Pacific Journal of Mathematics

IN THIS ISSUE—

Vol. 10, No. 4 December, 1960

PACIFIC JOURNAL OF MATHEMATICS

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

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

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

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

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

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

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

AN OPTIMUM CUBICALLY CONVERGENT ITERATIVE METHOD OF INVERTING A LINEAR BOUNDED OPERATOR IN HILBERT SPACE

M. ALTMAN

1. In paper [1] we considered a power series method of inverting a linear bounded operator in Hilbert space. This method is actually an iterative method with the same speed of convergence as a geometric progression. A product of two linear operators we shall call briefly a multiplication. Thus, in general, a power series approximative method has the following two properties:

(1) at each iteration we use one multiplication;

(2) the convergence is linear.

In paper [2] we considered an iterative method of inverting an arbitrary linear bounded operator in a Hilbert space. This method requires two multiplications at each iteration step, and the convergence is quadratic. In the present paper we give an iterative method of inverting an arbitrary linear bounded operator in a Hilbert space. This method requires three multiplications at each iteration step and is cubically convergent. Thus, the quadratically convergent method which requires two multiplications at each iteration step may be called the iterative hyperpower method of order two. Analogously, the cubically convergent iterative method which requires three multiplications at each iteration step may be called the iterative hyperpower method of order three. The following two problems arise now in a natural way:

(1) Is it possible to construct an iterative hyperpower method of any degree?

(2) To give a comparison of the hyperpower methods of different degrees, and to answer the question whether there exists an optimum method. As a criterion for a hyperpower method to be better we can assume the following:

The method I is better than the method II if after some iteration steps using the same amount of multiplications for both methods, the method I gives better accuracy. In this paper we construct a certain class of iterative hyperpower methods and for this class the answers to both questions mentioned above is positive. It turns out that the optimum method of this class is the iterative hyperpower method of degree three.

Let A be a linear (i.e. additive and homogeneous) bounded operator with the domain and the range in a Banach space X.

Let us assume that the operator A is non-singular, i.e. A has an

Received October 5, 1959. Based on research supported by O.N.R., U.S.A.

inverse A^{-1} defined on the space X. Let us suppose that the linear bounded operator R_1 is an approximate reciprocal of A. Suppose also that R_1 satisfies the following condition

$$\| I - AR_1 \| = a < 1 , \tag{1}$$

where I is the identity mapping of X

Let p be any positive integer such that $p \geqq 2$. We shall construct an iterative hyperpower method of degree p with the following property

$$I - AR_{n+1} = (I - AR_n)^p , \tag{2}$$

where (R_n) is the sequence of the approximate inverses of A. It is easy to see that this sequence can be defined as follows

$$R_{n+1} = R_n(I + T_n + T_n^2 + \cdots + T_n^{p-1}) , \tag{3}$$

where

$$T_n = I - AR_n , \qquad n = 1, 2, \cdots . \tag{4}$$

Multiplying both sides in (3) by A we get by (4)

$$AR_{n+1} = (I - T_n)(I + T_n + T_n^2 + \cdots + T_n^{p-1}) = I - T_n^p .$$

Hence we obtain the relationship (2).

Thus, we have the following theorem.

THEOREM 1. *The sequence of the approximate inverses R_n defined by formula* (3) *converges in the norm of operators toward the inverse of the non-singular operator A, provided that R_1 satisfies condition* (1). *The error estimate is given by the formula*

$$\| A^{-1} - R_{n+1} \| \leqq \| A^{-1} \| a^{p^n} \tag{5}$$

or

$$\| A^{-1} - R_{n+1} \| \leqq \| R_1 \| \frac{a^{p^n}}{1 - a} \tag{6}$$

Proof. Formula (2) gives by induction

$$AR_{n+1} = I - T_1^{p^n} . \tag{7}$$

Hence we get by (7)

$$R - R_{n+1} = RT_1^{p^n} \tag{8}$$

or

$$R - R_{n+1} = R_1(I - T_1)^{-1}T_1^{p^n} . \tag{9}$$

Formula (5) follows from (8) and formula (6) follows from (9).

For $p = 2$ formula (3) yields

$$R_{n+1} = R_n(2I - AR_n) . \tag{10}$$

This case was considered in [23]. For $p = 3$ we get

$$R_{n+1} = R_n(I + (I - AR_n) + (I - AR_n)^2) \tag{11}$$

or

$$R_{n+1} = R_n(3I - 3AR_n + (AR_n)^2) \tag{12}$$

Thus, we have a class of methods with the property (2).

The question is now if there is an optimum method in this class of methods. To compare two methods we shall use the criterion mentioned above, i.e. the method is better if using the same number of multiplications gives a better accuracy.

Let p and q be two different positive integers. Consider the correspondings methods M_p and M_q defined by the formula (3). At each iteration step the method M_p takes p multiplications and the method M_p takes q multiplications in the sense defined above. Suppose that after a certain number of iteration steps which is different for both methods we get the same number of multiplications which is equal to

$$mp = nq . \tag{13}$$

Then in virtue of (5) the accuracy of the methods M_p and M_q is given by the exponents p^m and q^n respectively. Suppose that

$$p^m > q^n .$$

Then we have by (13)

$$p^m > q^s ,$$

where

$$s = \frac{mp}{q} .$$

Hence, we have

$$p^{1/p} > q^{1/q} . \tag{14}$$

The inequality (14) shows that we obtain the optimum method M_p for p such that the function $p^{1/p}$ $(p = 2, 3, \cdots)$ achieves its maximum. This is the case when $p = 3$ since the maximum of the function

$$y = x^{1/x} , \quad x > 0$$

is attained at $x = e$.

2. We shall now apply Theorem 1 in order to find the approximate inverse of a linear bounded operator in a Hilbert space. Thus, we suppose that X is a Hilbert space H and A is a non-singular linear bounded operator with the domain and the range in H.

Let us begin with the case when A is a self-adjoint and positive definite operator, or, more precisely

$$A^* = A ,$$

where A^* is the adjoint of A, and A satisfies the condition

$$m(x, x) \leqq (Ax, x) \leqq M(x, x) ,$$

where $0 < m < M$, and m, M are the minimum and maximum eigenvalues of A respectively.

Consider the linear operator.

$$T_\alpha = I - \alpha A , \qquad 0 < \alpha < 2/M .$$

In virtue of the critical value theorem[1] we have

$$\frac{M - m}{M + m} \leqq \| T_\alpha \| = a_\alpha < 1 \quad \text{if} \quad 0 < \alpha < 2/M . \tag{13}$$

The minimum of the norm $\| T_\alpha \|$ is equal to

$$a_c = \| T_{\alpha_c} \| = \frac{M - m}{M + m} \tag{14}$$

and is reached precisely at the critical value α_c of A, i.e. for

$$\alpha = \alpha_c = \frac{2}{M + m} .$$

Thus, we get the following theorem.

THEOREM 2. *Let us suppose that A is a self-adjoint positive defined linear operator. If*

$$R_1 = \alpha I \quad \text{for} \quad 0 < \alpha < 2/M , \tag{15}$$

then the sequence of operators R_n determined by the iterative process in (3) *converges in the norm of the operators toward the inverse of A. The error estimate is given by the following formula*

$$\| A^{-1} - R_n \| \leqq \frac{1}{m} a_\alpha^{p^n} \tag{16}$$

[1] See [1], [2]

or

$$\| A^{-1} - R_n \| \leqq \frac{\alpha}{1 - a_\alpha} a_\alpha^{p^n} , \tag{17}$$

where $a_\alpha = \| T_\alpha \|$. *The convergence is best for the critical value of* A, i.e. *for* $\alpha = \alpha_c = 2/M + m$. *In this case* a_α *in formulae* (16) *and* (17) *should be replaced by* a_c *defined in* (14).

Putting $p = 3$ in Theorem 2 we get the theorem for the optimum method. Thus, we have

COROLLARY 1. *The iterative process defined by the formula* (11) *or* (12) *converges cubically toward the inverse of* A *provided that* R_1 *is defined by* (14). *The error estimate is given by formula* (16) *or* (17), *where* $p = 3$. *The convergence is best for the critical value of* A, i.e. *for* $\alpha = \alpha_c = 2/M + m$. *In this case* a_α *in formulae* (16) *and* (17) *should be replaced by* a_c *defined in* (14).

REMARK 1. The convergence of the iterative process is uniform with respect to α for any closed interval contained in the interval $0 < \alpha < 2/M$. Let us observe that α in (15) can be replaced by any number $1/K$, where K is greater than $\| A \|$. However, the convergence is faster when K is smaller. If the operator A is defined by a matrix

$$A = (a_{ij}) \qquad i, j = 1, 2, \cdots, k \tag{18}$$

satisfying the conditions of Theorem 2, then K can be replaced by any of the following numbers

$$\max_i \sum_{j=1}^{k} | a_{ij} | ; \quad \max_j \sum_{i=1}^{k} | a_{ij} | ; \quad \left(\sum_{i,j=1}^{k} | a_{ij} |^2 \right)^{1/2} . \tag{19}$$

However, the convergence is faster when K is smaller.

3. We shall now consider the general case when A is an arbitrary non-singular linear bounded operator in H.

Since the operator AA^* is self-adjoint and positive definite, we have the following inequalities

$$m^2(x, x) \leqq (AA^*x, x) \leqq M^2(x, x) ,$$

where $0 < m^2 < M^2$ and m^2, M^2 are the minimum and the maximum eigenvalues of AA^* respectively.

Let us consider the linear operator

$$T_\alpha = I - \alpha AA^* , \qquad 0 < \alpha < 2/M^2 .$$

Using the same argument as in § 2, we get the following inequalities instead of (13).

$$\frac{M^2 - m^2}{M^2 + m^2} \leqq \| T_\alpha \| = a_\alpha < 1 \quad \text{if} \quad 0 < \alpha < 2/M^2 . \tag{20}$$

The minimum of the norm $\| T_\alpha \|$ is reached at

$$\alpha = \alpha_c = \frac{2}{M^2 + m^2}$$

and is equal to

$$a_c = \| T_{\alpha_c} \| = \frac{M^2 - m^2}{M^2 + m^2} . \tag{21}$$

Thus we obtain the following theorem.

THEOREM 3. *If*

$$R_1 = \alpha A^* \quad \text{for} \quad 0 < \alpha < 2/M^2 , \tag{22}$$

then the sequence of operators R_n determined by the iterative process in (3) *converges in the norm of the operators toward the inverse of A. The error estimate is given by the formulae* (16) *or* (17), *where a_α should be replaced by the expression in* (18). *The convergence is best for*

$$\alpha = \alpha_c = \frac{2}{M^2 + m^2} .$$

For the error estimate in this case a_α in formulae (16) *and* (17) *should be replaced by a_c defined in* (21).

Putting $p = 3$ in Theorem 3 we get the theorem for the optimum method in general case. Thus we have

COROLLARY 2. *If R_1 is determined by* (22) *then the iterative process defined in* (11) *or* (12) *converges cubically toward the inverse of A. For the error estimate we have the formulae* (16) *or* (17), *where $p = 3$. The convergence is best for the critical value of AA^**, i.e. *for*

$$\alpha = \alpha_c = \frac{2}{M^2 + m^2} .$$

In this case a_α in formulae (16) *and* (17) *should be replaced by a_c determined in* (21).

REMARK 2. The convergence of the iterative process defined by Theorem 3 is uniform with respect to α for any closed interval contained in the interval $0 < \alpha < 2/M^2$. Let us remark that α in (22) can be replaced by any number $1/K$, where K is greater than $\|A\|^2$. However, the convergence is faster when K is smaller.

If the operator A is defined by the non-singular matrix in (18), then for K we can take any of the numbers in (19) with the matrix AA^* replacing the matrix A. We can also take for K any of the squared numbers in (19).

The table below shows the difference in rate of convergence between the following three method: I, II, III, where

I is the power series method considered in [1] (see page 52)

II is the quadratically convergent defined in (10)

III is the cubically convergent optimum method defined in (11) or (12).

Number of Iterations			Number of Multiplication			Accuracy ($a < 1$)		
I	II	III	I	II	III	I	II	III
6	3	2	6	6	6	a^6	a^8	a^9
12	6	4	12	12	12	a^{12}	a^{64}	a^{81}
18	9	6	18	18	18	a^{18}	a^{512}	a^{729}
24	12	8	24	24	24	a^{24}	a^{4096}	a^{6561}

References

1. M. Altman, *Inversion of non-singular linear bounded operators in Hilbert space with application to matrix calculus.*

2. ———, *Inversion of linear bounded operators in Hilbert space with application to matrix calculus.*

California Institute of Technology, Pasadena,
Academy of Sciences, Warsaw.

CRITERION FOR rTH POWER RESIDUACITY

N. C. ANKENY

The Law of Quadratic Reciprocity in the rational integers states: If p, q are two distinct odd primes, then q is a square $(\bmod p)$ if and only if $(-1)^{(p-1)/2}p$ is a square $(\bmod q)$.

One of the classical generalizations of the law of reciprocity is of the following type. Let r be a fixed positive integer, $\phi(r)$ denotes the number of positive integers $\leqq r$ which are relatively prime to r; p, q are two distinct primes and $p \equiv 1 \pmod r$. Then can we find rational integers $a_1(p), a_2(p), \cdots, a_h(p)$ determined by p, such that q is an rth power $(\bmod p)$ if and only if $a_1(p), \cdots, a_h(p)$ satisfy certain conditions $(\bmod q)$.

The Law of Quadratic Reciprocity states that for $r = 2$, we may take $a_1(p) = (-1)^{(p-1)/2}p$.

Jacobi and Gauss solved this problem for $r = 3$ and $r = 4$, respectively. Mrs. E. Lehmer gave another solution recently [2].

