r) - \sum_j \nu_j g(d_j, r) \,. \tag{4}$$

where $\{a_i\}$ and $\{b_j\}$ are the inverse images in Ω of α and δ respectively, and μ_i, ν_j are their multiplicities.

If $f(r)$ is a singularity of $t(\zeta, \alpha, \delta)$ the following proposition holds:

LEMMA 2.3. *If $f(r) = \alpha$, and if the Laurent expansion of $f(z)$ in the neighborhood of r is $f(z) = \sum_N^\infty c_i z_i$ with respect to the parametric disks at r and α fixed by $p(z, r)$ and $t(\zeta, \sigma, \alpha)$ respectively, then*

$$\lim_{z\to r} (Np(z, r) + t(f(z), \sigma, \alpha)) = \log \left|\frac{1}{c_N}\right| \,. \tag{5}$$

If $f(z) = \delta$, then, with the above expansion,

$$\lim_{z\to r} (-Np(z, r) + t(f(z), \sigma, \alpha)) = \log \left|\frac{1}{c_N}\right| \,. \tag{6}$$

Proof. We shall use the same symbol z for an arbitrary point on the surface and for its image under the mapping associated with the parametric disk under consideration. $t(f(z))$ and $p(z)$ will represent $t(f(z), \sigma, \alpha)$ and $p(z, r)$, and l_i, etc., constant coefficients. We set

$$q(z) = \exp\,[t(f(z)) + it^*(f(z))] \,;$$

this is single-valued in a neighborhood of r.

If $f(r) = \sigma$, the expansion in that neighborhood is

$$q(z) = \frac{1}{c_N} z^{-N} + - \sum_{N+1}^{\infty} l_i z^i \,.$$

Similarly, there is a neighborhood of r in which

$$r(z) = \exp\,[p(z) + ip^*(z)]$$

can be expanded as

$$r(z) = z + \sum_1^\infty m_i z^i .$$

Hence

$$\frac{1}{c_N} = \lim_{z \to r} q(z)\,(r(z))^N$$

which yields the first conclusion. The second is proved in the same way. This concludes the proof.

If we let λ equal N or $-N$ according as $f(r)$ is σ or α, then the function

$$t(f(z)) + \lambda p(z) - \sum_i{}' \mu_i g(s_i, z) + \sum_j{}' \nu_j g(a_i, z)$$

is harmonic on Ω, when the summations are over the inverse images in $\Omega \backslash r$. On applying Jensen's formula (4) and substituting from (5) or (6) we obtain the alternative expression

$$\log\left|\frac{1}{c_N}\right| = \frac{1}{2\pi}\int_{\partial\Omega} t(f(z))d^*p(z) + k\lambda + \sum_i{}' \mu_i g(s_i, r) - \sum_j{}' \nu_j g(a_i, r) .$$

We shall need the following property of subharmonic functions:

LEMMA 2.4. *Let u be an u.s.c. function on a region W.*

(i) *If u is subharmonic on W, then for every regular Ω whose closure is in W, and every $z \in \Omega$,*

$$u(z) \leqq \frac{1}{2\pi}\int_{\partial\Omega} u(w)d^*p_\Omega(w, z) . \tag{7}$$

(ii) *If for every $z \in W$, there is a regular Ω such that*

$$u(z) \leqq \frac{1}{2\pi}\int_{\partial h} u(w)d^*p_\Omega(w, z)$$

over every level line ∂h of $p_\Omega(w, z)$, then u is subharmonic on W.

Proof. To prove (i) we take an arbitrary Ω and $z \in \Omega$, and let $\{v_n\}$ be a descending sequence of continuous functions on $\partial\Omega$ tending to u. For any $w \in \partial\Omega$, we have by (2)

$$\overline{\lim_{z \to w}}\, u(z) \leqq \lim_{z \to w} \frac{1}{2\pi}\int_{\partial\Omega} v_n(w)d^*p(w, z)$$

for all v_n. By applying the monotone convergence theorem and the maximum principle we obtain the desired result.

For (ii) we let z_0 be an arbitrary point in the region, and choose a parametric disk about z_0. In terms of the associated unit disk the hypothesis yields

$$u(z_0) \leqq \frac{1}{2\pi} \int_0^{2\pi} u(z_0 + re^{i\theta})\, d\theta$$

for $0 < r \leqq 1$. The subharmonicity of u follows from the theory of functions on the plane.

We immediately obtain

COROLLARY 2.5. *If $u(z)$ is subharmonic on $\bar{\Omega}_2$ and $z \in \Omega_1 \subseteqq \Omega_2$, then*

$$\int_{\partial\Omega_1} u(w) d^* p_1(w, z) \leqq \int_{\partial\Omega_2} u(w) d^* p_2(w, z) \,, \tag{8}$$

where $p_i(w, z)$, $i = 1, 2$, is the capacity function on Ω_i.

3. Argument principle. Using the same notation as before we let Ω be a relatively compact regularly imbedded open set in the surface R, and $n(\Omega, \alpha)$, $n(\Omega, \delta)$ the number of inverse images (with multiplicities) in Ω of points α and δ in S that are not on the image of $\partial\Omega$. We have

$$\frac{1}{2\pi} \int_{\partial\Omega} d^* t(f(z), \alpha, \delta) = n(\Omega, \delta) - n(\Omega, \alpha) \tag{9}$$

where $\partial\Omega$ is oriented counter-clockwise, and t belongs to any principal family.

This statement follows from removing small disks at each of the inverse images of α and δ, applying Green's formula to $t(f(z), \alpha, \delta)$ over the remainder of Ω, and taking the limit as the disks shrink to points.

We choose $r \in \Omega$. If $p(z, r)$ is the capacity function on Ω, and k is its value on $\partial\Omega$, we let $\Omega_h = \{z \in \Omega \mid p(z, r) < h\}$ and ∂h be the boundary of Ω_h.

THEOREM 3.1. *If α and δ are not in the image of ∂h, and if α, δ and $f(r)$ are distinct, then*

$$\frac{1}{2\pi} \frac{d}{dh} \int_{\partial h} t(f(z), \alpha, \delta) d^* p(z, r) = n(h, \delta) - n(h, \alpha) \,, \tag{10}$$

where $n(h, \delta)$ and $n(h, \alpha)$ are the number of inverse images (with multiplicities) of δ and α in Ω_h, and t belongs to any principal family.

Proof. We let $t(f(z))$ and $p(z)$ represent $t(f(z), \alpha, \delta)$ and $p(z, r)$, $\{a_i\}$ and $\{d_j\}$ be the finite number of inverse images of α and δ in Ω_h.

There is a finite h' such that $\Omega_{h'}$ does not contain any of these inverse images. We remove small disks about the a_i's and d_j's and apply Green's formula to $t(f(z))$ and $p(z)$ over the remainder of $\Omega_h \backslash \bar{\Omega}_{h'}$. After evaluating and letting the disks shrink to points we obtain

$$\int_{\partial h - \partial h'} t(f(z)) d^*p(z) + 2\pi\Big[\sum_i \mu_i p(a_i) - \sum_j \nu_j p(d_j)\Big] \\ = h\int_{\partial h} d^*t(f(z)) - h'\int_{\partial h'} d^*t(f(z)) ,$$

since $p(z)$ is the capacity function on both Ω_h and $\Omega_{h'}$. In this relationship μ_i and ν_j are the multiplicities of the corresponding inverse images. The differentiation of this equation yields

$$\frac{d}{dh}\int_{\partial h} t(f(z)) d^*p(z) = \frac{d}{dh}\Big[h\int_{\partial h} d^*t(f(z))\Big] \\ = \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\Big[\varepsilon\int_{\partial h} d^*t(f(z)) - (h-\varepsilon)\int_{\partial(h-\varepsilon)-\partial h} d^*t(f(z))\Big] .$$

Since the last term vanishes for sufficiently small ε, we substitute from (9) and obtain the required relationship.

We note that (10) is an invariant property of principal families.

4. **Logarithmic capacities.** A logarithmic capacity of a compact set E properly contained in an arbitrary surface S can be defined in relation to any principal family $\mathscr{T}$ if $\sigma \in S\backslash E$. We let μ be a regular positive unit mass distribution on E. Since $t(\zeta, \eta, \sigma)$ is l.s.c. on E, we define the logarithmic potential of μ relative to $\mathscr{T}$ as

$$p_\mu(\eta) = \int_E t(\zeta, \eta, \sigma)\, d\mu(\zeta)$$

on $S\backslash\sigma$. The following proposition carries over from the plane:

LEMMA 4.1. *The logarithmic potential $p_\mu(\eta)$ is harmonic on $S\backslash(E\cup\sigma)$ and superharmonic on $S\backslash\sigma$. In the neighborhood of σ, $p_\mu(\eta) - \log|\eta-\sigma|$ is bounded.*

Proof. We let

$$t_n(\zeta, \eta, \sigma) = \min\{n, t(\zeta, \eta, \sigma)\} ,$$

and

$$p_{\mu n}(\eta) = \int_E t_n(\zeta, \eta, \sigma)\, d\mu(\zeta) .$$

By Lemmas 1.5 and 1.6, $t_n(\zeta, \eta, \sigma)$ is continuous in (ζ, η), $\zeta \in E$, $\eta \in S\backslash\sigma$, and as E is compact there is for any arbitrary point $\eta_0 \in S\backslash\sigma$ and $\varepsilon > 0$, a neighborhood Δ of η_0 such that

$$|t_n(\zeta, \eta, \sigma) - t_n(\zeta, \eta_0, \sigma)| < \varepsilon, \zeta \in E, \eta \in \Delta ,$$

It follows that $p_{\mu n}(\eta)$ is continuous and $p_\mu(\eta)$ l.s.c. on $S\backslash\sigma$.

Let Γ be a disk about η_0 such that $\partial\Gamma$ is a level line of $t(\zeta, \eta_0, \sigma)$. We orient $\partial\Gamma$ clockwise about η_0. Since $t(\zeta, \eta, \sigma)$ is bounded below for all $\zeta \in E, \eta \in \partial\Gamma$, and $-t(\zeta, \eta_0, \sigma)$ is the capacity function on Γ, we have by Corollary 1.2 and (2),

$$\frac{1}{2\pi}\int_{\partial\Gamma} d^*t(\eta, \eta_0, \sigma)\int_E t_n(\zeta, \eta, \sigma)\, d\mu(\zeta) = \int_E t'_n(\zeta, \eta_0, \sigma)\, d\mu(\zeta)$$

where for each ζ, $t'_n(\zeta, \eta, \sigma)$ is the harmonic function in η on Γ with boundary values $t_n(\zeta, \eta, \sigma)$. Further, by superharmonicity, $t'_n(\zeta, \eta_0, \sigma) \leqq t(\zeta, \eta_0, \sigma)$ for each ζ and for all n. We substitute in the above equation and apply the monotonic convergence theorem as $n \to \infty$. We obtain

$$\frac{1}{2\pi}\int_{\partial\Gamma} d^*t(\eta, \eta_0, \sigma)\int_E t(\zeta, \eta, \sigma)\, d\mu(\zeta) \leqq \int_E t(\zeta, \eta_0, \sigma)\, d\mu(\zeta)$$

and $p_\mu(\eta)$ is superharmonic by (7).

If $\eta_0 \notin E \cup \sigma$, Γ can be chosen such that $\bar{\Gamma} \subseteqq S\backslash(E \cup \sigma)$. Since $t(\zeta, \eta, \sigma)$ is harmonic on $\bar{\Gamma}$, the same method establishes the harmonicity of $p_\mu(\eta)$ on $S\backslash(E \cup \sigma)$ by (2) and the maximum principle.

To establish the final part of the lemma we need only note that by Lemmas 1.5 and 1.6 there is a neighborhood Δ of σ such that

$$t(\zeta, \eta, \sigma) = t(\eta, \zeta, \sigma) = \log|\eta - \sigma| + h(\eta, \zeta), \zeta \in E, \eta \in \Delta\backslash\sigma ,$$

where $h(\eta, \zeta)$ is bounded.

We deduce the following proposition:

COROLLARY 4.2. *If μ is as above and $f: R \to S$ is analytic, then, for a regular $\Omega \subset R$,*

$$\int_E d\mu(\zeta)\int_{\partial\Omega} t(f(z), \zeta, \sigma)\, d^*p(z, r) = \int_{\partial\Omega} d^*p(z, r)\int_E t(f(z), \zeta, \sigma)\, d\mu(\zeta) \tag{11}$$

where t belongs to any principal family $\mathscr{T}$ and p is the capacity function on Ω. The iterated integral is either finite or $+\infty$.

Proof. There exists a closed disk $D \subseteqq S$ about σ such that

(a) $D \cap E$ is void,

(b) $t(\alpha, \zeta, \sigma) < 0, \alpha \in D, \zeta \in E$ by 1.6,

(c) the intersection of D and the image of $\partial\Omega$ consists of a finite number (possibly zero) of Jordan arcs β_i, $i = 1, \cdots, n$, each of which passes through σ.

We divide $\partial\Omega$ into the inverse images γ_i, $i = 1, \cdots, n$, of β_i, and the remainder γ.

On γ the function $t(f(z), \zeta, \sigma)$ is uniformly bounded below for $\zeta \in E$ and we may apply Fubini's theorem to

$$\int_E d\mu(\zeta) \int_\gamma t(f(z), \zeta, \sigma)\, d^*p(z, r)\ .$$

The integral is either finite or $+\infty$.

For each i, we exhaust $\beta_i \backslash \sigma$ by a sequence of compact sets F_j. By (c) the restriction of $d^*p(z, r)$ to $f^{-1}(F_j) \cap \gamma_i$ induces a positive mass function on $F_j \subseteqq D\backslash\sigma$. Its logarithmic potential

$$p_{\mu ij}(\zeta) = \int_{f^{-1}(F_j)\cap\gamma_i} t(f(z), \zeta, \sigma)\, d^*p(z, r)$$

is harmonic on $S\backslash D$ by 4.1. By (b) the functions $p_{\mu ij}$ form a decreasing sequence; by Harnack's principle its limit

$$p_{\mu i}(\zeta) = \int_{\gamma_i} t(f(z), \zeta, \sigma)\, d^*p(z, r)$$

is either $-\infty$, or harmonic on $S\backslash D$.

We may assume that $p(z, r)$ is zero on $\partial\Omega$; then $\exp\{p(z, r) + ip^*(z, r)\}$ (choosing any branch of p^*) maps γ_i onto an arc of the unit circle. For any $\zeta_0 \in E$ we have

$$p_{\mu i}(\zeta_0) = \int_\alpha^\beta \log|e^{i\theta} - 1|\, d\theta + c$$

where c is some finite constant. Since this integral is bounded with respect to α and β, $p_{\mu i}(\zeta_0)$ is finite, and $p_{\mu i}(\zeta)$ is bounded on E.

Consequently, by (b), we may apply Fubini's theorem to

$$\int_E d\mu(\zeta) \int_{\gamma_i} t(f(z), \zeta, \sigma)\, d^*p(z, r)$$

for each i, and the integral is finite. Summing over γ and γ_i, we obtain the required relation.

5. **First main theorem.** To develop a first main theorem for analytic mappings $f: R \to S$ where R and S are arbitrary Riemann surfaces, we fix a point $\sigma \in S$ and define a principal family $\mathscr{T}$; we then select points $\tau \in S$ and $r \in R$ such that σ, τ and $f(r)$ are distinct. A parametric disk is selected at r.

Let $\overset{+}{t} = \max\{t, 0\}$. For a regular region $\Omega \subseteqq R$ such that $r \in \Omega$, the proximity function $m(\Omega, f)$, the counting function $N(\Omega, f)$ and the characteristic function $T(\Omega, f)$ of f on Ω are defined as

$$m(\Omega, f) = \frac{1}{2\pi}\int_{\partial\Omega} \overset{+}{t}(f(z), \sigma, \tau)\, d^*p(z, r)\ ,$$

$$N(\Omega, f) = \sum_i g(s_i, r)\ , \qquad \{s_i\} = f^{-1}(\sigma) \cap \Omega\ ,$$

$$T(\Omega, f) = m(\Omega, f) + N(\Omega, f)\ ,$$

where p and g are the capacity and Green's functions on Ω, and s_i is repeated in accordance with its multiplicity.

The proximity $m(\Omega, \alpha)$ and the counting functions $N(\Omega, \alpha)$ at the point α are defined as $m(\Omega, f)$ and $N(\Omega, f)$ when $\alpha = \sigma$; otherwise we define

$$m(\Omega, \alpha) = \frac{1}{2\pi}\int_{\partial\Omega} \overset{+}{t}(f(z), \alpha, \sigma)\, d^*p(z, r)\ ,$$

$$N(\Omega, \alpha) = \sum g(a_i, r)\ , \qquad \{a_i\} = f^{-1}(\alpha) \cap \Omega\ .$$

where a_i is repeated in accordance with its multiplicity.

The first main theorem reads:

THEOREM 5.1. *For every* $\alpha \in S\backslash f(r)$,

$$m(\Omega, \alpha) + N(\Omega, \alpha) = T(\Omega, f) + 0(1)\ ,$$

where $O(1)$ *is a bounded function with respect to* Ω.

Proof. When $\alpha = \sigma$ it is trivial; when $\alpha \neq \sigma$, Jensen's formula (4) is

$$\begin{aligned} t(f(r), \alpha, \sigma) = &\frac{1}{2\pi}\int_{\partial\Omega} \overset{+}{t}(f(z), \alpha, \sigma)\, d^*p(z, r) + \sum_i g(a_i, r) \\ &- \frac{1}{2\pi}\int_{\partial\Omega} \overset{+}{t}(f(z), \sigma, \alpha)\, d^*p(z, r) - \sum_i g(s_i, r)\ , \end{aligned}$$

which is

$$\begin{aligned} m(\Omega, \alpha) + N(\Omega, \alpha) = &\frac{1}{2\pi}\int_{\partial\Omega} \overset{+}{t}(f(z), \sigma, \alpha)\, d^*p(z, r) \\ &+ N(\Omega, f) + O(1)\ . \end{aligned} \tag{12}$$

For $\zeta \in S$ we define

$$q(\zeta) = \overset{+}{t}(\zeta, \sigma, \alpha) - \overset{+}{t}(\zeta, \sigma, \tau)\ .$$

There is a neighborhood Δ of σ in which both $t(\zeta, \sigma, \alpha)$ and $t(\zeta, \sigma, \tau)$ are positive. Hence in Δ,

$$q(\zeta) = t(\zeta, \sigma, \alpha) - t(\zeta, \sigma, \tau) = t(\zeta, \tau, \alpha)\ ,$$

which is bounded. Outside Δ, $q(\zeta)$ is obviously bounded. It follows that

$$\overset{+}{t}(f(z), \sigma, \alpha) = \overset{+}{t}(f(z), \sigma, \tau) + O(1) .$$

We conclude the proof by substituting this in (12).

We note that if $\mathscr{T}'$ and $\mathscr{T}''$ are principal families defined with respect to the same point σ, then the functions $t'(\zeta, \sigma, \tau)$ and $t''(\zeta, \sigma, \tau)$ belonging to these families differ by a bounded harmonic function. Consequently the corresponding characteristic functtions T_1 and T_2 are related by $T_1(\Omega, f) = T_2(\Omega, f) + O(1)$ where $O(1)$ is bounded with respect to Ω.

Before defining functions of bounded characteristic we shall develop an alternative representation of the characteristic function. For this purpose we prove the following lemma.

LEMMA 5.2. *$N(\Omega, \zeta)$ is continuous on $S\backslash f(r)$.*

Proof. Let α be an arbitrary point in $S\backslash f(r)$, and let $a_1, \cdots, a_q$ with multiplicities $\nu_1, \cdots, \nu_p$ be the inverse images of α in Ω.

We can construct open connected neighborhoods D', D of α in $S\backslash f(r)$, and E'_j, E_j of a_j in $\Omega\backslash r$ for every j, such that the following properties hold:

(a) Each neighborhood lies in a parametric disk about its associated point.

(b) Every inverse image of $\zeta \in D'\backslash\alpha$ is simple and ζ has ν_j inverse images in E'_j.

(c) Every $z \in E'_j\backslash a_j$ is simple.

(d) $\bar{E}_j \subseteqq E'_j$.

(e) Every $\zeta \in D\backslash\alpha$ has ν_j roots in E_j, and $\bar{D} \subseteqq D'$.

$(E'_j\backslash a_j, f)$ is a smooth covering surface of S. If $\gamma(t)$ is an arc in D from an arbitrary $\delta \in D$ to α, its path of determination $\gamma'(t)$ from an inverse image of δ in E_j cannot intersect $E'_j\backslash E_j$ and must tend to a_j. Similarly if the inverse image is not in an E'_j, $\gamma'(t)$ must tend to $\partial\Omega$. Hence every component of the inverse image of D that intersects Ω is either a neighborhood of some a_j or intersects $\partial\Omega$.

Let $D_0 = \{\xi \mid |\zeta - \alpha| < \rho_0\}$ be a disk in D in terms of the local coordinates. Let F_{j0} be the component of the inverse image of D_0 that contains a_j, let G_{j0}, $j = 1, \cdots, n$ be the components that intersect the inverse images b_j, $j = 1, \cdots, n$, of α on $\partial\Omega$, and let H_{j0}, $j = 1, \cdots, m$, be the other components that intersect Ω. The number of components is finite since $\partial\Omega$ and ∂D_0 are analytic curves.

We define a real-valued function $h_j(z)$ on H_{j0} by $h_j(z) = |f(z) - \alpha|$. For each H_{j0} there exists $r_j > 0$ such that $h_j(z) > r_j$ for $z \in H_{j0} \cap \Omega$, and there exists a positive $r_0 < r_j$ for all j. Let $D'_0 = \{\zeta \mid |\zeta - \alpha| < r_0\}$.

Let M be a uniform bound of the number of inverse images in Ω of $\zeta \in S$. For $\varepsilon > 0$ and every j, there exist neighborhoods $\Delta(b_j) \subseteqq G_{j0}$ of b_j such that

$$|g(z, r) - g(a_j, r)| < \frac{\varepsilon}{M}, \qquad z \in \Delta(a_j),$$

and

$$|g(z, r)| < \frac{\varepsilon}{M}, \qquad z \in \Delta(b_j)$$

where $g(z, r)$ is the Green's function on Ω and vanishes outside Ω. Then $|N(\Omega, \zeta) - N(\Omega, \alpha)| < \varepsilon$ in the intersection of D'_0 and the images of $\Delta(a_j)$ and $\Delta(b_j)$ for all j. This completes the proof.

LEMMA 5.3. *If μ is a regular positive unit measure on a compact set $E \subseteqq S \backslash \sigma$, and if $p_\mu(\eta) = \int_E t(\eta, \zeta, \sigma)\, d\mu(\zeta)$ is the logarithmic potential with respect to any family $\mathscr{T}$, then*

$$-p_\mu(f(r)) = -\frac{1}{2\pi}\int_{\partial\Omega} p_\mu(f(z))\, d^*p(z, r) + N(\Omega, f) - \int_E N(\Omega, \zeta)\, d\mu(\zeta)\,. \tag{13}$$

Proof. By Lemmas 1.5 and 5.2 we may integrate Jensen's formula (4) over E and obtain

$$\int_E t(f(r), \sigma, \zeta)\, d\mu(\zeta) = \frac{1}{2\pi}\int_E d\mu(\zeta)\int_{\partial\Omega} t(f(z), \sigma, \zeta)\, d^*p(z, r) + N(\Omega, f) - \int_E N(\Omega, \zeta)\, d\mu(\zeta)\,.$$

We apply (11) and obtain the required result, which is the natural generalization of Frostman's formula.

The characterization of $T(\Omega, f)$ that we need is a consequence of the next theorem.

For a fixed $\sigma, \tau \in S$ and $r \in R$ such that σ, τ and $f(r)$ are distinct, we shall write $t(\zeta)$ for $t(\zeta, \sigma, \tau)$, $t_m(\zeta)$ for $\max\{m, t(\zeta)\}$ and $p(z)$ for $p(z, r)$.

THEOREM 5.4. *If $E_m = \{\zeta \mid t(\zeta) = m\}$ where m is finite and t belongs to the principal family $\mathscr{T}_1(I)$ with respect to the identity partition, then*

(14)
$$t_m(f(r)) = \frac{1}{2\pi}\int_{\partial\Omega} t_m(f(z))\, d^*p(z) + N(\Omega, f) - \frac{1}{2\pi}\int_{E_m} N(\Omega, \zeta)\, d^*t(\zeta)\,.$$

Proof. We first prove this theorem for the case in which some extra hypotheses hold, and then remove the restrictions.

We assume that either S is closed or that S is a regular region containing the image of $\bar{\Omega}$ and that $m \neq \lim t(\zeta)$ as $\zeta \to \xi \in \partial S$. We choose a unit mass distribution on the compact set E_m (oriented clockwise about σ) such that $d\mu = 1/2\pi\, d^*t(\zeta)$. Its logarithmic potential is

(15)
$$p_\mu(\eta) = \frac{1}{2\pi}\int_{E_m} t(\zeta, \eta, \sigma)\, d^*t(\zeta)$$

E_m divides S into two components, one containing σ and the other τ; we shall call them the σ_m- and τ_m-components. If S is a regular region one of these components is a neighborhood of the ideal boundary; we suppose that it is the σ_m-component.

If $m < t(\eta) < \infty$, then the flux of $t(\zeta, \eta, \sigma)$ is zero over the boundary, $E_m \cup \partial S$, of the σ_m-component and is also zero over the boundary, ∂S, of S; since $E_m \cap \partial S$ is void, it follows that the flux over E_m is zero. As $t(\zeta)$ is constant on E_m, it follows from (15) that

$$-p_\mu(\eta) = -\frac{1}{2\pi}\int_{E_m} [t(\zeta, \eta, \sigma)\, d^*t(\zeta) - t(\zeta)\, d^*t(\zeta, \eta, \sigma)]\,.$$

The application of Green's formula to $t(\zeta, \eta, \sigma)$ and $t(\zeta)$ over the τ_m-component proves that the right-hand side equals $-t(\tau, \eta, \sigma) = t_m(\eta)$.

If $-\infty < t(\eta) < m$, we write $t(\zeta, \eta, \sigma) = t(\zeta, \tau, \sigma) + t(\zeta, \eta, \tau)$ in (15). The flux of $t(\zeta, \tau, \sigma)$ is 2π and the flux of $t(\zeta, \eta, \tau)$ is zero over the boundary, E_m, of the τ_m-component. The first integral equals $-m$. We add a zero term and obtain

$$-p_\mu(\eta) = m - \frac{1}{2\pi}\int_{E_m} [t(\zeta, \eta, \tau)\, d^*t(\zeta) - t(\zeta)\, d^*t(\zeta, \eta, \tau)]$$

from (15). We apply Green's formula to $t(\zeta, \eta, \tau)$ and $t(\zeta)$ over the σ_m-component, and it follows that $-p_\mu(\eta) = m - t(\sigma, \eta, \tau) = m = t_m(\eta)$ by (1).

We obtain the same results if we suppose that the τ_m-component is a neighborhood of the ideal boundary.

Since the application of Lemma 4.1 to (15) shows that $p_\mu(\eta)$ is continuous at τ, we conclude that $-p_\mu(\tau) = t_m(\tau)$.

If $\eta \in E_m$ we note that $t(\zeta, \eta, \sigma)$ is superharmonic in the neighborhood of E_m. We consider the level lines $E_{m-\varepsilon}$ and $E_{m+\varepsilon}$, $\varepsilon > 0$. For sufficiently small ε, either $t(\zeta)$ is the capacity function on the $\tau_{m-\varepsilon}$-, τ_m- and $\tau_{m+\varepsilon}$-components or $-t(\zeta)$ is the capacity function on the corresponding σ-components. In either case we apply (8) to (15) and obtain

$$t_{m-\varepsilon}(\eta) \leqq -p_\mu(\eta) \leqq t_{m+\varepsilon}(\eta) ,$$

which yields $-p_\mu(\eta) = t_m(\eta)$.

We substitute in (13) and obtain (14).

To remove the restrictions we shall denote the intersection of E_k and the image of $\bar{\Omega}$ by E'_k. Then E'_k is compact and

$$\int_{E'_k} N(\Omega, \zeta)\, d^*t(\zeta) = \int_{E_k} N(\Omega, \zeta)\, d^*t(\zeta) .$$

If S is a regular region and $m = \lim t(\zeta)$ as $\zeta \to \xi \in \partial S$, we take $\varepsilon_0 > 0$ sufficiently small that

$$\{\zeta \mid \operatorname{grad} t(\zeta) = 0 \text{ and } m + \varepsilon \leqq t(\zeta) \leqq m\} \subseteqq E_m .$$

For $\varepsilon_0 > \varepsilon > 0$ we map $E'_{m+\varepsilon}$ into E_m along the level lines of $t^*(\zeta)$. These are well defined as the different branches of $t^*(\zeta)$ differ by an additive constant. The mapping is one-to-one except that onto each of the finite number of zeros of grad $t(\zeta)$ on E_m is mapped a finite number (one more than the order of the zero) of points on $E'_{m+\varepsilon}$.

On the image of $\bar{\Omega}$ we set the measures $d\mu_\varepsilon = 1/2\pi\, d^*t(\zeta)$ on $E'_{m+\varepsilon}$, $0 < \varepsilon < \varepsilon_0$. By Helly's theorem there exists a limiting measure that is obviously on E'_m. By the continuity of the normal derivative of $t(\zeta)$ it is, under the above mapping, $d^*t(\zeta)$, a.e. Hence, if $N_q = \min(N, q)$, we obtain

$$\lim_{\varepsilon\to 0} \int_{E'_{m+\varepsilon}} N_k(\Omega, \zeta)\, d^*t(\zeta) \leqq \int_{E'_m} N(\Omega, \zeta)\, d^*t(\zeta) .$$

The opposite inequality is obtained by Fatou's lemma. Consequently,

$$\lim_{\varepsilon\to 0} \int_{E_{m+\varepsilon}} N(\Omega, \zeta)\, d^*t(\zeta) = \int_{E_m} N(\Omega, \zeta)\, d^*t(\zeta) .$$

We now establish (14) for m by applying it to $m + \varepsilon$, which is permissible, and letting $\varepsilon \to 0$.

If S is arbitrary we consider an exhaustion of S by regular regions W such that W contains σ, τ and the image of $\bar{\Omega}$. We denote by $t_W(\zeta) = t_W(\zeta, \sigma, \tau)$ the function in the $\mathscr{T}_1(I)$ family defined with respect to W, and we set $E_{Wm} = \{\zeta \in W \mid t_W(\zeta) = m\}$.

Let W_0 be a regular region containing the image of $\bar{\Omega}$. We first consider m such that $E_m \cap \bar{W}_0$ contains no zeros of grad $t(\zeta)$, and cover it with a finite number of parametric disks. We select $\varepsilon_0 > 0$, such that the set

$$F = \{\zeta \mid \varepsilon_0 \leqq t(\zeta) \leqq m + \varepsilon_0\} \cap \bar{W}_0$$

is contained in these disks and does not contain any zeros of grad $t(\zeta)$. On each disk we use $t(\zeta)$ and any branch of $t^*(\zeta)$ as local variables.

Since E_m and ∂W_0 are analytic manifolds, their intersection consists of a finite number of components. Consequently there exists a compact $F' \subseteqq F$ such that the intersection of E_m and the image of $\bar{\Omega}$ is contained in the interior of F', and that $\partial F'$ intersects E_m at a finite number of points, each of which has a neighborhood in which $\partial F'$ lies on a level line of $t^*(\zeta)$. We set $E_m^* = E_m \cap F'$

Since $t_W(\zeta)$ and its normal derivative tend uniformly on compact sets to $t(\zeta)$ and its normal derivative, there is for any $\varepsilon > 0$, a W_ε such that

$$E_{Wm} \subseteqq \{\zeta \mid m - \varepsilon < t(\zeta) < m + \varepsilon\} \cap F' ,$$

and that the maximum angle between E'_{Wm} and E_m is less than $\pi/2$, for $W \supseteqq W_\varepsilon$. For sufficiently small ε we can map E^*_{Wm} univalently onto E_m^* along the level lines of $t^*(\zeta)$.

We have set up the set we need for the proof. We apply (14) to the region W and let $W \to S$. It is only necessary to examine the convergence of the last term. On $\bar{W}_0$ we choose a set of measures $d\mu_W = 1/2\pi\, d^*t_W(\zeta)$ on E^*_{Wm}. For sufficiently large W,

$$\int_{E^*_{Wm}} N(\Omega, \zeta)\, d^*t_W(\zeta) = \int_{E_{Wm}} N(\Omega, \zeta)\, d^*t_W(\zeta) .$$

We apply Helly's theorem as before and obtain the necessary convergence. Consequently, the theorem holds for open S if there is no zero of grad $t(\zeta)$ on E_m.

If grad $t(\zeta)$ has a zero on E_m, we apply (14) to $E_{m+\varepsilon}$ and take the limit as $\varepsilon \to 0$. To obtain the convergence of the last term, we choose the set of measures $d\mu_\varepsilon = 1/2\pi\, d^*t(\zeta)$ on $E^*_{m+\varepsilon}$ and apply Helly's theorem. This completes the proof.

By taking $m = 0$ in (14) we immediately obtain a generalization of Cartan's formula:

COROLLARY 5.5. *If the characteristic $T(\Omega, f)$ is defined in terms of a principal family $\mathscr{T}_1(I)$, then*

$$T(\Omega, f) = \overset{+}{t}(f(r)) + \frac{1}{2\pi} \int_{E_0} N(\Omega, \zeta)\, d^*t(\zeta) . \tag{16}$$

As a side issue we shall strengthen Lemma 5.2.

LEMMA 5.6. *If $f(\partial\Omega)$ is the image of $\partial\Omega$, then $N(\Omega, \zeta)$ is LP on $S \backslash f(\partial\Omega)$.*

Proof. Let $\alpha \in S\backslash(f(\partial\Omega) \cup f(r))$. We take σ at α, a parametric disk Δ at α, and an arbitrary τ. Let $t(\zeta) = t(\zeta, \sigma, \tau) \in \mathscr{T}_1(I)$. There exists m_0 such that $\{\zeta \mid t(\zeta) \leqq m_0\} \subseteqq \Delta\backslash(f(\partial\Omega) \cup f(r))$.

For $m \geqq m_0$, (14) yields

$$m = \frac{1}{2\pi}\int_{\partial\Omega} md^*p(z) + N(\Omega, \alpha) - \frac{1}{2\pi}\int_{E_m} N(\Omega, \zeta)\, d^*t(\zeta)\,,$$

which is

$$N(\Omega, \alpha) = \frac{1}{2\pi}\int_{E_m} N(\Omega, \zeta)\, d^*t(\zeta)\,.$$

Since $-t(\zeta)$ is the capacity function on the neighborhood of α bounded by $m \geqq m_0$, the function $N(\Omega, \zeta)$ is harmonic on $S\backslash(f(\partial\Omega) \cup f(r)$.

Let the multiplicity of r be k. By the construction used in Lemma 5.2 there is for any n, a neighborhood of $f(r)$ such that each ζ therein has k inverse images in

$$\{z \mid g(z, r) > n\}\,,$$

and a uniformly bounded number of other inverse images, for all of which $g(z, r)$ is uniformly bounded above. Hence $N(\Omega, \zeta)$ has a logarithmic singularity with coefficient k. This completes the proof.

6. Functions of bounded characteristic.

The remark after Theorem 5.1 shows that if the characteristic function $T(\Omega, f)$ is bounded with respect to Ω when it is defined in terms of one principal family $\mathscr{T}$, then it is also bounded when defined in terms of another family. We shall show that this property is also independent of the points σ, τ, r, provided that σ, τ and $f(r)$ are distinct.

For a fixed family $\mathscr{T}_1(I)$ and a fixed τ, we define

$$x(\Omega, q) = \frac{1}{2\pi}\int_{\partial\Omega} \overset{+}{t}(f(z), \sigma, \tau)\, d^*p_\Omega(z, q)\,,$$

$$y(\Omega, q) = \sum_i g_\Omega(s_i, q)\,, \qquad \{s_i\} = f^{-1}(\sigma) \cap \Omega\,,$$

$$x'(\Omega, q) = \frac{1}{2\pi}\int_{\partial\Omega} \overset{+}{t}(f(z), \tau, \sigma)\, d^*p_\Omega(z, q)\,,$$

$$y'(\Omega, q) = \sum_i g_\Omega(t_i, q)\,, \qquad \{t_i\} = f^{-1}(\tau) \cap \Omega\,,$$

and $u = x + y$, $u' = x' + y'$, where Ω is a regular region in R, and s_i, t_i are repeated in accordance with their multiplicities.

LEMMA 6.1. *If Ω exhausts R, then the limits of $y(\Omega, q)$ and*

$u(\Omega, q)$ are either LP (harmonic with positive logarithmic singularities) functions or $+\infty$; if the limit of $u(\Omega, q)$ is LP, then the limit of $x(\Omega, q)$ is harmonic.

Proof. The classical method is employed. We first prove that $\Omega_1 \subseteqq \Omega_2$ implies $u(\Omega_1, q) \leqq u(\Omega_2, q)$. Let $z \in \bar{\Omega}_1$. We write $t(\zeta)$ for $t(\zeta, \sigma, \tau)$. If $\overset{+}{t}(f(z)) > 0$, then $\overset{+}{t}(f(z)) = t(f(z)) = u(\Omega_2, z) - u'(\Omega_2, z)$ by Jensen's formula (14). Hence $\overset{+}{t}(f(z)) \leqq u(\Omega_2, z)$ for all $z \in \bar{\Omega}_1$. Consequently,

$$\frac{1}{2\pi}\int_{\partial\Omega_1} \overset{+}{t}(f(z))\, d^*p_1(z, q) \leqq \frac{1}{2\pi}\int_{\partial\Omega_1}\left[u(\Omega_2, z) - \sum_i g_1(s_i, z)\right] d^*p(z, q)\ ,$$

which is, by transposition,

$$u(\Omega_1, q) \leqq u(\Omega_2, q)\ .$$

For any fixed Ω_0 we exhaust R by $\Omega \supseteqq \Omega_0$. By the application of Harnack's principle to $u(\Omega, q) - u(\Omega_0, q)$ over Ω_0, we find that the limit of $u(\Omega, q)$ is LP or $+\infty$ over Ω_0 and hence over R.

By the maximum principle, $y(\Omega_1, q) \leqq y(\Omega_2, q)$ when $\Omega_1 \subseteqq \Omega_2$, and the same proof carries through.

If the limit of $u(\Omega, q)$ is LP, so is that of $y(\Omega, q)$: further, both functions have the same singularities. By taking the limit of $x(\Omega, q) = u(\Omega, q) - y(\Omega, q)$, we obtain the harmonicity of the limit of $x(\Omega, q)$. This completes the proof.

THEOREM 6.2. *If $T(\Omega, f)$ is bounded with respect to Ω, then it is bounded for any choice of r, τ, σ if $f(r)$, σ and τ are distinct.*

Proof. A subscript indicates functions defined in terms of the new parameters.

(a) If r is changed to r_1 such that $f(r_1) \neq \sigma$, then $T_1(\Omega, f) = u(\Omega, r_1)$ is bounded since the limit of $u(\Omega, q)$ is LP.

(b) If τ is changed to τ_1, we have

$$|\, T_1(\Omega, f) - T(\Omega, f)\,| \leqq \frac{1}{2\pi}\int_{\partial\Omega}\left|\overset{+}{t}_1(f(z)) - \overset{+}{t}(f(z))\right| d^*p(z, r)\ .$$

The integrand is bounded since the function $q(\zeta)$ in the proof of Theorem 5.1 is bounded.

(c) If σ is changed to σ_1, we may by (b) take σ_1 as τ in defining $T(\Omega, f)$. From the definitions of the terms

$$T(\Omega, f) - (m_1(\Omega, \sigma) + N_1(\Omega, \sigma))$$

is a constant function. It follows from Theorem 5.1 that $T(\Omega, f) - T_1(\Omega, f) = O(1)$.

We have established the fact that the following class of functions is well-defined.

DEFINITION. *An analytic function $f: R \to S$, where R is an arbitrary open surface and S an arbitrary surface, is of bounded characteristic, $f \in MS$, if $T(\Omega, f)$ is bounded with respect to $\Omega \subseteqq R$.*

It follows from Lemma 6.1 that if $f \in MB$, then the limit of $y(\Omega, q)$ is a positive superharmonic function on R. Consequently R is hyperbolic and we set

$$N(R, \zeta, r) = \lim_{\Omega \to R} N(\Omega, \zeta) = \lim_{\Omega \to R} \sum_i g_\Omega(z_i, r)\ , \qquad \{z_i\} = f^{-1}(\zeta) \cap \Omega\ .$$

Since ([3] p. 429) $N(R, \zeta, r) = \sum_i g(z_i, r)$, $\{z_i\} = f^{-1}(\zeta)$, where g is Green's function on R, it follows ([3] p. 418) that the class MB is identical with the Lindelöfian maps. We are able to obtain a characterization in terms of $N(R, \zeta, q)$.

THEOREM 6.3. *If $f: R \to S$ is analytic, the following statements are equivalent*:

(a) $f \in MB$

(b) *there exists $s \in R$ and open $U \subseteqq S$ such that $N(R, \zeta, s) < \infty$ for $\zeta \in U$,*

(c) $N(R, \zeta, s) < \infty$, $s \in R$, $\zeta \in S \backslash f(r)$.

Proof. To prove that (b) implies (a) we select $\alpha \in U \backslash f(s)$ and a parametric disk $\varDelta$ at α such that $\bar{\varDelta} \subseteqq U \backslash f(s)$.

Set $\varDelta_n = \{\zeta \in \bar{\varDelta}_n \mid N(R, \zeta, s) \leqq n\}$ then by Lemma 5.2 $N(R, \zeta, s)$ is lower semi-continuous and $\varDelta_n$ is closed. Also $\bar{\varDelta} = \cup \varDelta_n$. By Baire's category theorem there exists M such that $\varDelta_M$ has an interior point.

Let $Q \subseteqq \varDelta_M$ be an open region, and $\tau \in Q \backslash (\sigma \cup f(s))$. We define a family $\mathscr{T}_1(I)$ at σ; $t(\zeta, \sigma, \tau)$ has a level line E in Q. There is a principal family $\mathscr{T}_1'(I)$ such that $E = \{\zeta \mid t'(\zeta) = 0\}$, $t' \in \mathscr{T}_1'(I)$. Substitution in (16) yields

$$T(\Omega, f) \leqq \overset{+}{t'}(f(s)) + \int_E M d^* t'(\zeta) < \infty$$

for all Ω. Hence $f \in MB$.

(c) implies (b) trivially. To show that (a) implies (c) we note that by Lemma 6.1, $N(\Omega, \zeta)$ is bounded above for $s \in R$, $\zeta \in S \backslash f(s)$, whenever $f \in MB$. This completes the proof.

An extremal decomposition characterization of MB functions is given by the following:

THEOREM 6.4. *An analytic $f: R \to S$ is of class MB if and only* *if $t(f(z))$ is the difference between two LP functions, where t may be from any principal family $\mathscr{T}$.*

Proof. From the proof of Lemma 6.1,

$$t(f(z)) = u(\Omega, z) - u'(\Omega, z)$$

for all Ω. If $f \in MB$, then the limits of u and u' are LP functions. This proves the necessity.

For the sufficiency we assume $t(f(z)) = v(z) - w(z)$, $v, w \in LP$.

The singularities of $u(\Omega, z)$ are positive singularities of $t(f(z))$, and so among the singularities of $v(z)$. Hence $v(z) - u(\Omega, z)$ is superharmonic on Ω and attains its minimum on $\partial\Omega$.

Let $w \in \partial\Omega$. By (2), $x(\Omega, z)$ is the harmonic function on Ω with boundary values $t(f(w))$ and

$$\lim_{z \to w} x(\Omega, z) = \overset{+}{t}(f(w))$$

for any approach to w; also $y(\Omega, z) \to 0$ as $z \to w$, and $v(w) \geqq t(f(w))$. Consequently $v(z) - u(\Omega, z) \geqq 0$ on Ω.

Since $v(z) \in LP$, there exists $r \in \Omega$ such that $v(r) < \infty$. Hence $u(\Omega, r)$ is bounded for all Ω, and $f \in MB$. This concludes the proof.

The integrand of the proximity function used by Sario [8] is

$$\begin{aligned} s(\zeta, \alpha) &= t(\zeta, \alpha, \sigma) + \log(1 + e^{2t(\zeta)})(1 + e^{2t(\alpha)}), \quad \alpha \neq \sigma \\ &= \log(1 + e^{2t(\zeta)}), \qquad \alpha = \sigma . \end{aligned}$$

where $t \in \mathscr{T}_0$ and $t(\zeta) = t(\zeta, \sigma, \tau)$. A comparison of the characteristic functions, evaluated at σ, shows that the functions of bounded characteristic with respect to Sario's characteristic function are the same as those treated above.

BIBLIOGRAPHY

1. L. Ahlfors L. Sario, *Riemann surfaces*, Princeton University Press, Princeton, 1960.
2. M. Heins, *On the Lindelöf principle*, Ann. of Math., (2) **61** (1955), 440-473.
3. ———, *Lindelöfian maps*, ibid. (2) **62** (1955), 418-446.
4. K. V. R. Rao, *Lindelöfian maps and positive harmonic functions*, Technical Report No. 28, ARO (D) Project 1517-M, April 1962, 48 pp.
5. ———, *Positive harmonic functions, Lindelöfian meromorphic functions and Jensen's inequality*, ibid. No. 31, 12 pp.
6. L. Sario, *Meromorphic functions and conformal metrics on Riemann surfaces*, Pacific J. Math., **12** (1962), 1079-1097.
7. ———, *Value distribution under analytic mappings of arbitrary Riemann surfaces*, Acta Math., **109** (1963), 1-10.
8. ———, *Second main theorem without exceptional intervals on arbitrary Riemann surfaces*, Mich. J. Math., **10** (1963), 207-219.

UNIVERSITY OF CALIFORNIA, LOS ANGELES

CLIFFORD VECTORS

CURTIS M. FULTON

In this paper we present a generalization of parallel vector fields in a Riemannian space. As it turns out, such fields exist in spaces of constant positive curvature.

Restricting ourselves to a Riemannian 3-space throughout, we need the oriented third-order tensor [3, p. 249]

$$\eta_{ijh} = [\mathrm{sgn}(g)g]^{1/2}\varepsilon_{ijh} .$$

whose covariant derivative vanishes [3, pp. 251–252]. The latter fact is best ascertained by the use of geodesic coordinates. If we write the determinant of the metric tensor with the aid of permutation symbols we also find without difficulty

$$g^{pq}\eta_{ijp}\eta_{khq} = g_{hj}g_{ik} - g_{hi}g_{jk} . \tag{1}$$

DEFINITION. *Let the direction of a vector field at any point be that of the unit vector V. The field is said to consist of Clifford vectors if*

$$V_{i,j} = L\eta_{ijh}V^h , \qquad L \neq 0 . \tag{2}$$

THEOREM. *If the Riemannian curvature K is constant and equal to L^2, the system of equations (2) is completely integrable. If, at any point, solutions of (2) exist in all directions, then $K = L^2 = const.$*

It is known that integrability conditions for (2) are obtained using covariant differentiation. Hence, on account of a Ricci identity [3, p. 83] and (1) we have

$$L_{,k}\eta_{ijh}V^h - L_{,j}\eta_{ikh}V^h + L^2(g_{hj}g_{ik} - g_{hk}g_{ij})V^h = R_{hijk}V^h . \tag{3}$$

If the Riemannian curvature is constant [3, p. 112],

$$R_{hijk} = K(g_{hj}g_{ik} - g_{hk}g_{ij}) \tag{4}$$

and conditions (3) are identically satisfied. This proves the first part of our theorem.

For proof of the second part we multiply (3) by $W^iV^jW^k$ and get

$$L^2(g_{hj}g_{ik} - g_{hk}g_{ij})V^hW^iV^jW^k = R_{hijk}V^hW^iV^jW^k .$$

Thus L^2 is the Riemannian curvature associated with the unit vectors

Received August 29, 1963.

V, W [3, p. 95]. Assume now that W is a solution of (2) and M the corresponding scalar factor. Then the above curvature is also equal to M^2. Continuing this process we conclude from Schur's theorem [3, p. 112] that the curvature is constant and because of (4) that $K = L^2$.

To conclude, we demonstrate a geometric property of Clifford vectors justifying the name chosen for them. Let $\boldsymbol{t}$ be the unit tangent to a geodesic and $\boldsymbol{U}$ a unit vector which undergoes a parallel displacement along the geodesic. Hence $U^i{}_{,j}t^j = 0$ and $\boldsymbol{U}$ remains in a plane passing through the geodesic [1, p. 161]. On the other hand, because of (2), $V_{i,j}t^it^j = 0$ which shows that a Clifford vector, propagaged along the geodesic, is inclined at a constant angle to it. Letting $\cos\theta = U^iV_i$, we see that

$$-\sin\theta d_s\theta = L\eta_{ijh}U^it^jV^h .$$

We now make the simplifying assumption that both $\boldsymbol{U}$ and $\boldsymbol{V}$ are perpendicular to $\boldsymbol{t}$. In this case the vector $\eta_{ijh}U^iV^h$ has the direction of $\boldsymbol{t}$ and using (1) we find its length to be $\sin\theta$. Thus $d_s\theta = \pm L$ and the Clifford vector rotates about the geodesic in either sense through an angle proportional to the displacement. This property may be used to define the Clifford parallels or paratactic lines in elliptic 3-space [2, p. 108].

References

1. A. Duschek, W. Mayer, *Lehrbuch der Differentialgeometrie*, Band II, Teubner, Leipzig und Berlin, 1930.

2. D. M. Y. Sommerville, *The elements of non-Euclidean geometry*, Dover, New York, 1958.

3. J. L. Synge, A. Schild, *Tensor calculus*, University of Toronto Press, Toronto, 1959.

University of California, Davis

MAXIMAL ALGEBRAS AND A THEOREM OF RADÓ

I. GLICKSBERG

1. A theorem of Radó [1, 4, 6, 9] asserts that a function f, continuous on the closed disc $D = \{z : |z| \leqq 1\}$, and analytic at all points of the interior of D where f doesn't vanish, is analytic on all the interior. One can of course take this as a statement about the uniformly closed algebra A_1—the disc algebra—formed by those f in $C(D)$ analytic on the interior of D, and in fact it is easy to restate the result in a form which makes sense for any function algebra. For let $T^1 = \{z : |z| = 1\}$, and call f locally approximable at z if f can be uniformly approximated by elements of A_1 on some neighborhood of z. Then it is clear that the result asserts that any f in $C(D)$, locally approximable at all z in $D \backslash (T^1 \cup f^{-1}(0))$, is in A_1.

Now since D can be viewed as the maximal ideal space of A_1, and T^1 as the Šilov boundary, we can formulate such an assertion for any uniformly closed algebra of functions—and, needless to say, it will fail in general.[1] But under appropriate maximality conditions the result does hold; in particular we shall show it holds for any uniformly closed function algebra A maximal on its Šilov boundary, provided the boundary is not all the maximal ideal space of A, and for intersections of such algebras.

This result holds as a consequence of two facts: Rossi's local maximum modulus principle [11], and a quite elementary lemma (2.1) which allows one to eliminate certain points as candidates for elements of the Šilov boundary of an algebra. In the original setting, where the elementary local maximum modulus principle for analytic functions can be used, our proof requires (beyond this lemma) only the fact that the disc algebra A_1 is a maximal subalgebra of $C(T^1)$ [7, 12]; no doubt it is no simpler than the proof given in [6]. However our arguments do establish some nontrivial variants of the result in the general setting (3.5, 3.6, 4.9), and, in particular, for functions analytic on polycylinders in C^n; deflated to the disc algebra almost all of these follow rather easily from Radó's result due to the topological simplicity of the one (complex) dimensional situation and the fact that there Radó's result can be applied locally.

One consequence of Radó's theorem is the fact that A_1 is *integrally closed* in $C(D)$, i.e., any f in $C(D)$ satisfying a polynomial equation

Received September 26, 1963. Work supported in part by the National Science Foundation through Grant GP 1876.

[1] For example, for the subalgebra of A_1 of those f with $f'(0) = 0$; $f(z) \equiv z$ is locally approximable off $f^{-1}(0)$, but not in the subalgebra.

$$f^n + a_{n-1}f^{n-1} + \cdots + a_0 = 0$$

with coefficients in A_1 must lie in A_1. This extends to our maximal algebras (§ 5), and, as a consequence, for every uniformly closed subalgebra A of $C(\mathscr{M})$, where $\mathscr{M}$, the maximal ideal space of A, properly contains the Šilov boundary of A, we have a larger subalgebra a $C(\mathscr{M})$ with the same Šilov boundary which is integrally closed in $C(\mathscr{M})$.

Another consequence of one of our variants of Radó's theorem is the analogue, for intersections of maximal algebras, of the elementary removably singularity theorem for analytic functions (§ 6); from this one also has an analogue of the elementary facts on the behavior of analytic functions near isolated singularities, valid for functions locally approximable on $\mathscr{M}$ less a point.

Finally, the main portion of our argument can be applied to yield an abstract version of Schwarz's lemma: for any algebra A, if $f, g \in A$ and f/g is bounded on $\mathscr{M}\backslash g^{-1}(0)$ then it is bounded by its supremum over the Šilov boundary. Various consequences of this are given in § 4.

The author is indebted to Kenneth Hoffman and John Wermer for many helpful comments; in particular it was Wermer who observed that the author's original version of 2.2 could be used to prove Radó's theorem, and suggested its use to obtain integral closure.

We shall use $\boldsymbol{C}$ for the complex numbers, $\boldsymbol{R}$ for the reals, and F^0 for the interior of a set F.

2. In all that follows $C(X)$ will denote the Banach space of all bounded complex continuous functions on the space X, and A will denote a closed separating subalgebra of some $C(X)$, containing the constants. In general we shall view any such algebra A as a closed subalgebra of $C(\mathscr{M})$, where $\mathscr{M}$ is the maximal ideal space of A; when there is any necessity we may write $\mathscr{M}_A$ for $\mathscr{M}$. A *closed* subset X of $\mathscr{M}$ is a *boundary* for A if every f in A assumes its maximum modulus over $\mathscr{M}$ on X; any boundary is just a superset of the Šilov boundary ∂ of A.

Let X be a boundary for A, and let F be a closed *non-void* subset of X. An f in A will be said to *peak within* X *on* F if $f(F) = 1$ while $|f| < 1$ on $X\backslash F$. As is easily seen a point m of X lies in the Šilov boundary ∂ of A if and only if for every open neighborhood V of m in X there is an f in A which peaks within X on a nonvoid subset of V. The following lemma is fundamental to our considerations.

LEMMA 2.1. *Let $X \subset \mathscr{M}$ be a boundary for A, and V a (relatively) open subset of X. Suppose $g \in A$ peaks within X on a nonvoid subset of V, and let $\alpha = \sup |g(X\backslash V)|$ (which is necessarily <1).*

Then any $f \in A$ vanishing on V also vanishes on the nonvoid open subset $U = \{m \in \mathscr{M} : |g(m)| > \alpha\}$ of $\mathscr{M}$.

Proof. Suppose $|g(m)| > \alpha$ and $f(m) \neq 0$. Let μ be a (normalized, nonnegative, regular Borel) measure on X representing m, so $h(m) = \int h d\mu$ for all h in A [7, p. 181]. Let ν be the complex measure $(1/f(m))f\mu$ (the ordinary product of function and measure), which again represents m since

$$\int h d\nu = \frac{1}{f(m)}\int hfd\mu = \frac{1}{f(m)} \cdot h(m)f(m)\,, \qquad h \in A.$$

Now set $h = (1/g(m)) \cdot g \in A$; since $|g(m)| > \alpha = \sup|g(X\setminus V)|$ we have $h(m) = 1 > \sup|h(X\setminus V)|$. Replacing h by a sufficiently high power of itself we can suppose $\sup|h(X\setminus V)| < 1/(2\|\nu\|)$, where $\|\nu\|$ is the total variation norm of ν, while $h(m)$ is still 1.

Since $f(V) = 0$ the measure $\nu = (1/f(m))f\mu$ is carried by $X\setminus V$, so

$$1 = h(m) = \int h d\nu = \int_{X\setminus V} h d\nu < \frac{1}{2\|\nu\|} \cdot \|\nu\| = \frac{1}{2}\,,$$

the desired contradiction.

Our main applications of 2.1 will be made via the following corollary, and usually with the set $\mathscr{F}$ a singleton.

COROLLARY 2.2. *Let $X \subset Y \subset \mathscr{M}$ be boundaries for A, V a relatively open subset of X, and $\mathscr{F}$ any subset of A. If V is contained in the topological boundary in Y of $\bigcap_{f\in\mathscr{F}} f^{-1}(0)$, then $V \cap \partial = \phi$.*

Suppose $V \cap \partial \neq \phi$, so some g in A peaks within X on a nonvoid subset F of V. Then each f in $\mathscr{F}$ must vanish on the open subset U of $\mathscr{M}$ given in 2.1, and $F \subset U$, so F lies in the interior of $\bigcap_{f\in\mathscr{F}} f^{-1}(0)$ in Y, not in its boundary.

For a boundary X for A, A is called *analytic on* X if every f in A vanishing on a nonvoid relatively open subset of X vanishes identically (on X, hence on ∂, hence on all of $\mathscr{M}$). In [5] an example was given of an algebra A analytic on ∂ but not on $\mathscr{M}$; the original purpose of 2.1 was to prove

COROLLARY 2.3. *If A is analytic on $\mathscr{M}$, A is analytic on ∂.*

Indeed if $f \in A$ vanishes on a relatively open subset V of ∂ then some g in A must peak within ∂ on a nonvoid subset of V, so that f vanishes on a nonvoid open subset of $\mathscr{M}$ by 2.1. Thus we have the more general assertion of

COROLLARY 2.4. *For any algebra A, an f in A vanishing on a nonvoid relatively open subset of ∂ vanishes on a nonvoid open subset of $\mathscr{M}$.*

In particular if $f^{-1}(0)$ is nowhere dense in $\mathscr{M}$ then $f^{-1}(0) \cap \partial$ is nowhere dense in ∂. Both 2.3 and 2.4 remain valid if $\mathscr{M}$ is replaced by any boundary for A, but neither need hold if ∂ is enlarged to an arbitrary boundary; for example both fail for the disc algebra A_1, with ∂ replaced by $X = T^1 \cup \{0\}$, and $\{0\}$ the relatively open subset of X.

As we shall see later (4.2), 2.1 yields some further information on zero sets of elements of algebras with $\mathscr{M} \neq \partial$.

Some simple variants of 2.1 are of interest, but will not be needed in what follows. For example

COROLLARY 2.5. *Let U and V be as in 2.1. Then any bounded sequence $\{f_n\}$ in A converging pointwise to zero on V converges pointwise to zero on U.*

For $\theta > 1$ let $U_\theta = \{m \in \mathscr{M} : |g(m)| \geqq \theta\alpha\}$. Then any bounded sequence $\{f_n\}$ in A converging uniformly to zero on V converges uniformly to zero on U_θ.

Proof. For the first part, suppose $f_n(m) \not\to 0$ for some m in U; replacing $\{f_n\}$ by a subsequence we can assume $f_n(m) \to c \neq 0$. Let μ again represent m, and let f be any weak* cluster point of $\{f_n\}$ in $L_\infty(\mu)$. Since a subnet of $\{f_n\}$ converges weak* to f we have

$$h(m)\int f d\mu = \int hf d\mu\ , \qquad h \in A\ ,$$

while $c = \lim f_n(m) = \lim \int f_n d\mu = \int f d\mu$. So for $h = g/g(m)$ we have, for all n,

$$c = h(m)^n \int f d\mu = \int h^n f d\mu\ .$$

But by dominated convergence, for any f' in $L_1(\mu)$ vanishing off V we have $\int f' f d\mu = \lim \int f_n f' d\mu = 0$, and thus $f = 0$ a.e. μ on V. So, since $\sup |h(X\setminus V)| < 1$,

$$c = \lim \int_{X\setminus V} h^n f d\mu = 0\ ,$$

our contradiction.

The second assertion is entirely elementary. With $m \in U_\theta$, and μ and h as before, we have $|h| \leqq 1/|g(m)| \leqq 1/\theta\alpha$ on X, and $\leqq \alpha/\theta\alpha =$

$1/\theta$ on $X\backslash V$. Thus

$$|f_n(m)| = \left|\int h^k f_n d\mu\right| \leqq \left|\int_V\right| + \left|\int_{X\backslash V}\right|$$
$$\leqq \left(\frac{1}{\theta\alpha}\right)^k \sup|f_n(V)| + \frac{1}{\theta^k}\|f_n\|;$$

since $\theta > 1$ the last term will be $<\varepsilon/2$ for some large k, and choosing $n \geqq N$ will then force the sum below ε.

We might note that there are trivial variants of this second assertion which allow $\{f_n\}$ to be unbounded, provided the sequence $\{\sup|f_n(V)|\}$ approaches zero rapidly enough. For example, if

$$\sup|f_n(V)| = o(\|f_n\|^{-\log\alpha\theta/\log\theta}),$$

as is easily verified.

3. Let X be a boundary for A. We shall call a function f, defined on part of X, *locally approximable* (*within*[2] *X by A*) *at* $x \in X$ if, for some neighborhood U *of x in X*, f is defined on U and is uniformly approximable there by elements of A; alternatively[3] $f|U \in (A|U)^-$, the closure in $C(U)$ of $A|U$. We shall say f is *locally approximable on* $Y \subset X$ if f is locally approximable at each point of Y; note that by definition the set of points of X at which a given function is locally approximable is open in X.

We have $\partial_A \subset X \subset \mathscr{M}_A$. Call A *relatively maximal in* $C(X)$ if $A|X \neq C(X)$ and no closed proper subalgebra B of $C(X)$ containing $A|X$ has $\partial_B = \partial_A$. (Since $\partial_A \subset \partial_B$ necessarily, we are requiring properly larger subalgebras of $C(X)$ to have properly larger Šilov boundaries.) Note that A is relatively maximal in $C(\partial_A)$ if and only if $A|\partial_A$ is a maximal closed subalgebra of $C(\partial_A)$; on the other hand if $X \neq \partial_A$ it follows quite simply from Zorn's lemma that there is a (necessarily proper) closed subalgebra $B \supset A$ of $C(X)$ with the same Šilov boundary which is relatively maximal in $C(X)$. (As we shall see later, an example of an algebra which is relatively maximal in $C(\mathscr{M})$ but not maximal is the algebra of functions in $C(D^n)$, analytic on the interior of D^n, the unit polycylinder in $\boldsymbol{C}^n$.)

The following simple observation will extend the range of our results.

LEMMA 3.1. *If A is relatively maximal in $C(X)$ and $X \subset Y \subset \mathscr{M}_A$ then A is relatively maximal in $C(Y)$.*

---

[2] We shall omit these terms when the algebra and boundary are clear.

[3] $f|U$ is the restriction of f to U, $A|U = \{g|U : g \in A\}$. Trivially the uniform closure $(A|U)^-$ of $A|U$ in $C(U)$ is isometrically isomorphic to the closure of $A|U^-$ in $C(U^-)$, and at times we may write $(A|U)^-$ where $(A|U^-)^-$ might also be used.

Suppose B is a larger subalgebra of $C(Y)$ with $\partial_B \subset \partial_A$, so that $\partial_B = \partial_A \subset X$. Then $B \mid X$ is closed in $C(X)$, and since we can assume $X \neq Y$, $B \mid X \neq C(X)$ since each point of $Y \backslash X$ provides a multiplicative linear functional on $B \mid X$. But A is relatively maximal in $C(X)$, so $A \mid X = B \mid X$, and each f in B coincides on $\partial_B = \partial_A$ with a g in A; since Y can clearly be viewed as a subset of $\mathscr{M}_B$, and $f - g \in B$ must vanish on all of $\mathscr{M}_B$ since it vanishes on ∂_B, $f = g$ on Y, and $B = A$.

The following is our direct extension of Radó's theorem.

THEOREM 3.2. *Suppose A is relatively maximal in $C(\mathscr{M}_A)$ with $\mathscr{M}_A \neq \partial_A$, or, more generally, is an intersection of closed subalgebras of $C(\mathscr{M}_A)$ each having a Šilov boundary which is a proper subset of $\mathscr{M}_A$ and each relatively maximal in $C(\mathscr{M}_A)$.*

Then any f in $C(\mathscr{M}_A)$ which is locally approximable on $\mathscr{M}_A \backslash (\partial_A \cup f^{-1}(0))$ is in A.

Proof. Consider first the special case in which A is relatively maximal in $C(\mathscr{M}_A)$, and let us write $\mathscr{M}$, ∂ for $\mathscr{M}_A$, ∂_A. Let B be the closed subalgebra of $C(\mathscr{M})$ generated by A and f.

For each m in $U = \mathscr{M} \backslash (\partial \cup f^{-1}(0))$ we have an open neighborhood U_m of m contained in U for which $f \mid U_m \in (A \mid U_m)^-$, so clearly $h \mid U_m \in (A \mid U_m)^-$ for any h in B. As a consequence $m \notin \partial_B$; for otherwise some h in B must peak within $\mathscr{M}$ on a subset of the open set U_m, so for some m' in U_m

$$| h(m') | > \sup | h(U_m^- \backslash U_m) | .$$

Since $h \mid U_m \in (A \mid U_m)^-$ this contradicts Rossi's local maximum modulus principle [11] (which asserts that $\partial_{(A\,U)^-} \subset U^- \backslash U$ for any open $U \subset \mathscr{M} \backslash \partial$).

Similarly for any m in $f^{-1}(0)^0$, the interior of $f^{-1}(0)$, $m \notin \partial$, we have a neighborhood $U_m \subset \mathscr{M} \backslash \partial$ on which $f \mid U_m = 0 \in (A \mid U_m)^-$, and we again conclude that $m \notin \partial_B$. So $\partial_B \subset \partial \cup F$, where F is the topological boundary of $f^{-1}(0)$ in $\mathscr{M}$.

Now $F \backslash \partial$ is a relatively open subset of the boundary $X = F \cup \partial$ for B, and $F \backslash \partial$ lies in the topological boundary F of $f^{-1}(0)$ in the subspace $Y = \mathscr{M}$ of $\mathscr{M}_B$; so 2.2 applies, showing $(F \backslash \partial) \cap \partial_B = \phi$, whence $\partial_B \subset \partial$. Since $\partial \subsetneq \mathscr{M}$, B is proper in $C(\mathscr{M})$, and since A is relatively maximal in $C(\mathscr{M})$, $B = A$. Thus $f \in A$ as desired.[5]

For the more general case[6] let $A = \bigcap A_\alpha$, where $\partial_{A_\alpha} \subsetneq \mathscr{M}$, and each A_α is relatively maximal in $C(\mathscr{M})$. Clearly $\partial \subset \partial_{A_\alpha}$, and $\mathscr{M}$ is

[4] Actually $A \mid X \neq C(X)$ is redundant if $X \neq \partial_A$, as will usually be the case.

[5] Radó's theorem for A_1 now follows from Wermer's maximality theorem [**7**, 12].

[6] Our discussion here (and in later sections) would be considerably simplified if one had a positive answer to the following open question, raised some time ago by Kenneth Hoffman: if $A \subset B \subset C(\mathscr{M}_A)$ and $\partial_B = \partial_A$, must $\mathscr{M}_B = \mathscr{M}_A$?

a subspace of $\mathscr{M}_{A_\alpha}$ properly containing ∂_{A_α}, so $\partial_{A_\alpha} \neq \mathscr{M}_{A_\alpha}$. Let $\rho_\alpha : \mathscr{M}_{A_\alpha} \to \mathscr{M}$ be the map dual to the injection of A into A_α, which we can of course view as a retraction of $\mathscr{M}_{A_\alpha}$ onto its subspace $\mathscr{M}$; finally let $h \to \hat{h}$ denote the Gelfand representation of A_α— for h in A, in particular, $\hat{h} = h \circ \rho_\alpha$.

Now trivially $f \circ \rho_\alpha \in C(\mathscr{M}_{A_\alpha})$ is locally approximable (by $A \circ \rho_\alpha$, hence) by $\hat{A_\alpha}$ on $\mathscr{M}_{A_\alpha} \backslash (\partial \cup (f \circ \rho_\alpha)^{-1}(0))$, so certainly on $\mathscr{M}_{A_\alpha} \backslash (\partial_{A_\alpha} \cup (f \circ \rho_\alpha)^{-1}(0))$. Since $\hat{A_\alpha}$ is relatively maximal in $C(\mathscr{M}_{A_\alpha})$ by 3.1, $f \circ \rho_\alpha \in \hat{A_\alpha}$ by our special case, whence $f = (f \circ \rho_\alpha) \mid \mathscr{M}$ is in $\hat{A_\alpha} \mid \mathscr{M} = A_\alpha$: since this holds for every α, $f \in A$, completing the proof of 3.2.

The argument of the special case of 3.2 is central to all that follows (and will be needed again). There, in distinction to the more general case, the only property of $\mathscr{M}$ that is used is the local maximum modulus principle; $\mathscr{M}$ could just as well be any boundary X fo which

$$\text{(3.1)} \qquad \partial_{(A\,U)^-} \subset U^- \backslash U, \text{ for all relatively open } U \text{ in } X \backslash (\partial \cup F)\,,$$

where F is the boundary of $f^{-1}(0)$ in X. Moreover 3.2 evidently yields a positive assertion about any algebra A with $\mathscr{M} \neq \partial$; it will be worthwhile later to combine these observations in the following corollary to our proof, in which $\mathscr{M}$ can be taken as X.

COROLLARY 3.3. *Let $f \in C(X)$, where X is a boundary for A for which* (3.1) *holds. Let f be locally approximable (within X) on $X \backslash (\partial \cup f^{-1}(0))$ and let B be the closed subalgebra of $C(X)$ generated by A and f. Then*

(a) $\partial_B = \partial$ *(so that $B = A$ if A is relatively maximal in $C(X)$), and*

(b) *local maximum modulus applies to B on X, i.e., for an open $U \subset X \backslash \partial$,*

$$\partial_{(B|U)^-} \subset U^- \backslash U\,.$$

If $X = \partial$ the assertions of 3.3 are of course vacuous. (a) is of course proved in 3.2, and also follows from (b), whose proof is simply a modification of that of 3.2. For if $x \in U$ is not in F, the boundary in X of $f^{-1}(0)$, then x has a neighborhood U_x with $U_x^- \subset U \backslash F$ for which $f \mid U_x \in (A \mid U_x)^-$, so $h \mid U_x \in (A \mid U_x)^-$ for any h in $(B \mid U)^-$; thus $x \notin \partial_{(B\,U)^-}$ as in 3.2. On the other hand if x is in F then $x \notin \partial_{(B\,U)^-}$ by 2.2, so (b) follows.

3.3 has the following consequences.

THEOREM 3.4. *Let $\mathscr{M}_A \neq \partial_A$. Then there is a closed subalgebra B of $C(\mathscr{M}_A)$ containing A, with $\partial_B = \partial_A$, for which any f in $C(\mathscr{M}_A)$, locally approximable by B on $\mathscr{M}_A \backslash (\partial_A \cup f^{-1}(0))$, must lie in B.*

Consider any chain of closed subalgebras B of $C(\mathscr{M}_A)$ which have the same Šilov boundary as A and to which local maximum modulus applies on $\mathscr{M}_A : \partial_{(B|U)^-} \subset U^-\backslash U$, for U open in $\mathscr{M}_A\backslash\partial_A$. By just the argument used in 3.2, if B_0 denotes the closure of the union of the elements of the chain then $\partial_{(B_0|U)^-} \subset U^-\backslash U$ for any open $U \subset \mathscr{M}_A\backslash\partial_A$, so $\partial_{B_0} = \partial_A$ and local maximum modulus applies to B_0 on $\mathscr{M}_A$. By Zorn's lemma then we have a closed subalgebra B of $C(\mathscr{M}_A)$ maximal with respect to these properties, with $A \subset B$. But now for an f in $C(\mathscr{M}_A)$ which is locally approximable (within $\mathscr{M}_A$) by B on $\mathscr{M}_A\backslash(\partial_A \cup f^{-1}(0))$ we have by 3.3 precisely the same properties for the algebra generated by B and f; thus the latter coincided with B, and $f \in B$.

The following extension of 3.2, which allows us to replace 0 by a countable subset of C, merely adds a category argument to that of 3.2. In the original setting of Radó's theorem it can be obtained by a local application of that result (and category).

THEOREM 3.5. *Let A be relatively maximal in $C(\mathscr{M})$ with $\mathscr{M} \neq \partial$. Let E be a countable subset of C, and $\{F_n\}$ a sequence of nowhere dense hull-kernel closed*[7] *subsets of $\mathscr{M}$. If $f \in C(\mathscr{M})$ is locally approximable on*

$$\mathscr{M}\backslash(\partial \cup f^{-1}(E) \cup (\bigcup F_n)) \tag{3.2}$$

then $f \in A$.

If A is only an intersection of relatively maximal subalgebras of $C(\mathscr{M})$, each having its Šilov boundary proper in $\mathscr{M}$, then the same assertion holds if $\bigcup F_n$ is closed, in particular if $\{F_n\}$ is finite.

Proof. Suppose first that A is relatively maximal, and let B be the closed subalgebra of $C(\mathscr{M})$ generated by A and f. Actually f is locally approximable on an open subset W of $\mathscr{M}\backslash\partial$ which contains (3.2), and also contains the open sets $f^{-1}(e)^0\backslash\partial$, $e \in E$, as well; and so for each $m \in W$ we have a neighborhood U_m of m, $U_m \subset \mathscr{M}\backslash\partial$, for which $f \mid U_m \in (A \mid U_m)^-$, whence $h \mid U_m \in (A \mid U_m)^-$ for all h in B. Since W is open we can conclude from the local maximum modulus principle that $\partial_B \cap W = \phi$ as before.

Suppose $m \in \partial_B\backslash\partial$, so $m \in \partial_B\backslash(\partial \cup W)$. Since W contains (3.2) and each set $f^{-1}(e)^0\backslash\partial$, such an m must lie in $\bigcup F_n$, or in $f^{-1}(E)\backslash\bigcup_{e\in E} f^{-1}(e)^0$, which is contained in the union of the boundaries of the sets $f^{-1}(e)$. Thus $\partial_B\backslash\partial$ is contained in a countable union of closed subsets of $\mathscr{M}$, and, by category, if $\partial_B\backslash\partial \neq \phi$ one of the sets $F_n \cap (\partial_B\backslash\partial)$ or

[7] Recall that a subset of $\mathscr{M} = \mathscr{M}_A$ is hull-kernel closed if and only if it is of the form $\bigcap_{g\in\mathscr{G}} g^{-1}(0)$, where $\mathscr{G} \subset A$.

(boundary $f^{-1}(e)) \cap (\partial_B \backslash \partial)$ has nonvoid interior V in the locally compact space $\partial_B \backslash \partial$, hence in ∂_B.

Now if $V \subset$ (boundary $f^{-1}(e)) \cap (\partial_B \backslash \partial)$ then $e - f$ is an element of B which vanishes on the relatively open subset V of ∂_B, while V lies in the boundary in $\mathscr{M}$ of $(e-f)^{-1}(0)$, so that 2.2 implies $V \cap \partial_B = \phi$, our contradiction. Similarly if $V \subset F_n \cap (\partial_B \backslash \partial)$, then since F_n is hull-kernel closed it has just the form of the intersection in 2.2; since F_n is nowhere dense in $\mathscr{M}$, V lies in boundary $F_n = F_n$, so 2.2 again yields a contradiction, and we conclude that $\partial_B \subset \partial$, whence $B = A$ and $f \in A$.

For the final assertion of 3.5 we consider $\mathscr{M}$ as a subspace of $\mathscr{M}_{A_\alpha}$ as in 3.2, with ρ_α our retraction of $\mathscr{M}_{A_\alpha}$ onto $\mathscr{M}$. Since now $\rho_\alpha^{-1} F_n$ need not be nowhere dense in $\mathscr{M}_{A_\alpha}$, we let $Y = \mathscr{M}_{A_\alpha} \backslash \bigcup \rho_\alpha^{-1}(F_n)^0$, and let B_α be the closed subalgebra of $C(Y)$ generated by (the restrictions of) $A_\alpha^\wedge$ and $f \circ \rho_\alpha$.

Our hypothesis that F_n is nowhere dense implies $\mathscr{M} \subset Y$ since $\rho_\alpha^{-1}(F_n)^0 \cap \mathscr{M} \subset F_n^0 = \phi$. And our hypothesis that $\bigcup F_n$ is closed implies $K = \bigcup \rho_\alpha^{-1} F_n = \rho_\alpha^{-1}(\bigcup F_n)$ is closed so that

$$U = \mathscr{M}_{A_\alpha} \backslash (\partial_{A_\alpha} \cup K)$$

is an open subset of $\mathscr{M}_{A_\alpha} \backslash \partial_{A_\alpha}$ contained in the subspace Y of $\mathscr{M}_{A_\alpha}$. Trivially $f \circ \rho_\alpha$ is locally approximable by $A_\alpha^\wedge$ on an open subset of Y which contains all points of U except (possibly) those lying in the boundaries of $(f \circ \rho_\alpha)^{-1}(e)$, $e \in E$. But now any m in U at which $f \circ \rho_\alpha$ is locally approximable has an open neighborhood U_m in $\mathscr{M}_{A_\alpha} \backslash \partial_{A_\alpha}$ with $U_m^- \subset U$ for which $h \mid U_m \in (A_\alpha \mid U_m)^-$, $h \in B$; since U_m is open in $\mathscr{M}_{A_\alpha}$, we know $m \notin \partial_B$ by just the argument of 3.2.

Thus $m \in \partial_B \backslash \partial_{A_\alpha}$ implies m lies in the boundary in Y of some $(f \circ \rho_\alpha)^{-1}(e)$, or in $K \backslash \bigcup \rho_\alpha^{-1}(F_n)^0 \subset \bigcup \{\rho_\alpha^{-1}(F_n) \backslash \rho_\alpha^{-1}(F_n)^0\}$, i.e., in the boundary of one of the sets $\rho_\alpha^{-1}(F_n) \cap Y$ in Y; and now the argument of the special case shows $\partial_B \subset \partial_{A_\alpha}$. By 3.1 $A_\alpha^\wedge$ is relatively maximal in $C(Y)$, so $B_\alpha = A_\alpha^\wedge \mid Y$, and since $\mathscr{M} \subset Y$, $B_\alpha \mid \mathscr{M} = A_\alpha^\wedge \mid \mathscr{M} = A_\alpha$, and $f \in A_\alpha$. Hence $f \in A$, completing the proof of 3.5.

As noted, the only point in the proof of the special case of 3.2 in which $\mathscr{M}$ had to be the full maximal ideal space of A, rather than a subset properly larger than the Šilov boundary, was in the application of local maximum modulus. In some special situations classical local maximum modulus can be applied, and we can then avoid using all of the maximal ideal space. For example, for $x \in X$, a boundary for A, call a *non-constant* map ρ_x of the open disc D^0 onto a subset of X containing x an *analytic disc through* x if $g \circ \rho_x$ is analytic for each g in A. Then

THEOREM 3.6. *Let $X \neq \partial$ be a boundary for A, and suppose A is relatively maximal in $C(X)$. Let $f \in C(X)$ and let F be the topological boundary of $f^{-1}(0)$ in X.*

Suppose that for every x in $U = X\backslash(\partial \cup F)$ there is an analytic disc ρ_x through x for which $f \circ \rho_x$ is analytic on D^0. Then $f \in A$.

As before, define B to be the closed subalgebra of $C(X)$ generated by A and f; for every disc ρ_x in our hypothesis we have $h \circ \rho_x$ analytic on D^0 for $h \in B$ as a uniform limit of analytic functions.

Now if $\partial_B \cap U$ is nonvoid then [10, p. 138] the open set U must contain a strong boundary point x of B, and since ρ_x is non-constant some g in B must peak within $\rho_x(D^0)$ on a proper subset containing x. So $g \circ \rho_x$ assumes its maximum modulus at a point of D^0, yet is non-constant and analytic on D^0; we conclude that $U \cap \partial_B$ is void, and $\partial_B \subset \partial \cup F$. Now the remainder of the proof of 3.2 applies.