In this paper I would like to develop the theory when r is a prime and $q \equiv 1 \pmod r$. I then show that q is an rth power $(\bmod p)$ if and only if a certain linear combination of $a_1(p), \cdots, a_{r-1}(p)$ is an rth power $(\bmod q)$. $a_1(p), \cdots, a_{r-1}(p)$ are determined by solving several simultaneous Diophantine equations. This determination appears mildly formidable and to make the actual numerical computations would certainly be so for a large r. (See Theorem B below.) Also given is a criterion for when r is an rth power $(\bmod p)$ in terms of a linear combination of $a_1(p), \cdots, a_{r-1}(p) \pmod{r^2}$. (See Theorem A below.)

It is possible by the methods developed in this paper to eliminate the conditions that r is a prime and $q \equiv 1 \pmod r$. This would complicate the paper a great deal, and the cases given clearly indicate the underlying theory.

Consider the following Diophantine equations in the rational integers:

$$(1) \qquad r\sum_{j=1}^{r-1} X_j^2 - \left(\sum_{j=1}^{r-1} X_j\right)^2 = (r-1)p^{r-2}$$

$$(2) \qquad \sum\nolimits_1^{(1)} X_{j_1}X_{j_2} = \sum\nolimits_i^{(1)} X_{j_1}X_{j_2} \qquad i = 2, \cdots, \frac{r-1}{2},$$

where $\sum_i^{(k)}$ denotes the sum over all $j_1, \cdots, j_{k+1} = 1, 2, \cdots, r-1$, with the condition $j_1 + \cdots + j_k - kj_{k+1} \equiv i \pmod r$.

Received April 24, 1959; in revised form January, 1960. This research was supported by the United States Air Force through the Air Force Office of Scientific Research of the Air Research and Development Command, under contract No. AF 18 (603)-90. Reproduction in whole or in part is permitted for any purpose of the United States Government.

$$(3) \qquad 1 + \sum_{j=1}^{r-1} X_j \equiv \sum_{j=1}^{r-1} jX_j \equiv 0 \pmod r$$

(4) not all of the $X_j \equiv 0 \pmod p$ and

$$\sum\nolimits_i^{(k)} X_{j_1} \cdots X_{j_{k+1}} - \sum\nolimits_0^{(k)} X_{j_1} \cdots X_{j_{k+1}} \equiv 0 \pmod{p^{r-k-1}}$$

for $k = 2, \cdots, r-2$; $i = 1, 2, \cdots, r-1$.

We shall prove in § II that there exist exactly $r-1$ distinct integral solutions of the equations (1) through (4). In particular let $\{X_j = a_j, j = 1, \cdots, r-1\}$ be a solution. Then we prove that the $a_j(p) = a_j$ satisfy our residuacity criterion, namely

THEOREM A. *r is an rth power* (mod p) *if and only if*

$$\sum_{j=1}^{r-1} ja_j + \frac{1}{2} ra_{r-1} \equiv 0 \pmod{r^2} .$$

THEOREM B. *If $q \equiv 1$* (mod r) *and h is any integer such that h^r is the least power of h which is $\equiv 1$* (mod q), *then q is an rth power* (mod q) *if and only if $\sum_{j=1}^{r-1} a_j h^j$ is an rth power* (mod q).

At the end of § II various special cases are considered.

In particular, for $q = 2, r = 5$, then 2 is a quintic power (mod p) if and only if $a_j \equiv a_{5-j}$ (mod 2), $j = 1, 2$.

For $q = 2, r = 7$, then 2 is a 7th power (mod p) if and only if $a_j \equiv 1$ (mod 2), $i = 1, \cdots, 6$.

Let $r = 3$. Then the solutions to the Diophantine equations (1) to (4) are (a_1, a_2) and (a_2, a_1), where

$$(5) \qquad p = a_1^2 - a_1a_2 + a_2^2,\ a_1 \equiv a_2 \equiv 1 \pmod 3 .$$

Multiplying (5) by 4 and grouping terms gives

$$4p = (a_1 + a_2)^2 + 3(a_1 - a_2)^2 .$$

Let $L = -a_1 - a_2$, $M = (a_1 - a_2)/3$. This gives the representation which Lehmer employs:

$$4p = L^2 + 27M^2,\ L \equiv 1 \pmod 3 .$$

Theorem A states that 3 is a cubic residue (mod p) if and only if $a_1 \equiv a_2$ (mod 9). This, in turn, is equivalent to M being divisible by 3, the condition quoted by Lehmer.

I. **Notation.** r denotes a prime number, ζ_r a primitive rth root of unity, Q the rational numbers, $Q(\zeta_r)$ the cyclotomic field over Q generated by ζ_r. For $j = 1, 2, \cdots, r-1$, σ_j are the automorphisms of $Q(\zeta_r)/Q$

such that $\sigma_j(\zeta_r) = \zeta_r^j$. $\sigma^{-1}(\zeta_r) = \zeta_r^{j'}$, where $jj' \equiv 1 \pmod r$. p denotes a positive rational prime $\equiv 1 \pmod r$, and $\chi_p = \chi$ will be any primitive rth power character (mod p).

$$g(\chi) = \sum_{n=1}^{p-1} \chi(n)\zeta_p^n$$

will be the Gaussian sum associated with χ_p. $\langle\alpha\rangle$ denotes the fractional part of α; i.e., $\langle\alpha\rangle = \alpha - [\alpha]$.

LEMMA 1. (i) $|g(\chi^k)|^2 = p$,
(ii) $g(\chi)^k g(\chi^{-k}) \in Q(\zeta_r)$,
(iii) $g(\chi)^r \in Q(\zeta_r)$, and
(iv) $\sigma_k(g(\chi)^r) = g(\chi^k)^r$
for $k = 1, 2, \cdots, r-1$.

Proof. (i) is the classical result about the absolute value of $g(\chi)$ and can easily be deduced from the definition of $g(\chi)$. (ii), (iii) and (iv) follow from Galois Theory using the relation $\sum_{n=1}^{p-1} \chi(n)\zeta_p^{nt} = \chi(t)^{-1}g(\chi)$ for any integer t prime to p.

LEMMA 2. *There exists a prime ideal $\mathfrak{p}$ in $Q(\zeta_r)$ dividing p such that* $(g(\chi^k)^r) = \sum_{j=1}^{r-1}\sigma_j^{-1}\mathfrak{p}^{r\langle kj/r\rangle}$.

Conversely, given any prime ideal $\mathfrak{p}_1$ in $Q(\zeta_r)$ dividing p, there exists a k such that

$$(g(\chi^k)^r) = \sum_{j=1}^{r-1} \sigma_j^{-1}\, \mathfrak{p}_1^j \,.$$

Proof. Lemma 2 is a result of Stickelberger. For a proof see Davenport and Hasse [1]. See especially the elegant proof on page 181–2. In $Q(\zeta_r)$, the ideal $(r) = (1-\zeta_r)^{r-1}$,

LEMMA 3. $(1-\zeta_r^t)(1-\zeta_r)^{-1} \equiv t \pmod{(1-\zeta_r)}$ and $r(1-\zeta_r^t)^{-r+1} \equiv -1 \pmod{(1-\zeta_r)}$ for $(t, r) = 1$.

Proof. The first fact follows as

$$(1-\zeta_r^t)(1-\zeta_r)^{-1} = \sum_{j=0}^{t-1} \zeta_r^j \equiv \sum_{j=0}^{t-1} 1 \equiv t \pmod{(1-\zeta_r)} \,.$$

The second follows from Wilson's Theorem as

$$\begin{aligned} r(1-\zeta_r^t)^{-r+1} &= \Big(\prod_{j=1}^{r-1} (1-\zeta_r^{jt})\Big)(1-\zeta_r^t)^{-r+1} \\ &= \prod_{j=1}^{r-1} (1-\zeta_r^{jt})(1-\zeta_r^t)^{-1} \equiv (r-1)! \equiv -1 \pmod{(1-\zeta_r)} \,. \end{aligned}$$

THEOREM 1. *For any t not divisible by r,*

$$g(\chi^t)^r + 1 \equiv r(1 - \chi(r)^{-t}) \pmod{(1 - \zeta_r)^{r+1}},$$

and consequently, $\chi(r) = 1$ if and only if

$$g(\chi^t)^r + 1 \equiv 0 \pmod{(1 - \zeta_r)^{r+1}}.$$

Proof. As

$$g(\chi) = \sum_{n=1}^{p-1} \chi(n)\zeta_p^n,$$

the binomial theorem yields

$$\begin{aligned} -g(\chi)^r &= \Big(-\sum_{n=1}^{p-1} \zeta_p^n + \sum_{n=1}^{p-1}(1 - \chi(n))\zeta_p^n\Big)^r = (1 + \sum_n (1 - \chi(n))\zeta_p^n)^r \\ &\equiv 1 + r\sum_n (1 - \chi(n))\zeta_p^n + \sum_n (1 - \chi(n))^r \zeta_p^{rn} \pmod{(1 - \zeta_r)^{r+1}}, \end{aligned}$$

as all other terms are divisible by at least $r(1 - \zeta_r)^2$. By Lemma 3, if $\chi(n) \neq 1$, $(1 - \chi(n))^{r-1} \equiv -r \pmod{(1 - \zeta_r)^r}$, and clearly, if $\chi(n) = 1$,

$$(1 - \chi(n))^r \equiv -r(1 - \chi(n)) \pmod{(1 - \zeta_r)^{r+1}}.$$

Thus,

$$\begin{aligned} -g(\chi)^r &\equiv 1 + r\Big(\sum_{n=1}^{p-1}(1 - \chi(n))\zeta_p^n - (1 - \chi(n))\zeta_p^{rn}\Big) \\ &\equiv 1 + r\sum_n (1 - \chi(n))\zeta_p^n - (1 - \chi(n)\chi(r)^{-1})\zeta_p^n \\ &\equiv 1 - r(1 - \chi(r)^{-1})\sum_n \chi(n)\zeta_p^n \\ &\equiv 1 - r(1 - \chi(r)^{-1})\sum_n \zeta_p^n \\ &\equiv 1 + r(1 - \chi(r)^{-1}) \pmod{(1 - \zeta_r)^{r+1}}. \end{aligned}$$

By (iv) of Lemma 1,

$$-g(\chi^t)^r = -\sigma_t(g(\chi)^r) \equiv 1 + r(1 - \chi(r)^{-t}) \pmod{(1 - \zeta_r)^{r+1}},$$

which completes the first statement of Theorem 1. The second statement in Theorem 1 then follows immediately.

Let q denote any positive rational prime other than r, f the least positive integer such that $q^f \equiv 1 \pmod{r}$, and $ef = r - 1$. Then in $Q(\zeta_r)$ the ideal $(q) = \mathfrak{A}_1\mathfrak{A}_2 \cdots \mathfrak{A}_e$, where the $\mathfrak{A}_j$ are prime ideals and

$$(6) \qquad \operatorname*{Norm}_{Q(\zeta_r),Q}(\mathfrak{A}_j) = q^f.$$

In the following let $\mathfrak{A}$ be any of the e prime divisors $\mathfrak{A}_j$, $j = 1, \cdots, e$.

THEOREM 2. *Let q, p, and r be distinct.*

Then

$$(7) \qquad g(\chi)^{q^f-1} \equiv \chi(q)^{-f} \pmod q .$$

Consequently $\chi(q) = 1$ *if and only if*

$$(8) \qquad g(\chi)^r \equiv \beta^r \pmod{\mathfrak{A}} \text{ for some } \beta \in Q(\zeta_r) .$$

Proof.
$$g(\chi)^{q^f} = \left(\sum_{n=1}^{p-1} \chi(n)\zeta_p^n\right)^{q^f}$$
$$\equiv \sum_{n=1}^{p-1} \chi(n)^{q^f}\zeta_p^{nq^f} \pmod q$$
$$\equiv \sum_n \chi(n)\zeta_p^{nq^f} \pmod q, \text{ as } r \mid q^f - 1 ,$$
$$\equiv \chi(q)^{-f} g(\chi) \pmod q .$$

Multiplying both sides of the above congruence by $\overline{g(\chi)}$, and noting (i) of Lemma 1, yields

$$pg(\chi)^{q^f-1} \equiv \chi(q)^{-f} p \pmod q \text{ or } g(\chi)^{q^f-1} \equiv \chi(q)^{-f} \pmod q ,$$

as p and q are distinct primes. Hence, we have proved (7).

Note that as $r \mid q^f - 1$, (7) becomes a congruence in $Q(\zeta_r)$. As $f \mid r-1$, $(f, r) = 1$, we have by (7) that $\chi(q) = 1$ if and only if $g(\chi)^{q^f-1} \equiv 1 \pmod{\mathfrak{A}}$.

(Note that $1 - \zeta_r^t \not\equiv 0 \pmod{\mathfrak{A}}$ unless $\zeta_r^t = 1$.)

If $g(\chi)^r \equiv \beta^r \pmod{\mathfrak{A}}$ for some $\beta \in Q(\zeta_r)$, then

$$g(\chi)^{q^f-1} \equiv \beta^{q^f-1} \equiv 1 \pmod{\mathfrak{A}}$$

by (6).

Conversely, if $g(\chi)^{q^f-1} \equiv 1 \pmod{\mathfrak{A}}$ then $(g(\chi)^r)^{(q^f-1)/r} \equiv 1 \pmod{\mathfrak{A}}$. By Lemma 1, $g(\chi)^r \in Q(\zeta_r)$. By (6) this implies $g(\chi)^r \equiv \beta^r \pmod{\mathfrak{A}}$. (Euler's Criterion for rth powers.)

In the above argument we must bear in mind that $g(\chi) \notin Q(\zeta_r)$.

II. In the last section we have developed a criterion for rth power residuacity in $Q(\zeta_r)$. From this we derive a criterion in the rational numbers Q, which is the purpose of Theorems A and B.

First let us assume that there is a rational integral solution $X_j = a_j$ of equations (1), (2), (3) and (4). In $Q(\zeta_r)$ define the algebraic integer $\alpha = \sum_{j=1}^{r-1} a_j \zeta_r^j$. We shall prove that α satisfies

$$(9) \qquad |\sigma_k(\alpha)|^2 = p^{r-2} , \qquad k = 1, 2, \cdots, r-1 .$$

$$(10) \qquad (p\alpha)^k \sigma_k(p\alpha)^{-1}$$

is also an algebraic integer in $Q(\zeta_r)$, for $k = 1, 2, \cdots, r-1$.