Other variations of this sort can be obtained. We have pointed out 3.6 mainly to note an apparently nontrivial variation of Radó's theorem which it yields for the polycylinder algebra—the algebra A_n of all functions continuous on the polycylinder D^n in C^n and analytic on its interior. Recall that for A_n, $\mathscr{M} = D^n$ and $\partial = T^n$; moreover if X is any closed subset of D^n containing the topological boundary of D^n in C^n (and thus a boundary for A_n) then[8] $A_n \mid X$ is relatively maximal in $C(X)$.

COROLLARY 3.7. *Suppose X is a closed subset of D^n containing the topological boundary of D^n in C^n. Suppose $f \in C(X)$, and through each point of $X\backslash T^n$ where f does not vanish we have an analytic disc in X on which f is analytic.*

Then f is an element of the polycylinder algebra A_n restricted to X.

(Note that we of course have analytic discs on which f is analytic through points of $f^{-1}(0)^0$. Here an analytic disc is simply an analytic map of D^0 into X, which need not be $(1-1)$, let alone bianalytic.)

Finally we should note that something slightly weaker than local

[8] This is no doubt well known; the proof for $n = 2$ is as follows, with $A = A_2 \mid X$. Suppose $A \subset B \subset C(X)$, and $\partial_B = \partial_A = T^2$. Each disc $D_0 = \{(z, w_0) : |z| \leqq 1\}$ with $|w_0| = 1$ lies in X and is a peak set of A (hence of B), since $(z, w) \to (1/2)(1 + \bar{w}_0 w)$ peaks there. Consequently [8, p. 227] $B \mid D_0$ is closed in $C(D_0)$ and $\partial_{B|D_0} \subset \partial_B \cap D_0 = T^2 \cap D_0 = \partial_{A|D_0}$. Since $A \mid D_0$ is the relatively maximal disc algebra and we now have $\partial_{B|D_0} = \partial_{A|D_0}$, we conclude that $A \mid D_0 = B \mid D_0$; thus $z \to b(z, w_0)$ is analytic on $|z| < 1$ for $b \in B$, $|w_0| = 1$. Similarly $w \to b(z_0, \omega)$ is analytic for $|z_0| = 1$. But now

$$\iint e^{i(m\theta + n\phi)} b(e^{i\theta}, e^{i\phi}) d\theta d\phi = 0$$

for n or $m > 0$, so $b = g \in A$ on $\partial_B = T^2$, whence $b - g$ must vanish on $X \subset \mathscr{M}_B$.

approximability can be used in its stead in 3.2–3.5: rather than insisting that f be uniformly approximable by elements of A on U_m (hence necessarily on U_m^-) as we have done, we need only insist on uniform approximation on

$$K = \{m\} \cup (U_m^- \backslash U_m) = \{m\} \cup \text{boundary } U_m. \tag{3.3}$$

For example, in 3.2, $f \mid U_m \in (A \mid U_m)^-$ was used only to show $(\mathscr{M} \backslash (\partial \cup f^{-1}(0))) \cap \partial_B = \phi$, and since [10, p. 138] strong boundary points are dense in ∂_B while $\mathscr{M} \backslash (\partial \cup f^{-1}(0))$ is open, it suffices to show $m \in \mathscr{M} \backslash (\partial \cup f^{-1}(0))$ cannot be a strong boundary point. But if m is a strong boundary point and f is uniformly approximable on (3.3) then $h \mid K \in (A \mid K)^-$ for all h in B while some h in B must have

$$| h(m) | > \sup | h(U_m^- \backslash U_m) | \tag{3.4}$$

since m is a strong boundary point not in $U_m^- \backslash U_m$. Now some h' in A satisfies (3.4) (since $h \mid K \in (A \mid K)^-$), contradicting local maximum modulus again.

It may be worthwhile to note what this yields for the disc algebra A_1: if $f \in C(D)$, and for each z in $D \backslash (T^1 \cup f^{-1}(0))$ there is an r_z, $0 < r_z \leqq \text{dist } (z, T^1 \cup f^{-1}(0))$ for which f can be approximated uniformly by polynomials on $\{z\} \cup \{z' : | z' - z | = r_z\}$, then $f \in A_1$. (Deleting $f^{-1}(0)$ everywhere, we have here simply a corollary to Wermer's maximality theorem for A_1 and the density of strong boundary points in ∂; from this the more general statement follows by Radó's theorem. Actually we can limit our z's to a dense countable set in $D \backslash (T^1 \cup f^{-1}(0))$ if we also assume that $r_z > k$ dist $(z, T^1 \cup f^{-1}(0))$ for some fixed $k > 0$.)

4. Schwarz's lemma. Our argument can also be applied to certain functions defined and continuous only on part of $\mathscr{M}$, for any algebra A. In particular, we have the following generalization of Schwarz's lemma (for $A = A_1$, take $g(z) \equiv z$), which has several consequences.

THEOREM 4.1. *Let f and g be in A and suppose f/g is bounded on $\mathscr{M} \backslash g^{-1}(0)$. Then*

$$\sup \left| \frac{f}{g} (\mathscr{M} \backslash g^{-1}(0)) \right| = \sup \left| \frac{f}{g} (\partial \backslash g^{-1}(0)) \right| \tag{4.1}$$

(In fact the assertion applies to the Gelfand representatives of any commutative Banach algebra.)

Proof. For each m in $U = \mathscr{M} \backslash (\partial \cup g^{-1}(0))$ let U_m be an open neighborhood of m with compact closure contained in U, chosen so small that $0 \in \boldsymbol{C}$ does not lie in the closed convex hull of $g(U_m^-)$. Then we have polynomials P_n for which $P_n(z) \rightarrow z^{-1}$ uniformly on $g(U_m)$, so that $(P_n \circ g) \mid U_m \rightarrow (1/g) \mid U_m$ in $C(U_m)$, and thus $(f/g) \mid U_m \in (A \mid U_m)^-$.

Letting B_0 be the uniformly closed subalgebra of $C(\mathscr{M}\backslash g^{-1}(0))$ generated by $A_0 = A \mid (\mathscr{M}\backslash g^{-1}(0))$ and f/g, we have

$$h \mid U_m \in (A \mid U_m)^- \tag{4.2}$$

for all h in B_0 and m in U.

Now let X be the closure of $\mathscr{M}\backslash g^{-1}(0)$ in $\mathscr{M}_{B_0}$, so that X is a boundary for the algebra $B_0\hat{}$ of Gelfand representatives of B_0. Set $B = B_0\hat{}\,|X$; B of course contains a continuous extension to X of each h in A_0, and of f/g, and we shall let h^* denote the extension of $h \mid (\mathscr{M}\backslash g^{-1}(0))$, for $h \in A$.

g^* cannot vanish on $\mathscr{M}\backslash g^{-1}(0)$. On the other hand g^* must vanish on $X\backslash\mathscr{M}$: for since $\mathscr{M}\backslash g^{-1}(0)$ is dense in X, $|g^*(x)| \geqq \varepsilon > 0$ implies x is in the closure of $\{m \in \mathscr{M} : |g(m)| \geqq \varepsilon\}$, which is already compact, so $x \in \mathscr{M}$. Consequently $g^{*-1}(0) = X\backslash\mathscr{M}$. Again since $\mathscr{M}\backslash g^{-1}(0)$ is dense in X, $g^{*-1}(0) = X\backslash\mathscr{M}$ must coincide with its boundary in X.

Now $\mathscr{M}\backslash g^{-1}(0) = X\backslash g^{*-1}(0)$ is open in X; on the other hand the imbedding of $\mathscr{M}\backslash g^{-1}(0)$ into X is a homeomorphism,[9] so that the relatively open subset $U = \mathscr{M}\backslash(\partial \cup g^{-1}(0))$ of $\mathscr{M}\backslash g^{-1}(0)$ is in fact relatively open (hence open) in X. Consequently each U_m is open in X and (4.2) suffices to show no m in U is in ∂_B, as in 3.2. So

$$\partial_B \subset X\backslash U \subset (X\backslash\mathscr{M}) \cup (\partial\backslash g^{-1}(0)) \subset (X\backslash\mathscr{M}) \cup F\,,$$

where F is the closure in X of $\partial\backslash g^{-1}(0)$, and $\partial_B \subset g^{*-1}(0) \cup F$.

If the relatively open subset $\partial_B\backslash F$ of ∂_B were nonvoid, then, since it lies in $g^{*-1}(0)$, hence in the boundary in X of this set, 2.2 would imply $\partial_B \cap (\partial_B\backslash F) = \phi = \partial_B\backslash F$; so $\partial_B\backslash F = \phi$, $\partial_B \subset F$ and trivially (4.1) follows. (Since the result applies to $A\hat{}$—with 1 adjoined if necessary—for any commutative Banach algebra A, the final assertion follows easily.)

Our first corollary to 4.1 gives some information about zero-sets which is quite familiar for the disc algebra: a (non-void) zero set $g^{-1}(0)$ ($g \in A$) disjoint from the Šilov boundary has a smallest neighborhood on which elements of A can vanish, while no f in A vanishing on $g^{-1}(0)$ can tend to zero faster than every power of g unless f vanishes on a neighborhood of $g^{-1}(0)$.

We first observe that (4.1) can be trivially improved to have $\partial\backslash(f^{-1}(0) \cup g^{-1}(0))$ in place of $\partial\backslash g^{-1}(0)$ on the right side of (4.1) (since f/g vanishes on $f^{-1}(0)\backslash g^{-1}(0)$).

[9] Trivially $B_0 \subset C(\mathscr{M}\backslash g^{-1}(0))$ implies the map of $\mathscr{M}\backslash g^{-1}(0)$ into $\mathscr{M}_{B_0}$ is continuous, while the map of $\mathscr{M}_{B_0}$ into $\mathscr{M}$ dual to the injection of A into B_0-restricted to the image of $\mathscr{M}\backslash g^{-1}(0)$ in $\mathscr{M}_{B_0}$-provides a continuous inverse.

COROLLARY 4.2. *Let f and g be in A, with $\phi \neq g^{-1}(0) \subset f^{-1}(0)$, and suppose* $\inf |g(\partial \backslash f^{-1}(0))| = \delta > 0$ *(which will of course be the case if $g^{-1}(0) \cap \partial = \phi$). If f/g^n is bounded on $\mathscr{M}\backslash g^{-1}(0)$ for every $n \geqq 1$ then f vanishes on*

$$g^{-1}(D_\delta^0)\,, \tag{4.3}$$

where D_δ^0 is the open disc about 0 of radius δ.

In particular, if $g^{-1}(0)$ is nonvoid and disjoint from the Šilov boundary, then any f in A vanishing on a neighborhood of $g^{-1}(0)$ vanishes on (4.3) with $\delta = \inf |g(\partial)|$.

By (4.1), modified as indicated,

$$\left|\frac{f}{g^n}\right| \leqq \sup\left|\frac{f}{g^n}(\partial\backslash(f^{-1}(0)\cup g^{-1}(0)))\right| = \sup\left|\frac{f}{g^n}(\partial\backslash f^{-1}(0))\right| \leqq \frac{\|f\|}{\delta^n}$$

on $\mathscr{M}\backslash g^{-1}(0)$ so that

$$|f(m)| \leqq \lim \|f\| \cdot \left|\frac{g(m)}{\delta}\right|^n = 0$$

if $0 \neq |g(m)| < \delta$. By hypothesis $f(m) = 0$ if $g(m) = 0$, so f vanishes on all of (4.3).

For convenience let us say $g \in A$ divides $f \in A$ if $f = gh$, $h \in A$.

COROLLARY 4.3. *Suppose A is analytic on $\mathscr{M}$ (§ 2), and $g \in A$ has $g^{-1}(0)$ nonvoid and disjoint from the Šilov boundary. Then if g divides a nonzero element f of A there is a largest integer n for which g^n divides f.*

Otherwise f/g^n is bounded for every n, and f must vanish on (4.3), hence on all of $\mathscr{M}$.

COROLLARY 4.4. *Suppose $A\,|\,\partial$ is an intersection of maximal closed subalgebras of $C(\partial)$, f and g are in A, and*[10] $\mathscr{M} \neq \partial \cup g^{-1}(0)$. *If f/g is bounded on $\mathscr{M}\backslash g^{-1}(0)$, and on ∂ has an extension in $C(\partial)$, then $f = gh$ for some h in A.*

---

[10] This hypothesis is superfluous if g does not vanish anywhere on ∂, but in general is essential to the result. For let $\mathscr{M}$ be the subset $(\{0\}\times D)\cup\{(r, z) : 0 \leqq r \leqq 1, |z| = 1\}$ of $\boldsymbol{R} \times \boldsymbol{C}$, and A all functions continuous on $\mathscr{M}$ and analytic on $\{0\} \times D^0$. Then $\mathscr{M}_A = \mathscr{M}$, $\partial_A = \mathscr{M}\backslash(\{0\} \times D^0)$ and setting $f(r, z) = r\bar{z}$, $g(r, z) = r$ we have $f/g(r, z) = \bar{z}$ so $f/g \notin A$. (If $0 \notin g(\partial)$ and $\mathscr{M} = \partial \cup g^{-1}(0)$ then each of the complementary sets ∂ and $g^{-1}(0)$ is open and closed; by a result of Šilov, or in fact by 3.2, the characteristic function of $g^{-1}(0)$ is an element of A. Since it vanishes on ∂ we conclude that $g^{-1}(0) = \phi$ and the assertion of 4.4 is vacuous.)

We shall only sketch the proof, which is quite similar to that of 4.1. Suppose first that $A\,|\,\partial$ is actually maximal. Let h_0 be an extension of f/g to $Y = \partial \cup (\mathscr{M}\backslash g^{-1}(0))$ with $h_0\,|\,\partial \in C(\partial)$; we now let B_0 be the uniformly closed algebra of *bounded* functions on Y generated by $A\,|\,Y$ and h_0. Of course we have $h\,|\,\partial$ and $h\,|\,(\mathscr{M}\backslash g^{-1}(0))$ continuous for $h \in B_0$, and this implies the $(1-1)$ map of Y into $\mathscr{M}_{B_0}$ is continuous when restricted to ∂ or $\mathscr{M}\backslash g^{-1}(0)$: we can view Y as a subset of $\mathscr{M}_{B_0}$, ∂ as a compact subspace.

As in 4.1 we let X be the closure of Y in $\mathscr{M}_{B_0}$, $B = \hat{B_0}\,|\,X$, and h^* the element of B corresponding to $h \in B_0$. If $g^*(x) \neq 0$ then for some $\varepsilon > 0$, x lies in the closure in X of $\{m \in \mathscr{M} : |\,g(m)\,| \geqq \varepsilon\}$, which is already compact in X as the continuous image of a compact subset of $\mathscr{M}\backslash g^{-1}(0)$. So $X\backslash g^{*-1}(0)$, an open subset of X, is contained in Y; thus $(X\backslash g^{*-1}(0))\backslash\partial$ is another subset of Y which is open in X. This last set is clearly the open subset $U = Y\backslash\partial = \mathscr{M}\backslash(\partial \cup g^{-1}(0))$ of Y, and U is open in X.

Now the imbedding of U into X is a homeomorphism (exactly as before; see footnote 9)), so any subset of U, open in $\mathscr{M}$, is open in X.

Consequently if we select, for $m \in U$, an open neighborhood U_m of m in $\mathscr{M}$ with $U_m^- \subset U$ (as in 4.1) for which $h^*\,|\,U_m = h\,|\,U_m \in (A\,|\,U_m)^-$, $h \in B_0$, then since U_m is in fact open in X the argument of 3.2 and local maximum modulus show $\partial_B \cap U_m = \phi$.

Thus $\partial_B \subset X\backslash U = X\backslash(Y\backslash\partial)$, so $\partial_B\backslash\partial \subset X\backslash Y$. But $X\backslash Y \subset g^{*-1}(0)$, as we have seen, so $X\backslash Y \subset g^{*-1}(0)\backslash\partial$, and $\partial_B\backslash\partial \subset g^{*-1}(0)\backslash\partial$. Since $Y\backslash\partial = \mathscr{M}\backslash(\partial \cup g^{-1}(0))$ is dense in $X\backslash\partial$ while g^* cannot vanish on this set, we clearly have $g^{*-1}(0)\backslash\partial$ contained in its boundary in X. So 2.2 applies to show $\partial_B\backslash\partial = \phi$, whence $\partial_B \subset \partial$ and $B\,|\,\partial$ is closed in $C(\partial)$.

By hypothesis $\mathscr{M}\backslash(\partial \cup g^{-1}(0)) \neq \phi$, so $\partial_B \subset \partial$ is proper in $\mathscr{M}_B$; thus $A\,|\,\partial \subset B\,|\,\partial \subsetneqq C(\partial)$, and by maximality $A\,|\,\partial = B\,|\,\partial$. Hence $h_0 = h$ on ∂ for some h in A, whence $f = gh$ on ∂, hence on all of $\mathscr{M}$.

Now if $A\,|\,\partial = \bigcap(A_\alpha\,|\,\partial)$ where each $A_\alpha\,|\,\partial$ is maximal in $C(\partial)$, then the preceding argument applied to A_α (with f and g taken in $C(\mathscr{M}_{A_\alpha})$) shows $h_0\,|\,\partial \in A_\alpha\,|\,\partial$; thus $h_0\,|\,\partial = h\,|\,\partial$ for some h in A, whence $f = gh$ on $\mathscr{M}$ as before.

COROLLARY 4.5. *Let $A\,|\,\partial$ be an intersection of maximal closed subalgebras of $C(\partial)$, and let g be an element of A with $\mathscr{M} \neq \partial \cup g^{-1}(0)$. Then any f in $C(\mathscr{M})$ with $fg \in A$ coincides on $\mathscr{M}\backslash g^{-1}(0)$ with an element of A.*

For fg/g is bounded on $\mathscr{M}\backslash g^{-1}(0)$ and on ∂ has the extension $f\,|\,\partial$ in $C(\partial)$, so that $fg = gh$ for some h in A by 4.4.

(If A is *analytic* on $\mathscr{M}$ (see § 2), $f \in A$; for then $g^{-1}(0)$ is nowhere dense in $\mathscr{M}$.)

Bishop [3, § 2, Lemma 3] has recently shown that (for any A) a point m in $\mathscr{M}\backslash\partial$ is represented by a (not necessarily unique) Jensen measure on ∂, i.e., there is a probability measure μ carried by ∂ for which Jensen's inequality holds:

$$\log|f(m)| \leqq \int \log|f|\,d\mu\,, \qquad f \in A\,.$$

(Applied to $f = e^{\pm g}$, $g \in A$, this yields $Re\,g(m) = \int Re\,g d\mu$ so that μ represents m on A.) As a consequence the argument of 4.4 yields

COROLLARY 4.5. *Suppose $f, g \in A$ and f/g is bounded on $\mathscr{M}\backslash g^{-1}(0)$ and on ∂ has an extension h_0 in $C(\partial)$. Then for each m in $\mathscr{M}\backslash(\partial \cup g^{-1}(0) \cup f^{-1}(0))$ there is a Jensen measure μ on ∂ representing m for which*

$$\int \log|g|\,d\mu - \log|g(m)| \leqq \int \log|f|\,d\mu - \log|f(m)|\,.$$

When (as in 4.4) f/g is actually the restriction of an element h of A, 4.5 follows trivially from Jensen's inequality for any Jensen measure μ representing m; for

$$\begin{aligned}\int \log|f|\,d\mu - \log|f(m)| &= \int \log|gh|\,d\mu - \log|gh(m)| \\ &= \left(\int \log|g|\,d\mu - \log|g(m)|\right) \\ &\quad + \left(\int \log|h|\,d\mu - \log|h(m)|\right)\end{aligned}$$

while the last term is nonnegative. In general we can construct the algebra B of 4.4, obtaining $\partial_B = \partial$ as there. Thus $m \in \mathscr{M}\backslash(\partial \cup g^{-1}(0))$, which provides an element of $\mathscr{M}_B\backslash\partial_B$, is represented on B by a Jensen measure μ on $\partial_B = \partial$ by Bishop's result. So

$$\log\left|\frac{f}{g}(m)\right| \leqq \int \log|h_0|\,d\mu$$

and

$$-\infty < \log|g(m)| \leqq \int \log|g|\,d\mu\,.$$

From the last we have $\partial \cap g^{-1}(0)$ a μ-null set so

$$\begin{aligned}\log|f(m)| - \log|g(m)| = \log\left|\frac{f}{g}(m)\right| &\leqq \int \log\left|\frac{f}{g}\right| d\mu \\ &\leqq \int \log|f|\,d\mu - \int \log|g|\,d\mu\end{aligned}$$

yielding 4.6.

If $A \mid \partial$ is not an intersection of maximal subalgebras of $C(\partial)$ and f, g are as in 4.4 one would not in general expect f/g to have an extension in $C(\mathscr{M})$— or even an extension to $\mathscr{M}$ continuous at all points of ∂. However this is the case if A has unique representing measures.[11]

COROLLARY 4.7. *Suppose each $m \in \mathscr{M}$ is represented by a unique (probability) measure on ∂. Let $f, g \in A$, with f/g bounded on $\mathscr{M}\backslash g^{-1}(0)$, and suppose that, on ∂, f/g has an extension in $C(\partial)$. Finally, suppose $\mathscr{M} \neq \partial \cup g^{-1}(0)$.*

Then f/g has an extension in $C(\mathscr{M})$.

Exactly as in 4.4 we form the closed subalgebra B of $C(X)$; X contains (a homeomorph of) ∂ and a continuous $(1-1)$ image of $\mathscr{M}\backslash g^{-1}(0)$ as before. Again we obtain $\partial_B \subset \partial$, so that each $m \in \mathscr{M}_B$ is represented by a probability measure μ_m on ∂, which is necessarily multiplicative on $A \subset B$, hence represents an element m' of $\mathscr{M}$; the map $m \rightarrow m'$ is of course nothing but the continuous map on $\mathscr{M}_B$ into $\mathscr{M}$ dual to the injection of A into B. But since representing measures for A are unique $m \rightarrow m'$ is $1-1$: for if m_1, m_2 both map into m' then $\mu_{m_1} = \mu_{m_2}$, whence $m_1 = m_2$.

Thus $\mathscr{M}_B$ is homeomorphic to a compact subset of $\mathscr{M}$ which necessarily contains $(\mathscr{M}\backslash g^{-1}(0)) \cup \partial$, so that h_0 (see 4.4) has a continuous extension to the closure of this set, hence to $\mathscr{M}$.

Actually in 4.1, 4.4 and 4.7 various other combinations of f and g (e.g., $f \exp(1/g)$) could be used in place of f/g. More generally f/g could be replaced by any h in $C(\mathscr{M}\backslash g^{-1}(0))$ which is locally approximable on $\mathscr{M}\backslash(\partial \cup g^{-1}(0))$, as is clear from their proofs. Thus

THEOREM 4.8. *Let $g \in A$ and suppose $h \in C(\mathscr{M}\backslash g^{-1}(0))$ is locally approximable on $\mathscr{M}\backslash(\partial \cup g^{-1}(0))$. Then*

$$\sup | h(\mathscr{M}\backslash g^{-1}(0)) | = \sup | h(\partial\backslash g^{-1}(0)) | .$$

Suppose that $\mathscr{M} \neq \partial \cup g^{-1}(0)$, while $h \mid \partial$ has an extension in $C(\partial)$. Then

(i) *If $A \mid \partial$ is an intersection of maximal subalgebras of $C(\partial)$,*

[11] More generally we could insist on uniqueness of the Jensen measure for each m (see [**3**, §2, Lemma 3]). An example where the assertion of 4.4 fails is the following which was pointed out by Wermer. Let $X = \{(z, w) \in \boldsymbol{C}^2 : |z| = 1 = |w|\}$, and A the closed subalgebra of $C(X)$ generated by the coordinate functions z, w, and all the functions w^m/z^n with $m > n > 0$. Then $\mathscr{M} = \{(z, w) \in \boldsymbol{C}^2 : |w| \leqq |z| \leqq 1\}$, the coordinate function $g = z$ vanishes only at $(0, 0)$ in $\mathscr{M}$, and w/z is bounded on $\mathscr{M}\backslash g^{-1}(0)$, but has no continuous extension to $\mathscr{M}$.

h is the restriction of an element of A.

(ii) *If each m in $\mathscr{M}$ has a unique representing measure on ∂ then h has an extension in $C(\mathscr{M})$.*

With sufficiently strong hypotheses we can also obtain an analogue of Radó's theorem in which continuity need not be assumed everywhere.

THEOREM 4.9. *Suppose $A\,|\,\partial$ is maximal in $C(\partial)$, $\mathscr{M} \neq \partial$. Let F be a relatively closed subset of $\mathscr{M}\backslash\partial$, E a countable subset of $\boldsymbol{C}$, and K a countable union of hull-kernel closed sets (for example, points) contained in F. Suppose $f \in C(\mathscr{M}\backslash F)$, f is locally approximable on $\mathscr{M}\backslash(\partial \cup F)$, $f^{-1}(e)$ is nowhere dense in $\mathscr{M}\backslash F$ for each e in E, and for every m in the boundary F_0 of F in $\mathscr{M}$, $m \notin K$, the cluster values of f at m lie in E, i.e.,*

$$\bigcap f(V\backslash F)^- \subset E$$

where the intersection is taken over all neighborhoods of m.

Then f is the restriction of an element of A.

Proof. Again for each m in $U = \mathscr{M}\backslash(\partial \cup F)$ we choose an open neighborhood U_m with $f\,|\,U_m \in (A\,|\,U_m)^-$, and let B_0 be the closed subalgebra of $C(\mathscr{M}\backslash F)$ generated by $A\,|\,(\mathscr{M}\backslash F)$ and f; thus

$$h\,|\,U_m \in (A\,|\,U_m)^- \tag{4.4}$$

for all m in U and h in B_0. We can again view $Y = \mathscr{M}\backslash F = \partial \cup (\mathscr{M}\backslash F)$ as a subset of $\mathscr{M}_{B_0}$, and ∂ as a compact subset. Let B be the restriction of the Gelfand representatives $B_0^\wedge$ to the closure X of Y in $\mathscr{M}_{B_0}$, and ρ the restriction to X of the map $\mathscr{M}_{B_0} \to \mathscr{M}$ dual to $A \to B_0$. Trivially $\rho(X)$ is the closure, in $\mathscr{M}$, of Y.

Now ρ is $1-1$ on $\rho^{-1}Y$, so $\rho^{-1}Y = Y$; for each h in B_0 is continuous on Y while for each x in X, $\hat{h}(x)$ is a cluster value of h at $\rho(x)$. Thus $\rho(x) = y \in Y$ implies $\hat{h}(x) = h(y) = \hat{h}(y)$, and $x = y$. (Since each U_m is open in $\mathscr{M}$, hence in $\rho(X)$, this implies $U_m = \rho^{-1}U_m$ is open in X.)

Each m in $\rho(X)\backslash Y$ lies in the boundary F_0 of F, clearly. Since each U_m is open in X, by local maximum modulus and (4.4) we have $\partial_B \cap U = \phi$, so $\partial_B \subset X\backslash U$; since $\rho^{-1}Y = Y$ and $Y = \partial \cup U$ we have $\rho(\partial_B) \subset \rho(X)\backslash U \subset \partial \cup (\rho(X)\backslash Y) \subset \partial \cup F_0$, so $\partial_B \subset \partial \cup \rho^{-1}(F_0)$. For each x in $\rho^{-1}(F_0)$, $\hat{f}(x)$ is a cluster value of f at $\rho(x)$, so that either $\hat{f}(x) \in E$ or $\rho(x) \in K = \bigcup_{i=1}^{\infty} K_i$ (where K_i is hull-kernel closed). Thus the locally compact space $\partial_B\backslash\partial \subset \rho^{-1}(K) \cup (\hat{f}^{-1}(E) \cap X)$, a countable union of closed subsets of X. By category, one of the sets $\rho^{-1}(K_i) \cap (\partial_B\backslash\partial)$ or $\hat{f}^{-1}(e) \cap (\partial_B\backslash\partial)$, $e \in E$, has a nonvoid relative interior V in $\partial_B\backslash\partial$ if $\partial_B\backslash\partial \neq \phi$.

Suppose $\rho^{-1}(K_i) \cap (\partial_B \backslash \partial)$ has nonvoid relative interior V. Then if $S = \{h \in A : h(K_i) = 0\}$, $K_i = \bigcap_{h \in S} h^{-1}(0)$, and since $\hat{f} = h \circ \rho$ on X, we have $\rho^{-1}(K_i) = \bigcap_{h \in S} \hat{h}^{-1}(0)$. Trivially $\rho^{-1}(K_i)$ is all boundary in X (since $\rho^{-1}(K_i) \cap Y = \phi$ and Y is dense), so 2.2 applies to yield the contradiction $V = \partial_B \cap V = \phi$.

Again if $f^{-1}(e) \cap (\partial_B \backslash \partial)$ has nonvoid relative interior V in $\partial_B \backslash \partial$, then $(e - \hat{f})^{-1}(0)$ contains V, and coincides with its boundary in X since $(e - \hat{f})^{-1}(0) \cap Y$ is nowhere dense in $Y = \mathscr{M} \backslash F$ by hypothesis, hence has a dense complement. Since this again yields $V = V \cap \partial_B = \phi$ by 2.2, we conclude that $\partial_B \backslash \partial = \phi$.

The remainder of the proof is now clear.

COROLLARY 4.10. *Suppose that the hypotheses of* 4.9 *are satisfied except for the requirement that* $A \mid \partial$ *be maximal in* $C(\partial)$. *Then the closed subalgebra of* $C(\mathscr{M} \backslash F)$ *generated by* f *and* $A \mid (\mathscr{M} \backslash F)$ *has the same Šilov boundary as* A.

5. **Integral closure.** For a boundary X of A we shall call A *integrally closed* in $C(X)$ if, when $a_0, a_1, \cdots, a_{n-1}$ are in A and $f \in C(X)$ then

$$p(f) = f^n + a_{n-1}f^{n-1} + \cdots + a_0 = 0 \quad \text{on } X \tag{5.1}$$

implies $f \in A$. We shall see that algebras to which 3.2 applies have this property for[12] $X = \mathscr{M}$ as a consequence of 3.2 and the implicit function theorem for analytic functions on C^n.

Recall that if F is analytic near $(z^0, w^0) = (z_1^0, \cdots, z_n^0, w^0)$ in C^{n+1}, $F(z^0, w^0) = 0$ and $(\partial F/\partial w) = F_{n+1}(z^0, w^0) \neq 0$ then, for some $\delta > 0$ and neighborhood V of z^0 in C^n, there is a unique function φ on V for which

$$F(z, \varphi(z)) = 0 \quad \text{and} \quad |\varphi(z) - w^0| < \delta\ ;$$

and φ is analytic on V. Consequently if

$$F(a_1, \cdots, a_n, f) = 0 \tag{5.2}$$

on a neighborhood of $m \in \mathscr{M}$, where $a_1, \cdots, a_n \in A, f \in C(\mathscr{M})$, and $a_i(m) = z_i^0, f(m) = w^0$, then

$$f = \varphi(a_1, \cdots, a_n)$$

near m. Thus f can be uniformly approximated by a power series in $a_1, \cdots, a_n$ near m, and for some neighborhood U_m of $m, f \mid U_m \in (A \mid U_m)^-$. So we have

[12] The same argument, using 3.3, yields this for any boundary X for which local maximum modulus applies to A on X, if A is relatively maximal in $C(X)$.

LEMMA 5.1. *Let* $a_1, \cdots, a_n \in A, f \in C(\mathscr{M})$, *and suppose* F *is analytic on a neighborhood of* $(a_1(m), \cdots, a_n(m), f(m))$ *in* C^{n+1} *while* (5.2) *holds on a neighborhood of* m. *Then* f *is locally approximable at* m *if* $F_{n+1}(a_1(m), \cdots, a_n(m), f(m)) \neq 0$.

We can now easily obtain the integral closure of the algebras in 3.2. Slightly more generally we have

THEOREM 5.2. *Suppose* A *is an intersection of relatively maximal subalgebras of* $C(\mathscr{M})$ *with Šilov boundaries proper in* $\mathscr{M}$. *If* $f \in C(\mathscr{M})$ *is locally approximable on* $\mathscr{M}\backslash\partial$ *outside the set where* (5.1) *holds then* $f \in A$; *in particular* A *is integrally closed in* $C(\mathscr{M})$.

Proof. f, and so $p(f)$, is locally approximable on $\mathscr{M}\backslash\partial$ except where $p(f) = 0$, so that $p(f) = a \in A$ by (3.2). Changing a_0, we can thus assume

$$p(f) = f^n + a_{n-1}f^{n-1} + \cdots + a_0 = 0$$

everywhere on $\mathscr{M}$. But now f is locally approximable off the set where

$$p'(f) = nf^{n-1} + (n-1)a_{n-1}f^{n-2} + \cdots + a_1 = 0$$

by 5.1, so $p'(f)$ is locally approximable off $p'(f)^{-1}(0)$, and $p'(f) \in A$ by 3.2. Continuing we finally have $(n!)f + a \in A$, and $f \in A$.

COROLLARY 5.3. *Suppose* A *satisfies the hypothesis of* 5.2, *while* $f \in A$ *does not have an nth root in* A *for some* $n > 1$. *Then* $\mathscr{M}\backslash f^{-1}(0)$ *is not simply connected (and if* $\mathscr{M}$ *is locally connected, some component of* $\mathscr{M}\backslash f^{-1}(0)$ *is not simply connected.)*

Finally, if A *is also analytic on* $\mathscr{M}$, $\mathscr{M}\backslash g^{-1}(0)$ *is connected for each* g *in* A.

If $\mathscr{M}\backslash f^{-1}(0)$ were simply connected we could find an h in $C(\mathscr{M}\backslash f^{-1}(0))$ with $h^n = f$ on $\mathscr{M}\backslash f^{-1}(0)$; setting $h = 0$ on $f^{-1}(0)$ we obtain an nth root of f in $C(\mathscr{M})$, and $h \in A$ by 5.2. (Similarly if the components of $\mathscr{M}\backslash f^{-1}(0)$ were simply connected we could find such an h on each component, and, if the components are open, we can combine these to again obtain an nth root in $C(\mathscr{M})$.)

Finally if A is also analytic on $\mathscr{M}$, and $\mathscr{M}\backslash g^{-1}(0) = U \cup V \neq \phi$, with U, V open and disjoint, then

$$h = \begin{cases} g \text{ on } U \cup g^{-1}(0) \\ -g \text{ on } V \end{cases}$$

defines an h in $C(\mathscr{M})$ which lies in A by 3.2; since $h+g$ or $h-g$ vanishes on a nonvoid open set (V or U) one vanishes identically. But $h=g$ implies $V=\phi$, $h=-g$ implies $U=\phi$, so $\mathscr{M}\backslash g^{-1}(0)$ must be connected.

Actually if A satisfies the hypothesis of 5.2 and is analytic on $\mathscr{M}$ then A is algebraically closed in $C(\mathscr{M})$ in the obvious sense. More generally such an A is *analytically closed* in $C(\mathscr{M})$ in the following sense.

Let $a_1, \cdots, a_n \in A$, $f \in C(\mathscr{M})$, and let F be a function analytic on a neighborhood in C^{n+1} of the range of the map

$$\rho\colon m \longrightarrow (a_1(m), \cdots, a_n(m), f(m))\ ;$$

despite our earlier notation we now let $F_k = [(\partial/\partial z_{n+1})^k F]\circ\rho$, $k \geqq 0$. Clearly F is not "independent of z_{n+1} on $\rho(\mathscr{M})$" if and only if $F_k(m) \neq 0$ for some m and $k \geqq 1$, and so we shall call A *analytically closed in* $C(\mathscr{M})$ if, for such a_i, f and F, with $F_k(m)\neq 0$ for some $k \geqq 1$ and m,

$$F(a_1, \cdots, a_n, f) = 0 \tag{5.3}$$

implies $f \in A$.

THEOREM 5.4. *If A is an intersection of relatively maximal subalgebras A_α of $C(\mathscr{M})$ each having its Šilov boundary proper in $\mathscr{M}$, and A is analytic on $\mathscr{M}$, then A is analytically closed in $C(\mathscr{M})$.*

For f is locally approximable on $\mathscr{M}\backslash(\partial \cup F_1^{-1}(0))$ by 5.1, so that F_1 is also, and $F_1 \in A$ by 3.2. Of course we may have $F_1 = 0$, but even then we know f (and so F_2) is locally approximable on $\mathscr{M}\backslash(\partial \cup F_2^{-1}(0))$, so that $F_2 \in A$ by 3.1; since not every $F_k = 0$ we have some F_k a nonzero element of A, and choosing k least, f is locally approximable on $\mathscr{M}\backslash(\partial \cup F_k^{-1}(0))$.

But now the final portion of 3.5 applies, with E void and $F_k^{-1}(0)$ our (single) hull-kernel closed subset of $\mathscr{M}$ (which is necessarily nowhere dense since $F_k \neq 0$ and A is analytic on $\mathscr{M}$).

For an algebra to which Radó's theorem applies the preceding argument shows (5.3) implies $F_k \in A$ for all k, and clearly we can replace $F_k^{-1}(0)$ in the proof by $K = \bigcap_k F_k^{-1}(0)$, with f locally approximable off this set; thus the hypothesis that K is nowhere dense is an adequate replacement for the analyticity of A on $\mathscr{M}$, yielding the first half of

THEOREM 5.5. *Suppose* (5.3) *holds with F appropriately analytic and $f \in C(\mathscr{M})$. Let $K = \bigcap F_k^{-1}(0)$.*

(1) *If A satisfies the hypotheses of 5.2 and K is nowhere dense, $f \in A$,*

(2) *If* $A|\partial$ *is maximal in* $C(\partial)$, *and* $\mathscr{M} \neq \partial \cup K$, *then* f *coincides with an element of* A *off the interior of* K.

Since $F_k \in A$ for all k, and f is locally approximable off K, (2) follows from 4.9 (with E void, and K the K of 4.9). Of course we could assume, as in 5.2, that 5.3 holds wherever f is not known to be locally approximable.

Actually any algebra A with $\mathscr{M} \neq \partial$ is contained in a subalgebra B of $C(\mathscr{M})$ given by 3.4 to which (1) applies, as is easily seen. In particular, B provides an integral closure of A in $C(\mathscr{M})$.

THEOREM 5.6. *Suppose* $\mathscr{M}_A \neq \partial_A$. *Then* A *is contained in a subalgebra* B *of* $C(\mathscr{M}_A)$ *which is integrally closed in* $C(\mathscr{M}_A)$ *and has* $\partial_B = \partial_A$. *Thus, in particular, if* $f \in C(\mathscr{M}_A)$ *satisfies* (5.1) *for* a_i *in* A, *the subalgebra of* $C(\mathscr{M}_A)$ *generated by* A *and* f *has* ∂_A *as its Šilov boundary.*

With B given by 3.4, the proof is precisely that of 5.2, with B in place of A.

Finally we should note that something stronger than integral closure in $C(\mathscr{M})$ holds for our intersections of relatively maximal algebra—we could require only that (5.1) holds locally on $\mathscr{M}\backslash\partial$, i.e., that each m in the (non-compact) space $\mathscr{M}\backslash\partial$ has a neighborhood on which an equation of the form (5.1) holds. Then, rather than invoking 3.2, we could simply show that for B, the subalgebra of $C(\mathscr{M})$ generated by A and f, one has $\partial_B \subset \partial$. (Indeed if $m \in \partial_B\backslash\partial$ and we choose p as in (5.1) of least possible degree with $p(f) = 0$ on a neighborhood U_m of m, then f and $p'(f)$ are locally approximable on $U_m\backslash p'(f)^{-1}(0)$, so m cannot lie in this open set—nor in its boundary by 2.2. Thus m is interior to $p'(f)^{-1}(0)$, so $p'(f)$ vanishes near m, contradicting the assumption that p had least degree.)

6. Removable singularities.

We next note an analogue of the elementary removable singularities theorem for analytic functions.

THEOREM 6.1. *Suppose* $A|\partial$ *is an intersection of maximal subalgebras* A_α *of* $C(\partial)$, $m \in \mathscr{M}\backslash\partial$, *and* f *is a bounded continuous function on* $\mathscr{M}\backslash\{m\}$ *which is locally approximable by* A *on* $\mathscr{M}\backslash(\partial \cup \{m\})$. *Then* $f \in A\,|\,(\mathscr{M}\backslash\{m\})$.

For each $\alpha, f \circ \rho_\alpha$ is locally approximable by $A_\alpha^\wedge$ on $\mathscr{M}_{A_\alpha}\backslash(\rho_\alpha^{-1}(m) \cup \partial)$, while $\rho_\alpha^{-1}(m)$ is a hull-kernel closed set[13] contained in $\mathscr{M}_{A_\alpha}\backslash\partial_{A_\alpha} = \mathscr{M}_{A_\alpha}\backslash\partial$.

[13] For $\{m\}$ is hull-kernel closed in $\mathscr{M}$ and $\rho_\alpha : \mathscr{M}_{A_\alpha} \to \mathscr{M}$ is continuous even when hull-kernel topologies are used.

Thus by 4.9 (with $E = \phi$, $F = K = \rho_\alpha^{-1}(m)$), $f \circ \rho_\alpha \in A_\alpha^\frown \mid (\mathscr{M}_A \backslash \rho_\alpha^{-1}(m))$, so $f \in A_\alpha \mid (\mathscr{M} \backslash \{m\})$ and $f \in A \mid (\mathscr{M} \backslash \{m\})$.

Trivially $\{m\}$ could just as well be any hull-kernel closed set in $\mathscr{M} \backslash \partial$. The result yields immediately an (imperfect) analogue of the elementary facts on behavior of analytic functions near isolated singularities.

COROLLARY 6.2. *Suppose $A \mid \partial$ is an intersection of maximal subalgebras of $C(\partial)$ and f is a continuous function on $\mathscr{M} \backslash \{m\}$, which is locally approximable on $\mathscr{M} \backslash (\partial \cup (m\})$. Then either*

(a) $f \in A \mid (\mathscr{M} \backslash \{m\})$

(b) $f = \text{const.} + 1/g$, $g \in A$, *and* $g^{-1}(0) = \{m\}$,

or

(c) *for each (deleted) open neighborhood V of m there is a compact K in $\boldsymbol{C}$ for which $f(V)$ is dense in $\boldsymbol{C} \backslash K$.*

Suppose (a) fails, so f cannot be bounded by 6.1. Let V be as in (c) and take $K = f(\mathscr{M} \backslash (V \cup \{m\}))$ which is compact. If (c) fails for this K then for some $z \in \boldsymbol{C} \backslash K$, $z - f$ is bounded away from zero on V; since it is also bounded away from zero on $\mathscr{M} \backslash (V \cup \{m\})$ (by the distance from z to K), $g = (1/(z - f)) \in C(\mathscr{M} \backslash \{m\})$, and is locally approximable on $\mathscr{M} \backslash (\partial \cup \{m\})$, so we can take g to be an element of A by 6.1. But now to obtain (b) we need only see that $g^{-1}(0) = \{m\}$; evidently g cannot vanish elsewhere, and if $g(m) \neq 0$ then g has a bounded inverse, whence f is bounded, contradicting our hypothesis that (a) fails.

Remark (Added in proof.) Wermer has pointed out the following completely elementary proof of 2.2, which actually applies if A is merely a multiplicative subsemigroup of $C(X)$. (For simplicity we shall suppose $\mathscr{F} = \{f\}$, a singleton):

For $x \in V$ we have a net $\{y_\delta\}$ in Y converging to x, with $f(y_\delta) \neq 0$ for each δ. Fixing δ, for $g \in A$ we have

$$\begin{aligned} |fg(y_\delta)| &\leqq \sup |fg(X)| = \sup |fg(X \backslash V)| \\ &\leqq \sup |f(X \backslash V)| \cdot \sup |g(X \backslash V)| \end{aligned}$$

so $|g(y_\delta)| \leqq c_\delta \sup |g(X \backslash V)|$, all g in A. Replacing g by its kth power and taking kth roots

$$|g(y_\delta)| \leqq c_\delta^{1/k} \sup |g(X \backslash V)|$$

whence $|g(y_\delta)| \leqq \sup |g(X \backslash V)|$. Since this holds for any δ,

$$|g(x)| \leqq \sup |g(X \backslash V)|, \ g \in A \ ,$$

so $X \backslash V$ is a boundary, and $V \cap \partial = \phi$.

An even shorter (but nonelementary) proof can be obtained using Bishop's result on the existence of Jensen (representing) measures [3], as Bishop observes in his forthcoming paper "Conditions for analyticity of certain sets" (§ 3).

REFERENCES

1. H. Behnke, K. Stein, *Modification komplexer Mannigfaltigkeiten und Riemannscher Gebiete*, Math. Ann., **124** (1951), 1-16.

2. E. Bishop, *A minimal boundary for function algebras*, Pacific J. Math., **9** (1959), 629-642.

3. E. Bishop, *Holomorphic completions, analytic continuation and the interpolation of seminorms*, Ann. of Math., **78** (1963), 468-500.

4. H. Cartan, *Sur une extension d'une theoreme de Radó*, Math. Ann., **125** (1952), 49-50.

5. I. Glicksberg, *A remark on the analyticity of function algebras*, Pacific J. Math., **13** (1963), 1181-1185.

6. E. Heinz, *Ein elementarer Beweis des Satzes von Radó-Behhnke-Stein-Cartan über analytischen Functionen*, Math. Ann., **131** (1956), 258-259.

7. Kenneth Hoffman, *Banach spaces of analytic functions*, Prentice-Hall, Englewood Cliffs, N. J., 1962.

8. K. Hoffman and I. M. Singer, *Maximal algebras of continuous functions*, Acta Math., **103** (1960), 217-241.

9. T. Radó, *Über eine nicht-fortsetzbare Riemannsche Mannigfaltigkeit*, Math. Z., **20** (1924), 1-6.

10. C. E. Rickart, *General theory of Banach algebras*, Van Nostrand, Princeton, 1960.

11. H. Rossi, *The local maximum modulus principle*, Ann. of Math., **72** (1960), 1-11.

12. J. Wermer, *On algebras of continuous functions*, Proc. Amer. Math. Soc., **4** (1953), 866-869.

UNIVERSITY OF WASHINGTON

MINIMUM PROBLEMS OF PLATEAU TYPE IN THE BERGMAN METRIC SPACE

KYONG T. HAHN

Dedicated to my teacher Professor C. Loewner on his seventieth birthday

1. Introduction. In this paper we are concerned with the existence of minimal surfaces with respect to the B-area (see § 4) and related problems in a bounded domain D in the space C^2 of two complex variables z_1, z_2.

Let $K_D(z, \bar{z})$, $z = (z_1, \cdots, z_n)$, be the Bergman kernel function of a bounded domain D in the space C^n of n complex variables. Throughout this paper, we assume $K_D(z, \bar{z})$ has the boundary value infinity at every point on the boundary of D. The kernel $K_D(z, \bar{z})$ enables us to define the Bergman metric

$$ds_D^2(z) = \sum_{\mu,\nu=1}^{n} T_{\mu\bar{\nu}}(z, \bar{z}) dz_\mu d\bar{z}_\nu, \quad T_{\mu\bar{\nu}} = \frac{\partial^2 \log K_D}{\partial z_\mu \partial \bar{z}_\nu}, \tag{1.1}$$

which is invariant with respect to pseudo-conformal mappings [4, pp. 51–53]. Using (1.1) we construct (see § 2) the complete Bergman metric space (D, d) over D and state a theorem for complete Riemannian spaces that for any two points in D, there exists a minimal curve with respect to d which connects the two points.

In § 3 we show that, if D is a plane domain bounded by finitely many boundary components $b_1, b_2, \cdots, b_n$, then there exists a minimal closed curve with respect to d among those curves which are homotopic to a fixed inner boundary component, say b_1, in $\overline{D(b_1)}$ (see § 3 for notation). If D is doubly connected, there exists a unique minimal closed curve in D. Furthermore, we prove a distortion theorem which gives bounds for the Bergman lengths of the minimal closed curves.

Analogous results are obtained in the case of two complex variables replacing the length by the B-area.

For a closed Jordan curve Γ in a complete metric space (D, d), we ask whether there exists a minimal surface with respect to the B-area which spans Γ. Answers to this question which constitute the main result of this paper are given in § 4.

As a generalization of § 3, we consider a domain D which is topologically equivalent to a product domain of the form $D_1 \times D_2$,

Received July 5, 1963. This paper represents the essential contents of the author's doctor thesis at Stanford University. The author wishes to express his thanks to Professor S. Bergman and Professor C. Loewner for their pertinent criticism in the preparation of this paper. This research was supported by Navy Contract Nonr-225(11) and NSF Grant 21344.

where D_k is a bounded domain as considered in § 3. When does there exist a minimal closed surface with respect to the B-area among those surfaces which are homotopic to T_1 in $\overline{D(T_1)}$ (see § 5 for notation)?

Answers are given in § 5. Distortion theorems for the minimal surfaces are given in § 6.

2. The Bergman metric space. A (continuous) curve c in D is said to be *regular* if it admits a regular (parametric) representation, i.e., there exists a continuously differentiable representation

$$G \mid I: z_k = G_k(t),\ k = 1, 2, \cdots, n,\ t \in I = [a, b]\ , \tag{2.1}$$

and dG_k/dt never vanish simultaneously at any $t \in I$. A curve c in D is said to be *piecewise regular* if it admits a piecewise regular representation, i.e., there exists a partition Δ: $a = t_0 < t_1 < \cdots t_{m-1} < t_m = b$ such that $G \mid [t_{k-1}, t_k]$ is regular for $k = 1, 2, \cdots, m$.

For a piecewise regular curve c given by (2.1), we define

$$L_D(c) = \int_a^b \Big[\sum_{\mu,\nu=1}^{n} T_{\mu\bar{\nu}}(G(t), \overline{G(t)}) \frac{dG_\mu}{dt} \frac{\overline{dG_\nu}}{dt} \Big]^{1/2} dt\ . \tag{2.2}$$

$L_D(c)$ is independent of the choice of piecewise regular representations of c. $L_D(c)$ will be called the *Bergman length* of c.

For any two points z^1 and z^2 in D, we define a distance function d by

$$d(z^1, z^2) = \inf_c L_D(c)\ , \tag{2.3}$$

where c runs over all piecewise regular curves which connect z^1 and z^2. Then the following theorem holds [15, § 16].

THEOREM 2.1. *d satisfies all the axioms for a metric and the metric space (D, d) is topologically equivalent to the metric space (D, ρ) with the Euclidean metric ρ. Moreover, the metric space (D, d) is finitely connected in the sense that every pair of points in D can be connected by a curve of finite Bergman length.*

The metric space (D, d) will be called the *Bergman metric space* over D. The significance of this metric space is that all metric properties are invariant under pseudo-conformal mappings.

We define the length (generalized) of a continuous curve c in D in the following way: For a partition $\Delta(I) = \{I_1, I_2, \cdots, I_m\}$, $I_k = [t_{k-1}, t_k]$, $k = 1, 2, \cdots, m$, of I, we define

$$\sigma(G; \varDelta(I)) = \sum_{k=1}^{m} \sigma(G; I_k), \quad \sigma(G; I_k) = d(G(t_k), G(t_{k-1})) .$$

Further, we define

$$\mathscr{L}_D(c) = \sup_{\varDelta} \sigma(G; \varDelta(I)) ,$$

where $\varDelta$ runs over all possible partitions of I. Then $\mathscr{L}_D(c)$ is independent of the choice of continuous representations of c. Clearly, the functional $\mathscr{L}_D$ is lower semi-continuous, i.e.,

$$\mathscr{L}_D(c) \leqq \liminf_{k\to\infty} \mathscr{L}_D(c_k), \text{ if } c_k \to c .$$

Further, for every piecewise regular curve c, $L_D(c) = \mathscr{L}_D(c)$ [15, § 16].

If $\mathscr{L}_D(c) < \infty$, c is said to be *rectifiable*. A curve is said to be *completely degenerated* if there is a representation $G \mid I$ such that G is constant on I. A representation $G \mid I$ is said to be *normal* if $\mathscr{L}_D(G; [t, t']) = t' - t$, for $t, t' \in I, t < t'$.

Let c be a rectifiable curve which is not completely degenerated. Then c admits a normal representation $G \mid [0, \mathscr{L}_D(c)]$, If we set $F(t) = G(t\mathscr{L}_D(c))$, $t \in I_0$, $I_0 = [0, 1]$, then $F \mid I_0$ is also a representation of c. Such a representation $F \mid I_0$ is called a *reduced representation* of c. For a closed curve, F is defined on $(-\infty, \infty)$ and is periodic of period 1. It is, therefore, enough to consider F on I_0. If $F \mid I_0$ is a reduced representation of a curve c, then the inequality

$$d(F(t), F(t')) \leqq \mathscr{L}_D(c) \mid t - t' \mid \tag{2.4}$$

holds for every $t, t' \in I_0$.

A metric space is called *complete* if every bounded infinite subset contains a limit point in the metric space. If D is homogeneous, (D, d) is always complete. Further, for every bounded generalized analytic polyhedron $D, (D, d)$ is complete. This is a result of S. Kobayashi (see [11] for details). For domains D in the space $\boldsymbol{C}^n$, $n \leqq 2$, Bergman has shown that the distance from a point in D to the boundary becomes infinite under certain hypothesis on the boundary of D [1], [6, Chap. III]. It is clear, in this case, that the metric space (D, d) is complete. Without going into great details in this direction, we shall assume in the sequel that the metric space (D, d) is always complete.

A curve K in (D, d) which connects z^1 and z^2 in D is called a *minimal curve* between z^1 and z^2 if $\mathscr{L}_D(K) \leqq \mathscr{L}_D(c)$ for all curves c connecting z^1 and z^2.

THEOREM 2.2. *For any two points z^1 and z^2, $z^1 \neq z_2$, in (D, d), there exists a minimal curve K between z^1 and z^2. Further, the* *minimal curves are analytic* (see [10] or [15, § 17]).

3. The existence of a minimal closed curve in a plane domain and its distortion theorem. We consider a multiply connected bounded domain D in the space $\boldsymbol{C}^1$ bounded by N Jordan closed curves $b_1, b_2, \cdots, b_n$, where b_n is the outer boundary component. Let (D, d) be the Bergman metric space derived from the Bergman metric

$$ds_D^2(z) = \boldsymbol{K}_D(z, \bar{z}) \, |\, dz \,|^2 \, . \tag{3.1}$$

It is assumed that (D, d) is complete. Then all the previous considerations, lemmas and theorems can be carried over to this case. We fix an inner boundary component of D, say b_1. Without loss of generality, we may assume b_1 to be a circle.

Let $\mathfrak{K}(D; b_1)$ be the class of all closed continuous curves c in D which are homotopic to b_1 in $\overline{D(b_1)}$, where $D(b_1)$ is a ring domain bounded by b_1 and b_n which contains the domain D, and $\overline{D(b_1)}$ is the closure of $D(b_1)$. A curve $K(D; b_1)$ in $\mathfrak{K}(D; b_1)$ which satisfies the condition $\mathscr{L}_D(K(D; b_1)) \leqq \mathscr{L}_D(c)$ for all $c \in \mathfrak{K}(D; b_1)$, will be called a *minimal closed curve* of D with respect to b_1. Due to the completeness of $(D; d)$ and the behavior of $\boldsymbol{K}_D(z, \bar{z})$ (described on page 943) on the boundary of D, we have

THEOREM 3.1. *There exists a minimal closed curve $K(D; b_1)$ of the domain D with respect to b_1. Further, it is analytic.*

Proof. Let $\gamma = \inf_c \mathscr{L}_D(c)$, where c runs over the class $\mathfrak{K}(D; b_1)$. Then $0 < \gamma < \infty$. There exists a minimizing sequence $\{c_k\}$ of rectifiable curves in $\mathfrak{K}(D; b_1)$. Let $G_k \,|\, I_0$ be the reduced representation of c_k. By (2.4), we have

$$d(G_k(t), G_k(t')) \leqq \mathscr{L}_D(c_k) \,|\, t - t' \,| \text{ for each } k \, ,$$

and $\{\mathscr{L}_D(c_k)\}$ has an upper bound δ which is finite. We choose an M such that $M^{1/2} > \delta / l(b_1)$, $l(b_1)$ is the Euclidean length of b_1. Then no c_k lies completely in $D - D_M$, $D_M = [z \,|\, \boldsymbol{K}_D(z, \bar{z}) \leqq M]$. Let

$$\rho = \max_{z_1 \, z_2 \in D_M} d(z_1, z_2) \, ,$$

then for every pair of positive integers p and q, we have $d(G_p(t), G_q(t)) < \rho + 2\delta$, $0 \leqq t \leqq 1$. Hence, we can select a subsequence $\{G_{k_i}\}$ of $\{G_k\}$ which converges uniformly to a continuous function G^0 on I_0. Let K be the closed curve whose representation is given by $G^0 \,|\, I_0$. Since $c_{k_i} \to K$, and by the lower semi-continuity of $\mathscr{L}_D$, we obtain $\mathscr{L}_D(K) = \gamma$. The analyticity of K is obvious.

THEOREM 3.2. *Every doubly connected domain has a unique minimal closed curve. It is analytic.*

Proof. We shall show first that annulus $Q = [z \mid r < |z| < 1]$ has a unique minimal closed curve given by $c_0 = [z \mid |z| = r^{1/2}]$. Let $P_1 = [z \mid r < |z| < r^{1/2}]$, $P_2 = [z \mid r^{1/2} < |z| < 1]$. If $c \cap P_1 = \phi$, it is immediate that $L_Q(c_0) \leqq L_Q(c)$, since the kernel function $\boldsymbol{K}_Q(z, \bar{z})$[1] assumes its minimum on c_0. If $c \cap P_2 = \phi$, by the conformal mapping $\zeta = r/z$, we have $\tilde{c} \cap P_1 = \phi$, where $\tilde{c}$ is the image curve of c under $\zeta = r/z$. Since $L_Q(\tilde{c}) = L_Q(c)$, $L_Q(c_0) \leqq L_Q(c)$ follows. If $c \cap P_1 \neq \phi$ and $c \cap P_2 \neq \phi$, we obtain two closed curves c_1, c_2 consisting of the subarcs of c and c_0 and such that $c_1 \cap P_2 = \phi$, $c_2 \cap P_1 = \phi$. By the previous arguments, $L_Q(c_i) \geqq L_Q(c_0)$, $i = 1, 2$. Since $L_Q(c_1) + L_Q(c_2) = L_Q(c) + L_Q(c_0)$, we have $L_Q(c_0) \leqq L_Q(c)$. Let D be a doubly connected domain. Then D can be mapped by a univalent analytic function $f(z)$ onto Q. It is clear that $f^{-1}(c_0)$ is the unique minimal closed curve of D with respect to the inner boundary component by the univalency of $f(z)$.

We consider a domain D in the z-plane which is bounded by $b_1 = [z \mid |z| = r]$, $b_N = [z \mid |z| = 1]$, and $(N-2)$ closed Jordan curves $b_2, \cdots, b_{N-1}$. The curves $b_2, \cdots, b_{n-1}$ lie in the domain bounded by b_1 and b_N.

Let $A_1 = [z \mid r < |z| < 1]$, $A_2 = [z \mid |z-a| < \rho, |z| > r]$[2], be exterior and interior domains of comparison for D, respectively, i.e., $A_1 \supset D \supset A_2$. Then

$$L_{A_1}(K(A_1)) \leqq L_D(K(D)) \leqq L_{A_2}(K(A_2)) , \tag{3.2}$$

where $K(A_1)$, $K(A_2)$ and $K(D)$ are minimal closed curves of A_1, A_2 and D with respect to b_1, respectively. It is an immediate consequence of the fact that if $B \subset A$, then $\boldsymbol{K}_B(z, \bar{z}) \geqq \boldsymbol{K}_A(z, \bar{z})$ for $z \in B$. The linear transformation

$$w = \frac{z-(a+\rho d)}{\rho - d(z-a)} , \qquad 0 < |a| \leqq \rho - r , \tag{3.3}$$

maps A onto $Q_R = [z \mid R < |z| < 1]$, where R is given by

---

1 A simple computation shows that the kernel function of Q,

$$\boldsymbol{K}_Q(z, z) = \frac{1}{\pi |z|^2}\left[\wp(2\log|z|;\ -2\log r,\ 2\pi i) + \frac{\zeta(\pi i;\ -2\log r,\ 2\pi i)}{\pi i}\right]$$

(see [**9**], [**18**]), where $\wp$ and ζ are the Weierstrass elliptic functions, assumes its minimum on c_0.

2 Here we choose a and ρ in such a way that $|z-a| < \rho$ contains b_1 but no other b_k, $k = 2, \cdots, N$, and A_2 to be the largest among such domains.

$$R = \left[\frac{r^2 - (a + \rho d)^2}{(\rho + ad)^2 - r^2 d^2}\right]^{1/2}, \tag{3.4}$$
$$d = \frac{r^2 - a^2 - \rho^2 + [(r^2 - a^2 - \rho^2)^2 - 4a^2\rho^2]^{1/2}}{2a\rho}.$$

Since $L_{A_2}(K(A_2)) = L_{Q_R}(K(Q_R))$, using (3.2), we obtain

THEOREM 3.3. *$E(r) \leqq (1/2)L_D(K(D)) \leqq E(R)$, where R is given by* (3.4) *and*

$$E(r) = [\pi\wp(\log r; -2\log r, 2\pi i) - i\zeta(\pi i; -2\log r, 2\pi i)]^{1/2},$$

$\wp$ and ζ are the Weierstrass elliptic functions.

The estimation of the bounds for the Bergman lengths of the minimal closed curves in Theorem 3.3 seems to be done only for a special domain. However, every multiply connected domain can always be mapped onto such a domain by a conformal mapping. Therefore, if we know the geometry of a given domain D, combining the various distortion theorems in the theory of conformal mappings and the result in Theorem 3.3, we can obtain various bounds for the Bergman lengths of the minimal curves for quite general domains.

4. The existence of a minimal surface which spans a given closed curve in (D, d). A surface S in the space C^2 is said to be *continuously differentiable* if it admits a continuously differentiable representation

$$G \mid Q_0 : z_k = G_k(u_1, u_2),\ k = 1, 2,\ (u_1, u_2) \in Q_0 = [0 \leqq u_1, u_2 \leqq 1].$$

A surface S is said to be *piecewise continuously differentiable* if it admits a piecewise continuously differentiable representation $G \mid Q_0$, i.e., there exists a partition $\Delta = \{\Delta_1, \Delta_2, \cdots, \Delta_m\}$ of Q_0 by rectilinear triangles Δ_k such that $G \mid \Delta_k$ is continuously differentiable, $k = 1, 2, \cdots, m$. The *ordinary B-area element* at a point (z_1, z_2) on a piecewise continuously differentiable surface S is defined by the equation [6, Chap. XI]

$$db_S(z) = \left|\frac{\partial(G_1, G_2)}{\partial(u_1, u_2)}\right| du_1 du_2. \tag{4.1}$$

The ordinary area element of S is given by the equation

$$da_S(z) = [g_{11}g_{22} - (Re\, g_{12})^2]^{1/2} du_1 du_2, \tag{4.2}$$
$$g_{\alpha\beta} = \sum_{i=1}^{2} \frac{\partial G_i}{\partial u_\alpha} \overline{\frac{\partial G_i}{\partial u_\beta}},\ \alpha, \beta = 1, 2.$$

Further (4.1) can also be written in the following form,

$$(4.1)' \qquad db_S(z) = [g_{11}g_{22} - |g_{12}|^2]^{1/2}\, du_1 du_2 .$$

Therefore, $da_S(z) \geqq db_S(z)$ at every point $z \in S$; the equality holds if and only if $Im\, g_{12} = 0$.

For a piecewise continuously differentiable surface S, the ordinary B-area is defined and given by the equation

$$(4.3) \qquad b(S) = \iint_{Q_0} \left| \frac{\partial(G_1, G_2)}{\partial(u_1, u_2)} \right| du_1 du_2 .$$

$b(S)$ is independent of the choice of piecewise continuously differentiable representations $G \mid Q_0$ of S. A surface S is said to be *analytic* if it admits an analytic representation $G \mid Q_0$, i.e., $\partial G_k/\partial \bar{w} = 0, k = 1, 2$, $w = u_1 + iu_2$.

For an analytic or an anti-analytic surface S, $b(S) = 0$. It is also clear that $b(S) = 0$ if and only if the tangent plane of S at every point is an analytic plane. A simple computation shows the following lemma:

LEMMA 4.1. *Thc following three conditions are equivalent*:

1) $b(S) = a(S)$,
2) $\dfrac{\partial(G_1, \bar{G}_1)}{\partial(u_1, u_2)} + \dfrac{\partial(G_2, \bar{G}_2)}{\partial(u_1, u_2)} = 0$ *at each point on* S,
3) $\oint_c \bar{G}_1 dG_1 + \bar{G}_2 dG_2 = 0$ *for every closed curve* c *on* S.

Let D be a bounded domain in the space $\boldsymbol{C}^2$ on which (D, d) is complete. The quantity

$$(4.4) \qquad dB_D(z) = [\boldsymbol{K}_D(z, \bar{z})]^{1/2}\, db_S(z),\ z = (z_1, z_2) ,$$

is invariant with respect to pseudo-conformal mappings and a monotone decreasing functional of D [6]. $dB_D(z)$ is called the *invariant B-area element* of S. For a piecewise continuously differentiable surface S in D, the invariant B-area of S is defined and given by the equation

$$(4.5) \qquad B_D(S) = \iint_{Q_0} [\boldsymbol{K}_D(G, \bar{G})]^{1/2} \left| \frac{\partial(G_1, G_2)}{\partial(u_1, u_2)} \right| du_1 du_2$$

and is independent of the choice of piecewise continuously differentiable representations $G \mid Q_0$ of S.

A surface S in D is said to satisfy the *condition (L) with respect to the metric d* if there exists a representation $G \mid Q_0$ of S for which there exists a constant $L(S) > 0$ depending only on S and satisfying the inequality

$$(4.6) \qquad \mathscr{L}_D(G; \sigma(w_1, w_2)) \leqq L(S) \mid w_1 - w_2 \mid$$

for every pair of points w_1; w_2 in Q_0; here $\sigma(w_1, w_2)$ is the line segment that joins w_1 and w_2 in Q_0, $w_k = u_1^{(k)} + iu_2^{(k)}$, $k = 1, 2$.

It is clear that $G \mid \partial(Q_0)$, where $\partial(Q_0)$ is the boundary of Q_0, is a representation of the boundary curve Γ of S and that Γ is rectifiable. It is also clear that every continuously differentiable surface S satisfies the condition (L) with respect to d.

We shall say that a surface S is *of class* $C'\mathfrak{R}(L, N, \Gamma)$ if S admits a continuously differentiable representation

$$G \mid Q_0 : z_k = G_k(w)\ , \quad k = 1, 2, w \in Q_0\ ,$$

which satisfies the following conditions:

(a) for a fixed positive constant L, $L(S) \leqq L$,

(b) for a fixed positive constant N,

$$\left| \frac{\partial G(w_1)}{\partial u_j} - \frac{\partial G(w_2)}{\partial u_j} \right| \leqq N \mid w_1 - w_2 \mid\ , \quad j = 1, 2, G = (G_1, G_2)\ ,$$

for every pair of points w_1, w_2 in Q_0,

(c) S spans a preassigned closed Jordan curve Γ in D in such a way that G is a one-to-one mapping on $\partial(Q_0)$.

A surface S_m is called *minimal surface* of the class $C'\mathfrak{R}(L, N, \Gamma)$ if $B_D(S_m) \leqq B_D(S)$ for all $S \in C'\mathfrak{R}(L, N, \Gamma)$.

THEOREM 4.1.[3] *For each L and N for which the class $C'\mathfrak{R}(L, N, \Gamma)$ is not empty, there exists a minimal surface S_m in the class.*

Proof. Let $\inf_S B_D(S) = \gamma$, where S runs over all surfaces in $C'\mathfrak{R}(L, N, \Gamma)$. Then $0 \leqq \gamma < \infty$. Hence, there exists a minimizing sequence $\{S_n\}$. Let $G^n \mid Q_0$ be a representation of S_n which satisfies conditions (a), (b) and (c). From (a) it follows that for any pair of positive integers p, q,

$$d(G^p(w), G^q(w)) \leqq 2\text{-}2^{3/2}\, L\ .$$

Therefore, $\{G^n(w)\}$ is equi-bounded. The equi-continuity of $\{G^n(w)\}$ follows from the inequality

$$d(G^n(w),\ G^n(w')) \leqq L \mid w - w' \mid \qquad \text{for any } w, w' \in Q_0 \text{ and all } n. \tag{4.7}$$

Hence, we can select a subsequence $\{G^m(w)\}$ of $\{G^n(w)\}$ which converges uniformly to a continuous function $G^0(w)$ defined in Q_0. Let $G^0 \mid Q_0$ define a surface S_0. Then it is clear that S_0 spans Γ in such a way that G^0 is a one-to-one mapping on $\partial(Q_0)$. The family $\{\partial G^m/\partial u_j\}$

[3] Replacing (a) by the condition (a′) on page 951, a result similar to Theorem 4.1 can be given (see Corollary 2).

of continuous functions $\partial G^m/\partial u_j$ is equi-bounded and equi-continuous by (b) for $j = 1, 2$. Therefore, we can select a subsequence $\{G^{m_i}(w)\}$ of $\{G^m(w)\}$ which converges uniformly to $G^0(w)$ and such that $\{\partial G^{m_i}/\partial u_j\}$ converges uniformly to a continuous function $\partial G^0/\partial u_j$ for $j = 1, 2$. This implies that S_0 is a continuously differentiable surface. In order to show $S_0 \in C'\mathfrak{R}(L, N, \Gamma)$, let c_{m_i} and c_0 be the image curves of a line segment $\sigma(w_1, w_2)$ which connects two points w_1 and w_2 in Q_0 under $G^{m_i}(w)$ and $G^0(w)$, respectively. Then c_{m_i} converges to c_0 and, hence, $\varliminf_{i\to\infty} L_D(c_{m_i}) \geqq L_D(c_0)$ by the lower semi-continuity of L_D. Since $L_D(c_{m_i}) \leqq L\,|\,w_1 - w_2\,|$ for all m_i, $L_D(c_0) \leqq L\,|\,w_1 - w_2\,|$. It is clear that $G^0(w)$ satisfies (b). Since the functional B_D is lower semi-continuous in $C'\mathfrak{R}(L, N, \Gamma)$ and $S_0 \in C'\mathfrak{R}(L, N, \Gamma)$, we have $B_D(S_0) = \gamma$. Thus S_0 is a minimal surface in the class $C'\mathfrak{R}(L, N, \Gamma)$.

REMARK. In the case that Γ lies on an analytic plane π and the portion $\tilde{\pi}$ of $\pi \cap \mathrm{D}$ enclosed by Γ is simply connected, $\tilde{\pi}$ is a minimal surface of $C'\mathfrak{R}(L, N, \Gamma)$ with some L and N, and $B_D(S_0) = 0$. In general, if there exists an analytic surface S in D which spans Γ, then S is a minimal surface with some L and N, and $B_0(S) = 0$.

Let $C'\mathfrak{R}(N, \Gamma)$ be the class of continuously differentiable surfaces in the space $\boldsymbol{C}^2$ which span a preassigned Jordan closed curve Γ in $\boldsymbol{C}^2$ and satisfy the condition (b). Then (b) implies condition (a) with respect to the Euclidean metric ρ for every surface in $C'\mathfrak{R}(N, \Gamma)$. Since $\boldsymbol{C}^2$ is complete with respect to ρ, the following corollary follows by the same procedure as in Theorem 4.1.

COROLLARY 1. *In the class $C'\mathfrak{R}(N, \Gamma)$, there exists a minimal surface S'_m in the sense that*

$$b(S'_m) \leqq b(S) \text{ for all } S \in C'\mathfrak{R}(N, \Gamma)\ .$$

Let $C'\mathfrak{R}_\alpha(N, \Gamma)$ be the class of continuously differentiable surfaces S in D which satisfy conditions (b), (c) and

(a′) for a preassigned real number α, $0 \leqq \alpha \leqq 1$,

$$\frac{db_S(z)}{da_S(z)} \geqq \alpha \text{ at every point } z \in S\ . \tag{4.8}$$

We notice that the class $C'\mathfrak{R}_\alpha(N, \Gamma)$ is motone decreasing with respect to α.

COROLLARY 2. *For a fixed $\alpha > 0$ and N for which $C'\mathfrak{R}_\alpha(N, \Gamma)$ is not empty there exists a minimal surface in the class.*

Proof. The B-areas $B_D(S_n)$ of S_n which belong to a minimizing sequence $\{S_n\}$ have a fixed uppper bound. Therefore, condition (a′) ensures the existence of an $M > 0$ such that every S_n lies completely in D_M, $D_M = [z \mid \boldsymbol{K}_D(z, \bar{z}) \leqq M]$. This implies condition (a) with some L, which depends on α and N. Hence, the corollary follows from the theorem.

5. The existence of minimal closed surfaces in (D, d). Let D_k be a domain in the space of one complex variable z_k bounded by n_k closed curves $b_1^{(k)}, b_2^{(k)}, \cdots, b_{n_k}^{(k)}$. Here $b_{n_k}^{(k)}$ is the outer boundary component of D_k and $b_1^{(k)}$ is an inner boundary component, which is a circle, i.e., $b_1^{(k)} = [z_k \mid |z_k| = r_k]$.

Let D be a domain in the space $\boldsymbol{C}^2$ which is topologically equivalent to the product domain $\tilde{D} = D_1 \times D_2$, and T_1 the topological image of $\tilde{T}_1 = b_1^{(1)} \times b_1^{(2)}$. A surface S in D which is homotopic to T_1 in $\overline{D(T_1)}$, where $D(T_1)$ is the topological image of $\tilde{D}(\tilde{T}_1) = D_1(b_1^{(k)}) \times D_2(b_2^{(k)})$ (see § 3 for notation), is a closed surface of the torus type and, hence, admits a doubly periodic representation

$$G \mid R^2 : z_k = G_k(u_1, u_2),\ k = 1, 2,\ (u_1, u_2) \in R^2\ ,$$
$$R^2 = (-\infty < u_1, u_2 < +\infty)\ ,$$

of periods 1. For our purposes, therefore, it is enough to consider G on the unit square Q_0 as a representation of S.

We shall say that a closed surface S is *of class $C'\mathfrak{K}_\alpha(N, T_1)$* if S is homotopic to T_1 in $\overline{D(T_1)}$ and admits a continuously differentiable representation $G \mid Q_0$ satisfying condition (a′) and (b) in § 4. By the same procedure as in Corollary 2 of Theorem 4.1, we can prove the following theorem for any fixed $\alpha > 0$.

THEOREM 5.1. *For each N for which the class $C'\mathfrak{K}_\alpha(N, T_1)$ is not empty, there exists a minimal closed surface $S_m(D)$ in the class.*

Let $D'\mathfrak{K}(\tilde{D}, \tilde{T}_1)$ be the class of all closed surfaces S of the form $S = c_1 \times c_2$ in D, where c_k is a piecewise continuously differentiable closed curve in D_k which is homotopic to $b_1^{(k)}$ in $\overline{D_k(b_1^{(k)})}$. For each $S \in D'\mathfrak{K}(\tilde{D}, \tilde{T}_1)$, we have $B_{\tilde{D}}(S) = L_{D_1}(c_1) \cdot L_{D_2}(c_2)$. It follows from the fact that $\boldsymbol{K}_{\tilde{D}}(z, \bar{z}) = \boldsymbol{K}_{D_1}(z_1, \bar{z}_1) \cdot \boldsymbol{K}_{D_2}(z_2, \bar{z}_2)$ [7]. Therefore, the following is an immediate consequence of Theorem 3.1.

THEOREM 5.2. *There exists a minimal surface $S_m(\tilde{D})$ of the class $D'\mathfrak{K}(\tilde{D}, \tilde{T}_1)$. It is given by $K(D_1) \times K(D_2)$, where $K(D_k)$ is a minimal closed curve of D_k with respect to $b_1^{(k)}$.*

Let $A = A_1 \times A_2$, where A_k is a doubly connected plane domain in the z_k-plane. Let $D'\mathfrak{R}_1(A, T)$ be the class of piecewise continuously differentiable closed surfaces S in A which are homotopic to $T = b_1^{(1)} \times b_2^{(2)}$ in $\bar{A}$, where $b_1^{(k)}$ is the inner boundary component of A_k, and satisfy the condition $da_S(z) = db_S(z)$.[4] Then the following theorem holds:

THEOREM 5.3. *There exists a unique minimal closed surface in the class $D'\mathfrak{R}_1(A, T)$. It is given by $K(A_1) \times K(A_2)$, where $K(A_k)$ is a minimal closed curve of A_k with respect to $b_1^{(k)}$.*

Proof. Let $A = Q = Q_1 \times Q_2$, $Q_k = [z_k \mid r_k < |z_k| < 1]$. We shall show $S_m(Q) = K(Q_1) \times K(Q_2)$ is a unique minimal closed surface of $D'\mathfrak{R}_1(Q, T)$. Let $P_{1k} = [z_k \mid r_k < |z_k| < r_k^{1/2}]$, $P_{2k} = [z_k \mid r_k^{1/2} \leqq |z_k| < 1]$. If $S \in P_{21} \times P_{22}$, it is immediate that $B_Q(S) \geqq B_Q(S_m)$. For any $S \in D'\mathfrak{R}_1(Q, T)$, S can be replaced by a surface $\tilde{S} \in D'\mathfrak{R}_1(Q, T)$ with $B_Q(S) = B_Q(\tilde{S})$ and lying in $P_{21} \times P_{22}$ by the pseudo-conformal mapping $z_k^* = r_k/z_k$, $k = 1, 2$. Thus, $B_Q(S) \geqq B_Q(S_m)$ for every $S \in D'\mathfrak{R}_1(Q, T)$. There exists a univalent analytic function $f_k(z_k)$ which maps A_k onto Q_k. Therefore, the pseudo-conformal mapping $w_k = f_k^{-1}(z_k)$ maps A onto Q and, hence, $S_m(Q)$ onto $S_m(A)$, $S_m(A) = K(A_1) \times K(A_2)$. The uniqueness of $S_m(A)$ is clear.

6. Bounds for the B-areas of minimal closed surfaces in the space (D, d). Using the method of exterior and interior domains of comparison, various bounds for the B-areas of minimal surfaces can be obtained. As we have considered in § 3, let D_k be bounded by $b_1^{(k)} = [z_k \mid |z_k| = r_k]$, $b_{n_k}^{(k)} = [z_k \mid |z_k| = 1]$, and $(n_k - 2)$ closed Jordan curves $b_2^{(k)}, \cdots, b_{n_k-1}^{(k)}$, which lie in the domain bounded by $b_1^{(k)}$ and $b_{n_k}^{(k)}$. Let $A_{1k} = [z_k \mid r_k < |z_k| < 1]$, $A_{2k} = [z_k \mid |z_k - a_k| < \rho_k, |z_k| > r_k]$, $0 < |a_k| \leqq \rho_k - r_k$, be exterior and interior domains of comparison for D_k, respectively. Then $A_j = A_{j1} \times A_{j2}$ can be used as exterior and interior domains of comparison of $D = D_1 \times D_2$, i.e., $A_1 \supset D \supset A_2$. Let $S_m(\tilde{D})$ and $S_m(A_j)$ be minimal surfaces of the classes $D'\mathfrak{R}(\tilde{D}, \tilde{T}_1)$ and $D'\mathfrak{R}(A_j, \tilde{T}_1)$, respectively. Then $B_{A_1}(S_m(A_1)) \leqq B_{\tilde{D}}(S_m(D)) \leqq B_{A_2}(S_m(A_2))$. Using this inequality, we have the following distortion theorem for minimal surfaces of the class $D'\mathfrak{R}(\tilde{D}, \tilde{T}_1)$.

THEOREM 6.1. $\prod_{k=1}^{2} E(r_k) \leqq (1/4) B_{\tilde{D}}(S_m(\tilde{D}, \tilde{T}_1)) \leqq \prod_{k=1}^{2} E(R_k)$, *where R_k is given in* (3.4) *with the corresponding subscript k and $E(r)$ is given in Theorem* 3.3.

By a construction of an interior domain of comparison for D in

[4] This is the case $\alpha = 1$ in (a′) (see (4.8) and Lemma 4.1).

Theorem 5.1, we can also obtain a distortion theorem for minimal surfaces $S_m(D)$ in Theorem 5.1 which gives us an upper bound. Suppose an interior domain of comparison for D is given by A_{2k}, then we have

THEOREM 6.2. *$B_D(S_m(D)) \leqq 4 \prod_{k=1}^{2} E(R_k)$, where R_k and $E(r)$ are given as in Theorem* 6.1.