To prove (9) we note that

$$\begin{aligned}|\alpha|^2 &= \Big(\sum_j a_j\zeta_r^j\Big)\Big(\sum_i a_i\zeta^{r-i}\Big)\\ &= \sum_{j,i} a_j a_i \zeta_r^{j-i}\\ &= \sum_{j=1}^{r-1} a_j^2 + \sum_{i=1}^{r-1}\big(\textstyle\sum_i^{(1)} a_{j_1}a_{j_2}\big)\zeta_r^i\,.\end{aligned}$$

By (2) all of the coefficients of ζ_r^i are equal, since for any i, the sums corresponding to i and $r-i$ are identical. Thus

$$\begin{aligned}|\alpha|^2 &= \sum_j a_j^2 - \textstyle\sum_1^{(1)} a_{j_1}a_{j_2}\\ &= \sum_j a_j^2 - (r-1)^{-1}\sum_{i=1}^{r-1}\textstyle\sum_i^{(1)} a_{j_1}a_{j_2}\\ &= r(r-1)^{-1}\sum_j a_j^2 - (r-1)^{-1}\sum_{i=0}^{r-1}\textstyle\sum_i^{(1)} a_{j_1}a_{j_2}\\ &= r(r-1)^{-1}\sum_{j=1}^{r-1} a_j^2 - (r-1)^{-1}\Big(\sum_{j=1}^{r} a_j\Big)^2\\ &= p^{r-2}\end{aligned}$$

by (1). Similarly $|\sigma_k(\alpha)|^2 = p^{r-2}$. Thus (1) and (2) imply (9).

Let k be a fixed integer $2 \leqq k \leqq r-1$. Then

$$\begin{aligned}(p\alpha)^k\sigma_k(p\alpha)^{-1} &= p^{k-1}\alpha^k\sigma_k(\alpha)^{-1}\\ &= p^{k-1}\alpha^k\sigma_{-k}(\alpha)\,|\,\sigma_k(\alpha)\,|^{-2}\\ &= p^{-r+k+1}\alpha^k\sigma_{-k}(\alpha)\end{aligned}\tag{11}$$

by (10). Now

$$\begin{aligned}\alpha^k\sigma_{-k}(\alpha) &= \big(\textstyle\sum a_j\zeta_r^j\big)^k\big(\textstyle\sum a_j\zeta_r^{-jk}\big)\\ &= \sum_{i=0}^{r-1}\big(\textstyle\sum_i^{(k)} a_{j_1}\cdots a_{j_{k+1}}\big)\zeta_r^i\\ &= \sum_{i=1}^{r-1}\big(\textstyle\sum_i^{(k)} - \textstyle\sum_0^{(k)}\big)\zeta_r^i\,.\end{aligned}\tag{12}$$

Condition (4) implies that each coefficient of ζ_r^i in (12) is divisible by p^{r-k-1}. Placing this information in (11) states that $(p\alpha)^k\sigma_k(p\alpha)^{-1}$ is an integer; thus proving (10).

(4) also tells us that p, but not p^2, divides $p\alpha$, as not all the coefficients of ζ_r^j *in* $\alpha = \sum_{j=1}^{r-1} a_j\zeta_r^j$ *are divisible by* p.

If we restate the above facts in terms of ideals, we have that $(p\alpha)$ is an integral ideal in $Q(\zeta_r)$ divisible only by the prime ideals which divide p.

There exists one prime ideal, say $\mathfrak{p}$, dividing p, which divides $p\alpha$ but $\mathfrak{p}^2$ does not divide $p\alpha$. All other prime factors of p in $Q(\zeta_r)$ are of the form $\sigma_i^{-1}\mathfrak{p}$. Hence,

(13) $$(p\alpha) = \sum_{i=1}^{r-1} \sigma_i^{-1}\mathfrak{p}^{d_i} \text{ where } d_1 = 1,\, d_i > 0 \ .$$

By (9)

$$\begin{aligned}(p\alpha)(\sigma_{-1}(p\alpha)) &= (p^2\,|\,\alpha\,|^2) = p^r \\ &= \Big(\prod_i \sigma_i^{-1}\mathfrak{p}^{d_i}\Big)\Big(\prod_i \sigma_{-1}\sigma_i^{-1}\mathfrak{p}^{d_i}\Big) \\ &= \prod_i \sigma_i^{-1}\mathfrak{p}^{d_i+d_{r-i}}\end{aligned}$$

or

(14) $$d_i + d_{r-i} = r \ .$$

By (10), $(p\alpha)^k\sigma_k(p\alpha)^{-1}$ is integral, or

$$\begin{aligned}(p\alpha)^k(\sigma_k(p\alpha))^{-1} &= \prod_i \sigma_i^{-1}\mathfrak{p}^{d_ik} \prod_i \sigma_k\sigma_i^{-1}\mathfrak{p}^{-d_i} \\ &= \prod \sigma_i^{-1}\mathfrak{p}^{d_ik-d_{ik}}\end{aligned}$$

is an integral ideal. (The index of d_{ik} is interpreted mod r.) Hence, $kd_i \geqq d_{ik}$.

As $d_1 = 1$, $k \geqq d_k$ for $k = 2, 3, \cdots, r-2$. By (14) this yields that $d_k = k$. By Lemma 2, we arrive at the fact that in terms of ideals

(15) $$(p\alpha) = (g(\chi^t)^r) \text{ for some } 1 \leqq t < r \ .$$

In proving (15) we have used (1), (2) and (4). We wish to prove that $p\alpha = g(\chi^t)^r$. To do this we now utilize (3). By (15) we have that for some unit $\eta \in Q(\zeta_r)$, $g(\chi^t)^r = \eta p\alpha$, or

(16) $$g(\chi^{tk})^r = \sigma_k(\eta p\alpha) = \sigma_k(\eta)\sigma_k(p\alpha) \ .$$

Taking the absolute value of both sides of (16) and utilizing (i) of Lemma 1 and (9) gives $p^r = |\,\sigma_k(\eta)\,|^2p^r$, or $|\,\sigma_k(\eta)\,|^2 = 1$. By a Theorem of Dirichlet on units (See [3] Theorem IV 9, A pp. 174), any unit which has all of its conjugates with absolute value 1 is then a root of unity. As $\eta \in Q(\zeta_r)$, $\eta = \pm\zeta_r^s$.

Now

$$\begin{aligned}\alpha = \sum_{j=1}^{r} a_j\zeta_r^j &= \sum_j a_j - \sum_j a_j(1-\zeta_r^j) \\ &\equiv \sum_j a_j - \sum_j ja_j(1-\zeta_r) \pmod{(1-\zeta_r)^2} \ ,\end{aligned}$$

by Lemma 3. As $p \equiv 1 \pmod r$, $p \equiv 1 \pmod{(1-\zeta_r)^2}$. By (3),

$$1 + \sum_j a_j \equiv \sum_j ja_j \equiv 0 \pmod r \ .$$

Hence, $p\alpha \equiv -1 \pmod{(1-\zeta_r)^2}$. By Theorem 1, $g(\chi^t)^r \equiv -1 \pmod{1-\zeta_r)^2}$. Therefore, $\eta \equiv 1 \pmod{(1-\zeta_r)^2}$. But $\eta = \pm\zeta_r^s \equiv \pm(1+s(1-\zeta_r)) \pmod{(1-\zeta_r)^2}$; i.e., $s \equiv 0 \pmod r$ and the $+$ sign holds. Hence, $\eta = 1$.

Therefore, if the a_j are any integral solution of (1), (2), (3) and (4), there exists an integer $1 \leqq t \leqq r-1$ such that

$$p\sum_{j=1}^{r-1} a_j\zeta_r^j = g(\chi^t)^r . \tag{17}$$

Conversely, given any integer $t, 1 \leqq t \leqq r-1$, and writing

$$g(\chi^t)^r = p\sum_{j=1}^{r-1} a_j\zeta_r^j ,$$

we can prove that the a_j are rational integers which satisfy (1), (2), (3), and (4). The proof is merely reversing the above steps we used in proving (17). By Lemma 2 the prime factorizations of $(g(\chi^s)^r)$ and $(g(\chi^t)^r)$, $1 \leqq s < t \leqq r-1$, are distinct, and thus $g(\chi^s)^r \neq g(\chi^t)^r$. Hence, we have shown that there are precisely $r-1$ rational integral solutions of (1), (2), (3), and (4).

We are now in a position to prove Theorems A and B. First for Theorem A.

Let a_j be an integral solution of (1) through (4). Then we have shown that $p\sum_{j=1}^{r-1} a_j\zeta_r^j = g(\chi^t)^r$ for some integer t relatively prime to r. By Theorem 1, the above states that $\chi(r) = 1$ if and only if $p\sum_j a_j\zeta_r^j \equiv -1 \pmod{(1-\zeta_r)^{r+1}}$.

Define b_s, $s = 0, 1, \cdots, r-2$, by $b_0 = -pa_{r-1}$, $b_s = p(a_s - a_{r-1})$, $s = 1, 2, \cdots, r-2$. Then

$$p\sum_{j=1}^{r-1} a_j\zeta_r^j = \sum_{s=0}^{r-2} b_s\zeta_r^s .$$

Further let

$$C_i = (-1)^i \sum_{s=i}^{r-2} \binom{s}{i} b_s ,$$

where $\binom{s}{i}$ is the binomial coefficient. Then

$$\begin{aligned} p\sum_{j=1}^{r-1} a_j\zeta_r^j = \sum_{s=0}^{r-2} b_s\zeta_r^s &= \sum_s b_s(1-(1-\zeta_r))^s \\ &= \sum_s b_s \sum_{i=0}^{s} (-1)^i\binom{s}{i}(1-\zeta_r)^i \\ &= \sum_{i=0}^{r-2} C_i(1-\zeta_r)^i . \end{aligned}$$

The first statement in Theorem 1 states that $g(\chi^t)^r + 1 \equiv 0 \pmod{(1-\zeta_r)^r}$. Hence,

$$\begin{aligned} \sum_{i=0}^{r-2} C_i(1-\zeta_r)^i + 1 &\equiv (C_0+1) + \sum_{i=1}^{r-2} C_i(1-\zeta_r)^i \\ &\equiv 0 \pmod{(1-\zeta_r)^r} \end{aligned}$$

This implies that $C_0 + 1 \equiv 0 \pmod{r^2}$. Hence,

$$\sum_{i=0}^{r-2} C_i(1-\zeta_r)^i \equiv C_1(1-\zeta_r) \pmod{(1-\zeta_r)^{r+1}}$$

or that $\chi(r) = 1$ if and only if

$$C_1 \equiv 0 \pmod{r^2} . \tag{18}$$

Now

$$\begin{aligned} C_1 &= (-1)\sum_{s=1}^{r-2}\binom{s}{1}b_s = -\sum_{s=1}^{r-2} sb_s \\ &= -p\sum_{s=1}^{r-2} s(a_s - a_{r-1}) \\ &= -p\sum_{s=1}^{r-2} sa_s + \frac{1}{2}p(r-2)(r-1)a_{r-1} \\ &\equiv -p\left(\sum_{s=1}^{r-1} sa_s + \frac{1}{2}ra_{r-1}\right) \pmod{r^2} . \end{aligned} \tag{19}$$

Equations (18) and (19) complete the proof of Theorem A.

Theorem B is also derived immediately from Theorem 2. If $q \equiv 1 \pmod r$, q a positive rational prime, then in $Q(\zeta_r)$, $(q) = \mathfrak{A}_1\mathfrak{A}_2 \cdots \mathfrak{A}_{r-1}$, where $\mathfrak{A}_j$ are prime ideals and $\mathrm{Norm}_{Q(\zeta_r),Q}\,\mathfrak{A}_j = q$.

We may take $0, 1, 2, \cdots, q-1$ as a set of residues $(\mathrm{mod}\,\mathfrak{A}_1)$. Hence, as $1 - \zeta_r^t \not\equiv 0 \pmod{\mathfrak{A}_1}$, unless $\zeta_r^t = 1$, $\zeta_r \equiv h \pmod{\mathfrak{A}_1}$, where h is a rational integer such that $h^r \equiv 1 \pmod q$.

Thus by Theorem 2, $\chi(q) = 1$ if and only if there is a $\beta \in Q(\zeta_r)$ such that $g(\chi^t)^r = p\sum_j a_j\zeta_r^j \equiv p\sum_j a_j h^j \equiv \beta^r \pmod{\mathfrak{A}_1}$.

We may take $\beta = b \in Q$ by the above remarks.

Hence, $\chi_p(q) = 1$ if and only if $\chi_q(p\sum_j a_j h^j) = 1$ where χ_q is a primitive rth power character $(\mathrm{mod}\,q)$.

If we had chosen another h_1 whose order was $r \pmod q$, then $h_1 \equiv h^t \pmod{\mathfrak{A}_1}$, and

$$p\sum_j a_j h_1^t \equiv p\sum_j a_j\zeta_r^{jt} \equiv g(\chi^t)^r \pmod{\mathfrak{A}_1} .$$

Thus, any h whose order $(\mathrm{mod}\,q)$ is r works equally well in Theorem B.

There are several special cases one can derive when $q \not\equiv 1 \pmod r$, in particular, when $q = 2$, and $r = 5, 7$.

If $q = 2$, $r = 5$, then in $Q(\zeta_r)$, 2 remains a prime because 2^4 is the least power of 2 congruent to $1 \pmod 5$. One can easily compute that the only elements in $Q(\zeta_5)$ which are fifth powers $(\mathrm{mod}\,2)$ are $1 = -\sum_{j=1}^4 \zeta_5^j$, $\zeta_5 + \zeta_5^{-1}$, and $\zeta_5^2 + \zeta_5^{-2} \pmod 2$. Hence, for $r = 5$, $\chi_p(2) = 1$ if and only if $a_j \equiv a_{5-j} \pmod 2$.