REMARK. For the product domain $Q = Q_1 \times Q_2$ of two annuli Q_1 and Q_2, $K(Q_1) \times K(Q_2)$ is not necessarily a minimal surface for the class $D'\mathfrak{R}_\alpha(Q, T)$ for a fixed $\alpha, 0 < \alpha < 1$.

BIBLIOGRAPHY

1. S. Bergman, *Über die Kernfunktion eines Bereiches und ihr Verhalten am Rande*, J. Reine Angew. Math., **169** (1933), 1-42, and **172** (1934), 89-128.
2. ———, *Sur quelques propriétés des transformations par un couple de fonctions de deux variables complexes*, Rend. Della R. Accad. Naz. Del Lincei, **19** (ser. 6) (1934), 474-478.
3. ———, *Zur Theorie von Pseudokonformen Abbildungen*, Recueil Math., nouv., **1** (43), (1936), 76-96.
4. ———, *Sur les functions orthogonales de plusieurs variables complexes*, Mém. des Sciences Math., **106** (1947).
5. ———, *Sur la fonction-noyau d'un domaine et ses applications dans la théorie des transformations pseudo-conformes*, Mém. des Sciences Math., **108** (1948).
6. ———, *The kernel function and conformal mapping*, Math. Surveys 5, Amer. Math. Soc., New York, 1950.
7. H. J. Bremermann, *Holomorphic continuation of the kernel function and the Bergman metric in several complex variables*, in Lectures on functions of a complex variable ed. by W. Kaplan, Univ. of Michigan Press, 1955.
8. H. Busemann, and M. Mayer, *On the foundations of calculus of variations*, Trans. Amer. Math. Soc., **49** (1941), 173-198.
9. B. Epstein, *The kernel-function and conformal invariants*, J. Math. and Mech., **7** (1958), 925-936.
10. D. Hilbert, *Über das Dirichletsche Princip*, Jahr. Dtsch. Math.-Verein., **8** (1900), 184-188.
11. S. Kobayashi, *Geometry of Bounded Domains*, Trans. Amer. Math. Soc., **92** (1959), 267-290.
12. C. B. Morrey Jr., *Quasi-convexity and the lower semi-continuity of multiple integrals*, Pacific J. Math., **2**, (1952), 25-53.
13. ———, *The parametric variational problem for double integrals*, Communications on Pure and Applied Math., **14** (1961), 569-575.
14. K. Reinhardt, *Über Abbildungen durch analytische Funktionen zweier Veranderlicher*, Math. Ann., **83** (1921), 211-255.
15. W. Rinow, *Die Innere Geometrie der Metrischen Räume*, Die Grundlehren der Math. Wissenschaften **105** (1961).
16. J. M. Stark, *On distortion in pseudo conformal mappings*, Pacific J. Math., **6** (1956), 565-582.
17. ———, *Minimum problems in the theory of pseudo-conformal transformations*

and their application to estimation of the curvature of the invariant metric, Pacific J. Math., **10** (1960), 1021-1038.
18. K. Zarankiewicz, *Über ein numerisches Verfahren zur konformen Abbildung zweifach zusammenhangender Gebiede,* Zeits. f. ang. Math. u. Mech. **14** (1934), 97-104.

STANFORD UNIVERSITY

A REPRESENTATION THEORY FOR A CLASS OF PARTIALLY ORDERED RINGS

ALLAN HAYES

The lattice ordered rings known as f-rings, introduced by Birkhoff and Pierce in [1], have been studied very intensively in the last few years. In particular Pierce has shown in [4] that the f-rings without nonzero nilpotents are precisely the (isomorphic images of) lattice ordered subdirect unions of totally ordered rings with integrity, and Johnson in [2] has gone on to prove that any Archimedean f-ring with no nonzero nilpotents can be represented as a lattice ordered ring of continuous extended realvalued functions on a locally compact Hausdorff space.

Since many commonly occurring examples of partially ordered rings are not lattice ordered it is natural to ask whether these two results can be generalised so as to be independent of the lattice structure. Such a generalisation is given here when multiplication is assumed commutative.

Theorem 1 characterises the subdirect unions of totally ordered commutative rings with integrity; Theorem 2 sharpens this result and Theorem 3 completes the programme by extending Johnson's representation.

The plan of the paper is as follows:

Section 1 is an introduction to the subject matter and methods of the paper; the succeeding three sections contain proofs of Theorems 1, 2 and 3 respectively and § 5 shows that for f-rings the representations given preserve the lattice structure.

1. Introduction. Throughout this paper "ring" will be an abbreviation for "commutative associative ring".

A *partially ordered* (or *po*-) *ring* is a ring whose elements are partially ordered in such a way that if $a \geqq b$ then $a + c \geqq b + c$ for all c and $ac \geqq bc$ for all $c \geqq 0$. Among the *po-rings* those with integrity (i.e. without divisors of zero) and a total ordering (the *toi-rings*) are particularly simple and it is our first aim to find out when a *po*-ring can suitably be built up from *toi*-rings. To make this more precise:

If $\{R_i\}_{i \in I}$ is a nonempty family of *toi*-rings their *direct union*, $\sum R_i$, is formed by taking the class of all functions $a: I \to \bigcup R_i$ with $a(i) \in R_i$ for all i, and defining addition by $(a + b)(i) = a(i) + b(i)$ for all i; multiplication by $(ab)(i) = a(i)b(i)$ for all i, and order by $a \geqq b$

Received September 16, 1963. This work was supported by N.S.F. Grant 3639-50-8476.

when $a(i) \geqq b(i)$ for all i. $\sum R_i$ is then a *po*-ring (in fact it is an *f*-ring). A *subdirect union* of the family $\{R_i\}_{i \in I}$ is a subring, R, of $\sum R_i$ satisfying $R(i) = R_i$ for all i, together with the partial ordering induced on it by the partial ordering of $\sum R_i$. If in addition, whenever R contains a it contains a^+, defined by $a^+(i) = a(i) \vee 0$ for all i, it is called a *lattice ordered* subdirect union of $\{R_i\}_{i \in I}$ (and is an *f*-ring).

A mapping, h, from one *po*-ring to another is called a *homomorphism* if it is a ring homomorphism such that $h(a) \geqq h(b)$ when $a \geqq b$: it is called an *isomorphism* if it is a ring isomorphism with $h(a) \geqq h(b)$ if and only if $a \geqq b$.

Suppose R is a *po*-ring and $\mathfrak{H}$ is a nonempty class of homomorphisms, h, of R onto *toi*-rings R_h respectively. Suppose further that if $a \in R$ and $a \ngeqq 0$ then there is an $h \in \mathfrak{H}$ with $h(a) < 0$. For any $a \in R$ let $\tilde{a}$ be the function on $\mathfrak{H}$ defined by $\tilde{a}(h) = h(a)$ for all $h \in \mathfrak{H}$. Then $\tilde{R} = \{\tilde{a}\colon a \in R\}$, with the natural induced structure, is a subdirect union of $\sum R_h$, and the map $a \to \tilde{a}$ is an isomorphism of R onto $\tilde{R}$.

To generate the homomorphisms needed we look at the *semirings* in R (i.e. the nonempty subsets, S, of R with $SS \cup (S + S) \subset S$). Under conditions stated in the next section, if $a_0 \ngeqq 0$ then maximalisation by Zorn's Lemma yields a semiring P, with $a_0 \notin P$ and $P'P' \subset -P'$,[1] which contains all $a \geqq 0$ and all squares in R. From this a homomorphism onto a *toi*-ring arises as follows:

$I = P \cap -P$ is a prime ring ideal in R. For,

(i) if $a, b \in I$ then clearly $a - b \in I$;

(ii) if $a \in I$ and $c \in R$ then $c \in P$ or $c \in -P$ (otherwise $-((-c)c) = c^2 \in P'$) and in either case $ac \in I$;

(iii) if $a \in I'$ and $b \in I'$ then $a \in P'$ or $-a \in P'$ and $b \in P'$ or $-b \in P'$; whence $ab \in P'$ or $-ab \in P'$ and certainly $ab \in I'$. Let h be the canonical homomorphism of R onto R/I, which is a ring with integrity. A simple calculation shows that $h(P)$ is a semiring, $h^{-1}(h(P)) = P$, $h(P) \cup -h(P) = h(R)$ and $h(P) \cap -h(P) = \{0\}$. So if we define $h(a) \geqq h(b)$ to mean $h(a) - h(b) \in h(P)$, (i.e. $a - b \in P$) then this is a total ordering making R/I into a *toi*-ring which is called the *quotient ring* of R by P and is denoted by R/P. Since P contains all $a \geqq 0$, $a_0 \in P'$ and $h^{-1}(hP) = P$, h is a homomorphism of R onto R/P and $h(a_0) < 0$.

It is convenient to write $a(P)$ for $h(a)$ and to use abbreviations similar to writing $a \geqq b, (P)$ for $a(P) \geqq b(P)$.

The representation of a *po*-ring as a ring of real valued functions on some set would be very useful. Unfortunately it seems difficult to find a simple general condition permitting this, which does not make all the functions used bounded. Nevertheless, a *po*-ring of the type here considered which is also Archimedean (that is $na \leqq b$, $n =$

[1] $P' = R \backslash P$.

$1, 2, \cdots$ implies $a \leqq 0$) can be represented using functions with values in the extended real numbers. The possibility of this is suggested by the observation that in a *toi*-ring R if $ab \geqq 0$ and $a > 0$ then $cb \geqq 0$ for all $c \geqq 0$ so that if

$$\begin{aligned} \bar{a} = \inf\{m/n\colon & \ m \text{ and } n \text{ are integers, } n > 0, \text{ and} \\ & mb \geqq nab \text{ for all } b > 0\} \\ (= \sup\{m/n\colon & \ m \text{ and } n \text{ are integers, } n > 0, \text{ and} \\ & mb \leqq nab \text{ for all } b > 0\}) \end{aligned}$$

it follows by routine calculations that $\bar{a} \geqq 0$ when $a \geqq 0$, $\overline{ab} = \bar{a}\bar{b}$ unless $\bar{a} = 0$ and $\bar{b} = \pm\infty$ or vice versa, and $\overline{a+b} = \bar{a} + \bar{b}$ unless $\bar{a}$ and $\bar{b}$ are infinite and of opposite sign. Here the infimum is taken in the extended reals and the infimum of the empty set is $+\infty$. The main problem is to guarantee that the substitution of $\bar{a}$ for a, which is usually far from being $(1-1)$, still leaves enough information for reconstruction of the original *po*-ring; it is here that the assumption that the ring is Archimedean is required.

The following notation will be standard for the rest of the paper:

If R is a *po*-ring then $R^+ = \{x\colon x \geqq 0\}$ is the class of *quasi positive* elements of R and $R^{++} = \{x\colon x > 0\}$ is the class of *positive* elements of R.

$\boldsymbol{Z}$ is the *po*-ring of integers.

$\boldsymbol{R}$ is the *po*-ring of real numbers and $\bar{\boldsymbol{R}}$ the "quasi *po*-ring" of the extended real numbers with the usual topology of the two point compactification.

If a set X is fixed in some context and $Y \subset X$ then Y' will denote $X \backslash Y$. The empty set is denoted by ϕ. The set with x as its only element will sometimes be denoted simply by x.

If A and B are subsets of a partially ordered set then $A \leqq B$ means that every element of A is less than or equal to every element in B.

2. f^*-rings. Lemma 1 below, on the semirings in a ring, is the key to the rest of the paper. It is used in this section to produce a characterisation of the isomorphic images of subdirect sums of *toi*-rings (Theorem 1).

A semiring S in a ring R is said to be *normal* with respect to a nonempty subset H of R if no expression of the form

$$\sum_{i=1}^{N} (-1)^{n_i+1} s_i a_{i,1} a_{i,2} \cdots a_{i,n_i} - a_1 a_2 \cdots a_{2q} - s \tag{1}$$

is zero, where each a is in H, each s is in S, each n is in $\boldsymbol{Z}^{++}$, q is in $\boldsymbol{Z}^{++}$ and N is in $\boldsymbol{Z}^{+}$.

If S contains all squares in R and $H = \{a\}$ then S is normal with respect to H if and only if $sa - a^{2n} \in S'$ for all $s \in S$ and all $n \in \boldsymbol{Z}^{++}$.

Normality of S with respect to H implies $H \subset S'$. For if $a \in H \cap S$ then $(-1)^{1+1}aa + (-1)^{1+1}aa - aa - aa = 0$.

A semiring P in a ring R is called *prime* if $P'P' \subset -P'$.

The usefulness of normality is due to the following result:

LEMMA. *If S is a semiring containing all squares in a ring R, and H is a nonempty subset of R then there is a prime semiring P in R with $P \supset S$ and $P' \supset H$ if and only if S is normal with respect to H.*

Proof. (i) If such a P exists then for any $a_1, a_2, \cdots, a_n \in P'$ and any $s \in S$, $(-1)^{n+1}sa_1a_2 \cdots a_n \leqq 0, (P)$ (see § 1 for this notation); and if n is even $-a_1a_2 \cdots a_n < 0, (P)$. So any expression of the form (1) is $<0, (P)$ and cannot be equal to zero.

(ii) Conversely, if S is normal with respect to H then Zorn's lemma shows that there is a maximal semiring, P, among the semirings containing S which are normal with respect to H. It will be proved that P is as required.

Since P contains all squares in R, if $x \in R$ then the semiring, P_x, generated by $P \cup \{x\}$ is $\boldsymbol{Z}^{+}x + xP + P$. So if $x \in P'$ and $y \in P'$, since neither P_x nor P_y is normal with respect to H, there are identities of the form

$$\sum_{i=1}^{N} (-1)^{n_i+1}(s_i' + s_i)a_{i,1} \cdots a_{i,n_i} - a_1a_2 \cdots a_{2q} - (s' + s) = 0$$

and

$$\sum_{j=1}^{M} (-1)^{m_j+1}(t_j' + t_j)b_{j,1} \cdots b_{j,m_j} - b_1b_2 \cdots b_{2r} - (t' + t) = 0\ ,$$

where every a and b is in H, every n and m is in $\boldsymbol{Z}^{++}$, q and r are in $\boldsymbol{Z}^{++}$, M and N are in $\boldsymbol{Z}^{+}$, every s and t is in P, every s' is in $\boldsymbol{Z}^{+}x + xP$ and every t' is in $\boldsymbol{Z}^{+}y + yP$.

Collection of the terms involving x, y respectively to one side of the equations (taking the rest to the other side) followed by multiplication of the new equalities yields, after rearrangement, the following,

$$\sum_{i=1}^{N}\sum_{j=1}^{M}(-1)^{n_i+m_j+2}s_i't_j'a_{i,1}\cdots a_{i,n_i}b_{j,1}\cdots b_{j,m_j}+s't'$$
$$+\sum_{i=1}^{N}(-1)^{n_i+2}s_i't'a_{i,1}\cdots a_{i,n_i}+\sum_{j=1}^{M}(-1)^{m_j+2}t_j's'b_{j,1}\cdots b_{j,m_j}$$
$$+\sum_{i=1}^{N}\sum_{j=1}^{M}(-1)^{n_i+m_j+1}s_it_ja_{i,1}\cdots a_{i,n_i}b_{j,1}\cdots b_{j,m_j}$$
$$+\sum_{i=1}^{N}(-1)^{n_i+1}s_ita_{i,1}a_{i,2}\cdots a_{i,n_i}$$
$$+\sum_{i=1}^{N}(-1)^{n_i+2r+1}s_ia_{i,1}\cdots a_{i,n}b_1\cdots b_{2r}$$
$$+\sum_{j=1}^{M}(-1)^{m_j+1}t_jsb_{j,1}\cdots b_{j,m_j}$$
$$+\sum_{j=1}^{M}(-1)^{m_j+2q+1}t_jb_{j,1}\cdots b_{j,m_j}a_1\cdots a_{2q}+(-1)^{2r+1}sb_1\cdots b_{2r}$$
$$+(-1)^{2q+1}ta_1\cdots a_{2q}-st-a_1a_2\cdots a_{2q}b_1b_2\cdots b_{2r}=0\,.$$

If $xy\in -P$ this would contradict the hypothesis that P is normal with respect to H.

It is clear that $P\supset S$ and $P'\supset H$, so the proof is complete.

COROLLARY. *If H has only one element, a, then there is a P as required if and only if $sa-a^{2n}\in S'$ for all $s\in S$ and all $n\in \boldsymbol{Z}^{++}$.*

The full force of Lemma 1 is not required until § 4; up to that point the corollary will be sufficient.

From now on A will always denote a *po*-ring, $\mathscr{S}$ the class of all semirings in A which contain A^+ and $\mathscr{P}$ the class of prime semirings in A which contain A^+. If $\mathscr{D}$ is a subset of $\mathscr{P}$ such that for any $a\notin A^+$ there is a $D\in\mathscr{D}$ with $a(D)<0$ then $\mathscr{D}$ will be said to be *distinguishing*.

A is called an *f^*-ring* if A^+ contains all squares in A and is normal with respect to every single point set $\{a\}$ with $a\notin A^+$.

We have:

THEOREM 1. *A is isomorphic to a subdirect union of toi-rings if and only if it is an f^*-ring.*

Proof. (i) If A is an f^*-ring then the Corollary to Lemma 1 shows that $\mathscr{P}$ is distinguishing, so that from the discussion in the previous section, A is isomorphic to a subdirect union of *toi*-rings $\{A/P\}_{P\in\mathscr{P}}$.

(ii) If A can be identified with a subdirect union R of *toi*-rings $\{R_i\}_{i\in I}$ then $a\in A\backslash A^+$ implies $a(i)<0$ for some $i\in I$, say $a(i_0)<0$. Consequently, if $s\in A^+$ and $n\in \boldsymbol{Z}^{++}$, $(sa-a^{2n})(i_0)<0$ and $sa-a^{2n}\notin A^+$.

Thus A is normal with respect to $\{a\}$. Also, for any $a \in A$, $(a^2)(i) = a(i)^2 \geqq 0$ for all $i \in I$, so $a^2 \in A^+$. Thus A is an f^*-ring.

3. Ring Archimedean f^*-rings. In this section a class of f^*-rings is introduced which includes the Archimedean f^*-rings and for which a considerable sharpening of Theorem 1 is possible (see Theorem 2 below).

A *po*-ring R is called *ring* (or *r*-) *Archimedean* if $\boldsymbol{Z}^+a + R^+a \leqq b$ implies $a \leqq 0$. An Archimedean *po*-ring is necessarily r-Archimedean, but the converse is not true, since every totally ordered field is r-Archimedean.

The following two measures of size will be used.

In any *toi*-ring R an element, a, is called a *ring* (r—) *order unit* if $\boldsymbol{Z}^+a + R^+a - R^+ = R$, and is called *ring* ($r$-) *infinitesimal* if $\boldsymbol{Z}^+a^2 + R^+a^2 \leqq |a|$. Notice that if for some $q > 0$, $(\boldsymbol{Z}^+|a| + R^+|a|)q \leqq q$ then a is r-infinitesimal and $(\boldsymbol{Z}^+|a| + R^+|a|)p \leqq p$ for all $p \geqq 0$. A *toi*-ring is r-Archimedean if and only if every positive element is an r-order unit.

The main result to be proved is:

THEOREM 2. *A necessary and sufficient condition that A be an r-Archimedean f^*-ring is that it be isomorphic to a subdirect union of r-Archimedean toi-rings with no nonzero r-infinitesimal elements.*

It will be convenient to divide up the proof into a number of lemmas.

LEMMA 2. *Let A be an r-Archimedean f^*-ring and $\mathscr{D}$ a distinguishing subclass of $\mathscr{P}$. If a (D) is r-infinitesimal in A/D for all $D \in \mathscr{D}$ such that $a \notin D$ then $a \geqq 0$.*

Proof. For each $D \in \mathscr{D}$ either (i) $a \geqq 0$ or (ii) $a < 0, (D)$ and $[\boldsymbol{Z}^+(-a) + A^+(-a)](-a) \leqq (-a), (D)$. In either case $[\boldsymbol{Z}^+(-a) + A^+(-a)]a^2 \leqq a^2(D)$. Therefore, since $\mathscr{D}$ is distinguishing, $[\boldsymbol{Z}^+(-a) + A^+(-a)]a^2 \leqq a^2$; whence, A being r-Archimedean, $(-a)^3 \leqq 0$, and in an f^*-ring this implies $-a \leqq 0$, i.e. $a \geqq 0$.

LEMMA 3. *In any toi-ring R if a is not r-infinitesimal then $|a|$ is an r-order unit.*

Proof. If $\boldsymbol{Z}^+|a| + R^+|a| \leqq b$ while $(n_0|a| + p_0|a|)|a| > |a|$ with $n_0 \in \boldsymbol{Z}^+$ and $p_0 \in R^+$, then $b > 0$ and $(n_0|a| + p_0|a|)b > b \geqq (n_0b + p_0b)|a| = (n_0|a| + p_0|a|)b$, which is impossible.

Let $\mathscr{M}$ be the class of maximal elements in $\mathscr{P}$ (under set inclusion).

LEMMA 4. *If $P \in \mathscr{P}$, $a \in P'$ and $|a(P)|$ is an r-order unit in A/P then no $Q \in \mathscr{P}$ can contain $P \cup \{a\}$, therefore there is an $M \in \mathscr{M}$ with $a \notin M \supset P$.*

Proof. Suppose such a Q does exist and take $q \in Q'$. Since $-a(P)$ is an r-order unit in A/P there are $n \in \boldsymbol{Z}^+$ and $p \in P$ such that $n[(-a) + p(-a)] \geqq q(P)$. So $n(-a) + p(-a) - q \in P$ and $q \in P + na + pa \subset Q$, contrary to the hypothesis that $q \in Q'$.

The three previous lemmas show that $\mathscr{M}$ is distinguishing for r-Archimedean f^*-rings. However, a stronger result is needed to prove the Theorem.

LEMMA 5. *In any toi-ring R the class, I, of r-infinitesimal elements is a prime ring ideal such that if $|c| \leqq |a|$ and $a \in I$ then $c \in I$.*

Proof. If $a \in I$ and $|c| < |a|$, then for any $n \in Z^+$ and $p, q \in R^+$, $(n|c| + p|c|)q \leqq (n|a| + p|a|)q \leqq q$, so $c \in I$.

If $a, b \in I$, $n \in \boldsymbol{Z}^+$ and $p, q \in R^+$, $(2n|a-b| + 2p|a-b|)q \leqq (2n|a|+2p|a|)q+(2n|b|+2p|b|)q \leqq 2q$, whence $(n|a-b|+p|a-b|)q \leqq q$ and $a - b \in I$.

If $a \in I$ and $e \in R$ then $ae \in I$, for if not then, by Lemma 3, there are $n \in \boldsymbol{Z}^+$ and $p \in R^+$ such that $n|ae| + p|ae| > 2|e|$. But, since $a \in I$, $|e| \geqq n|ae| + p|ae|$, and these two inequalities together yield the contradiction, $0 > |e|$.

I has now been proved to be an ideal: it remains to prove that it is prime.

If $a, b \in I'$ there are $m, n \in \boldsymbol{Z}^+$ and $p, q \in R^+$ such that for any $s > 0$, $(m|a| + p|a|)s > s$ and $(n|b| + q|b|)s > s$, whence, by multiplication $(mn|ab| + (mp + nq + pq)|ab|)s^2 > s^2 > 0$, and so $ab \in I'$.

Let $\mathscr{M}^* = \{M \in \mathscr{M}: A/M$ contains no nonzero r-infinitesimal elements$\}$.

Then we have:

LEMMA 6. *If $M \in \mathscr{M} \backslash \mathscr{M}^*$ then every element of A/M is r-infinitesimal.*

Proof. Let $I_M = \{x \in A: x(M)$ is r-infinitesimal$\}$ and let $P = I_M + M$. Lemma 5 shows immediately that P is a semiring containing A. Furthermore if $a, b \in P'$ then $-a(M)$ and $-b(M)$ are positive and non-r-infinitesimal in A/M. So $a(M)b(M)$ is positive and non-r-infinitesimal in A/M, and $-ab \in P'$.

The maximality of M and the supposition that $M \notin \mathscr{M}^*$ imply therefore that $P = A$. So if $a \in A$ there is a $b \in I_M$ with $|b(M)| \geqq |a(M)|$, whence $a(M)$ is r-infinitesimal.

The following simple result proves to be important.

LEMMA 7. *If a is a non-r-infinitesimal positive element of a toi-ring R then there is a $b \in R^+$ such that $b^2 > a$*

Proof. If $a^2 \geqq a$ there is nothing to prove. If $a^2 < a$ then, since a is not r-infinitesimal, there are $n \in \boldsymbol{Z}^+$ and $p \in R^+$ with $(na + pa)a > a$; whence $(na + pa)^2a^2 > a^2 > a^3$, $(na + pa)^2a^2 > a^3$ and $(na + pa)^2 > a$. So $na + pa$ may be taken for b.

Proof of Theorem 2.

(i) *Necessity.* $\mathscr{M}^*$ is a distinguishing subset of $\mathscr{P}$; for if $a \ngeqq 0$ Lemma 2 shows that there is a $P \in \mathscr{P}$ with $a \in P'$ and $a(P)$ not r-infinitesimal and by Lemma 4 there is an $M \in \mathscr{M}$ containing A^+ with $a \notin M$, so $\mathscr{M}$ is distinguishing. Lemma 6 and a second application of Lemma 4 show that $\mathscr{M}^*$ is distinguishing.

Reference to the introduction completes the proof.

(ii) *Sufficiency.* Suppose A is identified with a subdirect union of a family $\{R_i\}_{i\in I}$ of *toi*-rings without nonzero r-infinitesimal elements. If $a \in A$ satisfies $\boldsymbol{Z}^+a + aA^+ \leqq b$ and $a(i) > 0$ for some $i \in I$ then $\boldsymbol{Z}^+a(i) + p^2(i)a(i) \leqq b(i)$ for all $p \in A^+$; and by Lemma 7, $\boldsymbol{Z}^+a(i) + R_i^+a(i) \leqq b(i)$. So, since R_i is r-Archimedean, $a(i) \leqq 0$, contrary to hypothesis. Thus $a \leqq 0$ and A is r-Archimedean.

4. **Archimedean f^*-rings.** A ring of $\bar{\boldsymbol{R}}$-valued functions on a nonempty set X is a nonempty class, R, of $\bar{\boldsymbol{R}}$-valued functions on X such that

(i) If $\{f_i\}_{i\in I}$ is any finite subclass of R there is at least one point x in X where every $f_i(x)$ is finite.

(ii) If f, g and h are in R and $f(x) \geqq g(x)$ for all x where $h(x)$ is finite then $f(x) \geqq g(x)$ for all x in X.

(iii) If f and g are in R then there are functions s, p and n in R such that $s(x) = f(x) + g(x)$ whenever $f(x)$ and $g(x)$ are not infinite and of opposite sign, $p(x) = f(x)g(x)$ unless $f(x) = 0$ and $g(x) = \pm\infty$ or vice versa, and $n(x) = -f(x)$ for all x in X.

Condition (ii) shows that such s, p and n are unique, so they may be denoted by $f + g$, fg and $-f$ respectively.

Subsets of X of the form $\{x: f(x) = \pm\infty\}$ are called *nul-sets* (a name suggested by integration theory and Condition (ii)).

It is easily seen that any ring of $\bar{\boldsymbol{R}}$-valued functions on a set X

is an Archimedean f^*-ring. Conversely, if A is an Archimedean f^*-ring, and for each $a \in A$ $\bar{a}$ denotes the function $P \to \overline{a(P)}$ defined on $\mathscr{P}(\overline{a(P)}$ was defined in the Introduction), then Lemma 8 below and the remarks in the Introduction show that for any distinguishing subset $\mathscr{D}$ of $\mathscr{P}$ $\bar{A} \mid \mathscr{D} = \{\bar{a} \mid \mathscr{D} : a \in A\}$ is a ring of $\bar{R}$-valued functions on $\mathscr{D}$, and the map $a \to \bar{a} \mid \mathscr{D}$ is an isomorphism of A onto $\bar{A} \mid \mathscr{D}$.

If $\mathscr{D}$ is any subset of $\mathscr{P}$, $a, b \in A$ and $\lambda \in \bar{\boldsymbol{R}}$ it is convenient to adopt conventions similar to $\mathscr{D}(\bar{a} \geqq \lambda)$ for $\{D \in \mathscr{D} : \overline{a(D)} \geqq \lambda\}$ and $\mathscr{D}(a \geqq b)$ for $\{D \in \mathscr{D} : a(D) \geqq b(D)\}$.

LEMMA 8. *If A is an Archimedean f^*-ring and $\mathscr{D}$ is a distinguishing subset of $\mathscr{P}$ and if $\mathscr{D}(\bar{a} < \bar{b})$ is a nul-set then $a \geqq b$.*

Proof. There is a $c \in A^+$ with $\mathscr{D}(\bar{c} = \infty) \supset \mathscr{D}(\bar{a} < \bar{b})$; so $e \overset{\text{def}}{=} c + a^2 + b^2$ satisfies $\mathscr{D}(a \neq 0) \cup \mathscr{D}(b \neq 0) \subset \mathscr{D}(e \neq 0)$ and $\mathscr{D}(\bar{a} < \bar{b}) \cup \mathscr{D}(\bar{a} = \pm\infty) \cup \mathscr{D}(\bar{b} = \pm\infty) \subset \mathscr{D}(\bar{e} = \infty)$.

Consider the following three situations which may occur for a $D \in \mathscr{D}$:

(i) $b > a, (D)$ and $\bar{e}(D) = \infty$; whence $\boldsymbol{Z}^+(b - a) \leqq e(b - a), (D)$ and so $\boldsymbol{Z}^+(b - a)2e \leqq e^4 + (b - a)^2, (D)$.

(ii) $b > a, (D)$, and $\bar{e}(D) < \infty$; whence $\bar{a}(D)$ and $\bar{b}(D)$ are finite, $(\overline{b - a})(D) = 0$, and so $\boldsymbol{Z}^+(b - a)2e \leqq 2e, (D)$.

(iii) $b \leqq a, (D)$.

In all cases $\boldsymbol{Z}^+(b - a) \leqq e^4 + (b - a)^2 + 2e, (D)$. So $\boldsymbol{Z}^+(b - a)e \leqq e^4 + (b - a)^2 + 2e$ and, A being Archimedean, $(b - a)e \leqq 0$. This, in an f^*-ring with e as here defined, implies $b - a \leqq 0$, that is $a \geqq b$.

COROLLARY. *No nul-set can contain a nonempty set of the form $\mathscr{D}(\bar{a} > 0)$.*

Let $\mathscr{M}^{**} = \{M \in \mathscr{M}^* : \exists a \in A \text{ with } \bar{a}(M) \text{ nonzero}\}$.

Lemma 8 shows that $\mathscr{M}^{**}$ is distinguishing and so the mapping $a \to \bar{a} \mid \mathscr{M}^{**}$ is an isomorphism of A onto $\bar{A} \mid \mathscr{M}^{**}$.

Two natural topologies for $\mathscr{M}^{**}$, $\mathscr{T}_1$ with the sets of the form $\mathscr{M}^{**}(a > 0)$ as a subbase, and $\mathscr{T}_2$ with the sets of the form $\mathscr{M}^{**}(\bar{a} > 0)$ as a subbase, turn out to be the same.

LEMMA 9. *$\mathscr{T}_1 = \mathscr{T}_2 (= \mathscr{T}$ say). $\mathscr{T}$ is Hausdorff and is the weak topology induced on $\mathscr{M}^{**}$ by $\bar{A}$.*

Proof. $\mathscr{T}_2 \supset \mathscr{T}_1$, for if $M \in \mathscr{M}^{**}(a > 0)$ there is a $b \in A$ with $\bar{b}(M) > 0$, and since $a(M)$ is an r-order unit, there are, using Lemma 7, $n \in \boldsymbol{Z}^+$ and $e \in A^+$ such that $na + e^2a > b, (M)$. So $M \in \mathscr{M}^{**}(\overline{na + e^2a} > 0) \subset \mathscr{M}^{**}(na + e^2a > 0) \subset \mathscr{M}^{**}(a > 0)$. Conversely, $\mathscr{T}_1 \supset \mathscr{T}_2$, for if $M \in \mathscr{M}^{**}(\bar{a} > 0)$ then for some $n \in \boldsymbol{Z}^{++}$,

$M \in \mathscr{M}^{**}(\bar{a} > 1/n)$; so $na^3 > a^2, (M)$ and $M \in \mathscr{M}^{**}(na^3 - a^2 > 0) \subset \mathscr{M}^{**}(\bar{a} \geqq 1/n) \subset \mathscr{M}^{**}(\bar{a} > 0)$. $\mathscr{T}$ is Hausdorff. If $M_1, M_2 \in \mathscr{M}^{**}$ and $M_1 \neq M_2$ there are $a_1 \in M_1 \backslash M_2$ and $a_2 \in M_2 \backslash M_1$. Whence $a = a_1 - a_2 \in (-M_1') \cap M_2'$, that is $M_2 \in \mathscr{M}^{**}(a < 0)$ and $M_1 \in \mathscr{M}^{**}(a > 0)$.

Finally, $\mathscr{T}$ is the weak topology induced by $\bar{A}$ on $\mathscr{M}^{**}$. For, by definition, $\mathscr{T}$ is coarser than this weak topology. Conversely, if $\lambda > -\infty \mathscr{M}^{**}(\bar{a} \geqq \lambda) = \cap \{\mathscr{M}^{**}(s\overline{ae}^2 \geqq r\bar{e}^2)\colon r/s < \lambda,\ s > 0 \text{ and } e \in A\}$, and so is closed with respect to $\mathscr{T}$.

Next it is shown that $\mathscr{M}^{**}(\bar{a} \geqq \varepsilon)$ is compact for all $\varepsilon > 0$ and all $a \in A$.

It is sufficient to prove the following result.

Lemma 10. *If $a \in A$ then $\mathscr{M}^{**}(\bar{a} \leqq -1)$ is compact.*

Proof. Alexander's Theorem ([3] p. 139) shows that it is sufficient to prove that any cover of $\mathscr{M}^{***}(\bar{a} \leqq -1)$ by sets of the form $\mathscr{M}^{**}(c < 0)$, $c \in A$, has a finite subcover.

Accordingly, suppose C is a subset of A such that $\{\mathscr{M}^{**}(c < 0)\colon c \in C\}$ covers $\mathscr{M}^{**}(\bar{a} \leqq -1)$ and contains no finite subcover. A contradiction will be derived from this.

Consider any $M \in \mathscr{M}^{**}(\bar{a} \leqq -1)$ and any rational number m/n with $n > 0$, $m > 2$ and $2/3 < m/n < 1$. Since $\bar{a}(M) \leqq -1$, $naa^4 < -ma^4, (M)$ so $naa^4 + (m-2)a^4 < -2a^4 < -a^2, (M)$, that is $[na \cdot a^2 + (m-2)a^2]a^2 + a^2 < 0, (M)$. Thus $[na \cdot a^2 + (m-2)a^2]a^2 + a^2 \in N = \cap\{M'\colon \bar{a}(M) \leqq -1\}$.

Let $K = \{na \cdot a^2 + (m-2)a^2\colon m \geqq 2, n > 0 \text{ and } 2/3 < m/n < 1\}$.

If $\{c_i\}_{i=1}^r \subset C$ there is an $M \in \mathscr{M}^{**}(\bar{a} \leqq -1)$ with $\{c_i\}_{i=1}^r \subset M$. So the semiring, S, generated by $A^+ \cup C$ is normal with respect to N and there is a $P \in \mathscr{P}$ with $P \supset S$ and $P \cap N = \phi$. For any $k \in K$, $ka^2 + a^2 < 0, (P)$, so $k(P)$ is not r-infinitesimal in A/P. There is therefore an $M_0 \in \mathscr{M}$ with $M_0 \cap K = \phi$. Now for any element $na \cdot a^2 + (m-2)a^2$ of K $na \cdot a^2 + (m-2)a^2 < 0, (M_0)$; whence $\bar{a}(M_0) \leqq -(m-2)/n$. Consequently $\bar{a}(M_0) \leqq -1$, so $M_0 \in \mathscr{M}^{**}$, while $M_0 \supset C$, which is contrary to the hypothesis on C.

$\mathscr{M}^{**}$ may include semirings M such that $\bar{A}(M) \subset \{0, \pm\infty\}$. Lemma 8 shows that these are not algebraically significant (i.e. $\mathscr{M}^{***} \stackrel{\text{def}}{=} \{M \in \mathscr{M}^{**}\colon \exists a \in A \text{ with } \bar{a}(M) \notin \{0, \pm\infty\}\}$ is distinguishing). Considered as a subspace of the topological space $\{\mathscr{M}^{**}, \mathscr{T}\}$, $\mathscr{M}^{***}$ is a Hausdorff space. Further, since for all $a \in A$ and all $\lambda, \varepsilon \in R^+$, $\mathscr{M}^{**}(\lambda \geqq \bar{a} \geqq \varepsilon)$ is a closed, and therefore compact, subset of $\{\mathscr{M}^{**}, \mathscr{T}\}$ which is contained in $\mathscr{M}^{***}$. So $\mathscr{M}^{***}$ is a locally compact Hausdorff space; for if $D \in \mathscr{M}^{***}$ and $D \in \mathscr{M}^{**}(\bar{a} > 0)$ there is a $b \in A$ with $\infty > \bar{b}(D) > 0$, so $\mathscr{M}^{**}(\bar{a} \geqq 1/2\,\bar{a}(D) \wedge 1) \cap \mathscr{M}^{**}(2\bar{b}(D) \geqq \bar{b} \geqq 1/2\,\bar{b}(D))$ is a compact neighbourhood of D in $\mathscr{M}^{***}$.

The following analogue of [2] Theorem 4.1 has now been proved.

THEOREM 3. *If A is an Archimedean f^*-ring the mapping $a \to \bar{a} \mid \mathscr{M}^{***}$ is an isomorphism of A onto a ring $\bar{A} \mid \mathscr{M}^{***}$ of extended real valued functions on $\mathscr{M}^{***}$. The weak topology induced on $\mathscr{M}^{***}$ by $\bar{A} \mid \mathscr{M}^{***}$ is Hausdorff and locally compact and relative to it each set $\mathscr{M}^{***}(\lambda \geqq \bar{a} \geqq \varepsilon)$ with $a \in A$ and $\lambda, \varepsilon \in \boldsymbol{R}^{++}$ is compact. No function is infinite at every point of a nonempty set of the form $\mathscr{M}^{***}(\bar{a} > 0)$.*

The rest of Johnson's theorem seems to require that A be an f-ring.

5. **f-rings.** A commutative f-ring is a *po*-ring A which is lattice ordered in such a way that if $a \wedge b = 0$ then $ac \wedge b = 0$ for all $c \in A^+$.

An f-ring without nonzero nilpotents is an f^*-ring. For if $b, c \in A$ and $b \wedge c = 0$ then $bc \wedge bc = 0$, that is $bc = 0$. So for any $a \notin A^+$, $s \in A^+$ and $n \in \boldsymbol{Z}^{++}$, $sa - a^{2n} = sa^+ - sa^- - (a^+)^{2n} - (a^-)^{2n} \leqq sa^+ - (a^-)^{2n}$. And the latter expression is not in A^+ since $a^+ \wedge a^- = 0$ yields $sa^+ \wedge (a^-)^{2n} = 0$; whence $(sa^+ - (a^-)^{2n})^- = (a^-)^{2n} \neq 0$. Furthermore if A is an f^*-ring which is lattice ordered and such that $a \wedge b = 0$ implies $ab = 0$ then for any $P \in \mathscr{P}$, $(a \wedge b)(P) = a(P) \wedge b(P)$. For if $a \wedge b = c$ then $(a - c) \wedge (b - c) = 0$, so $(a - c)(b - c) = 0$; whence $(a - c)(P)(b - c)(P) = 0$. But A/P is a ring with integrity, so $(a - c)(P) = 0$ or $(b - c)(P) = 0$. Therefore, since $(a - c) \geqq 0$ and $(b - c) \geqq 0$, $(a - c)(P) \wedge (b - c)(P) = 0$ and $a(P) \wedge b(P) = c(P) = (a \wedge b)(P)$. Consequently the isomorphisms set up in Theorems 1 and 2 are isomorphisms onto a lattice ordered subdirect union of *toi*-rings which preserve lattice relations.

As for Theorem 3, it follows that for any $a, b \in A$ and any $M \in \mathscr{M}^{***}$, $\bar{a}(M) \wedge \bar{b}(M) = \bar{a} \wedge \bar{b}(M)$. Whence the sets $\{\mathscr{M}^{***}(a > 0)\}_{a \in A}$ form a basis for $\mathscr{T}$ and so does the class of sets $\{\mathscr{M}^{***}(\bar{a} > 0)\}_{a \in A}$. So each function a is finite on a dense subset of $\mathscr{M}^{***}$ (i.e. it is an *extended function* in the sense of [2]). Finally, Lemma 2.6 (ii) of [2] may be used to prove that the topology of $\mathscr{M}^{***}$ is precisely the weak topology induced by the bounded functions in $\bar{A} \mid \mathscr{M}^{***}$.

Note added in proof. Lemma 3, together with the remark at the end of the fourth paragraph of § 3, shows that for any toi ring, R, the following three properties are equivalent:

(i) R is r-Archimedean,

(ii) R has no nonzero r-infinitesimal elements,

(iii) Every element of R^{++} is an r-order unit.

So Theorem 2 can be sharpened. For example, we may omit "with no nonzero r-infinitesimal elements".

References

1. G. Birkhoff, and R. S. Pierce, *Lattice ordered rings*, An. Acad. Brasil Ci., **28** (1956), 41–69.
2. D. J. Johnson, *A representation theory for a class of Archimedean lattice ordered rings*, Proc. London Math. Soc., **12** (1962), 207–225.
3. J. L. Kelley, *General Topology*, Van Nostrand. (New York 1955).
4. R. S. Pierce, *Radicals in function rings*, Duke Math. J., **23** (1956), 253–261.

Purdue University

ON A GENERALIZED STIELTJES TRANSFORM

J. M. C. JOSHI

1. Introduction. In a series of recent papers [1–4] I have discussed various properties and inversion theorems etc. for the transform

$$(1.1) \qquad F(x) = \frac{\Gamma(\beta+\eta+1)}{\Gamma(\alpha+\beta+\eta+1)} \times \int_0^\infty (xy)^\beta {}_1F_1(\beta+\eta+1;\ \alpha+\beta+\eta+1;\ -xy) f(y)dy$$

where $f(y) \in L(0, \infty)$, $\beta \geqq 0$, $\eta > 0$.

$$F(x) = A\int_0^\infty (xy)^\beta F(x, y) f(y)dy$$

where, for convenience, we denote $\Gamma(\beta+\eta+1)/\Gamma(\alpha+\beta+\eta+1)$ by A and ${}_1F_1(a; b; -xy)$ by $F(x, y)$, a and b standing respectively for $\beta+\eta+1$ and $\alpha+a$. For $\alpha=\beta=0$ (1.1) reduces to the well known Laplace Transform

$$(1.2) \qquad F(x) = \int_0^\infty e^{-xy} f(y)dy\ .$$

The transform (1.1), which may be called a generalization of Laplace Transform, arises when we apply Kober's [5] operators of Fractional Integration [6] to $x^\beta e^{-x}$.

The object of the present paper is to give a generalization of Stieltjes Transform, to give an inversion theorem for it and to use that inversion theorem to obtain an inversion theorem for the transform (1.1). In another paper (to appear elsewhere) I have found out inversion operators directly for (1.1).

2. Generalized Stieltjes transform. We prove

THEOREM 2.1. *If*

$$(2.1) \qquad \phi(s) = \int_0^\infty e^{-sx} F(x)dx$$

where $F(x)$ is given by the convergent integral (1.1), *then*

$$(2.2) \qquad \phi(s) = \frac{A\Gamma(\beta+1)}{s}\int_0^\infty \left(\frac{y}{s}\right)^\beta F\left(a, \beta+1; b; -\frac{y}{s}\right) f(y)dy$$

Received, July, 1, 1963.

provided that $\beta \geqq 0$, $\eta > 0$ and $f(y) \in L(0, \infty)$.

Proof. We have

$$\begin{aligned}\phi(s) &= A\int_0^\infty e^{-sx}dx\int_0^\infty (xy)^\beta {}_1F_1(a; b; -xy)f(y)dy \\ &= A\int_0^\infty y^\beta f(y)dy\int_0^\infty x^\beta e^{-sx}{}_1F_1(a; b; -xy)dx\end{aligned}$$

on changing the order of integration, which is easily seen to be justified under the conditions stated, since [7, page 59]

$$ {}_1F_1(a; b; -x) = \frac{\Gamma(b)}{\Gamma(b-a)}x^{-a}\{1 + 0[|x)^{-1}\} \qquad (x \to \infty)$$

and

$$ {}_1F_1(a; b; -x) = 0(1) \qquad (x \to 0).$$

Therefore [7, page 43]

$$\phi(s) = \frac{A\Gamma(\beta+1)}{s}\int_0^\infty \left(\frac{y}{s}\right)^\beta F\left(a, \beta+1; b; -\frac{y}{s}\right)f(y)dy$$

under the conditions stated.

COROLLARY 2.1(a). *When $\beta = 0$, $\eta = 2m$, $\alpha = -m - k + (1/2)$, $\phi(s)$ reduces to the generalization of Stieltjes Transform*

$$\begin{aligned}(2.3) \qquad \phi(s) = &\frac{\Gamma(2m+1)}{\Gamma\left(m - k + \frac{3}{2}\right)} \\ &\times \frac{1}{s}\int_0^\infty F\left(2m+1, 1; m - k + \frac{3}{2}; -\frac{y}{s}\right)f(y)dy\end{aligned}$$

introduced by Varma [8]

COROLLARY 2.1(b). *When $\alpha = \beta = 0$, then $\phi(s)$ reduces to the well known Stieltjes Transform* [9, page 323]

$$(2.4) \qquad \phi(s) = \int_0^\infty (s+y)^{-1}f(y)dy\,.$$

COROLLARY 2.1(c). *When $\beta = 0$, $\alpha = -\eta = 1 - \sigma$, $\phi(s)$ reduces to another generalization of Stieltjes Transform* [9, page 328]

$$(2.5) \qquad \chi(s) = \frac{\phi(s)}{\Gamma(\sigma)s^{\sigma-1}} = \int_0^\infty \frac{f(y)}{(s+y)^\sigma}dy\,.$$

3. Generalized Stieltjes transform as convolution transform. In this section we will find out an inversion operator for the generalized Stieltjes Transform (2.2) by putting it into the form of Convolution Transform. The Convolution Transform with kernel $G(x)$ of the function $\phi(x)$ into $f(x)$ is defined as [10, page 4]

$$f(x) = \int_{-\infty}^{\infty} G(x-t)\phi(t)dt \ . \tag{3.1}$$

The corresponding inversion function $E(x)$, which serves to invert the transform, is defined by the equation

$$[E(x)]^{-1} = \int_{-\infty}^{\infty} G(y)e^{-xy}dy \ .$$

If $\phi(s)$ be defined as in (2.2), we have

$$-\phi'(s) = \frac{A\Gamma(\beta+1)}{s^{\beta+2}}\int_0^{\infty} F\left(a, \beta+2; b; -\frac{y}{s}\right)(sy)^{\beta} f(y)dy$$

because, by Euler's theorem on homogeneous functions,

$$S\left(\frac{\partial}{\partial s}\right)\left[\left(\frac{s}{y}\right)^{-\beta-1} F\left(a, \beta+1; b; -\frac{y}{s}\right)\right]$$
$$= -y\left(\frac{\partial}{\partial y}\right)\left[\left(\frac{s}{y}\right)^{-\beta-1} F\left(a, \beta+1; b; -\frac{y}{s}\right)\right]$$

or

$$\left(\frac{\partial}{\partial s}\right)\left[\left(\frac{y}{s}\right)^{\beta} \cdot \frac{1}{s} F\left(a, \beta+1; b; -\frac{y}{s}\right)\right]$$
$$= -\frac{1}{s^{\beta+2}}\left(\frac{\partial}{\partial y}\right)\left[y^{\beta+1} F\left(a, \beta+1; b; -\frac{y}{s}\right)\right]$$

and

$$\left(\frac{\partial}{\partial y}\right)[y^{\beta+1}F(a, \beta+1; b; y)] = y^{\beta}F(a, \beta+2; b; y) \ .$$

Therefore

$$-e^s\phi'(e^s) = A\Gamma(\beta+1)\int_{-\infty}^{\infty} e^{-(s-y)(\beta+1)}F(a, \beta+2; b; e^{-(s-y)})f(e^y)dy$$

or

$$\xi(s) = A\Gamma(\beta+1)\int_{-\infty}^{\infty} e^{-(s-y)(\beta+1)}F(a, \beta+2; 2; e^{-(s-y)})\zeta(y)dy$$

where

$$\xi(s) \equiv -e^s\phi'(e^s)$$

and

$$\zeta(s) \equiv f(e^s) .$$

Therefore the inversion function $E(x)$ is given by the equation

$$\begin{aligned}\frac{1}{E(x)} &= A\Gamma(\beta+1)\int_{-\infty}^{\infty} e^{-y(\beta+x+1)}F(a, \beta+2; b; -e^{-y})dy \\ &= \frac{\Gamma(\eta-x)\Gamma(\beta+x+1)\Gamma(1-x)}{\Gamma(\alpha+\eta-x)}\end{aligned}$$

provided that

$$b \neq 0, -1, -2, \cdots, Re(1-x) > 0, Re(\eta - x) > 0$$

and

$$Re(\beta + x + 1) > 0$$

since [11, page 79]

$$\int_0^\infty z^{-s-1}F(a, b; d; -z)dz = \frac{\Gamma(a)\Gamma(a+s)\Gamma(b+s)\Gamma(-s)}{\Gamma(a)\Gamma(b)\Gamma(d+s)}$$

if

$$Res < 0, Re(a+s) > 0, Re(b+s) > 0 ,$$

and $d \neq 0$ or a negative integer.

Therefore,

$$E(D)\{\xi(s)\} = \zeta(s)$$

or

$$\frac{\Gamma(\alpha+\eta-D)}{\Gamma(\beta+1+D)\Gamma(1-D)}\{-e^s\phi'(e^s)\} = f(e^s) , \qquad D \equiv \frac{d}{ds}$$

and we shall give definite meaning to the operations involved. Now

$$\frac{1}{\Gamma(1-x)} = \lim_{n\to\infty} n^x \prod_{k=1}^{n}\left(1 - \frac{x}{k}\right)$$

and

$$\frac{\Gamma(\alpha+\eta-x)}{\Gamma(\eta-n)\Gamma(\beta+x+1)} = \lim_{n\to\infty}\frac{n^{\alpha-\beta-x}}{\Gamma(n+2)}\prod_{k=0}^{n}\frac{(D-\eta-k)(D+\beta+1+k)}{(D-\alpha-\eta-k)} .$$

Also we have [10, page 66]

$$\prod_{k=1}^{n-1}\left(1-\frac{D'}{k}\right)[e^x F(e^x)] = \frac{(-)^{n-1}}{(n-1)!}e^{nx}F^{(n-1)}(e^x)$$

and

$$\prod_{k=0}^{n}(D'+a+k)[e^{-(a+n)x}F(e^x)] = e^{-(a-1)x}F^{(n+1)}(e^x)$$

$$\prod_{k=0}^{n}(D'+a-k)[e^{-ax}F(e^x)] = e^{(n+1-a)x}F^{(n+1)}(e^x)$$

$$\prod_{k=0}^{n}(D'+a-k)^{-1}[e^{(n+1-a)x}F(e^x)] = e^{-ax}F^{(-n-1)}(e^x)$$

where $F^{(-n-1)}(x)$ denotes a function $\psi(x)$ such that

$$\left(\frac{d}{dx}\right)^{n+1}[\psi(x)] = F(x)\,, \qquad D' \equiv \frac{d}{dx}\,.$$

Using the above relations,

$$\begin{aligned} &E(D)\{-e^s\phi'(e^s)\} \\ &\equiv (-)^n n^{\alpha-\beta}e^{(\alpha+\eta)s}D_1^{-n-1}e^{-\alpha s}D_1^{n+1}e^{-(\eta+\beta)s} \\ &\times D_1^{n+1}e^{(2n+\beta+1)s}\phi^{(n)}(e^s) = f(e^s)\,, \qquad D_1 \equiv \frac{d}{de^s},\ (n\to\infty)\,. \end{aligned}$$

Returning to original variables, we have.

$$\begin{aligned} \text{(3.2)} \qquad &\lim_{n\to\infty}(-)\frac{n\Gamma(n+\alpha)}{\Gamma(n+\beta)\Gamma(n)\Gamma(n+2)} \\ &\times S^{\alpha+\eta}D^{-n-1}s^{-\alpha}D^{n+1}s^{-(\eta+\beta)}D^{n+1}s^{2n+\beta+1}\phi^{(n)}(s)\,. \end{aligned}$$

We thus have.

THEOREM 3.1. *$f(s) \in C\cdot B$ on $0 < s < \infty$ and if the integral* (2.2) *converges, then* (3.2) *holds for* $s > 0$.

COROLLARY 3.1(a). *When $\beta = 0$, $\alpha = -m-k+(1/2)$, $\eta = 2m$ we have the corresponding result for Varma's Transform.*

COROLLARY 3.1(b). *When $\alpha = \beta = 0$ we have the Theorem* 9.4 *of Hirschman and Widder* [10, page 69].

COROLLARY 3.1(c). *Similarly for $\alpha = -\eta = 1-\sigma$ and $\beta = 0$ we have a theorem for* (2.5).

4. Application to generalized Laplace transform. We may now use inversion formula derived above to obtain a new inversion of the Generalized Laplace Transform (1.1). For we have, as above

$$(4.1)\qquad \phi(s) = \frac{A\Gamma(\beta+1)}{s}\int_0^\infty \left(\frac{y}{s}\right)^\beta F\left(a_1\beta+1; b; -\frac{y}{s}\right)f(y)dy\ .$$

Therefore if we invert the integral (4.1) we get $f(y)$. But

$$\phi(s) = \int_0^\infty e^{-sn}F(x)dx\ .$$

Therefore

$$\begin{aligned}\phi^{(n-1)}(s) &= (-)^{n-1}\int_0^\infty e^{-sx}x^{n-1}F(x)dx \\ &= \frac{(-)n-1}{s^n}\int_0^\infty e^{-y}y^{n-1}F\left(\frac{y}{s}\right)dy\end{aligned}$$

by a simple change of variable. But the repeated use of the theorem

$$\left(\frac{\partial}{\partial x}\right)\left[\frac{x^{n+\beta-1}}{y^{n+\beta}}f\left(\frac{y}{x}\right)\right] = -\left(\frac{\partial}{\partial y}\right)\left[\frac{x^{n+\beta-2}}{y^{n+\beta-1}}f\left(\frac{y}{x}\right)\right]$$

gives

$$\left(\frac{\partial}{\partial x}\right)^n\left[\frac{x^{n+\beta 1}}{y^{n+\beta}}f\left(\frac{y}{x}\right)\right] = (-)^n\left(\frac{\partial}{\partial y}\right)^n\left[\frac{x^{\beta-1}}{y^\beta}f\left(\frac{y}{x}\right)\right].$$

Therefore,

$$D^n s^{2n+\beta-1}\varphi^{(n-1)}(s) = (-)^{2n-1}s^{n-1+\beta}\int_0^\infty e^{-sx}D^n[x^\beta F(x)]dx\ .$$

Similarly,

$$\begin{aligned}&D^n s^{-(\eta+\beta)}D^n s^{2n+\beta-1}\phi^{(n-1)}(s) \\ &\quad = (-)^n\int_0^\infty e^{-sx}s^{-(\eta+1)}f_1(x)dx\end{aligned}$$

where, for convenience, we write

$$f_1(x) = x^{-n-\eta-1}D'^n x^{\eta+\beta+2n}D'^n(x^\beta F(x))\ .$$

Then

$$\begin{aligned}&D^{-n}s^{-d}D^n s^{-(\eta+\beta)}D^n s^{2n+\beta-1}\varphi^{(n-1)}(s) \\ &\quad = \int_0^\infty e^{-sn}s^{n-1}(sn)^{-\eta-\alpha-1}D'^{-n}\{x^{\eta+\alpha+1-n}f_1(x)\}dx\ .\end{aligned}$$

Therefore finally we have,

$$\lim_{n\to\infty}(-)^n \frac{\Gamma(n-1+\alpha)}{\Gamma(n-1+\beta)\Gamma(n+1)\Gamma(n-1)}$$
$$\times\, s^{\alpha+\eta}D^{-n}s^{-\alpha}D^n s^{-(\eta+\beta)}D^n s^{2n+\beta-1}\varphi^{(n-1)}(s)$$
$$=\lim_{n\to\infty}(-)^n \frac{\Gamma(n-1+\alpha)}{\Gamma(n-1+\beta)\Gamma(n+1)\Gamma(n-1)}$$
$$\times\int_0^\infty e^{-sx}s^{n-1}x^{-\eta-\alpha-}D'^{-n}x^{\alpha}$$
$$\times\, D'^n x^{\eta+\beta+2n}D'^n\{x^\beta F(x)\}dx = f(s)\cdots(A)\,.$$

We have thus proved

THEOREM 4.1. *If $f(x)\in L$ in $0<x<\infty$ and if $F(x)$ is defined by the convergent integral* (1.1) *then the result* (A) *holds for almost all positive values of s.*

COROLLARY 4.1. *When $\alpha=\beta=0$ we have Theorem* 25(a) *of Widder* [9, page 385].

I am indebted to Dr. K. M. Saksena for guidance and help in the preparation of the paper.

REFERENCES

1. J. M. C. Joshi, *Convergence Theorems for a Generalization of Laplace Transform*, accepted for publication in Nieuw Archief Voor Wiskunde.
2. ———, *Abelian Theorem for a Generalization of Laplace Transform*, accepted for publication in Collectanea Mathematica, Barcelona, Vol. 15.
3. ———, *Real Inversion Theorem for a Generalization of Laplace Transform*, accepted for publication in Collectanea Mathematica, Barcelona Vol. 14.
4. ———, *A complex Inversion Theorem for a Generalization of Laplace Transform* communicated for publication to Collectanea Mathematica.
5. H. Kober, Quart Jour, Math, (Oxford) **11** (1940), 193-211.
6. A. Erdelyi, Rendi Conti Del Seminario Mathematico, **10** (1950), 217-234.
7. L. J. Stater, *Confluent Hypergeometric Functions*, Cambridge University Press, 1960.
8. R. S. Varma, Proceedings of the National Academy of Sciences, India, **20** (1951) 209-216.
9. D. V. Widder, *The Laplace Transform*, Princeton University Press, 1946.
10. Hirschman and Widder, *The Convolution Transform*, Princeton University Press, 1955.
11. A. Erdelyi, etc. *Higher Transcendental Function*, Vol. 1, 1953.

D. S. B. GOVT. COLLEGE, NAINITAL, INDIA

INVERSION AND REPRESENTATION THEOREMS FOR A GENERALIZED LAPLACE TRANSFORM

J. M. C. JOSHI

1. Introduction. In a series of recent papers I have discussed various properties and inversion theorems etc. for the transform

$$F(x) = \frac{\Gamma(\beta+\eta+1)}{\Gamma(\alpha+\beta+\eta+1}\int_0^\infty (xy)^\beta {}_1F_1(\beta+\eta+1; \alpha+\beta+\eta+1; -xy)f(y)dy \; . \tag{1.1}$$

where $f(y) \in L0, \infty)$, $\beta \geqq 0$, $\eta > 0$.

$$= A\int_0^\infty (xy)^\beta \psi(x, y)f(y)dy$$

where for convenience we denote $\Gamma(\beta+\eta+1)/\Gamma(\alpha+\beta+\eta+1)$ by A and ${}_1F_1(a; b; -xy)$ by $\psi(xy)$; a and b standing respectively for $\beta+\eta+1$ and $a+\alpha$. For $\alpha=\beta=0$ (1.1) reduces to the wellknown Laplace transform

$$F(x) = \int_0^\infty e^{-xy}f(y)dy \; . \tag{1.2}$$

The transform (1.1), which may be called a generalization of the Laplace transform, arises if we apply Kober's operators of fractional integration [2] to the function $x^\beta e^{-x}$[1].

The object of the present paper is to obtain an inversion and a representation theorem for the transform (1.1) by using properties of Kober's operators defined below.

2. Definition of operations. The operators given by Kober are defined as follows.

$$I^+_{\eta,\alpha}[f(x)] = \frac{1}{\Gamma(\alpha)}x^{-\eta-\alpha}\int_0^x (x-u)^{\alpha-1}u^\eta f(u)du$$

$$K^-_{\zeta\,\alpha}[f(x)] = \frac{1}{\Gamma(\alpha)}x^\zeta\int_n^\infty (u-x)^{\alpha-1}u^{-\zeta-\alpha}f(u)du$$

where $f(x) \in L_p(0, \infty)$, $1/p + 1/q = 1$, if $1 < p < \infty$ and $1/p$ or $1/q$ 0 if p or $q=1$, $\alpha > 0$, $\zeta > -(1/p)$, $\eta > -(1/q)$.

The Mellin transform $\bar{M}f(x)$ of a function $f(x) \in L_p(0, \infty)$ is defined as

Received November 4, 1963.

$$\bar{M}f(x) = \int_0^\infty f(x)x^{it}du \qquad (p = 1)$$

and

$$= \overset{\text{index}V}{\lim_{x\to\infty}} \int_{1/x}^{x} f(x)^{it-1/q}dn \qquad (p > 1)\,.$$

The inverse Mellin transform $M^{-1}\phi(t)$ of a function $\phi(t) \in L_q(-\infty, \infty)$ is defined by

(2.1) $$M^{-1}\phi(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \phi(t)x^{-it}dt \qquad (q = 1)$$

and

$$= \frac{1}{2\pi} \overset{\text{index}p}{\lim_{T\to\infty}} \int_{-T}^{T} \phi(t)x^{-it-1/p}dt \qquad (q > 1)\,.$$

If Mellin transform is applied to Kober's operators and the orders of integrations are interchanged we obtain, under certain conditions

$$\bar{M}\{I_{\eta}^{+}{}_{\alpha}f(x)\} = \frac{\Gamma\left(\eta + \frac{1}{q} - it\right)}{\Gamma\left(\alpha + \left\{\eta + \frac{1}{q} - it\right\}\right]}\bar{M}f(x)$$

and

$$\bar{M}\{K_{\zeta}^{-}{}_{\alpha}f(x)\} = \frac{\Gamma\left(\zeta + \frac{1}{p} + it\right)}{\Gamma\left[\alpha + \left(\zeta + \frac{1}{p} + it\right)\right]}\bar{M}f(x)\,.$$

But

$$\bar{M}(e^{-x}\cdot x^{\beta}) = \int_0^\infty e^{-x}x^{\beta+it-1/q}dx = \Gamma\left(\beta + it + \frac{1}{p}\right), \quad \text{if } Re\left(\beta + \frac{1}{p}\right) > 0\,.$$

Therefore

$$\bar{M}\{I_{\eta,\alpha}^{+}(x^{\beta}e^{-x})\} = \frac{\Gamma\left[\left(\eta + \frac{1}{q} - it\right)\right]\Gamma\left(\beta + \frac{1}{p} + it\right)}{\Gamma\left[\alpha + \left\{\eta + \frac{1}{q} - it\right\}\right]}$$

and

$$\bar{M}\{K^-_{\zeta,\alpha}(x^\beta e^{-x})\} = \frac{\Gamma\left(\beta + it + \frac{1}{p}\right) P\left(\zeta + it + \frac{1}{p}\right)}{\Gamma\left[\alpha + \left\{\zeta + \frac{1}{p} + it\right\}\right]} .$$

By (2.1) we then have

$$I^+_{\eta,\alpha}(x^\beta e^{-x}) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{\Gamma\left(\eta + \frac{1}{q} - it\right)\Gamma\left(\beta + \frac{1}{p} + it\right)}{\Gamma\left[\alpha + \left(\eta + \frac{1}{q} - it\right)\right]} x^{-it-1/p} dt \tag{2.2}$$

and

$$K^-_{\zeta,\alpha}(x^\beta e^{-x}) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{\Gamma\left(\zeta + \frac{1}{p} + it\right)\Gamma\left(\beta + \frac{1}{p} + it\right)}{\Gamma\left[\alpha + \left(\zeta + \frac{1}{p} + it\right)\right]} x^{-it-1/p} dt ,$$

provided that $1/p > 0$, $\eta + 1/q > 0$ and $\zeta + 1/p > 0$.

3. **Inversion theorem.** We now define an inversion operator which will serve to invert (1.1).

An operator is defined for integral values of n by the relations

$$W_0[G(x)] = G(x) ,$$

$$W_n[G(x)] = (-)^n n^{\beta+n+1}\left(\frac{d}{dx}\right)^n [x^{-\beta}G(x)], \ (n = 1, 2, \cdots)$$

$$Q_{n,t}[G(x)] = \frac{1}{\Gamma(n + 1 + \beta - \alpha)}[W_n[G(x)]]_{n=n/t} (n = 1, 2, \cdots) .$$

THEOREM 3.1. *If $f(t)$ is bounded in $(0 < t < \infty)$ then, provided that the integral* (1.1) *converges, $\eta > 0$, $\beta \geqq 0$*

$$f(t) = \lim_{n\to\infty} Q_{n,t}[F(x)]$$

for almost all positive t.

Proof. Let x be any number greater than zero. Then, since the integral (1.1) converges, we can differentiate under the integral sign. Also (2.2) gives

$$\left(\frac{d}{dx}\right)[x^{-\beta}I_{n,\alpha}(x^\beta e^{-x})] = -x^{-\beta}I_{\eta+1,\alpha}[x^\beta e^{-x}] . \tag{3.1}$$

Using this relation we get

$$W_n[F(n)] = (-)^n n^{\beta+n+1}\int_0^\infty x^{-\beta} y^n I_{\eta+n,\alpha}\{(xy)^\beta e^{-xy}\} f(y)dy$$
$$= \frac{\Gamma(\beta+\eta+n+1)}{\Gamma(\alpha+\beta+\eta+n+1)}\int_0^\infty y^{\beta+n}{}_1F_1(\beta+\eta+n+1;$$
$$\alpha+\beta+\eta+n+1-xy)f(y)dy\ .$$

Therefore

$$Q_{n,t}\{F(x)\}$$
$$= \frac{\Gamma(\beta+\eta+1)}{\Gamma(\alpha+\beta+\eta+1)}\left(\frac{n}{t}\right)^{\beta+n+1}\frac{1}{\Gamma(n+\beta+1-\alpha)}$$
$$\times\int_0^\infty y^{\beta+n}{}_1F_1(n+\beta+\eta+1;\alpha+\beta+\eta+1+n;\ -xy)f(y)dy$$
$$= \frac{1}{\Gamma(n+\beta+1-\alpha)}\left(\frac{n}{t}\right)^{\beta+n+1}\frac{\Gamma(a)}{\Gamma(b)}$$
$$\times\int_0^\infty y^{\beta+n}{}_1F_1(a+n;b+n;\ -xy)f(y)dy$$

in the notation of § 1.

$$= \frac{\Gamma(a+n)}{\Gamma(b+n)\Gamma(n+\beta+1-\alpha)}\left(\frac{n}{t}\right)^{n+\beta+1}$$
$$\times\int_0^\infty (tv)^{n+\beta}{}_1F_1(a+n;b+n;\ -nv)f(tv)dt$$
$$= \frac{\Gamma(a+n)}{\Gamma(b+n)\Gamma(n+\beta+1-\alpha)}\left(\frac{n}{t}\right)^{n+\beta+1}$$
$$\times\int_0^\infty v^{n+\beta}{}_1F_1(\beta+\eta+n+1;\alpha+\beta+\eta+n+1;\ -nv)f(tv)dt$$

by a simple change of variable. Now by using a result of Slater [4] we have

$$\frac{\Gamma(a+n)}{\Gamma(b+n)}{}_1F_1(a+n;b+n;\ -v)\sim(nv)^{a-b}e^{-nv}\qquad(n\to\infty).$$

Therefore

$$\lim_{n\to\infty} Q_{n,t}\{F(n)\} = \lim_{n\to\infty}\frac{n^{\beta+n+1-\alpha}}{\Gamma(n+\beta+1-\alpha)}\int_0^\infty v^{n+\beta-\alpha}e^{-nv}f(tv)dv\ .$$

But [3] we have for almost all positive t

$$\lim_{n\to\infty}\frac{n^{\beta+n+1-\alpha}}{\Gamma(n+\beta+1-\alpha)}\int_0^\infty y^{n+\beta-\alpha}e^{-ny}\{f(ty)-f(t)\}dy = 0$$

and so we have our theorem.

5. **Representation theorem.** In this section we propose to give a set of necessary and sufficient conditions for the representation of a function as an integral of the form (1.1). We shall need a lemma which we now prove.

LEMMA 4.1. *If n is a positive integer and x and t are positive variables then*

$$\left(\frac{\partial}{\partial t}\right)^n\left[t^{\beta+n-1}I_{\eta,\alpha}\left\{\left(\frac{x}{t}\right)^\beta e^{-x/t}\right\} = \frac{n^n}{t^{n+1-\beta}}I_{\eta+n,\alpha}\left\{\left(\frac{x}{t}\right)^\beta e^{-x/t}\right\}.$$

Proof. It is plain that

$$\left(\frac{t}{x}\right)^{\beta+n-1}I_{\eta,\alpha}\left\{\left(\frac{x}{t}\right)^\beta e^{-x/t}\right\}$$

is a homogeneous function of zero order. Therefore applying Euler's theorem we get

$$t\left(\frac{\partial}{\partial t}\right)\left[\left(\frac{t}{x}\right)^{\beta+n-1}I_{\eta,\alpha}\left\{\left(\frac{x}{t}\right)^\beta e^{-x/t}\right\}\right] + n\left(\frac{\partial}{\partial x}\right)\left[\left(\frac{t}{x}\right)^{\beta+n-1}I_{\eta,\alpha}\left\{\left(\frac{x}{t}\right)^\beta e^{-x/t}\right\}\right] = 0$$

or

$$\left(\frac{\partial}{\partial t}\right)\left[\frac{t^{\beta+n-1}}{x^{\beta+n}}I_{\eta,\alpha}\left\{\left(\frac{x}{t}\right)^\beta e^{-x/t}\right\} = -\left(\frac{\partial}{\partial x}\right)\left[\frac{t^{\beta+n-2}}{x^{\beta+n-1}}I_{\eta,\alpha}\left\{\left(\frac{x}{t}\right)^\beta e^{-x/t}\right\}\right]$$

or

$$\begin{aligned}\frac{\partial^2}{\partial t^2}\left[\frac{t^{\beta+n-1}}{x^{\beta+n}}I_{\eta,\alpha}\left\{\left(\frac{x}{t}\right)^\beta e^{-x/t}\right\}\right] &= -\frac{\partial^2}{\partial t\partial x}\left[\frac{t^{\beta+n-2}}{x^{\beta+n-1}}I_{\eta,\alpha}\left\{\left(\frac{x}{t}\right)^\beta e^{-x/t}\right\}\right]\\ &= -\left(\frac{\partial}{\partial x}\right)\left[\frac{\partial}{\partial t}\left\{\frac{t^{\beta+n-2}}{x^{\beta+n-1}}I_{\eta,\alpha}\left\{\left(\frac{x}{t}\right)^\beta e^{-x/t}\right\}\right\}\right]\\ &= (-)^2\frac{\partial^2}{\partial x^2}\left[\frac{t^{\beta+n-3}}{x^{\beta+n-2}}I_{\eta,\alpha}\left\{\left(\frac{x}{t}\right)^\beta e^{-x/t}\right\}\right].\end{aligned}$$

Proceeding in the same manner we have

$$\frac{\partial^n}{\partial t^n}\left[\frac{t^{\beta+n-1}}{x^{\beta+n}}I_{\eta,\alpha}\left\{\left(\frac{x}{t}\right)^\beta e^{-n/t}\right\}\right] = \frac{t^{\beta-n-1}}{x^\beta}I_{\eta+n,\alpha}\left\{\left(\frac{x}{t}\right)^\beta e^{-x/t}\right\}\right]$$

using (3.1).

THEOREM 4.1. *The necessary and sufficient conditions that a given function $F(x)$ may have the representation* (1.1) *with $f(y)$ bounded and $Re\,\eta > 0$ $Re\,\beta \geqq 0$ are that*

(i) *$F(x)$ has derivatives of all orders in $0 < x < \infty$.*

(ii) *$F(x)$ tends to zero as x tends to infinity and*
(iii) $|Q_{n,t}\{F(x)\}| < M$ *for all integral* n $(0 < t < \infty)$.

Proof. First let us suppose that $F(x)$ has the representation (1.1). Under the conditions of the theorem it is obvious that all the derivatives of $F(x)$ exist. Also

$$\begin{aligned} F(x) &\leqq M' \frac{\Gamma(\beta+\eta+1)}{\Gamma(\alpha+\beta+\eta+1)} \\ &\quad \times \int_0^\infty (xy)^\beta {}_1F_1(\beta+\eta+1;\, \alpha+\beta+\eta+1;\, -xy)dy \\ &= \frac{M'\Gamma(\eta)\Gamma(\beta+1)}{x\Gamma(\alpha+\eta)} \end{aligned}$$

since $f(y)$ is bounded. So $F(x)$ tends to zero as x tends to infinity. To prove the necessity of (iii) we see, as in Theorem 3.1, that

$$|Q_{n,t}\{F(x)\}| \leqq \left\{\frac{n^{\beta+n+1-\alpha}}{\Gamma(n+\beta+1-\alpha)}\int_0^\infty v^{n+\beta-\alpha}e^{-nv}dv\right\}\left\{\operatorname*{lub}_{0\leqq t<\infty} |f(tv)|\right\} = M .$$

To prove the sufficiency let us suppose that the conditions are satisfied. If we now set

$$J_n = \int_0^\infty I_{\eta,\alpha}\{(xy)^\beta e^{-xy}\}Q_{n,y}\{F(x)\}dy$$

we have

$$\begin{aligned} J_n &= \frac{1}{\Gamma(n+1+\beta-\alpha)}\int_0^\infty \frac{n}{t^2} I_{\eta,\alpha}\left\{\left(\frac{nx}{t}\right)^\beta e^{-nx/t}\right\} W_n\{F(x)\}dn \\ &= (-)^n \int_0^\infty nt^{n+\beta-1} I_{\eta,\alpha}\left\{\left(\frac{nx}{t}\right)^\beta e^{-nx/t}\right\}\left(\frac{d}{dt}\right)^n \{t^{-\beta}F(t)\}dt . \end{aligned}$$

It will be seen in the course of the arguement that this integral exists. Integrating by parts we have

$$\begin{aligned} J_n &= \frac{(-)^n n}{\Gamma(n+\beta+1-\alpha)}\left[t^{n+\beta-1}I_{\eta,\alpha}\left\{\left(\frac{nn}{t}\right)^\beta e^{-nn/t}\right\}\left(\frac{d}{dt}\right)^{n-1}\{t^{-\beta}F(t)\}\right]_0^\infty \\ &\quad + \frac{(-)^{n-1}n}{\Gamma(n+1+\beta-\alpha)}\int_0^\infty \left(\frac{d}{dt}\right)^{n-1}\{t^{-\beta}F(t)\}\left(\frac{\partial}{\partial t}\right)\{t^{n+\beta-1}I_{\eta,\alpha}\phi\}dt \end{aligned}$$

where

$$\phi \equiv \left(\frac{nx}{t}\right)^\beta e^{-nx/t} .$$

Now

$$
\begin{aligned}
I_{\eta\alpha}\phi &= 0(t^{\eta+1}) \qquad (t \to 0) \\
&= 0(1) \quad \beta = 0(t \to \infty) \\
&= 0(1) \quad \beta > 0(t \to \infty)
\end{aligned}
$$

for [1]

$$
I_{\eta,\alpha}(\phi) = \frac{\Gamma(\beta+\eta+1)}{\Gamma(\alpha+\beta+\eta+1)}\left(\frac{nx}{t}\right)^{\beta} {}_1F_1\left(\beta+\eta+1; \alpha+\beta+\eta+1; -\frac{nx}{t}\right).
$$

Also the hypotheses of the theorem by implications mean that

$$
F(x) = 0(x^{-1})
$$

and in general

$$
F^{(n)}(x) = 0(x^{-n-1})
$$

and

$$
\begin{aligned}
&\left(\frac{d}{dt}\right)^{n-1}[t^{-\beta}F(t)] \\
&\quad = \{(-)^{n-1}\beta(\beta+1)\cdots(\beta+n-2)t^{-\beta-n+1}F(t) + \cdots t^{-\beta}F^{(n-1)}(t)\} .
\end{aligned}
$$

Therefore the integrated part

$$
= 0[t^{\eta+1}\{A_1F(t) + \cdots t^{n-1}F^{(n-)}(t)\}] \to 0 \quad \text{as} \quad t \to 0 .
$$

Also it is

$$
= 0[A_1F(t) + \cdots tF^{(n-1)}(t)] \to 0 \qquad \text{as} \quad t \to \infty .
$$

Therefore the integrated part is zero and integrating by parts again

$$
\begin{aligned}
J_n = &\frac{(-)^{n-1}n}{\Gamma(n+\beta+1-\alpha)}\left[\frac{\partial}{\partial t}(t^{n+\beta-1}I_{\eta\alpha}\phi)\left(\frac{d}{dt}\right)^{n-2}\{t^{-\beta}F(t)\}\right]_0^{\infty} \\
&+ \frac{(-)^{n-2}n}{\Gamma(n+\beta+1-\alpha)}\int_0^{\infty}\left(\frac{d}{dt}\right)^{n-2}\{t^{-\beta}F(t)\}\frac{\partial^2}{\partial t^2}(t^{n+\beta-1}I_{\eta,\alpha}\phi)dt .
\end{aligned}
$$

Now

$$
\left(\frac{\partial}{\partial t}\right)\{t^{\beta+n-1}I_{\eta,\alpha}\phi\} = [(n-1)t^{\beta+n-2}I_{\eta,\alpha}\phi + \cdots + nnt^{\beta+n-3}I_{\eta+1,\alpha(\varphi)}]
$$

and

$$
\begin{aligned}
&\left(\frac{d}{dt}\right)^{n-2}\{t^{-\beta}F(t)\} \\
&\quad = \{(-)^{n-2}\beta(\beta+1)\cdots(\beta+n-3)t^{-\beta-n+2}F(t) + \cdots t^{-\beta}F^{(n-2)}(t)\} .
\end{aligned}
$$

Therefore as before the integrated part again approaches zero when t tends to zero and t tends to infinity. Proceeding in the same manner we obtain

$$J_n = \frac{n}{\Gamma(n+\beta+1-\alpha)}\int_0^\infty t^{-\beta}F(t)\frac{\partial^n}{\partial t^n}\{t^{\beta+n-1}I_{\eta,\alpha}\phi\}dt$$
$$= \frac{n}{\Gamma(n+\beta+1-\alpha)}\int_0^\infty t^{-\beta}F(t)\frac{(nx)^n}{t^{n+1}}t^\beta I_{\eta+n,\alpha}(\phi)dt$$

by the Lemma 4.1. Hence

$$J_n = \frac{n^{n+\beta+1}n^{n+\beta}\Gamma(a)}{\Gamma(n+\beta+1-\alpha)\Gamma(b)}\int_0^\infty t^{-\beta-n-1}{}_1F_1\Big(a;\,b;\,-\frac{nx}{t}\Big)F(t)dt\,.$$

It is clear that this integral exists under the hypotheses of the theorem and therefore all the previous integrals exist. By a simple substitution this gives on using the asymptotic expansion of ${}_1F_1(a;b;x)$ [4]

$$J_n \sim \frac{n^{\beta+n+1}n^{n+\beta}}{\Gamma(n+\beta+1-\alpha)}\int_0^\infty u^{\beta+n-1}e^{-nxu}F\Big(\frac{1}{u}\Big)du\,.$$

Let

$$(1/u)F\Big(\frac{1}{u}\Big) \equiv \psi(u)\,.$$

Now

$$(1/u)F(1/u) = 0(1) \quad (u \to \infty) \quad \text{and} \quad F\Big(\frac{1}{u}\Big) = 0(1) \quad (u \to 0)\,.$$

Hence it is easily seen
(i) $\psi(u) \in L$ $(1/R \leqq t < R)$ for every $R > 1$.
(ii) $\int_1^\infty \psi(u)e^{-cu}du$ converges for any fixed $c > 0$, and
(iii) $\int_0^1 u\psi(u)du$ also converges. Therefore [3]

$$\lim_{n\to\infty} J_n = \frac{1}{u}\psi\Big(\frac{1}{u}\Big) = F(u)\,.$$

Now if

$$\chi(x,y) = \frac{\Gamma(a)}{\Gamma(b)}(xy)^\beta {}_1F_1(a;\,b;\,-xy)\,.$$

Then $\chi(xy) \in L$ in $0 \leqq y < \infty$ under the conditions assumed for the convergence of (1.1). Therefore by a theorem on weak compactness of a set of functions [5] the inequalities in the hypothesis (iii) of the theorem imply the existence of a subset $\{n_i\}$ of the positive integers and a bounded function $f(y)$ such that

$$\lim_{i\to\infty}\int_0^\infty [Q_{n_i,y}\{F(x)\}]\chi(x, y) = \int_0^\infty \chi(x, y)f(y)dy \ .$$

Hence

$$F(x) = \int_0^\infty \chi(x, y)f(y)dy$$

and the theorem is established.

I am indebted to Dr. K. M. Saksena for guidance and help in the preparation of the paper.

References

1. A. Erdelyi, *On some functional transformations*, Rend. del Semin. Matematico. **10** (1950-51), 217-234.
2. H. Kober, *On fractional integrals and derivatives,* Quart. Jour. Math. **11** (1940), 193-211.
3. J. M. C. Joshi, *A singular integral and a real inversion theorem for a generalized Laplace transform,* to appear.
4. L. J. Slater, *Confluent Hypergeometric Functions*, Cambridge University Press, 1960.
5. D. V. Widder, *The Laplace Transform,* Princeton University Press 1946.

EXTREMAL ELEMENTS OF THE CONVEX CONE B_n OF FUNCTIONS

E. K. McLachlan

Let B_0 be the set of nonnegative real continuous on $[0, 1]$, let B_1 be the set of functions belonging to B_0 such that $\Delta_h^1 f(x) = f(x + h) - f(x) \geqq 0$, $h > 0$, for $[x, x + h] \subset [0, 1]$, and let B_n, $n > 1$ be the set of functions belonging to B_{n-1} such that $\Delta_h^n f(x) \geqq 0$ for $[x, x + nh] \subset [0, 1]$ [1]. Since the sum of two functions in B_n belongs to B_n and since a nonnegative real multiple of a B_n function is a B_n function, the set of B_n functions form a convex cone. It is the purpose of this paper to give the extremal elements [2] of this cone, to prove that they are not dense in a compact convex set that does not contain the origin but meets every ray of the cone, and to show that for the functions of the cone an integral representation in terms of extremal elements is possible. The intersection of the B_n cones is the well-known class of functions, the absolutely monotonic functions. Thus the set of these functions form a convex cone also. The extremal elements for this convex cone are given too.

In some correspondence with the author relative to the convex cone B_2, Professor F. F. Bonsall noted that the extremal elements of B_2 were the indefinite integrals of the characteristic functions that are extremal elements of the weak closure of B_1. Professor Bonsall guessed that successive integration would give the extremal elements of B_n. This proved to be a very good guess, and the author gratefully acknowledges the assistance of these comments.

In the following discussion the vertex of the convex cone is not considered as an extremal element.

1. The convex cone B_0. For $f \in B_0$, then take $f_1(x) = x f(x)$ and $f_2 = f - f_1$. Then f is the sum of functions in B_0 that are not proportional to f. Therefore, B_0 has no extremal elements.

2. The convex cone B_1. For $f = c > 0$ and $f = f_1 + f_2$ where f_1 and $f_2 \in B_1$ then $0 = \Delta_h^1 f(x) = \Delta_h^1 f_1(x) + \Delta_h^1 f_2(x)$ implies $\Delta_h^1 f_i(x) = 0$ for $i = 1, 2$ and $[x, x + h] \subset [0, 1]$. Therefore $f_i = c_i$, $c_i > 0$, $i = 1, 2$, where $c_1 + c_2 = c$. Hence f is an extremal element of B_1. Now $f = c > 0$ belongs also to B_n for $n > 1$. The set B_n is a subcone of B_1 and hence $f = c$ is again an extremal element of B_n.

If f is not constant then $f(0) = m$ and $f(1) = M$ and a non-proportional decomposition can be given by taking $f_1(x) = \min(f(x), (1/2)(M + m))$

Received October 23, 1963.

and $f_2 = f - f_1$.

3. The convex cone B_2. The functions of B_2 are exactly the non-negative, nondecreasing and convex functions on $[0, 1]$ [5].

Again the positive constant functions are extremal functions. If $f \in B_2$, f is not constant and $f(0) > 0$ then take $f_1 = f(0)$ and $f_2 = f - f_1$. In so doing f_1 and $f_2 \in B_2$ and f_1 and f_2 are not proportional to f. Since this same technique still can be used for B_n, $n > 2$, the only extremal elements of B_n such that $f(0) > 0$ are the positive constant functions.

If $f(x) = 0$, $x \in [0, \xi]$ and $m(x - \xi)$ for $x \in (\xi, 1]$ where $0 \leqq \xi < 1$ and $m > 0$, then for $f = f_1 + f_2$ it follows that f_1 and f_2 are zero where f is zero and f_1 and f_2 are linear where f is linear. Thus f_1 and f_2 are proportional to f and f is therefore extremal.

If $f(x) = 0$, $x \in [0, \xi_1]$, $m_1(x - \xi_1)$ for $x \in (\xi_1, \xi_2]$, $\cdots$,

$$\sum_{i=1}^{k} m_i(x - \xi_i)$$

for $x \in (\xi_k, 1]$ where $0 < \xi_1 < \xi_2 < \cdots < \xi_k < 1$ and $m_i > 0$ for $i = 1, 2, \cdots, k$, for $k > 1$ then $f \in B_2$. Let $f_1(x) = 0$, for $x \in [0, \xi_1]$, $f_1(x) = m_1(x - \xi_1)$ for $(\xi_1, 1]$ and $f_2 = f - f_1$. Then f_1 and $f_2 \in B_2$ and both are not proportional to f.

Finally, if f is not any of the above functions, but f belongs to B_2, let $\xi_1 = \inf \{x: f(x) > 0\}$. Then $0 \leqq \xi_1 < 1$. On $[\xi_1, 1]$, f is convex, $f(\xi_1) = 0$ and $f(1)$ is finite. Furthermore, the right-hand derivative at ξ_1, $f'_+(\xi_1)$ is finite and in $[\xi_1, 1] f'_-$, the left-hand derivative, must take on more than a finite number of values since f is not polygonal on $[\xi_1, 1]$. Thus there exist ξ_2, $\xi_1 < \xi_2 \leqq 1$ such that on $[\xi_1, \xi_2] f'_+$ is not piecewise linear on three or more non-overlapping segments whose union is $[\xi_1, \xi_2]$ and $f'_-(\xi_2)$ is finite. By Lemma 4 of a paper by the author [4], there exist convex, nonnegative and nondecreasing functions f_1 and f_2 different from f on $[\xi_1, \xi_2]$ such that f_1 and f_2 have the same values and the same derivatives at the end-points as f and $f = \alpha f_1 + (1 - \alpha) f_2$ for some α, $0 < \alpha < 1$. Thus define f_1 and f_2 equal to f on the complement of $[\xi_1, \xi_2]$ relative to $[0, 1]$ and then αf_1 and $(1 - \alpha) f_2$ belong to B_2 and both are not proportional to f.

Thus the extremal elements of B_2 are positive constant functions and those f such that $f(x) = 0$, $x \in [0, \xi]$ and $f(x) = m(x - \xi)$ for $x \in [\xi, 1]$ where $0 \leqq \xi < 1$ and $m > 0$. Designate this latter function by $f(\xi, 1;)$ for $m = 1$.

4. The convex cone B_n, $n > 2$. The function f, such that $f(x) = 0$, $x \in [0, \xi]$, $f(x) = m(x - \xi)^{n-1}$, $x \in [\xi, 1]$, $0 \leqq \xi < 1$ and $m > 0$, that is $m f(\xi, n - 1;)$ belongs to B_n and is an extremal element of B_n.

Already $mf(\xi, 1;)$ belongs to B_2. Now by induction it shall be shown that $mf(\xi, n-1;) \in B_n$ for $n > 2$. In fact, it is true in general that if $f \in B_{n-1}$ and if

$$F(x) = \int_0^x f(t)\,dt\ ,$$

then $F \in B_n$. For if $\Delta_h^k f(x) \geqq 0$ for $k = 0, \cdots, n-1$ then

$$\Delta_h^k F(x) = \Delta_h^{k-1} \int_x^{x+h} f(t)\,dt = \Delta_h^{k-1} f(\xi) > 0$$

where $x < \xi < x - h$ and $k = 0, \cdots, n$. Thus since

$$mf(\xi, n-1; x) = \int_0^x (n-1) mf(\xi, n-2; t)\,dt$$

and since by the induction hypothesis $(n-2)mf(\xi, n-2;) \in B_{n-1}$, it follows that $mf(\xi, n-1;) \in B_n$.

Similarly, by induction it shall be shown that $f = mf(\xi, -1;)$ is an extremal element of B_n It has already been shown that $mf(\xi, 1;)$ is an extremal element of B_{n-1} for any $m > 0$ and for $0 \leqq \xi < 1$. Now let $f = mf(\xi, n-1;) = f_1 + f_2$ where f_1 and f_2 belong to B_n. For $n > 2$, functions in B_n have derivatives, f_1' and f_2' on $[0, 1)$ (See [5] Chapter IV) and the functions f_1' and f_2' belong to B_{n-1} on $[0, \delta]$ for any δ, $0 < \delta < 1$. Take $\delta < 1$ such that $\xi < \delta$, then by the induction hypothesis it follows that f_1' and f_2' are proportional to $f' = (n-1)m$ $f(\xi, n-2;)$ on $[0, \delta]$. Hence $f_i(x) = \lambda_i f(x) + c_i$, $x \in [0, \delta]$, $0 \leqq \lambda_i$, where c_i is a constant for $i = 1, 2$. Since $f_1(0) = f_2(0) = (n-1)m$ $f(\xi, n-2; 0) = 0$ it follows that $c_i = 0$, $i = 1, 2$ and hence f_1 and f_2 are proportional to f on $[0, \delta]$ for any δ, $0 < \delta < 1$. However, since f, f_1 and f_2 are continuous on $[0, 1]$, it follows then that f_1 and f_2 are proportional to f on $[0, 1]$. Therefore, $mf(\xi, n-1;)$ is an extremal element of B_n.

Notice that like the positive constant functions these functions $mf(\xi, n-1;)$ for $\xi = 0$, that is the functions $mf(0, n-1;)$ belong to B_n for all n since its derivatives of all orders exist and are nonnegative on $[0, 1]$. However, if $\xi > 0$, let s and k be integers such that $s > k$ and let x and h be such that $x + (s-2)h = \xi$, $0 \leqq x < x + sh \leqq 1$. Then

$$\Delta_h^s mf(\xi, k; x) = m[(2h)^k - s(h)^k] = mh^k(2^k - s)\ .$$

Hence, if $s > 2^k$, then the expression on the right is negative and thus $mf(\xi, k;) \notin B_s$. This means that whereas $mf(\xi, n-1;) \in B_n$ it does not belong to B_j for $j > 2^{n-1}$.

It remains only to show that the functions of B_n other than the positive constant functions of the form $m f(\xi, k;)$, $0 \leqq \xi < 1$, $m > 0$, $k = 1, 2, \cdots, n-1$ that belong to B_n are not extremal elements of B_n.

It is known that f' exists and is a continuous function on $[0, 1)$. If f' can be extended to be a continuous function on $[0, 1]$, that is, if $\lim f'(x)$ as $x \to 1^-$ exists and is finite, then $f' \in B_{n-1}$. By assuming the induction hypothesis on n, there exist functions g_1 and g_2 belonging to B_{n-1} such that $f' = g_1 + g_2$ and g_1 and g_2 are not proportional to f'. Let $f_i(x) = \int_0^x g_i(t)\,dt$, $i = 1, 2$. Thus f_1 and f_2 belong to B_n and they are not proportional to f. For if $f_1 = \lambda_1 f$, $\lambda_1 \geqq 0$, then $f_1' = \lambda_1 f' = g_1$. This clearly violates what is known about g_1. Hence such a function f is not an extremal element of B_n.

Finally, suppose that $f \in B_n$ and $\lim f'(x) = +\infty$ as $x \to 1^-$. Then the following must be true: $f', f'', \cdots, f^{(n-2)}$ and $f_+^{(n-1)}$, the right-hand derivative of $f^{(n-2)}$ are defined on $[0, 1)$; each of them approaches $+\infty$ as x approaches one from the left; and $\Delta_h^k f^{(j)}(x) \geqq 0$ for $0 \leqq x < 1$, $j = 1, 2, \cdots, n-1$, (with the special understanding for $j = n-1$), $k = 0, 1, 2, \cdots, n-j$. Denote by $B_{n-j}[0, 1)$ the set of real functions ϕ of $[0, 1)$ $\Delta_h^k \phi(x) \geqq 0$, $0 \leqq x < 1$, $k = 0, 1, \cdots, n-j$ for $j = 1, 2, \cdots, n-1$ such that $\phi(x) \to +\infty$ as $x \to 1^-$. The functions $B_{n-j}[0, 1)$ form a convex cone and $f^{(j)} \in B_{n-j}[0, 1)$ for $j = 1, \cdots, n-1$. By an argument similar to the one given earlier, the indefinite integral of a function F in $B_m[0, 1)$ belongs to $B_{m+1}[0, 1)$ if $\int_0^x F(t)\,dt \to +\infty$ as $x \to 1^-$. Also if g, g_1 and $g_2 \in B_m[0, 1)$, $g = g_1 + g_2$, and g_1 and g_2 are not proportional to g, then the indefinite integrals of g_1 and g_2 are not proportional to g. Not that if $g = g_1 + g_2$ as above and if $\int_0^{1-} S(t)\,dt$ is finite, then the same will be true of $\int_0^{1-} g_i(t)\,dt$ for $i = 1, 2$. If the $\lim g(t) = +\infty$ as $t \to 1^-$ and $\int_0^{1-} g_i(t)\,dt = +\infty$ then the same will be true of $\int_0^{1-} g_i(t)\,dt$ for $i = 1, 2$ if there exists constants $\gamma_i > 0$, $i = 1, 2$ such that $g_i(t) \geqq \gamma_i\, g(t)$ for some δ, $0 < \delta < 1$. For the case when $\int_0^{1-} g_i(t)\,dt$ is finite then f_i where $f_i(x) = \int_0^x g_i(t)\,dt$, $i = 1, 2$ can be extended into a function that is continuous on $[0, 1]$. Hence f_1 and f_2 will belong to B_{m+1}.

Thus the object is to find two functions g_1 and g_2 that belong to $B_1[0, 1)$, such that $f_+^{(n-1)} = g_1 + g_2$, g_1 and g_2 are not proportional to $f_+^{(n-1)}$, and such that $g_i(t) \geqq \lambda_i f_+^{(n-1)}(t)$, $\delta \leqq t < 1$, $\delta > 0$. Then f_1 given by

$$f_i(x) = \int_0^x \int_0^{t_{n-2}} \cdots \int_0^{t_2} \int_0^{t_1} g_i(t)\,dt\,dt_1 \cdots dt_{n-2}\,,$$

$i = 1, 2$ belong to B_n and give a nonproportional decomposition of f. The lemma below shows how the functions g_1 and g_2 with the desired properties can be constructed.

LEMMA. *Given f on $[0, 1)$ such that f is right continuous, nonnegative, nondecreasing and $f(x) \rightarrow +\infty$ as $x \rightarrow 1^-$. There exist two functions f_1 and f_2 on $[0, 1)$ that are right continuous, nonnegative and nondecreasing, $f = f_1 + f_2$, f_1 and f_2 are not proportional to f, and $f_i(x) \geqq \gamma_i f(x)$ on $[\delta, 1)$ for some $0 < \delta < 1$ and $\gamma_i > 0$, $i = 1, 2$.*

Proof. All the discontinuities of f must be jump discontinuities. If the point $x = 1$ is an accumulation point of the discontinuities of f, then there exist c_1, c_2 and c_3, $0 < c_1 < c_2 < c_3 < 1$ such that f has a jump of θ_i at c_i, $\theta_i > 0$, $i = 1, 2, 3$. Take $\theta = (1/2) \min(\theta_1 \theta_2 \theta_3)$. Let f_1 be such that $f_1(x) = (1/2)(f(x) - \theta)$, $c_1 \leqq x < c_2$, $f_1(x) = (1/2)(f(x) + \theta)$, $c_2 \leqq x < c_3$ and $f_1(x) = (1/2) f(x)$ otherwise. Take $f_2 = f - f_1$. Then f_1 and f_2 have the required properties.

If the point $x = 1$ is not an accumulation point of the discontinuities then there exists δ, $0 < \delta < 1$ such that f is continuous on $[\delta, 1)$. Let ξ be a point such that $f(\xi) = f(\delta) + 1$, then $\delta \leqq \xi < 1$. Take f_1 such that $f_1(x) = (1/2)f(x)$, $0 \leqq x < \xi$ and $f_1(x) = (1/3)(f(x) - f(\delta) - 1) + (1/2)(f(\delta) + 1)$, $\xi \leqq x < 1$. Let $f_2 = f - f_1$. Then again f_1 and f_2 have the required properties.

5. **Absolutely monotonic functions.** The continuous functions f on $[0, 1]$ such that $f^{(k)}(x) \geqq 0$ for $0 < x < 1$, $k = 0, 1, 2, \cdots$ were called *absolutely monotonic* functions by Bernstein. These functions clearly form a convex cone of functions on $[0, 1]$. Since the functions f belonging to B_n, $n > 2$, have $f^{(k)}(x) \geqq 0$, $k = 0, 1, \cdots, n - 2$, it follows that $\bigcap_{n=0}^{\infty} B_n$ is contained in the set of absolutely monotonic functions. Since the continuous functions f on $[0, 1]$ such that $f^{(k)}(x) \geqq 0$, $k \leqq n$ on $(0, 1)$ have $\Delta_h^k f(x) \geqq 0$ for $k \leqq n$, then $\bigcap_{n=0}^{\infty} B_n$ is the set of absolutely monotonic functions. Denote this set by B_∞

From the earlier remarks it is clear that c_0, $c_1 x$, $c_2 x^2$, $\cdots$ belong to B_∞ for $c_i > 0$, $i = 0, 1, 2, \cdots$ and they are indeed extremal elements of B_∞. Since any $f \in B_\infty$ is absolutely monotonic on $[0, 1)$ it follows that

$$f(x) = \sum_{n=0}^{\infty} f^{(n)}(0)(x^n/n!) , \quad 0 \leqq x < 1 .$$

Consequently, if as many as two terms are nonzero in the series expansion, then take f_1 equal to one of the two nonzero terms and $f_2 = f - f_1$. Then clearly f_1 and f_2 belong to B_∞ and f has a nonproportional decomposition. Hence the only extremal elements of B_∞ are the functions $c_i x^i$, $c_i > 0$, $i = 0, 1, 2, \cdots$.

The following theorem summarizes all of the results up to this point.

THEOREM. *The convex cone B_0 has no extremal elements. The* *functions* $f = c > 0$, *where* c *is a constant, are extremal elements of* B_n, $n = 1, 2, 3, \cdots$. *The function* $m\,f(\xi, n-1; x) = 0$ *for* $0 \leqq x < \xi$ *and* $m(x-\xi)^{n-1}$ *for* $\xi \leqq x \leqq 1$, $m > 0$, $0 \leqq \xi < 1$ *are extremal elements of* B_n, $n = 2, 3, \cdots$. *The only other extremal elements of* B_n, $n = 2, 3, \cdots$ *are those functions* $m\,f(\xi, k;)$, $k = 1, 2, \cdots, n-2$ *that belong to* B_n. *The extremal elements of the convex cone* B_∞, *the absolutely monotonic functions, are the functions of the form* $c_i x^i$, $c_i > 0$, $i = 0, 1, 2, \cdots$.

6. **Integral representations.** The set of functions $B_n - B_n$, $n \geqq 1$, form a linear space containing the convex cone B_n. Using the topology of simple convergence $B_n - B_n$ becomes a locally convex space. Let C_n be the set of functions f of B_n such that $f(1) = 1$. Clearly, C_n meets every ray of C_n once and only once and does not meet the origin in $B_n - B_n$, that is the zero function. Furthermore, C_n is convex. Each function f of C_n is such that $0 \leqq f(x) \leqq 1$ for all $0 \leqq x \leqq 1$ since f is nonnegative and nondecreasing. It follows by use of the Tychonoff theorem that C_n is contained in a compact set in $B_n - B_n$, namely $\{f : f \in B_n - B_n, 0 \leqq f(x) \leqq 1, 0 \leqq x \leqq 1\}$. Thus C_n is compact, if it can be shown that C_n is closed. This will be done by showing the complement of C_n is open.

If $g \in B_n \backslash C_n$ then $g(1) \neq 1$. The set

$$V(1; \varepsilon) + g = \{f : f \in B_n - B_n, |f(1) - g(1)| < \varepsilon\}$$

where $\varepsilon = (1/2)\,|1 - g(1)|$ is an open set about g that fails to meet C_n. If $g \notin B_n$ then there exists x_0, k and h such that $\Delta_h^k g(x_0) = \delta < 0$. Now

$$\Delta_h^k g(x_0) = \sum_{j=0}^{k} (-1)^j \binom{k}{j} g(x_0 + (k-j)\,h)\,.$$

Consider

$$\begin{aligned} V &= V(x_0, x_0 + h, \cdots, x_0 + kh; \varepsilon) + g \\ &= \{f : f \in B_n - B_n, |f(x_0 + jh) - g(x_0 + jh)| < \varepsilon, j = 0, 1, \cdots, k\}\,. \end{aligned}$$

where $\varepsilon = 2^{-(k+1)}(-\delta)$. Then V does not meet C_n since for if $f \in V$

$$\begin{aligned} \Delta_h^k f(x_0) &= \Delta_h^k(f(x_0) - g(x_0)) + \Delta_h^k g(x_0) \\ &< |\Delta_h^k(f(x_0) - g(x_0))| + \Delta_h^k g(x_0) \\ &< \sum_{j=0}^{k} \binom{k}{j} |f(x_0 + (k-j)h) - g(x_0 + (k-j)h)| + \delta \\ &< \varepsilon \sum_{j=0}^{k} \binom{k}{j} + \delta \\ &= \varepsilon\, 2^k + \delta \\ &= (1/2)\,\delta < 0\,. \end{aligned}$$

Hence $f \notin B_n$.

Thus by Theorem 39.4 of Choquet [3], it follows that for any function f_0 in C_n there exists a nonnegative measure μ_0 on the closure of the extreme points of C_n such that $f_0(x)\int d\mu_0 = \int f(x)\,d\mu_0$. Since C_n meets every ray of the cone B_n and does not contain the origin, it follows that each function of B_n is a scalar multiple of such a representation.

If the set of extremal elements of C_n are dense in C_n, then the above result would be of no interest, but this is not the case. Consider $g_0(x) = (1/2) + 2^{n-2} f(1/2, n-1; x)$. Then g_0 belongs to B_n since it is the sum of two functions in B_n. Notice further that $g_0(1) = 1$ and hence $g_0 \in C_n$. The neighborhood of g_0,

$$\begin{aligned} V_0 &= V(0, 1; 1/8) + g_0 \\ &= \{f : f \in B_n - B_n, \, |f(i) - q_0(i)| < (1/8), \, i = 0, 1\}\,, \end{aligned}$$

does not meet any extreme point of C_n. Any positive constant function of C_n is $f(x) = 1$ for all x and hence $f(0) > 5/8$ at $x = 0$. Any function of the form $m\,f(\xi, k;)$ that belongs to zero at $x = 0$ and hence does not belong to V_0.

7. Remarks. Choquet [3] discusses convex cones of functions related to the cones discussed here. The main difference is that the differences, $\Delta_h^k f(x)$, alternate in sign as k takes on successive integral values in the cones that Choquet considered.

REFERENCES

1. F. F. Bonsall, *Semi-alqebras of continuous functions*, Proceedings of the International Symposium on Linear Spaces, (1961), 101–114.
2. N. Bourbaki, *Espaces vectoriels topologiques*, Act. Sci. Ind. no. 1189, Paris, 1953.
3. G. Choquet, *Theory of capacities*, Annales de l'Institut Fourier, **5** (1953 and 1954), 131–2S6.
4. E. K. McLachlan, *Extremal elements of the convex cone of semi-norms*, Pacifie J. Math., **13** (1963), 1335–1342.
5. D. V. Widder, *The Laplace transform*, Princeton Mathematics Series, **6** (1946).

OKLAHOMA STATE UNIVERSITY

CONTRIBUTIONS TO BOOLEAN GEOMETRY OF p-RINGS

ROBERT A. MELTER

1. Introduction. In a paper in this journal [7], J. L. Zemmer proposed two problems relating to the geometry of the Boolean metric space of a p-ring. (A p-ring is a ring R in which $px = 0$ and $x^p = x$ for some positive prime p, and all $x \in R$. The axioms of a p-ring imply its commutativity.) The first problem asked for necessary and sufficient conditions in order that a subset of such a space (hereafter called a p-space) be a metric basis; the second problem was the determination of congruence indices for p-spaces, with respect to the class of Boolean metric spaces. The present paper contains solutions to these questions as well as a brief discussion of certain properties of the group of motions of a p-space, and an introduction to analytic geometry in a p-space. The reader is referred to Zemmer's paper for definitions not contained herein.

2. Metric bases for p-spaces. Let us recall the following definition.

DEFINITION 2.1. A subset S of a Boolean metric space M is called a *metric basis*, if and only if x, y in M and $d(x, s) = d(y, s)$ for all $s \in S$ imply $x = y$.

Let R be a p-space and B its Boolean ring of idempotents. It is well known that B is a subdirect sum of $GF(2)$ [6]. Denote by B^* the complete direct sum of these same rings.

Associate with every subset S of R a subset $\bar{S}$ of B^* defined as follows:

Let $S_{j,k}$ be the subring of B^* consisting of those elements z of B^* having the property

$$z \subseteqq \bigcap_{s \in S} (s - j)^{p-1}(s - k)^{p-1}$$

for $j, k = 0, 1, 2, \cdots, p-1,\ j \neq k$.

Let

$$\bar{S} = \bigcup_{set} S_{j,k}[j < k;\ j, k = 0, 1, 2, \cdots, p-1] .$$

THEOREM 2.1. *Let R be a p-space with Boolean ring of idempotents B. If S is a subset of R then S is a metric basis for R if*

Received August 12, 1963. The contents of this paper formed a part of the author's University of Missouri Doctoral Dissertation, written under the direction of Professor Joseph L. Zemmer.

and only if $\bar{S} \cap B = 0$, *where* $\cap$ *indicates set intersection.*

Proof. A sequence of lemmas will be established, followed by the demonstration of the theorem itself.

LEMMA 2.2. *Let* w, s, b, d *be elements of a* p*-ring such that* $w^2 = w$, *and* $w \subseteqq (s-b)^{p-1} \cap (s-d)^{p-1}$, *then* $(s-dw)^{p-1} = (s-bw)^{p-1}$.

Proof. By the binomial expansion

$$\begin{aligned}(s - dw)^{p-1} \\ &= s^{p-1} - (p-1)s^{p-2}dw + \frac{(p-1)(p-2)}{2}s^{p-3}d^2w^2 + \cdots + d^{p-1}w^{p-1} \\ &= s^{p-1} - (p-1)s^{p-2}dw + \frac{(p-1)(p-2)}{2}d^2ws^{p-3} + \cdots + d^{p-1}w \\ &= w(s-d)^{p-1} - ws^{p-1} + s^{p-1} .\end{aligned}$$

Similarly $(s-bw)^{p-1} = w(s-b)^{p-1} - ws^{p-1} + s^{p-1}$. Hence $(s-dw)^{p-1} - (s-bw)^{p-1} = w[(s-d)^{p-1} - (s-b)^{p-1}]$. But $w \subseteqq (s-b)^{p-1} \cap (s-d)^{p-1}$ implies $w(s-b)^{p-1} = w(s-d)^{p-1} = w$ and hence $w[(s-b)^{p-1} - (s-d)^{p-1}] = w - w = 0$, and thus $(s-dw)^{p-1} = (s-bw)^{p-1}$, which establishes the lemma.

LEMMA 2.3. *Let* x, y, s, f, g *be elements of a* p*-ring such that* $(x-s)^{p-1} = (y-s)^{p-1}$, *and* $(f-g)^{p-1} = 1$, *then* $\overline{(x-f)^{p-1}}\,\overline{(y-g)^{p-1}} \subseteqq (s-f)^{p-1}(s-g)^{p-1}$ *where the bar over an idempotent indicates its complement in the Boolean ring of idempotents.*

Proof. Let

$$\begin{aligned} a &= (x-s)^{p-1} & t &= (y-g)^{p-1} \\ b &= (y-s)^{p-1} & u &= (s-f)^{p-1} \\ r &= (x-f)^{p-1} & v &= (s-g)^{p-1} \end{aligned}$$

and recall that $1 = (f-g)^{p-1}$. By hypothesis $a = b$ and using the fact that the mapping $x \rightarrow x^{p-1}$ is a strong Boolean valuation the following inequalities are obtained:

$$\begin{array}{lll} a \subseteqq r \cup u & b = a \subseteqq t \cup v & 1 \subseteqq u \cup v \\ u \subseteqq r \cup a & v \subseteqq b \cup t = a \cup t & \end{array}$$

but $1 \subseteqq u \cup v$ implies $u \cup v = 1$, or equivalently

$$* \qquad u + v + uv = 1 ,$$

the addition taking place in the Boolean ring of idempotents.

But then,

$$1 = u \cup v \subseteqq r \cup a \cup t = 1$$
$$1 = u \cup v \subseteqq r \cup t \cup v = 1$$
$$1 = u \cup v \subseteqq r \cup t \cup u = 1 .$$

Let $c = (r \cup t)$, then $c \cup u = 1$ and $c \cup v = 1$ or $c + u + uc = 1$ and $c + v + cv = 1$. Adding the two last equalities it follows that $(u + v)(1 + c) = 0$, $(u + v)(1 + r + t + rt) = 0$, or $(u + v)(1 + r)(1 + t) = 0$. But by * $(u + v) = (1 + uv)$ so that $(1 + uv)(1 + r)(1 + t) = 0$, and in turn $(1 + uv)\bar{r}\bar{t} = 0$ or $\bar{r}\bar{t}uv = \bar{r}\bar{t}$. Returning to the original symbols, this is equivalent to

$$\overline{(x - f)^{p-1}}\,\overline{(y - g)^{p-1}} \subseteqq (s - f)^{p-1}(s - g)^{p-1}$$

which establishes the lemma.

LEMMA 2.4. *Let x, y be elements of a p-ring such that $(x - y)^{p-1} \neq 0$. Then elements f, g, can be selected from the summands of the identity, $0, 1, 2, \cdots, p - 1$ such that*

(i) $(f - g)^{p-1} = 1$, *and*

(ii) $\overline{(x - f)^{p-1}}\,\overline{(y - g)^{p-1}} \neq 0$.

Proof. From the hypothesis it is clear that $x \neq y$. If the p-ring is considered as a subring of the ring of all functions on a set X with values in $GF(p)$, then there is some element t_0 of X such that $x(t_0) \neq y(t_0)$. Let f and g correspond to the functions $f(t) \equiv x(t_0)$ for all $t \in X$ and $g(t) \equiv y(t_0)$ for all $t \in X$. It will be shown that f and g satisfy the conditions set forth by the conclusion of the lemma. Clearly f and g are distinct for every t, and hence $(f - g)^{p-1} = 1$. But $(x - f)(t_0) = (y - g)(t_0) = 0$, so that $\overline{(x - f)^{p-1}}(t_0) = \overline{(y - g)^{p-1}}(t_0) = 1$, and $\overline{(x - f)^{p-1}}\,\overline{(y - g)^{p-1}} \neq 0$.

Proof of Theorem 2.1.

Necessity. Suppose S is a metric basis and $\bar{S} \cap B \ni w \neq 0$. Then w is an element of some $S_{j,k}$, say $S_{b,d}$. Consider bw and dw. Since b and d are distinct and at least one is a unit in the p-ring, $bw \neq dw$. But then by Lemma 2.2 $(s - dw)^{p-1} = (s - bw)^{p-1}$, that is bw and dw have the same distances from every element of S contradicting the assertion that S was a metric basis.

Sufficiency. Suppose $\bar{S} \cap B = 0$ and S is not a metric basis. Then there are elements x, y, of R such that $d(x, s) = d(y, s)$ for all $s \in S$, and $x \neq y$. By Lemma 2.4 there are summands of the identity f, g, such that

$$(f-g)^{p-1}=1 \qquad \text{and} \qquad \overline{(x-f)^{p-1}}\,\overline{(y-g)^{p-1}} \neq 0 .$$

But by Lemma 2.3

$$\overline{(x-f)^{p-1}}\,\overline{(y-g)^{p-1}} = w \subseteqq (s-f)^{p-1}(s-g)^{p-1}$$

for all $s \in S$, that is $w \in S_{f,g}$ or $w \in \bar{S}$, so that $0 \neq w \in \bar{S} \cap B$. This contradiction terminates the proof of Theorem 2.1.

An examination of the proof of Theorem 2.1 reveals that the role played by the set of summands of the identity can be taken by any equilateral p-tuple with side 1. Further, if $\bar{S} \cap B = 0$ with respect to a given equilateral p-tuple with side 1, then $\bar{S} \cap B = 0$ with respect to every equilateral p-tuple with side 1.

A restatement of the theorem can be given which exposes its content of a metric characterization of metric bases.

THEOREM 2.5. *Let R be a p-space with distance algebra B. A subset S of R is a metric basis for R if and only if there exists an equilateral p-tuple with side 1, $\{v_1, v_2, \cdots, v_p\}$, such that the distance algebra does not contain a nonzero element w such that $w \subseteqq \bigcap_S d(s, v_i)d(s, v_j)$ $[i \neq j,\ i, j = 1, 2, \cdots, p]$. (The intersection is to be formed in the Boolean completion of the distance algebra).*

The statement of Theorem 2.5 can be somewhat simplified in a p-space for which the distance algebra is a complete Boolean algebra.

THEOREM 2.6. *Let R be a p-space with complete distance algebra B. A subset S of R is a metric basis for R if and only if there exists an equilateral p-tuple with side 1, $\{v_1, v_2, \cdots, v_p\}$, such that $\bigcap_S d(s, v_i)d(s, v_j) = 0$, $i \neq j$.*

A similar result obtains if S is any finite subset of an arbitrary p-space.

THEOREM 2.7. *Let R be a p-space and S a finite subset. Then S is a metric basis for R if and only if there exists an equilateral p-tuple with side 1, $\{v_1, v_2, \cdots, v_p\}$ such that $\bigcap_S d(s, v_i)d(s, v_j) = 0$ $[i \neq j]$.*

A useful algebraic interpretation of Theorem 2.7 is incorporated in the following Theorem 2.8.

THEOREM 2.8. *Let R be a p-space. Consider the p-ring R as a subdirect sum of $GF(p)$, that is as a set of "sequences" with terms* *in* $GF(p)$. *Then if* S *is a finite subset of* R, S *is a metric basis for* R *if and only if the set of* kth *terms of elements of* S *contains at least* $p-1$ *distinct elements of* $GF(p)$, *for every* k.

COROLLARY 1. *A set of* $p-1$ *elements of a* p-*space forms a metric basis if and only if it is equilateral of side* 1.

COROLLARY 2. *A metric basis for a* p-*space contains at least* $p-1$ *elements.*

COROLLARY 3. *Every element of an autometrized Boolean algebra forms a metric basis.*

Corollary 3 was originally discovered by Ellis [1].

Ellis [2] quotes a conjecture due to J. Gaddum that in a metric space any equilateral set containing the maximal number of elements forms a metric base provided the space is complete and convex.

In a p-space the maximal equilateral sets have exactly p-elements. These sets are metric bases if and only if they have side 1, that is that they are maximal with respect both to number of sides and to common distance.

It is interesting to note that in a p-space even though every metric basis must contain at least $p-1$ points, there are *infinite minimal metric bases*, that is infinite metric bases such that no proper subset is also a metric basis. The following example illustrates such a case.

Example 2.1. Let R be a 3-space in which the distance algebra B is the complete direct sum of countably many copies of $GF(2)$. Let S be the set of atoms in B. Then S is a metric basis for R, but no proper subset of S has this property.

We concluded this section with a brief study of superposability properties of metric bases in p-spaces.

It is known that every congruence between two finite subsets of a p-space can be extended to a motion. The following example illustrates that this conclusion cannot be extended to metric bases.

EXAMPLE 2.2. Let $[0, 1)$ be the right open interval on the real line. Let B denote the class of all subsets of $[0, 1)$ that are unions of finitely many right open intervals $[a, b)$, $0 \leqq a \leqq 1$, $0 \leqq b \leqq 1$, where a and b are *rational* numbers. Then B is an atom-free Boolean algebra whose Boolean operations are the usual set operations [4]. Furthermore, B is not a complete Boolean algebra. For example, the set X

of open intervals of the form $[0, a)$ where $a < \frac{\sqrt{2}}{2}$ has no least upper bound.

Represent this Boolean algebra as "sequences" of zeros and ones indexed by the continuum from 0 to 1. Then a typical element of X will appear as follows:

$$\left(1, 1, 1, 1, \cdots 1, \cdot\cdot 0, 0, 0, 0, 0, \cdot\cdot \frac{\sqrt{2}}{2} \cdot\cdot 0, 0, 0, 0, 0, \cdots\right).$$

A typical element of the set X^* of upper bounds of X will appear as

$$\left(1, 1, 1, 1, 1, \cdots\cdots\cdots 1, 1, 1, \cdot\cdot \frac{\sqrt{2}}{2} \cdot\cdot 1, 1, 0, 0, 0, \cdots\right).$$

and a typical element of the set Y of complements of elements of X^* will appear as

$$\left(0, 0, 0, \cdots\cdots\cdots\cdots 0, 0, \cdot\cdot \frac{\sqrt{2}}{2} \cdot\cdot 0, 0, 1, 1, 1, \cdots\right).$$

It is clear that the sets X and Y have the same cardinality since they are both infinite subsets of a countable set.

Let $x \rightarrow f(x)$ be any one-to-one correspondence between X and Y. Zemmer [7] has shown that in a p-space with B as Boolean algebra of idempotents there is a congruence which cannot be extended to a motion, between the sets A and C defined as follows: A contains 0, and for each x in X the element $x + f(x)$. C contains 0, and for each x in X the element $x + 2f(x)$. The congruence F between A and C takes 0 into 0 and $x + f(x)$ into $x + 2f(x)$. It will be shown, moreover, that in the 3-ring with B as Boolean algebra of idempotents the sets A and B are metric bases. Theorem 2.1 can be applied. Since $0 \in A$, it is clear that $\bigcap_{a \in A} d(a, 0)d(a, 2)$ and $\bigcap_{a \in A} d(a, 0)d(a, 1)$ are both equal to zero. However, since for any coordinate less than the $\sqrt{2}/2$th there is a 1 in x for some x in X and for any coordinate greater than the $\sqrt{2}/2$th there is a 1 in some y in Y and since $xy = 0$, $\bigcap_{a \in A} d(a, 1)$ (in the complete direct sum) is the atom with a 1 in the $\sqrt{2}/2$th coordinate, but since B itself is atom free, this implies that there are no elements z of B such that $z \subseteqq \bigcap_{a \in A} d(a, 1)d(a, 2)$ and hence by Theorem 2.1 A is a metric basis. A similar argument shows that C is also a metric basis, which establishes the example.

3. Imbedding and characterization theorems.

DEFINITION 3.1. Let $\{S\}$ be a class of Boolean metric spaces. Then a Boolean metric space R is said to have *congruence indices*

(n, k) with respect to $\{S\}$ provided evey member of $\{S\}$ containing more than $n+k$ distinct points, is congruently imbeddable in R, whenever every n of its poinits are imbeddable in R.

DEFINITION 3.2. A space R is said to have *congruence order* n with respect to $\{S\}$ provided it has congruence indices $(n, 0)$ with respect to $\{S\}$.

(It is understood that the distance algebras of members of the comparison class are isomorphic with the distance algebra of the space R.)

The following series of theorems will establish that a p-space with Boolean algebra of idempotents B where B is a complete direct sum of $GF(2)$ has best congruence order $p+1$ with respect to the class of all Boolean metric spaces (S, B, d). Theorem 3.4 generalizes a theorem due to Ellis [1].

LEMMA 3.1. *If A and B are congruent metric bases for a Boolean metric p-space R and if $f: A \rightarrow B$ is a congruence between the two sets, which can be extended to a motion, then the extension is unique.*

Proof. Suppose f and g are distinct motions which agree on A; then there is an $x \in R$ such that $f(x) \neq g(x)$. But for all $a \in A$,

$$\begin{aligned} d(f(x), f(a)) = d(x, a) &= d(g(x), g(a)) , \\ &= d(g(x), f(a)) , \end{aligned}$$

which contradicts the assumption that B is a metric basis.

LEMMA 3.2. *If A is a metric basis, for a Boolean metric p-space, and A and B are superposable then B is also a metric basis.*

Proof. Let f be a motion which takes A onto B. Suppose B is not a metric basis, then there are elements x, y, of R such that $x \neq y$, and $d(x, b) = d(y, b)$ for all $b \in B$. But then $d(f^{-1}(x), f^{-1}(b)) = d(f^{-1}(y), f^{-1}(b))$ for all $f^{-1}(b)$ in A, and since f^{-1} is, in particular, one-to-one, this contradicts the assertion that A is a metric basis.

COROLLARY. *If A is a finite metric basis for a Boolean metric p-space, and A and B are congruent, then B is also a metric basis.*

Proof. This follows immediately from the lemma and the corollary to Theorem 5 of [7].

If $\{S_1, S_2, \cdots, S_k\}$ and $\{t_1, t_2, \cdots, t_k\}$ are subsets of a Boolean metric space the statement

$S_1, S_2, \cdots, S_k \approx t_1, t_2, \cdots, t_k$ is to indicate that the mapping which takes S_i into t_i $(i = 1, 2, \cdots, k)$ is a congruence.

LEMMA 3.3. *If $\{r'_1, r'_2, \cdots, r'_{p-1}\}$ is a metric basis for a Boolean metric space and*

$$r'''_1, r'''_2, \cdots, r'''_{p-1}, x''' \approx r'_1, r'_2, \cdots, r'_{p-1}, x'$$
$$r'''_1, r'''_2, \cdots, r'''_{p-1}, y''' \approx r'_1, r'_2, \cdots, r'_{p-1}, y'$$

then

$$r'''_1, r'''_2, r'''_3, \cdots, r'''_{p-1}, x''', y''' \approx r'_1, r'_2, r'_3, \cdots, r'_{p-1}, x'y' .$$

Proof. Consider the unique motion which takes

$$\{r'_1, r'_2, \cdots, r'_{p-1}, x'\} \quad \text{into} \quad \{r'''_1, r'''_2, \cdots, r'''_{p-1}, x'''\} .$$

Such a motion exists since by the corollary to Theorem 5 of [7] any congruence between two finite sets can be extended to a motion. If $A \subset B$ and A is a metric basis, then B is also a metric basis. Hence $\{r'_1, r'_2, \cdots, r'_{p-1}, x'\}$ and $\{r'''_1, r'''_2, \cdots, r'''_{p-1}, x'''\}$ are superposable, and by the corollary to Lemma 3.2, $\{r'''_1, r'''_2, \cdots, r'''_{p-1}, x'''\}$ also forms a metric basis and then by Lemma 3.2 the congruence

$$r'''_1, r'''_2, \cdots, r'''_{p-1}, \cdots, x''' \approx r'_1, r'_2, \cdots, r'_{p-1}, x'$$

can be uniquely extended to a motion. Suppose that this motion takes y' into y^* where $y^* \neq y'''$. Then

$$r'''_1, r'''_2, r'''_{p-1}, y''' \approx r'''_1, r'''_2, \cdots, r'''_{p-1}, y^*$$

which contradicts the fact that $\{r'''_1, r'''_2, \cdots, r'''_{p-1}\}$, being congruent to a metric basis are themselves a metric basis by the corollary to Lemma 3.2.

THOREM 3.4. *A Boolean metric space S with distance algebra B is congruently imbeddable in the p-space R with Boolean algebra of idempotents B if*:

(i) *S contains $p-1$ points congruent with a metric basis of R,*
(ii) *Every $p+1$ points of S are congruently imbeddable in R.*

Proof. Let $\{\rho_1, \rho_2, \cdots, \rho_{p-1}\}$ be a $p-1$ tuple of S congruent with $\{r_1, r_2, \cdots, r_{p-1}\}$ a metric basis in R, that is

$$\rho_1, \rho_2, \cdots, \rho_{p-1} \approx r_1, r_2, \cdots, r_{p-1} . \tag{1}$$

Let ρ_p be another point of S. Then there exists $\{r'_1, r'_2, \cdots, r'_{p-1}, r'_p\}$ in S, such that

(2) $$\rho_1, \rho_2, \rho_3, \cdots, \rho_p \approx r_1', r_2', \cdots, r_p'$$

and by the corollary to Lemma 3.2 $\{r_1', r_2', \cdots, r_{p-1}'\}$ is a metric basis.

Let $\zeta \in S$. Then again there exists $\{r_1'', r_2'', \cdots, r_p'', x''\} \in R$ such that

(3) $$\rho_1, \rho_2, \cdots, \rho_p, \zeta \approx r_1'', r_2'', \cdots, r_p'', x''$$

and therefore

(4) $$r_1', r_2', \cdots, r_p' \approx r_1'', r_2'', \cdots, r_p'' .$$

Let x' be the image of x'' under the unique motion which preserves congruence (4). Thus there is defined a single-valued mapping $x' = x'(\zeta)$ of S into R, and

(5) $$\rho_1, \rho_2, \cdots, \rho_p, \zeta \approx r_1', r_2', \cdots, r_p', x' .$$

It remains to show that distances are preserved.

Let $\zeta, \eta \in S$ and let x, y be the corresponding elements in R. Now

(6) $$\rho_1, \rho_2, \cdots, \rho_{p-1}, \zeta, \eta \approx r_1''', r_2''', \cdots, r_{p-1}''', x''', y''' \in R$$

for some $p+1$ tuple $\{r_1''', r_2''', \cdots, r_{p-1}''', x''', y'''\} \in R$. Then using Lemma 3.3, (5), (6) and the fact that $r_1', r_2', \cdots, r_p', y' \approx \rho_1, \rho_2, \cdots, \rho_p, \eta$ it follows that

$$\rho_1, \rho_2, \cdots, \rho_{p-1}, \zeta, \eta \approx r_1''', r_2''', \cdots, r_{p-1}''', x''', y''' \approx r_1', r_2', \cdots, r_{p-1}' x', y'$$

and hence $d(\zeta, \eta) = d(x', y')$.

THEOREM 3.5. *Let S be a Boolean metric space with distance algebra B, then every p-tuple of S is imbeddable in the p-space R with Boolean ring of idempotents B.*

COROLLARY. *Every finite Boolean metric space is imbeddable in a p-space, for some prime p.*

Proof. Let $\{s_1, s_2, \cdots, s_p\}$ be a p-tuple in S. Let q_{ij} denote $d(s_i, s_j)$. Consider the following set of $p-1$-tuples of elements of B:

$$\begin{aligned}
s_1' &= (0, 0, \cdots, 0)\\
s_2' &= (q_{12}, 0, \cdots, 0)\\
s_3' &= (q_{13}\overline{q_{24}}, q_{13}q_{23}, 0, \cdots, 0)\\
s_4' &= (q_{14}\overline{q_{24}}, q_{14}q_{24}\overline{q_{34}}, q_{14}q_{24}q_{34}, 0, \cdots, 0)\\
s_j' &= (q_{1j}\overline{q_{2j}}, q_{1j}q_{2j}\overline{q_{3j}}, q_{1j}q_{2j}q_{3j}\overline{q_{4j}}, \cdots, q_{j}q_{2j}\cdots\overline{q_{j-1,j}}, q_{1j}q_{2j}\cdots q_{j-1,j}, 0, \cdots, 0)\\
s_p' &= (q_{1p}\overline{q_{2p}}, q_{1p}q_{2p}\overline{q_{3p}}, \cdots, q_{1p}q_{2p}\cdots\overline{q_{p-1,p}}, q_{1p}q_{2p}\cdots q_{p-1,p}) .
\end{aligned}$$

It is clear that the s'_i are $p-1$-tuples of pairwise orthogonal elements of B and therefore by Theorem 1 of [7] correspond to elements of R. It remains to show that the mapping $\lambda : s_i \rightarrow s'_i$ is an isometry. Let $q'_{ij} = d(s'_i, s'_j)$.

Consider the rings B and R in their subdirect sum representations. In order to show that λ is an isometry it is sufficient to show that q'_{ij} has a zero in a given component if and only if q_{ij} has a zero in that same component. Let Q_{ij} and Q'_{ij} represent the αth component of q_{ij} and q'_{ij}, respectively. Let $\bar{S}_j$ represent the entry in the αth component of the subdirect sum representation of s'_j.

Assertion. $Q_{ij} = 0$ if and only if $Q'_{ij} = 0$.

It is clear that $Q_{1j} = 0$ if and only if $Q'_{1j} = 0$, $[j = 1, 2, \cdots, p]$. Suppose, therefore, that $i, j \neq 1$.

Suppose that $Q_{ij} = 0$ and assume without loss of generality that i is less than j. Then $\bar{S}_j$ is equal to x where $0 \leqq x \leqq i-1$. But if $\bar{S}_j = x - 1$ where $1 < x < i$ then $Q_{tj} = 1$ for $t = 1, 2, \cdots, x-1$, and $Q_{xj} = 0$ which implies that $Q_{ni} = 1$ for $n = 1, 2, \cdots, x-1$. (For if $Q_{ni} = 0$ for some $n, [n = 1, 2, \cdots, x-1]$ then by the triangle inequality $Q_{ni} = 0$, $Q_{ij} = 0$ imply $Q_{nj} = 0$ which is a contradiction.) But then since $Q_{xj} = 0$, $Q_{ij} = 0$ imply $Q_{xi} = 0$, $\bar{S}_i = x - 1$, and $Q'_{ij} = 0$.

Now, still under the hypothesis that $Q_{ij} = 0$ it remains to show, in order to complete the proof of the necessity of the assertion that if $S_j = 0$, then $\bar{S}_i = 0$. But if $\bar{S}_j = 0$, $Q_{1j} = 0$. (For suppose $\bar{S}_j = 0$ and $Q_{1j} = 1$, then there must be an r, $[r = 2, 3, \cdots, j-1]$ such that $Q_{rj} = 0$. But then by examining the term in s'_j involving Q_{rj} it is seen that there must be a v strictly less than r such that $Q_{vj} = 0$, and proceeding by induction $Q_{1j} = 0$, contrary to hypothesis). But $Q_{1j} = 0$ and $Q_{ij} = 0$ imply by the triangle inequality that $Q_{1i} = 0$ and hence $\bar{S}_i = 0$ which completes the proof of the necessity of the assertion.

To demonstrate the sufficiency of the assertion it must be shown that if $Q'_{ij} = 0$, then $Q_{ij} = 0$.

If $Q_{ij} = 0$, then $\bar{S}_i = \bar{S}_j = x$, where x is an integer mod p. Assume without loss of generality that $i < j$ and suppose $x \neq 0$, $i - 1$. Then $Q_{x-1,j} = 0$, $Q_{x-1,i} = 0$ which together imply that $Q_{ij} = 0$. If $x = i - 1$, it is clear from examining the term in S_j involving $\overline{Q_{ij}}$ that $Q_{ij} = 0$, and lastly if $x = 0$, $Q_{1j} = 0$, and $Q_{1i} = 0$; hence by the triangle inequality $Q_{ij} = 0$. This completes the proof of the theorem.

To clarify the proof, it seems worthwhile to establish the theorem without using the subdirect sum formulation, in a particular instance. Thus let $\{s_1, s_2, s_3\}$ be a Boolean metric triple. Then

$$\begin{aligned} s_1' &= (0, 0)\ , \\ s_2' &= (q_{12}, 0)\ , \\ s_3' &= (q_{13}\overline{q_{23}}, q_{13}q_{23})\ . \end{aligned}$$

Since the sum of the coordinates in a Boolean vector representation is the distance from the origin it is clear that $q_{12} = q_{12}'$. By the same token

$$q_{13}' = q_{13}(q_{23} + \overline{q_{23}}) = q_{13}\ .$$

Lastly $q_{23}' = d(s_3' - s_2', 0)$. The Boolean vector representation of $s_3' - s_2'$ is (a_1, a_2) where

$$\begin{aligned} a_1 &= q_{12}q_{13}q_{23} + q_{13}\overline{q_{23}}\,\overline{q_{12}} \\ a_2 &= q_{12}q_{13}q_{23} + q_{12}\overline{q_{12}\overline{q_{23}} + q_{13}q_{23}} \end{aligned}$$

so that $q_{23}' = a_1 + a_2$, which upon simplification gives

$$\begin{aligned} q_{23}' &= q_{12} + q_{13} + q_{12}q_{13}q_{23} \\ &= q_{23} \end{aligned}$$

since in any Boolean metric space the product of the lengths of the sides of a triangle is equal to their sum.

Before indicating the procedure for imbedding $p + 1$-tuples, a definition of a chain of integers and some lemmas concerning these chains will be presented.

DEFINITION 3.3. Let i, j be positive integers such that $i \leqq j$. An (i, j) chain is any finite sequence of positive integers such that

(1) The sequence has exactly j terms,

(2) The first element in the sequence is 1, and the last is i,

(3) The terms in the sequence are selected from the integers $1, 2, \cdots, j$,

(4) If r and s are integers which occur in the sequence and r is less than s, then the first occurrence of r precedes the first occurrence of s. Every integer between r and s must occur if r and s occur.

Let $x_1, x_2, \cdots, x_j$ be an (i, j) chain. Define a metric on this chain by letting $d(x_a, x_b) = r_{ab} = 1$ if $x_a \neq x_b$ and $d(x_a, x_b) = r_{ab} = 0$ if $x_a = x_b$.

LEMMA 3.6. *Let $s_1, s_2, \cdots, s_v$ be a v-tuple in a Boolean metric space. Let $t_{ij} = d(s_i, s_j)$ and let T_{ij} denote the αth component in the subdirect sum representation of t_{ij}. Then there exists a unique (i, v) chain Γ such that $r_{ab} = T_{ab}$, $a, b = 1, 2, \cdots, v$.*

Proof. By induction on v. For $v = 1$ the theorem is trivially satisfied. Suppose then that $\{s_1, s_2, \cdots, s_k\}$ is a Boolean metric k tuple and $x_1, x_2, \cdots, x_k$ is the unique chain such that $r_{ab} = T_{ab}$. If $T_{w,k+1} = 1$ for $w = 1, 2, \cdots, k$, let x_{k+1} be the next integer not already used in the chain. This integer is uniquely determined and $r_{ab} = T_{ab}$ $a, b = 1, 2, \cdots, k+1$. On the other hand if $T_{\bar{w},k+1} = 0$ where $1 \leqq \bar{w} \leqq k+1$, let $x_{k+1} = x_{\bar{w}}$. x_{k+1} is uniquely determined, for if $T_{\bar{w},k+1} = 0$ and $T_{\bar{\bar{w}},k+1} = 0$, then by the triangle inequality $T_{\bar{w},\bar{\bar{w}}} = 0$ and so $x_{\bar{w}} = x_{\bar{\bar{w}}} = x_{k+1}$ and hence $r_{\bar{\bar{w}},k+1} = 0$. If $r_{a,k+1} = 0$, then $x_a = x_{k+1} = x_{\bar{w}}$, hence, $T_{a,\bar{w}} = 0$, which with the hypothesis $T_{\bar{w},k+1} = 0$, yields $T_{a,k+1} = 0$, which completes the proof.

DEFINITION 3.4. Let $\{p_1, p_2, \cdots, p_k\}$ be a finite subset of a Boolean metric space. Then the *distance product* of this subset is defined to be

$$\prod_{i \neq j} d(p_i, p_j) .$$

THEOREM 3.7. *Let S be a Boolean metric space with distance algebra B. A $p+1$-tuple K of S is imbeddable in the p-space R with Boolean ring of idempotents B if and only if the distance product of K is zero.*

Proof. The necessity is easily established. Let $\{t_1, t_2, \cdots, t_{p+1}\}$ be points of a p-space. In the αth component of the subdirect sum representation, each of the t_r must contain one of the elements of $GF(p)$. Thus in this αth component, for some c, d, t_c and t_d have the same element of $GF(p)$, and hence the distance product has a zero in the αth component. Since this is true for every α, the distance product of $\{t_1, t_2, \cdots, t_{p+1}\}$ is zero.

To establish the sufficiency of the condition, let $\{s_1, s_2, \cdots, s_j\}$ be a Boolean metric j-tuple and let C_{ij} be an arbitrary (i, j) chain. Denote by q_{ab} the distance $d(s_a, s_b)$ and let $C^*_{i\,j}$ be the product

$$\prod_{a,b \leqq j} g(q_{ab})$$

where $g(q_{ab}) = \overline{q_{ab}}$, if the ath and bth terms in C_{ij} are identical; $g(q_{ab}) = q_{ab}$, if the ath and bth terms in C_{ij} differ. Let $\{s_1, s_2, s_3, \cdots, s_{p+1}\}$ be a Boolean metric $p+1$ tuple with distance product zero. Define a set of $p-1$-tuples of B as follows:

$$t_J = (t_J^1, t_J^2, \cdots, t_J^I, \cdots, t_J^{P-1}) \qquad (J = 1, 2, \cdots, p+1)$$

where t_J^I is equal to zero if $I > J-1$, otherwise t_J^I is the Boolean algebra union of all the elements of B of the form $C^*_{I+1,J}$.

Let T_J^I denote the αth component in the subdirect sum representation of t_J^I.

In order to show that the mapping $s_J \rightarrow t_J$ is a mapping into a p-ring, it is sufficient to establish that $T_J^n T_J^m = 0$, if $n \neq m$. But this follows at once from the fact that $T_J^n = 1$ if and only if there is an $(n+1, J)$ chain $x_1, x_2, \cdots, x_J$ such that the αth component of $d(s_a, s_b)$ is equal to $d(x_a, x_b)$ $[a, b = 1, 2, \cdots, J]$, for it follows from Lemma 3.6 that two (i, j) chains are isometric if and only if they are identical.

Since $T_J^I = 1$ if and only if t_J has an I in the αth component and also if and only if $\{s_1, s_2, \cdots, s_J\}$ is such that for a unique $(I+1, J)$ chain $y_1, y_2, \cdots, y_J$, $d(y_a, y_b)$ is equal to the αth component of $d(s_a, s_b)$, $(a, b = 1, 2, \cdots, J)$, it follows that $\{R_1 + 1, R_2 + 1, \cdots, R_{p+1} + 1\}$, where R_k is the αth component of t_k, is the unique chain such that $d(R_m + 1, R_n + 1)$ is equal to the αth component of $d(s_m, s_n)$ $(m, n = 1, 2, \cdots, p+1)$ and hence the αth component of $d(t_a, t_b) = 0$ if and only if the αth component of $d(s_a, s_b) = 0$. This completes the proof of the theorem.

Recall that if B is a Boolean ring, B^* designates the complete direct sum of those $GF(2)$ used in the subdirect sum representation of B.

LEMMA 3.8. *Let S be Boolean metric space with distance algebra B, in which the distance product of every $p+1$ paints is zero. Then S is congruent with a subset of a Boolean metric space S^* with distance algebra B^*, such that B is isomorphic with a subalgebra of B^*, the distance product of every $p+1$ points of S^* is zero, and S^* contains an equilateral $p-1$-tuple of side* 1.

Proof. Let $\{t_1, t_2, \cdots, t_n\}$ be a maximal equilateral set of side 1 in S. If $n \geqq p-1$, no further proof is needed. If $n < p-1$, consider B in its subdirect sum representation and let B^* be the complete direct sum of the $GF(2)$ used to represent B. Let S^* be the set union of S and an element σ. Define a distance d' in S^* as follows: if $x, y \in S$, $d'(x, y) = d(x, y)$, $d'(\sigma, \sigma) = 0$. For $x \in s$, define $d'(x, \sigma) = q'_{x\sigma}$ by giving its αth component $Q'_{x\sigma}$ as follows: If for all $w \in S$, the αth component of $d(w, t_i) = 0$ for some $i = 1, 2, \cdots, n$ then $Q'_{x\sigma} = 1$ for all $x \in S$. If there is a w_α such that the αth component of $d(w_\alpha, t_i) = 1$ for all $i = 1, 2, \cdots, n$, then let $Q'_{x\sigma} = 0$ if and only if $d(x, w_\alpha)$ has a zero in the αth component.

To show that S^* is a Boolean metric space, observe that it is clear that if r, s are elements of S^*, with $r = s$, then $d'(r, s) = 0$. If $d'(r, s) = 0$, it is evident that $r = s$ if r and s are both elements of S. Suppose then that $d'(x, \sigma) = 0$ where $x \in S$. But then in the αth component $d(x, w_\alpha)$ has a zero, where w_α is such that the αth component of $d(w_\alpha, t_i) = 1$ for $i = 1, 2, \cdots, n$, by the triangle inequality. Since this is true for every α, $\{t_1, t_2, \cdots, t_n, x\}$ is an equilateral set of side 1, contrary to hypothesis. The symmetry of d' follows at once from its definition. For the triangle inequality the only triples which need be studied are those of the form (x, y, σ). But, referring now to the αth component, if $d(x, y) = 0$, $d(y, \sigma) = 0$ then $d(y, w_\alpha) = 0$, hence $d(x, w_\alpha) = 0$ and $d(x, \sigma) = 0$ and if $d(x, \sigma) = 0$, $d(y, \sigma) = 0$ then $d(x, w_\alpha) = 0$, $d(y, w_\alpha) = 0$, and $d(x, y) = 0$. In all other cases $d(x, y)$, $d(x, \sigma)$, $d(y, \sigma)$ clearly form a metric triple, because x, y, σ, is a Boolean metric triple unless in some component two of $d(x, \sigma)$, $d(y, \sigma)$, $d(x, y)$ are equal to zero and the third is equal to one.

To show that $\{t_1, t_2, \cdots, t_n\}$ form an equilateral set of side 1, suppose this is not the case, then in some αth component, for some i, $d(\sigma, t_i) = 0$, but then $d(\sigma, w_\alpha) = 0$, hence $d(w_\alpha, t_i) = 0$, contrary to the definition of w_α.

In verifying that the distance product of every $p + 1$ points of S^* is zero, it is sufficient to consider $p + 1$ tuples $\{r_1, r_2, \cdots, r_p, \sigma\}$, $[r_i \in S]$ where in some αth component, the distance products of the r's is one. But if the αth component of $d'(r_i, \sigma)$ is one for $i = 1, 2, \cdots, p$, then either there is for every i, a j, $[j = 1, 2, \cdots, n]$ where $n < p - 1$, such that $d'(r_i, t_j)$ has a zero in the αth component (which implies that for some i, j, k, $d'(r_i, t_k)$, $d'(r_j, t_k)$ have zeros in the αth component and so $d'(r_i, r_j)$ has a zero in the αth component, contrary to hypothesis). On the other hand, if there exists a w_α such that in the αth component $d'(w_\alpha, t_j) = 1$ for all j, $[j = 1, 2, \cdots, n]$, and $d'(\sigma, r_i)$ has a 1 in the αth component, then $d'(w_\alpha, r_i)$ has a 1 in the αth component. But then $\{r_1, r_2, \cdots, r_p, w_\alpha\}$ is a $p + 1$ tuple in S with distance product different from zero.

Continuing in this manner a space containing an equilateral $p - 1$ tuple of side 1 is obtained.

THEOREM 3.9. *Let S be a Boolean metric space with distance algebra B and let R^* be the p-space with Boolean ring of idempotents B^*. The space S is congruently imbeddable in R^* if and only if the distance product of every $p + 1$ points of S is equal to zero.*

Proof. By hypothesis the distance product of every $p + 1$ points of S is zero. Then by Lemma 3.8, S is congruently contained in a Boolean metric space S^*, with distance algebra B^*, containing an equilateral $p - 1$ tuple of side 1, and in which the distance product of every $p + 1$ points is zero. By Lemma 3.7, every $p + 1$ points of S are imbeddable in R^*, and by Theorem 3.4, S^* is congruently imbeddable in R^*, and hence S is congruently imbeddable in R^*. This establishes the sufficiency of the condition and the necessity follows immediately from Theorem 3.7.

COROLLARY 1. *S is congruently imbeddable in R^* if and only if every $p+1$ points of S are congruently imbeddable in the p-space R, with Boolean ring of idempotents B.*

COROLLARY 2. *R^* has congruence order $p+1$ with respect to the class of all Boolean metric spaces (S, B^*, d).*

LEMMA 3.10. *A p-space does not have congruence order p.*

Proof. Let M be a Boolean metric space of any cardinality in which the distance of every two distinct points is one. Then M has every p points imbeddable in a given p-space, but M itself need not be.

THEOREM 3.11. *A p-space R^*, with distance algebra B^* has best congruence order $p+1$ with respect to the class of all Boolean metric spaces.*

Proof. By Corollary 2 of Theorem 3.9 the best congruence order of R^* is less than or equal to $p+1$, but by Lemma 3.10 the congruence order is greater than p.

Another topic of interest in distance geometry is psuedo sets.

DEFINITION 3.5. A $p+1$ tuple T in a Boolean metric space is said to be a *pseudo-p-space* $p+1$ *tuple* if every p points of T are imbeddable in a p-space but T is not.

THEOREM 3.12. *A Boolean metric $p+1$ tuple is either imbeddable in a p-space or is a pseudo-p-space $p+1$ tuple.*

Theorem 3.9 gives a solution to the congruent imbedding problem of determining necessary and sufficient conditions in order that a Boolean metric space be isometric with a subspace of a p-space. In order to obtain a characterization of Boolean metric spaces themselves one method is to first categorize those subspaces of a given p-space which are themselves p-spaces among the class of all subspaces of the p-space. This is accomplished in the following two theorems.

THEOREM 3.13. *Let R be a Boolean metric p-space with distance algebra B. Let S be a subspace of R. Then a necessary and sufficient condition that S be a p-space is that:*

(1) *There exists a subalgebra $\bar{B}$ of B such that S contains an equilateral $p-1$ tuple with side 1 of $\bar{B}:\{t_1, t_2, \cdots, t_{p-1}\}$,*

(2) *There is a one-to-one correspondence between the elements of* S, *and the set of pairwise orthogonal* $p-1$ *tuples*: $\{c_1, c_2, \cdots, c_{p-1}\}$ *of elements of* $\bar{B}$, *such that for* $x \in S$, $d(x, t_i) = \overline{c_i}$.

Proof. The necessity is clear, since for any sub-p-space, $\{t_1, t_2, \cdots, t_{p-1}\}$ can be taken as summands of the identity and the c_i are then the "coordinates" in a Boolean vector representation.

Sufficiency. If the conditions of the theorem are satisfied the set of $p-1$ tuples of c's form a p-ring, which is a subring of the original ring.

THEOREM 3.14. *Let* S *be a Boolean metric space with distance algebra* B. *A necssary and sufficient condition that* S *be a* p-*space is that*:

(1) *The distance product of every* $p+1$ *points of* S *is zero and for some subalgebra* $\bar{B}$ *of* B

(2) S *contains an equilateral* $p-1$ *tuple of side* 1 *in* $\bar{B}$

(3) *There is a one-to-one correspondence between the elements of* S, *and the set of pairwise orthogonal* $p-1$ *tuples*: $\{c_1, c_2, \cdots, c_{p-1}\}$ *of elements of* $\bar{B}$, *such that for* $x \in S$, $d(x, t_i) = \overline{c_i}$.

Proof. By Theorem 3.9, S is a subspace of a p-space, but by Theorem 3.13, S is then a p-space.

4. Properties of the group of motions. This section is devoted to developing certain properties of the group of motion of p-spaces.

THEOREM 4.1. *In a* p-*space every rotation about the origin is a product of a finite number of involutions.*

Proof. Let R be a p-space and B its distance algebra. Let $x \rightarrow f(x)$ be a rotation about the origin on R, and M the matrix corresponding to f. Then $M = (a_{ij})$ is a $(p-1) \times (p-1)$ matrix with elements in B satisfying $a_{ik}a_{ij} = 0$, $j \neq k$, and $a_{ij}a_{kj} = 0$, $i \neq k$, and $MM' \neq I$, where $a_{ij} \in B$.

For $b \in B$, denote by b^r, the rth component of b in the subdirect sum representation of B, and define $M^{(r)} = (a^r_{ij})$.

Then the set $\{M^{(r)}\}$, $r \in \mathscr{R}$, consists of at most $(p-1)!$ different matrices each of which is a permutation matrix. Clearly

$$M^{(r)} = M_1^{(r)} \cdot M_2^{(r)} \cdot \cdots M_{p-2}^{(r)}$$

where the elements on the right are transposition matrices.

Whence $M^{(r)}$ can be transformed into $M_k^{(r)}$ by a certain permutation of its columns.

Let M_{rk} be the matrix which results from applying these same column operations to M.

Let Z_r be the product of those elements in M corresponding to the 1's in $M^{(r)}$. Let $Z_r^{(s)}$ be the sth component of Z_r, and note that $Z_r^{(s)} = 1$ if and only if $M^{(r)} = M^{(s)}$.

Let M_{rk}^* be the matrix obtained from M_{rk} by multiplying every element by z_r and then adding $\bar{Z}_r$ to the elements along the main diagonal, i.e., $M_{rk}^* = Z_r M_{rk} + \bar{Z}_r I$.

Denote the matrix of tth components of M_{rk}^* by $M_{rk}^{*(t)}$.

It follows that:

$$\begin{aligned} M_{rk}^{*(t)} &= M_k^{(r)} \qquad \text{if} \quad M^{(r)} = M^{(t)} \\ M_{rk}^{*(t)} &= I \qquad \text{if} \quad M^r \neq M^t . \end{aligned}$$

(From the definition of M_{rk}^*, if $M^{(r)} = M^{(t)}$, $M_{rk}^{*(t)} = M_{rk}^{(t)}$, and from the definition of M_{rk}, $M_{rk}^{(t)} = M_{rk}^{(r)}$, which is equal to $M_k^{(r)}$ by the definition of $M_k^{(r)}$, that is, $M_{rk}^{*(t)} = M_k^{(r)}$).

Thus

$$\begin{aligned} \prod_k M_{rk}^{*(t)} &= M^{(r)} \qquad \text{if} \quad M^{(r)} = M^{(t)} \\ \prod_k M_{rk}^{*(t)} &= I \qquad \text{if} \quad M^{(r)} \neq M^{(t)} \qquad (k = 1, 2, \cdots, p-2) . \end{aligned}$$

Now select a minimal set of r's, $L = (r_1, r_2, \cdots, r_m)$ such that each $M^{(r)} = M^{r_j}$ for some $r_j \in L$. Then

$$M = \prod M_{r_j k}^* \qquad (k = 1, 2, \cdots, p-2, r_j \in L) .$$

To show this, observe that

$$M^{(\lambda)} = \prod M_{r_j k}^{*(\lambda)} \qquad (k = 1, 2, \cdots, p-2, r_j \in L) .$$

Let $r_\beta \in L$ be such that $M^{(\lambda)} = M^{(r_\beta)}$. Then

$$\begin{aligned} \prod M_{r_j k}^{*(\lambda)} &= (\prod M_{r_\beta k}^{*(\lambda)})(\prod_{j \neq \beta} M_{r_j k}^{*(\lambda)}) \\ &= M^{(r_\beta)} \cdot I \\ &= M^{(r_\beta)} = M^{(\lambda)} . \end{aligned}$$

COROLLARY. *Every motion which leaves zero fixed in a* 3-*space is a reflection. Every reflection in a* 3-*space therefore has determinant equal to* -1.

The proof of Theorem 4.1. suggests that there is a close relationship between the group of motions of a p-space, and permutation groups. Indeed it is the case that the group of motions is a subgroup of the direct product of permutation groups on $p-1$ letters. This will be made precise in the following two theorems.

DEFINITION 4.1. Let B be a Boolean ring. Consider B as a subdirect sum of $GF(2)$. Let φ be a group of permutations on p-symbols and G_φ the full direct product of φ of the same cardinality and number of summands as B. For $b \in B$, and $P \in \varphi$, let $g(P, b)$ be the element of G_φ, which effects the permutation P where b has 1's and the identity permutation elsewhere. Denote by $G_\varphi(B)$ the subgroup of G_φ generated by the set of elements $g(P, b)$, $P \in \varphi$, $b \in B$.

THEOREM 4.2. *Let R be a p-space with Boolean ring of idempotents B. Then the group of motions of R which leave zero fixed is $G_T(B)$ where T is the symmetric group on $p-1$ symbols.*

Proof. Let M be a motion matrix for R. In the proof of Theorem 4.1 it was shown that M can be written as a product of matrices $M^*_{r_jk}$, but these matrices correspond to motions of the form $g(t, b)$ where t is a transposition.

COROLLARY. *Let R be a p-space. Then the group of motions of R is $G_S(B)$, where S is the group of permutations on p symbols.*

Proof. Let $f(x)$ be a motion, then $f(x) = xM + b$. It has been shown in the theorem that the rotation is an element of $G_T(B)$ and hence of $G_S(B)$. Consider now the translation $t(x) = x + t$. It can be written as the product of translations as $t_1(x) \cdot t_2(x) \cdot \cdots, \cdot t_{p-1}(x)$ where $t_i(x) = x + i(1-(t-i))^{p-1}$ which are elements of $G_S(B)$.

On the other hand it must be shown that every element of G_SB is a motion. It suffices to show that every $g(P, b)$ is a motion. Thus let $g(P, b)$ be given. If P fixes zero, the result follows from the theorem. If P does not fix zero, let $0'$ be the image of zero under P. Consider the permutation q: $x \rightarrow x - 0'$ of the integers mod p. Then $g(pq, b)$ is a motion and has a matrix M, and $f(x) = xM + 0'b$ corresponds to $g(p, b)$.

THEOREM 4.3. *Let R be a 3-space with Boolean ring of idempotents B. Then every motion f on R which leaves 0 fixed is of the form $f(x) = ax$ where a is a unit in the 3-ring.*

Proof. It follows from Theorem 4 of [7] that $f(x) = xM$ where $M = (a_{ij})$ $i, j = 1, 2$, and $a_{ij} \in B$. Further

$$M = \begin{pmatrix} a & 1+a \\ 1+a & a \end{pmatrix}$$

Suppose then that $x = (x_1, x_2)$, and so

$$\begin{aligned} f(x) &= (ax_1 + (1 + a)x_2, (1 + a)x_1 + ax_2) \\ &= (x_1, x_2)\cdot(a, 1 + a) \end{aligned}$$

where $(a, 1 + a)(a, 1 + a) = (1, 0) = 1$.

5. **Analytic geometry in p-spaces.** If a rectangular coordinate system is introduced in a Euclidean plane E, a point P can be represented as a pair (x, y) of real numbers. One then seeks to describe geometrically the loci of equations of the form $y = f(x)$, and conversely, given a geometric description of a plane set, to find the equation of which it is the corresponding locus. But a point P in the Euclidean plane may also be considered to be represented by the single complex number $z = x + iy$. Here the question is not so much the investigation of the loci of equations of the from $f(z) = 0$; a study is rather made of the way in which geometric properties change or remain invariant under transformations $w = f(z)$ of the plane into itself. It is the purpose of the following remarks to exhibit theorems which illustrate that an analytic geometry for p-spaces may be developed in a manner analogous with both of the methods discussed above for Euclidean plane geometry.

Suppose, therefore, that R is a p-space. Since the elements of the p-ring R are in one-to-one correspondence with the points of the p-space R, every function $f(x)$ defined for all x in the p-ring R and having values in the p-ring R induces a mapping of the p-space R into itself. This mapping need not of course preserve distances, and in general will not even be one-to-one. Theorem 5.2 establishes necessary and sufficient conditions that a polynomial function defined on a p-ring R induce a motion on the corresponding p-space.

The following theorem, which was first established in 1882 is needed for the proof.

THEOREM 5.1. *Raussnitz* [6]. *Let* $f(x) = a_{-1}x^{p-1} + a_0x^{p-2} + a_1x^{p-3} + \cdots + a_{p-2}$ *be a polynomial where* $a_i \in GF(p)$, $(i = -1, 0, 1, 2, \cdots, p-1)$. *Then a necessary and sufficient condition that* $f(0), f(1), \cdots, f(p-1)$ *be distinct is that* (i) *the determinant* $R(k)$ *be equal to zero for* $k = 0, 1, \cdots, a_{p-2} - 1, a_{p-2} + 1, \cdots, p - 1$ *where*

$$R(k) = \begin{vmatrix} a_0 & a_1 & a_2 & \cdots & a_{p-3} & a_{p-2} - k \\ a_1 & a_2 & a_2 & & a_{p-2} - k & a_0 \\ \cdots\cdots & & & & & \\ \cdots\cdots & & & & & \\ a_{p-2} - k & a_0 & a_1 & & a_{p-4} & a_{p-3} \end{vmatrix}$$

and (ii) $a_{-1} = 0$.

THEOREM 5.2. *Let R be a p-space. Then a necessary and sufficient condition that the polynomial*

$$P(x) = a_{-1}x^{p-1} + a_0x^{p-2} + a_1x^{p-3} + \cdots + a_{p-2}\,,$$

where the a_i $(i = -1, 0, 1, \cdots, p-1)$ are elements of the p-ring R, induce a motion on the p-space R is that

(i) $a_{-1} = 0$

and

(ii) $\bar{R}(c_k) = 0$ $(k = 0, 1, 2, \cdots, p-1)$ *where*

$$\bar{R}(c_k) = \begin{vmatrix} a_0 & a_1 & a_2 & \cdots & a_{p-3} & a_{p-2} - c_k \\ a_1 & a_2 & a_3 & & a_{p-2} - c_k & a_0 \\ \cdots\cdots & & & & & \\ \cdots\cdots & & & & & \\ a_{p-2} - c_k & a_0 & a_1 & & a_{p-4} & a_{p-3} \end{vmatrix}$$

and $c_k = -(a_{p-2} - k)^{p-1} + k + 1$. (Note that $R(k)$ has integer arguments whereas the arguments of $\bar{R}(c_k)$ are elements of a p-ring).

Proof. Suppose that the polynomial $P(x)$ corresponds to a motion M on the p-space R, and consider the p-ring R as a subdirect sum of $GF(p)$. Then the elements of R may be represented as $(r_1, r_2, \cdots, r_t, \cdots)$ where the $r_t \in GF(p)$. Clearly M induces a permutation p_t on the components r_t, *f*or every t. If for $x_1 \in R$ and $x_2 \in R$, $r_t^1 = r_t^2$ and $M(r_t^1) \neq M(r_t^2)$, then $d(x_1, x_2)$ will have a zero in its tth component while $d(M(x_1), M(x_2)$ will have a one in the tth component contradicting the assumption that M is distance preserving. The uniqueness of p_t is a consequence of the fact that the motion M is a well defined mapping. Let $a_{j,t}$ be the tth component of a_j in the subdirect sum representation of R. Then the polynomial

$$P_t(x) = a_{-1,t}x^{p-1} + a_{0,t}x^{p-2} + a_{1,t}x^{p-3} + \cdots + a_{p-2,t}$$

must represent the permutation p_t on the elements of $GF(p)$. Hence by Theorem 5.1, $a_{-1,t} = 0$, for all t, so that $a_{-1} = 0$. Also, $R(k_t) = 0$ for all t, and $k = 0, 1, 2, \cdots, a_{p-2} - 1, a_{p-2,t} + 1, \cdots, p-1$. Notice however that $c_{k,t}$ ranges over $0, 1, 2, \cdots, a_{p-2,t} - 1,, a_{p-2,t} + 1, \cdots, p-1$ as k takes on the values $0, 1, \cdots, p-1$. Thus $\bar{R}(c_k) = 0$, for $k = 0, 1, 2, \cdots, p-1$, and the necessity of conditions (i) and (ii) is established.