For $q = 2$, $r = 7$, then $2^3 \equiv 1 \pmod 7$. Hence, in $Q(\zeta_7)$, $(2) = \mathfrak{A}_1\mathfrak{A}_2$ where $\mathrm{Norm}\,\mathfrak{A}_i = 8$. For $\alpha \equiv \beta^7 \pmod{\mathfrak{A}_1}$, $\beta \not\equiv 0 \pmod{\mathfrak{A}_1}$, and $\beta \in Q(\zeta_7)$

implies $\alpha \equiv 1 \pmod{\mathfrak{A}_1}$. Hence, for $r = 7$, $\chi_p(2) = 1$ if and only if $a_j \equiv 1 \pmod 2$ for $j = 1, \cdots, 6$.

One could easily generalize this to the case when $r = 2^s - 1$. Then $\chi_p(2) = 1$ if and only if $a_j \equiv 1 \pmod 2$ for $j = 1, \cdots, r-1$.

BIBLIOGRAPHY

1. H. Davenport and H. Hasse, *Die Nullstellen der Kongruenzzetafunktionen in gewissen zyklischen Fällen,* Journal für die reine und angewandte Mathematik, Band CLXXII, (1935), 151-182.

2. E. Lehmer, *Criteria for cubic and quartic residuacity,* Mathematika, **5**, Part 1, (1958), 20-29.

3. H. Weyl, *Algebraic Theory of Numbers,* Annals of Mathematical Studies, Princeton University Press, 1940.

ON INVARIANT PROBABILITY MEASURES I

J. R. BLUM[1] AND D. L. HANSON[2]

1. Introduction. Let Ω be a set and let $\mathscr{A}$ be a σ-algebra of subsets of Ω. Let T be a one-to-one bimeasurable transformation mapping Ω onto itself. T then induces the group of transformations $\{T^i, i = 0, \pm 1, \cdots\}$ defined in the usual way. If $A \in \mathscr{A}$, T^iA is defined to be the set of images of the elements of A under the transformation T^i.

Let $\mathscr{P}$ be the class of probability measures defined on $\mathscr{A}$ for which T is invariant, i.e. if P is a probability measure defined on $\mathscr{A}$ then $P \in \mathscr{P}$ if and only if $PA = PTA$ for every $A \in \mathscr{A}$. Let $\mathscr{A}_1$ be the subclass of $\mathscr{A}$ which is invariant under T; a set $A \in \mathscr{A}$ belongs to $\mathscr{A}_1$ if and only if $A = TA$. It is trivial to verify that $\mathscr{A}_1$ is sub-σ-algebra of $\mathscr{A}$. Finally let $\mathscr{P}_1$ be the subclass of $\mathscr{P}$ for which T is ergodic, i.e. if $P \in \mathscr{P}$ then $P \in \mathscr{P}_1$ if and only if $PA = 0$ or $PA = 1$ for every $A \in \mathscr{A}_1$.

In § 2. several results are proved, concerning the structure of the class $\mathscr{P}$. These are not new, although several of them do not seem to have appeared in the literature. The main theorem of this paper is in § 3 where it is shown that each element of $\mathscr{P}$ can be represented as a convex combination of the extreme points of $\mathscr{P}$. Several consequences of this theorem are pointed out.

2. Some properties of the class $\mathscr{P}$.

THEOREM 1. *Let P and Q be elements of $\mathscr{P}$. Suppose $PA = QA$ for $A \in \mathscr{A}_1$. Then $P \equiv Q$.*

Proof. Let $\mu = P - Q$. Then μ is a completely additive set function defined on $\mathscr{A}$. If μ is not identically zero, there exists $A \in \mathscr{A}$ such $\mu(A) > 0$ and $\mu(A) \geqq \mu(B)$ for all $B \in \mathscr{A}$. This follows from the Hahn decomposition theorem. Write $\mu(A) = \alpha + \beta$, where $\alpha = \mu(A - A \cap TA)$ and $\beta = \mu(A \cap TA)$. Since $\mu(A - A \cap TA) = \mu(TA - A \cap TA)$ we have $\mu(A \cup TA) = 2\alpha + \beta$. Now if $\alpha < 0$, then $\mu(A \cap TA) > \mu(A)$ and A is not maximal, and if $\beta < 0$ then $\mu(A - A \cap TA) > \mu(A)$ and A is not maximal. Consequently $\alpha \geqq 0$ and $\beta \geqq 0$. But if A is maximal then $\alpha + \beta \geqq 2\alpha + \beta$. Hence $\alpha = 0$ and $\mu(A \cup TA) = \mu(A)$. By the same argument we show that $\mu(T^{-1}A \cup A \cup TA) = \mu(A)$ and it follows by in-

Received December 9, 1959.

[1] Supported in part by the Office of Ordnance Research, U. S. Army, under Contract No. DA-33-008-ORD-965.

[2] Work done while the author was a Standard Oil Co. Fellow at Indiana University.

duction that $\mu(B_n)=\mu(A)$ for every positive integer n, where $B_n = \bigcup_{i=-n}^{n} T^i A$. Now B_n is an increasing sequence of sets. Let $B = \lim_{n\to\infty} B_n$. Then $\mu(B) = \mu(A) > 0$. But clearly $B = \bigcup_{i=-\infty}^{\infty} T^i A \in \mathscr{A}_1$ and μ is zero on $\mathscr{A}_1$. Consequently we have a contradiction and the theorem is proved.

Suppose now that $P \in \mathscr{P}_1$ and $Q \in \mathscr{P}$ and suppose also that Q is absolutely continuous with respect to P. Then if $A \in \mathscr{A}_1$ we have $PA = 0$ or $PA = 1$ and hence Q agrees with P on $\mathscr{A}_1$. Thus the theorem applies and we have

COROLLARY 1. *If $P \in \mathscr{P}_1, Q \in \mathscr{P}$, and Q is absolutely continuous with respect to P then $Q \equiv P$.*

Theorem 1 also furnishes an elegant proof of a result which was proved by Lamperti [3], and in a special situation by Harris [1]. Suppose P and Q are both ergodic, i.e. $P \in \mathscr{P}_1$ and $Q \in \mathscr{P}_1$. Then either P and Q are orthogonal or for each $A \in \mathscr{A}$ for which $PA = 1$ we have $Q(A) > 0$. Now suppose $A \in \mathscr{A}_1$ and $PA = 1$. Then if Q is not orthogonal to P and since $Q \in \mathscr{P}_1$ we must have $Q(A) = 1$ and it follows that $P = Q$ on $\mathscr{A}_1$. We have

COROLLARY 2. *If $P \in \mathscr{P}_1, Q \in \mathscr{P}_1$, then either $P \equiv Q$ or P is orthogonal to Q.*

In § 3, we shall show that this result can be considerably generalized.

THEOREM 2. *$\mathscr{P}$ is a convex set. $P \in \mathscr{P}_1$ if and only if P is an extreme point of $\mathscr{P}$.*

Proof. The first statement is obvious. Suppose $P \in \mathscr{P}_1$ and suppose we may represent P in the form $P \equiv \alpha P_1 + (1-\alpha)P_2$ where $0 < a < 1$ and $P_i \in \mathscr{P}, i = 1, 2$. Then clearly P_1 and P_2 are absolutely continuous with respect to P and it follows from Corollary 1 that $P_1 \equiv P_2 \equiv P$. Thus if $P \in \mathscr{P}_1$ it is an extreme point of $\mathscr{P}$. Conversely if $P \notin \mathscr{P}_1$ there exists a set $B \in \mathscr{A}_1$ with $0 < PB < 1$. Then we may write $P \equiv \alpha P_1 + (1-\alpha)P_2$ where $\alpha = PB$, and for $A \in \mathscr{A}$ we have $P_1(A) = P(A \cap B)/P(B)$ and $P_2(A) = P(A \cap B^c)/P(B^c)$. It is easily verified that P_1 and P_2 are invariant probability measures and it follows that P is not an extreme point of $\mathscr{P}$.

Theorem 2 strongly suggests that it may be possible to obtain the elements of $\mathscr{P}$ as convex combinations of the extreme points of $\mathscr{P}_1$. Under a rather mild assumption this is in fact true, as will be shown in the next section. Examples of the kind of theorem we have in mind were proved by Hewitt and Savage [2].

3. The representation theorem.

Throughout part of this section we shall assume that if $A \in \mathscr{A}_1$ and if $PA = 0$ for every $P \in \mathscr{P}_1$ then

$PA = 0$ for every $P \in \mathscr{P}$. Clearly such a condition is necessary for a convex representation theorem and the condition can actually be verified in many examples of interest.

Suppose now that $P \in \mathscr{P}_1$. Theorem 1 tells us that P has a unique invariant extension from $\mathscr{A}_1$ to $\mathscr{A}$. This suggests that if $A \in \mathscr{A}$ we should be able to determine PA by knowing only the values of P on $\mathscr{A}_1$. A proof of this statement follows from the individual ergodic theorem.

THEOREM 3. *Let $A \in \mathscr{A}$. For every α with $0 \leqq \alpha \leqq 1$ there exists a set $A'_\alpha \in \mathscr{A}_1$ such that if $P \in \mathscr{P}_1$ then $PA = \alpha$ if and only if $PA'_\alpha = 1$.*

Proof. Let $f_S(x)$ be the set characteristic function of the set S. Let $A \in \mathscr{A}$, and α be given. For every positive integer n define $g_{n,A}(x) = 1/n \sum_{i=1}^{n-1} f_A(T^i x)$, and define $A'_\alpha = \{x \mid \lim_{n\to\infty} g_{n,A}(x) = \alpha\}$. Clearly $A'_\alpha \in \mathscr{A}_1$ and the individual ergodic theorem implies that $PA = \alpha$ if and only if $PA'_\alpha = 1$, whenever $P \in \mathscr{P}_1$.

Using the same technique we can prove

THEOREM 4. *Let $A \in \mathscr{A}$. For every α with $0 \leqq \alpha \leqq 1$ there exists a set $A_\alpha \in \mathscr{A}_1$ such that if $P \in \mathscr{P}_1$ then $PA \leqq \alpha$ if and only if $PA_\alpha = 1$.*

Let $A \in \mathscr{A}_1$. Define π_A by $\pi_A = \{P \in \mathscr{P}_1 \mid PA = 1\}$. Let Π be the collection of all such sets π_A i.e. $\Pi = \{\pi_A \mid A \in \mathscr{A}_1\}$. The following facts are easily verified:

(i) $\pi_\Omega = \mathscr{P}_1$

(ii) $[\pi_A]^c = \pi^c$

(iii) $\pi \bigcup_n A_n = \bigcup_n \pi A_n$

where A and each A_n is an element of $\mathscr{A}_1$. Since $\mathscr{A}_1$ is a σ-algebra it follows that Π is a σ-algebra. Now let $Q \in \mathscr{P}$. We define a set function μ_Q on Π by $\mu_Q(\pi_A) = Q(A)$.

We shall show that under the assumption at the beginning of this section μ_Q is in fact a probability measure defined on Π. Clearly $\mu_Q(\pi_A) \geqq 0$ for each π_A, and $\mu_Q(\mathscr{P}_1) = \mu_Q(\pi_\Omega) = Q(\Omega) = 1$. Now suppose $\{\pi_{A_n}\}$ is a sequence of disjoint elements of π. It is easily verified that this is the case if and only if $PA_n \cap A_m = 0$ for every pair of sets A_n, A_m in $\mathscr{A}_1$ with $n \neq m$ and for every $P \in \mathscr{P}_1$. It follows from the assumption that $Q(A_n \cap A_m) = 0$ for $n \neq m$. Hence $\mu_Q\{\bigcup_n \pi A_n\} = Q(\bigcup_n A_n) = \sum_n Q(A_n) = \sum_n \mu_Q\{\pi_{A_n}\}$ and we have shown that μ_Q is a probability measure defined on Π. We summarize in

THEOREM 5. *If Π and μ_Q are defined as above then Π is a σ-algebra of subsets of $\mathscr{P}_1$. Under the assumption at the beginning of this section μ_Q is a probability measure defined on Π.*

THEOREM 6. *Let $A \in \mathscr{A}$. Consider the function $f_A(P)$ defined on $\mathscr{P}_1$ and with values $f_A(P) = PA$. Then $f_A(P)$ is measurable with respect to Π.*

Proof. We must show that for every α with $0 \leqq \alpha \leqq 1$ we have $\{P \in \mathscr{P}_1 \mid f_A(P) \leqq \alpha\} = \{P \in \mathscr{P}_1 \mid PA \leqq \alpha\} \in \Pi$. But it follows from Theorem 4 that $\{P \in \mathscr{P}_1 \mid PA \leqq \alpha\} = \pi_{A \leftarrow A_\alpha}$ where $A_\alpha \in \mathscr{A}_1$ is the set guaranteed by Theorem 4, and the theorem follows.

Since $f_A(P)$ is bounded and measurable it is clearly integrable with respect to any probability measure defined on Π. Now let $Q \in \mathscr{P}$ and μ_Q be the corresponding probability measure defined on Π. For each $A \in \mathscr{A}$ define $Q'(A)$ by

$$Q'(A) = \int_{\mathscr{P}_1} f_A(P) d\mu_Q = \int_{\mathscr{P}_1} PAd\mu_Q \,.$$

It follows immediately from this definition that Q' is an invariant probability measure defined on $\mathscr{A}$. But if $A \in \mathscr{A}_1$ we have $Q'(A) = \mu_Q\{\pi_A\} = Q(A)$. Hence $Q' = Q$ on $\mathscr{A}_1$ and it follows from Theorem 1 that $Q' \equiv Q$. Furthermore suppose we know that $Q(A) = \int_{\mathscr{P}_1} PAd\mu$, where μ is some probability measure defined on Π. Then if $A \in \mathscr{A}_1$ we have $Q(A) = \int_{\mathscr{P}_1} PAd\mu = \mu\{\pi_A\} = \mu_Q\{\pi_A\}$, i.e. $\mu \equiv \mu_Q$. We state these results in

THEOREM 7. *Suppose the assumption at the beginning of the section holds. Then for every $Q \in \mathscr{P}$ there exists a unique probability measure μ_Q defined on Π such that*

$$Q(A) = \int_{\mathscr{P}_i} P(A) d\mu_Q \textit{ for every } A \in \mathscr{A} \,.$$

We shall refer to Theorem 7 as the representation theorem, and the rest of this section is devoted to exploring some consequences of this theorem. One immediate consequence is a generalization of Corollary 2 to Theorem 1.