On the other hand, suppose that conditions (i) and (ii) are satisfied by $P(x)$. It will first be shown that the polynomial $P^*(x)$ where

$P^*(x) = P(x) - a_{p-2}$ also satisfies conditions (i) and (ii). For each t, the polynomial $P_t(x)$ satisfies the conditions of Theorem 5.1 and hence $P(x)$ induces a permutation p_t on the tth component of the subdirect sum representation of the p-ring R. But in each component $P^*(x)$ also induces a permutation and since conditions (i) and (ii) of Theorem 5.1 are necessary conditions, $P^*(x)$ satisfies conditions (i) and (ii) of Theorem 5.2. It is clear that if $P^*(x)$ is a motion, so also is $P(x)$, and thus it is sufficient to consider polynomials $P(x)$ for which $a_{p-2} = 0$.

Since there are only a finite number of different permutations on the elements of $GF(p)$, it is possible to choose a finite set of distinct permutations

$$\{q_1, q_2, \cdots, q_s\} = \Gamma$$

in such a way that for each t, p_t is equal to one of the q_j. Note that $1 \leqq s \leqq p!$. Now, with each permutation q_j, there is associated at most a finite number of polynomials

$$Q_{jk}(x) = i_{0,k}^{(j)} x^{p-2} + i_{1,k}^{(j)} x^{p-3} + \cdots + i_{p-2,k}^{(j)} x \qquad [k = 1, 2, \cdots, w_j]$$

in $GF(p)\ [x]$ which satisfy the conditions of Theorem 5.1 and such that $q_j(i) = Q_{jk}(i)$, $i = 0, 1, \cdots, p-1$, $k = 1, 2, \cdots, w_j$.

Define b_{j_k}, an element in the Boolean ring of idempotents, as follows:

$$b_{j_k} = (a_0 - i_{0,k}^{(j)})^{p-1} \cup (a_1 - i_{1,k}^{(j)})^{p-1} \cup \cdots \cup (a_{p-3} - i_{p-2,k}^{(j)})^{p-1} .$$

This element has a zero in those components t of the subdirect sum representation of the Boolean ring of idempotents, where

$$a_{h,t} = i_{h,k}^{(j)} \qquad [h = 0, 1, \cdots, p-3]$$

and has a 1 in the other components. Let

$$b_j = 1 + \prod_{k=1}^{w_j} b_{j_k} ,$$

and note that b_j has a 1 in those components t where the permutation $q_j = p_t$ and zeros elsewhere.

Define a matrix $M = (m_{ij})$ as follows:

$$m_{ij} = b_{j_1} + b_{j_2} + \cdots + b_{j_v}$$

where $q_{j_1}, q_{j_2}, \cdots, q_{j_r}, \cdots, q_{j_v}$ are those elements of Γ which satisfy $q_{j_r}(i) = j$, and $m_{ij} = 0$ if there are no such permutations in Γ. It can be seen that m_{ij} has a 1 in the tth component if and only if $P_t(i) = j$. Since the b_j are pairwise orthogonal and a permutation is a one-to-one onto map, it is clear that M satisfies the conditions for a motion matrix and $P(x) \equiv xM$.

To illustrate the second point of view in analytic geometry reference will be made to the particular instance of a 3-space, although similar results could be obtained for larger primes.

It follows from the Boolean vector representation of p-rings that a 3-ring can be represented as the set of all pairwise orthogonal ordered pairs (x, y) of elements from its Boolean ring of idempotents. Thus the pair (x, y) can be considered as coordinates for points in the 3-space. The locus of all points of the 3-space, whose coordinates satisfy an equation of the form $Ax + By + C = 0$, where $A \cup B = 1$, is called a *linear set*. (The indicated operations are those of the Boolean ring of idempotents).

THEOREM 5.3. *A linear set is a circle of radius $A + B + C$.*

Proof. Denote by Ω the linear set associated with the equation $Ax + By + C = 0$. Then if $(x, y) \in \Omega$,

$$d[(x, y), (1 + B, 1 + A)] = A + B + C .$$

For

$$\begin{aligned} d[(x, y), (1 + B, 1 + A)] &= d[((1 + B, 1 + A) - (x, y)), 0] \\ &= d[(c, d), 0] = c + d \end{aligned}$$

where

$$\begin{aligned} c &= (1 + A)x + y(1 + A + 1 + B + 1) + (1 + B)(1 + x + y) \\ d &= (1 + B)y + x(1 + A + 1 + B + 1) + (1 + A)(1 + x + y) \end{aligned}$$

hence

$$c + d = Ax + By + A + B = A + B + C .$$

Also if $d[(1 + B, 1 + A), (x, y)] = A + B + C$ then from the above

$$[d(x, y), (1 + B, 1 + A)] = Ax + By + A + B$$

and hence $Ax + By + C = 0$.

COROLLARY. *The form $A + B + C$ is a complete set of invariants for linear sets under motions.*

The following theorem illustrates a connection between the geometry of a p-space and the geometry of its Boolean ring of idempotents.

THEOREM 5.4. *If R is a p-space and B the corresponding Boolean ring of idempotents, then B itself is a Boolean metric space and is isometric to the set of idempotents of R, considered as a sub-*

space of R. Further, any motion on B, can be extended to a motion on R.

Proof. In an autometrized Boolean ring, the distance between two elements is the ring sum. But if x and y are idempotents in a ring their sum in the Boolean ring of idempotents is $x + y - 2xy$. But it is easy to see that if x and y are idempotents in a p-ring $x + y - 2xy = (x - y)^{p-1}$. Hence the distance between two idempotents is the same, whether the set of idempotents is considered as a subspace of the p-space, or as forming a Boolean ring itself.

If f is a motion on B, then the motion $f^*(x) = xM + f(0)$ is a motion on R which coincides with f on B, where the matrix $M = (m_{ij})$ is defined as:

$$m_{11} = m_{p-1,p-1} = \overline{f(0)}\ , \qquad m_{1,p-1} = m_{p-1,1} = f(0)\ ,$$

$m_{ii} = 1$ for $i \neq 1, p-1$, and all other elements in the matrix equal to zero.

REFERENCES

1. David Ellis, *Autometrized Boolean algebras I, II*, Canadian J. Math. **3** (1951), 87–93 and 145–147.

2. ———, *Geometry in abstract distance spaces*, Mathematical Publications (Debrecen), (1951), 1–25.

3. R. F. Keller, *A lattice valuation for simple and semi-simple algebras*, Doctoral Dissertation, University of Missouri (1962).

4. C. J. Penning, *Boolean Metric spaces*, Doctoral Thesis, Technische Hogeschool te Delft, (1960).

5. G. Raussnitz, *Zur theorie der congruenzen hoheren grades*, Mathematische und Naturwissenschaftliche Berichte aus Ungarn, **I** (1882–1883), 266–278.

6. M. H. Stone, *The theory of representations for Boolean algebras*, Trans. Amer. Math. Soc. **40** (1936), 37–111.

7. J. L. Zemmer, *Some remarks on p-rings and their Boolean geometry*, Pacific J. Math., **6** (1956), 193–208.

UNIVERSITY OF RHODE ISLAND

BASIC SEQUENCES AND THE PALEY-WIENER CRITERION

JAMES R. RETHERFORD

1. Introduction. Throughout the paper X will denote a complete metric linear space (i.e., a complete topological linear space with topology derived from a metric d with the property that $d(x, y) = d(x - y, 0)$, for all $x, y \in X$) or some specialization thereof over the real or complex field; $\|x\|$ will denote $d(x, 0)$; and if $\{x_n\}$ is a sequence in X, $[x_n]$ will denote the closed linear span of the elements $\{x_n\}_{n \in \omega}$.

A sequence $\{x_n\}$ is said to be a *basic sequence of vectors* if $\{x_n\}$ is a basis of vectors of the space $[x_n]$, i.e., for each $x \in [x_n]$ there corresponds a unique sequence of scalars $\{a_i\}$ such that

$$x = \sum_{i=1}^{\infty} a_i x_i \;, \tag{1.1}$$

the convergence being in the topology of X. We say that the basis is unconditional if the convergence in (1.1) is unconditional. It is well known that if $\{x_n\}$ is a basic sequence of vectors, then every $x \in [x_n]$ can be represented in the form $x = \sum_{i=1}^{\infty} f_i(x) x_i$ where $\{f_i\}$ is the sequence of continuous coefficient functionals biorthogonal to $\{x_i\}$ (Arsove [1, p. 368], Dunford and Schwartz [4, p. 71]).

Similarly, we say that a sequence $\{M_i\}$ of nontrivial subspaces of a complete metric linear space X is a *basis of subspaces* of X, if for each $x \in X$, there corresponds a unique sequence $\{x_i\}$, $x_i \in M_i$ for each i, such that

$$x = \sum_{i=1}^{\infty} x_i \;. \tag{1.2}$$

This concept has been studied by Fage [5], Markus [9], and others in separable Hilbert space and by Grimblyum [6] and McArthur [10] in complete metric linear spaces. We say that the basis of subspaces is *unconditional* if the convergence in (1.2) is unconditional.

If $\{M_i\}$ is a basis of subspaces for X, for each $i \in \omega$ define E_i from X into X by $E_i(x) = x_i$ where $\sum_{i=1}^{\infty} x_i$ is the unique representation of $x \in X$. E_i is a projection (linear and idempotent); $E_i E_j = 0$ if $i \neq j$; the range of E_i is M_i; for each $x \in X$, $x = \sum_{i=1}^{\infty} E_i(x)$ and if $E_i(x) = 0$ for each i, then $x = 0$. $\{M_i\}$ will be called a *Schauder basis of subspaces* if each E_i is continuous.

September 24, 1962. This research was supported by the Air Force Office of Scientific Research.

A sequence $\{M_i\}$ of non-trivial subspaces of X is a (*unconditional*) *basic sequence of subspaces* if $\{M_i\}$ is a (unconditional) basis of subspaces of $[M_i]$, the closed linear span of $\bigcup_{i\in\omega} M_i$. If $\{M_i\}$ is a basic sequence of subspaces and $x \in [M_i]$ then $x = \sum_{i=1}^{\infty} E_i(x)$, where E_i is now defined on $[M_i]$.

The classical Paley-Wiener theorem can be formulated in X as follows.

1.3. THEOREM. *Let $\{x_n\}$ and $\{y_n\}$ be sequences in X and let λ be a real number $(0 < \lambda < 1)$ such that*

$$\left\|\sum_{n=1}^{m} a_n(x_n - y_n)\right\| \leqq \lambda \left\|\sum_{n=1}^{m} a_n x_n\right\| \tag{1.3a}$$

holds for arbitrary scalars $a_1, \cdots, a_m$. Then (1) *if $\{x_n\}$ is a basis so is $\{y_n\}$;* (2) *if $\{x_n\}$ is fundamental* (i.e., $[x_n] = X$) *so is $\{y_n\}$.*

Recently Arsove [1] showed that Theorem 1.3 is valid in a complete metric linear space. It is the purpose of this paper to show that this result and results similar to those of Pollard [13], Hilding [7], and Nagy [11] (all of which generalize condition 1.3a) are valid for basic sequences of subspaces in X. As a corollary to Theorem 4.3 we obtain a new version of the Paley-Wiener theorem.

The author wishes to express his gratitude to Professor C. W. McArthur for his help and encouragement in the preparation of this paper.

2. Basic sequences of subspaces.

Special cases of the following lemma have been used by Hilding [7, p. 93], Nagy [11, p. 76], and others to prove theorems similar to Theorems 2.3 and 2.4.

2.1. LEMMA. *Let $\{M_i\}$ and $\{N_i\}$ be sequences of nontrivial subspaces of the complete metric linear space X. Suppose that for each $i \in \omega$ there exists a one-to-one linear transformation T_i of M_i onto N_i and suppose further that there are positive numbers m, M such that*

$$m\left\|\sum_{i=1}^{p} x_i\right\| \leqq \left\|\sum_{i=1}^{p} T_i(x_i)\right\| \leqq M\left\|\sum_{i=1}^{p} x_i\right\| \tag{2.1a}$$

holds for arbitrary $x_i \in M_i$, $i = 1, \cdots, p$. Then
(i) *there is a linear homeomorphism T of $[M_i]$ onto $[N_i]$ such that the restriction of T to M_i equals T_i for each $i \in \omega$ and such that*

$$m\,\|x\| \leqq \|T(x)\| \leqq M\,\|x\|\,, \text{ for all } x \in [M_i]\,. \tag{2.1b}$$

(ii) *$\{M_i\}$ is a (unconditional) basic sequence of subspaces if and only if $\{N_i\}$ is a (unconditional) basic sequence of subspaces.*

Proof. Let X_0 denote the space of finite linear combinations of $\bigcup_{i\in\omega} M_i$. These, of course, are reducible to the form $\sum_{i=1}^n x_i$, $x_i \in M_i$. If $x_i, x_i' \in M_i$, $i = 1, \cdots, p$ and $\sum_{i=1}^p x_i = \sum_{i=1}^p x_i'$ then from 2.1a it follows that $\sum_{i=1}^p T_i(x_i) = \sum_{i=1}^p T_i(x_i')$. Thus we may define a linear transformation S from X_0 into $[N_i]$ by $S(\sum_{i=1}^p x_i) = \sum_{i=1}^p T_i(x_i)$ and have $m\|x\| \leqq \|S(x)\| \leqq M\|x\|$, for all $x \in X_0$. It is clear that S restricted to M_i is equal to T_i and that S is continuous. Thus defined on a dense subset of $[M_i]$, S has a unique linear extension T to $[M_i]$ satisfying 2.1b. From 2.1b it follows that T is one-to-one and T^{-1} is continuous. We show T is onto $[N_i]$.

Let $y \in [N_i]$. Then $y = \lim_k g_k$ where g_k is of the form $g_k = \sum_{i=1}^{n(k)} y_i^{(k)}$, $y_i^{(k)} \in N_i$, $i = 1, \cdots, n(k)$. For each such $y_i^{(k)}$ there is a unique $x_i^{(k)} \in M_i$ such that $T_i(x_i^{(k)}) = y_i^{(k)}$. Let $h_k = \sum_{i=1}^{n(k)} x_i^{(k)}$. Then from 2.1b, $\|h_p - h_q\| \leqq (1/m)\|g_p - g_q\|$, so $\{h_k\}$ is Cauchy and there is an $x_0 \in [M_i]$ such that $\{h_k\} \to x_0$. Clearly, $T(x_0) = y$.

To verify (ii) suppose $\{M_i\}$ is basic, i.e., a basic sequence of subspaces. Let $y \in [N_i]$. Then $y = T(x)$ for some $x \in [M_i]$. x has a unique expansion $x = \sum_{i=1}^\infty x_i$, $x_i \in M_i$ and $y = \sum_{i=1}^\infty T(x_i)$, $T(x_i) \in N_i$. Now if $y = \sum_{i=1}^\infty y_i$, $y_i \in N_i$, then $y_i = T(x_i')$ for some unique $x_i' \in M_i$. Hence $0 = T(\sum_{i=1}^\infty x_i - x_i')$ which implies $x_i = x_i'$. Since the expansion for y is unique, it follows that $\{N_i\}$ is basic. The converse follows from (i) in the same way. If in the preceding argument $\{M_i\}$ had been assumed an unconditional basis of subspaces for $[M_i]$ then the series $\sum_{i=1}^\infty x_i$ would have been unconditionally convergent to x and since T is a linear homeomorphism it follows that $\sum_{i=1}^\infty T(x_i)$ would be unconditionally convergent.

2.2. DEFINITION. *Two sequences $\{x_i\}$ and $\{y_i\}$ (in the given order) in X are said to have the property*:

(P–W) (*for Paley-Wiener*) *if there is a real number λ $(0 < \lambda < 1)$ such that $\|\sum_{i=1}^n a_i(x_i - y_i)\| \leqq \lambda \|\sum_{i=1}^n a_i x_i\|$ holds for arbitrary scalars $a_1, a_2, \cdots, a_n$;*

(P–H) (*for Pollard-Hilding*) *if for each positive real number k, there are real numbers $\lambda_1, \lambda_2 (0 \leqq \lambda_i < \min[1; 2^{1-1/k}], i = 1, 2)$ such that*

$$\left\|\sum_{i=1}^n a_i(x_i - y_i)\right\| \leqq \left[\lambda_1 \left\|\sum_{i=1}^n a_i x_i\right\|^k + \lambda_2 \left\|\sum_{i=1}^n a_i y_i\right\|^k\right]^{1/k}$$

holds for arbitrary scalars $a_1, \cdots, a_n$;

(N) (*for Nagy*) *if there are real numbers λ', μ, ν $(0 \leqq \lambda' < 1,\ 0 \leqq \nu < 1,\ 0 \leqq \mu,\ \mu^2 \leqq [1 - \lambda'][1 - \nu])$ such that*

$$\left\|\sum_{i=1}^{n} a_i(x_i - y_i)\right\|^2 \leqq \lambda'\left\|\sum_{i=1}^{n} a_i x_i\right\|^2 + \mu\left\|\sum_{i=1}^{n} a_i x_i\right\|\cdot\left\|\sum_{i=1}^{n} a_i y_i\right\| + \nu\left\|\sum_{i=1}^{n} a_i y_i\right\|^2$$

holds for arbitrary scalars $a_1, \cdots, a_n$.

If $k = 1$ and $\lambda_1 = \lambda_2$ property P–H reduces to

$$\left\|\sum_{i=1}^{n} a_i(x_i - y_i)\right\| \leqq \lambda\left[\left\|\sum_{i=1}^{n} a_i y_i\right\| + \left\|\sum_{i=1}^{n} a_i x_i\right\|\right] \tag{2.2a}$$

where $\lambda = \lambda_1 = \lambda_2$.

2.3. LEMMA. *If* $\{x_n\}$ *and* $\{y_n\}$ *are sequences in* X *with property* P–W, P–H *or* N *then* 2.2a *holds, with* λ $(0 < \lambda < 1)$ *an appropriately chosen constant.*

Proof. That property P–W implies 2.2a is obvious. If $\{x_n\}$, $\{y_n\}$ have property P–H, let $\lambda = [\max(\lambda_1, \lambda_2)]^{1/k}$; if $\{x_n\}$, $\{y_n\}$ have property N let $\lambda = [\max(\lambda', \mu, \nu)]^{1/2}$.

2.4. THEOREM. *Suppose* $\{M_i\}$ *and* $\{N_i\}$ *are sequences of nontrivial subspaces of* X *and suppose that for each* $i \in \omega$, T_i *is a one-to-one linear transformation of* M_i *onto* N_i. *Suppose further that there is a* $\lambda(0 < \lambda < 1)$ *such that*

$$\left\|\sum_{i=1}^{n} (x_i - T_i(x_i))\right\| \leqq \lambda\left(\left\|\sum_{i=1}^{n} x_i\right\| + \left\|\sum_{i=1}^{n} T_i(x_i)\right\|\right) \tag{2.4a}$$

holds for arbitrary $x_i \in M_i$, $i = 1, \cdots, n$. *Then*
(i) *there is a linear homeomorphism* T *of* $[M_i]$ *onto* $[N_i]$ *such that* T *restricted to* M_i *equals* T_i *for each* i *and such that*

$$[(1 - \lambda)/(1 + \lambda)]\,\|x\| \leqq \|T(x)\| \leqq [(1 + \lambda)/(1 - \lambda)]\,\|x\| \tag{2.4b}$$

for each $x \in [M_i]$;
(ii) $\{M_i\}$ *is a (unconditional) basic sequence of subspaces if and only if* $\{N_i\}$ *is a (unconditional) basic sequence of subspaces.*

Proof.

$$\begin{aligned}\left\|\sum_{i=1}^{n} T_i(x_i)\right\| &\leqq \left\|\sum_{i=1}^{n} (T_i(x_i) - x_i)\right\| + \left\|\sum_{i=1}^{n} x_i\right\| \\ &\leqq (1 + \lambda)\left\|\sum_{i=1}^{n} x_i\right\| + \lambda\left\|\sum_{i=1}^{n} T_i(x_i)\right\|,\end{aligned}$$

i.e.,

$$\left\|\sum_{i=1}^{n} T_i(x_i)\right\| \leqq [(1 + \lambda)/(1 - \lambda)]\left\|\sum_{i=1}^{n} x_i\right\|.$$

Similarly,

$$\left\| \sum_{i=1}^{n} x_i \right\| \leqq [(1+\lambda)/(1-\lambda)] \left\| \sum_{i=1}^{n} T_i(x_i) \right\| .$$

Thus

$$[(1-\lambda)/(1+\lambda)] \left\| \sum_{i=1}^{n} x_i \right\| \leqq \left\| \sum_{i=1}^{n} T_i(x_i) \right\| \leqq [(1+\lambda)/(1-\lambda)] \left\| \sum_{i=1}^{n} x_i \right\| .$$

The conclusions follow from Lemma 2.1.

2.5. COROLLARY. *Suppose* $\{M_i\}$ *and* $\{N_i\}$ *are sequences of non-trivial subspaces of* X *and suppose that for each* $i \in \omega$, T_i *is a one-to-one linear transformation of* M_i *onto* N_i. *Suppose further that* $\{x_i\}$ *and* $\{T_i(x_i)\}$ *have property* P–W, P–H *or* N, *for arbitrary* $x_i \in M_i$ *(observe that since* $x_i \in M_i$ *is arbitrary,* x_i *and* $T_i(x_i)$ *include the scalar* a_i *for each* i*) then the conclusions of Theorem* 2.4 *hold. In particular, if Property* P–W *holds and* $\{M_i\}$ *is a basis of subspaces for* X, *so is* $\{N_i\}$.

Proof. The first part of the corollary follows from Lemma 2.3. Arsove [1, p. 367] has shown how to prove the other assertion of the corollary. We repeat the proof for completeness.

Since Property P–W holds there exists a linear operator T from X into X satisfying $\| x - T(x) \| \leqq \lambda \| x \|$, $x \in X$ and such that T restricted to M_i equals T_i. Let $A = T - I$, where I is the identity operator. A is continuous at each $x \in X$ and furthermore $\| A^n(x) \| \leqq \lambda^n \| x \|$ for each $x \in X$ and positive integer n. Thus a linear operator U of X onto X may be defined by $U(x) = \sum_{n=0}^{\infty} (-A^n(x))$, $x \in X$. It follows that $\| U(x) \| \leqq (1-\lambda)^{-1} \| x \|$, so U is continuous. Given $y \in X$, let $x = U(y)$. Then $y = (I + A)x = T(x)$ so T is onto X. Thus $\{N_i\}$ is a basis of subspaces for X.

3. **Basic sequences of vectors.** If X has a basis of vectors $\{x_n\}$, then $\{x_n\}$ induces in a natural way a basis of subspaces $\{M_i\}$ for X. We have only to define M_i to be the span of the single element x_i (denoted by $sp(x_i)$). From the remarks in the introduction we have $x = \sum_{i=1}^{\infty} f_i(x) x_i$ for each $x \in X$, so $E_i(x) = f_i(x) x_i$. Since $h(a) = a x_i$ is a linear homeomorphism of the scalar field into X and $f_i(x)$ is a continuous linear functional it follows that E_i is continuous for each $i \in \omega$ and so $\{M_i\}$ is a Schauder basis of subspaces for X. Thus, for Schauder bases of vectors, we obtain the following theorems as corollaries to the theorems of § 2.

3.1. THEOREM. *Suppose $\{x_i\}$ and $\{y_i\}$ are nontrivial (i.e., $x_i \neq 0$, $y_i \neq 0$, for each $i \in \omega$) sequences in X and suppose there is a $\lambda (0 < \lambda < 1)$ such that*

$$\left\| \sum_{i=1}^{n} a_i(x_i - y_i) \right\| \leqq \lambda \left(\left\| \sum_{i=1}^{n} a_i x_i \right\| + \left\| \sum_{i=1}^{n} a_i y_i \right\| \right) \tag{3.1a}$$

holds for arbitrary scalars $a_1, \cdots, a_n$. Then,
(i) there exists a linear homeomorphism T of $[x_i]$ onto $[y_i]$ such that $T(x_i) = y_i$ for each $i \in \omega$, and
(ii) $\{x_i\}$ is a (unconditional) basic sequence of vectors if and only if $\{y_i\}$ is a (unconditional) basic sequence of vectors.

Proof. Let $M_i = sp(x_i)$ and $N_i = sp(y_i)$. Define a linear operator T_i from M_i onto N_i by $T_i(ax_i) = ay_i$ where a is an arbitrary scalar. Clearly, T_i is one-to-one and continuous. 3.1a can be rewritten

$$\left\| \sum_{i=1}^{n} (x_i' - T_i(x_i')) \right\| \leqq \lambda \left(\left\| \sum_{i=1}^{n} x_i' \right\| + \left\| \sum_{i=1}^{n} T_i(x_i') \right\| \right) \tag{3.1b}$$

for arbitrary $x_i' \in M_i$, $i = 1, \cdots, n$. The conclusions follow from Theorem 2.4.

Thus in particular, if $\{x_n\}$ and $\{y_n\}$ are nontrivial sequences in X with property P–W, P–H or N, the conclusions of 3.1 are valid.

We have remarked that if $\{x_n\}$ and $\{y_n\}$ have property P–W and $\{x_n\}$ is a basis of vectors for X, then $\{y_n\}$ is a basis of vectors for X. From 3.1 it follows that if $\{x_n\}$ is an unconditional basis of vectors for X, then $\{y_n\}$ is an unconditional basis of vectors for X.

4. Basic sequences in Banach spaces.

From Grinblyum [6] the following can be derived (a proof is given in [10]).

4.1. LEMMA. *Let $\{M_i\}$ be sequence a of nontrivial closed subspaces in a Banach space X. $\{M_i\}$ is a Schauder basis of subspace for $[M_i]$ if and only if there is a $K \geqq 1$ such that for arbitrary $p, q \in \omega$, $p \leqq q$ we have $\| \sum_{i=1}^{p} x_i \| \leqq K \| \sum_{i=1}^{q} x_i \|$, for arbitrary $x_i \in M_i$, $i = 1, \cdots, q$.*

4.2. LEMMA. *Let $\{M_i\}$ be a sequence of nontrivial closed subspaces of a Banach space X. $\{M_i\}$ is an unconditional Schauder basis of subspaces of $[M_i]$ if and only if there is a $K \geqq 1$ such that for arbitrary finite sets of positive integers F, F' with $F \subset F'$ we have $\| \sum_{i \in F} x_i \| \leqq K \| \sum_{i \in F'} x_i \|$, for arbitrary $x_i \in M_i$.*

4.3. THEOREM. *Suppose $\{M_i\}$ and $\{N_i\}$ are sequences of closed nontrivial subspaces of a Banach space X.*

(1) *If there is a* $\lambda (0 < \lambda < 1)$ *such that for an arbitrary finite set of integers* F' *and arbitrary* $y_i \in N_i$, $i \in F'$, *there exists* $x_i \in M_i$, $i \in F'$ *such that*

$$\left\| \sum_{i \in F} (y_i - x_i) \right\| \leqq \lambda \left[\left\| \sum_{i \in F} x_i \right\| + \left\| \sum_{i \in F} y_i \right\| \right] \tag{4.3a}$$

holds for arbitrary $F \subset F'$ *then* $\{N_i\}$ *is an unconditional (Schauder) basic sequence of subspaces if* $\{M_i\}$ *is an unconditional (Schauder) basic sequence of subspaces;*
(2) *if there is a* $\lambda (0 < \lambda < 1)$ *such that for arbitrary* $q \in \omega$ *and arbitrary* $y_1, \cdots, y_q, y_i \in N_i$, $i = 1, \cdots, q$ *there exist* $x_1, \cdots, x_q, x_i \in M_i$, $i = 1, \cdots, q$ *such that*

$$\left\| \sum_{i=1}^{p} (y_i - x_i) \right\| \leqq \lambda \left[\left\| \sum_{i=1}^{p} x_i \right\| + \left\| \sum_{i=1}^{p} y_i \right\| \right] \tag{4.3b}$$

holds for all $p \leqq q$ *then* $\{N_i\}$ *is a (Schauder) basic sequence of subspaces if* $\{M_i\}$ *is a (Schauder) basic sequence of subspaces.*

Proof. We prove (2). The proof of (1) is analogous using Lemma 4.2 instead of 4.1.

Suppose $\{M_i\}$ be a basis of subspaces for $[M_i]$. By Lemma 4.1 there is a $K \geqq 1$ such that

$$\left\| \sum_{i=1}^{p} x_i \right\| \leqq K \left\| \sum_{i=1}^{q} x_i \right\|, \ x_i \in M_i, \ p \leqq q .$$

We have

$$\left\| \sum_{i=1}^{p} y_i \right\| \leqq \left\| \sum_{i=1}^{p} (y_i - x_i) \right\| + \left\| \sum_{i=1}^{p} x_i \right\|$$

and from (4.4b) it follows that

$$\left\| \sum_{i=1}^{p} y_i \right\| \leqq \frac{1+\lambda}{1-\lambda} \left\| \sum_{i=1}^{p} x_i \right\| .$$

Also

$$\left\| \sum_{i=1}^{q} x_i \right\| \leqq \frac{1+\lambda}{1-\lambda} \left\| \sum_{j=1}^{q} y_i \right\| .$$

Thus we have

$$\left\| \sum_{i=1}^{p} y_i \right\| \leqq \left[\frac{1+\lambda}{1-\lambda} \right]^2 K \left\| \sum_{i=1}^{q} y_i \right\| .$$

Thus by Lemma 4.1, $\{N_i\}$ is a basis of subspaces for $[N_i]$.

4.4. COROLLARY. *Let* $\{x_i\}$ *and* $\{y_i\}$ *be non-trivial sequences in a Banach space* X.
(1) *If there is a* $\lambda(0 < \lambda < 1)$ *such that for an arbitrary finite set of indices* F' *and arbitrary scalars* $\{a_i\}$, $i \in F'$, *there exist scalars* $\{b_i\}$, $i \in F'$, *such that*

$$\left\| \sum_{i \in F} (a_i y_i - b_i x_i) \right\| \leqq \lambda \left[\left\| \sum_{i \in F} a_i y_i \right\| + \left\| \sum_{i \in F} b_i x_i \right\| \right] \tag{4.4a}$$

holds for arbitrary $F \subset F'$ *then* $\{y_i\}$ *is an unconditional (Schauder) basic sequence of vectors if* $\{x_i\}$ *is an unconditional (Schauder) basic sequence of vectors;*
(2) *if there is a* $\lambda(0 < \lambda < 1)$ *such that for arbitrary* $q \in \omega$ *and arbitrary scalars* $a_1, \cdots, a_q$ *there are scalars* $b_1, \cdots, b_q$ *such that*

$$\left\| \sum_{i=1}^{p} (a_i y_i - b_i x_i) \right\| \leqq \lambda \left[\left\| \sum_{i=1}^{p} b_i x_i \right\| + \left\| \sum_{i=1}^{p} a_i y_i \right\| \right] \tag{4.4b}$$

holds for all $p \leqq q$ *then* $\{y_i\}$ *is a (Schauder) basic sequence of vectors if* $\{x_i\}$ *is a(Schauder) basic sequence of vectors.*

Proof. Let $M_i = sp(x_i)$, $N_i = sp(y_i)$ and apply the preceeding theorem.

4.4 is a new form of the Paley-Wiener theorem for we no longer require the coefficients of x_i and y_i to be the same. We could now define properties similar to properties P–W, P–H and N by merely asserting the existence of a scalar b_i to replace the coefficient of x_i in each of the properties defined in 2.2. It is easy to see that these new forms of properties P–W, P–H and N imply the hypotheses of corollary 4.5.

It is unknown to the author whether $[x_n]$ is linearly homeomorphic to $[y_n]$ or not.

REFERENCES

1. M. G. Arsove, *The Paley-Wiener theorem in metric linear spaces*, Pacific J. Math., **10** (1960), 365–379.
2. C. Bessaga and A. Pelczynski, *On bases and unconditional conditional convergence of series in Banach spaces*, Studia Math., **17** (1958), 151–164.
3. M. M. Day, *Normed linear spaces*, Springer-Verlag, Berlin, 1958.
4. N. Dunford and J. T. Schwartz, *Linear Operators*, Part I, Interscience Publishers, New York, 1958.
5. M. K. Fage, *The rectification of bases in Hilbert space*, Dokl. Akad. Nauk., SSSR (N.S.) **75** (1950), 1053–1056 (In Russian).
6. M. M. Grinblyum, *On the representation of a space of type B in the form of a direct sum of subspaces*, Dokl. Akad. Nauk. SSSR (N.S.) **70** (749–752.

7. S. H. Hilding, *Note on completeness theorems of Paley-Wiener type*, Ann. of Math., (2) **49** (1948), 953-955.
8. M. Krein, D. Milman and M. Rutman, *On a property of a basis in a Banach space*, Khark. Zap. Matem. Obsh. (4), **16** (1940), 182. (In Russian with English Resume).
9. A. S. Markus, *A basis of root vectors of a dissipative operator*, Soviet Math.-Doklady, Amer. Math. Soc. Trans. **1** (1960), 599-602.
10. C. W. McArthur. *Infinite direct sums in complete metric linear spaces* (to appear).
11. B. Sz. Nagy, *Expansion theorems of Paley-Wiener type*, Duke Math. J., **14** (1947), 975-978.
12. R. E. A. C. Paley and N. Wiener, *Fourier transforms in the complex domain*, New York, 1934.
13 H. Pollard, *Completeness theorems of Paley-Wiener type*, Ann. of Math. (2) **45** (1944), 738-739.

FLORIDA STATE UNIVERSITY

QUASI-POSITIVE OPERATORS

D. W. SASSER

1. Introduction. The classical results of Perron and Frobenius ([6], [7], [12]) assert that a finite dimensional, nonnegative, non-nilpotent matrix has a positive eigenvalue which is not exceeded in absolute value by any other eigenvalue and the matrix has a nonnegative eigenvector corresponding to this positive eigenvalue. If the matrix has strictly positive entries, then there is a positive eigenvalue which exceeds every other eigenvalue in absolute value, and the corresponding space of eigenvectors is one-dimensional and is spanned by a vector with strictly positive coordinates. Numerous generalizations of these results to order-preserving linear operators acting in ordered linear spaces have appeared in recent years; a short bibliography is included at the end of this paper. In this paper a generalization in a different direction is obtained which reduces, in the finite dimensional case, to the assertion that the Perron-Frobenius theorems hold if it is only required that all but a finite number of the powers of the matrix satisfy the given conditions. The principal results are theorems of the Perron-Frobenius type which are applicable to any compact linear operator (the compactness condition is weakened somewhat), acting in an ordered real Banach space B, which satisfies a condition weaker than order-preserving. In addition, the results apply to the case when the "cone" of positive elements in B has no interior.

2. Preliminaries. Throughout the sequel, B will denote a real Banach space with norm $\|\cdot\|$. The complex extension of B, $\tilde{B}$, is the complex Banach space $\tilde{B} = \{x + iy \mid x, y \in B\}$ with the obvious definitions of addition and complex scalar multiplication and the norm in $\tilde{B}$ is $\| x + iy \| = \sup_\theta \| \cos\theta \cdot x + \sin\theta \cdot y \|$. If T is a (real) linear operator on B into B, the (complex) linear operator $\tilde{T}$ on $\tilde{B}$ into $\tilde{B}$ is defined by $\tilde{T}(x + iy) = Tx + iTy$. T is bounded if and only if $\tilde{T}$ is bounded, in which case $\| T \| = \| \tilde{T} \|$. The spectrum, $\sigma(T)$, and the resolvent, $\rho(T)$, are defined to be the corresponding sets associated with the operator $\tilde{T}$. We denote the spectral radius of T by r_T, $r_T = \lim_{n\to\infty} \| T^n \|^{1/n} = \sup_{\lambda \in \sigma(T)} |\lambda|$ (provided $\| T \| < \infty$).

In all of our results there will be a basic assumption that the linear operator under consideration is quasi-compact, a notion which we will now define. A bounded linear operator T is compact (also called completely continuous) if each sequence Tx_1, Tx_2, $\cdots$, with

Received September 13, 1962. This work was performed under the auspices of the United States Atomic Energy Commission.

$\|x_i\| \leqq 1$, $i = 1, 2, \cdots$, has a convergent subsequence. T is quasi-compact if there exists a positive integer n and a bounded linear operator V such that $T^n - V$ is compact and $r_V < r_T^n$.[1] There are a number of properties possessed by quasi-compact operators some of which we state now without proof.[2] If $\lambda_0 \in \sigma(T)$ and $|\lambda_0| = r_T$, then λ_0 is an isolated point in $\sigma(T)$ and is in the point spectrum, i.e., $(\lambda_0 I - \tilde{T})$ is not one-to-one. The resolvent operator, $R(\lambda, T) \equiv (\lambda I - \tilde{T})^{-1}$, exists in a neighborhood of λ_0 (excluding λ_0) and, in this neighborhood, $R(\lambda, T)$ has a Laurent series expansion of the form

$$R(\lambda, T) = \sum_{k=1}^{n(\lambda_0)} \frac{(\lambda_0 I - \tilde{T})^{k-1}}{(\lambda - \lambda_0)^k} P(\lambda_0, T) + \sum_{k=0}^{\infty} (\lambda - \lambda_0)^k A_k(\lambda_0, T)$$

where $A_k(\lambda_0, T)$ is a bounded linear operator and the series on the right is convergent in the uniform operator topology. The integer $n(\lambda_0)$ is the index of λ_0, i.e., $n(\lambda_0)$ is the smallest integer n such that $\{x \mid (\lambda_0 I - \tilde{T})^{n+1} x = 0\} = \{x \mid (\lambda_0 I - \tilde{T})^n x = 0\}$. $P(\lambda_0, T)$ is a projection onto the finite dimensional space $\{x \mid (\lambda_0 I - \tilde{T})^{n(\lambda_0)} x = 0\}$. The minimal property of $n(\lambda_0)$ implies that $(\lambda_0 I - \tilde{T})^{n(\lambda_0)-1} P(\lambda_0, T) \neq 0$.

We recall that for an arbitrary bounded linear operator, the resolvent $R(\lambda, T) = (\lambda I - \tilde{T})^{-1}$ is an analytic function of λ for $\lambda \in \rho(T)$ and the expansion $R(\lambda, T) = \sum_{k=0}^{\infty} (1/\lambda)^{k+1} \tilde{T}^k$ is valid for $|\lambda| > r_T$.

3. Quasi-positive operators. A *cone* in B is a convex set K which contains λx for all $\lambda \geqq 0$ if it contains x. K is a *proper* cone if $x \in K$ and $-x \in K$ imply $x = 0$. A cone K induces an ordering $\geqq$ in B with $x \geqq y$ if and only if $x - y \in K$. This transitive ordering satisfies

(1) if $x \geqq y$, $u \geqq v$, then $x + u \geqq y + v$,

(2) if $x \geqq y$ and $\lambda \geqq 0$, then $\lambda x \geqq \lambda y$, and

(3) $x \geqq y$ if and only if $-y \geqq -x$.

If the cone is proper, then the ordering satisfies, in addition,

(4) if $x \geqq y$ and $y \geqq x$, then $x = y$.

We will use the notation $x > y$ to denote $x \geqq y, x \neq y$. Associated with a cone K is a closed cone K^+ in the conjugate space B^* of continuous, real-valued, linear functions on B, consisting of those $x^* \in B^*$ with the property that $x^*(x) \geqq 0$ for all $x \in K$. K^+ is a proper cone if and only if the linear space spanned by K is dense in B (a set with this property is called *fundamental*). This is an easy consequence of the Hahn-Banach theorem on the extension of linear functionals. We will use the notations $x^* \geqq y^*$ and $x^* > y^*$ to denote $x^* - y^* \in K^+$

[1] Note that a compact operator is quasi-compact if and only if it has a positive spectral radius.

[2] For details, see Yu. L. Smvl'yan, *Completely continuous perturbations of operators*, Amer. Math. Soc. Translations **10**, 341–344.

and $x^* - y^* \in K^+$, $x^* \neq y^*$, respectively. An element $x > 0$ ($x^* > 0$) will be called *strictly positive* if $x^*(x) > 0$ for all $x^* > 0$ ($x^*(x) > 0$ for all $x > 0$).

The following theorem is a characterization of a closed cone and its interior (when the latter is nonvoid) in terms of K^+. The proof may be found, for example, in [11] (Theorem 1.3 and its corollaries, pg. 16).

THEOREM 1. *Let K be a closed cone in B. Then $x \in K$ if and only if $x^*(x) \geqq 0$ for all $x^* \geqq 0$. If K has a nonvoid interior, then*

(1) *x is in the interior of K if and only if x is strictly positive and*

(2) *for each x on the boundary of K there exists an $x^* > 0$ such that $x^*(x) = 0$.*

COROLLARY. *If K is a closed proper cone, K^+ is a total set of functionals,* i.e., *for each $x \neq 0$, $x \in B$, there exists $x^* > 0$ such that $x^*(x) \neq 0$.*

Proof. Since either $x \notin K$ or $-x \notin K$ if $x \neq 0$, this follows immediately from Theorem 1.

A linear operator T on B into B will be called *positive with respect to a cone* K if $TK \subseteqq K$. In the absence of ambiguity we will simply say T is positive. In our applications K will be a closed cone and in this case, in view of Theorem 1, T is positive if and only if $x^*(Tx) \geqq 0$ for all $x \geqq 0$, $x^* \geqq 0$. Since $Tx \geqq 0$ if $x \geqq 0$, we have $x^*(T^2x) \geqq 0$ and, in general, $x^*(T^nx) \geqq 0$ for all n and all $x \geqq 0$, $x^* \geqq 0$. We define T to be *quasi-positive* if for each pair $x \geqq 0$, $x^* \geqq 0$, there exists an integer $n(x, x^*) \geqq 1$ such that $x^*(T^nx) \geqq 0$ if $n \geqq n(x, x^*)$. We define T to be *strictly quasi-positive* if for each pair $x > 0$, $x^* > 0$, there exists an integer $n(x, x^*) \geqq 1$ such that $x^*(T^nx) > 0$ if $n \geqq n(x, x^*)$. Finally we define T to be *strongly quasi-positive* if it is not nilpotent[3] and for each pair $x > 0$, $x^* > 0$, $\liminf_{n\to\infty} x^*(T^nx)/\| T^n \| > 0$.

4. Spectral properties.

Throughout this section, K will denote a closed proper cone in B and K will be assumed to be fundamental. T will denote a quasi-compact bounded linear operator with spectral radius 1. This restriction on the spectral radius is for convenience only and the results given may be interpreted for a general (quasi-compact) bounded linear operator S with spectral radius $r_S > 0$ by considering the operator $T = (1/r_S)\, S$ which has spectral radius 1.

[3] An operator T is nilpotent if $T^n = 0$ for some n.

THEOREM 2. *If T is quasi-positive and quasi-compact with spectral radius 1, then $1 \in \sigma(T)$ and the index of 1 is not exceeded by the index of any other point $\lambda \in \sigma(T)$, $|\lambda| = 1$.*

Proof. Assume that $1 \in \rho(T)$. Since $\rho(T)$ is open and $R(\lambda, T)$ is analytic in λ for $\lambda \in \rho(T)$, it follows that the function $g(\lambda) = x^*(R(1/\lambda, T)x)$, $x > 0$, $x^* > 0$, is analytic for $1/\lambda \in \rho(T)$, in particular for λ in some neighborhood of 1. Moreover, $R(\lambda, T) = \sum_{k=0}^{\infty}(1/\lambda)^{k+1}\tilde{T}^k$ if $|\lambda| > 1$, hence $g(\lambda) = \sum_{k=0}^{\infty}\lambda^{k+1}x^*(T^k x)$ if $|\lambda| < 1$. A theorem of Pringsheim states that if a power series has nonnegative coefficients and converges in the open unit disk, either 1 is a singularity of the series or the series has radius of convergence greater than 1.[4] Clearly it is sufficient to assume that all but a finite number of the coefficients are nonnegative. Since $x^*(T^n x) \geqq 0$ if $n \geqq n(x, x^*)$, and $g(\lambda)$ is analytic in a neighborhood of 1, we conclude that the series $\sum_{k=0}^{\infty}\lambda^{k+1}x^*(T^k x)$ converges in $|\lambda| < 1 + \delta$ for some $\delta > 0$. By assumption $r_T = 1$, hence $R(\lambda, T)$ has a singularity somewhere on $|\lambda| = 1$, say at λ_0. Since T is quasi-compact, the expansion

$$R(\lambda, T) = \sum_{k=1}^{n} \frac{(\lambda_0 I - \tilde{T})^{k-1}}{(\lambda - \lambda_0)^k} P(\lambda_0, T) + \sum_{k=0}^{n} (\lambda - \lambda_0)^k A_k(\lambda_0, T)$$

is valid for $0 < |\lambda - \lambda_0| < \delta'$, where $n = n(\lambda_0)$ is the index of λ_0, and $(\lambda_0 I - \tilde{T})^{n-1}P(\lambda_0, T) \neq 0$. We may choose $x > 0$ such that $(\lambda_0 I - \tilde{T})^{n-1}P(\lambda_0, T)x = y \neq 0$ since K is fundamental and by Theorem 1 we may choose $x^* > 0$ such that $x^*(y) \neq 0$. It follows easily that

$$g(\lambda) = (\lambda/\lambda_0)^n(1/\lambda_0 - \lambda)^{-n}h(\lambda)\,, \quad |1/\lambda - \lambda_0| < \delta\,,$$

where $h(\lambda)$ is analytic and $h(1/\lambda_0) = x^*(y) \neq 0$. Thus g has a pole at $1/\lambda_0$ which contradicts the fact that g has a Taylor's series about the origin with radius of convergence greater than 1. Our assumption that $1 \in \rho(T)$ leads to a contradiction, hence $1 \in \sigma(T)$.

Now let the index of 1 be n. It is easy to see that $\lim_{\lambda\to 1}(\lambda - 1)^k R(\lambda, T) = 0$ if $k > n$. It follows that for $|\lambda| > 1$, $\lim_{\lambda\to 1}(\lambda - 1)^k \sum_{m=0}^{\infty}(1/\lambda)^{m+1}x^*(T^m x) = 0$ for every pair $x > 0$, $x^* > 0$ and clearly this implies $\lim_{\lambda\to 1}(\lambda - 1)^k \sum_{m=j}^{\infty}(1/\lambda)^{m+1}x^*(T^m x) = 0$ if $k > n$ and $j \geqq 0$. If $\lambda_0 \in \sigma(T)$, $|\lambda_0| = 1$ and λ_0 has index l, then $\lim_{\lambda\to\lambda_0}(\lambda - \lambda_0)^l R(\lambda, T) \neq 0$. We may choose $x > 0$ and $x^* > 0$ such that $\lim_{\lambda\to\lambda_0}(\lambda - \lambda_0)^l x^*(R(\lambda, T)x) \neq 0$ and it follows that for $|\lambda| > 1$, $\lim_{\lambda\to\lambda_0}(\lambda - \lambda_0)^l \sum_{m=j}^{\infty}(1/\lambda)^{m+1}x^*(T^m x) \neq 0$. Let $\lambda_0 = e^{i\varphi}$, $\lambda = \rho e^{i\varphi}$, $\rho > 1$. If $j \geqq n(x, x^*)$, $|(\lambda - \lambda_0)^l \sum_{m=j}^{\infty}(1/\lambda)^{m+1}x^*(T^m x)| \leqq (\rho - 1)^l \sum_{m=j}^{\infty}(1/\rho)^{m+1}x^*(T^m x)$. The expression on the right in this last inequality tends to zero as

[4] See Titchmarsh, *Theory of Functions*, pg. 214. Acknowledgement is due here to S. Karlin for the essence of the proof in Theorem 2 (see [**10**], Theorem 4).

ρ tends to 1 if $l > n$, hence $l \leqq n$. This completes the proof.

THEOREM 3. *If T is quasi-positive and quasi-compact with spectral radius* 1, *there exist elements $u > 0$ and $u^* > 0$ such that $Tu = u$, $T^*u^* = u^*$.*[5]

Proof. By Theorem 2, $1 \in \sigma(T)$. We have

$$R(\lambda, T) = \sum_{k=1}^{n} \frac{(I - \tilde{T})^{k-1}}{(\lambda - 1)^k} P(1, T) + \sum_{k=0}^{\infty} (\lambda - 1)^k A_k(1, T)$$

where $P(1, T)$ is a projection onto the finite-dimensional space $\{x \mid (I - \tilde{T})^n x = 0\}$ and $(I - \tilde{T})^{n-1} P(1, T) \neq 0$. Let $\Gamma = (I - \tilde{T})^{n-1} P(1, T)$. It is easy to see that $R(\lambda, T)B \subseteqq B$ for λ real. Since $\Gamma = \lim_{\lambda \to 1} (\lambda - 1)^n R(\lambda, T)$, it follows that $\Gamma B \subseteqq B$. Also $\tilde{T}\Gamma = \Gamma\tilde{T} = \Gamma$. Let $x \geqq 0$, $x^* \geqq 0$ be arbitrary and let $N = n(x, x^*)$. If $\lambda > 1$, we have $x^*(T^N R(\lambda, T)x) = \sum_{m=0}^{\infty} (1/\lambda)^{m+1} x^*(T^{N+m})x \geqq 0$. It follows that for $\lambda > 1$, $x^*(T^N \Gamma x) = \lim_{\lambda \to 1} (\lambda - 1)^n \sum_{m=0}^{\infty} (1/\lambda)^{m+1} x^*(T^{N+m} x) \geqq 0$. Since $T^N \Gamma = \Gamma$, Γ is a positive operator. We choose $v > 0$ such that $\Gamma v = u \neq 0$. Then $u > 0$ and $Tu = T\Gamma v = \Gamma v = u$. We choose $v^* > 0$ such that $v^*(u) > 0$. Letting $u^* = \Gamma^* v^*$, we see that for $x \geqq 0$, $u^*(x) = (\Gamma^* v^*)(x) = v^*(\Gamma x) \geqq 0$ since $v^* > 0$ and Γ is a positive operator. Hence $u^* \geqq 0$, and since $u^*(v) = (\Gamma^* v^*)(v) = v^*(\Gamma v) = v^*(u) > 0$, $u^* > 0$. Finally, we have $\Gamma T = \Gamma$ which implies $T^* \Gamma^* = \Gamma^*$, hence $T^* u^* = T^*(\Gamma^* v^*) = \Gamma^* v^* = u^*$ which completes the proof.

For strictly quasi-positive operators we obtain stronger results in the next two theorems.

THEOREM 4. *If T is strictly quasi-positive and quasi-compact with spectral radius* 1, *then $1 \in \sigma(T)$, 1 has index one and $\tilde{T}$ has a representation of the form $\tilde{T} = \sum_{j=1}^{m} \lambda_j P_j + S$ where $\lambda_1 = 1$, $|\lambda_j| = 1$, $P_j^2 = P_j$, $SP_j = P_j S = 0$, $j = 1, 2, \cdots, m$, $P_i P_j = 0$ if $i \neq j$, and $r_s < 1$.*

Proof. By Theorem 2, $1 \in \sigma(T)$. By Theorem 3, there exists $u^* > 0$ such that $T^* u^* = u^*$ and for $x > 0$, $u^*(x) = u^*(T^n x) > 0$ if $n \geqq n(x, u^*)$, hence u^* is strictly positive. Let the index of 1 be n. Then $\Gamma = \lim_{\lambda \to 1} (\lambda - 1)^n R(\lambda, T) \neq 0$. For $\lambda > 1$ and arbitrary x we have

$$\begin{aligned} u^*(\Gamma x) &= \lim_{\lambda \to 1} (\lambda - 1)^n \sum_{k=0}^{\infty} (1/\lambda)^{k+1} u^*(T^k x) = \lim_{\lambda \to 1} u^*(x)(\lambda - 1)^n \sum_{k=0}^{\infty} (1/\lambda)^{k+1} \\ &= u^*(x) \lim_{\lambda \to 1} (\lambda - 1)^{n-1} = 0 \end{aligned}$$

[5] T^* is the adjoint of T, defined on B^* by $(T^* x^*)(x) = x^*(Tx)$.

unless $n = 1$. In proving Theorem 3 we showed that Γ is a positive operator, hence there exists $x > 0$ such that $\Gamma x > 0$ and therfore $u^*(\Gamma x) > 0$. It follows that $n = 1$. By Theorem 2, every $\lambda_0 \in \sigma(T)$, $|\lambda_0| = 1$, has index 1 and hence $P(\lambda_0, T) = \lim_{\lambda\to\lambda_0}(\lambda - \lambda_0)R(\lambda, T)$ exists and is a projection onto the finite dimensional space $\{x \mid (\lambda_0 I - \tilde{T})x = 0\}$. Let $\lambda_1 = 1, \lambda_2 \cdots, \lambda_m$ be an enumeration of the points in $\sigma(T)$ with absolute value 1 and let $P_j = P(\lambda_j, T)$. Since $\tilde{T}$ commutes with $R(\lambda, T)$ and $P_j = \lim_{\lambda\to\lambda_j}(\lambda - \lambda_j)R(\lambda, T)$, it follows that $\tilde{T}$ commutes with P_j. For $i \neq j$ we have $\lambda_i P_i P_j = \tilde{T}P_i P_j = P_i \tilde{T} P_j = \lambda_j P_i P_j$, hence $P_i P_j = 0$. Define the bounded linear operator S by the equation $\tilde{T} = \sum_{j=1}^m \lambda_j P_j + S$. Since $\tilde{T}P_j = P_j\tilde{T} = \lambda_j P_j$, $P_j^2 = P_j$ and $P_i P_j = 0$ if $i \neq j$, it follows that $P_j S = SP_j = 0$. This implies $\tilde{T}^n = \sum_{j=1}^m \lambda_j^n P_j + S^n$. Suppose $r_S \geqq 1$. T is quasi-compact, hence $\tilde{T}^n = U + V$ for some n where U is compact and $r_V < 1$. The operator U' defined by $U'x = Ux - \sum_{j=1}^m \lambda_j^n P_j x$ is compact[6] and $S^n = U' + V$. Therefore S is quasi-compact. Let $\lambda \in \sigma(S)$, $|\lambda| = r_S \geqq 1$. Then $Sx = \lambda x$ for some $x \in \tilde{B}$, $x \neq 0$. Since $P_j S = SP_j = 0$, it follows that $\tilde{T}x = \lambda x$ and therefore for some j, $\lambda = \lambda_j$ and $P_j x = x$. This implies $Sx = SP_j x = 0$, a contradiction. Therefore $r_S < 1$ and the proof is complete.

Before stating our next result, we state the following lemma which is easily proved.

LEMMA 1. *If E is a finite dimensional real Banach space, K is a cone in E and K is fundamental, then K contains an open set.*

THEOREM 5. *If T is strictly quasi-positive and quasi-compact with spectral radius* 1, *the eigenspace for T corresponding to the eigenvalue* 1 *is one-dimensional.*

Proof. By Theorem 4 we have $\tilde{T} = \sum_{j=1}^m \lambda_j P_j + S$ where P_j is a projection onto the eigenspace corresponding to λ_j, $\lambda_1 = 1$, $|\lambda_j| = 1$, $P_j S = SP_j = 0$, $j = 1, 2, \cdots, m$ and $P_i P_j = 0$ if $i \neq j$. By a theorem of Kronecker, there exists a sequence $n_1, n_2 \cdots$ of positive integers such that $\lim_{k\to\infty} \lambda_j^{n_k} = 1$, $j = 1, 2, \cdots, m$.[7] Since $r_S < 1$, it follows that $\lim_{n\to\infty} \| S^n \| = 0$. This implies $\lim_{k\to\infty} \tilde{T}^{n_k} = \sum_{j=1}^m P_j$. Let $P = \sum_{j=1}^m P_j$. For $x \in B$ we have $Px = \lim_{k\to\infty} T^{n_k}x$, hence $PB \subseteqq B$. For $x \geqq 0$ and $x^* \geqq 0$, $x^*(Px) = \lim_{k\to\infty} x^*(T^{n_k} x) \geqq 0$, hence P is a positive operator. Consider the finite dimensional real Banach space PB with closed proper cone PK. Since K is fundamental in B, it is clear that PK is fundamental in PB. Therefore, by Lemma 1, PK contains an open set (open relative to PB). Since T is strictly quasi-positive, every

[6] The compact operators from an ideal in the algebra of bounded linear operators and any bounded operator with a finite dimensional range is compact.

[7] See, for example, Hardy & Wright, *The Theory of Numbers*, Oxford Univ. Press.

non-trivial fixed vector of T in K is strictly positive. By Theorem 3, there exists $u > 0$ such that $Tu = u$. Let $Tx = x, x \neq 0$. We wish to show u and x are linearly dependent and for this purpose we may assume $x \notin K$ (otherwise replace x by $-x$). It is clear that $u \in PK$ and $x \in PB$. Let $t_0 = \sup\{t \mid u + tx \in PK\}$. Since u is in the interior of PK and $x \notin PK$, it is easy to see that $0 < t_0 < \infty$ and that $u + t_0x$ is on the boundary of PK. Hence, by Theorem 1, there exists $x^* \in (PK)^+$ such that $x^*(u + t_0x) = 0$. We extend x^* to $y^* \in B^*$ by defining $y^*(y) = x^*(Py)$. Since $PK \subseteqq K$, it follows that $y^* \in K^+$. We have $P(u+t_0x) = u + t_0x$, hence $y^*(u + t_0x) = x^*(u_0 + t_0x) = 0$. Now $u + t_0x$ is a fixed vector of T which is not strictly positive, hence $u + t_0x = 0$, which completes the proof.

Our next result is a characterization of strongly quasi-positive operators.

THEOREM 6. *If T is quasi-compact with spectral radius* 1, *then T is strongly quasi-positive if and only if the following conditions are satisfied*:

(1) $1 \in \sigma(T)$ *and* 1 *is the only point in $\sigma(T)$ with absolute value one*,

(2) *the eigenspace for T corresponding to the eigenvalue* 1 *is one-demensional and is spanned by a strictly positive element u*,

(3) *there exists a strictly positive element u^* such that $T^*u^* = u^*$.*

Proof. In Theorems 3, 4, 5 we have seen that if T is strictly quasi-positive (in particular, if it is strongly quasi-positive), then $1 \in \sigma(T)$ and (2) and (3) hold. There remains to show 1 is the only point in $\sigma(T)$ with absolute value one. We define the operator $P = \sum_{j=1}^{m} P_j$ as in Theorem 5 and recall that PB is a finite dimensional real Banach space with closed proper cone PK containing interior elements. Let $\lambda = e^{i\theta}$ be a point in $\sigma(T)$ and let $\tilde{T}(x + iy) = e^{i\theta}(x + iy)$ for some x, y in B, not both zero. It is easy to see that $Px = x$ and $Py = y$, hence $x \in PB$ and $y \in PB$. At least one of the four elements $x + y$, $x - y$, $y - x$, $-x - y$ must be not in PK since otherwise $x + y = 0$, $x - y = 0$, hence $x = y = 0$. Therefore $ax + by \notin PK$ for some choice of $a = \pm 1$ and $b = \pm 1$. Now choose $t > 0$ such that $u + t(ax + by) = v$ is on the boundary of PK. By Theorem 1, there exists $x^* \in (PK)^+$, $x^* \neq 0$, such that $x^*(v) = 0$. We extend x^* to $y^* \in K^+ : y^*(y) = x^*(Py)$. Now choose a sequence of positive integers $n_1, n_2, \cdots$ such that $\lim_{k\to\infty} e^{in_k\theta} = 1$. It follows that $\lim_{k\to\infty} T^{n_k}v = v$. Since $r_T = 1$, we have $\| T^n \| \geqq 1$ for all n and hence if $v > 0$,

$$\liminf_{n\to\infty} y^*(T^n v) \geqq \liminf_{n\to\infty} y^*(T^n v)/\| T^n \| > 0 .$$

This is impossible since $\lim_{k\to\infty} y^*(T^{n_k}v) = y^*(v) = 0$. Therefore $v = 0$, i.e., $ax + by = -(1/t)u$. Since $\tilde{T}(x + iy) = e^{i\theta}(x + iy)$, it follows that $u^*(x) + iu^*(y) = e^{i\theta}(u^*(x) + iu^*(y))$. This implies either $e^{i\theta} = 1$ or $u^*(x) = u^*(y) = 0$. The second alternative is incompatible with $ax + by = -(1/t)u$ since $u^*(u) > 0$. Therefore $e^{i\theta} = 1$ and the necessity of (1), (2), (3) is proved.

Now let T satisfy conditions (1), (2), (3). We assume without loss of generality that u^* is normalized so that $u^*(u) = 1$. Define the bounded linear operator S by $Tx = u^*(x)u + Sx$. As in Theorem 4, it can be shown that $r_S < 1$. We have $Su = Tu - u^*(u)u = u - u = 0$ and it follows that $T^n x = u^*(x)u + S^n x$. Since $r_S < 1$, $\| S^n \| \leqq M$ for all n and hence $\| T^n \| \leqq \| u^* \| \| u \| + \| S^n \| \leqq M'$ for all n. Moreover, $S^n x \to 0$ as $n \to \infty$ for all x. Hence if $x > 0$ and $x^* > 0$,

$$\liminf_{n\to\infty} x^*(T^n x)/\| T^n \| \geqq \liminf_{n\to\infty} (u^*(x)x^*(u) + x^*(S^n x))/M'$$
$$\geqq u^*(x)x^*(u)/M' > 0 .$$

Therefore T is strongly quasi-positive and the theorem is proved.

THEOREM 7. *Assume that B is a lattice*[8] *with respect to the ordering given by K. Then Theorem 6 is true if "strongly quasi-positive" is replaced by "strictly quasi-positive."*

Proof. Conditions (1), (2) and (3) in Theorem 6 imply T is strongly quasi-positive, hence, a fortiori, T is strictly quasi-positive. Now suppose T is strictly quasi-positive. Then $1 \in \sigma(T)$ and (2), (3) hold. It is easy to see from the representation of Theorem 4, $\tilde{T} = \sum_{j=1}^{m} \lambda_j P_j + S$, that $\| T^n \|$ is bounded independently of n. Hence, by a theorem of Krein-Rutman ([11], Theorem 8.1 and corollary), every $\lambda \in \sigma(T)$, $|\lambda| = 1$, is a root of unity. It is easily verified that every power of T is quasi-compact and strictly quasi-positive, hence the eigenspace for T^n corresponding to the eigenvalue 1 is one-dimensional for all n. If $\tilde{T}x = \lambda x$, $|\lambda| = 1$, $\lambda^n = 1$, then $\tilde{T}^n x = \lambda^n x = x$ and it follows that $\lambda = 1$ which completes the proof.

An immediate consequence is the following corollary.

COROLLARY. *If B is a lattice, every strictly quasi-positive and quasi-compact operator is strongly quasi-positive.*

The conclusion of this corollary is not true in general as we will illustrate by an example. Let B be three-dimensional (real) Euclidean

[8] I.e., each pair of elements in B has a greatest lower bound and a least upper bound.

space, $B = \{(x_1, x_2, x_3)\}$, and let $K = \{(x_1, x_2, x_3) \mid x_1^2 + x_2^2 \leqq x_3^2, x_3 \geqq 0\}$. If we interpret "to the right" to mean any direction in which the x_3 coordinate is increasing, each non-trivial element $x^* \in K^+$ is represented by a plane through the origin whose unit normal at the origin directed to the right lies in K. Let T be a rotation about the x_3 axis through θ radians where θ and 2π are incommensurable. It is clear that $\| T^n \| = 1$ for all n and that $TK \subseteqq K$. To show that T is strictly quasi-positive it suffices to consider $x^* \in K^+$ which is represented by a plane tangent to K. If p is in the interior of K, $T^n p$ is in the interior for all n, hence $x^*(T^n p) > 0$. Now let p be on the boundary of K. There exists exactly one point q which has the same x_3 coordinate as p and such that $x^*(q) = 0$. Since θ and 2π are incommensurable, there is at most one value of n such that $T^n p = q$. Therefore, $x^*(T^m p) > 0$ for all m sufficiently large and, hence, T is strictly quasi-positive. If p is on the boundary of K, so is $T^n p$ for all n. We can pick a sequence $n_1, n_2, \cdots$ such that $T^{n_k} p$ converges to a point q on the boundary of K and there exists $x^* \in K^+$ such that $x^*(q) = 0$, $x^* \neq 0$. This shows T is not strongly quasi-positive.

BIBLIOGRAPHY

1. T. Ando, *Positive linear operators in semi-ordered linear spaces*, J. Fac. Science, Hokkaido University, Ser. I, XIII, (1957), 214-228.

2. G. Birkhoff, *Extensions of Jentzsch's theorem*, Trans. Amer. Math. Soc., **85** (1957), 219-228.

3. F. F. Bonsall, *Endomorphisms of partially ordered vector spaces*, J. London Math. Soc., **30** (1955), 133-144.

4. ———, *Endomorphism of a partially ordered vector space without order unit*, J. London Math. Soc., **30** (1955), 144-153.

5. ———, *Linear operators in complete positive cones*, Proc. London Math. Soc., **3**, VIII (29), (1958), 53-75.

6. G. Frobenius, *Über Matrizen aus positiven Elementen*, Sitz. Ber. Preuss. Academie der Wiss. Berlin (1908), 471-476, (1909), 514-518.

7. ———, *Über Matrizen aus nicht negativen Elementen*, Sitz. Ber. Preuss. Akademie der Wiss. Berlin (1912), 456-477.

8. E. Hopf, *Über lineare Integralgleichungen mit positiven Kern*, Sitz. Ber. Preuss. Akademie der Wiss. Berlin, XVIII (1928), 233-245.

9. R. Jentzsch, *Über Integralgleichungen mit positiven Kern*, Crelles J., **141** (1912), 235-244.

10. S. Karlin, *Positive Operators*, J. Math., **8** (1959), 907-938.

11. M. Krein and M. Rutman, *Linear operators leaving invariant a cone in a Banach space*, Uspehi Matem. Navk (1948), 3-95, Amer. Math. Soc. Translation No. 26.

12. O. Perron, *Zur Theorie der Matrizen*, Math. Ann., **64** (1907), 248-263.

13. H. Schaefer, *Positive Transformationen in lokal konvexen halbgeordneten Vektorräumen*, Math. Ann., **129** (1955), 323-329.

14. ———, *Some spectral properties of positive linear operators*, Pacific J. Math., **10** (1960), 1009-1019.

SANDIA LABORATORY, ALBUQUERQUE, NEW MEXICO

ON THE STRUCTURE OF INFRAPOLYNOMIALS WITH PRESCRIBED COEFFICIENTS

O. SHISHA

Introduction. The main result of this paper is Theorem 5 which deals with the structure of infrapolynomials with prescribed coefficients. This theorem was quoted (without proof) in a previous paper [Shisha and Walsh, 1961][1], and was used there to prove a few results concerning the geometrical location of the zeros of some infrapolynomials with prescribed coefficients [loc. cit., Theorems 11, 12, 16, 17]. Two similar results are given here in Theorem 6.

We refer the reader to the Introduction of the last mentioned paper for a review of the development of the concept of infrapolynomial. Here we shall just mention two of the underlying definitions.

A. Let n and q be natural numbers ($q \leq n$), $n_1, n_2, \cdots, n_q$ integers such that $0 \leq n_1 < n_2 \cdots < n_q \leq n$, and S a set in the complex plane[2]. An nth *infrapolynomial on S with respect to* $(n_1, n_2, \cdots, n_q)$ is a polynomial $A(z) \equiv \sum_{\nu=0}^{n} a_\nu z^\nu$ such that no $B(z) \equiv \sum_{\nu=0}^{n} b_\nu z^\nu$ exists, satisfying the following properties.

(1) $B(z) \not\equiv A(z)$,

(2) $b_{n_\nu} = a_{n_\nu}$ $(\nu = 1, 2, \cdots, q)$,

(3) $|B(z)| < |A(z)|$ whenever $z \in S$ and $A(z) \neq 0$, and

(4) $B(z) = 0$ whenever $z \in S$ and $A(z) = 0$.

B. Let n be a natural number. *A simple n-sequence* is a sequence having one of the forms

$$(0, 1, \cdots, k, n-l, n-l+1, \cdots, n) \; [k \geq 0,\ l \geq 0,\ k+l+2 \leq n],$$
$$(0, 1, \cdots, k) \; [0 \leq k < n],\ (n-l, n-l+1, \cdots, n) \; [0 \leq l < n].$$

Theorem 5 may yield information on the location of the zeros of an nth infrapolynomial $A(z)$ on a set S with respect to a simple n-sequence σ. For it allows (under quite general conditions) to set $A(z) \equiv B(z)\, D(z)$ where $D(z)$ is a polynomial all of whose zeros lie in S, whereas $B(z)$ is a divisor of a polynomial $Q(z)$ whose structure is given by the theorem. By studying the location of the zeros of $Q(z)$, one may get information on the location of the zeros of $A(z)$. By this method, Theorems 11, 12, 16, 17 [loc. cit.] were proved. (Compare

Received September 12, 1963. This research was supported (in part) by the U.S Air Force through the Air Force Office of Scientific Research.

[1] Dates in square brackets refer to the bibliography.

[2] We deal throughout this paper with the *open* plane of complex numbers.

also the proof of Theorem 6 below.)

Theorem 5 is a generalization of Fekete's structure theorem [1951], and we use his method of proof [cf. also Fekete 1955]. The concept of a "juxtafunction" (Definition 1) is a generalization of Fekete's "nearest polynomial" [1955], later termed "juxtapolynomial" [Walsh and Motzkin 1957]. Theorems 1–4 and Lemmas 1–4 are contained in the author's Ph. D. thesis [1958]; they are needed for the proof of Theorem 5, and they generalize previous results of Fekete [1951, 1955]. The principal results of the present paper were published by the author (without proof) in abstracts (1958a, 1959, 1961].

1. DEFINITION 1. *Let S be a set in the complex plane and let Π be a set of complex functions defined on*[3] *S such that whenever $f_1 \in \Pi$, $f_2 \in \Pi$ and c_1, c_2 are complex numbers, then*[4] *$c_1 f_1 + c_2 f_2 \in \Pi$. Let f be a complex function defined on S. A juxtafunction to f on S with respect to Π is an element p of Π having the property: there does not exist a $q \in \Pi$ satisfying*

(a) $q(z) \neq f(z)$ *for at least one* $z \in S$,

(b) $|f(z) - q(z)| < |f(z) - p(z)|$ *whenever* $z \in S$ and $p(z) \neq f(z)$,

(c) $q(z) = f(z)$ *whenever* $z \in S$ and $p(z) = f(z)$.

EXAMPLES A. Let $S(\neq \varnothing)$ be[5] a closed and bounded set in the complex plane. Let $f, p_1, p_2, \cdots, p_n, \mu$ be complex functions with domain S which are continuous on S, and assume, furthermore, that $\mu(z) \neq 0$ throughout S. For every complex function ψ with domain S which is continuous on S, let $\|\psi\| = \max[|\mu(z)\psi(z)|, z \text{ on } S]$. It is known that there exist complex numbers $\lambda_1^*, \lambda_2^*, \cdots, \lambda_n^*$ such that for every complex $\lambda_1 \lambda_2, \cdots, \lambda_n$,

$$\left\| f - \sum_{\nu=1}^{n} \lambda_\nu^* f_\nu \right\| \leq \left\| f - \sum_{\nu=1}^{n} \lambda_\nu f_\nu \right\| . \tag{1}$$

Consider the linear space Π of all linear combinations (with complex coefficients) of $p_1, p_2, \cdots, p_n$. Then $p = \sum_{\nu=1}^{n} \lambda_\nu^* p_\nu$ is a juxtafunction to f on S with respect to Π. Indeed, suppose that some $q = \sum_{\nu=1}^{n} \lambda_\nu' p_\nu$ satisfies (a), (b) and (c) of Definition 1. Let ζ be a point of S such that

$$\|f - q\| = |\mu(\zeta)(f(\zeta) - q(\zeta))| .$$

Then by (a) $q(\zeta) \neq f(\zeta)$, and therefore, by (c), $p(\zeta) \neq f(\zeta)$. From (b) we get $\|f - \sum_{\nu=1}^{n} \lambda_\nu' f_\nu\| = \|f - q\| = |\mu(\zeta)(f(\zeta) - q(\zeta))| < |\mu(\zeta)(f(\zeta)$

[3] i.e. their domains include S.

[4] The domain of $c_1 f_1 + c_2 f_2$ is the intersection of those of f_1 and f_2.

[5] $\varnothing$ denotes the empty set.

$- p(\zeta)) | \leq \| f - q \| = \| f - \sum_{\nu=1}^{n} \lambda_\nu^* f_\nu \|$, contradicting (1).

B. Let $f, p_1, p_2, \cdots, p_n$ be real functions with domain $S = [0, 1]$, continuous there, and assume furthermore that $p_1, p_2, \cdots, p_n$ are orthonormal on $[0, 1]$. Let Π be again the set of all linear combinations (with complex coefficients) of $p_1, p_2, \cdots p_n$. Let $\lambda_\nu^* = \int_0^1 f(x) p_\nu(x) dx$ ($\nu = 1, 2, \cdots, n$). Then $p = \sum_{\nu=1}^{n} \lambda_\nu^* p_\nu$ is a juxtafunction to f on S with respect to Π. Indeed, if $p = f$, then the last assertion follows from Lemma 1 below. We thus assume that $p(x_0) \neq f(x_0)$ for some $x_0 \in [0, 1]$. Suppose there exists a $q = \sum_{\nu=1}^{n} \lambda_\nu p_\nu$ satisfying (a), (b) and (c) of Definition 1. Then $|f(x) - q(x)| \leq |f(x) - p(x)|$ throughout $[0, 1]$, and $|f(x_0) - q(x_0)| < |f(x_0) - p(x_0)|$. Thus

$$\int_0^1 \left[f(x) - \sum_{\nu=1}^{n} Re(\lambda_\nu) p_\nu(x) \right]^2 dx < \int_0^1 \left[f(x) - \sum_{\nu=1}^{n} \lambda_\nu^* p_\nu(x) \right]^2 dx ,$$

contradicting the least squares property of the Fourier coefficients λ_ν^*.

LEMMA 1. *Let S and Π be as in Definition 1 and let f be an element of Π with domain S. Then f is the unique function with domain S which is a juxtafunction to f on S with respect to Π.*

Proof. f is such a juxtafunction, since (a) and (c) of Definition 1 are mutually contradictory when p is f. If p (with domain S) belongs to Π and $p \neq f$, then $q = \frac{1}{2}(p + f)$ belongs to Π and satisfies (a), (b) and (c), so that p is not a juxtafunction to f on S with respect to Π.

THEOREM 1.

Hypotheses.

1. $S (\neq \emptyset)$ is a closed and bounded set in the complex plane, $f, p_1, p_2, \cdots, p_n$ are complex functions defined and continuous on[6] S.

2. *Π is the set of all complex functions defined on S which can be represented throughout S as linear combinations (with complex coefficients) of the p_ν's.*

3. *p is a juxtafunction to f on S with respect to Π, and $p(z) \neq f(z)$ throughout S.*

[6] As the domain of f may properly include S, its continuity on S means that if $a \in S$, and if $(a_j)_{j=1}^{\infty}$ is a sequence of points of S converging to a, them $\lim_{j \to \infty} f(a_j) = f(a)$. Similarly for $p_1, p_2, \cdots, p_n$ and in Lemma 2.

Conclusion. There exist distinct points $z_1, z_2, \cdots, z_m$ of S $(1 \leq m \leq 2n+1)$ and positive $\lambda_1, \lambda_2, \cdots, \lambda_m$ such that:

(I). $p(z)$ is a juxtafunction to f on $s = \{z_1, z_2, \cdots, z_m\}$ with respect to Π,

(II). *No* complex $b_1, b_2, \cdots, b_n$ exist such that $|f(z) - \sum_{\nu=1}^{n} b_\nu p_\nu(z)| < |f(z) - p(z)|$ throughout s,

(III). $\sum_{\mu=1}^{m} \lambda_\mu p_\nu(z_\mu)/\{f(z_\mu) - p(z_\mu)\} = 0,\ \nu = 1, 2, \cdots, n.$

REMARK 1. Observe that (I) is implied by (II).

For the proof of Theorem 1 we shall need two lemmas.

LEMMA 2. *Let $S(\neq \emptyset)$ be a closed and bounded set in the complex plane, and Π a set of complex functions, defined and continuous on S such that whenever $f_1 \in \Pi$, $f_2 \in \Pi$, and c_1 and c_2 are complex numbers, then $c_1 f_1 + c_2 f_2 \in \Pi$. Let f be a complex function defined and continuous on S, and let p be an element of Π such that $p(z) \neq f(z)$ throughout S. A necessary and sufficient condition for the existence of a $q \in \Pi$ satisfying throughout S*

$$|f(z) - q(z)| < |f(z) - p(z)| \tag{2}$$

is the existence of an $r \in \Pi$, satisfying throughout S

$$|f(z) - p(z) - r(z)| < |f(z) - p(z) + r(z)|\ . \tag{3}$$

Proof of Lemma 2.

Necessity. Let $r = q - p$. Then throughout S

$$|f(z) - p(z) - r(z)| < |f(z) - p(z)| < |f(z) - p(z)|\{2 - |f(z) - q(z)| \times |f(z) - p(z)|^{-1}\} \leq |2\{f(z) - p(z)\} - \{f(z) - q(z)\}| = |f(z) - p(z) + r(z)|\ .$$

Sufficiency. We use the fact that if a, b are arbitrary complex numbers, the inequalities $|a - b| < |a + b|$, $Re(b\,\bar{a}) > 0$, are equivalent. Since throughout S

$$Re[r(z)/\{f(z) - p(z)\}] = |f(z) - p(z)|^{-2}\ Re[r(z)\overline{\{f(z) - p(z)\}}] > 0\ ,$$

we have there $\alpha\,|r(z)/\{f(z) - p(z)\}|^2 < 2Re[r(z)/\{f(z) - p(z)\}]$ where $\alpha = \min\,[|\{f(z) - p(z)\}/r(z)|^2 Re(r(z)/\{f(z) - p(z)\}),\ z \text{ on } S]$. Let $q = p + \alpha r$. Then throughout S,

$$|f(z) - q(z)| = |f(z) - p(z)|\,|1 - \alpha r(z)\{f(z) - p(z)\}^{-1}| = |f(z) - p(z)| \times [1 + \alpha^2|r(z)/(f(z) - p(z))|^2 - 2\alpha Re\{r(z)(f(z) - p(z))^{-1}\}]^{1/2} < |f(z) - p(z)|\ .$$

LEMMA 3. *Let the Hypotheses* 1, 2 *of Theorem* 1 *hold, and let* p *be an element of* Π *such that* $p(z) \neq f(z)$ *throughout* S. *For every* $z \in S$, *let* $F(z)$ *denote the point* $(x_1(z), y_1(z), x_2(z), y_2(z), \cdots, x_n(z), y_n(z))$ *of the* (*real*) *Euclidean* $2n$-*space* E_{2n}, *where* $x_\nu(z)$ *is the real part and* $y_\nu(z)$ *the imaginary part of* $p_\nu(z)\,\overline{\{f(z) - p(z)\}}$. *A necessary and sufficient condition for the existence of a* $q \in \Pi$ *satisfying* (2) *throughout* S, *is that the point* $\Omega_{2n} = (0, 0, \cdots 0)$ *of* E_{2n} *does not belong to the convex hull* H *of*[7] $F(s)$.

Proof of Lemma 3.

Necessity. By Lemma 2 there exists an $r \in \Pi$ such that (3), i.e. the inequality

$$Re[r(z)\overline{\{f(z) - p(z)\}}] > 0 \tag{3a}$$

holds throughout S. Let $s_1, t_1, s_2, t_2, \cdots, s_n, t_n$ be reals such that throughout S, $r(z) = \sum_{\nu=1}^{n}(s_\nu - it_\nu)p_\nu(z)$. Then throughout S we have

$$\sum_{\nu=1}^{n} s_\nu x_\nu(z) + t_\nu y_\nu(z) > 0 \tag{4}$$

and thus $F(s)$ is a subset of the half-space

$$s_1x_1 + t_1x_2 + \cdots + s_nx_{2n-1} + t_nx_{2n} > 0\ . \tag{5}$$

Therefore H is also a subset of this half-space, and consequently $\Omega_{2n} \notin H$.

Sufficiency. Since H is compact and $\Omega_{2n} \notin H$, we can find a half-space (5) containing $F(S)$. Thus (4) holds for every $z \in S$. Setting $r = \sum_{\nu=1}^{n}(s_\nu - it_\nu)p_\nu$, we have throughout S, (3a), and therefore (3). Thus, by Lemma 2, there exists a $q \in \Pi$ satisfying (2) throughout S.

Proof of Theorem 1. f cannot belong to Π, for otherwise, by Lemma 1, the restrictions of f and of p to S would coincide, contradicting Hypothesis 3. By Definition 1, there does not exist a $q \in \Pi$ satisfying (2) throughout S. Using notations of the last lemma, it follows that $\Omega_{2n} \in H$. By a well known theorem of Carathéodory there exist in $F(S)$ distinct points $A_1, A_2, \cdots, A_m$ $(m \leq 2n + 1)$ and there exist positive $\Lambda_1, \Lambda_2, \cdots, \Lambda_m$ such that

$$\Omega_{2n} = \sum_{\mu=1}^{m} \Lambda_\mu A_\mu\ . \tag{6}$$

Let

[7] $F(s)$ is, as usual, the set of all $F(z)$, $z \in S$.

$$A_\mu = F(z_\mu)\ ,\ z_\mu \in S \qquad (\mu = 1, 2, \cdots, m). \tag{7}$$

Then the z_μ are distinct, and from (6) we get by taking components,

$$\sum_{\mu=1}^{m} \Lambda_\mu p_\nu(z_\mu)\{\overline{f(z_\mu) - p(z_\mu)}\} = 0 \qquad (\nu = 1, 2, \cdots, n)\ . \tag{8}$$

Thus

$$\sum_{\mu=1}^{m} \lambda_\mu p_\nu(z_\mu)/\{f(z_\mu) - p(z_\mu)\} = 0 \qquad (\nu = 1, 2, \cdots, n)$$

where $\lambda_\mu = \Lambda_\mu |f(z_\mu) - p(z_\mu)|^2 > 0$ $(\mu = 1, 2, \cdots, m)$. Let $s = \{z_1, z_2, \cdots, z_m\}$, and let π be the set of all functions defined on s which can be represented throughout s as linear combinations (with complex coefficients) of the p_ν. Obviously $p \in \pi$, since $p \in \Pi$. From (6) and (7) it follows that Ω_{2n} belongs to the convex hull of $F(s)$ and therefore, by Lemma 3 (taking there s in place of S and π in place of Π) there does not exist a $q \in \pi$ satisfying (2) throughout s. This concludes the proof.

REMARK 2. Suppose that one of the p_ν in Theorem 1 equals throughout S a constant $c(\neq 0)$. Then from (8) we obtain $\sum_{\mu=1}^{m} \Lambda_\mu\{f(z_\mu) - p(z_\mu)\} = 0$. Thus 0 belongs to the convex hull of the image of s (and a fortiori of S) under $f - p$. [Compare Motzkin and Walsh 1953, § 2, and Fekete 1955, § 18].

REMARK 3. Let $s' = \{z_1, z_2, \cdots, z_M\}$ be a finite set in the complex plane and suppose that $f, p_1, p_2, \cdots, p_n$ are complex functions defined on s'. Let π' be the set of all complex functions representable throughout s' as a linear combination with complex coefficients of $p_1, p_2, \cdots, p_n$. Let p be an element of π' such that $p(z) \neq f(z)$ throughout s', and suppose there exist nonnegative reals $\lambda'_1, \cdots, \lambda'_M$ (not all zero) such that

$$\sum_{\mu=1}^{M} \lambda'_\mu p_\nu(z_\mu)/\{f(z_\mu) - p(z_\mu)\} = 0 \quad (\nu = 1, 2, \cdots, n)\ .$$

Then there does not exist a $q \in \pi'$ such that (2) holds throughout s'. Indeed, we have

$$\sum_{\mu=1}^{M} \Lambda'_\mu p_\nu(z_\mu)\{\overline{f(z_\mu) - p(z_\mu)}\} = 0 \quad (\nu = 1, 2, \cdots, n)$$

where Λ'_μ are nonnegative reals, not all zero. Therefore (using notations of Lemma 3) Ω_{2n} belongs to the convex hull of $F(s')$. By Lemma 3, there does not exist a $q \in \pi'$ satisfying (2) throughout s'. Consequently, p is a juxtafunction to f on s' with respect to π'.

THEOREM 2. *Let the hypotheses of Theorem* 1 *hold and suppose furthermore that* $f - p, p_1, p_2, \cdots ; p_n$ *are real valued throughout* S. *Then the inequality* $1 \leq m \leq 2n + 1$ *in the conclusion of Theorem* 1 *can be replaced by* $1 \leq m \leq n + 1$.

Theorem 2 is proved with the aid of the following lemma, in the same way that Theorem 1 was proved with the aid of Lemma 3.

LEMMA 4. *Let the hypotheses* 1, 2 *of Theorem* 1 *hold, let* p *be an element of* Π *such that* $f(z) \neq p(z)$ *throughout* S, *and suppose that* $f - p, p_1, p_2, \cdots, p_n$ *are real throughout* S. *For every* $z \in S$, *let* $F_1(z)$ *denote the point* $(p_1)(z)\{f(z) - p(z)\},\ p_2(z)\{f(z) - p(z)\}, \cdots, p_n(z)\{f(z) - p(z)\})$ *of the (real) Euclidean n-space* E_n. *A necessary and sufficient condition for the existence of a* $q \in \Pi$ *satisfying* (2) *throughout* S, *is that the point* $\Omega_n = (0, 0, \cdots, 0)$ *of* E_n *does not belong to the convex hull of* $F_1(S)$.

The proof of the last lemma is analogous to that of Lemma 3.

We shall make frequent use of the concept of unisolvence. We mention therefore the following

DEFINITION 2. *Let* S *be a set in the complex plane, and* $(p_\nu(z))_{\nu=1}^n$ *a finite sequence of complex functions defined on* S. *The sequence will be called unisolvent on* S *if and only if for every complex* $c_1, c_2, \cdots, c_n$ *(not all zero) the set of all* $z \in S$ *for which* $\sum_{\nu=1}^n c_\nu p_\nu(z) = 0$, *contains less than* n *points.*

REMARK 4. Thus $(p_\nu(z))_{\nu=1}^n$ is unisolvent on S if and only if this sequence is linearly independent on every n-point subset of S. A simple example is the sequence $(z^{\nu-1})_{\nu=1}^n$, which is unisolvent on every subset of the complex plane. A unisolvent sequence has been termed also (for an important particular case) a "Tchebycheff system". Other terms used in this connection are "Haar system" and "interpolational system".

THEOREM 3. *Let the hypotheses of Theorem* 1 *hold and suppose that each of the sequences* $(p_\nu(z))_{\nu=1}^j (j = 1, 2, \cdots, n)$ *is unisolvent on* S. *Then the inequalities*

$$1 \leq m \leq 2n + 1 \tag{9}$$

in Theorem 1, *can be replaced by the sharper estimate* $n + 1 \leq m \leq 2n + 1$. *Furthermore, if the additional hypothesis of Theorem* 2 *is* *made too,* (9) *can be replaced by* $m = n + 1$.

Proof. Choose distinct points $z_1, z_2, \cdots, z_m$ of S and positive $\lambda_1, \lambda_2, \cdots, \lambda_m$ such that (I), (II) and (III) of Theorem 1 hold, where $1 \leq m \leq 2n + 1$ and where, furthermore, $1 \leq m \leq n + 1$ in case the additional hypothesis of Theorem 2 holds. We shall prove that $n + 1 \leq m$. Indeed: suppose $m \leq n$. Then since $(p_\nu(z))_{\nu=1}^m$ is unisolvent on S, the determinant whose jth row is $p_1(z_j)\ p_2(z_j) \cdots p_m(z_j)$ is different from zero. Therefore there exist constants $c_1, \cdots, c_m$ such that $f(z) = \sum_{\nu=1}^n c_\nu p_\nu(z)$ throughout s. Let π have the same meaning as in the proof of Theorem 1; then $f \in \pi$. By Theorem 1, (II), p is a juxtafunction to f on s with respect to π. By Lemma 1 (with S replaced by s, Π by π, and f by the restriction of our f to s) we have $f(z) = p(z)$ throughout s, contradicting hypothesis 3 of Theorem 1.

2. We apply now Theorems 1, 2 and 3 to nth infrapolynomials (cf. the Introduction).

THEOREM 4. *Let* n *and* q *be natural numbers* $(q \leq n)$, $n_1, n_2, \cdots, n_q$ *integers such that* $0 \leq n_1 < n_2 \cdots < n_q \leq n$, *and* S *a closed and bounded set in the complex plane. Let* $A(z)$ $(\neq 0$ *throughout* $S)$ *be an* nth *infrapolynomial on* S *with respect to* $(n_1, \cdots, n_q)$. *Then*[8] *there exist distinct points* $z_1, z_2, \cdots, z_m$ *of* S,

$$1 \leq m \leq 2(n - q) + 3 \tag{10}$$

and positive $\lambda_1, \lambda_2, \cdots, \lambda_m$ *such that* $A(z)$ *is an* nth *infrapolynomial on* $s = \{z_1, z_2, \cdots, z_m\}$ *with respect to* $(n_1, n_2, \cdots, n_q)$ *and such that*

$$\sum_{\mu=1}^m \lambda_\mu z_\mu^{l_\nu} / A(z_\mu) = 0 \qquad (\nu = 1, 2, \cdots, n + 1 - q) \tag{11}$$

where $l_1, l_2, \cdots, l_{n+1-q}$ $(l_1 < l_2 \cdots < l_{n+1-q})$ *are the elements of* $\{0, 1, \cdots, n\} - \{n_1, n_2, \cdots, n_q\}$. *If the polynomials* $A(z), z^{l_1}, \cdots, z^{l_{n+1-q}}$ *are real valued throughout* S, *then* (10) *can be replaced by* $1 \leq m \leq n + 2 - q$. *If each of the sequences* $(z^{l_\nu})_{\nu=1}^j$ $(j = 1, 2, \cdots, n + 1 - q)$ *is unisolvent on* S, *then* (10) *can be replaced by*

$$n - q + 2 \leq m \leq 2(n - q) + 3\,. \tag{12}$$

If the polynomials $A(z)$, $z^{l_1}, \cdots, z^{l_{n+1-q}}$ *are real valued throughout* S *and each of the sequences* $(z^{l_\nu})_{\nu=1}^j$ $(j = 1, 2, \cdots, n + 1 - q)$ *is unisolvent on* S, *then* (10) *can be replaced by* $m = n - q + 2$.

REMARK 5. If $(n_1, n_2, \cdots, n_q)$ of Theorem 4 is a simple n-sequ-

[8] As is easily seen, S cannot be empty. [Cf. Shisha and Walsh, 1961, footnote 7 on p. 117].

ence (cf. the Introduction) and if, in case $n_1 = 0$, $0 \notin S$, then as is easily seen, the sequences $(z^{l_\nu})_{\nu=1}^{j}$ ($j = 1, 2, \cdots, n+1-q$) are unisolvent on S.

Proof of Theorem 4. Let Π be the set of all complex functions defined on S which are expressible throughout S as linear combinations of $z^{l_1}, z^{l_2}, \cdots, z^{l_{n+1-q}}$ with complex coefficients, and let $f(z) \equiv \sum_{\nu=1}^{q} a_{n_\nu} z^{n_\nu}$, $p(z) \equiv -\sum_{\nu=1}^{n+1-q} a_{l_\nu} z^{l_\nu}$. It is easily seen that $p(z)$ is a juxtafunction to f on S with respect to Π. Therefore, by Theorem 1 there exist distinct points $z_1, \cdots, z_m$ ($m \leq 2(n+1-q)+1 = 2(n-q)+3$) of S and positive $\lambda_1, \lambda_2, \cdots, \lambda_m$ such that (11) holds, and such that no complex $b_1, b_2, \cdots, b_{n+1-q}$ exist satisfying

$$\left| \sum_{\nu=1}^{q} a_{n_\nu} z^{n_\nu} - \sum_{\nu=1}^{n+1-q} b_\nu z^{l_\nu} \right| < \left| A(z) \right|$$

throughout $s = \{z_1, z_2, \cdots, z_m\}$. Thus $A(z)$ is an nth infrapolymial on s with respect to $(n_1, n_2, \cdots, n_q)$. The rest of Theorem 4 follows from Theorems 2 and 3.

REMARK 6. Let $n, n_1, n_2, \cdots, n_q$ be integers ($q \leq n$, $0 \leq n_1 < n_2 \cdots < n_q \leq n$), $A(z) \equiv \sum_{\nu=0}^{n} a_\nu z^\nu$ a polynomial, $z_1, z_2, \cdots, z_M$ points of the complex plane, and $\lambda'_1, \lambda'_2, \cdots, \lambda'_M$ ($\sum_{\mu=1}^{M} \lambda'_\mu > 0$) nonnegative reals such that $A(z_\mu) \neq 0$ ($\mu = 1, 2, \cdots, M$), and such that $\sum_{\mu=1}^{M} \lambda'_\mu z_\mu^{l_\nu} / A(z_\mu) = 0$ ($\nu = 1, 2, \cdots, n+1-q$), where the l_ν have the same meaning as in Theorem 4. Then $A(z)$ is an nth infrapolynomial on $s' = \{z_1, z_2, \cdots, z_M\}$ with respect to $(n_1, n_2, \cdots, n_q)$. Indeed: let f and p be as in the last proof, and let π' be the set of all complex functions representable throughout s' as a linear combination (with complex coefficients) of $z^{l_1}, z^{l_2}, \cdots, z^{l_{n+1-q}}$. The asserted conclusion follows from Remark 3.

We give now the following structure theorem which is the main result of this paper.

THEOREM 5. *Let n and q ($1 \leq q \leq n$) be integers, and σ a simple n-sequence of q elements. Let S be a closed and bounded set in the complex plane, and in case $0 \in \sigma$, assume that $0 \notin S$. Let $A(z)$ ($\not\equiv 0$) be an nth infrapolynomial on S with respect to σ, and let $B(z)$ ($\neq 0$ throughout S) be a divisor of $A(z)$. Assume also that the degree*[9] *r of $B(z)$ is $\geq q$. Then $B(z)$ is a divisor of some*

$$Q(z) \equiv P(z)g(z) + z^K \sum_{\mu=1}^{M-q+2} \lambda_\mu g(z)/(z - z_\mu)\ . \tag{13}$$

[9] By degree of a polynomial ($\not\equiv 0$) we mean its exact degree. The polynomial 0 is assigned the degree-1.

Here M is an integer satisfying $r \leq M \leq 2r - q + 1$, the z_ν are distinct points of S, $g(z) \equiv \prod_{\mu=1}^{M-q+2}(z - z_\mu)$, the λ_μ are positive reals with $\sum_{\mu=1}^{M-q+2}\lambda_\mu = 1$, $P(z)$ is a polynomial of degree $\leq q - 1$ such that $P(z)g(z) + z^{K+M-q+1}$ is of degree $\leq M$, and K is $\min[\nu, \nu \notin \sigma, \nu = 0, 1, 2, \cdots]$.

REMARK 7. As will be seen from the proof of Theorem 5, if S and the coefficients of $B(z)$ are real, the inequality $r \leq M \leq 2r - q + 1$ of the theorem can be replaced by the equality $M = r$.

In the proof of Theorem 5 use will be made of the following

LEMMA 5. *Let n, q, σ and K be as in the last theorem, let S be a set in the complex plane, and let $A(z)$ $(\not\equiv 0)$ be an* nth *infrapolynomial on S with respect to σ. Let $B(z)$ be a polynomial of degree $r(\geq q)$ dividing $A(z)$. Then $B(z)$ is an* rth *infrapolynomial on S with respect to σ_0, where σ_0 is that simple r-sequence of q elements for which* $K = \min[\nu, \nu \notin \sigma_0, \nu = 0, 1, 2, \cdots]$.

The proof of Lemma 5 is straightforward and may be omitted.

Proof of Theorem 5. By Lemma 5, $B(z)$ in an rth infrapolynomial on S with respect to the sequence σ_0 defined there. We choose (cf. Theorem 4 and Remark 5) distinct points $z_1, z_2, \cdots, z_m$ of S and positive $\lambda_1, \lambda_2, \cdots, \lambda_m$ such that $\sum_{\mu=1}^{m}\lambda_\mu = 1$ and

$$\sum_{\mu=1}^{m}\lambda_\mu z_\mu^\rho / B(z_\mu) = 0 \tag{14}$$

for every integer ρ satisfying $0 \leq \rho \leq r$, $\rho \notin \sigma_0$. Here m is an integer satisfying $r - q + 2 \leq 2(r - q) + 3$, and in case S and the coefficients of $B(z)$ are real we may take $m = r - q + 2$. Set

$$g(z) \equiv \prod_{\mu=1}^{m}(z - z_\mu), \quad N(z) \equiv \sum_{\mu=1}^{m}\lambda_\mu z_\mu^{r-q+K+1} g(z)/\{B(z_\mu)(z - z_\mu)\}\,. \tag{15}$$

If μ and ν are integers, $1 \leq \mu \leq m$, $0 \leq \nu \leq r - q + K$, then

$$[\lambda_\mu z_\mu^{r-q+K+1} g(z)/\{B(z_\mu)(z - z_\mu)\}]_{z=0}^{(\nu)} = -\sum_{j=0}^{\nu}\lambda_\mu z_\mu^{r-q+K-j}\binom{\nu}{j} j\,!\, g^{(\nu-j)}(0)/B(z_\mu)$$

(the equality is obvious if $z_\mu = 0$, and otherwise it is obtained by Leibnitz's rule for differentiating a product). Therefore, from (15) we get

$$N^{(\nu)}(0) = -\sum_{j=0}^{\nu}\binom{\nu}{j} j\,!\, g^{(\nu-j)}(0)\sum_{\mu=1}^{m}\lambda_\mu z_\mu^{r-q+K-j}/B(z_\mu) \tag{16}$$
$$(\nu = 0, 1, \cdots, r - q + K)\,.$$

Since $\{0, 1, \cdots, r\} - \{\sigma_0\} = \{r - q + K - j\}_{j=0}^{r-q}$, therefore (16) and (14) yield $N^{(\nu)}(0) = 0$, $\nu = 0, 1, \cdots, r - q$. Hence we can write $N(z) \equiv z^{r-q+1}M_1(z)$ where $M_1(z)$ is a polynomial (of degree $\leq m - 2$). Let

$$M_2(z) \equiv \sum_{\mu=1}^{m} \lambda_\mu z_\mu^K g(z)/\{B(z_\mu)(z - z_\mu)\} .$$

By (14),

$$\sum_{\mu=1}^{m} \lambda_\mu z_\mu^K / B(z_\mu) = 0$$

and therefore the degree of $M_2(z)$ is $\leq m - 2$. For every z_j different from zero we have by (15), $M_1(z_j) = z_j^{-r+q-1}N(z_j) = \lambda_j z_j^K g'(z_j)/B(z_j) = M_2(z_j)$. Since there are at least $m - 1$ such z_j, we have $M_1(z) \equiv M_2(z)$. Consider now the polynomial

$$R(z) \equiv B(z)M_2(z) - \sum_{\mu=1}^{m} \lambda_\mu z_\mu^K g(z)/(z - z_\mu) .$$

For $j = 1, 2, \cdots, m$ we have $R(z_j) = B(z_j)M_2(z_j) - \lambda_j z_j^K g'(z_j) = 0$. Therefore we can write $R(z) \equiv g(z)U(z)$, where $U(z)$ is some polynomial. Also, the relation $N(z) \equiv z^{r-q+1}M_2(z)$ and the definition of $R(z)$ imply that the degree of the latter is $\leq m + q - 2$. Therefore the degree of $U(z)$ is at most $q - 2$. If $K \geq 1$, then the relation

$$B(z)M_2(z) \equiv g(z)U(z) + \sum_{\mu=1}^{m} \lambda_\mu z_\mu^K g(z)/(z - z_\mu)$$

yields, upon putting $z_\mu^K = [z + (z_\mu - z)]^K$ and developing the last right member,

$$B(z)M_2(z) \equiv g(z)[U(z) + A_{K-1}(z)] + z^K \sum_{\mu=1}^{m} \lambda_\mu g(z)/(z - z_\mu) ,$$

where $A_{K-1}(z)$ is a polynomial of degree $K - 1$. The last relation (with $A_{K-1}(z) \equiv 0$) holds also when $K = 0$. We set now $P(z) \equiv U(z) + A_{K-1}(z)$, and get that $B(z)$ is a divisor of

$$Q(z) \equiv P(z)g(z) + z^K \sum_{\mu=1}^{m} \lambda_\mu g(z)/(z - z_\mu) .$$

The degree of $Q(z)$, i. e. of $B(z)M_2(z)$, is $\leq m + q - 2$. Thus the degree of $P(z)$ is $\leq q - 1$, and that of $P(z)g(z) + z^{K+m-1}$ is $\leq m + q - 2$. We set now $M = m + q - 2$, and observe that the conclusions of the theorem are all satisfied.

REMARK 8. The polynomial $Q(z)$ of (13) is an Mth infrapolynomial on $\{z_1, z_2, \cdots, z_{M-q+2}\}$ with respect to σ_1, where σ_1 is that simple

M-sequence of q elements for which $\min[\nu, \nu \notin \sigma_1, \nu = 0, 1, 2, \cdots] = K$. This follows from Theorem 1 of Shisha and Walsh [1961].

THEOREM 6. *Let S be a closed and bounded set in the complex plane, $A(z) \equiv \sum_{\nu=0}^{n} a_\nu z^\nu$ $(n \geq 1,\ a_n \neq 0)$ an nth infrapolynomial on S with respect to $(n-1)$, and suppose that $A(z) \neq 0$ throughout S. Then:*

(a) *Every zero ζ of $A(z)$ is of the form*

$$c(\zeta) - \lambda(\zeta)[a_{n-1}/a_n] \tag{17}$$

where $c(\zeta)$ belongs to the convex hull of S and where $0 \leq \lambda(\zeta) \leq 1$.[10]

(b) *Suppose that S lies in a closed disc $C: |z - a| \leq r$ (≥ 0). Then all zeros of $A(z)$ belong to $C \cup C_1$, where C_1 is the closed disc $|z - [a - (a_{n-1}/a_n)]| \leq r$. If C and C_1 are disjoint then $A(z)$ has at least $n-1$ zeros belonging to C. [Multlplicities are always being counted].*

Proof. We choose distinct points $z_1, z_2, \cdots, z_m$ of S and positive $\lambda_1, \lambda_2, \cdots, \lambda_m$ $(m \leq 2n+1)$ such that $\sum_{\mu=1}^{m} \lambda_\mu = 1$ and $\sum_{\mu=1}^{m} \lambda_\mu z_\mu^\rho / A(z_\mu) = 0$ for all integers ρ with $0 \leq \rho \leq n$, $\rho \neq n-1$. Then $1 = \sum_{\mu=1}^{m} \lambda_\mu A(z_\mu)/A(z_\mu) = \sum_{\mu=1}^{m} \lambda_\mu a_{n-1} z_\mu^{n-1}/A(z_\mu)$, and so

$$\sum_{\mu=1}^{m} \lambda_\mu z_\mu^{n-1}/A(z_\mu) = 1/a_{n-1}\,.$$

We set $g(z) \equiv \prod_{\mu=1}^{m}(z - z_\mu)$, $N(z) \equiv \sum_{\mu=1}^{m} \lambda_\mu z_\mu^{n-1} g(z)/\{A(z_\mu)(z - z_\mu)\} \equiv a_{n-1}^{-1} z^{m-1} + \cdots$. We follow the proof of Theorem 5 from the sentence following (15). Again we have $N^{(\nu)}(0) = 0$ for every ν satisfying $0 \leq \nu \leq n-2$. Thus we may set $N(z) \equiv z^{n-1} M_1(z)$, where $M_1(z) \equiv a_{n-1}^{-1} z^{m-n} + \cdots$ is some polynomial. Let $M_2(z) \equiv \sum_{\mu=1}^{m} \lambda_\mu g(z)/\{A(z_\mu)(z - z_\mu)\}$. If $n = 1$, then $M_2(z) \equiv N(z) \equiv M_1(z)$. If $n > 1$ then for each z_j different from zero, $M_1(z_j) = \lambda_\mu g'(z_j)/A(z_j) = M_2(z_j)$, and since there are at least $m-1$ such z_j and $M_1(z)$ and $M_2(z)$ are of degrees $\leq m-2$, we have again $M_2(z) \equiv M_1(z)$. Consider now the polynomial $R(z) \equiv A(z) M_2(z) - \sum_{\mu=1}^{m} \lambda_\mu g(z)/(z - z_\mu) \equiv (a_n/a_{n-1}) z^m + \cdots$ For $j = 1, 2, \cdots, m$, $R(z_j) = 0$, and therefore $R(z) \equiv (a_n/a_{n-1}) g(z)$. Thus, $A(z)$ is a divisor of $Q(z) \equiv (a_n/a_{n-1}) g(z) + \sum_{\mu=1}^{m} \lambda_\mu g(z)/(z - z_\mu)$. Let ζ be a zero of $A(z)$. Then $g(\zeta) \neq 0$, and thus $a_n/a_{n-1} + \sum_{\mu=1}^{m} \lambda_\mu/(\zeta - z_\mu) = 0$. Since $\sum_{\mu=1}^{m} \lambda_\mu/(\zeta - z_\mu)$ can be written [Shisha and Walsh 1961, Lemma on p. 127] as $\lambda(\zeta)/(\zeta - c(\zeta))$ where $c(\zeta)$ and $\lambda(\zeta)$ are as required in (a) of our theorem, ζ is of the form (17). Suppose now that S lies in a closed disc $C: |z - a| \leq r(\geq 0)$. Then by a theorem due to J. L. Walsh [cf. 1922, Theorem VI; see also Shisha and Walsh 1961, p. 147] all zeros of $Q(z)$ lie in

[10] Thus ζ belongs to the set swept by the convex hull of S while being displaced, the displacement being given by the vector $-a_{n-1}/a_n$.

$C \cup C_1$, and if C and C_1 are disjoint, the number of zeros of $Q(z)$ in them is, respectively, $m-1$ and 1. From this follow the conclusions of part (b) of our theorem.

BIBLIOGRAPHY

1. M. Fekete 1951. *On the structure of extremal polynomials,* Proc. Natl. Acad. Sci. U.S. **37**, pp. 95-103. 1955. *On the structure of polynomials of least deviation,* Bull. Research Council Israel **5**A, pp. 11-19.

2. T. S. Motzkin and J. L. Walsh 1953. *On the derivative of a polynomial and Chebychev approximation,* Proc. Amer. Math. Soc. **4**, pp. 76-87.

3. O. Shisha 1958. *Properties of extremal systems of points* (mimeographed), Ph. D. thesis. 1958a. *Nearest functions (abstract)*, Notices Amer. Math. Soc. **5**, No. 7, p. 853. 1959. *On generalized infrapolynomials of a certain type* (abstract), Ibid. **6**, No. 3, p. 283. 1961. *A structure theorem for infrapolynomials with prescribed coefficients* (abstract), Ibid. **8**, No. 1, p. 66.

4. O. Shisha and J. L. Walsh 1961. *The zeros of infrapolynomials with some prescribed coefficients,* J. d'analyse Math. **9**, pp. 111-160.

5. J. L. Walsh 1922. *On the location of the roots of certain types of polynomials,* Trans. Amer. Math. Soc. **24**, pp. 163-180. 1958. *On infrapolynomials with prescribed constant term,* J. Math. Pures Appl. **37**, pp. 295-316.

6. J. L. Walsh and T. S. Motzkin 1957. *Polynomials of best approximation on a real finite point set,* Proc. Natl. Acad. Sci. U. S. **43**, pp. 845-846.

7. J. L. Walsh and M. Zedek 1956. *On generalized Tchebycheff polynomials,* Ibid. **42**, pp. 99-104.

AEROSPACE RESEARCH LABORATORIES, WRIGHT-PATTERSON AIR FORCE BASE
HARVARD UNIVERSITY AND
NATIONAL BUREAU OF STANDARDS

ON COMPARABLE MEANS

O. SHISHA AND G. T. CARGO

1. Let $-\infty < a < b < \infty$, and let Φ denote the set of all functions, continuous and strictly monotone in $[a, b]$. For every $\varphi \in \Phi$, every positive integer n, every $x_1, x_2, \cdots, x_n$ of $[a, b]$, and every positive $q_1, q_2, \cdots, q_n$ with $\sum_{\nu=1}^{n} q_\nu = 1$, we consider the mean

$$M_\varphi(x_1, x_2, \cdots, x_n \mid q_1, q_2, \cdots, q_n) = \varphi^{-1}(\textstyle\sum_{\nu=1}^{n} q_\nu \varphi(x_\nu)) .$$

Let ψ and χ be elements of Φ. We write

$$(1) \qquad M_\psi \leqq M_\chi$$

if and only if the inequality $M_\psi(x_1, x_2, \cdots, x_n \mid q_1, q_2, \cdots, q_n) \leqq M_\chi(x_1, x_2, \cdots, x_n \mid q_1, q_2, \cdots, q_n)$ holds for every $n \geqq 1$, every $x_1, x_2, \cdots, x_n$ of $[a, b]$, and every positive $q_1, q_2, \cdots, q_n$ with $\sum_{\nu=1}^{n} q_\nu = 1$.

A well-known necessary and sufficient condition for (1) to hold is that $\chi(\psi^{-1}(x))$ be convex in $[\psi(a), \psi(b)]$ (or $[\psi(b), \psi(a)]$) if χ is increasing, and that $\chi(\psi^{-1}(x))$ be concave there if χ is decreasing.

It is not difficult to see that (1) holds if and only if $M_\psi(x_1, x_2 \mid q_1, q_2) \leqq M_\chi(x_1, x_2 \mid q_1, q_2)$ for every x_1, x_2 of $[a, b]$ and every positive q_1, q_2 with $q_1 + q_2 = 1$, which in turn holds if and only if $M_\psi(x_1, x_2 \mid 1/2, 1/2) \leqq M_\chi(x_1, x_2 \mid 1/2, 1/2)$ for every x_1, x_2 of $[a, b]$.

Similarly, we write

$$(2) \qquad M_\psi < M_\chi$$

if and only if the inequality

$$M_\psi(x_1, x_2, \cdots, x_n \mid q_1, q_2 \cdots, q_n) < M_\chi(x_1, x_2, \cdots, x_n \mid q_1, q_2 \cdots, q_n)$$

holds for every $n \geqq 2$, every $x_1, x_2, \cdots, x_n$ (not all equal) of $[a, b]$, and every positive $q_1, q_2, \cdots, q_n$ with $\sum_{\nu=1}^{n} q_\nu = 1$. A necessary and sufficient condition for (2) to hold is that $\chi(\psi^{-1}(x))$ be strictly convex in $[\psi(a), \psi(b)]$ (or $[\psi(b), \psi(a)]$) if χ is increasing, and that $\chi(\psi^{-1}(x))$ be strictly concave there if χ is decreasing. Also, (2) holds if and only if $M_\psi(x_1, x_2 \mid q_1, q_2) < M_\chi(x_1, x_2 \mid q_1, q_2)$ for every x_1, x_2 ($\neq x_1$) of $[a, b]$ and every positive q_1, q_2 with $q_1 + q_2 = 1$, which in turn holds if and only if $M_\psi(x_1, x_2 \mid 1/2, 1/2) < M_\chi(x_1, x_2 \mid 1/2, 1/2)$ for every x_1 and x_2 ($\neq x_1$) of $[a, b]$.

2. In this paper we give simple criteria for the validity of (1)

Received December 30, 1963. The work of the second author was supported by the National Science Foundation through grant NSF-GP 1086.

and of (2), and then we give a few applications.

THEOREM 1. *Let* ψ *and* χ *be elements of* Φ *differentiable in* (a, b), *and let* $\psi' \neq 0$ *there. A necessary and sufficient condition for* (1) *to hold is that* χ'/ψ' *be nondecreasing in* (a, b) *if* ψ *and* χ *are monotone in the same sense, and that* χ'/ψ' *be nonincreasing there if* ψ *and* χ *are monotone in opposite senses.*

Proof. Consider the function $u(x) \equiv \chi(\psi^{-1}(x))$. Let J denote the open interval joining $\psi(a)$ to $\psi(b)$, and let $\bar{J}$ be the closure of J. For every $\xi \in J$, we have

$$u'(\xi) = \chi'(\psi^{-1}(\xi))/\psi'(\psi^{-1}(\xi)) \,. \tag{3}$$

Suppose that ψ and χ are monotone in the same sense. Then (1) holds if and only if $u(x)$ is convex in $\bar{J}$ in case χ increases, and if and only if $u(x)$ is concave there in case χ decreases. So (1) holds if and only if $u'(x)$ is nondecreasing in J in case ψ increases, and if and only if $u'(x)$ is nonincreasing there in case ψ decreases. From this, with the aid of (3), one easily infers that (1) is equivalent to χ'/ψ' being nondecreasing in (a, b). Similariy one shows that (1) is equivalent to χ'/ψ' being nonincreasing in (a, b), if ψ and χ are monotone in opposite senses.

One can modify Theorem 1 by replacing in it (1) by (2), "nondecreasing" by "strictly increasing," and "nonincreasing" by "strictly decreasing."

3. Given a function ψ, one may construct by means of Riemann-Stieltjes integrals functions χ such that $M_\psi \leqq M_\chi$. In fact, we have the following

THEOREM 2. *Let* ψ *be a real function, continuous in* $[a, b]$ *and differentiable in* (a, b). *Let* $m(x)$ *be nondecreasing or nonincreasing in* $[a, b]$, *continuous in* (a, b), *and suppose* $m(x)\psi'(x) \neq 0$ *throughout* (a, b). *Let* C *be a real constant, and for every* $x \in [a, b]$ *let*

$$\chi(x) = C + \int_a^x m(t)d\psi(t) \,.$$

Then ψ *and* χ *belong to* Φ. *If* $m(x)$ *is positive in* (a, b) *and nondecreasing in* $[a, b]$, *or negative in* (a, b) *and nonincreasing in* $[a, b]$, *then* $M_\psi \leqq M_\chi$. *Otherwise,* $M_\chi \leqq M_\psi$.

Proof. Since $\psi' \neq 0$ in (a, b), by a well known property of the derivative, ψ' is either positive throughout (a, b), or negative throughout (a, b). Thus ψ is strictly monotone in $[a, b]$. Also, by well-known properties of the Riemann-Stieltjes integral, χ is continuous in $[a, b]$, and $\chi'(x) = m(x)\psi'(x)$ throughout (a, b) (and so χ is strictly monotone in $[a, b]$). If $m(x)$ is positive in (a, b) and nondecreasing in $[a, b]$, then ψ and χ are monotone in the same sense in $[a, b]$, χ'/ψ' is nondecreasing in (a, b), and hence by Theorem 1, $M_\psi \leqq M_\chi$. Similarly the rest of Theorem 2 follows.

Theorem 2 can be modified by replacing in it "nondecreasing" by "strictly increasing," "nonincreasing" by "strictly decreasing," "$M_\psi \leqq M_\chi$" by "$M_\psi < M_\chi$," and "$M_\chi \leqq M_\psi$" by "$M_\chi < M_\psi$."

4. A converse of Theorem 2 is the following

THEOREM 3. *Let ψ and χ be elements of Φ differentiable in (a, b), and suppose $\psi' \neq 0$ there. Suppose, furthermore, that $M_\psi \leqq M_\chi$. Then there exists a function $m(x)$, nondecreasing in (a, b) if ψ and χ are monotone in the same sense, and nonincreasing there if ψ and χ are monotone in opposite senses, such that throughout* $[a, b]$

$$\chi(x) = \chi(a) + \int_a^x m(t)\psi'(t)dt \qquad \text{(a Lebesgue integral)} . \tag{4}$$

Proof. For every $x \in (a, b)$, let $m(x) = \chi'(x)/\psi'(x)$. By Theorem 1, $m(x)$ has the monotonicity property steated in Theorem 3. Now for every $x \in [a, b]$

$$\chi(x) - \chi(a) = \int_a^x \chi'(t)dt = \int_a^x m(t)\psi'(t)dt$$

(cf. [5], Theorems 269 (p. 188) and 264 (p. 183)).

REMARK. Observe that the integral in (4) can be written, under appropriate conditions, as a Riemman-Stieltjes integral: $\int_a^x m(t)\,d\psi(t)$. [Cf. loc. cit, Theorem 322.1 (p. 254), and 322 (p. 253)].

Theorem 3 remains valid if we replace in it "$M_\psi \leqq M_\chi$" by "$M_\psi < M_\chi$," "nondecreasing" by "strictly increasing," and "nonincreasing" by "strictly decreasing."

5. It is known that if the end-point a is positive and $r < s$, $rs \neq 0$, then $M_{x^r} < M_{x^s}$, and $M_{x^{-|r|}} < M_{\log x} < M_{x^{|r|}}$. Consequently, if $a > 0$ then for every real r $(\neq 0, 1)$, $M_{(x^r)'} < M_{x^r}$, and $M_{(\log x)'} < M_{\log x}$. The question thus arises: Under what conditions on a function φ does one have $M_{\varphi'} < M_\varphi$ (or $M_{\varphi'} \leqq M_\varphi$)?

THEOREM 4. *A necessary and sufficient condition for a real* *function* φ *to fulfill the conditions* (α)–(γ) *below is that* $\varphi(x)$ *should be (throughout* $[a, b]$*) of one of the forms* $A + \int_a^x \exp C(t)dt$, $A - \int_a^x \exp C(t)dt$, $A + \int_a^x \exp\{-C(t)\}dt$, $A - \int_a^x \exp\{-C(t)\}dt$, *where* A *is a real number, and* $C(t)$ *is a function, continuous and convex in* $[a, b]$, *differentiable in* (a, b), *and satisfying there* $C'(x) < 0$.

(α) φ *is twice differentiable in* (a, b), $\varphi'(a)$ *and* $\varphi'(b)$ *exist as right and left hand derivatives, respectively,* $\varphi'(a)\varphi'(b) \neq 0$, *and* φ' *is continuous in* $[a, b]$.

(β) $\varphi'\varphi'' \neq 0$ *throughout* (a, b) *(and hence* φ *and* φ' *are strictly monotone in* $[a, b]$*).*

(γ) $M_{\varphi'} \leqq M_\varphi$.

Proof.

Necessity. By Theorem 1, φ'/φ'' is either positive and nondecreasing in (a, b), or negative and nonincreasing there. Thus, φ''/φ' is either positive and nonnincreasing in (a, b), or negative and nondecreasing there. In the first case we set $C(x) = -\log|\varphi'(x)|$ (in $[a, b]$). Then $C(x)$ is continuous in $[a, b]$ and $C'(x) < 0$ in (a, b). Also $C'(x)$ is nondecreasing in (a, b), and, therefore, $C(x)$ is convex in $[a, b]$. Either for every $x \in [a, b]$, $\varphi(x) = \varphi(a) + \int_a^x \exp\{-C(t)\}dt$, or for every $x \in [a, b]$, $\varphi(x) = \varphi(a) - \int_a^x \exp\{-C(t)\}dt$. In the second case, we set $C(x) = \log|\varphi'(x)|$ (in $[a, b]$). Then $C(x)$ is continuous in $[a, b]$, $C'(x) < 0$ in (a, b), and, again, $C(x)$ is convex in $[a, b]$. Either for every $x \in [a, b]$, $\varphi(x) = \varphi(a) + \int_a^x \exp C(t)dt$, of for every $x \in [a, b]$, $\varphi(x) = \varphi(a) - \int_a^x \exp C(t)dt$.

Sufficiency. (α) and (β) clearly hold. Also, by the convexity of $C(t)$, $C'(t)$ is nondecreasing in (a, b). Now, either throughout (a, b), $\varphi'/\varphi'' = \{C'(t)\}^{-1}$, or throughout (a, b), $\varphi'/\varphi'' = -\{C'(t)\}^{-1}$. In the first case, φ' and φ are monotone in opposite senses, and φ'/φ'' is nonincreasing in (a, b). In the second case, φ' and φ are monotone in the same sense, and φ'/φ'' is nondecreasing in (a, b). In either case, by Theorem 1, $M_{\varphi'} \leqq M_\varphi$.

Theorem 4 can be modified by replacing in it "convex" by "strictly convex," and "$M_{\varphi'} \leqq M_\varphi$" by "$M_{\varphi'} < M_\varphi$."

THEOREM 5. *Let* φ *be strictly monotone in* $[a, b]$ *and three-times differentiable in* (a, b). *Let* φ' *be continuous in* $[a, b]$ *(where* $\varphi'(a)$

and $\varphi'(b)$ *are right and left hand derivatives, respectively). Let* $\varphi'' \neq 0$ *throughout* (a, b). *A necessary and sufficient condition for* $M_{\varphi'} \leqq M_\varphi$ *to hold is that* $\varphi''^2 \geqq \varphi'\varphi'''$ *throughout* (a, b) *if* φ' *and* φ *are monotone in the same sense, and that* $\varphi''^2 \leqq \varphi'\varphi'''$ *throughout* (a, b) *if* φ' *and* φ *are monotone in opposite senses.*

Theorem 5 follows easily from Theorem 1 by considering the derivative of φ'/φ''.

Similarly, under the hypotheses of Theorem 5, $M_{\varphi'} < M_\varphi$ holds, if $\varphi''^2 > \varphi'\varphi'''$ throughout (a, b) and φ and φ' are monotone in the same sense, and also if $\varphi''^2 < \varphi'\varphi'''$ throughout (a, b) and φ and φ' are monotone in opposite senses.

As an example, let $a = 0$, $b = \pi/2$, $\varphi(x) \equiv \cos x$. φ and φ' are monotone in the same sense in $[0, \pi/2]$, and $\varphi''^2 = \cos^2 x > -\sin^2 x = \varphi'\varphi'''$ throughout $(0, \pi/2)$. Therefore, $M_{-\sin x} < M_{\cos x}$, i.e., $M_{\sin x} < M_{\cos x}$.

6. In a previous paper [3] the authors studied, for given positive $q_1, q_2, \cdots, q_n$ (with $\sum_{\nu=1}^n q_\nu = 1$), the ratio

$$(5)\quad \begin{cases} F(x_1, x_2, \cdots, x_n) \\ \quad = M_\chi(x_1, x_2, \cdots, x_n \mid q_1, q_2, \cdots, q_n)/M_\psi(x_1, x_2, \cdots, x_n \mid q_1, q_2, \cdots, q_n) \end{cases}$$

where $0 < a$, $\psi(x) \equiv x^r$, $\chi(x) \equiv x^s$ $(r < s,\ rs \neq 0)$.

Their purpose was to find an upper bound for F in

$$I = \{(x_1, x_2, \cdots, x_n) :\ a \leqq x_k \leqq b,\ k = 1, 2, \cdots, n\}\ .$$

A crucial step was to show that if X^* is a point of I such that $F(X^*) = \max\{F(X) : X \in I\}$, then X^* is necessarily a vertex of I. In particular, X^* cannot be an interior point of I. This last property holds under quite general conditions:

THEOREM 6. *Let* ψ *and* χ *be elements of* Φ, *differentiable in* (a, b), *and satisfying* $\psi'\chi' \neq 0$ *there. Assume* $0 \notin [a, b]$, $M_\psi < M_\chi$. *Let* $q_1, \cdots, q_n$ $(n > 1)$ *be given positive numbers with* $\sum_{\nu=1}^n q_\nu = 1$, *and let* I *be as in the last paragraph. Let* F *of* (5) *attain its maximum in* I *at a point* $X^* = (x_1^*, \cdots, x_n^*)$ *of* I. *Then* X^* *is not an interior point of* I.

Proof. Suppose that some x_j^* satisfies $a < x_j^* < b$. Then $(\partial F/\partial x_j)_{x_\nu = x_\nu^*,\ \nu=1,2,\cdots,n} = 0$, i.e.,

$$\left[\psi^{-1}\Big(\sum_{\nu=1}^n q_\nu \psi(x_\nu^*)\Big)\right]^{-2}\left[q_j\chi'(x_j^*)\psi^{-1}\Big(\sum_{\nu=1}^n q_\nu\psi(x_\nu^*)\Big)\Big/\chi'\Big(\chi^{-1}\Big(\sum_{\nu=1}^n q_\nu\chi(x_\nu^*)\Big)\Big)\right.$$
$$\left. - q_j\psi'(x_j^*)\chi^{-1}\Big(\sum_{\nu=1}^n q_\nu\chi(x_\nu^*)\Big)\Big/\psi'\Big(\psi^{-1}\Big(\sum_{\nu=1}^n q_\nu\psi(x_\nu^*)\Big)\Big)\right] = 0\ .$$

Thus

$$\chi'(x_j^*)/\psi'(x_j^*) = \left[\chi^{-1}\left(\sum_{\nu=1}^{n} q_\nu \chi(x_\nu^*)\right)\chi'\left(\chi^{-1}\left(\sum_{\nu=1}^{n} q_\nu \chi(x_\nu^*)\right)\right)\right]$$
$$\times \Big/ \left[\psi^{-1}\left(\sum_{\nu=1}^{n} q_\nu \psi(x_\nu^*)\right)\psi'\left(\psi^{-1}\left(\sum_{\nu=1}^{n} q_\nu \psi(x_\nu^*)\right)\right)\right].$$

Let C denote the right hand side of the last equality. If both x_j^* and x_k^* are interior points of $[a, b]$, then $\chi'(x_j^*)/\psi'(x_j^*) = C = \chi'(x_k^*)/\psi'(x_k^*)$, and hence, by the strict monotonicity of χ'/ψ' [see the end of § 2], $x_j^* = x_k^*$. Thus, if X^* were an interior point of I, we would have $x_1^* = x_2^* = \cdots = x_n^*$, and therefore

$$1 = F(x_1^*, x_2^*, \cdots, x_n^*) = \max\{F(X) : X \in I\} > 1 .$$

References

1. E. F. Beckenbach, *On the inequality of Kantorovich* (*abstract*), Notices, Amer. Math. Society, **10**, No. 5 (1963), 440.

2. E. F. Beckenbach and R. Bellman, *Inequalities*, Springer-Verlag (1961).

3. G. T. Cargo and O. Shisha, *Bounds on Ratios of Means*, Journal of Research of the National Bureau of Standards, Vol. 66B, No. 4, pp. 169–170 (1962).

4. G. H. Hardy, J. E. Littlewood, and G. Pólya, *Inequalities* (second edition), Cambridge University Press (1952).

5. H. Kestelman, *Modern theories of integration* (second revised edition), Dover Publications, 1960.

AEROSPACE RESEARCH LABOLATORIES, WRIGHT-PATTERSON AFB, OHIO
AND
SYRACUSE UNIVERSITY

A CHARACTERIZATION OF WEAK* CONVERGENCE

MAURICE SION

1. Introduction. Let X be a locally compact, Hausdorff space and $\{\mu_i ; i \in D\}$ be a net of Radon measures on X (in the sense of Caratheodory). The weak* or vague limit of this net is the Radon measure ν such that

$$\lim_i \int f d\mu_i = \int f d\nu$$

for every continuous function f vanishing outside some compact set. In this paper, we construct in § 3 a Radon measure φ^* from a given base $\mathscr{B}$ for the topology of X and $\liminf_i \mu_i$ and then, in § 4, we give necessary and sufficient conditions for φ^* to be the weak* limit of the μ_i. In particular, if the latter exists then it is the φ^* generated when $\mathscr{B}$ is the family of all open sets.

The measure φ^* is obtained from another measure φ by a standard regularizing process. The definition of φ easily extends to abstract spaces but that of φ^* makes essential use of the topology. Thus, it is of some importance to know when $\varphi = \varphi^*$, that is, when a measure constructed through an abstract process from the μ_i turns out to be, in the topological situation, the weak* limit of the μ_i. In Theorem 3.3 we give a condition for $\varphi = \varphi^*$ and in § 5 we give an example to show that the condition cannot be eliminated.

We refer to standard texts such as Halmos [1], Kelley [2], and Munroe [3] for the elementary properties and concepts of topology and measure theory used in this paper.

2. Notation.

2.1 ω denotes the set of natural numbers.

2.2 0 denotes both the empty set and the smallest number in ω.

2.3 μ is a Caratheodory (outer) measure on X if and only if μ is a function on the family of all subsets of X such that $\mu 0 = 0$ and

$$0 \leqq \mu A \leqq \sum_{n \in \omega} \mu B_n \leqq \infty \quad \text{whenever } A \subset \bigcup_{n \in \omega} B_n \subset X .$$

2.4 For μ *a* Caratheodory measure on X, A is μ-measurable if and only if $A \subset X$ and for evey $T \subset X$

$$\mu T = \mu(T \cap A) + \mu(T - A) .$$

2.5 For X a topological space, μ is a Radon measure on X if and

Received September 26, 1963. This work was supported by the U. S. Air Force Office of Scientific Research.

only if μ is a Caratheodory measure on X such that:

(i) open sets are μ-measurable,

(ii) if C is compact then $\mu C < \infty$,

(iii) if α is open then $\mu\alpha = \sup\{\mu C\,;\, C \text{ compact},\ C \subset \alpha\}$,

(iv) if $A \subset X$ then $\mu A = \inf\{\mu\alpha\,;\, \alpha \text{ open},\ A \subset \alpha\}$.

2.6 For X a topological space, $C_0(X)$ is the family of all real-valued continuous functions on X vanishing outside some compact set.

2.7 $(D, <)$ is a directed set if and only if $D \neq 0$, D is partially ordered by $<$ so that for any $i, j \in D$ there exists $k \in D$ with $i < k$ and $j < k$.

2.8 A net is a function on a directed set.

2.9 $\bar{A}$ denotes the closure of A.

3. **The lim inf measure.** Let X be a regular topological space; $\mathscr{B}$ be a base for the topology of X, closed under finite unions and intersections; $(D, <)$ be a directed set and, for each $i \in D$, μ_i be a Radon measure on X.

For each $\alpha \in \mathscr{B}$, let

$$g\alpha = \varliminf_{i \in D} \mu_i\alpha\ (= \sup_{j \in D} \inf_{\substack{i \in D \\ j<i}} \mu_i\alpha) < \infty$$

and let φ be the Caratheodory measure on X generated by g and $\mathscr{B}$ (see method I of Munroe [3]), i.e. for each $A \subset X$,

$$\varphi A = \inf\left\{\sum_{\alpha \in H} g\alpha\,;\ H \text{ countable},\ H \subset \mathscr{B},\ A \subset \bigcup_{\alpha \in H} \alpha\right\}.$$

As we show in § 5, φ need not be a Radon measure even when X is compact and Hausdorff. For this reason, for any $A \subset X$ let

$$\varphi^* A = \inf_{\substack{\alpha \text{ open} \\ A \subset \alpha}}\ \sup_{\substack{C \text{ compact} \\ C \subset \alpha}} \varphi C\,.$$

We then have the following:

3.1 THEOREM. *φ is a Caratheodory measure on X such that:*

(i) *if A and B are disjoint, closed, compact sets then $\varphi(A \cup B) = \varphi A + \varphi B$.*

(ii) *if $A \subset X$ then $\varphi A = \inf\{\varphi\alpha;\ \alpha \text{ open},\ A \subset \alpha\}$.*

(iii) *if C is compact and for every $\alpha \in \mathscr{B}$, $g\alpha = \lim_i \mu_i\alpha$ then*

$$\varphi C = \inf\{g\alpha;\ \alpha \in \mathscr{B},\ C \subset \alpha\}\,.$$

3.2 THEOREM. *φ^* is a Radon measure on X such that:*

(i) $\varphi^* \leqq \varphi$.

(ii) *if C is compact then $\varphi^* C = \varphi C$.*

3.3 THEOREM. *If every open set in X is the countable union of compacta then $\varphi^* = \varphi$.*

Proofs

Proof of 3.1

(i) Let A, B be closed, compact and $A \cap B = 0$. Since X is regular and $\mathscr{B}$ is closed to finite unions, there exist $\alpha, \beta \in \mathscr{B}$ such that $A \subset \alpha$, $B \subset \beta$ and $\alpha \cap \beta = 0$. Given $\varepsilon > 0$, choose $\gamma_n \in \mathscr{B}$ for $n \in \omega$ so that $A \cup B \subset \bigcup_{n \in \omega} \gamma_n$ and

$$\sum_{n \in \omega} g\gamma_n \leqq \varphi(A \cup B) + \varepsilon .$$

Let $\gamma'_n = \gamma_n \cap \alpha$ and $\gamma''_n = \gamma_n \cap \beta$. Then $\gamma'_n, \gamma''_n \in \mathscr{B}$, $A \subset \bigcup_{n \in \omega} \gamma'_n$, $B \subset \bigcup_{n \in \omega} \gamma''_n$ and hence

$$\varphi A + \varphi B \leqq \sum_{n \in \omega} (g\gamma'_n + g\gamma''_n) \leqq \sum_{n \in \omega} g\gamma_n \leqq \varphi(A \cup B) + \varepsilon .$$

Since ε is arbitrary and φ is a Caratheodory measure we have $\varphi(A \cup B) = \varphi A + \varphi B$.

(ii) Let $A \subset X$. If $\varphi A = \infty$ then the conclusion is trivial. So, let $\varphi A < \infty$ and $\varepsilon > 0$. Then there exists a countable $H \subset \mathscr{B}$ such that $A \subset \bigcup_{\alpha \in H} \alpha$ and

$$\sum_{\alpha \in H} g\alpha \leqq \varphi A + \varepsilon$$

and therefore

$$\varphi(\bigcup_{\alpha \in H} \alpha) \leqq \sum_{\alpha \in H} \varphi\alpha \leqq \sum_{\alpha \in H} g\alpha \leqq \varphi A + \varepsilon .$$

(iii) Suppose for every $\alpha \in \mathscr{B}$, $g\alpha = \lim_i \mu_i \alpha$. Then for $\alpha_0, \cdots, \alpha_n$ in $\mathscr{B}$ we have

$$\begin{aligned}\sum_{k=0}^{n} g\alpha_k &= \lim_i \sum_{k=0}^{n} \mu_i \alpha_k \\ &= \lim_i \mu_i\Big(\bigcup_{k=0}^{n} \alpha_k\Big) \\ &= g\Big(\bigcup_{k=0}^{n} \alpha_k\Big) .\end{aligned}$$

Hence for any compact C,

$$\varphi C = \inf \{g\alpha\, ;\, \alpha \in \mathscr{B},\, C \subset \alpha\} .$$

Proof of 3.2

(i) Clearly, for any compact C, $\varphi C < \infty$ and, for any open α,

$$\varphi^*\alpha = \sup\{\varphi C\,;\ C \text{ compact},\ C \subset \alpha\} \leqq \varphi\alpha\,.$$

Thus, for any $A \subset X$, using 3.1 (ii) we have

$$\begin{aligned}\varphi^* A &= \inf\{\varphi^*\alpha\,;\ \alpha \text{ open},\ A \subset \alpha\}\\ &\leqq \inf\{\varphi\alpha\,;\ \alpha \text{ open},\ A \subset \alpha\}\\ &= \varphi A\,.\end{aligned}$$

(ii) For any compact C and open $\alpha \supset C$, we have $\varphi C \leqq \varphi^*\alpha$, hence $\varphi C \leqq \varphi^* C$. By (i) then $\varphi^* C = \varphi C$.

(iii) To see that φ^* is a Radon measure, we now only need to check that open sets are φ^*-measurable. Let α be open, $T \subset X$ and $\varepsilon > 0$. Let T' be open, $T \subset T'$ and $\varphi^* T' < \varphi^* T + \varepsilon$. Note that if C is compact, β is open and $C \subset \beta$ then, by regularity, $\bar{C} \subset \beta$. Thus, since $T' \cap \alpha$ is open, there exists a closed, compact $C_1 \subset T' \cap \alpha$ with $\varphi^*(T' \cap \alpha) \leqq \varphi C_1 + \varepsilon$. Also, since $T' - C_1$ is open, there exists a closed compact $C_2 \subset T' - C_1$ with $\varphi^*(T' - C_1) \leqq \varphi C_2 + \varepsilon$. Then

$$\begin{aligned}\varphi^*(T \cap \alpha) + \varphi^*(T - \alpha) &\leqq \varphi^*(T' \cap \alpha) + \varphi^*(T' - C_1)\\ &\leqq \varphi C_1 + \varphi C_2 + 2\varepsilon\\ &= \varphi(C_1 \cup C_2) + 2\varepsilon \quad (\text{by } 3.1\,(\text{i}))\\ &\leqq \varphi^* T' + 2\varepsilon\\ &\leqq \varphi^* T + 3\varepsilon\,.\end{aligned}$$

Proof of 3.3. We need only show that $\varphi^* A = \varphi A$ for open A. Given such A, by assumption, $A = \bigcup_{n\in\omega} C_n$ where the C_n are compact and $C_n \subset C_{n+1}$. Because of regularity, we may assume that the C_n are closed compact. We shall show that $\varphi A = \lim_n \varphi C_n$. To this end, let $\varepsilon > 0$ and define α_n and C'_n by recursion as follows: let $C' = C_0$ and, for any $n \in \omega$, let α_n be open, $C'_n \subset \alpha_n$, $\varphi\alpha_n \leqq \varphi C'_n + \varepsilon/2^{n+1}$ and

$$C'_{n+1} = C_{n+1} - \bigcup_{j=0}^{n} \alpha_j\,.$$

Then the C'_n are closed compact, mutually disjoint and $A \subset \bigcup_{n\in\omega}\alpha_n$. Thus,

$$\begin{aligned}\varphi A &\leqq \sum_{n\in\omega} \varphi\alpha_n \leqq \sum_{n\in\omega} \varphi C'_n + \varepsilon\\ &= \lim_N \sum_{n=0}^{N} \varphi C'_n + \varepsilon = \lim_N \varphi\Big(\bigcup_{n=0}^{N} C'_n\Big) + \varepsilon\\ &\leqq \lim_N \varphi C_N + \varepsilon\,.\end{aligned}$$

4. Weak* convergence. Let X be a locally compact, Hausdorff space, $\mathscr{M}$ be the family of Radon measures on X, μ be a net in $\mathscr{M}$. It is well known that $\mathscr{M}$ can be identified with the set of positive linear functionals on $C_0(X)$ so that the weak* or vague limit of the μ_i is defined by

4.1. DEFINITION. $(W^*)\text{-}\lim_i \mu_i = \nu$ if and only if $\nu \in \mathscr{M}$ and, for every $f \in C_0(X)$,

$$\lim_i \int f d\mu_i = \int f d\nu \ .$$

On the other hand, for any base $\mathscr{B}$ for the topology of X, let

4.2. DEFINITION. $\mathscr{B}\text{-}\underline{\mathrm{Lim}}_i\, \mu_i$ be the measure φ^* defined in § 3. If $\mathscr{B}$ is the family of all open sets then we simply write $\underline{\mathrm{Lim}}_i\, \mu_i$ instead of $\mathscr{B}\text{-}\underline{\mathrm{Lim}}_i\, \mu_i$.

We then have the following:

4.3. THEOREM. *$(W^*)\text{-}\lim_i \mu_i$ exists if and only if there exists a base $\mathscr{B}$ for the topology of X, closed under finite unions and intersections, such that, for every $\alpha \in \mathscr{B}$, $\lim_i \mu_i \alpha < \infty$, in which case,*

$$(W^*)\text{-}\lim_i \mu_i = \mathscr{B}\text{-}\underline{\mathrm{Lim}}_i\, \mu_i = \underline{\mathrm{Lim}}_i\, \mu_i \ .$$

The proof of this theorem is given in Lemmas A, B, C, D, E below. A restricted version of Lemma B was proved by Wulfsohn [4].

LEMMA A. *Let $\nu \in \mathscr{M}$ and*

$$\mathscr{B} = \{\alpha :\ \alpha \textit{ is open, } \bar{\alpha} \textit{ is compact and } \nu \text{ (boundary } \alpha) = 0\} \ .$$

Then $\mathscr{B}$ is a base for the topology of X and is closed under finite unions and intersections.

Proof. Let A be open and $a \in A$. Then there exists $f \in C_0(X)$ such that: $0 \leqq f(x) \leqq 1$ for $x \in X$, $f(a) = 1$ and $f(x) = 0$ for $x \notin A$. Since $\int f d\nu < \infty$, there exists $0 < t < 1$ such that $\nu(f^{-1}\{t\}) = 0$. Let $\alpha = \{x : f(x) > t\}$. Then α is open, $a \in \alpha \subset A$ and boundary $\alpha = f^{-1}\{t\}$ so that $\alpha \in \mathscr{B}$. Thus, $\mathscr{B}$ is a base. It is closed to finite unions and intersections since boundary $(\alpha \cup \beta) \cup$ boundary $(\alpha \cap \beta) \subset$ boundary $\alpha \cup$ boundary β for any open α, β.

LEMMA B. *$(W^*)\text{-}\lim_i \mu_i = \nu$ if and only if $\nu \in \mathscr{M}$ and $\lim_i \mu_i \alpha = \nu\alpha$ for every open α with $\bar{\alpha}$ compact and ν (boundary $\alpha) = 0$.*

Proof. Let $(W^*)\text{-}\lim_i \mu_i = \nu$, α be open, $\bar{\alpha}$ compact, ν (boundary

$\alpha) = 0$. For any compact $C \subset \alpha$, let $f \in C_0(X)$, $0 \leqq f(x) \leqq 1$ for all $x \in X$, $f(x) = 1$ for $x \in C$, $f(x) = 0$ for $x \notin \alpha$. Then

$$\nu C \leqq \int f d\nu = \lim_i \int f d\mu_i \leqq \varliminf_i \mu_i \alpha .$$

Hence

$$\nu\alpha \leqq \varliminf_i \mu_i \alpha .$$

Now, since ν (boundary $\alpha) = 0$, given $\varepsilon > 0$, let β be open, $\bar{\alpha} \subset \beta$ and $\nu\beta \leqq \nu\bar{\alpha} + \varepsilon = \nu\alpha + \varepsilon$. Let $f \in C_0(x)$, $0 \leqq f(x) \leqq 1$ for $x \in X$, $f(x) = 1$ for $x \in \bar{\alpha}$, $f(x) = 0$ for $x \notin \beta$. Then

$$\varlimsup_i \mu_i \alpha \leqq \lim_i \int f d\mu_i = \int f d\nu \leqq \nu\beta \leqq \nu\alpha + \varepsilon .$$

Thus,

$$\nu\alpha = \lim_i \mu_i \alpha .$$

Conversely, suppose $\nu \in \mathscr{M}$ and $\lim_i \mu_i \alpha = \nu\alpha$ for every open α with $\bar{\alpha}$ compact and ν (boundary $\alpha) = 0$. Let $f \in C_0(X)$, $\varepsilon > 0$. Then there exist $t_k \neq 0$ for $k = 0, \cdots, n$ such that $t_k < t_{k+1}$, $t_0 \leqq f(x) \leqq t_n$ for $x \in X$, $\nu(f^{-1}\{t_k\}) = 0$ and

$$\sum_{k=0}^{n-1} t_{k+1} \nu\alpha_k - \varepsilon \leqq \int f d\nu \leqq \sum_{k=0}^{n-1} t_k \nu\alpha_k + \varepsilon$$

where

$$\alpha_k = \{x : t_k < f(x) < t_{k+1}\}$$

so that α_k is open, $\bar{\alpha}_k$ is compact and ν (boundary $\alpha_k) = 0$. Then $\lim_i \mu_i \alpha_k = \nu\alpha_k$ and

$$\begin{aligned} \int f d\nu &\leqq \lim_i \sum_{k=1}^{n-1} t_k \mu_i \alpha_k + \varepsilon \\ &\leqq \varliminf_i \int f d\mu_i + \varepsilon . \end{aligned}$$

Now, let β_k be open, $\bar{\beta}_k$ be compact, ν (boundary $\beta_k) = 0$, $\bar{\alpha}_k \subset \beta_k$ and $\nu\beta_k \leqq \nu\alpha_k + \varepsilon/(n \,|\, t_{k+1} |)$. Then $\lim_i \mu_i \beta_k = \nu\beta_k$ and

$$\begin{aligned} \varlimsup_i \int f d\mu_i &\leqq \lim_i \sum_{k=0}^{n-1} t_{k+1} \mu_i \beta_k \\ &= \sum_{k=0}^{n-1} t_{k+1} \nu\beta_k \\ &\leqq \sum_{k=0}^{n-1} t_{k+1} \nu\alpha_k + \varepsilon \\ &\leqq \int f d\nu + 2\varepsilon . \end{aligned}$$

LEMMA C. *If* $(W^*)\text{-}\lim_i \mu_i = \nu$ *and*

$$\mathscr{B} = \{\alpha : \alpha \textit{ is open, } \bar{\alpha} \textit{ is compact, } \nu \text{ (boundary } \alpha) = 0\}$$

then

$$\nu = \mathscr{B}\text{-}\underline{\mathrm{Lim}}_i\, \mu_i \, .$$

Proof. Let $g\alpha = \underline{\lim}_i\, \mu_i\alpha$ for any $\alpha \in \mathscr{B}$, φ be the measure generated by g and $\mathscr{B}$ (see § 3). Then, in view of Lemma B and 3.1 (iii), for any compact $C \subset X$,

$$\varphi C = \inf\{g\beta\,;\, \beta \in \mathscr{B}\,;\, C \subset \beta\}\,.$$

Now, for any open $\alpha \supset C$ there exists, by Lemma A, $\beta \in \mathscr{B}$ with $C \subset \beta \subset \alpha$. Therefore, using Lemma B, and the outer regularity of ν, we have

$$\begin{aligned} \nu C &= \inf\{\nu\alpha\,;\ \alpha \text{ open, } C \subset \alpha\} \\ &= \inf\{\nu\beta\,;\ \beta \in \mathscr{B},\ C \subset \beta\} \\ &= \inf\{g\beta\,;\ \beta \in \mathscr{B},\ C \subset \beta\} \\ &= \varphi C\,. \end{aligned}$$

Hence, for any $A \subset X$,

$$\begin{aligned} \nu A &= \inf_{\substack{\alpha \text{ open} \\ A \subset \alpha}} \ \sup_{\substack{C \text{ compact} \\ C \subset \alpha}} \nu C \\ &= \inf_{\substack{\alpha \text{ open} \\ A \subset \alpha}} \ \sup_{\substack{C \text{ compact} \\ C \subset \alpha}} \varphi C = \mathscr{B}\text{-}\underline{\mathrm{Lim}}_i\, \mu_i A\,. \end{aligned}$$

LEMMA D. *Let* $\mathscr{B}$ *be a base for the topology of* X, *closed under finite unions and intersections, such that for any* $\alpha \in \mathscr{B}$, $\lim_i \mu_i\alpha < \infty$. *Then*

$$\mathscr{B}\text{-}\underline{\mathrm{Lim}}_i\, \mu_i = (W^*)\text{-}\lim_i \mu_i\,.$$

Proof. For $\alpha \in \mathscr{B}$, let $g\alpha = \underline{\lim}_i\, \mu_i\alpha = \lim_i \mu_i\alpha$, φ be the measure generated by g and $\mathscr{B}$ and $\varphi^* = \mathscr{B}\text{-}\underline{\mathrm{Lim}}_i\, \mu_i$ (see § 3). Then, by Theorem 3.2, $\varphi^* \in \mathscr{M}$. Let α be open, $\bar{\alpha}$ compact, $\varphi^*(\text{boundary } \alpha) = 0$. By 3.2 (ii), we have

$$\varphi^*\alpha = \varphi^*\bar{\alpha} = \varphi\bar{\alpha}$$

and by 3.1 (iii),

$$\varphi\bar{\alpha} = \inf\{g\beta\,;\ \beta \in \mathscr{B},\ \bar{\alpha} \subset \beta\}\,.$$

Given $\varepsilon > 0$, let $\beta \in \mathscr{B}$, $\bar{\alpha} \subset \beta$ and $g\beta \leqq \varphi^*\alpha + \varepsilon$. Then

$$\overline{\lim_i}\, \mu_i\alpha \leqq \lim_i \mu_i\beta = g\beta \leqq \varphi^*\alpha + \varepsilon\,.$$

On the other hand, let C be compact, $C \subset \alpha$ and $\varphi^*\alpha < \varphi^*C + \varepsilon = \varphi C + \varepsilon$. Then there exists $\gamma \in \mathscr{B}$ such that $C \subset \gamma \subset \alpha$ and therefore

$$\varphi C \leqq g\gamma = \lim_i \mu_i\gamma \leqq \underline{\lim_i}\, \mu_i\alpha\,.$$

Thus,

$$\overline{\lim_i}\, \mu_i\alpha \leqq \varphi^*\alpha \leqq \underline{\lim_i}\, \mu_i\alpha$$

so that $\lim_i \mu_i\alpha = \varphi^*\alpha$. By Lemma B then $\varphi^* = (W^*)\text{-}\lim_i \mu_i$.

LEMMA E. *Let $\mathscr{B}$ be a base for the topology of X, closed under finite unions and for every $\alpha \in \mathscr{B}$, $\lim_i \mu_i\alpha < \infty$. Then*

$$\mathscr{B}\text{-}\underline{\operatorname{Lim}}_i\, \mu_i = \underline{\operatorname{Lim}}_i\, \mu_i\,.$$

Proof. For any open α, let $g\alpha = \underline{\lim}_i\, \mu_i\alpha$, φ_1 be the measure generated by g and $\mathscr{B}$ and φ_2 be the measure generated by g and the family of all open sets. We have to show that for any compact C, $\varphi_1 C = \varphi_2 C$. Now, clearly $\varphi_2 C \leqq \varphi_1 C$. Suppose $\varphi_2 C < \infty$ and $\varepsilon > 0$. Let α_i be open for $i = 0, \cdots, n$, $C \subset \bigcup_{i=0}^n \alpha_i$ and

$$\sum_{i=1}^n g\alpha_i \leqq \varphi_2 C + \varepsilon\,.$$

For each $x \in C$ there exists $\beta \in \mathscr{B}$ such that $x \in \beta \subset \alpha_i$ for some $i = 0, \cdots, n$. Since C is compact, there is a finite family $H \subset \mathscr{B}$ which covers C and is a refinement of $\{\alpha_0, \cdots, \alpha_n\}$. For each i, let β_i be the union of all those elements in H which are contained in α_i. Then $\beta_i \in \mathscr{B}$, $\beta_i \subset \alpha_i$ and $C \subset \bigcup_{i=0} \beta_i$. Thus,

$$\varphi_1 C \leqq \sum_{i=0}^n g\beta_i \leqq \sum_{i=0}^n g\alpha_i \leqq \varphi_2 C + \varepsilon\,.$$

5. **Remarks.** Let $\mathscr{B}$, g, φ be as in § 3. The following example shows that φ need not be a Radon measure.

Let X be the set of all ordinals up to and including the first uncountable ordinal Ω. Then, in the order-topology, X is compact Hausdorff. For each $i < \Omega$, let μ_i be the point mass at i, that is, $\mu_i\alpha = 1$ if $i \in \alpha$ and $\mu_i\alpha = 0$ if $i \notin \alpha$. Let

$$\mathscr{B} = \{\alpha\,;\ \alpha \text{ is open and } \Omega \notin (\bar{\alpha} - \alpha)\}.$$

For any $\alpha \in \mathscr{B}$, if $\Omega \notin \alpha$ then α is countable and hence $g\alpha = \underline{\lim}_i\, \mu_i\alpha = 0$; if $\Omega \in \alpha$ then $g\alpha = 1$. Let $A = X - \{\Omega\}$. Then A is open and, being uncountable, for any countable family $H \subset \mathscr{B}$ which covers A there exists $\alpha \in H$ with $g\alpha = 1$. Thus, $\varphi A = 1$. On the other hand, if C is compact $C \subset A$ then C is countable and hence $\varphi C = 0$. Thus,

$$\varphi A \neq \sup \{\varphi C\, ;\ C \text{ compact},\ C \subset A\}\,.$$

Note, however, that if, instead of taking $\mathscr{B}$ as above, we let $\mathscr{B}$ be the family of all open sets in X then there exist uncountable, disjoint $\alpha, \beta \in \mathscr{B}$ with $A = \alpha \cup \beta$. Then $g\alpha = g\beta = 0$ so that $\varphi A = 0$. In this case, φ is the point mass at Ω and $\varphi = \varphi^*$.

We are unable to determine if this holds true in general for compact or locally compact Hausdorff spaces, i.e. if $\varphi = \varphi^*$ whenever $\mathscr{B}$ is the family of all open sets in X.

REFERENCES

1. P. R. Halmos, *Measure Theory*, Van Nostrand, 1950.
2. J. L. Kelley, *General Topology*, Van Nostrand, 1955.
3. M. E. Munroe, *Introduction to Measure and Intergration*, Addison-Wesley, 1953.
4. Aubrey Wulfsohn, *A note on the vague topology for measures*, Proc. Cambridge Phil. Soc., **58** (1962), 421–422.

A PERMANENT INEQUALITY FOR POSITIVE FUNCTIONS ON THE UNIT SQUARE

MORTON L. SLATER AND ROBERT J. THOMPSON

Introduction. During the past few years the van der Waerden conjecture on the minimum of the permanent of a doubly stochastic matrix has received considerable attention. (See Marcus and Newman [1] and [2], Marcus and Minc [1], among others.) This conjecture states that if A is a doubly stochastic matrix, i.e. if

$$a_{ij} \geqq 0, \sum_{i=1}^{n} a_{ij} = \sum_{j=1}^{n} a_{ij} = 1 ,$$

then the permanent of A is $\geqq n!\, n^{-n}$. (The permanent of A is $\sum \prod a_{i\sigma(i)}$, where the summation is taken over all permutations σ in the symmetric group.) Despite the seemingly elementary character of the conjecture, it is, so far as the present authors are aware, still unresolved in general, although it has been settled in some special cases. (See the above references.)

An implication of the conjecture is that some term of the permanent expansion must be greater than or equal to n^{-n}. This was established by Marcus and Minc [1] in 1962. Specifically they showed that if $\prod a_{ii}$ is not exceeded by any other term in the permanent expansion, then

$$(1) \qquad \sum \log a_{ii} \geqq \sum \sum a_{ij} \log a_{ij} \geqq \mathrm{n} \log n^{-1} .$$

The second inequality above is a simple application of Jensen's inequality using the convex function $x \log x$; the first inequality is the key to the problem. It is the extension of this inequality to functions defined on the unit square that is referred to in the title of this paper. We will show in § 4 that under suitable hypotheses

$$(2) \qquad \infty > \int_0^1 \log f(x, x)\, dx \geqq \int_0^1 \int_0^1 f(x, y) \log f(x, y)\, dxdy \geqq 0 .$$

The proof of (2) (and incidentally a new proof of (1)) is based ultimately on the following theorem:

THEOREM 1. *Let S be an arbitrary set and $f(p, q)$ a real-valued function defined on $S \times S$ with the following property:*

(C) *if $p_1, \cdots, p_n$ is any finite sequence of points in S, not necessarily distinct, then*

Received September 12, 1963. This work was performed under the auspices of the United States Atomic Energy Commission.

$$f(p_1, p_2) + f(p_2, p_3) + \cdots + f(p_{n-1}, p_n) + f(p_n, p_1) \leqq 0 .$$

Then there exists a real valued function φ *defined on* S *such that for all* $(p, q) \in S \times S$

$$f(p, q) \leqq \varphi(p) - \varphi(q) .$$

Furthermore, given any $s \in S$, *we may determine* $\varphi(p)$ *so that for all* $p \in S$

$$f(p, s) \leqq \varphi(p) \leqq -f(s, p) , \quad \text{and } \varphi(s) = 0 .$$

This theorem for finite sets S is essentially contained in a paper by S. N. Afriat [1] which appeared in 1963 in connection with a study of empirical preference analysis in economics. Theorem 1 was discovered independently by the authors in their study of the van der Waerden conjecture; it is very closely related to the linear programming dual of a theorem proved by Garret Birkhoff [1], which states that the doubly stochastic matrices are the convex hull of the permutation matrices. Indeed it was this last fact which persuaded us that Theorem 1 could be applied directly to the van der Waerden conjecture. In § 1 we will give a proof of this theorem which differs essentially from that for the finite case given by Afriat; it is certainly much shorter.

The proof of (2) to be given in §§ 3 and 4 will depend on Theorem 1 and on the following "Arzela type" compactness result proved by M. Riesz. We state it, for reference, in the form that we shall use it. It is also convenient to state here a partial converse of the Fubini theorem proved by L. Tonelli.

THEOREM A (M. Riesz). *Let* M *be a set of functions in* L (0, 1). *If*

1° *there exists a constant* K *such that for all* $x(t) \in M$

$$\int_0^1 |x(t)|\, dt \leqq K ,$$

and if

2° *for every* $\varepsilon > 0$, *there is a* $\delta > 0$ *such that for all* $x(t) \in M$ *and all* h *for which* $|h| < \delta$

$$\int_0^1 |x(t+h) - x(t)|\, dt < \varepsilon ,$$

then the set M *is conditionally compact in the sense of the metric of* L. A proof of the above result can be found in Nemyckii [1].

THEOREM B (Fubini converse: L. Tonelli). *Let* $f(x, y)$ *be measurable* *on the unit square. If for almost all* x, $|f(x, y)|$ *is summable as a function of* y, *and if*

$$\int_0^1 dx \int_0^1 |f(x, y)|\, dy$$

exists as an iterated integral and is finite, then $f(x, y)$ *is summable on the unit square.*

A proof of this theorem is in McShane [1].

1. Proof of theorem 1. Define $g(p, q) = f(p, q)$ for $p \neq q$ and $g(p, p) = 0$. Then g satisfies condition (C) and $f \leqq g$. Choose a fixed $s \in S$ and define

$$\varphi(p) = lub\{g(p, q_1) + g(q_1, q_2) + \cdots + g(q_{n-1}, q_n) + g(q_n, s)\}$$

where the least upper bound is taken over all finite sequences $q_1, \cdots, q_n$ selected from S. Since g satisfies (C) the finite sum is always $\leqq -g(s, p)$, and so the least upper bound is finite. Now fix $q_1 = q$ and let the remaining q_i range unrestricted. The definition of φ yields at once $\varphi(p) \geqq g(p, q) + \varphi(q)$ so that $f(p, q) \leqq g(p, q) \leqq \varphi(p) - \varphi(q)$ as claimed. Finally $f(p, s) \leqq g(p, s) \leqq \varphi(p) \leqq -g(s, p) \leqq -f(s, p)$, which completes the proof.

It may be worth remarking that if the range of f is any conditionally complete lattice ordered group, the proof goes through unchanged.

2. Proof of the matrix theorem. In this section we give a proof of inequality (1) based on Theorem 1. Suppose as stated in the introduction that the $n \times n$ matrix A is doubly stochastic and that $\prod a_{ii} \geqq \prod a_{i\sigma(i)}$ for all permutations σ. It is technically convenient to assume for the moment also that $a_{ij} > 0$.

Let $b_{ij} = \log a_{ij} - \log a_{ii}$; then b_{ij} as a function on $S \times S$, $S = \{1, 2, \cdots, n\}$, is easily seen to satisfy condition (C). (This follows readily from $b_{ii} = 0$ and $\sum b_{i\sigma(i)} \leqq 0$ for all σ.) Hence there exists a vector c_i such that $b_{ij} \leqq c_i - c_j$. Thus

$$\log a_{ij} \leqq \log a_{ii} + c_i - c_j\ , \qquad i, j = 1, \cdots, n\ ,$$

so that

$$a_{ij} \log a_{ij} \leqq a_{ij} \log a_{ii} + a_{ij}\, c_i - a_{ij}\, c_j\ .$$

If we now sum first with respect to j and then with respect to i, the vector c_i drops out and we have

$$\sum \sum a_{ij} \log a_{ij} \leqq \sum \log a_{ii}\ .$$

The positivity restriction of the a_{ij} is easily removed by a simple continuity argument.

3. Functions on the unit square. In this and the following section we shift our attention from the discrete matrix situation of § 2 and study an analogous situation on the unit square.

Let I denote the half open unit interval $[0, 1)$ and $\mathscr{T}$ the class of one-to-one measure preserving transformations of I onto I. We will prove the following theorem:

THEOREM 2. *Let $f(x, y)$ be a measurable function on $I \times I$ which satisfies*

$1°$ *for all* $T \in \mathscr{T}, f(x, Tx) \in L(I)$ *and* $\int_0^1 f(x, Tx)\,dx \leqq 0,$

and

$2°$ *the limit as* $\delta \to 0$ *of* $\int_0^1 |f(x, x+\delta)|\,dx = 0.$

(The function $f(x, y)$ is defined outside $I \times I$ to be periodic of period one in x and y.) Then there exists a function $\varphi \in L\ (I)$ such that for almost all $(x, y) \in I \times I$

$$f(x, y) \leqq \varphi(x) - \varphi(y)\ .$$

The proof of Theorem 2 requires two lemmas. (Throughout this section we will assume that $1°$ and $2°$ above hold.)