THEOREM 8. *Let $Q_i \in \mathscr{P}$, $i = 1, 2$. Then Q_1 and Q_2 are orthogonal if and only if the corresponding measured μ_{Q_1} and μ_{Q_2} are orthogonal.*

Proof. Suppose Q_1 and Q_2 are orthogonal. Let B be a set such that $Q_1(B) = 1 = Q_2(B^c)$ and let $A = \bigcup_{i=-\infty}^{\infty} T^i B$. Then $A \in \mathscr{A}_1$ and $Q_1(A) = 1 = Q_2(A^c)$ and we obtain $1 = \mu_{Q_1}\{\pi_A\} = \mu_{Q_2}\{(\pi_A)^c\}$. Thus μ_{Q_1} and μ_{Q_2} are orthogonal. Conversely if μ_{Q_1} and μ_{Q_2} are orthogonal there is a set $A \in \mathscr{A}_1$ such that $1 = \mu_{Q_1}\{\pi_A\} = Q_1(A)$ and $0 = \mu_{Q_2}\{\pi_A\} = Q_2(A)$ and the theorem is proved.

Another interesting consequence of the theorem is the obvious fact that if $A \in \mathscr{A}$ and if $PA = 1$ for each $P \in \mathscr{P}_1$ then $Q(A) = 1$ for each $Q \in \mathscr{P}$. Thus the individual ergodic theorem for arbitrary invariant measures is an immediate consequence of that theorem for ergodic measures. Furthermore Theorem 7 throws some light on the evaluation of the limiting function in the individual ergodic theorem. Let $Q \in \mathscr{P}$ and let $f(x)$ be defined on Ω and measurable with respect to $\mathscr{A}$. Let $f_n(x) = 1/n \sum_{i=0}^{n-1} f(T^i x)$. Then if $f \in L_1(Q)$ the ergodic theorem states that $\lim_{n\to\infty} f_n(x) = f^*(x)$ say, exists on a set of Q-measure one. It is clear that f^* is invariant i.e. $f^*(Tx) = f^*(x)$ for all x for which f^* exists. If f is also integrable with respect to $P \in \mathscr{P}_1$ then f^* is constant on a set of P-measure one, and we have

$$Q\{x \mid f^*(x) \leqq u\} = \int_{\mathscr{P}_1} Px\{\mid f^*(x) \leqq u\} d\mu_Q = \mu_Q\{P \in \mathscr{P}_1 \mid f^* \leqq u\} ,$$

In particular we conclude f^* is a constant, say c, on a set of Q-measure one if and only if $\mu_P[P \in \mathscr{P}_1 \mid P\{x \mid f^*(x) = c\}] = 1$.

Finally, suppose f is again measurable with respect to $\mathscr{A}$. Let $Q \in \mathscr{P}$ and suppose $\mu_Q P\left\{\in \mathscr{P}_1 \mid \int_\Omega |f| \, dP < \infty\right\} = 1$. Then we can easily prove

THEOREM 8. *If* $\int_\Omega |f| \, dP$ *is an integrable function of* P *(with respect to* μ_Q*) then* $f \in L_1(Q)$ *and*

$$\int_\Omega f dQ = \int_{\mathscr{P}_1} \left[\int_\Omega f dP\right] d\mu_Q .$$

REFERENCES

1. T. E. Harris, *On chains of infinite order*, Pacific J. Math., **5** (1955), 707–724.
2. E. Hewitt and L. J. Savage, *Symmetric measures on Cartesian products*, TAMS, **80** (1955), 470–501.
3. J. Lamperti, *Stationary measures for certain stochastic processes*, Pacific J. Math., **8** (1958), 127–132.

SANDIA CORPORATION AND
INTERNATIONAL BUSINESS MACHINES CORPORATION

POSITIVE OPERATORS COMPACT IN AN AUXILIARY TOPOLOGY

F. F. BONSALL

Of the several generalizations to infinite dimensional spaces of the Perron-Frobenius theorem on matrices with non-negative elements, two are outstanding for their freedom from ad hoc conditions.

THEOREM A (Krein and Rutman [3] Theorem 6.1). *If the positive cone K in a partially ordered Banach space E is closed and fundamental, and if T is a compact linear operator in E that is positive (i.e., $TK \subset K$) and has non-zero spectral radius ρ, then ρ is an eigenvalue corresponding to positive eigenvectors of T and of T^*.*

THEOREM B ([4] p. 749 [1] p. 134). *If the positive cone K in a partially ordered normed space E is normal*[1] *and has interior points, and if T is a positive linear operator in E, then the spectral radius is an eigenvalue of T^* corresponding to a positive eigenvector.*

In [2], we have proved the following generalization of Theorem A.

THEOREM C. *Let the positive cone K in a normed and partially ordered space E be complete, and let T be a positive linear operator in E that is continuous and compact in K. If the partial spectral radius μ of T is non-zero, then μ is an eigenvalue of T corresponding to a positive eigenvector.*

Also in [2], we have developed a single method of proof of Theorems A, B, C which exploits the fact that the resolvent operator is a geometric series, and thus avoids the use of complex analysis or any other deep method.

In [5] (Theorems (10.4), (10.5)), Schaefer has further extended these results by showing that (A) and (C) remain valid for operators in locally convex spaces, with suitable definitions of spectral radius and partial spectral radius.

Our aim in the present article is to unify these theorems still further. We prove a single theorem (Theorem 1) that contains Theorem C (and hence A), and also contains Theorem B except in the case $\rho = 0$, for which an extra gloss is needed (Theorem 2). The central idea is that instead of being compact in K in the norm topology, T maps the part of the unit ball in K into a set that is compact with respect to a

Received February 19, 1960.

[1] K is said to be a *normal* cone if there exists a positive constant κ such that

$$\|x + y\| \geqq \kappa\|x\| \qquad (x, y \in K) .$$

second linear topology, this topology being related to the norm topology in a certain way. This idea is, in essence, derived from the recent paper [6] of Schaefer, though his conditions are too restrictive for our purpose. Again we use only elementary real analysis of the kind used in [2]. After proving our two main theorems, we exhibit a number of examples of situations in which these theorems are applicable.

NOTATION. We suppose that E is a normed and partially ordered real linear space with norm $\|\cdot\|$, norm topology τ_N, and positive cone K; i.e., K is a non-empty set satisfying the axiom:

(i) $x, y \in K$, $\alpha \geqq 0$ imply $x + y$, $\alpha x \in K$,

(ii) $x, -x \in K$ imply $x = 0$.

We write $x \leqq y$ or $y \geqq x$ to denote that $y - x \in K$.

We suppose that K is complete with respect to the norm. However, we do not require that E be complete, so that there is no real loss of generality in supposing that $E = K - K$, and we shall therefore suppose that this is the case. We exclude the trivial case in which $K = (0)$.

We denote by B the intersection of K with the closed unit ball in E, i.e., $B = \{x : x \in K \text{ and } \|x\| \leqq 1\}$, and suppose that T is a linear operator in E that is positive ($TK \subset K$) and partially bounded (i.e., $\|Tx\|$ is bounded on B). We denote the partial bound of T by $p(T)$ i.e.,

$$p(T) = \sup \{\|Tx\| : x \in B\} ,$$

and by μ the partial spectral radius

$$\mu = \lim_{n\to\infty} \{p(T^n)\}^{1/n} .$$

We are indebted to H. H Schaefer for several helpful suggestions, and in particular for pointing out that substantial simplification can be obtained by introducing a second norm q into E defined as follows. Let B_0 denote the convex symmetric hull of B, i.e.,

$$B_0 = \{\alpha x + \beta y\colon\ x, y \in B, |\alpha| + |\beta| = 1\} ,$$

and let q be the gauge functional of B_0,

$$q(x) = \inf \{\lambda\colon\ \lambda > 0 \text{ and } \lambda^{-1}x \in B_0\} .$$

It is easily verified that q is a norm in E, that $q(x) \geqq \|x\|$ $(x \in E)$, and that $q(x) = \|x\|$ $(x \in K)$. Also the completeness of K with respect to the given norm implies that E and K are complete with respect to q.

Given a positive operator T, the partial bound and the partial spectral radius are the usual operator norm and spectral radius for the operator T in the Banach space (E, q). For $\lambda > \mu$, the resolvent operator

$R_\lambda = (\lambda I - T)^{-1}$ is given by the series

$$R_\lambda = \frac{1}{\lambda}I + \frac{1}{\lambda^2}T + \frac{1}{\lambda^3}T^2 + \cdots .$$

which converges in the operator norm for (E, q), and is a partially bounded positive operator.

We suppose that we are given a second linear topology τ in E, such that K is (τ)-closed and T is (τ)-continuous in K.

DEFINITION. Given a subset A of K, we say that τ is *sequentially stronger than* τ_N *at* 0 *relative to* A if 0 is a (τ_N)-cluster point of each sequence of points of A of which it is a (τ)-cluster point.

THEOREM 1. *If TB is contained in a (τ)-compact set, τ is sequentially stronger than τ_N at* 0 *relative to TB, and $\mu > 0$, then there exists a non-zero vector u in K with $Tu = \mu u$.*

THEOREM 2. *If B is contained in a (τ)-compact set, and τ is sequentially stronger than τ_N at* 0 *relative to B, then there exists a non-zero vector u in K with $Tu = \mu u$.*

Since $TB \subset p(T)B$, Theorem 2 is contained in Theorem 1 except when $\mu = 0$.

The proofs of these theorems will depend on the following two lemmas. Lemma 1, which is needed in the proof of Lemma 2, is repeated from [2] in order to make the present paper self-contained.

LEMMA 1. *Let $\{a_n\}$ be an unbounded sequence of non-negative real numbers. Then there exists a subsequence $\{a_{n_k}\}$ such that*

(i) $a_{n_k} > k \qquad (k = 1, 2, \cdots),$

(ii) $a_{n_k} > a_j \qquad (j < n_k, k = 1, 2, \cdots).$

Proof. By induction. With $n_1, \cdots, n_{k-1}$ chosen to satisfy (i) and (ii), let n_k be the smallest positive integer r with $a_r > a_{n_{k-1}} + k$.

LEMMA 2. *If TB is contained in a (τ)-compact set, and τ is sequentially stronger than τ_N at* 0 *relative to TB, then*

$$\lim_{\lambda \to \mu + 0} p(R_\lambda) = \infty .$$

Proof. Suppose that the conditions of the lemma are satisfied, but that $p(R_\lambda)$ does not tend to infinity as λ decreases to μ. Then there exists a positive constant M such that $p(R_\nu) \leq M$ for some ν greater than and arbitrarily close to μ.

The case $\mu = 0$ is easily settled. For if $\mu = 0$, then

$$\lambda R_\lambda x \geqq x \qquad (\lambda > 0,\ x \in K),$$

and letting λ tend to zero through values for which $p(R_\lambda) \leqq M$, we obtain $-x \in K$, $K = (0)$. This is the trivial case that we have excluded.

Suppose now that $\mu > 0$. Then we may choose λ, ν with

$$0 < \lambda < \mu < \nu < \lambda + M^{-1}$$

and with $p(R_\nu) \leqq M$. With this choice of λ, ν the series

$$R_\nu + (\nu - \lambda)R_\nu^2 + (\nu - \lambda)^2 R_\nu^3 + \cdots$$

converges in operator norm for the Banach space (E, q) to a partially bounded positive operator S with

$$Sx = \lambda^{-1}x + \lambda^{-1}TSx \qquad (x \in K).$$

Thus

$$Sx \geqq \lambda^{-1}TSx \qquad (x \in K),$$

and therefore

(1) $$Sx \geqq \lambda^{-(n+1)}T^n x \qquad (x \in K,\ n = 1, 2, \cdots).$$

Since $\lim_{n\to\infty} p(\lambda^{-(n+1)}T^n) = \infty$, and since the partial bound of a positive operator coincides with its operator norm in (E, q), the principle of uniform boundedness implies that there exists a point $x \in E$ with $q(\lambda^{-(n+1)}T^n x)$ unbounded. Since $E = K - K$, it follows that there exists $w \in K$ for which the sequence $(\|\lambda^{-(n+1)}T^n w\|)$ is unbounded. Therefore, by Lemma 1, there exists a subsequence such that

(2) $$\lim_{k\to\infty} \|\lambda^{-(n_k+1)}T^{n_k}w\| = \infty ,$$

(3) $$\|\lambda^{-(n_k+1)}T^{n_k}w\| \geqq \|\lambda^{-n_k}T^{n_k-1}w\| .$$

Since

$$\|T^{n_k}w\| \leqq p(T)\,\|T^{n_k-1}w\| ,$$

we also have

(4) $$\lim_{k\to\infty} \|\lambda^{-n_k}T^{n_k-1}w\| = \infty .$$

Let $y_k = \|T^{n_k-1}w\|^{-1}T^{n_k-1}w$. Then, by (1), there exists $z_k \in K$ with

(5) $$\|\lambda^{-n_k}T^{n_k-1}w\|^{-1}Sw = \lambda^{-1}Ty_k + z_k \qquad (k = 1, 2, \cdots).$$

By (4) and (5), we have

(6) $$\lambda^{-1}Ty_k + z_k \to 0 \quad (\tau) .$$

Since $y_k \in B$ and TB is contained in a (τ)-compact set, the sequence $(\lambda^{-1}Ty_k)$ has a (τ)-cluster point y in K. By (6), $-y$ is a (τ)-cluster point of (z_k), and since $z_k \in K$ and K is (τ)-closed, $-y \in K$. Thus $y = 0$, and 0 is a (τ)-cluster point of (Ty_k). But τ is sequentially stronger than τ_N at 0 relative to TB, and so 0 is a (τ_N)-cluster point of (Ty_k). But this is absurd, for, by (3),

$$\| Ty_k \| \geqq \lambda \| y_k \| = \lambda .$$

Proofs of Theorems 1 *and* 2. Since $TB \subset p(T)B$, Lemma 2 is available under the conditions of each theorem, and gives

$$\lim_{\lambda \to \mu + 0} p(R_\lambda) = \infty .$$

Then, applying the principle of uniform boundedness as in the proof of Lemma 2, we see that there exists a sequence (λ_n) converging decreasingly to μ, and a point w in K with $\| w \| = 1$ and

$$\lim_{n \to \infty} \| R_{\lambda_n} w \| = \infty ,$$

and we may suppose that $R_{\lambda_n} w \neq 0$ $(n = 1, 2, \cdots)$. Let $\alpha_n = \| R_{\lambda_n} w \|^{-1}$, and $u_n = \alpha_n R_{\lambda_n} w$. Then

$$\mu u_n - Tu_n = (\mu - \lambda_n) u_n + \alpha_n w . \tag{8}$$

Under the conditions of Theorem 2, the proof is easily completed. For, since $u_n \in B$ and B is contained in a (τ)-compact set, it follows from (8) that

$$\mu u_n - Tu_n \to 0 \quad (\tau) .$$

Also (u_n) has a (τ)-cluster point u in K, and since T is (τ)-continuous in K, we have

$$\mu u - Tu = 0 .$$

We have $u \neq 0$, for otherwise 0 is a (τ_N)-cluster point of (u_n), which is absurd, since $\| u_n \| = 1$.

Finally, suppose that the conditions of Theorem 1 are satisfied. Then, by (8),

$$(\mu I - T)Tu_n = T(\mu I - T)u_n = (\mu - \lambda_n)Tu_n + \alpha_n Tw .$$

Since TB is contained in a (τ)-compact set, it follows that

$$(\mu I - T)Tu_n \to 0 \quad (\tau) ,$$

and (Tu_n) has a (τ)-cluster point v in K. Therefore, by the (τ)-continuity of T,

$$(\mu I - T)v = 0\ .$$

If $v = 0$, then 0 is a (τ_N)-cluster point of (Tu_n). But, by (8),

$$\mu u_n - Tu_n \to 0 \quad (\tau_N)\ ,$$

and so 0 is a (τ_N)-cluster point of (μu_n). Since $\mu \neq 0$ and $\| u_n \| = 1$, this is absurd. Hence $v \neq 0$, and the proof is complete.

It will be noticed that the preceding theorems and lemmas remain true if compactness is replaced by countable compactness, no change in the proofs being required. It may be of interest to remark that under the conditions of Theorem 2, K is a normal cone. However, since this fact is not needed for our main purpose, we omit its proof.

EXAMPLE 1. Taking $\tau = \tau_N$ in Theorem 1, we obtain Theorem C, and hence, as we have seen in [2], Theorem A also.

EXAMPLE 2. Suppose that there exists a subset A of K with the following properties:

(i) Given $x \in E$ with $\| x \| \leqq 1$, there exists $a \in A$ with $-a \leqq x \leqq a$.

(ii) TA is contained in a (τ_N)-compact set.[2]

Let E^* denote the usual dual space of continuous linear functionals on the normed space E, and let K^* denote the dual cone of all elements of E^* that are non-negative on K. Then K^* is a norm complete positive cone in E^*, and we denote by B^* the intersection of K^* with the closed unit ball in E^*.

For each φ in E^*, let $T^*\varphi$ be defined as usual by

$$(T^*\varphi)(x) = \varphi(Tx) \qquad (x \in E).$$

Since T is not necessarily a bounded operator in E, $T^*\varphi$ may fail to belong to E^*. However, $T^*K^* \subset K^*$, and T^* is a partially bounded operator in $K^* - K^*$. For, given $\varphi \in B^*$ and $x \in E$ with $\| x \| \leqq 1$, there exists $a \in A$ with $-a \leqq x \leqq a$, and therefore

$$-\varphi(Ta) \leqq \varphi(Tx) \leqq \varphi(Ta)\ .$$

Since TA is contained in a (τ_N)-compact set, the set $\{\| Ta \| : a \in A\}$ has a finite upper bound M and so $|\varphi(Tx)| \leqq M$, $\| T^*\varphi \| \leqq M$, $T^*B^* \subset MB^*$, T^* is partially bounded. It is easily seen that T^* is weak*-continuous in K^* and that K^* is weak*-closed.

We shall show that if the partial spectral radius μ^* of T^* is not zero, then Theorem 1 is applicable to the operator T^* in the space $K^* - K^*$ with the weak* topology as the auxiliary topology τ. This will prove the existence of a non-zero element ψ of K^* with

[2] In Examples 2, 3 no auxiliary topology is needed in E, but an auxiliary topology will appear in the dual space.

$$T^*\psi = \mu^*\psi .$$

Since T^* maps B^* into the weak*-compact set MB^*, we need only prove that the weak* topology is sequentially stronger than the norm topology at 0 relative to T^*B^*. To prove this, let $\varphi_n \in B^*$ $(n = 1, 2, \cdots)$, and suppose that 0 is a weak*-cluster point of the sequence $(T^*\varphi_n)$. Since TA is contained in a (τ_N)-compact set, given $\varepsilon > 0$, there exist $a_1, \cdots, a_r$ in A such that for each point a in A there is some k $(1 \leqq k \leqq r)$ with

$$\| Ta - Ta_k \| < \varepsilon/2 . \tag{9}$$

Since 0 is a weak*-cluster point of $(T^*\varphi_n)$, there exists an infinite set Λ of positive integers such that

$$| (T^*\varphi_n)(a_k) | < \varepsilon/2 \qquad (k = 1, \cdots, r;\ n \in \Lambda), \tag{10}$$

i.e.,

$$| \varphi_n(Ta_k) | < \varepsilon/2 \qquad (k = 1, \cdots, r;\ n \in \Lambda).$$

By (9) and (10), we have

$$| \varphi_n(Ta) | < \varepsilon \qquad (a \in A,\ n \in \Lambda). \tag{11}$$

Given $x \in E$ with $\| x \| \leqq 1$, there exists $a \in A$ with $-a \leqq x \leqq a$, and so, by (11),

$$| (T^*\varphi_n)(x) | = | \varphi_n(Tx) | \leqq \varphi_n(T_a) < \varepsilon \qquad (n \in \Lambda),$$

$$\| T^*\varphi_n \| \leqq \varepsilon \qquad (n \in \Lambda).$$

Therefore 0 is a norm-cluster point of $(T^*\varphi_n)$, and we have proved that Theorem 1 is applicable.

EXAMPLE 3. Suppose that there exists a subset A of K with the following properties:

(i) Given $x \in E$ with $\| x \| \leqq 1$, there exists $a \in A$ with $-a \leqq x \leqq a$.

(ii) A is contained in a (τ_N)-compact set.

Let K^*, B^*, T^* be defined as in Example 2. Given $\varphi \in B^*$ and $x \in E$ with $\| x \| \leqq 1$, there exists $a \in A$ with $-a \leqq x \leqq a$, and therefore

$$| \varphi(Tx) | \leqq \varphi(Ta) \leqq \| Ta \| \leqq p(T) \| a \| .$$

Since A is contained in a (τ_N)-compact set, $\| a \|$ is bounded on A, and T^* is a partially bounded mapping of K^* into itself.

We show that Theorem 2 is applicable to the operator T^*. Since K^* is weak*-closed, B^* is weak*-compact, and T^* is weak*-continuous in K^*, we need only prove that the weak* topology is sequentially stronger than the norm topology at 0 relative to B^*. This is proved by an argument similar to that in Example 2, but using A in place of TA.

It follows that there exists a non-zero element ψ of K^* with $T^*\psi = \mu^*\psi$, where μ^* is the partial spectral radius of T^*.

In particular, the conditions of this example are satisfied with A consisting of a single point if K contains an interior point in the normed space E. Thus Theorem B is contained in this example, and hence in Theorem 2.

EXAMPLE 4. Theorem 1 of Schaefer [6] is a case of our Theorem 2. In this case the topology τ is given, and Schaefer constructs a norm in $K - K$ in such a way that

$$\|x\| = f(x) \qquad (x \in K),$$

where f is a certain (τ)-continuous linear functional. Since f is (τ)-continuous, it is easily verified that τ is sequentially stronger than τ_N at 0 relative to B.

REFERENCES

1. F. F. Bonsall, *Endomorphisms of partially ordered vector spaces*, J. London Math. Soc., **30** (1955), 133-144.

2. ———, *Linear operators in complete positive cones*, Proc. London Math. Soc., (3) **8** (1958), 53-75.

3. M. G. Krein and M. A. Rutman, *Linear operators leaving invariant a cone in a Banach space* (Russian), Uspehi Mat. Nauk (N.S.), 3, No. 1, **23** (1948), 3-95, English translation: American Math. Soc., Translation 26.

4. M. Krein, *Sur les opérations linéaires transformant un certain ensemble conique en lui-même*, C.R. (Doklady) Acad. Sci. U.R.S.S. (N.S.), **23** (1939), 749-752.

5. H. Schaefer, *Halbgeordnete lokalkonvexe Vektorräume*. II, Math. Annalen **138** (1959), 259-286.

6. ———, *Some spectral properties of positive linear operators*, Pacific J. Math., **10** (1960), 1009-1019.

DURHAM UNIVERSITY
NEWCASTLE-ON-TYNE

SUMMABILITY OF DERIVED CONJUGATE SERIES

B. J. Boyer

1. Introduction. In a recent paper ([3] it was shown that the summability of the successively derived Fourier series of a *CP* integrable function could be characterized by that of the Fourier series of another *CP* integrable function. It is the purpose of the present paper to give analogous theorems for the successively derived conjugate series of a Fourier series.

2. Definitions. The terminology used in [3] will be continued in this paper. In addition let us define:

$$\psi(t) = \psi(t, r, x) = \frac{1}{2}[f(x+t) + (-1)^{r-1}f(x-t)] \tag{1}$$

$$Q(t) = \sum_{i=0}^{[\frac{r-1}{2}]} \frac{\bar{a}_{r-1-2i}}{(r-1-2i)!} t^{r-1-2i} \tag{2}$$

$$g(t) = r!t^{-r}[\psi(t) - Q(t)] \tag{3}$$

The rth derived conjugate series of the Fourier series of $f(t)$ at $t = x$ will be denoted by $D_rCFSf(x)$, and the nth mean of order (α, β) of $D_rCFSf(x)$ by $\bar{S}^r_{\alpha,\beta}(f, x, n)$.

3. Lemmas.

Lemma 1. *For* $\alpha = 0, \beta > 1$ *or* $\alpha > 0, \beta \geqq 0$, *and* $r \geqq 0$,

$$\bar{\lambda}^{(r)}_{1+\alpha,\beta}(x) = -\pi^{-1}r!(-x)^{r+1} + 0(|x|^{-1-\alpha}\log^{-\beta}|x|) + 0(|x|^{-r-2}) \text{ as } |x| \to \infty .$$

This is a result due to Bosanquet and Linfoot [2].

Lemma 2. *For* $\alpha > 0, \beta \geqq 0$ *or* $\alpha = 0, \beta > 0$ *and*

$$r \geqq 0,\ x^r\bar{\lambda}^{(r)}_{1+\alpha+r,\beta}(x) = \sum_{i,j=0}^{r} B^r_{ij}(\alpha, \beta)\bar{\lambda}_{1+\alpha+r-i,\beta+j}(x) ,$$

where the B^r_{ij} *are independent from* x *and have the properties*:

(i) $B^r_{ij}(\alpha, 0) = 0$ *for* $j \geqq 1$;

(ii) $B^r_{r0}(\alpha, \beta) \neq 0$;

(iii) $\sum_{i,j=0}^{r} B^r_{ij}(\alpha, \beta) = (-1)^r r!$.

Received January 19, 1960.

The proofs of (i) and (ii) will be found in [3], Lemma 2, taking the imaginary parts of the equations there. Part (iii) follows immediately from the first part of the lemma and Lemma 1.

LEMMA 3. *For* $n > 0, \alpha = 0, \beta > 1$ *or* $\alpha > 0, \beta \geqq 0,$ *and* $r \geqq 0,$

$$\left(\frac{d}{dt}\right)^r \left\{2B\pi^{-1}\sum_{\nu\leqq n}\left(1-\frac{\nu}{n}\right)^\alpha \log^{-\beta}\left(\frac{C}{1-\frac{\nu}{n}}\right)\sin \nu t\right\}$$

$$= 2n^{r+1}\sum_{k=-\infty}^{\infty}\bar{\lambda}^{(r)}_{1+\alpha,\beta}[n(t+2k\pi)] .$$

Proof. Smith ([6], Lemma 6) has shown that for every odd, Lebesgue integrable function, $Z(t)$, of period 2π,

$$\bar{S}_{\alpha,\beta}(Z, 0, n) = -2n\int_0^\infty Z(t)\bar{\lambda}_{1+\alpha,\beta}(nt)dt .$$

Since the right side of this equation can be written

$$-2n\int_0^\pi Z(t)\sum_{k=-\infty}^{\infty}\bar{\lambda}_{1+\alpha,\beta}[n(t+2k\pi)]dt$$

for every such $Z(t)$, the lemma is true for $r = 0$. For $r \geqq 1$ the interchange of $(d/dt)^r$ and $\sum_{-\infty}^{\infty}$ is justified by uniform convergence.

The following lemma is a direct consequence of Lemma 3:

LEMMA 4. *Let* $f(x) \in CP[-\pi, \pi]$ *and be of period* 2π. *For* $n > 0$ *and* $\alpha = 0, \beta > 1$ *or* $\alpha > 0, \beta \geqq 0,$

$$\bar{S}^r_{\alpha,\beta}(f, x, n) = 2(-n)^{r+1}\int_0^\pi \psi_r(t)\sum_{k=-\infty}^{\infty}\bar{\lambda}^{(r)}_{1+\alpha,\beta}[n(t+2k\pi)]dt .$$

LEMMA 5. *For* $\alpha \geqq 0, \beta \geqq 0, n > 0$ *and* $r \geqq 0,$

$$n^{r+1}\int_0^\infty Q(t)\bar{\lambda}^{(r)}_{1+\alpha+r,\beta}(nt)dt = 0 ,$$

where $Q(t)$ *is defined by* (2).