LEMMA 1. *Let $E \subset I$ be the union of a finite number of disjoint intervals and let $T \in \mathscr{T}$ be such that $TE = E$. Then*

$$\int_E f(x, Tx)\,dx \leqq 0\ . \tag{3}$$

Proof. We may assume that the intervals of E are semi-open (open on the right), so that the same is true of the finite set of non-continuous intervals that compose $I - E$. Let $J = [a, b)$ be one such interval of $I - E$. Define a measure preserving transformation U_n on J as follows: set $\delta_n = (b - a)/2n$ and

$$\begin{aligned} U_n x &= x + \delta_n\ , \quad a + 2(k-1)\delta_n \leqq x < a + (2k-1)\delta_n\ ; \\ U_n x &= x - \delta_n\ , \quad a + (2k-1)\delta_n \leqq x < a + 2k\,\delta_n\ , \\ &\qquad\qquad\qquad\qquad\qquad\qquad k = 1, \cdots, n\ . \end{aligned}$$

Then

$$\left| \int_J f(x, U_n x)\,dx \right| \leqq \int_0^1 |f(x, x+\delta_n)|\,dx + \int_0^1 |f(x, x-\delta_n)|\,dx$$
$$\to 0 \text{ as } n \to \infty \text{ by } 2° \text{ of Theorem } 2\ .$$

If we define U_n similarly on each of the finite set of $J \subset I - E$, and $U_n x = Tx$ for $x \in E$, then $U_n \in \mathscr{T}$ and

$$\int_{I-E} f(x, U_n x)\, dx + \int_E f(x, Tx)\, dx \leqq 0 ,$$

by 1° of Theorem 2. Since $\int_{I-E} f(x, U_n x)\, dx \to 0$, the result follows.

Lemma 2. *Let $f(x, y)$ be as in Theorem 2. Define for $0 < \lambda < 1$*

$$f(x, y; \lambda) = \frac{1}{\lambda}\int_0^\lambda f(x + t, y + t)\, dt . \tag{4}$$

Then $f(x, y; \lambda)$ satisfies condition (C) *of Theorem* 1 *on $I \times I$.*

Proof. We prove the lemma for the function $\lambda f(x, y; \lambda)$. Define $F(x_1, x_2, \cdots, x_n; \lambda) = \lambda f(x_1, x_2; \lambda) + \cdots + \lambda f(x_n, x_1; \lambda) \equiv F(x; \lambda)$. We will show that given any ordered set $x = (x_1, \cdots, x_n)$, $F(x; \lambda) \leqq 0$ for all $0 < \lambda < 1$. The following two easily verified properties of $F(x; \lambda)$ will be required:

(5a) given any finite ordered set x, there are finite ordered sets $x^{(i)}$, each of which has distinct components, and elements x_j, such that identically in λ

$$\begin{aligned} F(x; \lambda) = F(x^{(1)}; \lambda) + \cdots + F(x^{(k)}; \lambda) \\ + F(x_1, x_1; \lambda) + \cdots + F(x_p, x_p; \lambda) ; \end{aligned}$$

(5b) identically in x

$$F(x; \lambda) = F(x; \lambda_1) + F(x + \lambda_1; \lambda_2) + \cdots + F(x + \lambda_1 + \cdots + \lambda_{k-1}; \lambda_k) ,$$

where $\lambda = \lambda_1 + \cdots + \lambda_k$.

(We leave to the reader the verification of the above.)

As a consequence of Lemma 1 ($F(x_j, x_j; \lambda) \leqq 0$) and (5a), it will suffice to prove $F(x; \lambda) \leqq 0$ when the components of x are distinct. Suppose then that $x = (x_1, \cdots, x_n)$, $x_i \neq x_j$ for $i \neq j$, $0 \leqq x_i < 1$, and consider for the moment the x_i rearranged in increasing order, say $y_1, \cdots, y_n$. We define $\lambda_* = \operatorname{Min}\{y_2 - y_1, y_3 - y_2, \cdots, y_n - y_{n-1}, y_1 + 1 - y_n\}$, and note that $\lambda_* > 0$ by our conditions on the x_i. Suppose first that $0 < \lambda \leqq \lambda_*$, and let E be the set of points $x_i + t$ ($i = 1, \cdots, n$; $0 \leqq t < \lambda$) reduced modulo 1. For $0 \leqq t < \lambda$ define $T(x_i + t) = x_{i+1} + t$, $i = 1, \cdots, n - 1$ and $T(x_n + t) = x_1 + t$, where again all numbers are reduced modulo 1. Since $\lambda \leqq \lambda_*$, T is well defined on E and $TE = E$. For $x \in I - E$, define $Tx = x$, and we have $T \in \mathscr{T}$. By the periodicity of f,

$$F(x; \lambda) = \int_E f(x, Tx)\, dx, \text{ which is } \leqq 0 \text{ by Lemma 1 .}$$

We have shown, then, that

$$\text{for } 0 < \lambda \leqq \lambda_* , \quad F(x; \lambda) \leqq 0 . \tag{6}$$

Finally, since for $0 < \lambda < 1$ we may write $\lambda = k\lambda_* + r$ where k is a nonnegative integer and $0 \leqq r < \lambda_*$, we see that (5b) and (6) complete the proof. (This is equivalent to iterating T k times with $\lambda = \lambda_*$ and then using T with $\lambda = r$.)

Before staring the proof of Theorem 2 we make a heuristic remark about hypothesis 2°. If $f(x, y) \leqq \varphi(x) - \varphi(y), f(x, x) = 0$, and all the functions are smoothly differentiable, then the surfaces $z = f(x, y)$ and $z = \varphi(x) - \varphi(y)$ are tangent along $y = x$, and so $\varphi(x)$ is determined (up to an additive constant) by $\varphi'(x) = f_1(x, x)$. This suggests strongly that the "nature" of φ in general is determined by the behavior of $f(x, y)$ in the neighborhood of $y = x$. This will become clear in the proof that follows; later we will mention some consequences to φ of altering 2°.

We proceed now to the proof of Theorem 2. By Theorem 1 and Lemma 2 we know that for each $\lambda, 0 < \lambda < 1$, and for any $s \in I$, we can find a function $\varphi(x; s, \lambda)$ such that for all $(x, y) \in I \times I$

$$\begin{aligned} f(x, y; \lambda) &\leqq \varphi(x; s, \lambda) - \varphi(y; s, \lambda) , \\ f(x, s; \lambda) &\leqq \varphi(x; s, \lambda) \leqq -f(s, x; \lambda) , \end{aligned} \tag{7}$$

and

$$\varphi(s; s, \lambda) = 0 .$$

The remainder of the proof will be devoted to analyzing the (conditional) compactness of the family $\{\varphi(x; s, \lambda)\}$ in $L\ (I)$.

Theorem A (Riesz-Arzela) tells us that conditional compactness is implied by equicontinuity and uniform boundedness. We have from (7)

$$f(x, y; \lambda) \leqq \varphi(x; s, \lambda) - \varphi(y; s, \lambda) \leqq -f(y, x; \lambda) , \tag{8}$$

so that

$$\begin{aligned} |\varphi(x + \delta; s, \lambda) - \varphi(x; s, \lambda)| &\leqq |f(x + \delta, x; \lambda)| + |f(x, x + \delta; \lambda)| \\ &\leqq \frac{1}{\lambda} \int_0^1 \{|f(x + \delta, x)| + |f(x, x + \delta)|\}\, dx . \end{aligned} \tag{9}$$

Thus by 2°, $\varphi(x; s, \lambda)$ is continuous and hence measurable. Furthermore from the first inequality of (9) and Theorem B we have easily

$$\begin{aligned} &\int_0^1 |\varphi(x + \delta; s, \lambda) - \varphi(x; s, \lambda)|\, dx \\ &\qquad \leqq \int_0^1 \{|f(x + \delta, x)| + |f(x, x + \delta)|\}\, dx , \end{aligned} \tag{10}$$

so that the entire family $\{\varphi(x; s, \lambda)\}$ is equicontinuous (L).

Uniform boundedness (L) is more of a problem. We have found it necessary to choose an appropriate sub-family, and this will be done in the following paragraphs.

Since $f(x, y)$ is measurable on $I \times I$ we conclude from 2° and Theorem B that there exists a number $a > 0$ such that f is summable on the set P bounded by the lines $x = 0$, $x = 1$, $y = x \pm a$. We define $\bar{f}(x, y) = f(x, y)$ on P and all points in the plane congruent to P modulo one in x and y; elsewhere we set $\bar{f}(x, y) = 0$.

We will choose $s_1 \in I$ so that $0 \leqq s_1 < a$, and both (11) and (12) are satisfied:

(11) as $\lambda \longrightarrow 0$

$$\lim \frac{1}{\lambda}\int_{s_1}^{s_1+\lambda} dx \int_0^1 |\bar{f}(x, y)|\, dy = \int_0^1 |\bar{f}(s_1, y)|\, dy < \infty ,$$

and

$$\lim \frac{1}{\lambda}\int_{s_1}^{s_1+\lambda} dy \int_0^1 |\bar{f}(x, y)|\, dx = \int_0^1 |\bar{f}(x, s_1)|\, dx < \infty ;$$

and

(12) as $n \longrightarrow \infty$, for almost all $x \in I$,

$$\lim f_n(s_1, x) = f(s_1, x) , \quad \text{and } \lim f_n(x, s_1) = f(x, s_1) ,$$
$$\text{where } f_n(x, y) \equiv f(x, y; n^{-1}) .$$

For almost all $s \in I$ (11) holds since $\bar{f} \in L(P)$ and so $\in L\,(I \times I)$. Similarly, (12) is valid for almost all $s \in I$ by the fundamental theorem of calculus. (We introduce f_n in (12) to avoid some possible measurability difficulties.) Thus s_1 can certainly be chosen as required.

We will now show that the family $\{\varphi(x; s_1, n^{-1})\}$ is uniformly bounded (L). We choose $s_2, \cdots, s_k$ so that

(13) $$s_1 < s_2 < \cdots < s_k < 1 ,$$
$$s_{i+1} - s_i < 2a \text{ for } i = 1, \cdots, k-1 , \quad \text{and } 1 - s_k < a ;$$

(14) s_i satisfies (11) when s_1 is replaced by s_i, $i = 2, \cdots, k$; and finally

(15) as $n \longrightarrow \infty$,

$$\lim f_n(s_1, s_i) = f(s_1, s_i) ,$$

and

$$\lim f_n(s_i, s_1) = f(s_i, s_1) , \qquad i = 2, \cdots, k .$$

Now define $[a_1, b_1) = [0, s_1 + a)$, $(a_k, b_k) = (s_k - a, 1)$, and $(a_i, b_i) =$

$(s_i - a, s_i + a)$, $i = 2, \cdots, k-1$. The union of these intervals covers I. Write $\varphi_n(x; s_1)$ for $\varphi(x; s_1, n^{-1})$. Then by (8)

$$(16) \qquad |\varphi_n(x; s_1)| \leqq |f_n(x, s_i)| + |f_n(s_i, x)| + |\varphi_n(s_i; s_1)|, \quad \text{for } x \in I \text{ and } i = 1, \cdots, k.$$

Hence

$$(17) \qquad \begin{aligned} &\int_{a_i}^{b_i} |\varphi_n(x; s_1)|\, dx \\ &\leqq \int_{a_i}^{b_i} |f_n(x, s_i)|\, dx + \int_{a_i}^{b_i} |f_n(s_i, x)|\, dx + (b_i - a_i)|\varphi_n(s_i, s_1)|, \\ &\leqq A_i + B_i + C_i, \quad \text{where}, \end{aligned}$$

for $1 \leqq i \leqq k$, by (14)

$$A_i = lub\left\{n \int_{s_i}^{s_i+n^{-1}} dy \int_0^1 |\bar{f}(x, y)|\, dx\right\} < \infty,$$

$$B_i = lub\left\{n \int_{s_i}^{s_i+n^{-1}} dx \int_0^1 |\bar{f}(x, y)|\, dy\right\} < \infty;$$

$$C_1 = 0, \text{ by } (7),$$

and for $2 \leqq i \leqq k$, by (15)

$$C_i = (b_i - a_i)\, lub\, \{|f_n(s_i, s_1)| + |f_n(s_1, s_i)|\} < \infty.$$

Since

$$(18) \qquad \int_0^1 |\varphi_n(x; s_1)|\, dx \leqq \sum \int_{a_i}^{b_i} |\varphi_n(x; s_1)|\, dx,$$

we have established uniform boundedness (L) and Theorem A applies. We have then that some subsequence $\{\varphi_{n_i}(x; s_1)\}$ converges to $\varphi(x)$ (say) in L and $f_{n_i}(x, y)$ converges to $f(x, y)$ for almost all $(x, y) \in I \times I$. Since for all (x, y), $f_n(x, y) \leqq \varphi_n(x; s_1) - \varphi_n(y; s_1)$, Theorem 2 follows.

We now return to our remark preceding the proof of the theorem. We have just seen that the fact that φ is in $L(I)$ has been determined by condition 2°. It is reasonable to expect that a strengthening of 2° should lead to a "smoothing" of φ, and this is indeed the case. If 2° is replaced by

"2°_p for fixed $p\,(1 \leqq p < \infty)$ the limit as $\delta \to 0$ of

$$\int_0^1 |f(x, x + \delta)|^p\, dx = 0\text{''},$$

then $\varphi \in L_p(I)$. The modification of the proof consists of invoking the L_p version of Theorem A, which is also to be found in Nemyckii [1]. Finally if we replace 2° by

"2°_∞ the limit as $\delta \to 0$ of $\operatorname{ess\,sup}_x |f(x, x+\delta)| = 0$" ,

then $\varphi \in C(I)$. (The classical Arzela or Ascoli theorem is used.)

4. **The permanent theorem in $L(I \times I)$.** In this section we state and prove the $L(I \times I)$ analog of the discrete theorem of § 2.

THEOREM 3. *Suppose that $f(x, y)$ defined and measurable on $I \times I$ has the following properties*:

1° $f(x, y) > 0$ *and* $\int_0^1 f(x, y)\, dx = \int_0^1 f(x, y)\, dy = 1$, *for all* x, y;

2° *for all* $T \in \mathscr{T}$, $f(x, Tx)$ *is measurable,*

$$\log \frac{f(x, Tx)}{f(x, x)} \in L(I) ,$$

and

$$\int_0^1 \log \frac{f(x, Tx)}{f(x, x)}\, dx \leqq 0 ;$$

and

3°
$$\int_0^1 \left| \log \frac{f(x, x+\delta)}{f(x, x)} \right| dx \to 0 , \quad \text{as } \delta \to 0 .$$

Then $f \log f \in L(I \times I)$ *and*

$$\infty > \int_0^1 \log f(x, x)\, dx \geqq \int_0^1 \int_0^1 f(x, y) \log f(x, y)\, dxdy \geqq 0 . \tag{19}$$

Proof. Conditions 2° and 3° above suffice for the application of Theorem 2 to the function $\log [f(x, y)/f(x, x)]$: there exists $\varphi(x) \in L(I)$ such that for almost all x, y

$$\log \frac{f(x, y)}{f(x, x)} \leqq \varphi(x) - \varphi(y) . \tag{20}$$

If we multiply by $f(x, y)$ and rearrange, we find

$$\begin{aligned} -\frac{1}{e} &\leqq f(x, y) \log f(x, y) \\ &\leqq f(x, y) \log f(x, x) + \varphi(x) f(x, y) - \varphi(y) f(x, y) , \end{aligned} \tag{21}$$

where the first inequality is a consequence of $-1/e = glb\, x \log x$ for $x > 0$. Now, as functions of y, $f(x, y) \log f(x, x)$ and $\varphi(x)f(x, y)$ both $\in L$ by 1° above. Again, if we apply Theorem B to $\varphi(y)f(x, y)$, integrating first with respect to x, we see that $\varphi(y)f(x, y) \in L(I \times I)$, and so for almost all x, that function is summable as a function of y.

Thus by (21), $f(x, y) \log f(x, y)$ is summable y for almost all x, and integrating gives

$$0 \leqq \int_0^1 f(x, y) \log f(x, y)\, dy \tag{22}$$
$$\leqq \log f(x, x) + \varphi(x) - \int_0^1 \varphi(y) f(x, y)\, dy .$$

(The first inequality above is Jensen: $\psi\left(\int_0^1 f\, dy\right) \leqq \int_0^1 \psi(f)\, dy$, where $\psi(x) = x \log x$.) Hence

$$\log f(x, x) \geqq \int_0^1 \varphi(y) f(x, y)\, dy - \varphi(x) , \tag{23}$$

and so $\log f(x, x)$ is bounded below by a summable function. Now, since by 1° and Theorem B $f \in L(I \times I)$, it follows that for almost all δ, $f(x, x+\delta) \in L(I)$. We choose δ so that $f(x, x+\delta) \in L(I)$. Since $\log f(x, x+\delta) < f(x, x+\delta)$, and since by 2° $\log f(x, x+\delta) - \log f(x, x) \in L(I)$, we see that

$$\log f(x, x) \leqq f(x, x+\delta) + \log [f(x, x)/f(x, x+\delta)] , \tag{24}$$

and so $\log f(x, x)$ is also bounded above by a summable function; hence $\log f(x, x) \in L(I)$. Returning to (21) we apply Theorem B and have $f \log f \in L(I \times I)$; then integrating both sides of (22) φ drops out and we have (19) as asserted.

References

1. S. N. Afriat, [**1**] *The system of inequalities* $a_{rs} > X_r - X_s$, Proc. Camb. Phil. Soc., **59** (1963), 125–133.

2. G. Birkhoff, [**1**] *Tres observaciones sobre el algebra lineal*, Univ. Nac. Tucumán Rev. Ser. **A5** (1946), 147–151.

3. M. Marcus, and H. Minc, [**1**] *Some results on doubly stochastic matrices*, Proc. Amer. Math. Soc., **13** (1962), 571–579.

4. M. Marcus, and M. Newman, [**1**] *On the minimum of the permanent of a doubly stochastic matrix*, Duke Math. J., **26** (1959), 61–72.—[**2**] *The permanent function as an inner product*, Bull. Amer. Math. Soc., **67** (1961), 223–224.

5. E. J. Mc Shane, [**1**] *Integration*, Princeton Univ. Press, Princeton, New Jersey (1947).

6. V. V. Nemyckiĭ, [**1**] *The method of fixed points in analysis*, Uspekhi Mat. Nauk., **1** (1936), 141–174.

SANDIA LABORATORY, ALBUQUERQUE, NEW MEXICO

ON FIXED POINTS OF AUTOMORPHISMS OF CLASSICAL LIE ALGEBRAS

DAVID A. SMITH

1. Introduction. Let A be the automorphism group of a semi-simple Lie algebra $\mathfrak{L}$ over an algebraically closed field of characteristic zero. Let $n(A_i)$ denote the minimal multiplicity of 1 as characteristic root for elements of a connected (algebraic) component A_i of A, and let $m(A_i)$ denote the minimal dimension of fixed point spaces for elements of A_i. Jacobson showed in [3] that $n(A_i) = m(A_i)$, and determined these numbers. It is the purpose of this paper to extend these results to automorphisms of classical Lie algebras over essentially arbitrary fields, using the method of Chevalley [1], as extended by Steinberg [10], for associating such algebras with semi-simple complex Lie algebras.

Throughout the paper fields of characteristics 2 and 3 will be excluded without further mention. The results obtained here are valid in some cases in characteristics 2 and 3, but exclusion of these cases permits considerable simplification of the exposition. All vector spaces in this paper are finite dimensional.

2. Lie algebras and automorphism groups. Let $\mathfrak{L}_C$ be a semi-simple Lie algebra over the complex field C. Let $\mathfrak{H}_C$ be a Cartan subalgebra of $\mathfrak{L}_C$, and let e_i, f_i, h_i $(1 \leqq i \leqq l)$ be a canonical set of generators; i.e. the h_i form a basis for $\mathfrak{H}_C$, and

$$(1)\qquad \begin{aligned} [h_ih_j] &= 0\ , \\ [e_if_j] &= \delta_{ij}h_i\ , \\ [e_ih_j] &= A_{ji}e_i\ , \\ [f_ih_j] &= -A_{ji}f_i\ , \end{aligned}$$

where (A_{ij}) is the Cartan matrix of $\mathfrak{L}_C$. Let $\alpha_i(h_j) = A_{ji}$ for $i, j = 1, 2, \cdots, l$. Then $\pi \equiv \{\alpha_1, \alpha_2, \cdots, \alpha_l\}$ is a fundamental system of roots (of $\mathfrak{L}_C$ with respect to $\mathfrak{H}_C$), and the e_i (respectively, f_i) are root vectors for the α_i (respectively, $-\alpha_i$).

For each (nonzero) root α, let $\mathfrak{L}_\alpha$ denote the root space of α, and let h_α be the unique element of $[\mathfrak{L}_\alpha, \mathfrak{L}_{-\alpha}]$ such that $\alpha(h_\alpha) = 2$. In particular, $h_{\alpha_i} = h_i$, $1 \leqq i \leqq l$.

THEOREM (Chevalley [1]). *$\mathfrak{L}_C$ contains a complete set $\{e_\alpha\}$ of root vectors for the* (nonzero) *roots α such that*

Received October 10, 1963. The research reported here formed part of a dissertation presented for the degree of Doctor of Philosophy in Yale University. The author wishes to thank Professor Nathan Jacobson, who directed this research.

$$(2) \qquad [e_\alpha e_{-\alpha}] = h_\alpha \quad \textit{for all } \alpha\,;$$

$$(3) \qquad [e_\alpha e_\beta] = \pm(r+1)e_{\alpha+\beta}\,,$$

for all roots α, β *such that* $\alpha+\beta$ *is a root, where* r *is the largest integer* q *such that* $\beta - q\alpha$ *is a root.*

It is easily seen from Chevalley's proof of this theorem that the set $\{e_\alpha\}$ may be taken to contain the e_i and f_i, $1 \leqq i \leqq l$. Furthermore, the h_α are integral linear combinations of the h_i, and the roots are integral linear combinations of the α_i, so the set $\{h_i \mid 1 \leqq i \leqq l\} \cup \{e_\alpha \mid \alpha$ a nonzero root$\}$ is a basis for $\mathfrak{L}_C$ with an *integral* multiplication table contained in (1)–(3) and the relations

$$(4) \qquad [e_\alpha h_i] = \alpha(h_i)e_\alpha\,.$$

Such a basis $\{h_i, e_\alpha\}$ (containing the e_i and f_i) will be called a *Chevalley basis* for $\mathfrak{L}_C$. Henceforth a particular Chevalley basis will be assumed fixed. When it is convenient to do so, linear transformations in $\mathfrak{L}_C$ will be identified with their matrices relative to this basis.

Let K be an arbitrary field, and form a Lie algebra $\mathfrak{L}$ over K, related to $\mathfrak{L}_C$, as in [1]: $\mathfrak{L}$ is the tensor product (over the integers) of the additive group of K with the additive group generated by the Chevalley basis $\{h_i, e_\alpha\}$ of $\mathfrak{L}_C$; $\mathfrak{L}$ is equipped with the multiplication table (1)–(4) after identifying $1_K \otimes e_\alpha$ with e_α, etc. Thus the h_α, e_α, etc., are now thought of as elements of $\mathfrak{L}$, but observe that the subscripts still refer to roots of $\mathfrak{L}_C$.

Let $\mathfrak{H} = \sum_1^l Kh_i$. $\mathfrak{H}$ is an abelian subalgebra of $\mathfrak{L}$, and the roots of $\mathfrak{L}$ relative to $\mathfrak{H}$ are the linear functions $\bar\alpha$ defined by $\bar\alpha(h_\beta) =$ the class modulo the characteristic of K of $\alpha(h_\beta)$.

We follow the approach of Steinberg [10] in relating the Lie algebras $\mathfrak{L}$ of Chevalley with the Lie algebras of classical type of Mills and Seligman [4]. First let $\mathfrak{L}_C$ be simple. Then we have [10, 2.6]:

(a) No h_α is in the center $\mathfrak{Z}$ of $\mathfrak{L}$.

(b) $\mathfrak{Z} = \{h \in \mathfrak{H} \mid \bar\alpha(h) = 0$ for all roots α of $\mathfrak{L}_C\}$.

(c) If $\bar{\mathfrak{L}} = \mathfrak{L}/\mathfrak{Z}$, and $\bar{\mathfrak{H}} = \mathfrak{H}/\mathfrak{Z}$, then $\bar{\mathfrak{L}}$ is simple and $\bar{\mathfrak{H}}$ is a Cartan subalgebra of $\bar{\mathfrak{L}}$.

More generally, if $\mathfrak{L}_C$ is only semi-simple, then $\mathfrak{L}_C = \mathfrak{L}_{1,C} \oplus \cdots \oplus \mathfrak{L}_{r,C}$, where the $\mathfrak{L}_{i,C}$ are (non-abelian) simple ideals in $\mathfrak{L}_C$. Thus $\mathfrak{L} = \mathfrak{L}_1 \oplus \cdots \oplus \mathfrak{L}_r$, where the $\mathfrak{L}_i$ are the Lie algebras of Chevalley corresponding to the $\mathfrak{L}_{i,C}$, and are non-abelian ideals in $\mathfrak{L}$. The center $\mathfrak{Z}_i$ of $\mathfrak{L}_i$ is as described in (b), and the center $\mathfrak{Z}$ of $\mathfrak{L}$ is $\mathfrak{Z}_1 \oplus \cdots \oplus \mathfrak{Z}_r$. Furthermore, $\mathfrak{L}/\mathfrak{Z} \cong (\mathfrak{L}_1/\mathfrak{Z}_1) \oplus \cdots \oplus (\mathfrak{L}_r/\mathfrak{Z}_r)$. Every such algebra $\bar{\mathfrak{L}} = \mathfrak{L}/\mathfrak{Z}$ will be called a *classical Lie algebra.* (These are essentially the Lie algebras of classical type of Mills and Seligman, although some additional algebras over fields of characteristics 2 and 3 can be obtained by the process described here.)

If $\mathfrak{L}_C$ is simple, $\mathfrak{Z} \neq 0$ if and only if $\mathfrak{L}_C$ is of type A_l and the characteristic p of K divides $l+1$. In this case, $\mathfrak{Z}$ is one-dimensional [8, § 1].

Let A_C denote the automorphism group of $\mathfrak{L}_C$. As an algebraic group, A_C has a decomposition

$$A_C = A_0 \cup A_1 \cup \cdots \cup A_{r-1} \tag{5}$$

into connected (algebraic) components, where A_0 is the component of the identity automorphism. (The terminology of algebraic groups will be seen to be more natural here than that of topological groups.)

An *automorphism of the Cartan matrix* (A_{ij}) of $\mathfrak{L}_C$ is a permutation s of the numbers $1, 2, \cdots, l$ such that $A_{ij} = A_{s(i), s(j)}$ for all i, j. Associated with such a permutation s is a unique automorphism σ of $\mathfrak{L}_C$ such that $e_i^\sigma = e_{s(i)}, f_i^\sigma = f_{s(i)}$, $i = 1, 2, \cdots, l$ [2, p. 280]. Following Steinberg, we call σ a *graph* automorphism of $\mathfrak{L}_C$. The set F of graph automorphisms is a finite group, and the elements of $F = \{1, \sigma_1, \cdots, \sigma_{r-1}\}$ form a system of coset representatives of A_0 in A_C [2, Chapter IX; 3, Corollary to Theorem 6]:

$$A_C = A_0 \cup \sigma_1 A_0 \cup \cdots \cup \sigma_{r-1} A_0 \,. \tag{6}$$

This decomposition coincides with (5), and the number r of algebraic components is also the order of F.

For each root α and each complex number t, let $x_\alpha(t)$ denote the automorphism $\exp(t \text{ ad } e_\alpha)$ of $\mathfrak{L}_C$. The significance of the Chevalley basis for automorphisms is that the matrix of every $x_\alpha(t)$ has entries which are polynomials in t with integer coefficients [1]. Let $x_\alpha(\xi)$ denote the matrix obtained from $x_\alpha(t)$ by replacing the complex parameter t by an indeterminate ξ. We can then replace ξ by an arbitrary element t of K to obtain a matrix over K, again denoted $x_\alpha(t)$. Considered as a linear transformation of $\mathfrak{L}$ relative to the Chevalley basis, $x_\alpha(t)$ is an automorphism.

We also introduce certain diagonal (relative to the Chevalley basis) automorphisms of $\mathfrak{L}$. Let k be any homomorphism of the additive group generated by the roots of $\mathfrak{L}_C$ into the multiplicative group K^*. We associate with k the automorphism $\eta(k)$ of $\mathfrak{L}$ defined by $h\eta(k) = h$ for $h \in \mathfrak{H}$, $e_\alpha \eta(k) = k(\alpha) e_\alpha$ for α a root of $\mathfrak{L}_C$. In particular, we can associate a homomorphism k with each $t \in K^*$ and each root α of $\mathfrak{L}_C$ by defining $k(\beta) = t^{\beta(h_\alpha)}$ for each root β. The corresponding automorphism will be denoted $z_\alpha(t)$.

Next we associate automorphisms of $\mathfrak{L}$ with the graph automorphisms of $\mathfrak{L}_C$. Let σ be a graph automorphism with associated permutation s. We have $h_i^\sigma = [e_i^\sigma, f_i^\sigma] = [e_{s(i)}, f_{s(i)}] = h_{s(i)}$, so σ permutes the h_i's. For an arbitrary root $\gamma = \sum k_i\alpha_i$, let $\gamma' = \sum k_i\alpha_{s(i)}$. γ' is a root [2, p. 122, XVI] and one can show that $e_\gamma^\sigma = \pm e_{\gamma'}$. This is done by induction on the level (i.e. $\sum |k_i|$) of γ. Hence, relative to the Chevalley basis, the matrix of σ has only the numbers 0, 1, -1 as entries (and in fact, exactly one nonzero entry in each row and column). Thus the matrix of σ defines an automorphism σ of $\mathfrak{L}$ over K. These automorphisms will also be called graph automorphisms.

The automorphism group of $\bar{\mathfrak{L}}$ is isomorphic to the automorphism group of $\mathfrak{L}$ [10, p. 1122]. We will therefore identify automorphisms of $\mathfrak{L}$ with their induced automorphisms in $\bar{\mathfrak{L}}$, but all references to matrices will mean relative to the Chevalley basis in $\mathfrak{L}$.

The group G of Chevalley is the group of automorphisms of $\mathfrak{L}$ (or $\bar{\mathfrak{L}}$) generated by the $x_\alpha(t)$ for all roots α and $t \in K$ and the $\eta(k)$ for all homomorphisms k of the additive group generated by the roots into K^*.

THEOREM (Steinberg). *If A is the automorphism group of $\mathfrak{L}$ (or $\bar{\mathfrak{L}}$), G the Chevalley group, and $F = \{1, \sigma_1, \cdots, \sigma_{r-1}\}$ the group of graph automorphisms, then G is normal in A, and*

$$A = G \cup \sigma_1 G \cup \cdots \cup \sigma_{r-1} G \tag{7}$$

is the coset decomposition of A over G.

Steinberg proves this theorem in [10] only for the case of $\mathfrak{L}_C$ simple, but the extension to the semi-simple case is straightforward if one considers the action of A in $\bar{\mathfrak{L}}$. The analogy between equations (7) and (6) is clear; in fact, they coincide if K is an algebraically closed field of characteristic zero. However (7) is also analogous to (5) by the following result.

THEOREM (Ono [5, Theorem 3]). *If K is infinite, and the Killing form of $\mathfrak{L}_C$ is nondegenerate modulo the characteristic of K, then G is the algebraic component of 1 in A, and (7) is the decomposition of A into connected algebraic components.*

3. **Indices of automorphism groups.** For each component (or coset) A_i of A_C define the index $n(A_i)$ to be the minimal multiplicity of the characteristic root 1 for elements of A_i. For each $\eta \in A_C$, let $\mathfrak{F}(\eta)$ denote the subspace of $\mathfrak{L}_C$ of η-fixed points. Define another index $m(A_i)$ to be the minimal $\dim \mathfrak{F}(\eta)$, $\eta \in A_i$. We have [3, Theorem 6 and Corollary, Theorem 10]:

THEOREM (Jacobson). *Let σ_i be the unique element of F in A_i,* *and let* s_i *be the associated automorphism of the Cartan matrix. Then* $n(A_i) = m(A_i) =$ *the number of cycles in the decomposition of* s_i *into disjoint cycles.*

COROLLARY. $n(A_0) = l = \dim \mathfrak{H}_C$, *and* $0 < n(A_i) < l$ *if* $i \neq 0$.

In view of Steinberg's theorem in the previous section, it is reasonable to ask for the relationship between $n(A_i)$ and both the minimal multiplicity $n(\sigma_i G)$ of 1 as characteristic root and the minimal dimension $m(\sigma_i G)$ of fixed point spaces for elements of $\sigma_i G$ in the automorphism group A of $\mathfrak{L}$. (Obviously a distinction between $\mathfrak{L}$ and $\mathfrak{L}$ must be maintained here; we will consider $\bar{\mathfrak{L}}$ in § 4.)

In the sequel we will make use of the subgroup G' of G generated by the automorphisms $x_\alpha(t)$ for α a root of $\mathfrak{L}_C$ and $t \in K$. For each root α and each $t \in K^*$, $z_\alpha(t) \in G'$, and if K is algebraically closed, $G' = G$ [1, § IV].

THEOREM 1. *Let* $\mathfrak{L}_C$, A_C, A_i, K, $\mathfrak{L}$, A, G, *and* σ_i *be as defined above. Then* $n(\sigma_i G) \geqq m(\sigma_i G) \geqq n(A_i)$.

Proof. The first inequality is clear. We first assume K is algebraically closed, so that G is generated by the $x_\alpha(t)$. We have seen that an arbitrary element η of A can be written as a product of exactly one $\sigma_i \in F$ and certain $x_\alpha(t_j)$'s in some order. Thinking now of matrices, η is then a specialization of a corresponding product $\eta(\xi)$ of matrices σ_i, $x_\alpha(\xi_j)$, where the ξ's are indeterminates, one for each x-type factor. Since the entries of $x_\alpha(\xi_j)$ are polynomials in ξ_j with integer coefficients, $\eta(\xi)$ is a matrix whose entries are polynomials in certain indeterminates $\xi_1, \xi_2, \cdots, \xi_m$ with integer coefficients.

The number m of indeterminates appearing in a matrix $\eta(\xi)$ depends not only on the automorphism η but on the choice of a representation of η as a product of the generators; this number plays no special role here, but it must not be assumed to be constant.

The integer coefficients of the polynomial entries of $\eta(\xi)$ may be chosen so that specialization of the ξ_j to complex numbers t_j gives an element $\eta(t)$ of A_C, and the choice of σ_i determines the component in which $\eta(t)$ lies.

Let σ_i be fixed, and let $l_i = n(A_i)$. The fact that $l_i \leqq \dim \mathfrak{F}(\eta)$ for $\eta \in A_i$ can be expressed as follows: for every specialization $\xi_j \to t_j \in C$, rank $(\eta(t) - I) \leqq n - l_i$, where $n = \dim \mathfrak{L}_C = \dim \mathfrak{L}$. A similar statement can be made for $\eta(\xi)$, for if $\eta(\xi) - I$ had a nonzero minor of size $> n - l_i$, that minor would be a polynomial and would remain nonzero under some specialization $\xi_j \to t_j \in C$. Hence we see that for every $\eta(\xi)$ corresponding to σ_i (i.e. for every element $\eta \in \sigma_i G$ and for every representation of η as a product of σ_i and certain of the other generators) we have rank $(\eta(\xi)-I) \leqq n - l_i$. But then specializing $\xi_j \to t_j \in K$, the rank of such a matrix certainly cannot increase. Hence rank $(\eta - I) \leqq n - l_i$ for every $\eta \in \sigma_i G$, or in other words $m(\sigma_1 G) \geqq l_i$.

Now drop the assumption of algebraic closure on K, and let Ω be the algebraic closure of K. If η is an arbitrary element of $\sigma_i G$, then the extension of η to an automorphism of $\mathfrak{L}_\Omega$ is still in the component of $A(\mathfrak{L}_\Omega)$ corresponding to σ_i. This is clear, because $\eta = \sigma_i \tau$, $\tau \in G$, and τ can be expressed as a product of the generators of G, whose extensions to $\mathfrak{L}_\Omega$ are elements of $G(\mathfrak{L}_\Omega)$. Hence $\dim \mathfrak{F}(\eta) = \dim \mathfrak{F}(\eta_\Omega) \geqq l_i$ for $\eta \in \sigma_i G$. This completes the proof of Theorem 1.

THEOREM 2. *Let $\mathfrak{L}_C$, A_C, A_i, K, $\mathfrak{L}$, A, G, and σ_i be as in Theorem 1, and suppose further that K is infinite. Then $m(\sigma_i G) = m(A_i)$. For $i = 0$, $n(G) = n(A_0) = l$. If, in addition, the characteristic of K does not divide the length of any cycle in the permutation associated with σ_i, then $n(\sigma_i G) = n(A_i)$. In particular, this is the case if $\mathfrak{L}_C$ is simple.*

Proof. For the Chevalley group itself, we consider the diagonal automorphisms (or matrices) $z_\alpha(t) = \operatorname{diag}\{1, 1, \cdots, 1, \cdots, t^{\beta(h_\alpha)}, \cdots\}$, where each of the first l elements is 1, and the following entries are of the form $t^{\beta(h_\alpha)}$ where β runs through all the roots of $\mathfrak{L}_C$. For some selection of $t_1, t_2, \cdots, t_l \in K$, to be determined presently, let $\eta = \prod_1^l z_{\alpha_i}(t_i)$. The diagonal entries of η after the lth one are of the form $\prod_1^l t_i^{\beta(h_i)}$. For each root β, some $\beta(h_i) \neq 0$. Thus each of these entries is a rational expression in the t_i which is not identically 1. Since K is infinite, we can choose $t_1, \cdots, t_l$ so that none of the diagonal entries of η after the lth one is 1. (This can be expressed as a polynomial condition of degree $\leqq 3(n-l)$, where $n = \dim \mathfrak{L}$, since $|\beta(h_i)| \leqq 3$.) Thus η is an element of G for which $l = \dim \mathfrak{F}(\eta) =$ the multiplicity of 1 as characteristic root.

Now consider an element $\sigma \neq 1$ in F. σ maps $\mathfrak{H}$ into itself, and also maps the subspace $\mathfrak{S}$ spanned by the root vectors $\{e_\beta\}$ into itself. In $\mathfrak{H}$, σ acts as a permutation of the h_i, and in $\mathfrak{S}$ (as noted above) the matrix of σ has only 0, ± 1 as entries, and exactly one nonzero entry in each row and column. If η is chosen as in the previous paragraph, we have $\sigma\eta \,|\, \mathfrak{H} = \sigma \,|\, \mathfrak{H}$ (where the bar denotes restriction), and $\sigma\eta \,|\, \mathfrak{S}$ has nonzero entries where $\sigma \,|\, \mathfrak{S}$ does and each of these entries will be $\pm$ one of the entries of $\eta \,|\, \mathfrak{S}$. If K is infinite, then the t_i selected to define η can be chosen to satisfy not only the conditions imposed above, but also the condition that 1 not be a characteristic root of $\sigma\eta \,|\, \mathfrak{S}$.

Next consider the permutation matrix $\sigma \,|\, \mathfrak{H}$. For a suitable arrangement of the basis $h_1, \cdots, h_l$ of $\mathfrak{H}$, this matrix consists of diagonal blocks, where each block is the matrix of a cyclic permutation. Let T be a linear transformation in a k-dimensional space which cyclically permutes a basis $u_1, u_2, \cdots, u_k$. Then the fixed point space of T is spanned by $u_1 + u_2 + \cdots + u_k$. The characteristic polynomial of T (up to sign) is $(\lambda - 1)(\lambda^{k-1} + \lambda^{k-2} + \cdots + \lambda + 1)$. 1_K is a root of the second factor if and only if $k \cdot 1_K = 0$. Thus the multiplicity of 1 as characteristic root of T is 1 if and only if the characteristic of K does not divide k.

We have demonstrated that each cycle of s contributes exactly one dimension to the fixed point space of $\sigma \mid \mathfrak{H}$, and, if the characteristic does not divide the length of the cycle, exactly 1 to the multiplicity of 1 as characteristic root. If $\mathfrak{L}_C$ is simple, only cycles of lengths $\leqq 3$ occur, which completes the proof of Theorem 2.

COROLLARY. *Let $\mathfrak{L}$ be a split semi-simple Lie algebra over an arbitrary field of characteristic zero, and let $A = G \cup \sigma_1 G \cup \cdots \cup \sigma_{r-1} G$ be the automorphism group of $\mathfrak{L}$. Then $m(\sigma_i G) = n(\sigma_i G) =$ the number l_i of cycles in the decomposition of the permutation s_i. For G itself, $l_0 = l$, the dimension of a Cartan subalgebra, and for $i \neq 0$, $0 < l_i < l$.*

REMARKS. (a) The corollary extends the results of Jacobson [3] beyond the algebraically closed case. Part of this is essentially contained in [3] in remarks following Theorem 10.

(b) The decomposition of A in the corollary is also the decomposition into connected algebraic components, by Ono's theorem in § 2.

We will consider in the remaining sections the extent to which the exclusion of small fields is necessary to obtain the conclusions of Theorem 2. In particular, we will answer this explicity for the Chevalley group for algebras of types A, B, C, and D.

There is also the question of how these results may be extended to the algebras $\bar{\mathfrak{L}}$, in the case where one or more components are of type A_l, $p \mid l + 1$. In the following section we will obtain explicit results in the case where $\mathfrak{L}_C$ itself is simple of type A_l, $p \mid l + 1$.

4. Algebras of type A. Let $\mathfrak{L}_C$ be simple of type A_l. Then $\mathfrak{L}$ can be taken to be the Lie algebra of all $(l + 1) \times (l + 1)$ matrices of trace 0 over K. If A is any nonsingular $(l + 1) \times (l + 1)$ matrix, then the mapping $X \rightarrow A^{-1}XA$ is an automorphism η of $\mathfrak{L}$. This automorphism is in G, by [9, § 2] and the last paragraph of the proof of Theorem 1.

THEOREM 3. *If $\mathfrak{L}_C$ is of type A_l and K is any field (of charac-*

teristic $\neq 2, 3$), *then* $m(G) = l$. *If* $|K| > l + 1$, *then* $n(G) = l$.

Proof. Let η be an automorphism given by conjugation by a cyclic matrix A. The space of all matrices commuting with A (i.e. all polynomials in A) has dimension $l + 1$, since the minimum polynomial of A has degree $l + 1$. $\mathfrak{F}(\eta)$ is the intersection of this space with $\mathfrak{L}$, and has dimension l.

An alternate approach to selecting an $\eta \in G$ gives a slightly weaker result, but also gives an automorphism having 1 as characteristic root with multiplicity l. Let $\eta : X \to A^{-1}XA$ where $A = \operatorname{diag}\{a_1, a_2, \cdots, a_{l+1}\}$, the a_i being all distinct and all different from 0. This requires $|K| > l + 1$. Take as basis for $\mathfrak{L}$ the matrix units e_{ij}, $i \neq j$, and the diagonal matrices $h_i = e_{i+1,i+1} - e_{ii}$, $1 \leqq i \leqq l$. Then $h_i^\eta = h_i$, and $e_{ij}^\eta = a_i^{-1}a_je_{ij}$. Since $a_i^{-1}a_j \neq 1$ for $i \neq j$, we have $l = \dim \mathfrak{F}(\eta) =$ the multiplicity of 1 as characteristic root, which completes the proof.

Now suppose the characteristic p of K divides $l + 1$. Then $\mathfrak{L}$ has one-dimensional center $\mathfrak{Z}$ consisting of scalar multiples of the identity matrix. A more convenient basis than the one listed above is obtained by replacing h_l by $I = lh_1 + (l - 1)h_2 + \cdots + 2h_{l-1} + h_l$, and taking this to be the first basis vector. The cosets of the remaining basis vectors then form a basis for $\bar{\mathfrak{L}} = \mathfrak{L}/\mathfrak{Z}$.

Since $l > 1$, we have one nontrivial graph automorphism σ with associated permutation $(1, l)(2, l - 1) \cdots$, in which the number of cycles is $[(l + 1)/2]$. We will denote by $\bar{n}(G)$ the minimal multiplicity of 1 as characteristic root for elements of G acting in $\bar{\mathfrak{L}}$, and similarly define $\bar{n}(\sigma G)$, $\bar{m}(G)$, $\bar{m}(\sigma G)$.

THEOREM 4. *Let* $\bar{\mathfrak{L}}$ *be a* (*simple*) *classical Lie algebra of type* A_l *over a field* K *of characteristic* p, *where* $p \mid l + 1$. *Let* $A = G \cup \sigma G$ *be the automorphism group of* $\bar{\mathfrak{L}}$. *Then* $\bar{n}(G) \geqq \bar{m}(G) \geqq l - 1$, *and* $\bar{n}(\sigma G) \geqq \bar{m}(\sigma G) \geqq [(l + 1)/2]$. *If* $|K| > l + 1$, *then* $\bar{n}(G) = \bar{m}(G) = l - 1$, *and if* K *is infinite, then* $\bar{n}(\sigma G) = \bar{m}(\sigma G) = [(l + 1)/2]$.

Proof. We observe first that $I^\sigma = (lh_1 + (l - 1)h_2 + \cdots + 2h_{l-1} + h_l)^\sigma = lh_l + (l - 1)h_{l-1} + \cdots + 2h_2 + h_1 = -I$. Every element of the subgroup G' of G acts by a conjugation in $\mathfrak{L}$ [6, (3.5)], so I is a fixed point of every element of G'. G is generated by G' and certain automorphisms leaving $\mathfrak{H} = \sum Kh_i$ pointwise fixed, so I is fixed under every element of G. On the other hand, if $\eta = \sigma\tau$, $\tau \in G$, then $I^\eta = (-I)^\tau = -I$, so I is not fixed under η.

Relative to the bases chosen above for $\mathfrak{L}$ and $\bar{\mathfrak{L}}$, every automorphism η of $\mathfrak{L}$ has a matrix of the form

$$A = \left[\begin{array}{c|c} a_1 & 0 \cdots 0 \\ \hline a_2 & \\ \cdots & B \\ a_n & \end{array}\right] \tag{8}$$

where B is the matrix of the induced automorphism $\bar{\eta}$ in $\bar{\mathfrak{L}}$. We have just seen that $a_1 = 1$ if $\eta \in G$ and $a_1 = -1$ if $\eta \in \sigma G$. For any η, the characteristic polynomial of A is

$$f(\lambda\,;\eta) = (\lambda - a_1)f(\lambda\,;\bar{\eta})\,, \tag{9}$$

where $f(\lambda\,;\bar{\eta})$ is the characteristic polynomial of B. Thus for $\eta \in G$, the multiplicity of 1 as characteristic root of $\bar{\eta}$ is exactly 1 less than that for η. In particular, if $|K| > l + 1$, $\bar{n}(G) \leqq l - 1$.

Now for $\eta \in G$, $\mathfrak{Z} \subseteqq \mathfrak{F}(\eta)$, hence $\dim \overline{\mathfrak{F}(\eta)} = \dim \mathfrak{F}(\eta) - 1$ (where the bar denotes image under $\mathfrak{L} \to \bar{\mathfrak{L}}$). Clearly $\overline{\mathfrak{F}(\eta)} \subseteqq \mathfrak{F}(\bar{\eta})$, so $l - 1 \leqq \bar{m}(G) \leqq \bar{n}(G)$. Again, if $|K| > l + 1$, $\bar{n}(G) = l - 1$.

On the other hand, if $\eta \in \sigma G$, $\mathfrak{Z} \cap \mathfrak{F}(\eta) = 0$, so $\dim \mathfrak{F}(\eta) = \dim \overline{\mathfrak{F}(\eta)} \leqq \dim \mathfrak{F}(\bar{\eta})$, and $\bar{m}(\sigma G) \geqq [(l + 1)/2]$. By (9), the multiplicity of 1 as characteristic root must be the same for η and $\bar{\eta}$. Hence if K is infinite, then $\bar{n}(\sigma G) = \bar{m}(\sigma G) = [(l + 1)/2]$.

5. Simple algebras of types B, C, D. Let $\mathfrak{L}_C$ be simple of type B_l, C_l, or D_l. Then $\mathfrak{L}$ can be taken to be the Lie algebra of $n \times n$ matrices X over K ($n = 2l$ or $2l + 1$) such that $X = -S^{-1}X'S$, where X' is the transpose of X, and S is

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & I_l \\ 0 & I_l & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & I_l \\ -I_l & 0 \end{bmatrix}, \quad \text{or} \quad \begin{bmatrix} 0 & I_l \\ I_l & 0 \end{bmatrix}$$

in the respective cases B, C, or D. If A is any matrix such that $ASA' = S$, then $X \to A^{-1}XA$ is an automorphism of $\mathfrak{L}$, and, as for type A_l, is in the Chevalley group. We will select in each case a diagonal matrix A which defines an automorphism of $\mathfrak{L}$ having l-dimensional fixed point space, after discarding a suitable number of small fields. The orthogonality condition requires that A be of the form diag $\{a_1, a_2, \cdots, a_l, a_1^{-1}, a_2^{-1}, \cdots, a_l^{-1}\}$ in cases C and D and of the form diag $\{1, a_2, a_3, \cdots, a_{l+1}, a_2^{-1}, \cdots, a_{l+1}^{-1}\}$ in case B.

THEOREM 5. *Let $\mathfrak{L}$ be a simple classical Lie algebra of type B_l, C_l, or D_l over a field K, and let G be its Chevalley group. Then $n(G) = m(G) = l$ if $|K| > 2l$, $2l + 1$, or $2l - 1$ in the respective cases B_l, C_l, D_l.*

Proof. First consider case C. Denoting matrix units by e_{ij}, a basis for $\mathfrak{L}$ [7, § XVII] is

$$\begin{aligned}
h_i &= e_{ii} - e_{i+l,\,i+l} && ; \\
e_{(-i,\,j)} &= e_{ij} - e_{j+l,\,i+l}\,, && i \neq j\,; \\
e_{(-i,\,-j)} &= e_{i,\,j+l} + e_{j,\,i+l}\,, && i < j\,; \\
e_{(i,\,j)} &= e_{i+l,\,j} + e_{j+l,\,i}\,, && i < j\,; \\
e_{(-2i)} &= e_{i,\,i+l} && ; \\
e_{(2i)} &= e_{i+l,\,i} && ;
\end{aligned}$$

where in all cases $i, j = 1, 2, \cdots, l$. If we choose A as above, then conjugation by A acts diagonally, leaving the h_i fixed, and the remaining diagonal elements have the forms $a_i^{-1}a_j$, $a_i^{-1}a_j^{-1}$, $a_i a_j$ $(i \neq j)$, a_i^{-2}, a_i^2. Hence we wish to choose the a_i so that no a_i is 0, 1, -1, or $a_j^{\pm 1}$ for $j \neq i$; in other words, so that

$$\textstyle\prod_1^l a_i(a_i^2 - 1) \prod_{i<j}(a_i - a_j)(a_i a_j - 1) \neq 0 \,.$$

The left-hand side of this inequality is a polynomial of degree $2l + 1$ in each of the a_i. Thus there exist such elements in K if $|K| > 2l + 1$.

The details for types B and D are similar, and appropriate bases are given in [7, § XVII]. For type B the same conditions are obtained except that some a_i may be -1. Hence $|K| > 2l$ suffices. For type D, both 1 and -1 are allowed, so $|K| > 2l - 1$ suffices.

REMARK. Professor G. B. Seligman has communicated to the author a proof that $m(G) = l$ when $\mathfrak{L}$ is of type B_l, C_l, or D_l, over any field K of characteristic $\neq 2$ or 3. His proof is a natural analog of the first part of the proof of Theorem 3, although the details are naturally more complicated. As in Theorem 3, this approach does not yield $n(G) = l$.

REFERENCES

1. C. Chevalley, *Sur certains groupes simples*, Tôhoku Math. J. (2) **7** (1955), 14-66.
2. N. Jacobson, *Lie Algebras*, Interscience, New York, 1962.
3. ———, *A note on automorphisms of Lie algebras*, Pacific J. Math., **12** (1962), 303-315.
4. W. H. Mills and G. B. Seligman, *Lie algebras of classical type*, J. Math. Mech., **6** (1957), 519-548.
5. T. Ono, *Sur les groupes de Chevalley*, J. Math. Soc. Japan **10** (1958), 307-313.
6. R. Ree, *On some groups defined by Chevalley*, Trans. Amer. Math. Soc., **84** (1957), 392-400.
7. G. B. Seligman, *On Lie algebras of prime characteristic*, Amer. Math. Soc. Memoir No. 19, 1956.
8. ———, *Some remarks on classical Lie algebras*, J. Math. Mech., **6** (1957), 549-

558.
9. ———, *On automorphisms of Lie algebras of classical type* II, Trans. Amer. Math. Soc., **94** (1960), 452-482.
10. R. Steinberg, *Automorphisms of classical Lie algebras,* Pacific J. Math., **11** (1961), 1119-1129.

DUKE UNIVERSITY

HOMOGENEOUS QUASIGROUPS

SHERMAN K. STEIN

A mathematical system whose group of automorphisms is transitive we will call homogeneous. If the group of automorphisms is doubly transitive, then we will call the system doubly homogeneous. We examine here homogeneous and doubly homogeneous finite quasigroups.

We prove that there are no homogeneous quasigroups whose order is twice an odd number (Theorem 1.1). As the quasigroups satisfying the identity $X(YZ) = XY \cdot XZ$ show, there are homogeneous quasigroups of all other orders ([5], p. 236).

We then examine doubly homogeneous quasigroups and show that they are intimately connected with nearfields (Theorem 2.2). Since all finite nearfields are known, we thus have a complete description of the doubly homogeneous quasigroups.

In the last two sections we obtain various equivalent descriptions of double homogeneity and apply them to the construction of block designs and models for certain identities.

1. Homogeneous quasigroups.

In this section two theorems are obtained that generalize results concerning distributive quasigroups.

THEOREM 1.1. *There is no homogeneous quasigroup of order* $4k + 2$.

Proof. Let $(Q, \circ)$ be a homogeneous quasigroup of order $4k + 2$. We first construct out of this quasigroup an idempotent homogeneous quasigroup of order $4k + 2$.

Define $f: Q \to Q$ by $f(x) = x \circ x$. We assert that f is onto Q, and hence a bijection. Indeed, let a be a fixed element of Q, $b = a \circ a$, c an arbitrary element of Q, g an automorphism of $(Q, \circ)$ such that $g(b) = c$. We then have

$$c = g(b) = g(a \circ a) = g(a) \circ g(a) = f(g(a)) .$$

Thus f is onto Q.

We thus can define a quasigroup $(Q, \odot)$, isotopic to $(Q, \circ)$, by $f(x) \odot f(y) = x \circ y$. Since $f(x) \odot f(x) = x \circ x = f(x)$, $(Q, \odot)$ is idempotent. Moreover, if g is an automorphism of $(Q, \circ)$, it is also an automorphism of $(Q, \odot)$, since

Received January 11, 1963, and in revised form September 5, 1963.

This work was partly sponsored by the Air Force Office of Scientific Research.

$$g(f(x)) \odot f(y)) = g(x \circ y) = g(x) \circ g(y)$$

and

$$g(f(x) \odot g(f(y)) = (g(x \circ x)) \odot (g(y \circ y)) = (g(x) \circ g(x)) \odot (g(y) \circ g(y))$$
$$= f(g(x)) \odot f(g(y)) = g(x) \circ g(y) .$$

Thus $(Q, \odot)$ is an idempotent homogeneous quasigroup of order $4k + 2$. By ([5], p. 237), such quasigroups do not exist, and the theorem is proved.

As was shown in [6], if Q is a left-distributive quasigroup, then there is a quasigroup A' orthogonal to it. The next theorem generralizes this fact. The proof makes use of the notion of transversal for a quasigroup, $(Q, \circ)$, of order n. A transversal for $(Q, \circ)$ is a set $T \subset Q \times Q$, $T = \{(x_1, y_1), \cdots, (x_n, y_n)\}$ such that $x_i = x_{i'}$ implies $i = i'$, $y_j = y_{j'}$ implies $j = j'$, and $x_i \circ y_i = x_j \circ y_j$ implies $i = j$. It is easily seen that *there is a quasigroup orthogonal to* $(Q, \circ)$ *if and only if there are* n disjoint transversals for Q.

THEOREM 1.2. *If* $(Q, \circ)$ *is a quasigroup of order* n *possessing a transitive set of* n *automorphisms, then there is a quasigroup orthogonal to it.*

Proof. Let $\phi_1, \phi_2, \cdots, \phi_n$ be a transitive set of n automorphisms of $(Q, \circ)$ and $Q = \{b_1, b_2, \cdots, b_n\}$. We shall define n disjoint transversals for Q, $T(1), T(2), \cdots, T(n)$, where $T(k) \subset Q \times Q$, $k = 1, 2, \cdots, n$. Select $a \in Q$ and let

$$T(k) = \{(\phi_i(a), \phi_i(b_k)) \mid 1 \leqq i \leqq n\} .$$

The first coordinates of the n elements of $T(k)$ are distinct and so are the second coordinates; also $T(i) \cap T(j) = \phi$ if $i \neq j$.

It must be shown that $\phi_i(a) \circ \phi_i(b_k) = \phi_j(a) \circ \phi_j(b_k)$ implies that $i = j$. From the assumed equation it follows that $\phi_i(a \circ b_k) = \phi_j(a \circ b_k)$. Since the n automorphisms $\phi_1, \cdots, \phi_n$ are transitive on a set of n elements, it follows that if ϕ_i and ϕ_j agree on a single element of Q then $\phi_i = \phi_j$; thus $\phi_i = \phi_j$, and the theorem is proved.

2. Relations between doubly homogeneous quasigroups and nearfields. Consider a finite doubly homogeneous groupoid $(G, \circ)$. For any order n the two groupoids defined by $x \circ y = x$ or $x \circ y = y$ are doubly homogeneous (in fact any bijection of G is an automorphism of $(G, \circ)$). Also the groupoid of order 2 given by $1 \circ 1 = 2$, $2 \circ 2 = 1$, $1 \circ 2 = 2$, $2 \circ 1 = 1$, and its transpose are doubly homogeneous. We will show that the only other doubly homogeneous groupoids are quasigroups

THEOREM 2.1. *A doubly homogeneous groupoid* $(G, \circ)$ *is either*:
(i) *The groupoid defined by* $x \circ y = x$, *for all* $x, y \in G$,
(ii) *The groupoid defined by* $x \circ y = y$ *for all* $x, y \in G$,
(iii) *An idempotent doubly homogeneous quasigroup, or*
(iv) *A groupoid isomorphic to the groupoid defined above.*

Proof. First let us show that if the order of G is at least 3, then $(G, \circ)$ is idempotent. To do so, let $c, d \in G$, $c \neq d$, $c \circ c = d$. Let $e \in G$, $e \neq c, d$, and ϕ be an automorphism of $(G, \circ)$ such that

$$\phi(c) = c, \phi(d) = e .$$

Then we have

$$c \circ c = d \text{ and } c \circ c = \phi(c) \circ \phi(c) = \phi(c \circ c) = \phi(d) = e ,$$

a contradiction that implies $c \circ c = c$.

Assume that $a, b \in G$, $a \neq b$. If $a \circ b = a$, then the double homogeneity of $(G, \circ)$ implies that $x \circ y = x$ for all $x, y \in G$. Similarly, if $a \circ b = b$, then $x \circ y = y$ for all $x, y \in G$.

Consider finally the case, $a \circ b = c$, $c \neq a, b$. Double homogeneity implies that the equations $A \circ Y = C$ and $X \circ B = C$ have solutions, X, Y if $A \neq C$, $B \neq C$. Combining this with the idempotency of $(G, \circ)$, we see that if $(G, \circ)$ has order at least 3, then it is a quasigroup.

The case of order 2 is left to the reader.

In view of Theorem 2.1, we will examine doubly homogeneous quasigroups.

In the rest of this paper we will generally assume that all quasigroups are idempotent. An idempotent quasigroup that can be generated by two elements will be called a two-generated quasigroup. A two-quasigroup is a doubly homogeneous two-generated quasigroup. We will show that two-quasigroups and finite nearfields are closely related.

A finite near field, S, consists of a finite set S and two binary operations, $+$ and ., defined on all of S. The operation $+$ is an abelian group, the operation ., restricted to $S - \{0\}$ is a group, and left distributivity holds, $a(b + c) = ab + ac$. From these conditions it follows that $a0 = 0 = 0a$ and $(-1)a = -a = a(-1)$ (see [8, pp. 188–190]), and that the equation $ax + bx = c$ has a unique solution if $a + b \neq 0$. Moreover, it is implicit in [8] that a finite nearfield has a primitive element.

THEOREM 2.2. *If* $(S, \circ)$ *is a two-quasigroup, then there is a nearfield* $(S, +, .)$ *and primitive element* k *such that* $x \circ y = x + (y - x)k$.

The automorphisms of $(S, \circ)$ *are of the form* $\phi(x) = a + bx$.

Proof. The group G of automorphisms of $(S, \circ)$ is doubly transitive and only the identity automorphism fixes two elements of S. Such a group of permutations on a finite set determines a near field as follows ([9], p. 25, [2], pp. 385–388).

The elements of G leaving no elements fixed, together with the identity transformation, form an abelian, simply transitive normal subgroup N of G. Select an element $0 \in S$. We define $x + y$ as follows. There is a unique $\sigma \in N$, such that $\sigma(0) = x$; define $x + y$ to be $\sigma(y)$.

We define $x \cdot y$ as follows. Select $1 \in S$, $1 \neq 0$. Define $x \cdot y$ to be $\tau(y)$ where $\tau(0) = 0$, $\tau(1) = x$. Then $(S, +, .)$ is a nearfield. Moreover, since $\sigma(x) = x + b$ and $\tau(x) = ax$ $(a \neq 0)$ are automorphisms of $(S, \circ)$, then so is $\phi(x) = ax + b$. Since there are (n) $(n - 1)$ such ϕ's, where n is the cardinality of S, it follows that every automorphism of $(S, \circ)$ has the form $\phi(x) = ax + b$.

Next, we express the quasigroup $(S, \circ)$ in terms of the nearfield $(S, +, .)$ just constructed. Let $0 \circ 1 = k$. If $x, y \in S$, $x \neq y$, let ϕ be the automorphism of $(S, 0)$ such thar $\phi(0) = x$, $\phi(1) = y$, that is, $\phi(u) = x + (y - x)u$ for all $u \in S$. Then we have

$$x \circ y = \phi(0) \circ \phi(1) = \phi(0 \circ 1) = \phi(k) = x + (y - x)k, \quad (x \neq y).$$

Since $x \circ x = x + (x - x)k$, $(S, \circ)$ is of the asserted form.

COROLLARY 2.3. *A commutative two-quasigroup* $(Q, \circ)$ *is of* (*odd*) *prime order*, p, *and is expressible in terms of* $GF(p)$, *the Galois field of* p *elements, by the formula* $x \circ y = (x + y)/2$.

Proof. $(Q, \circ)$ is expressible in terms of a nearfield $(Q, +, .)$ by the formula $x \circ y = x + (y - x)k$. Since $(Q, \circ)$ is commutative, $0 \circ 1 = 1 \circ 0$. Thus

$$k = 0 \circ 1 = 1 \circ 0 = 1 - k\,,$$

hence

$$k + k = 1\,.$$

By left distributivity $k \cdot 2 = 1$. Now, the element 1 in any finite nearfield generates a Galois field with a prime number of elements, say p elements. The equation $k \cdot 2 = 1$ shows that $p \neq 2$ and that k is an element of that Galois field. Since k is a primitive element of $(Q, +, .)$, we see that $(Q, +, .)$ is the Galois field with p elements, and $x \circ y = x + (y - x)(1/2) = (x + y)/2$.

The next corollary relates doubly homogeneity to the identity

$(x \circ y) \circ (z \circ w) = (x \circ z) \circ (y \circ w)$, which has several names, including "the medial law".

COROLLARY 2.4. *A two-generated quasigroup* $(S, \circ)$ *of prime order* p, *is medial if and only if it is doubly homogeneous.*

Proof. If $(S, \circ)$ is doubly homogeneous, then it is of the form $x + (y - x)k$, for some nearfield. But the only near fields of prime order are the Galois fields. Thus $x \circ y = (1 - k)x + ky$ and a simple computation shows that satisfies the identity $(x \circ y) \circ (z \circ w) = (x \circ z) \circ (y \circ w)$ Hence $(S, \circ)$ is medial.

Conversely, if $(S, \circ)$ is medial, it is of the form $x \circ y = A(x) + B(y)$ where $(S, +)$ is an abelian group on p elements, and A and B are automorphisms of $(S, +)$ such that $A(x) + B(x) = x$, for all $x \in S$ (see [4]). But $(S, +)$ can be imbedded in the larger structure $(S, +, .)$, the Galois field of p elements, in such a way that every automorphism, ϕ, of $(S, +)$ is of the form $\phi x = ax$ for some $a \in S$. Thus $A(x) = (1 - k)x$ and $B(x) = kx$ for some k. Hence we have $x \circ y = x + (y - x)k$ and so $(S, \circ)$ is doubly homogeneous.

THEOREM 2.5. *Let* $(S, +, .)$ *be a finite nearfield and* $k \in S$, $k \neq 0, 1$. *Define a binary operation* $\circ$ *on* S *by* $x \circ y = x + (y - x)k$. *Then* $(S, \circ)$ *is a doubly homogeneous quasigroup.* $(S, \circ)$ *is a two-quasigroup if and only if* k *is a primitive element of* S.

Proof. It is easy to see that $(S, \circ)$ is a quasigroup. For example, if $x \circ y = x' \circ y$, then

$$x + (y - x)k = x' + (y - x')k$$

and so

$$(x - x') = (x - y)k + (y - x')k \ .$$

But we also have

$$(x - x') = (x - y)1 + (y - x')1 \ .$$

By the definition of a nearfield and the fact that $k \neq 1$, we obtain $x = x'$.

For $a, b \in S$, $a \neq 0$, define $\phi\colon\ S \to S$ by $\phi(x) = ax + b$. Each ϕ is an automorphism of $(S, \circ)$ and the collection of all such ϕ's is doubly transitive on S. Thus $(S, \circ)$ is a doubly homogeneous quasigroup.

If $(S, \circ)$ is a two-quasigroup, it is generated, as a quasigroup, by any two of its elements, in particularly by $\{0, 1\}$. Now, the nearfield in $(S, +, .)$ generated by k contains 0 and 1; thus k is a primitive element of $(S, +, .)$ Finally, we must show that if k is a primitive element of $(S, +, .)$, then $\{0, 1\}$ generates the quasigroup $(S, \circ)$. To do so, let $(T, \circ)$ be the subquasigroup of $(S, \circ)$ generated by $\{0, 1\}$. We will show that $T = S$.

First of all, $(T, \circ)$ is doubly homogeneous. Indeed, if $a, b \in T$, $a \neq b$, and ϕ is an automorphism of $(S, \circ)$ such that $\phi(0) = a$, $\phi(1) = b$, then $\phi(T)$ is contained in the quasigroup generated by $\{a, b\}$. Since T and $\phi(T)$ have the same cardinality, $\phi(T) = T$, and $\phi \mid T$ is an automorphism of $(T, \circ)$, taking 0 into a, and 1 into b.

Thus, by Theorem 2.2, $(T, \circ)$ is related to a nearfield $(T, \oplus, \odot)$ by the formula $x \circ y = x \oplus (y \ominus x) \odot k'$, where $(T, \oplus, \odot)$ can be chosen to have the same 0 and 1 as $(S, +, .)$ [$\ominus$ denotes subtraction in $(T, \oplus, \odot)$]. We will show that $\oplus$ and $\odot$ are restrictions of $+$ and $.$, and thus $(T, \oplus \odot)$ is a subnearfield of $(S, +, .)$.

Note first that since $0 \oplus (1 \ominus 0) \odot k' = 0 \circ 1 = 0 + (1 - 0)k$, we have $k = k'$ and thus $k \in T$. Next we will show that $x \oplus y = x + y$ and $x \odot y = x \cdot y$ for all $x, y \in T$.

For $x = 0$, it is obvious that $x \oplus y = x + y$. Let $x \in T$, $x \neq 0$, and $\phi\colon S \to S$ be the automorphism of $(S, \circ)$ given by $\phi(y) = x + y$. Then $\phi \mid T$ is an automorphism of $(T, \circ)$ without fixed points. Thus $(\phi \mid T)y = u \oplus y$ for some fixed $u \in T$ and all $y \in T$. Since $u = (\phi \mid T)\ (0) = \phi(0) = x$, we have $u = x$. Thus $x \oplus y = (\phi \mid T)y = \phi(y) = x + y$ for all $x, y \in T$.

To show $x \odot y = x \cdot y$ for all $x, y \in T$, we proceed similarly. For $x = 0$ or 1 the statement is trivial. Let $x \neq 0, 1$, $x \in T$. Let $\phi : S \to S$ be defined by $\phi(y) = x \cdot y$. Then $\phi \mid T$ is an automorphism of $(T, \circ)$ with the one fixed element, 0. Thus $(\phi \mid T)(y) = u \odot y$ for some u. Since $u = u \odot 1 = (\phi \mid T)(1) = \phi(1) = x \cdot 1 = x$, we have $u = x$. Hence $x \odot y = (\phi \mid T)y = \phi(y) = x \cdot y$, for all $x, y \in T$.

Thus $(T, \oplus, \odot)$ is a subnearfield of $(S, +, .)$ and contains the element k. Since k is a primitive element of the nearfield S, we must have $S = T$. Thus $(S, \circ)$ is generated by $\{0, 1\}$ and therefore is a two-quasigroup.

COROLLARY 2.6. *If k is a primitive element of a nearfield S, then $\{0, k\}$ generates S by the single binary operation $x \circ y = x + (y - x)k$.*

The relation between quasigroups and near fields is shown further in the following theorems. For simplicity if k is an element of a nearfield $(Q, +, .)$, then the quasigroup $(Q, \circ)$ defined by $x \circ y = x + (y - x)k$ we denote $Q(k)$.

THEOREM 2.7. *If $(Q, +, .)$ is a nearfield, $k, k' \in Q$ and $\phi\colon Q \to Q$*

is an automorphism of $(Q, +, .)$ *such that* $\phi(k) = k'$, *then* ϕ *is an isomorphism between* $Q(k)$ *and* $Q(k')$.

Proof. Let $\circ$ denote multiplication in $Q(k)$ and $\odot$ denote multiplication in $Q(k')$. Then $\phi(x \circ y) = \phi(x + (y - x)k) = \phi(x) + (\phi(y) - \phi(x))k' = \phi(x) \odot \phi(y)$. Thus ϕ is an isomorphism of $Q(k)$ onto $Q(k')$.

The next theorem is the converse of Theorem 2.7.

THEOREM 2.8. *If* $(Q, +, .)$ *is a nearfield*, k, k' *are primitive elements of* $(Q, +, .)$, *and* $Q(k)$ *is isomorphic to* $Q(k')$, *then there is an automorphism* ϕ *of* $(Q, +, .)$ *such that* $\phi(k) = k'$.

Proof. Let $\alpha: Q(k) \to Q(k')$ be an isomorphism between the quasigroups $Q(k)$ and $Q(k')$. Let $\circ$ and $\odot$ be the operations in $Q(k)$, $Q(k')$ respectively. Since $Q(k)$ is doubly homogeneous, we may assume that $\alpha(0) = 0$ and $\alpha(1) = 1$. Then

$$\alpha(k) = \alpha(0 \cdot 1) = \alpha(0) \odot \alpha(1) = 0 \odot 1 = k'$$

We will show that α is an automorphism of $(Q, +, .)$.

Let σ be an antomorphism of $Q(k)$ defined by $\sigma(x) = x + b$, $b \neq 0$. Then, $\alpha\sigma\alpha^{-1}$, being an automorphism of $Q(k')$ and having no fixed elements, is of the form $x \to x + c$ for some fixed c. Thus $\alpha\sigma(t) = \alpha(t) + c$ for all $t \in Q$; equivalently, $\alpha(t + b) = \alpha(t) + c$. In particular, $\alpha(b) = \alpha(0 + b) = \alpha(0) + c = c$, and we have $\alpha(t + b) = \alpha(t) + \alpha(b)$. That is, α is an automorphism of $(Q, +)$.

Similarly, let $\sigma: Q(k) \to Q(k)$ be given by $\sigma(x) = ax$. Since σ is an automorphism of $Q(k)$ with $\sigma(0) = 0$, $\alpha\sigma\alpha^{-1}$ is an automorphism τ of $Q(k')$ with $\tau(0) = 0$. Thus $\tau(x) = a'x$ for some $a' \in Q$. We have $\alpha\sigma(x) = \tau\alpha(x)$, or equivalently, $\alpha(ax) = a'\alpha(x)$. But $\alpha(1) = 1$; hence $\alpha(a) = \alpha(a \cdot 1) = a'\alpha(1) = a' \cdot 1 = a'$. Thus $\alpha(ax) = \alpha(a)\alpha(x)$, and α is an automorphism of $(Q, \cdot)$. This ends the proof.

As another application of Theorem 2.2 we have

THEOREM 2.9. *A left-distributive two-quasigroup is medial* (*hence right-distributive*).

Proof. Let $(Q, \circ)$ be a left-distributive two-quasigroup. By Theorem 2.2, $x \circ y = x + (y - x)k$ for some nearfield $(Q, +, .)$. Since left translation by 0 is an automorphism of $(Q, \circ)$, there exist $a, b \in Q$ such that

$$0 \circ x = a + bx \qquad \text{for all } x \in Q.$$

Thus $xk = a + bx$ for all $x \in Q$.

It is easy to see that $a = 0$ and $b = k$, by letting $x = 0, 1$. Thus $xk = kx$ for all $x \in Q$. Since k is a primitive element of $(Q, +, .)$, the nearfield in question is commutative, hence a field. The theorem follows immediately.

It might be remarked that a quasigroup and its conjugates [5] have the same automorphisms. Thus the conjugate of a two-quasigroup is a two-quasigroup. If $x \circ y = z$ then two of the six conjugate operations, α and β, are defined by $x\alpha z = y$ and $y\beta z = x$. Here α and β denote division on the left and right respectively. It turns out that α and β are easily expressed in terms of the nearfield describing $\circ$. For if $x \circ y = x + (y - x)k = z$, then $y = x + (z - x)k^{-1}$. Also, it can be shown that if $x \circ y = z$ then $x = y + (z - y)(1 - k)^{-1}$.

3. Two-homogeneity and identities. Let Q be a finite idempotent quasigroup and $\Phi(Q)$ be the identities valid on Q [7]. Let F be the free groupoid on two generators x, y and $F(Q)$ be the homomorphic image of F obtained from F through factoring F by $\Phi(Q)$. That is, define an equivalence relation, $\sim$, on F as follows: If U, $V \in F$ and $U = V$ is an identity valid on Q, write $U \sim V$. Then $F(Q)$ is $F/\sim$. It is easily seen that $F(Q)$ is a finite idempotent quasigroup. Note also that if $U \in F$ and $a, b \in Q$, then replacement of x and y in U by a and b defines an element in Q; we denote this element, $U(a, b)$. We may denote U itself as $U(x, y)$. If $U \in F$, then U determines a unique element of $F(Q)$, denoted $\tilde{U}$.

THEOREM 3.1. *Let Q be a quasigroup generated by $\{a, b\}$, and assume that for all U, $V \in F$ such that $U(a, b) = V(a, b)$, one also has the identity $U(x, y) = V(x, y)$ valid on Q. Then Q is isomorphic to $F(Q)$. The converse holds.*

Proof. Since Q satisfies all the identities that $F(Q)$ satisfies, there is a homomorphism $h: F(Q) \to Q$ such that $h(\tilde{x}) = a$, $h(\tilde{y}) = b$. Also we can define a function $k: Q \to F(Q)$ by setting $k(a) = \tilde{x}$ and $k(b) = \tilde{y}$, and extending this assignment to a homomorphism. (The possibility of defining this k is equivalent to the hypothesis made on a and b.) Clearly h and k are inverse to each other, hence isomorphisms.

Conversely, assume that $h: F(Q) \to Q$ is an isomorphism. Let $a = h(\tilde{x})$, $b = h(\tilde{y})$. If $U(a, b) = V(a, b)$, then $h[\widetilde{U(x, y)}] = h[\widetilde{V(x, y)}]$. Since h is an injection, $\widetilde{U(x, y)} = \widetilde{V(x, y)}$. Thus $U \sim V$, which was to be proved.

COROLLARY 3.2. A two-quasigroup Q is isomorphic to $F(Q)$.

It should be noted that for a quasigroup Q, $F(Q)$ is doubly homogeneous if and only if it is generated by any pair of elements. And when $F(Q)$ is a two-quasigroup, any two elements of Q generate a quasigroup isomophic to $F(Q)$.

COROLLARY 3.3. *Two two-quasigroups are isomorphic if and only if they have the same identities in two variables.*

COROLLARY 3.4. *A two-generated quasigroup Q is doubly homogeneous if and only if for all distinct $a, b \in Q$ and all distinct $c, d \in Q$, $U(a, b) = V(a, b)$ is equivalent to $U(c, d) = V(c, d)$ for all terms U, V in two variables.*

Proof. Clearly, if $\{a, b\}$ generates Q, so does $\{c, d\}$. Then apply Theorem 3.1 and the remarks preceding Corollary 3.3.

As already mentioned, a two-quasigroup is defined by its identities in two variables. In fact, if Q is a two-quasigroup of order n, then Q can be defined by $n^2 - n + 1$ identities, namely the identity $X^2 = X$ and an identity corresponding to each product $u_i(a, b) \cdot u_j(a, b) = u_k(a, b)$, $i \neq j$, where each element of Q is represented in the form $u_s(a, b)$ for some term in a and b. Let us consider, for example, the only two-quasigroup of order four, Q, given by:

	a	b	ab	ba
a	a	ab	ba	b
b	ba	b	a	ab
ab	b	ba	ab	a
ba	ab	a	b	ba

Since $a \cdot ab = ba$ and $ab \cdot ba = a$, Q satisfies the identities:

(i) $X \cdot XY = YX$ and

(ii) $XY \cdot YX = X$. We will show that (i) and (ii) are sufficient to reconstruct the multiplication table for Q. This will be useful in § 4.

THEOREM 3.5. *A finite groupoid Q' satisfying the identities* (i), *$X \cdot XY = XY$ and* (ii), *$XY \cdot YX = X$ is a quasigroup. Moreover any two distinct element $a, b \in Q'$ generate a quasigroup Q'' described by the preceding multiplication table.*

Proof. Let L and R be a left-and right translation in Q' by the same element. By (i), $LL = R$. We prove that L is an injection.

Let $c, d, e \in Q'$ and $cd = ce$. We will show that $d = e$. We have $c \cdot cd = c \cdot ce$ and, by (i), $dc = ec$. Thus $dc \cdot cd = ec \cdot ce$. By (ii),

$d = e$. Thus Q' is a quasigroup.

Since Q' satisfies $X \cdot XY = YX$, it satisfies $X \cdot XX = XX$. Since Q' is a quasigroup it must therefore satisfy $XX = X$; thus Q' is idempotent.

We next show that distinct elements of Q' do not commute. Assume that $c, d \in Q'$, $cd = dc$. Then, by (ii) we have $c = cd \cdot dc = dc \cdot cd = d$.

Now let us examine the quasigroup Q'' generated by a and b. First of all, Q'' is an idempotent quasigroup and $ab \neq ba$. Thus Q'' has at least the four distinct elements a, b, ab, ba. We will show that Q'' has no more elements.

From (i) and (ii) we obtain $XY(XY \cdot YX) = XY \cdot X$, hence $YX \cdot XY = XY \cdot X$ and thus $Y = XY \cdot X$. From $Y = XY \cdot X$ follows $Y = X \cdot YX$ [7]. Also, $XY \cdot Y = XY(XY \cdot X) = X \cdot XY = YX$.