Proof. If $r = 0$, then $Q(t) = 0$. For $r \geqq 1$ and $i = 0, 1, \cdots [r-1/2]$, the truth of the lemma follows from the equation:

$$\int_0^\infty x^{r-1-2i}\bar{\lambda}^{(r)}_{1+\alpha+r,\beta}(x)dx = 0 ,$$

which is easily verified by means of $r - 1 - 2i$ integrations by parts.

The final two lemmas of this section give the appropriate representation of the nth mean of $D_rCFSf(x)$.

LEMMA 6. *Let $f(x) \in C_\lambda P[-\pi, \pi]$ and be of period 2π. Let m, $0 \leqq m \leqq \lambda + 1$, be an integer for which $\Psi_m(t) \in L[0, \pi]$. Then, for $\alpha = m$, $\beta > 1$ or $\alpha > m$, $\beta \geqq 0$ and $r \geqq 0$,*

$$\bar{S}^r_{\alpha+r,\beta}(f, x, n) = 2(-n)^{r+1}\int_0^\pi [\psi(t) - Q(t)]\bar{\lambda}^{(r)}_{1+\alpha+r,\beta}(nt)dt + C_r + o(1)$$

as $n \to \infty$, where

$$C_r = 2\pi^{-1}(-1)^{r+1}\int_0^\pi \psi(t)\left(\frac{d}{dt}\right)^r\left[\frac{1}{2}ctn\frac{1}{2}t - t^{-1}\right]dt \tag{4}$$
$$+ 2r!\pi^{-1}\int_\pi^\infty t^{-r-1}Q(t)dt .$$

Proof. It follows from Lemmas 4 and 5 that

$$\bar{S}^r_{\alpha+r,\beta}(f, x, n) = 2(-n)^{r+1}\int_0^\pi [\psi(t) - Q(t)]\bar{\lambda}^{(r)}_{1+\alpha+r,\beta}(nt)dt$$
$$+2(-n)^{r+1}\int_0^\pi \psi(t)\sum_{-\infty}^{\infty}{}'\bar{\lambda}^{(r)}_{1+\alpha+r,\beta}[n(t + 2k\pi)]\,dt \tag{5}$$
$$+ -2(-n)^{r+1}\int_\pi^\infty Q(t)\bar{\lambda}^{(r)}_{1+\alpha+r,\beta}(nt)dt$$
$$= I_1 + I_2 + I_3 .$$

Since the degree of $Q(t)$ is $r - 1$, Lemma 1 shows that

$$I_3 = 2r!\pi^{-1}\int_\pi^\infty t^{-r-1}Q(t)dt + o(1) . \tag{6}$$

Let us define:

$$J(n, t) = 2(-n)^{r+1}\sum_{-\infty}^{\infty}{}'\{\bar{\lambda}^{(r)}_{1+\alpha+r,\beta}[n(t + 2k\pi)]$$
$$- (-1)^r r!\pi^{-1}[n(t + 2k\pi)]^{-r-1}\} .$$

Again appealing to Lemma 1, we see that $\lim_{n\to\infty} (\partial/\partial t)^j J(n, t) = 0$ uniformly for $t \in [0, \pi]$ and $j = 0, 1, \cdots, m$.

With the aid of the well-known cotangent expansion I_2 may be written:

$$I_2 = \int_0^\pi \psi(t)J(n, t)dt + (-1)^{r+1}2\pi^{-1}\int_0^\pi \psi(t)\left(\frac{d}{dt}\right)^r \tag{7}$$
$$\left[\frac{1}{2}ctn\frac{1}{2}t - t^{-1}\right]dt .$$

But after m integrations by parts, it is seen that

$$\int_0^\pi \psi(t)J(n, t)dt = o(1) . \tag{8}$$

The lemma now follows from equations (5), (6), (7), and (8).

A particular, but useful, case of Lemma 6 is

LEMMA 7. *Let* $f(x) \in C_\lambda P[-\pi, \pi]$ *and be of period* 2π. *If* $g(t) \in C_\mu P[0, \pi]$, *where* $g(t)$ *is defined by* (3), *then*

$$\bar{S}_{\alpha,\beta}(g, 0, n) = -2n\int_0^\pi g(t)\bar{\lambda}_{1+\alpha,\beta}(nt)dt$$
$$-2\pi^{-1}\int_0^\pi g(t)\left(\frac{1}{2}ctn\frac{1}{2}t - t^{-1}\right)dt + o(1)$$

for $\alpha = 1 + \xi, \beta > 1$ *or* $\alpha > 1 + \xi, \beta \geqq 0$, *where* $\xi = \min[\mu, \max(r, \lambda)]$.

The hypothese of Lemma 6 are fulfilled, because $t^r g(t) \in C_\lambda P[0, \pi]$ implies $G_{1+\xi}(t) \in L[0, \pi]$ by Lemma 6 of [3].

4. Theorems.

THEOREM 1. *Let* $f(x) \in C_\lambda P[-\pi, \pi]$ *and be of period* 2π. *If there exist constants* $\bar{a}_{r-1-2i}$, $i = 0, 1, \cdots [r - 1/2]$, *such that*

(i) $g(t) \in C_\mu P[0, \pi]$ *for some integer* μ;

(ii) $CFSg(0) = s(\alpha, \beta)$ *for* $\alpha = 1 + \xi, \beta > 1$ *or* $\alpha > 1 + \xi, \beta \geqq 0$, *where* $\xi = \min[\mu, \max(r, \lambda)]$;

then $D_r CFSf(x) = S(\alpha + r, \beta)$, $s = \pi^{-1}\int_0^\pi g(t)ctn(1/2)tdt$ *and*

$$S = -2\pi^{-1}\int_0^\pi t^{-1}g(t)dt + C_r ,$$

where C_r *is defined by equation* (4).

THEOREM 2. *Let* $f(x) \in C_\lambda P[-\pi, \pi]$ *and be of period* 2π. *If* $D_r CFSf(x) = S(\alpha + r, \beta)$ *for* $\alpha = 1 + \lambda, \beta > 1$ *or* $\alpha > 1 + \lambda, \beta \geqq 0$, *then there exist constants* $\bar{a}_{r-1-2i}$, $i = 0, 1, \cdots [r - 1/2]$, *such that*

(i) $g(t) \in C_\mu P[0, \pi]$ *for some integer* μ:

(ii) $CFSg(0) = s(\alpha', \beta')$, *where*

$\alpha' = 1 + \xi, \beta' > 1$ *if* $1 + \lambda \leqq \alpha < 1 + \xi$ *or* $\alpha = 1 + \xi, \beta \leqq 1$ $\alpha' = \alpha$, $\beta' = \beta$ *if* $\alpha = 1 + \xi, \beta > 1$ *or* $\alpha > 1 + \xi, \beta \geqq 0$, *and* ξ, s *and* S *have the values given in Theorem* 1.

Before passing to the proofs of these theorems, let us observe that the existence of the constants $\bar{a}_{r-1-2i}$ implies their uniqueness from the definition of $g(t)$. In fact, it can be shown that the $\bar{a}_{r-1-2i}$ are given by

$$D_{r-1-2i}FSf(x) = \bar{a}_{r-1-2i}(C), \qquad i = 0, 1, \cdots \left[\frac{r-1}{2}\right].$$ [1]

[1] Bosanquet ([1], Theorem 1) has shown this for $f(x)$ Lebesgue integrable and (C) replaced by Abel summability.

In addition it can be shown that when $f(x) \in L$, the sum, S, of $D_r CFSf(x)$ may be written

$$S = -2\pi^{-1}\int_{\to 0\,(C)}^{\infty} t^{-1}g(t)dt \ .$$ [2]

Proof of Theorem 1. That $s = -\pi^{-1}\int_0^\pi g(t)ctn(1/2)tdt$ follows from the consistency of (α, β) summability and a result due to Sargent ([4], Theorem 3). Therefore, both $g(t)ctn(1/2)t$ and $t^{-1}g(t)$ are CP integrable over $[0, \pi]$.

From Lemma 7 we have

$$(9) \qquad \bar{S}_{\alpha,\beta}(g, 0, n) - s = -2n\int_0^\pi g(t)[\bar{\lambda}_{1+\alpha,\beta}(nt) - (\pi nt)^{-1}]dt + o(1) \ .$$

The left side of (9) is $o(1)$ by hypothesis. By consistency equation (9) remains valid if α is replaced by $\alpha + r - i$ and β by $\beta + j, i, j = 0, 1, \cdots r$. Therefore,

$$-2n\int_0^\pi g(t)\sum_{i,j=0}^{r} B_{ij}^r(\alpha, \beta)[\bar{\lambda}_{1+\alpha+r-i,\beta+j}(nt) - (\pi nt)^{-1}]dt = o(1) \ .$$

With the aid of Lemmas 2 and 6, the last equation becomes

$$\bar{S}_{\alpha+r,\beta}^r(f, x, n) = -2\pi^{-1}\int_0^\pi t^{-1}g(t)dt + C_r + o(1) \ .$$

This completes the proof of Theorem 1.

Proof of Theorem 2. Due to the length of this proof and its similarity to the proof of Theorem 2, ([3]), only a brief outline of the proof will be given.

Putting $Q(t) = 0, \beta = 0$ and $p > \alpha + r$ in Lemma 6 and integrating the right-hand side of the resulting equation $\lambda + 1$ times, one can show that

$$D_{r+\lambda+1}CFS(\Psi_{\lambda+1}, 0, n) \text{ is summable } (C, p) \ .$$

A result due to Bosanquet ([1], Theorem 1) and the stepwise procedure employed in the proof of Theorem 2 ([3], equations 18 through 22) lead to the conclusion: $t^{-r-1}[\psi_r(t) - Q(t)] \in CP[0, \pi]$ for an appropriate polynomial $Q(t)$, i.e., $t^{-1}g(t) \in CP[0, \pi]$. From this statement and a results due to Sargent ([4], Theorem 3), $g(t) \in C_\mu P[0, \pi]$ for some integer μ and $CFSg(0) = s(C)$, where $s = \pi^{-1}\int_0^\pi g(t)ctn(1/2)tdt$.[3]

[2] Ibid. The difference in sign is due to the distinction between allied and conjugate series.

[3] The CP integrability of $g(t)ctn(1/2)t$ is equivalent to that of $t^{-1}g(t)$.

That S, the $(\alpha+r, \beta)$ sum of $D_r CFSf(x)$, has the value

$$-2\pi^{-1}\int_0^\pi t^{-1}g(t)dt + C_r$$

follows immediately from Theorem 1 and the consistency of the summability scale.

Thus, it remains to prove only the order relations (α', β') in (ii) of the theorem. A straightforward calculation using the representations in Lemmas 6 and 7, the properties of the $B^r_{ij}(\alpha, \beta)$ in Lemma 2, and the consistency of the summability scale applied to $D_r CFSf(x)$, leads to the following equations:

$$\sum_{i,j=0}^{r} B^r_{ij}(\alpha'+k, \beta')\Big[\bar{S}_{\alpha'+k+r-i,\beta'+j}(g, 0, n) - \pi^{-1}\int_0^\pi g(t)ctn\frac{1}{2}tdt\Big] = o(1)\ ,$$

for $k = 0, 1, 2, \cdots$.

The expression in brackets may be considered the nth mean of order $(\alpha'+k+r-i, \beta'+j)$ of a series formed from $CFSg(0)$ by altering the first term. Since this series is summable (C) to 0, then Lemma 8 [3] shows that $CFSg(0) = s(\alpha', \beta')$.

The following theorem gives a sufficient condition for the (α, β) summability of $CFSg(0)$ for $\beta \neq 0$. Since the proof follows the usual lines for Riesz summability, it is omitted.

THEOREM 3. *Let $g(t)$ be an odd function of period 2π. If $t^{-1}g(t) \in C_k P[0, \pi]$, where k is a non-negative integer, then*

$$CFSg(0) = -\pi^{-1}\int_0^\pi g(t)ctn\frac{1}{2}tdt(1+k, \beta), \beta > 1\ .$$

As an application of these theorems it can be shown that

$$D_r CFSf(0, m) = S(1+m+2r, \beta), \beta > 1\ ,$$

where $f(x; m)$ is either $x^{-m}\sin x^{-1}$ or $x^{-m}\cos x^{-1}$, $m = 0, 1, 2, \cdots$.

The following results may be deduced from Theorems 1 and 2. It is assumed that $f(x) \in C_\lambda P[-\pi, \pi]$ and is of period 2π. The values of S and s, when either exists, and ξ are given in Theorem 1.

(A). If $g(t) \in C_\mu P[0, \pi]$, then for $\alpha = 1+\xi, \beta > 1$ or $\alpha > 1+\xi$, $\beta \geqq 0$, $D_r CFSf(x) = S(\alpha+r, \beta)$ if and only if $CFSg(0) = s(\alpha, \beta)$.

(B). For $\alpha = 1+\max(r, \lambda), \beta > 1$ or $\alpha > 1+\max(r, \lambda), \beta \geqq 0$, $D_r CFSf(x) = S(\alpha+r, \beta)$ if and only if $g(t) \in CP[0, \pi]$ and $CFSg(0) = s(\alpha, \beta)$.

These results generalize, to various degrees, results obtained by Takahashi and Wang [7] and Bosanquet [1].

A weak, but none the less interesting, form of these results is

(C). If $f(x) \in CP[-\pi, \pi]$ and is of period 2π, then in order that $D_r CFSf(x)$ be summable (C), it is necessary and sufficient that $g(t) \in CP[0, \pi]$ and $CFSg(0)$ be summable (C).