From these identities follow: $aa = a$, $bb = b$, $ab \cdot ab = ab$, $ba \cdot ba = ba$; $a \cdot ab = ba$, $a \cdot ba = b$, $b \cdot ab = a$, $b \cdot ba = ab$; $ab \cdot a = b$, $ab \cdot b = ba$, $ab \cdot ba = a$; $ba \cdot a = ab$, $ba \cdot b = a$, $ba \cdot ab = b$. Thus Q' has only the four elements a, b, ab, ba. Moreover its multiplication table is the one already given.

4. Block designs and quasigroups. By a pairwse balanced incomplete block design on a set S we will mean a family of subsets $B_1, B_2, \cdots, B_r$ of S, each containing the same number of elements, $k \geq 3$, such that each pair of elements of S is a subset of exactly one of the B's. If $(S, \circ)$ is a doubly homogeneous quasigroup, then the two-generated subquasigroups of S form a pairwise balanced incomplete block design (for brevity, block design). Calling the cardinality of S, v, we then have a doubly transitive block design $B(k, v)$ where k, incidentally, is a power of a prime. The following theorems show various relations between block designs and algebraic aspects of quasigroups.

THEOREM 4.1. *A two-generated quasigroup Q is doubly homogeneous (hence a two-quasigroup) if and only if the two-generated subquasigroups of $Q \times Q$ all have the same order.*

Proof. Assume that Q is a two-quasigroup of cardinality q. Consider the quasigroup $Q^* \subset Q \times Q$ generated by $\{(a, c), (b, d)\}$, where a, b, c, and d are distinct. Let $\pi : Q \times Q \to Q$ be the projection defined by $\pi(q_1, q_2) = q_1$. Then $\pi(Q^*) = Q$ since Q is generated by any two of its elements, in particular, a and b. Now, for any U and V, terms in the variables x and y, $U((a, c), (b, d)) = V((a, c), (b, d))$ if and only if, $U(a, b) = V(a, b)$ and $U(c, d) = V(c, d)$. By Corollary 3.4, $U(a, b)$

$= V(a, b)$ if and only if $U(c, d) = V(c, d)$. Thus π is an isomorphism onto Q, and $\{(a, c), (b, d)\}$ generates a quasigroup of order q. Special cases such as $\{(a, b), (b, b)\}$, $\{(a, b), (a, b)\}$ or $\{(a, b), (c, a)\}$ are easily disposed of.

Conversely, assume that Q^*, of order q, is two-generated and that every two elements of $Q \times Q$ generate a quasigroup of the same order, necessarily q. We will show that Q is doubly homogeneous. Let $\{a_1, a_2\}$ and $\{b_1, b_2\}$ be two distinct pairs of elements of Q, $a_1 \neq a_2$, $b_1 \neq b_2$. Then $a = (a_1, a_2)$ and $c = (b_1, a_2)$ generate a quasigroup of order q; thus a and $b = (b_1, b_2)$ generate a quasigroup Q^* such that $\pi(Q^*) = Q$. This implies that two elements of Q^* are equal if their first coordinates are equal. Thus $U(a_1, b_1) = V(a_1, b_1)$ is equivalent to $U(a_2, b_2) = V(a_2, b_2)$. By Corollary 3.4, Q is a two-quasigroup.

The notion of two-quasigroup can be used to give a simple proof of the following combinatorial theorem due to Skolem [1, p. 183].

THEOREM 4.2 *If k is a prime power and $B(k, v_1)$ and $B(k, v_2)$ exist, then $B(k, v_1v_2)$ exists.*

Proof. Let $B(k, v_i)$ be a block design on the set S_i, $i = 1, 2$. Select a two-quasigroup Q of order k. On each block of $B(k, v_1)$ and $B(k, v_2)$ define a quasigroup isomorphic to Q. This defines on S_i a quasigroup Q_i, $i = 1, 2$, such that every two elements of S_i generate a quasigroup isomorphic to Q. Every two elements of $Q_1 \times Q_2$ generate a quasigroup R satisfying all the identities that Q satisfies. Since $Q = F(Q)$, R is a homomorphic image of Q. As a two-quasigroup contains no proper subquasigroups, (other than those with one element), R is isomorphic to Q. This shows that on $S_1 \times S_2$ there is a $B(k, v_1v_2)$.

THEOREM 4.3. *There is a quasigroup of order v satisfying the identities $X \cdot XY = YX$ and $XY \cdot YX = X$ if and only if $v = 12n + 1$ or $v = 12n + 4$.*

Proof. Recalling the example at the end of § 3 and the argument in the proof of Theorem 4.2, we see that such quasigroups exist if and only if there is a $B(4, v)$. As Hanani proved in [3], a $B(4, v)$ exists if and only if $v = 12n + 1$ or $v = 12n + 4$.

Similar reasoning shows that if an identity in two letters has a two-quasigroup model of order k, and there is a $B(k, v)$, then the identity has a model of order v. In particular, since $X \cdot XY = YX$ has a two-quasigroup model of order 5, it has, by [3], models of all orders of the form $20n + 1$ or $20n + 5$ (except possibly 141).

REFERENCES

1. R. C. Bose and S. S. S. Shrikhande, *On the composition of incomplete block designs*, Canadian Math. J., **12** (1960) 177-188.
2. M. Hall, *The theory of groups*, Macmillan, NY, 1959.
3. H. Hanani, *The existence and construction of balanced incomplete block designs*, Ann. Math. Stat., **32** (1961), 361-386.
4. D. C. Murdoch, *Structure of abelian quasigroups*, Trans. Amer. Math. Soc., **49** (1951), 392-409.
5. S. Stein, *On the foundations of quasigroups*, Trans. Amer. Math. Soc., **85** (1957), 228-256.
6. ———, *Left-distributive quasigroups*, Proc. Amer. Math. Soc., **10** (1959), 557-558.
7. ———, *Finite models of identities*, ibid, **14** (1963), 216-222.
8. H. Zassenhaus, *Über endliche Fastkörper*, Abh. Math. Sem. Hamburg, **11** (1935), 187-220.
9. ———, *Kennzeichnung endlicher linearer Gruppen als Permutationsgruppen*, ibid, pp. 17-40.

UNIVERSITY OF CALIFORNIA AT DAVIS

ON THE LOCATION OF THE ZEROS OF SOME INFRAPOLYNOMIALS WITH PRESCRIBED COEFFICIENTS

J. L. WALSH AND O. SHISHA

1. Various results have been obtained regarding the zeros of infrapolynomials with prescribed coefficients. (See e.g. [Walsh, 1958], [Walsh and Zedek, 1956], [Fekete and Walsh, 1957], [Shisha and Walsh, 1961, 1963], and [Shisha, 1962]). Our purpose in the present note is twofold:

(i) to contribute more deeply to that study, making use of some properties of polynomials and rational functions, and

(ii) conversely, further to show how results concerning infrapolynomials can be used in the investigation of some rational functions and in particular some combinations of a polynomial and its derivative.

2. We repeat here the underlying definition. Let n and q be natural numbers ($q \leqq n$), $n_1, n_2, \cdots, n_q$ integers such that $0 \leqq n_1 < n_2 \cdots < n_q \leqq n$, and S a pointset in the (open) complex plane. An nth infrapolynomial on S with respect to $(n_1, n_2, \cdots, n_q)$ is a polynomial $A(z) \equiv \sum_{\nu=0}^{n} a_\nu z^\nu$ having the property: There does not exist a polynomial $B(z) \equiv \sum_{\nu=0}^{n} b_\nu z^\nu$ such that $B(z) \not\equiv A(z)$, $b_{n_\nu} = a_{n_\nu}$ $(\nu = 1, 2, \cdots, q)$, $|B(z)| < |A(z)|$ whenever $z \in S$ and $A(z) \neq 0$, and $B(z) = 0$ whenever $z \in S$ and $A(z) = 0$.

3. Of special importance among the above sequences $(n_1, n_2, \cdots, n_q)$, are "simple n-sequences" [Shisha and Walsh, 1961]. Given a natural number n, we define a "simple n-sequence" to be a sequence having one of the forms $(0, 1, \cdots, k, n-l, n-l+1, \cdots, n)$ $[k \geqq 0, l \geqq 0, k + l + 2 \leqq n]$; $(0, 1, \cdots, k)$ $[0 \leqq k < n]$; $(n-l, n-l+1, \cdots, n)$ $[0 \leqq l < n]$. We shall consider nth infrapolynomials on some special sets S with respect to simple n-sequences σ. The sets S will consist of $n - s + 2$ points, where s is the number of elements of σ, and S will be required not to contain the origin, in case σ contains zero. As explained in the Introduction to the last mentioned paper, this particular situation is of special importance, as the general case is to a large extent reducible to it, and as these particular nth infrapolynomials are closely related to certain combinations of a polynomial and its derivative. Numerous results on such combinations exist in the literature.

Received May 2, 1963. This research was supported (in part) by the U.S. Air Force through the Air Force Office of Scientific Research.

4. THEOREM. *Let n be a natural number, σ a simple n-sequence, s its number of elements. Let $S = \{z_1, z_2, \cdots, z_{n-s+2}\}$ be a set of $n - s + 2$ (distinct) points of the (open) complex plane, and set $g(z) \equiv \prod_{\nu=1}^{n-s+2}(z - z_\nu)$. In case $\sigma = (0, 1, \cdots, k)$ or $\sigma = (0, 1, \cdots, k, n - l, n - l + 1, \cdots, n)$ set $K = k + 1$. In case $0 \notin \sigma$, set $K = 0$.* (Thus $K = \min[\nu, \nu \notin \sigma, \nu = 0, 1, 2, \cdots]$). *Also, in case $0 \in \sigma$, assume $0 \notin S$. Let $A(z) \equiv \sum_{\nu=0}^{n} a_\nu z^\nu$ be an nth infrapolynomial on S with respect to σ.*

Then [by Theorem 1, Shisha and Walsh, 1961] *one can set*

$$A(z) \equiv P(z)g(z) + \alpha z^K \sum_{\nu=1}^{n-s+2} \lambda_\nu g(z)/(z - z_\nu) . \tag{1}$$

Here $\lambda_1, \lambda_2, \cdots, \lambda_{n-s+2}$ are nonnegative reals with $\sum_{\nu=1}^{n-s+2}\lambda_\nu = 1$, α is a complex number, and $P(z)$ is a polynomial of degree[1] *$\leqq s - 1$ such that $P(z)g(z) + \alpha z^{K+n-s+1}$ is of degree $\leqq n$.*[2]

I. *Let S be contained in a disc $C: |z - c| \leqq r$. Then every zero $\zeta (\notin C)$ of $A(z)$ satisfies*

$$| P(\zeta) + (\overline{\zeta - c})\alpha\zeta^K/\{| \zeta - c |^2 - r^2\} | \leqq r | \alpha\zeta^K |/\{| \zeta - c |^2 - r^2\} . \tag{2}$$

If $K = 0$, and if a zero ζ of $A(z)$ satisfies $r < \rho_1 \leqq |\zeta - c| \leqq \rho_2$, then $|\alpha|/\{\rho_2 + r\} \leqq |P(\zeta)| \leqq |\alpha|/(\rho_1 - r)$ i.e. *(in case $\alpha \neq 0$ and $P(z)$ is not a constant) ζ lies in the closed interior of the lemniscate $|P(z)| = |\alpha|/(\rho_1 - r)$, and in the closed exterior of the lemniscate $|P(z)| = |\alpha|/\{\rho_2 + r\}$.*

II. *Let $P(z) \equiv \beta z^t + \gamma z^{t-1} + \cdots$ $(t \geqq 0, \beta \neq 0)$, and suppose that S and all the zeros of $P(z)$ lie in some closed disc C, and that $\alpha \neq 0$, $K = 0$. Let $w_1, w_2, \cdots, w_{t+1}$ be distinct solutions of $w^{t+1} = -\alpha/\beta$. Then every zero $(\notin C)$ of $A(z)$ lies in $\bigcup_{\nu=1}^{t+1}(w_\nu + C)$.*[3]

III. *Suppose that $A(z)$ is a real polynomial,*[4] *and that $\alpha \neq 0$. Assume, furthermore, that $P(z)/(\alpha z^K)$ is of the form $A + \sum_{\nu=1}^{p} A_\nu z^\nu + \sum_{\nu=1}^{q} B_\nu z^{-\nu}$ with all $Re(A_\nu) \leqq 0$ and all $Re(B_\nu) \geqq 0$. Let z_0 be a non-real zero of $A(z)$ satisfying $|\arg z_0| \leqq \min(\pi/p, \pi/q)$.*[5] *Then z_0 belongs to at least one (Jensen) disc*

$$\left| z - \frac{1}{2}(z_\nu + \overline{z_\nu}) \right| \leqq \frac{1}{2} | z_\nu - \overline{z_\nu} | . \tag{3}$$

[1] Degree of a polynomial means its exact degree. The polynomial 0 is assigned the degree -1.

[2] One can show that α and $P(z)$ are uniquely determined, and in case $\alpha \neq 0$, so are $\lambda_1, \lambda_2, \cdots, \lambda_{n-s+2}$.

[3] $w_\nu + C$ denotes the closed disc consisting of all points $w_\nu + z$, $z \in C$.

[4] i.e. the coefficients of $A(z)$ are real.

[5] arg denotes the principal value of the argument.

In particular, if $p = q = 1$, *every non-real zero of* $A(z)$ *belongs to at least one of these discs.*

IV. *Suppose that* $A(z)$ *is a real polynomial,* $\alpha \neq 0$, *and that* $P(z)/(\alpha z^K)$ *is of the form* $\sum_{\nu=0}^{p} A_\nu z^\nu + \sum_{\nu=1}^{q} B_\nu z^{-\nu}$ $(p \geqq 0,\ q \geqq 2)$ *with all* $Re(A_\nu) \leqq 0$ *and all* $Re(B_\nu) \geqq 0$. *Suppose furthermore that* $\lambda_\nu > 0$ *implies* $Re(z_\nu) > 0$ $(\nu = 1, 2, \cdots, n - s + 2)$. *Let* z_0 *be a non-real zero of* $A(z)$ *satisfying* $|\arg z_0| \leqq \min\{\pi/(p+1), \pi/(q-1)\}$. *Then*:

A. *There exists a* ν, $1 \leqq \nu \leqq n - s + 2$, $Im(z_\nu) \neq 0$, *such that* z_0 *belongs to the closed interior of the circle passing through* z_ν *and* $\overline{z_\nu}$ *and tangent to the line* $0z_\nu$.

B. *If neither* z_0 *nor* $\overline{z_0}$ *belongs to* S, *one can choose* ν *so that* $\lambda_\nu > 0$, *and therefore* $Re(z_0) > 0$.

V. *Suppose that* S *is a real set contained in a finite interval* $J: x_1 \leqq x \leqq x_2$, *that* $A(z)$ *is a real polynomial, and that* $K = 0$. *Suppose* $P(z)$ *is of the form* $\beta z^t + \gamma z^{t-1} + \cdots$ $(t \geqq 0, \beta \neq 0)$, *and that all zeros of* $P(z)$ *lie in the above interval. Then every real zero* $(\notin J)$ *of* $A(z)$ *is of the form* $\xi + \omega$ *where* $\xi \in J$ *and* ω *is a real number satisfying* $\omega^{t+1} = -\alpha/\beta$. *Thus, if* t *is odd and* $\alpha\beta > 0$, *all real zeros of* $A(z)$ *lie in* J.

5. *Proof of Part I.* Let ζ $(\notin C)$ be a zero of $A(z)$. Then by (1),

$$P(\zeta) + \alpha\zeta^K \sum_{\nu=1}^{n-s+2} \lambda_\nu/(\zeta - z_\nu) = 0 .$$

By a result due to Walsh [cf. 1950, § 1.5.1, Lemma 1]

$$\sum_{\nu=1}^{n-s+2} \lambda_\nu/(\zeta - z_\nu) = 1/(\zeta - z'), \quad z' \in C . \tag{4}$$

By an elementary mapping property of the function $1/z$ we have

$$|1/(\zeta - z') - (\overline{\zeta - c})/\{|\zeta - c|^2 - r^2\}| \leqq r/\{|\zeta - c|^2 - r^2\} ,$$

from which (2) follows. The rest of part I is easily obtained from (2).

Proof of Part II. Let ζ $(\notin C)$ be a zero of $A(z)$. Again we have a relation (4), which implies $P(\zeta)(\zeta - z') = -\alpha$. Furthermore, the last left hand side can be written [Walsh, 1922] $\beta(\zeta - \eta)^{t+1}$ with $\eta \in C$. Hence $\zeta \in \bigcup_{\nu=1}^{t+1} (w_\nu + C)$.

Proof of Part III. We may assume $g(z_0) \neq 0$, $g(\overline{z_0}) \neq 0$. Since $\overline{A(\overline{z_0})} = A(z_0) = 0$, we have by (1),

$$\begin{aligned} 0 &= \overline{P(\overline{z_0})/(\alpha \overline{z_0}^K)} + \sum_{\nu=1}^{n-s+2} \lambda_\nu/(z_0 - \overline{z_\nu}) \\ &= \bar{A} + \sum_{\nu=1}^{p} \overline{A_\nu} z_0^\nu + \sum_{\nu=1}^{q} \overline{B_\nu} z_0^{-\nu} + \sum_{\nu=1}^{n-s+2} \lambda_\nu/(z_0 - \overline{z_\nu}) \\ &= A + \sum_{\nu=1}^{p} A_\nu z_0^\nu + \sum_{\nu=1}^{q} B_\nu z_0^{-\nu} + \sum_{\nu=1}^{n-s+2} \lambda_\nu/(z_0 - z_\nu) \\ &= 2Re(A) + \sum_{\nu=1}^{p} 2Re(A_\nu) z_0^\nu + \sum_{\nu=1}^{q} 2Re(B_\nu) z_0^{-\nu} \\ &\quad + \sum_{\nu=1}^{n-s+2} \lambda_\nu\{(z_0 - z_\nu)^{-1} + (z_0 - \overline{z_\nu})^{-1}\} . \end{aligned}$$

By theorem 21 [Shisha and Walsh, 1961], there exists a ν (with $\lambda_\nu > 0$) such that z_0 lies in (3).

Similarly, using Theorem 22 [loc. cit.] one proves Part IV.[6]

Proof of Part V. Let ζ ($\notin J$) be a real zero of $A(z)$. Then $P(\zeta) + \alpha \sum_{\nu=1}^{n-s+2} \lambda_\nu/(\zeta - z_\nu) = 0$. Now, $\sum_{\nu=1}^{n-s+2} \lambda_\nu/(\zeta - z_\nu)$ can be written as $1/(\zeta - x')$, $x' \in J$. Also, since all zeros of $P(z)$ lie in J, one can set $P(\zeta)(\zeta - x') = \beta(\zeta - \xi)^{t+1}$, $\xi \in J$. Setting $\omega = \zeta - \xi$, we have $\zeta = \xi + \omega$, $\omega^{t+1} = -\alpha/\beta$.

6. We apply now our results to some special cases. We continue to assume the contents of the first paragraph of the Theorem. Thus, the contents of the second paragraph of the Theorem hold, too.

(a) Suppose $\sigma = (n)$. If $a_n = 0$ then $A(z) \equiv 0$, for otherwise the polynomial $B(z) \equiv 0$ would fulfill the properties stated at the end of § 2. We thus assume that $a_n \neq 0$. Then $a_n^{-1}A(z)$ is an infrapolynomial ("Extremalpolynom") on S in the sense of Fekete and von Neumann [1922]. Also one easily sees that $P(z) \equiv 0$, $\alpha = a_n$. By a known result [loc. cit., p. 138, cf. also Fejér 1922] all zeros of $A(z)$ belong to the convex hull of S. Thus Parts I, II and V of the Theorem do not apply. Parts III and IV do apply; but they can be derived from known results [Fekete and von Neumann 1922 p. 138, and Walsh 1958 p. 305]. Thus, if z_0 is a non-real zero of $A(z)$, and if $A(z)$ is a real polynomial, then z_0 belongs to at least one of the discs (1). If, in addition, $\lambda_\nu > 0$ implies $Re(z_\nu) > 0$ ($\nu = 1, 2, \cdots, n+1$), then A and B of Part IV hold.

(b) Suppose $\sigma = (n-1, n)$. Then $s = 2$, $K = 0$ and [Shisha and Walsh 1961, p. 146]

[6] Observe that if (i) $A(z)$ is a real polynomial, (ii) $\alpha \neq 0$, and (iii) S is symmetric in the axis of reals, then (i) α is real, (ii) $\lambda_\nu = \lambda_\mu$ if $z_\nu = \overline{z_\mu}$, and (iii) $g(z)$ and $P(z)$ are real polynomials. Indeed, suppose $z_\nu = \overline{z_\mu}$. Then (1) yields $\alpha z_\nu^K \lambda_\nu g'(z_\nu) = A(z_\nu) = \overline{A(z_\mu)} = \bar{\alpha} z_\nu^K \lambda_\mu g'(z_\nu)$. Thus, if α is real, $\lambda_\nu = \lambda_\mu$. To prove that α is real, choose ν_0, μ_0 so that $\lambda_{\nu_0} > 0$ and $z_{\nu_0} = \overline{z_{\mu_0}}$. Then $(\lambda_{\nu_0} + \lambda_{\mu_0}) Im(\alpha) = 0$, and therefore $Im(\alpha) = 0$. From (1) we see now that $P(z)g(z)$ is a real polynomial; therefore, so is $P(z)$.

$$A(z) \equiv a_n g(z) + \left(a_{n-1} + a_n \sum_{\nu=1}^{n} z_\nu\right) \sum_{\nu=1}^{n} \lambda_\nu g(z)/(z - z_\nu) ,$$

$\lambda_\nu \geqq 0$, $\sum_{\nu=1}^{n} \lambda_\nu = 1$. Thus, $P(z) \equiv a_n$ and $\alpha = a_{n-1} + a_n \sum_{\nu=1}^{n} z_\nu$. One can apply Part I. Part II implies that if $a_n \neq 0$ and if S lies in a closed disc C, then every zero ($\notin C$) of $A(z)$ lies in $-(\alpha/a_n) + C$. This, however, is a known result [loc. cit. Theorem 14, cf. also Walsh 1922 Theorem VI]. Again, the information we obtain from Part III follows from known results [Fekete and Walsh 1957 Theorem X, Fekete and von Neumann 1922 p. 138]. Assume that $A(z)$ is a real polynomial, $\alpha \neq 0$, $\lambda_\nu > 0$ implies $Re(z_\nu) > 0$ ($\nu = 1, 2, \cdots, n$), and $Re(a_n/\alpha) \leqq 0$ (i.e. if $a_n \neq 0$ then $\sum_{\nu=1}^{n} Re(z_\nu) \leqq -a_{n-1}/a_n$). By Part IV if z_0 is an arbitrary non-real zero of $A(z)$, then the conclusions A and B there hold. Finally, one can apply also Part V.

(c) Suppose $\sigma = (n-2, n-1, n)$. Then $s = 3$, $K = 0$. We set $P(z) \equiv \rho + \tau z$, so that (1) yields

$$a_n = \tau , \quad a_{n-1} = \rho - \tau \sum_{\nu=1}^{n-1} z_\nu , \quad a_{n-2} = -\rho \sum_{\nu=1}^{n-1} z_\nu + \tau \sum_{1 \leqq j < k \leqq n-1} z_j z_k + \alpha .$$

Thus, setting $\sigma_1 = \sum_{\nu=1}^{n-1} z_\nu$, $\sigma_2 = \sum_{1 \leqq j < k \leqq n-1} z_j z_k$, we have

$$A(z) \equiv (\rho + \tau z) g(z) + \alpha \sum_{\nu=1}^{n-1} \lambda_\nu g(z)/(z - z_\nu) ,$$

where

$$\rho = a_n \sigma_1 + a_{n-1} , \quad \tau = a_n , \quad \alpha = a_{n-2} + \sigma_1(a_n \sigma_1 + a_{n-1}) - a_n \sigma_2 .$$

We may apply Parts I–V. For example, suppose that $A(z)$ is a real polynomial, that $\alpha \neq 0$, and that either $a_n = 0$, or $a_n \neq 0$ and

$$(a_{n-2}/a_n) + (a_{n-1}/a_n)\, Re(\sigma_1) + Re(\sigma_1^2 - \sigma_2) \leqq 0 .$$

Then $Re(\tau/\alpha) \leqq 0$, and therefore, by Part III, every non-real zero of $A(z)$ belongs to at least one of the discs (3).

(d) Suppose $\sigma = (n-3, n-2, n-1, n)$. Here $s = 4$, $K = 0$. We set $P(z) \equiv \rho + \sigma_0 z + \tau z^2$, and from (1) we get

$$a_n = \tau , \quad a_{n-1} = \sigma_0 - \tau \sum_{\nu=1}^{n-2} z_\nu , \quad a_{n-2} = \rho - \sigma_0\left(\sum_{\nu=1}^{n-2} z_\nu\right) + \tau \sum_{1 \leqq j < k \leqq n-2} z_j z_k ,$$

$$a_{n-3} = -\rho\left(\sum_{\nu=1}^{n-2} z_\nu\right) + \sigma_0\left(\sum_{1 \leqq j < k \leqq n-2} z_j z_k\right) - \tau\left(\sum_{1 \leqq j < k < m \leqq n-2} z_j z_k z_m\right) + \alpha .$$

Thus, setting $\sigma_1 = \sum_{\nu=1}^{n-2} z_\nu$, $\sigma_2 = \sum_{1 \leqq j < k \leqq n-2} z_j z_k$, $\sigma_3 = \sum_{1 \leqq j < k < m \leqq n-2} z_j z_k z_m$, we have[7]

$$A(z) \equiv (\rho + \sigma_0 z + \tau z^2) g(z) + \alpha \sum_{\nu=1}^{n-2} \lambda_\nu g(z)/(z - z_\nu) ,$$

[7] Observe that if $n = 4$, $\sum_{1 \leqq j < k < m \leqq n-r} z_j z_k z_m$ is zero, being an empty sum.

where

$$\tau = a_n\,, \quad \sigma_0 = a_{n-1} + a_n\sigma_1\,, \quad \rho = a_{n-2} + (a_{n-1} + a_n\sigma_1)\sigma_1 - a_n\sigma_2\,,$$
$$\alpha = a_{n-3} + (a_{n-2} + a_{n-1}\sigma_1 + a_n\sigma_1^2 - 2a_n\sigma_2)\sigma_1 - a_{n-1}\sigma_2 + a_n\sigma_3\,.$$

Here again we can use I–V of the Theorem. For example, suppose S is contained in a disc $C: |z - c| \leqq r$. By I, if a zero ζ of $A(z)$ satisfies $r < \rho_1 \leqq |\zeta - c| \leqq \rho_2$, then

$$|\alpha|/(\rho_2 + r) \leqq |\rho + \sigma_0 z + \tau z^2| \leqq |\alpha|/(\rho_1 - r)\,.$$

By II, if $\alpha\tau \neq 0$, if C contains also the zeros of $P(z) \equiv \rho + \sigma_0 z + \tau z^2$, and if w_1, w_2, w_3 are distinct zeros of $w^2 + \alpha/\tau$, then every zero ($\notin C$) of $A(z)$ lies in $\bigcup_{\nu=1}^{3}(w_\nu + C)$.

7. The following theorem is due to Marden [contained in his Theorem (1, 1), 1949]. *Let $z_1, z_2, \cdots, z_m$ be (distinct) points of the (open) complex plane, let $\mu_1, \mu_2, \cdots, \mu_m$ be positive numbers, and let $A_0, A_1, \cdots, A_{p-1}$ ($p \geqq 1$) be arbitrary complex numbers. Let*

$$F(z) \equiv \sum_{\nu=0}^{p-1} A_\nu z^\nu + \sum_{\nu=1}^{m} \mu_\nu/(z - z_\nu)\,,$$

and set $S = \{z_1, z_2, \cdots, z_m\}$. Let T be the set of those zeros of $F(z)$ at which S subtends an angle $< \pi/(p+1)$. Then the number of points of T (each counted according to its multiplicity) is $\leqq p$. From this follows a result on the zeros of combinations of the form $Q(z) \equiv P(z)f(z) + f'(z)$ where $f(z)$ and $P(z)$ are polynomials. (See loc. cit. Theorem (4.3)).

Using known results on infrapolynomials, we can derive Marden's theorem very easily. For the theorem is obviously true if all the A_ν are zero. Furthermore, one obviously may assume that $A_{p-1} \neq 0$, $m > 1$. Set $g(z) \equiv \prod_{\nu=1}^{m}(z - z_\nu)$, $\mu = \sum_{\nu=1}^{m}\mu_\nu$, $\lambda_\nu = \mu_\nu/\mu$ ($\nu = 1, 2, \cdots, m$). Consider the polynomial

$$A(z) \equiv A_{p-1}^{-1}\,g(z)F(z) \equiv \left(\sum_{\nu=0}^{p-1}(A_\nu/A_{p-1})z^\nu\right)g(z) + \mu A_{p-1}^{-1}\sum_{\nu=1}^{m}\lambda_\nu g(z)/(z - z_\nu)$$
$$\equiv z^{m+p-1} + \cdots$$

which by Theorem 1 of [Shisha and Walsh, 1961] is an $(m + p - 1)$th infrapolynomial on S with respect to $(m - 1, m, \cdots, m + p - 1)$. By a theorem due to Zedek [cf. Zedek 1955, Walsh and Zedek 1956, and Fekete and Walsh 1957] the number of points of T (which is the number of zeros of $A(z)$, multiplicities taken into account, at which S subtends an angle $< \pi/(p+1)$) is $\leqq p$.

BIBLIOGRAPHY

1. L. Fejér. 1922. *Über die Lage der Nullstellen von Polynomen, die aus Minimumforderungen gewisser Art entspringen*, Math. Ann. **85**, pp. 41–48.
2. M. Fekete. 1951. *On the structure of extremal polynomials*, Proc. Nat'l. Acad. Sci. U. S. **37**, pp. 95–103.
3. M. Fekete and J. L. von Neumann. 1922. *Über die Lage der Nullstellen gewisser Minimumpolynome*, Jber. Deutsch. Math. Verein. **31**, pp. 125–138.
4. M. Fekete and J. L. Walsh. 1957. *On restricted infrapolynomials*, J. Analyse Math. **5**, pp. 47–76.
5. M. Marden. 1949. *On the zeros of rational functions having prescribed poles, with applications to the derivative of an entire function of finite genre*, Trans. Amer. Math. Soc. **66**, pp. 407–418.
6. O. Shisha. 1962. *An extension of Jensen's theorem for the derivative of a polynomial and for infrapolynomials*, J. of Res. of the Nat'l. Bur. of Standards, **66B**, pp. 53–55.
7. O. Shisha and J. L. Walsh. 1961. *The zeros of infrapolynomials with some prescribed coefficients*, J. Analyse Math. **9**, pp. 111–160.
8. J. L. Walsh. 1922. *On the location of the roots of certain types of polynomials*, Trans. Amer. Math. Soc. **24**, pp. 163–180.
9. ———, 1950. *The location of critical points of analytic and harmonic functions*, Colloquium Publications, Amer. Math. Soc. **34**.
10. ———, 1958. *On infrapolynomials with prescribed constant term*, J. Math. Pures Appl. **37**, pp. 295–316.
11. J. L. Walsh and O. Shisha. 1963. *The zeros of infrapolynomials with prescribed values at given points*, Proc. Amer. Math. Soc., **14**, pp. 839–844.
12. J. L. Walsh and M. Zedek. 1956. *On generalized Tchebycheff polynomials*, Proc. Nat'l. Acad. Sci. U. S. **42**, pp. 99–104.
13. M. Zedek. 1955. *Fejér's theorem on the zeros of extremal polynomials generalized*. Preliminary report. Bull. Amer. Math. Soc. **61**, p. 49.

HARVARD UNIVERSITY
AEROSPACE RESEARCH LABORATORIES, WRIGHT-PATTERSON AIR FORCE BASE
AND
NATIONAL BUREAU OF STANDARDS

HOMOMORPHISMS OF d-SIMPLE INVERSE SEMIGROUPS WITH IDENTITY

R. J. WARNE

Munn determined all homomorphisms of a regular Rees matrix semigroup S into a Rees matrix semigroup S^* [3, 2]. This generalized an earlier theorem due to Rees [7, 2].

We consider the homomorphism problem for an important class of d-simple semigroups.

Let S be a d-simple inverse semigroup with identity. Such semigroups are characterized by the following conditions [1, 4, 2].

A1: S is d-simple.
A2: S has an identity element.
A3: Any two idempotents of S commute.

It is shown by Clifford [1] that the structure of S is determined by that of its right unit semigroup P and that P has the following properties:

B1: The right cancellation law hold in P.
B2: P has an identity element.
B3: The intersection of two principal left ideals of P is a principal left ideal of P.

Two elements of P are L-equivalent if and only if they generate the same principal left ideal.

Since any homomorphic image of a d-simple inverse semigroup with identity is a d-simple inverse simigroup with identity [5], we may limit our discussion to homomorphisms of S into S^* where S^*, as well as S, is of this type.

In § 1, we consider two such semigroups S and S^* with right unit semigroups P and P^* respectively. We determine the homomorphisms of S into S^* in terms of certain homomorphism of P into P^*, and we show that S is isomorphic to S^* if and only if P is isomorphic to P^*.

In § 2, we show that if P is a semigroup satisfying B1 and B2 on which L is a congruence relation then P is a Schreier extension of its group of units U by P/L and that P/L satisfies B1, B2, and has a trivial group of units. P satisfies B3 if and only if P/L satisfies B3. The converse of this theorem is also given. In this case, we determine the homomorphisms of P into P^* in terms of the homomor-

Received June 25, 1963.

phisms of U into U^* and those of P/L into P^*/L^* and give the corresponding isomorphism theorem. In § 3, some examples are given.

It is a pleasure to acknowledge several helpful conversations with Professor A. H. Clifford.

Section 1. *The correspondence between the homomorphism of S and those of P.*

We first summarize the construction of Clifford referred to in the introduction.

Let S be any semigroup with identity element. We say that the two elements are R-equivalent if they generate the same principal right ideal: $aS = bS$. L-equivalent elements are defined analogously. Two elements a and b are called d-equivalent if there exists an element of S which is L-equivalent to a and R- equivalent to b (This implies the existence of an element of S which is R-equivalent to a and L-equivalent to b.) We shall say that S is d-simple if it consists of a single class of d-equivalent elements.

Now let P be any semigroup satisfying B1, B2 and B3. From each class of L-equivalent elements of P, let us pick a fixed representative. B3 states that if a and b are elements of P, there exists c in P such that $Pa \cap Pb = Pc$. c is determined by a and b to within L-equivalence. We define avb to be the representative of the class to which c belongs. We observe also that

$$avb = bva \; . \tag{1.1}$$

We define a binary operation x by

$$(axb)b = avb \tag{1.2}$$

for each pair of elements a, b of P.

Now let $P^{-1}oP$ denote the set of ordered pairs (a, b) of elements of P with equality defined by

(1.3) $(a, b) = (a', b')$ if $a' = \rho a$ and $b' = \rho b$ where ρ is a unit in P (ρ has a two sided inverse with respect to 1, the identity of P).

We define product in $P^{-1}oP$ by

$$(a, b)(c, d) = ((cxb)a, (bxc)d) \; . \tag{1.4}$$

Clifford's main theorem states: *Starting with a semigroup P satisfying* B1, 2, 3 *equations* (1.2), (1.3), *and* (1.4) *define a semigroup $P^{-1}oP$ satisfying* A1, 2, 3. P *is isomorphic with the right unit subsemigroup* of $P^{-1}oP$ (the right unit subsemigroup of $P^{-1}oP$ is the set of elements of $P^{-1}oP$ having a right inverse with respect to 1. This set is easily shown to be a semigroup). *Conversely, if S is a semigroup satisfying* A1, 2, 3 *its right unit subsemigroup P satisfies* B1, 2, 3 *and S is isomorphic with $P^{-1}oP$.*

The following results are also obtained:

The elements $(1, a)$ of $P^{-1}oP$ constitute a subsemigroup thereof isomorphic to P. We have

(1.5) $$(1, a)(1, b) = (1, ab) \text{ for } a, b \text{ in } P.$$

The ordered pair $(1, 1)$ is the identify of $P^{-1}oP$, i.e.

(1.6) $$(a, b)(1, 1) = (1, 1)(a, b) = (a, b) \text{ for } a, b \text{ in } P.$$

The right inverse of $(1, a)$ is $(a, 1)$, i.e.

(1.7) $$(1, a)(a, 1) = (1, 1) \text{ for a in } P.$$

(1.8) $$(a, c) = (a, 1)(1, c) \text{ for all } a \text{ and } c \text{ in } P.$$

We identity S with $P^{-1}oP$ and P with $\{(1, a): a \text{ in } P\}$.

(1.9) $(avb)c = \rho(acvbc)$ where a, b, and c are in P and ρ is a unit in P.

(1.10) The idempotent elements of $P^{-1}oP$ are just those elements of the form (a, a) where a in P.

(1.11) $$(a, a)(b, b) = (avb, avb) \text{ for all } a, b \text{ in } P.$$

(1.12) $aLb(a, b$ in $P)$ if and only if $a = \rho b$ where ρ is a unit in P.

Let P and P^* be semigroups satisfying B1, and B2 and B3. Let v and u be the 'join' operations on P and P^* respectively defined on page 2. Let N be a homomorphism of P into P^*. N is called a *semilattice homomorphism* (or sl-homomorphism) if

(1.13) $$P^*((avb)N) = P^*(aN) \cap P^*(bN)$$

i. e. $(avb)N\ LaNubN$ in P^*.

It is easily seen that we always have $P^*((avb)N) \subseteq P^*(aN) \cap P^*(bN)$. However, the reverse inclusion is not generally valid. For example, we might have $P = G^+$, $P^* = G^{*+}$, where G and G^* are lattice-ordered groups. An order-preserving homomorphism of G into G^* need not preserve the lattice operations.

THEOREM 1.1. *Let S and S^* be semigroups satifying* A1, A2, *and*

A3, *and let P and P^* be their right unit subsemigroups, Let N, be a* sl-*homomorphism of P into P^*, and let k be an element of P^*.*

For each element (a, b) of S, define

$$(a, b)M = [(aN)k, (bN)k] \tag{1.14}$$

the square brackets indicating an element of S^*. Then M is a homomorphism of S into S^*. Conversely, every homomorphism of S into S^* is obtained in this fashion.

PROOF. To show that M is single valued, let $(a, b) = (a', b')$. Then, $a' = \rho a$ and $b' = \rho b$ where ρ is a unit in P by (1.3). Thus, $a'N = \rho NaN$ and $b'N = \rho NbN$. Thus, since ρN is a unit of P^*, $(a, b)M = (a', b')M$ by (1.3). To show that M is a homomorphism let $\times$ and $\otimes$ be the operations defined on P and P^* respectively by (1.2). Thus, using (1.2), (1.9), (1.13), and (1.12) obtain $((rN)k \otimes (nN)k)(nN)k = (rN)k \quad u(nN)k = w(rNunN)k = w\rho^* \quad ((rvn)N)k = w\rho^*(((r \times n)n)N)k = w\rho^*((r \times n)N)(nN)k$ where w and ρ^* are units in P^*. Thus, from B1,

$$(rN)k \otimes (nN)k = w\rho^*((r \times n)N) . \tag{1.15}$$

Now, from (1.2), (1.1), and (1.15), we have $((nN)k \otimes (rN)k) \ (rN)k = (nN)k \ u(rN)k = (rN)k \ u(nN)k = w\rho^* \ ((rvn)N)k = w\rho^* \ ((nvr)N)k = w\rho^* \ (((n \times r)r)N)k = w\rho^* \ ((n \times r)N) \ (rN) \ k$. Therefore, by B1,

$$(nN)k \otimes (rN)k = w\rho^* \ ((n \times r)N) . \tag{1.16}$$

Thus, by (1.14), (1.4), (1.15), (1.16), and (1.3), $(m, n)M(r, s)M = [(mN)k, (nN)k] \ [(rN)k, (sN)k] = [((rN)k \otimes (nN)k) \ (mN)k, \ ((nN)k \otimes (rN)k) \ (sN)k] = [w\rho^*((r \times n)N) \ (mN)k, \ w\rho^* \ ((n \times r)N) \ (sN)k] = [((r \times n)m)Nk, ((n \times r)s)Nk] = ((r \times n)m, (n \times r)s)M = ((m, n) \ (r, s))M$. Conversely, let M be a homomorphism of S into S^*. Then, by (1.6) and (1.10),

$$(1, 1)M = [k, k] \tag{1.17}$$

for some k in P^*. Now suppose that $(1, n)M = [a, b]$ and $(n, 1)M = [c, d]$ for n in P. It thus follows from (1.7) and (1.6) that $[a, b] [c, d] [a, b] = [a, b]$ and $[c, d] [a, b] [c, d] = [c, d]$. From (1.8) and (1.7), it easily follows that $[a, b] [b, a] [a, b] = [a, b]$ and $[b, a] [a, b] [b, a] = [b, a]$. Hence, $[b, a]$ and $[c, d]$ are inverses of $[a, b]$ (2, p. 27). Therefore, it follows from a theorem of Munn and Penrose (4; 2, p. 28, Theorem 1.17) that $[b, a] = [c, d]$. Thus

$$\begin{aligned}(1, n)M &= [a, b]\\ (n, 1)M &= [b, a]\end{aligned} \tag{1.18}$$

Now, from (1.7), (1.17), and (1.18), $[a, b][b, a] = [k, k]$. Thus, from (1.8) and (1.7), we have $[a, a] = [k, k]$. Hence, by (1.3), $a = \rho k$ where ρ is a unit of P^*. Therefore, by (1.18) and (1.3),

$$(1.19) \qquad \begin{aligned}(1, n)M &= [\rho k, b] = [k, \rho^{-1}b] = [k, c]\\ (n, 1)M &= [b, \rho k] = [\rho^{-1}b, k] = [c, k]\end{aligned}$$

where $c = \rho^{-1}b$. Now, again using (1.8) and (1.7), $[c, k][k, c] = [c, c]$. Thus, by (1.11), $[k, k]\ [c, c] = [kuc, kuc] = [c, c]$. Therefore, by (1.3) (1.12), $P^*(kuc) = P^*c$. Hence, by the definition of u, $P^*k \cap P^*c = P^*c$ and $P^*c \subseteq P^*k$. Thus, we may write $c = B_n k$ where B_n in P^*. Thus, from (1.19), we have

$$(1.20) \qquad \begin{aligned}(1, n)M &= [k, B_n k]\\ (n, 1)M &= [B_n k, k]\ .\end{aligned}$$

It follows easily from (1.8), (1.20) and (1.7) that

$$(1.21) \qquad (m, n)M = [B_m k, B_n k]\ .$$

Thus, to complete the proof, we must show that $n \rightarrow B_n$ is a homomorphism of P into P^* and that $P^*\ (B_m\ u\ B_n) \subseteq P^*B_{mvn}$. It follows from (1.20), (1.3), and (B1) that $n \rightarrow B_n$ is single valued. To show that $n \rightarrow B_n$ is a homomorphism we first note that from (1.5) and (1.20), $[k, B_m k][k, B_n k] = [k, B_{mn}k]$. Thus, by (1.4)

$$(1.22) \qquad [(k \otimes B_m k)k, (B_m k \otimes k)B_n k] = [k, B_{mn}k]\ .$$

From (1.2), the definition of u, and (1.12)

$$(1.23) \qquad (k \otimes B_m k)\ B_m k = ku\ (B_m k) = wB_m k$$

where w is a unit of P^*. Thus, by (B1)

$$(1.24) \qquad k \otimes (B_m k) = w\ .$$

By virtue of (1.2), (1.1), and (1.23), $((B_m k \otimes k)k = (B_m k)\ uk = ku\ (B_m k) = wB_m k$. Hence, by (B1),

$$(1.25) \qquad (B_m k) \otimes k = wB_m\ .$$

If we substitute (1.24) and (1.25) in (1.22), we obtain $[wk, wB_m B_n k] = [k, B_{mn}k]$. Hence, from (1.3) and (B1), we have $B_m B_n = B_{mn}$. We now show that $P^*(B_m uB_n) = P^*B_{mvn}$. From (1.4), $(1, m)\ (n, 1) = (n \times m, m \times n)$. Hence, it follows from (1.21), (B1), and (B2) that $[k, B_m k]$ $[B_n k, k] = [B_{n\times m}k, B_{m\times n}k]$. Thus, by virtue of (1.4), $[((B_n k) \otimes (B_m k))k, ((B_m k) \otimes (B_n k))k] = [B_{n\times m}k, B_{m\times n}k]$. Hence, by (1.3) and (B1), $(B_n k) \otimes (B_m k) = \rho^*{}_1 B_{n\times m}$ where $\rho^*{}_1$ is a unit of P^*. Thus, by (1.2), $B_n kuB_m k = ((B_n k) \otimes (B_m k))\ B_m k = \rho^*{}_1 B_{n\times m} B_m k = \rho^*{}_1 B_{(n\times m)m} k = \rho^*{}_1 B_{nvm} k$. Therefore, by (B1) and (1.9), $\rho'\ (B_n u B_m) = \rho^*{}_1 B_{nvm}$ where ρ' is a unit of P^*. Hence $P^*(B_n u B_m) = P^* B_{nvm}$.

THEOREM 1.2. *Let S, P, S^*, and P^* be as in Theorem 1.1. Let Ω be the set of isomorphisms of P onto P^*. Define $(m, n)M_N = [mN, nN]$ for N in Ω. Then $\{M_N : N \text{ in } \Omega\}$ is the complete set of isomorphisms of S onto S^*. Hence, $N \rightarrow M_N$ is a one-to-one correspondence between the isomorphisms of P onto P^* and those of S onto S^* and S is isomorphic to S^* if and only if P is isomorphic to P^*. The group of automorphisms of P is isomorphic to the group of automorphisms of S.*

PROOF. We first show that $P^*\ (aNubN) \subseteq P^*\ ((avb)N)$ for a, b in P and for any isomorphism N of P onto P^*. It is easy to see that $Pa \subseteq Pb$ if and only if $P^*(aN) \subseteq P^*(bN)$. Since $aNubN = zN$ for some z in P, $P^*zN = P^*(aN) \cap P^*(bN) \subseteq P^*(aN)$, $P^*(bN)$ by the definition of u. Thus, $Pz \subseteq P(avb)$ by the definition of v and the desired result follows. Therefore, by Theorem 1.1, M_N is a homomorphism of S into S^*. To show it is one-to-one let $(m, n)M_N = (p, q)M_N$, i. e. $[mN, nN] = [pN, qN]$. Thus, using (1.3), we may show that $mN = (\rho' p)N$ and $nN = (\rho' q)N$ where ρ' is a unit of P. Thus, by (1.3), $(m, n) = (p, q)$. Clearly, M_N maps S onto S^*. Conversely, let M be an isomorphism of S onto S^*. By Theorem 1.1, $(m, n)M = [(mN)k, (nN)k]$ where k in P^* and N is a homomorphism of P into P^*. Now, it follows from (1.6), (B1), and (B2) that $(1, 1)\ M = [k, k] = [1^*, 1^*]$ where 1^* is the identity of P^*. Thus, by (1.3), k is a unit of P^*. Now, let $nA = k^{-1}\ (nN)k$ for all n in P. It is easily seen that A is a homomorphism of P into P^*. Now, by (B1), (B2), and (1.3), we have

$$(1.26)\qquad \begin{aligned} (m, 1)M &= [(mN)k, k] = [k^{-1}(mN)k, 1^*] = [mA, 1^*] \\ (1, m)M &= [k, (mN)k] = [1^*, k^{-1}(mN)k] = [1^*, mA]\,. \end{aligned}$$

Thus, from (1.26) and (1.3), we have $mA = nA$ implies $m = n$. Let a be in P^*. Then, by the remarks on page 3, it follows that $[1^*, a] = (1, m)M$ for some m in P. Hence, by (1.26) and (1.3), $a = mA$. Therefore A is an isomorphism of P onto P^*. From (1.26) and (1.8), we have $(m, n)M = [mA, nA]$. Thus, $M = M_A$.

Section 2. *A reduction of the homomorphism problem by an application of Schreier extensions.*

We first will briefly review the work of Rédei [6] on the Schreier extension theory for semigroups (we actually give the right-left dual of his construction.). Let G be a semigroup with identity e. We consider a congruence relation n on G and call the corresponding division of G into congruence classes a *compatible class division* of G. The class H containing the identity is said to be the *main class* of the division. H is easily shown to be a subsemigroup of G. The division is called *right normal* it and only if the classes are of the form,

$$Ha_1, Ha_2, \cdots \ (a_1 = e) \tag{2.1}$$

and $h_1\, a_i = h_2\, a_i$ with h_1, h_2 in H implies $h_1 = h_2$. The system (2.1) is shown to be uniquely determined by H. H is then called a *right normal divisor* of G and G/n is denoted by G/H.

Let G, H, and S be semigroups with identity. Then, if there exists a right normal divisor H' of G such that $H \cong H'$ and $S \cong G/H'$, G is said to be a Schreier extension of H by S.

Now, let H and S be semigroups with identities E and e respectively. Consider $H \times S$ under the following multiplication:

$$(A, a)\ (B, b) = (AB^a a^b, ab)\ (A, B \text{ in } H;\ a, b \text{ in } S) \tag{2.2}$$

in which

$$a^b,\ B^a \text{ (in } H)$$

designate functions of the arguments a, b and B, a respectively, and are subject to the conditions

$$a^e = E,\ e^a = E,\ B^e = B,\ E^a = E\ . \tag{2.3}$$

We call $H \times S$ under this multiplication a Schreier product of H and S and denote it by HoS.

Redéi's main theorem states:

THEOREM 2.1 (Rédei). *A Schreier product $G = HoS$ is a semigroup if and only if*

$$(AB)^c = A^c B^c\ (A, B \text{ in } H:\ c \text{ in } S) \tag{2.4}$$

$$(B^a)^c c^a = c^a B^{ca}\ (B \text{ in } H;\ a, c \text{ in } S) \tag{2.5}$$

$$(a^b)^c c^{ab} = c^a (ca)^b\ (a, b,\ c \text{ in } S) \tag{2.6}$$

are valid. These semigroups (up to an isomorphism) are all the Schreier extensions of H by S and indeed the elements (A, e) form a right normal divisor H' of G for which

$$\begin{aligned} G/H' &\cong S\ (H'(E,\ a) \to a) \\ H' &\cong H\ ((A,\ \mathrm{e}) \to A) \end{aligned} \tag{2.7}$$

are valid.

THEOREM 2.2 *Let U be a group with identity E and let S be a semigroup satisfying* B1 *and* B2 (*denote its identity by e*) *and suppose S has a trivial group of units. Then every Schreier extension $P = UoS$ of U by S satisfies* B1 *and* B2 (*the identity is (E, e)*) *and the group of units of P is $U' = \{(A, e): A$ in $U\} \cong U$. Furthermore L is a congruence relation on P and $P/L \cong S$. P satisfies* B3 *if and only if S satisfies* B3.

Conversely, let P be a semigroup satisfying B1 *and* B2 *on which L is a congruence relation. Let U be the group of units of P. Then U is a right normal divisor of P and $P/U \cong P/L$. Thus, P is a Schreier extension of U by P/L. P/L satisfies* B1 *and* B2 *and has a trivial group of units.*

REMARK. Hence if P is any semigroup satisfying B1 and B2 with group of units U such that L is a congruence relation on P, we will write $P = (U, P/L, a^b, A^b)$ in conjunction with Theorem 2.1 and 2.2. (We note that L is a right regular equivalence relation on any semigroup) a^b, A^b will be called the function pair belonging to P.

REMARK. A theorem of Rees [8, Theorem 3.3] is a special case of the above theorem.

Proof. It follows easily from (2.2) and (2.3) that P satisfies B1 and has identity (E, e). From Theorem 2.1, $U' \cong U$. Now, suppose (A, a) is a unit of P. Then, $(A, a)\ (B, b) = (E, e)$ for some (B, b) in P. Hence by (2.2), $ab = e$. Thus, by (B1), (B2), and the fact that the group of units of S is e, $a = b = e$, and (A, a) in U'. From (2.2) and (2.3), every element of U' is a unit of P.

Next, we determine the principal left ideals of P. From (2.2), we have

(2.8) $$P(A, a) = \{(BA^b b^a, ba): B \text{ in } U,\ b \text{ in } S\}$$

$= \{(C, ba): C$ in U, b in $S\}$.
Since $P(A, a)$ just depends on a, we may write $P(A, a) = P_a$ for all A in U.

Next, we show that

(2.9) $$(A, a)\ L\ (B, b) \text{ if and only if } a = b\ .$$

Now, from (2.8), $(A, a)\ L\ (B, b)$ implies $b = xa$ and $a = yb$ for some x, y in S Thus, by B1, $xy = yx = e$, and since S has a trivial group of units, $x = y = e$. Thus, $a = b$. The converse is evident from (2.8). It follows easily from (2.9) and (2.2) that L is a congruence relation. $L_{(E,a)}$ will denote the L-class of P containing (E, a). It is easily seen that the mapping $L_{(E,a)} \to a$ is an isomorphism of P/L onto S. Now suppose S satisfies B3, i.e. a, b in S implies there exists c in S such that

$$Sa \cap Sb = Sc . \tag{2.10}$$

From (2.10) and (2.8),

$$P_a \cap P_b = P_c \tag{2.11}$$

and P satisfies B3. If P satisfies B3, it follows from (2.8) and (2.11) that S satisfies B3.

Now, let P be a semigroup satisfying B1 and B2 with group of units U on which L is a congruence relation. By (1.12) (this is shown without using B3) U is the congruence class mod L containing the identity 1 of P, i.e. U is the main class of the compatible class division of P given by L. If a in P, $L_a = Ua$ from (1.12). If $\rho_1 a = \rho_2 a$ a where ρ_1, ρ_2 in U, then $\rho_1 = \rho_2$ by B1. Thus, U is a right normal divisor of P and $P/U \cong P/L$. Hence, P is a Schreier extension of U by P/L. By virtue of (1.12) and (B1), P/L satisfies B1.

Let $a \to \bar{a}$ be the natural homomorphism of P onto P/L. Then, $\bar{1}$ is the identity of P/L. Let $\bar{a}$ be a unit of $\bar{P}$. Then, by (1.12), (B1), and (B2), a is in U. Hence, $\bar{a} = \bar{1}$. Therefore, P/L has a trivial group of units.

THEOREM 2.3. *Let $P = (U, P/L, a^b, A^b)$ and $P^* = (U^*, P^*/L^*, b^c, B^c)$ be semigroups satisfying* B1 *and* B2 *on which L and L^* are congruence relations. U and a^b, A^b denote the unit group and function pair of P. U^* and b^c, B^c denote the unit group and function pair of P^*. P/L is the factor semigroup of P mod L and P^*/L^* is the factor semigroup of P^* mod L^*. Let f be a homomorphism of U into U^*, g be a homomorphism of P/L into P^*/L^*, and h be a function of P/L into U^*. Suppose f, g and h are subject to the following conditions:*

$$(ah)\ (bh)^{(ag)}(ag)^{(bg)} = (a^b f)(ab)h \tag{2.12}$$

$$(bh)(Af)^{(bg)} = (A^b f)(bh) . \tag{2.13}$$

For each (A, a) in P define

$$(A,\ a)M = [(Af)(ah),\ ag] \tag{2.14}$$

where the square brackets denote elements of P^*. Then M is a homomorphism of P into P^* Conversely, every homomorphism of P into P^* is obtained in this fashion. M is an isomorphism if and *only if f and g are isomorphisms.*

Proof. Clearly, M is single valued. From (2.14), (2.2), (2.4), (2.13) and (2.12), we have

$$(A, a)M\ (B, b)M = [Af)(ah),\ ag]\ [(Bf)(bh),\ bg] =$$
$$= [(Af)(ah)((Bf)(bh))^{(ag)}(ag)^{(bg)}, ag \,.\, bg] = [(Af)(ah)(Bf)^{ag}(bh)^{ag}(ag)^{bg}, (ab)_g]$$
$$= [(Af)(B^a f)(ah)(bh)^{ag}(ag)^{bg}, (ab)_g] = [(Af)(B^a f)(a^b f)(ab)h, (ab)_g]$$
$$[(AB^a a^b)f\,(ab)h, (ab)_g] = (AB^a a^b, ab)M = ((A, a)(B, b))M \,.$$

Thus, M is a homomorphism of P into P^*. Conversely, let M be any homomorphism of P into P^*. It follows from B1 and B2 that $UM \subseteq U$.* Thus, by Theorem 2.2, we may let

$$(2.15) \qquad (A, e)M = [Af, e^*]$$

where e and e^* denote the identities of P/L and P^*/L^* respectively. Clearly, f is a mapping of U into U^*. It follows easily from (2.15), (2.2) and (2.3) that f is a homomorphism of U into U^*. Let E be the identity of U. Then,

$$(2.16) \qquad (E, a)M = [ah, ag] \,.$$

Clearly, h is a function of P/L into U^* and g is a function of P/L into P^*/L^*. From (2.2) and (2.3), $(A, a) = (A, e)(E, a)$. Thus, by (2.15), (2.16), (2.2), and (2.3)

$$(2.17) \quad (A, a)M = (A, e)M\,(E, a)M = [Af, e^*][ah, ag] = [(Af)(ah), ag] \,.$$

From (2.2) and (2.3), we have $(E, a)(E, b) = (a^b, ab)$. Thus, by (2.17), we have $[ah, ag]\,[bh, bg] = [(a^b f)(ab)h, (ab)g]$. Therefore, by (2.2)

$$(2.18) \qquad [(ah)(bh)^{ag}(ag)^{bg}, (ag)(bg)] = [(a^b f)(ab)h, (ab)g] \,.$$

From (2.18), it follows that g is a homomorphism and (2.12) is satisfied. From (2.2) and (2.3), we have $(E, b)(A, e) = (A^b, b)$. Thus, from (2.17) and (2.15), $[bh, bg]\,[Af, e^*] = [(A^b f)(bh), bg]$. Hence, (2.13) follows from (2.2) and (2.3).

Suppose M is an isomorphism of P onto P^*. Therefore, by (2.14) $(A, a)M = [(Af)(ah), ag]$ where f is a homomorphism of U into U^*, h is a single valued mapping of P/L into U^* and g is a homomorphism P/L into P^*/L^*. It is easy to see that $UM = U^*$. Thus, by virtue of theorem 2.2, if B in U^*, there exists A in U such that $(A, e)M = [B, e^*]$. Thus, by (2.15), $Af = B$ and f maps U onto U^*. By (2.15), f is one-to-one and hence is an isomorphism of U onto U^*. To show g is one-to-one, let

$$ag = bg \ . \tag{2.19}$$

There exists x in U^* such that

$$x(bh) = ah \ . \tag{2.20}$$

Now, by (2.2) and (2.3), $(xf^{-1}, e)(E, b) = (xf^{-1}, b)$. Hence, by (2.15), (2.14), (2.2), (2.3), (2.19) and (2.20), $(xf^{-1}, b)M = [x, e^*][bh, bg] = [x(bh), bg] = [ah, ag] = (E, a)M$. Hence, $a = b$. It follows immediately from (2.14) that g maps P/L onto P^*/L^* and hence g is an isomorphism of P/L onto P^*/L^*.

Conversely, suppose there exists an isomorphism f of U onto U^*, an isomorphism g of P/L onto P^*/L^* and a single valued mapping h of P/L into U^* such that (2.12) and (2.13) are satisfied. Therefore, by (2.14), $(A, a)M = [(Af)(ah), ag]$ is a homomorphism of P into P^*. It is easily seen that M is one-to-one. Let $[B, b]$ be in P^*. Now there exists a in P/L such that $b = ag$ and A in U such that $(Af)(ah) = B$. Hence $(A, a)M = [B, b]$ by (2.14).

REMARK. If $ah = E^*$, where E^* is the identity of U^*, then (2.12) and (2.13) simplify greatly:

$$(ag)^{bg} = a^b f \ , \tag{2.12$'$}$$

$$(Af)^{bg} = A^b f \ . \tag{2.13$'$}$$

Professor Clifford remarks that we can bring this about by making a new choice of representative elements in P or in P^*, respectively, in the following two cases: if the range of h is contained in the range of f; or if $ag = a'g$ (a, a' in P/L) implies $ah = a'h$.

Section 3. Examples. We give some examples to illustrate the theory.

EXAMPLE 1. The bicyclic semigroup "C" [2, p. 43] consists of all pairs of nonnegative integers with multiplication given by

$$(i, j)(k, s) = (i + k - \min\ (j, k)\, j + s - \min\ (j, k)) \ , \tag{3.1}$$

A complete set of endomorphisms of "C" is given by

$$(i, j)M_{(t,\,k)} = (ti + k, tj + k)(i, j \text{ are nonnegative integers}) \tag{3.2}$$

where (t, k) runs through all ordered pairs of nonnegative integers.

The only automorphism of 'C' is the identity.

EXAMPLE 2. Let G be any group of order greater than or equal to two with identity E. Let I_0 be the nonnegative integers under the usual addition. Consider $P = GxI_0$ under the following multiplication.

$$(A, a)(B, b) = (AB^a, a + b) \tag{3.3}$$

where $B^a = B$ if $a = 0$
$B^a = E$ if $a \neq 0$.

P is a semigroup satisfying (B1), (B2), (B3) which is not left cancellative. Let S be the semigroup corresponding to P in Clifford's main theorem. Let h be a mapping of I_0 into G such that $oh = E$ and $ah = (a + b)h$ for all $a \neq 0$. Let f be an automorphism of G. Then,

(3.4) $((A, a), (B, b))M = (((Af)(ah), a), ((Bf)(bh), b))$ where (A, a) ,

(B, b) in P is an automorphism of S. Conversely every automorphism of S is obtained in this fashion.

One obtains similar results if I_0 is replaced by the positive part of any lattice ordered group.

EXAMPLE 3. Let G^+ be the positive part of any lattice ordered group G. Let S be the semigroup corresponding to G^+ in Clifford's main theorem. Then there exists a one-to-one correspondence between the automorphisms M of S and the order preserving automorphisms N of G. This correspondence is given by

$$(m, n)M = (mN, nN) \ (m \text{ and } n \text{ in } G^+) .$$

REFERENCES

1. A. H. Clifford, *A class of d-simple semigroups*, Amer. J. Math., **75** (1953), 547-556.
2. A. H. Clifford and G. B. Preston, *The algebraic theory of semigroups*, Math. Surveys No. 7, Amer. Math. Soc., Providence, R. I., 1962.
3. W. D. Munn, *Semigroups and their algebras*, Dissertation, Cambridge University, 1955.
4. W. D. Munn and R. Penrose, *A note on inverse semigroups*, Proc. Cambridge Philos. Soc., **51** (1955), 369-399.
5. G. B. Preston, *Inverse semi-groups*, J. London, Math. Soc., **29** (1954), 396-403.
6. L. Redéi, *Die Verallgemeinerung der Schreierschen Erweiterungstheorie*, Acta Sci. Math., Szeged **14** (1952), 252-273.
7. D. Rees, *On semi-groups*, Proc. Cambridge Phil. Soc., **36** (1940), 387-400.
8. ———, *On the ideal structure of a semi-group satisfying a cancellation law*, Quarterly J. Math. Oxford Ser., **19** (1948), 101-108.

WEST VIRGINIA UNIVERSITY
MORGANTOWN, WEST VIRGINIA

LINEAR TRANSFORMATIONS ON GRASSMANN SPACES

R. WESTWICK

1. Let U denote an n-dimensional vector space over an algebraically closed field F, and let G_{nr} denote the set of nonzero pure r-vectors of the Grassmann product space $\bigwedge^r U$. Let T be a linear transformation of $\bigwedge^r U$ which sends G_{nr} into G_{nr}. In this note we prove that T is nonsingular, and then, by using the results of Wei-Liang Chow in [1], we determine the structure of T.

For each $z = x_1 \wedge \cdots \wedge x_r \in G_{nr}$, we let $[z]$ denote the r-dimensional subspace of U spanned by the vectors $x_1, \cdots, x_r$. By Lemma 5 of [1], two independent elements z_1 and z_2 of G_{nr} span a subspace all of whose nonzero elements are in G_{nr} if and only if $\dim([z_1] \cap [z_2]) = r - 1$; that is, if and only if $[z_1]$ and $[z_2]$ are adjacent. If $V \subseteqq \bigwedge^r U$ is a subspace such that each nonzero vector in V is in G_{nr} and if V is maximal (that is, not contained in a larger such subspace) then $\{[z] \mid z \in V, z \neq 0\}$ is a maximal set of pairwise adjacent r-dimensional subspaces of U. These sets of subspaces are of two types; namely, the set of all r-dimensional subspaces of U containing a common $(r-1)$-dimensional subspace, and the set of all r-dimensional subspaces of an $(r+1)$-dimensional subspace of U. We adopt the usual convention of calling these sets of subspaces maximal sets of the first and second kind respectively. We will let A_r denote the set of those maximal V which determine a set of pairwise adjacint subspaces of the first kind, and we will let B_r denote the set of those maximal V which determine a set of pairwise adjacent subspaces of the second kind.

2. In this section we prove that if T sends each member of B_r into a member of B_r then T is nonsingular.

Let $U_1, \cdots, U_t$ be k-dimensional pairwise adjacent subspaces of U and let $z_i \in G_{nk}$ be such that $[z_i] = U_i$ for $i = 1, \cdots, t$. Then $\{U_1, \cdots, U_t\}$ is said to be independent if and only if $\{z_1, \cdots, z_t\}$ is an independent subset of $\bigwedge^k U$. We note the following facts concerning an independent set $\{U_1, \cdots, U_t\}$. If it is of the first kind (in the sense of the previous section) then there is an independent set of vectors $\{x_1, \cdots, x_{k-1}, y_1, \cdots, y_t\}$ of U such that for $i = 1, \cdots, t$, $U_i = \langle x_1, \cdots, x_{k-1}, y_i\rangle \cdot \langle\cdots\rangle$ denotes the linear subspace spanned by the vectors enclosed. If it is of the second kind, then there is an independent set of vectors $\{x_1, \cdots, x_{k+1}\}$ such that $U_i = \langle x_1, \cdots, x_{i-1}, x_{i+1}, \cdots, x_{k+1}\rangle$, for $i = 1, \cdots, t$. It is easily

Received July 2, 1963. The author is indebted to M. Marcus for his encouragement and help.

deduced from this that $\dim(\bigwedge^r U_1 + \cdots + \bigwedge^r U_t)$ is equal to $t\binom{k-1}{r-1} + \binom{k-1}{r}$ or $\sum_{i=0}^{t-1}\binom{k-i}{r-1}$ according as the set of subspaces $\{U_i\}$ is of the first or second kind. We adopt the usual convention that $\binom{m}{n} = 0$ if $m < n$. Finally, if the set $\{U_1, \cdots, U_t\}$ is not independent, then for some i, $\bigwedge^r U_i \subseteqq \bigwedge^r U_1 + \cdots + \bigwedge^r U_{i-1}$. In fact, the choice of i such that $\{z_1, \cdots, z_{i-1}\}$ is independent and $z_i \in \langle z_1, \cdots, z_{i-1}\rangle$ will do.

We require the

LEMMA 1. *Let $\{U_1, \cdots, U_{s+1}\}$ be a set of pairwise adjacent k-dimensional subspaces of U. Suppose further that the set is independent and is of the second kind. Let $V \subseteqq \bigwedge^r U_1 \cdots + \bigwedge^r U_{s+1}$ be a subspace with dimension $\binom{k-s}{r-s}$, where $s \leqq r \leqq k$. Then there is a set $\{V_1, \cdots, V_s\}$ of pairwise adjacent k-dimensional subspaces of U such that $V \cap (\bigwedge^r V_1 + \cdots + \bigwedge^r V_s) \neq \{0\}$.*

Proof. Let $m = \binom{k-s}{r-s}$ and let $\{z_1, \cdots, z_m\}$ be a basis of V. Choose an independent set of vectors $\{x_1, \cdots, x_{k+1}\}$ of U such that for $i = 1, \cdots, s+1$, $U_i = \langle x_1, \cdots, x_{i-1}, x_{i+1}, \cdots, x_{k+1}\rangle$. We can write

$$z_i = z_1^i + x_1 \wedge \cdots \wedge x_{s-1} \wedge x_s \wedge z_2^i + x_1 \wedge \cdots \wedge x_{s-1} \wedge x_{s+1} \wedge z_3^i$$

where

$$z_1^i \in \bigwedge^r U_1 + \cdots + \bigwedge^r U_{s-1} \quad \text{and} \quad z_2^i, z_3^i \in \bigwedge^{r-s} \langle x_{s+2}, \cdots, x_{k+1}\rangle$$

for $i = 1, \cdots, m$. In the case that $s = 1$, we take $z_1^i \in \bigwedge^r \langle x_3, \cdots, x_{k+1}\rangle$. In the case that $s = r$, we take $z_2^i, z_3^i \in F$. If $\{z_2^1, \cdots, z_2^m\}$ or $\{z_3^1, \cdots, z_3^m\}$ is dependent, then we can form a linear combination of $z_1, \cdots, z_m$ which will be in $\bigwedge^r U_1 + \cdots \bigvee^r{}_{s-1} + \bigwedge^r U_{s+1}$ or $\bigwedge^r U_1 + \cdots + \bigwedge^r U_{s-1} + \bigwedge^r U_s$ respectively. If, on the other hand, both sets are independent then each is a basis of $\bigwedge^{r-s}\langle x_{s+2}, \cdots, x_{k+1}\rangle$ since $\dim(\bigwedge^{r-s}\langle x_{s+2}, \cdots, x_{k+1}\rangle) = \binom{k-s}{r-s} = m$. Let $z_2^i = \sum_{j=1}^m a_{ij} z_3^j$, $i = 1, \cdots, m$. Choose $\lambda \neq 0$ and $b_i \in F$, not all equal to zero, such that

$$\lambda b_j = \sum_{i=1}^m b_i a_{ij}\,, \qquad j = 1, \cdots, m\,.$$

Then

$$0 \neq \sum_{j=1}^m b_j z_j = \sum_{j=1}^m z_1^j + \sum_{j=1}^m x_1 \wedge \cdots \wedge x_{s-1} \wedge (x_s + \lambda^{-1} x_{s+1}) \wedge b_j z_2^j$$
$$\in \bigwedge^r U_1 + \cdots + \bigwedge^r U_{s-1} + \bigwedge^r V_1$$

where $V_1 = \langle x_1 \cdots, x_{s-1}, x_s + \lambda^{-1}x_{s+1}, x_{s+2}, \cdots, x_{k+1}\rangle$. The subspaces

$U_1, \cdots, U_{s-1}, V_1$ are pairwise adjacent and so the Lemma is proved.

The nonsingularity of T is now proved as follows. Let W be a subspace of U. We prove, by induction on the dimension of W, that T is one-to-one on $\bigwedge^r W$ and that the image of $\bigwedge^r W$ under T is $\bigwedge^r W'$ for some subspace W' of U with $\dim(W) = \dim(W')$. When $\dim(W) = r + 1$ this is clear since we are assuming that B_r is sent into B_r by T. Suppose that the statement has been proved for k-dimensional subspaces, and consider a $(k+1)$-dimensional subspace W of U. Let s be the largest integer such that for any set $\{W_1, \cdots, W_s\}$ of pairwise adjacent k-dimensional subspaces of W, T is one-to-one on $\bigwedge^r W_1 + \cdots + \bigwedge^r W_s$. If $s \geqq r + 1$ then T is one-to-one on $\bigwedge^r W$, since in this case, for an independent set $\{W_1, \cdots, W_s\}$ we must have $\bigwedge^r W = \bigwedge^r W_1 + \cdots + \bigwedge^r W_s$. Suppose then that $1 \leqq s \leqq r$ and let $\{U_1, \cdots, U_{s+1}\}$ be any set of $s+1$ pairwise adjacent k-dimensional subspaces of W. If the set is dependent then T is one-to-one $\bigwedge^r U_1 + \cdots + \bigwedge^r U_{s+1}$ since we may drop one of the terms. Therefore we assume that the set is independent. Choose k-dimensional subspaecs $U_1', \cdots, U_{s+1}'$ such that $T(\bigwedge^r U_i) = \bigwedge^r U_i'$ for $i = 1, \cdots, s+1$. For each $j \leqq s$, T maps $\bigwedge^r U_1 + \cdots + \bigwedge^r U_j$ onto $\bigwedge^r U_1' + \cdots + \bigwedge^r U_j'$. Therefore, since T is one-to-one on $\bigwedge^r U_1 + \cdots + \bigwedge^r U_s$, the set $\{U_1', \cdots, U_s'\}$ is independent. Furthermore, the set $\{U_1', \cdots, U_{s+1}'\}$ is also independent. If not, then the image under T of both $\bigwedge^r U_1 + \cdots + \bigwedge^r U_s$ and $\bigwedge^r U_1 + \cdots \bigwedge^r U_{s+1}$ is $\bigwedge^r U_1' + \cdots + \bigwedge^r U_s'$. But then the dimension of the null space of T in $\bigwedge^r U_1 + \cdots + \bigwedge^r U_{s+1}$ is at least as large as the difference in the dimensions of $\bigwedge^r U_1 + \cdots + \bigwedge^r U_{s+1}$ and $\bigwedge^r U_1 + \cdots + \bigwedge^r U_s$, that is, $\binom{k-s}{r-s}$. We apply Lemma 1 to contradict the choice of s. It follows that T is one-to-one on all of $\bigwedge^r W$. Finally, let $\{W_1, \cdots, W_{k+1}\}$ be an independent set of k-dimensional pairwise adjacent subspaces of W (necessarily of the second kind). Let W_i' be chosen so that $T(\bigwedge^r W_i) = \bigwedge^r W_i'$. It follows easily that $\{W_1', \cdots, W_{k+1}'\}$ is of the second kind also, so that the image of $\bigwedge^r W$ is $\bigwedge^r W'$ where W' is the $(k+1)$-dimensional subspace of U containing $W_1', \cdots, W_{k+1}'$. By taking $W = U$ we see that T is one-to-one on $\bigwedge^r U$.

3. It is necessary to investigate whether a general T does necessarily send each element of B_r into B_r. For the cases $n > 2r$, $n < 2r$, this is proved directly, using Lemma 2. The case $n = 2r$ requires a more delicate argument, given at the end of this section; there it is shown that if some element of B_r is sent into B_r by T, then T sends B_r into B_r.

LEMMA 2. *Let $r < n$ and let V_1 and V_2 be in A_r such that $V_1 \cap V_2 \neq \{0\}$. Then, if $V \subseteqq V_1 + V_2$ and $\dim(V) = n - r$, we have $V \cap G_{nr} \neq \phi$.*

Proof. Let U_i be the $(r-1)$-dimensional subspace of U determined by V_i for $i = 1, 2$. Since $V_1 \cap V_2 \neq \{0\}$, either $U_1 = U_2$ or $\dim(U_1 \cap U_2) = r - 2$.

If $U_1 = U_2$ then $V_1 = V_2$, so that in this case it is clear that $V \cap G_{nr} \neq \phi$.

Suppose that $\dim(U_1 \cap U_2) = r - 2$ and let $\{x_1, \cdots, x_{r-2}\}$ be a basis of this intersection. Choose y_i such that $U_i = \langle x_1, \cdots, x_{r-2}, y_i \rangle$ for $i = 1, 2$. Choose u_i and v_i in U, $i = 1, \cdots, n - r$, such that

$$\{z_i = x_1 \wedge \cdots \wedge x_{r-2} \wedge (y_1 \wedge u_i + y_2 \wedge v_i) \mid i = 1, \cdots, n - r\}$$

forms a basis of V. If

$$\{x_1, \cdots, x_{r-2}, y_1, y_2, v_1, \cdots, v_{n-r}\} \quad \text{or} \quad \{x_1, \cdots, x_{r-2}, y_1, y_2, u_1, \cdots, u_{n-r}\}$$

is dependent, then there is a linear combination of the z_i which is in V_1 or V_2 respectively. If, on the other hand, both sets are independent, then they are both bases for U and we may write

$$u_i = w_i + c_i y_2 + \sum_{j=1}^{n-r} a_{ij} v_j \,, \qquad i = 1, \cdots, n - r \,,$$

where $w_i \in \langle x_1, \cdots, x_{r-2}, y_1 \rangle$ and $c_i, a_{ij} \in F$. We note that $\det(a_{ij}) \neq 0$ so we can choose $\lambda \neq 0$ and b_j for $j = 1, \cdots, n - r$, not all zero, such that $\lambda b_j = \sum_{i=1}^{n-r} b_i a_{ij}$. Then

$$0 \neq \sum_{j=1}^{n-r} b_j z_j = x_1 \wedge \cdots \wedge x_{r-2} \wedge (y_1 + \lambda^{-1} y_2) \wedge \left[\left(\sum_{j=1}^{n-r} b_j c_j\right) y_2 + \lambda \sum_{j=1}^{n-r} b_j v_j\right]$$

is an element of $V \cap G_{nr}$. This proves the Lemma.

For $n \neq 2r$ the image under T of an element of B_r is an element of B_r. For $n < 2r$ this is clearly so since the subspaces of $\bigwedge^r U$ in B_r have dimension $r + 1$, which is greater than the dimension $(n - r + 1)$ of the subspaces in A_r.

For $n > 2r$ we proceed as follows. The image of an A_r is an A_r. Suppose that the image of a $W \in B_r$ is a subspace of a $V \in A_r$. Choose two elements V_1 and V_2 of A_r such that $V_1 \cap V_2 \neq \{0\}$ and $\dim(V_1 \cap W) = \dim(V_2 \cap W) = 2$. One does this by choosing V_1 and V_2 so that the $(r-1)$-dimensional subspaces of U determined by them are adjacent subspaces of the $(r+1)$-dimensional subspace determined by W. Now, $T(V_1) = T(V_2) = V$ since each is in A_r and each intersects V in at least two dimensions. Therefore $T(V_1 + V_2) = V$ and so the null space of T in $V_1 + V_2$ has dimension equal to $(2n - 2r + 1) - (n - r + 1) = n - r$. By Lemma 2, it follows that the null space of T intersects G_{nr} which contradicts the hypothesis that T sends G_{nr} into G_{nr}.

In the case that $n = 2r$ the image of a B_r may be an A_r since the dimensions are equal. However, we prove that if some B_r is sent into a B_r by T, then the image of each B_r is a B_r. Suppose not. Then we can choose $(r+1)$-dimensional subspaces W_1 and W_2 of U such that $T(\bigwedge^r W_1) \in A_r$ and $T(\bigwedge^r W_2) \in B_r$. Furthermore, we can choose W_1 and W_2 adjacent, so that $\dim(W_1 \cap W_2) = r$. Choose three distinct elements V_1, V_2, and V_3 of A_r such that the $(r-1)$-dimensional subspaces of U determined by these elements are contained in $W_1 \cap W_2$. Then $\dim(V_i \cap \bigwedge^r W_j) = 2$ for $i = 1, 2, 3$ and $j = 1, 2$, so that $T(V_i)$ intersects $T(\bigwedge^r W_j)$ in at least two dimensions for each i, j. This implies that each $T(V_i)$ is equal to one of $T(\bigwedge^r W_j)$ and so two of them are equal. The argument of the previous paragraph now leads to a contradiction.

4. By essentially the same argument as used by Chow in [1] to prove his Theorem 1, we can prove that; if S is a nonsingular linear transformation of $\bigwedge^r U$ sending G_{nr} into G_{nr}, and if the image of each B_r is a B_r, then S is a compound. (By a compound we mean a linear transformation of $\bigwedge^r U$ which is induced by a linear transformation of U.)

In the case that $n \neq 2r$ it follows that T is necessarily a compound. For $n = 2r$, T is a compound if some B_r is sent into a B_r. If we let T_0 denote a linear transformation of $\bigwedge^r U$ induced by a correlation of the r-dimensional subspaces of U, then T_0 is nonsingular and sends G_{nr} onto G_{nr}. The image of each A_r under T_0 is a B_r. Therefore, if a B_r is sent by T into an A_r, the T_0T is a compound. We have proved the

THEOREM. *Let U be an n-dimensional vector space over an algebraically closed field and let T be a linear transformation of $\bigwedge^r U$ which sends G_{nr} into G_{nr}. Then T is a compound except, possibly, when $n = 2r$, in which case T may be the composite of a compound and a linear transformation induced by a correlation of the r-dimensional subspaces of U.*

REFERENCE

1. Wei-Liang Chow, *On the Geometry of Algebraic Homogeneous Spaces,* Annals of Math., **50** (1949), 32-67.

Pacific Journal of Mathematics
Vol. 14, No. 3 July, 1964