References

1. L. S. Bosanquet, *On the derived allied series of a Fourier series*, Proc. London Math. Soc. (2), **49** (1947), 63-76.
2. L. S. Bosanquet and E. H. Linfoot, *Generalized means and the summability of Fourier series*, Quart. Journal of Math. (Oxford), **2** (1931), 207-229.
3. B. J. Boyer, *On the summability of derived Fourier series*, Pacific J. Math. (To be published.)
4. W. L. C. Sargent, *On the summability* (*C*) *of allied series*, Proc. London Math. Soc., **50** (1949), 330-348.
5. A. H. Smith, *On the summability of derived conjugate series of the Fourier-Lebesgue type*, Bull. Amer. Math. Soc., **40** (1934), 406-412.
6. ———, *Summability of conjugate Fourier series*, Duke Math. Journal. **4** (1938), 270-276.
7. T. Takahashi and F. T. Wang, *Some notes on trigonometric series*, Tohoku Math. Journal, **41** (1935), 169-187.

Florida State University

A NOTE ON A PROBLEM OF FUCHS

D. L. BOYER

In [1] Fuchs has asked (problem 3) the cardinality of the set of all pure subgroups of an Abelian group. The purpose of this paper is to settle the question for nondenumerable Abelian groups. $|A|$ will denote the cardinality of the set A.

THEOREM. *Let G be a nondenumerable Abelian group, and let $\mathscr{P}$ be the collection of all pure subgroups, P, of G with $|P| = |G|$. Then $|\mathscr{P}| = 2^{|G|}$.*

Proof. Let T be the torsion subgroup of G. If $|T| < |G|$, then $|G/T| = |G|$ and by a result of Walker [3, Theorem 4], G/T, and hence G, has $2^{|G|}$ pure subgroups of order $|G|$.

If $|T| = |G|$, then we write T in the form $T = \sum_{i,\alpha} \oplus Z_\alpha(p_i^\infty) \oplus \sum_p \oplus R_p$, where the R_p are reduced primary groups and $\sum_{i,\alpha} \oplus Z_\alpha(p_i^\infty)$ is the maximal divisible subgroup of T.

If the above decomposition of T has $|G|$ summands then the theorem follows.

If the above decomposition has fewer than $|G|$ summands, then $|\sum_p \oplus R_p| = |G|$.

We first consider the case that there exists a prime, p, such that $|R_p| = |G|$. Let B be a basic subgroup of R_p. If $|B| < |R_p|$, then $|R_p/B| = |G|$ and $R_p/B = \sum_{\alpha \in A} \oplus Z_\alpha(p^\infty)$ with $|A| = |G|$. Thus the theorem holds for R_p/B, and hence also for G. If $|B| = |R_p|$, then since B is the direct sum of cyclic groups, $B = \sum_{\alpha \in A} \oplus C_\alpha$, it follows that $|A| = |G|$. Thus the theorem follows for B and hence for G. Finally, if $|R_p| < |G|$ for all p, we let[1] $R' = \sum_{p_i} \oplus R_{p_i}$, where the sum is taken over all primes, p_i, such that $|R_{p_i}| > \aleph_0$. Then $|R'| = |G| = \sum |R_{p_i}|$. We have proved above that for each p_i, R_{p_i} has $2^{|R_{p_i}|}$ pure subgroups, $P(i)$ of order $|R_{p_i}|$. For each i, choose $P(i) \subset R_{p_i}$ with $|P(i)| = |R_{p_i}|$. Then $P = \sum \oplus P(i)$ is a pure subgroup of R' with $|P| = |R'|$, and the number of subgroups formed in this way is $2^{|G|}$.

REFERENCES

1. L. Fuchs, *Abelian groups*, Hungarian Academy of Science (1958), Budapest.
2. W. Scott, *Groups and cardinal numbers*, Amer. J. Math., **74** (1952), 187–197.
3. E. Walker, *Subdirect sums and infinite Abelian groups*, Pacific J. Math., **9** (1959), 287–291.

FRESNO STATE COLLEGE

Received October 22, 1959.

[1] This is exactly the method used by Scott, [2].

THE ENVELOPES OF HOLOMORPHY OF TUBE DOMAINS IN INFINITE DIMENSIONAL BANACH SPACES

H. J. BREMERMANN

1. Introduction. Let B be a Banach space with the strong topology generated by the norm. An open and connected set is called a *domain*. Let f be a complex valued functional defined in a domain D of a complex Banach space B_c. Let L be a finite dimensional translated complex linear subspace of $B_c : L = \{z \mid z = z_0 + \tau_1 a_1 + \cdots \tau_n a_n\}$ where $z_0, a_1, \cdots, a_n$ are fixed elements $\tau_1, \cdots, \tau_n$ complex parameters. (In the following we will call L an "affine subspace"). f is called "G-holomorphic" (=Gâteaux-holomorphic) if and only if the restriction of f to the intersection $D \cap L$ of D with any finite dimensional affine subspace L of B_c is holomorphic (in the ordinary sense). (Compare Hille-Phillips [7], Soeder [9], Bremermann [5].)

A functional that is G-holomorphic and locally bounded is called "F-holomorphic" (Fréchet-holomorphic). For finite dimension the notions (ordinary) "holomorphic function" and "G- and F-holomorphic functional" coincide. (The theory of holomorphic functionals in finite dimensional Banach spaces is equivalent to the theory of n complex variables.) For infinite dimension, in general, there exist already linear (and hence G-holomorphic) functionals that are not locally bounded (and hence not F-holomorphic).

In Bremermann [5] it has been shown that the phenomenon of "simultaneous holomorphic continuation," well known for n complex variables, persists for infinite dimension even for the very general G-holomorphic functionals: There exist domains such that all G-holomorphic functionals can be continued into a larger domain.

A domain for which a G-holomorphic functional exists that cannot be continued is called (in analogy to finite dimension) a "domain of G-holomorphy." In Bremermann [5] it has been shown that a domain of G-holomorphy is "pseudo-convex" (in a sense which is a natural extension from finite dimension).

We will apply these notions in the following to infinite dimensional tube domains and moreover we will show that it is possible to define and to determine the envelope of holomorphy of tube domains.

Finite dimensional tube domains and their envelopes of holomorphy have been studied by Bochner [1], Bochner-Martin [2], Hitotumatu [8], and Bremermann [3], [4]. It has been shown that a tube domain is pseudo-convex if and only if it is convex, and that the envelope of

Received January 11, 1960. This research has in part been supported by the Office of Naval Research under Contract Nonr 447 (17).

holomorphy of any tube domain is its convex envelope. The former property has been extended to infinite dimension in [5]. We extend here the latter. *To the author's knowledge this is the first time that the envelope of holomorphy of a class of infinite dimensional domains has been determined.* At the same time the proof given in the following is simpler than some previous proofs for finite dimension.

2. Tube domains, envelopes of holomorphy. Let B_c be a complex Banach space that is split into a real and imaginary part, such that every $z \in B_c$ is written

$$z = x + iy, \text{ where } x \in B_r,\ y \in B_r ,$$

where B_r is a real Banach space. Then a *domain* T_X *is called a tube domain with basis* X *if and only if it is of the form* $T_X = \{z \mid x \in X, y$ arbitrary$\}$, *where* X *is a domain in* B_r.

Obviously, T_X is convex if and only if X is convex, and X is convex if and only if the intersection of X with every finite dimensional affine subspace L_r of B_r is convex. ($L_r = \{x \mid x = x_0 + t_1 a_1 + \cdots, t_n a_n\}$, where $x_0, a_1, \cdots, a_n$ are fixed elements in B_r, and $t_1, \cdots, t_n$ real parameters).

It is somewhat difficult to define the envelope of holomorphy for arbitrary domains. Already for finite dimension it may not be schlicht. (Comp. [3], [6]). However, for finite dimension the following is true. Let D be a given domain. Suppose we have a domain $E(D)$ with the following properties:

(I) Every function holomorphic in D can be continued as a (single-valued) holomorphic function to $E(D)$.

(II) To every finite boundary point z_0 of $E(D)$ there exists a function that is holomorphic throughout $E(D)$ and is singular at z_0. If $E(D)$ has these properties, then $E(D)$ is the envelope of holomorphy of D.

Analogously, if we have an infinite dimensional domain D and a domain $E(D)$ with the properties (I) and (II) (with respect to G-holomorphic functionals), then we call $E(D)$ the *envelope of G-holomorphy of D.*

3. Proof of the main theorem. Let T_X be a tube domain that is not convex. Then, there exists an affine subspace

$$L_r = \{x \mid x = x_0 + t_1 a_1 + \cdots t_n a_n\}$$

($x_0, a_1, \cdots, a_n \in B_r, t_1, \cdots, t_n$ real parameters) such that $X \cap L_r$ is not convex.

Now it would be possible that $X \cap L_r$ is not connected and each connected component is convex (for instance if L_r is one-dimensional).

If X is not convex, then there exist two points x_1 and x_2 that cannot be connected by a straight line segment in X. However, X is connected, and even arcwise connected. Hence we can connect x_1 and x_2 by an arc in X, and even by a "polygon" that is by finitely many straight line segments. The polygonal arc spans a finite dimensional affine subspace L_r and the connected component of $L_r \cap X$ that contains x_1 and x_2 is not convex since x_1 and x_2 cannot be connected by a straight line.

Thus $L_r \cap X$ has a connected component that is not convex. Hence there exists a point x_3 on the boundary of $L_r \cap X$ and a line segment s containing x_3 such that s is locally a supporting line segment of the complement of $L_r \cap X$. In particular, x_3 and s can be chosen such that in a neighborhood of x_3 the line segment s has with the boundary $\partial(X \cap L_r)$ only the point x_3 in common.

Let the equation of the line containing s be

$$s = \{x \mid x = x_3 + bt\} ,$$

where b is a fixed element in B_r, t a real parameter. Let b be normalized such that $\|b\| = 1$. This real line lies in the analytic plane:

$$A = \{z \mid z = x_3 + b\tau\} ,$$

where τ is a complex parameter.

Let S_ρ be a disc on A with center at x_3, radius ρ:

$$S_\rho = \{z \mid z = x_3 + b\tau, |\tau| < \rho\} .$$

If ρ is small enough, then S_ρ will lie entirely in T_X, except for the points

$$\{z \mid z = x_3 + ibt, |t| < \rho, t \text{ real}\} .$$

We now apply the following lemma (which is an immediate consequence of the "fundamental Lemma" 3.1 (and 3.2) of [5] and Theorem 6.3 of [6]).

To formulate the lemma we need the distance function $d_D(z)$ which is defined as follows: Given a domain D, then

$$d_D(z) = \sup r \ni \{z' \mid \|z - z'\| < r\} \subset D ,$$

in other words $d_D(z)$ is the distance of the points z from the boundary of D, measured in the norm of B_c.

LEMMA. *Let $h(z)$ be the solution of the boundary value problem*

$$h(\tau) = \log d_{T_X}(x_3 + b\tau) \text{ for } |\tau| = \rho ,$$
$$h(\tau) \text{ harmonic for } |\tau| < \rho .$$

Then any function that is G-holomorphic in T_X can be continued G-holomorphically into the point set:

$$C = \{z \mid z' = x_3 + \tau b, |\tau| < \rho, ||z - z'|| < e^{h(\tau)}\} .$$

(We note that even though $\log d_{T_X}(x)$ becomes infinite at the two points $z = x_3 \pm i\rho b$, the solution of the boundary value problem exists and is finite for all $|\tau| < \rho$).

The pointset C is a neighborhood of the point $z = x_3$. In particular it contains the points $||z - x_3|| < e^{h(0)}$, and $e^{h(0)} \neq 0$. This continuation procedure can be repeated at any point $z = x_3 + iy$, where y is arbitrary. We always get the same neighborhood, independently of y, because the function $d_{T_X}(x_3 + iy)$ and hence h does not depend upon y. Hence any function G-holomorphic in T_X can not only be continued into a larger domain but into a larger tube domain $T_{X'}$, that means $X \subset X'$, $X \neq X'$.

We have to observe however one difficulty: If the intersection $X \cap \{X \mid ||x - x_3|| < e^{h(0)}\}$ consists of more than one component, then continuation into $T_{X'}$ with $X' = X \cup \{x \mid ||x - x_3|| < e^{h(0)}\}$ could possibly be such that the continued function would no longer be single-valued in $T_{X'}$. In order to keep the continuation single-valued we remove from X' all components of $X \cap \{x \mid ||x - x_3|| < e^{h(0)}\}$ except the one that intersects S_ρ. In this way the continuation remains single-valued.

Thus we have the result: If T_X is a tube domain such that X is not convex, then any function that is G-holomorphic can be continued G-holomorphically (and single-valued) into a larger tube domain with basis X'. Then we can apply the same result to $T_{X'}$, and obviously the process can be iterated as long as the enlarged tube is not yet convex. Thus we have proved:

Given a tube domain T_X, then any function that is G-holomorphic in T_X can be continued G-holomorphically into the convex envelope of T_X.

(The convex envelope of T_X equals $T_{C(X)}$, where $C(X)$ is the convex envelope of X.)

On the other hand there exists to every boundary point z_0 of $T_{C(X)}$ a supporting affine subspace of B_c and a linear functional $l(z)$ that becomes zero exactly on the affine subspace. (This is an immediate consequence of the Hahn-Banach theorem.) The functional $1/l(z)$ is then G-holomorphic in $T_{C(X)}$ and becomes singular at z_0. Hence we have shown:

To every boundary point z_0 of a *convex* tube domain there exists a functional that is G-holomorphic in the domain and singular at z_0. The two statements combined give:

THEOREM. *Let T_X be a tube domain in a complex Banach space (of arbitrary dimension). Then the envelope of G-holomorphy of T_X is the convex envelope of T_X, which equals $T_{C(X)}$, where $C(X)$ is the convex envelope of X.*

Bibliography

1. S. Bochner, *A theorem on analytic continuation of functions in several variables*, Annals of Math., **39**, (1938), 14–19.
2. S. Bochner and W. T. Martin, *Several Complex Variables*, Princeton, 1948.
3. H. J. Bremermann, *Die Holomorphiehüllen der Tuben-und Halbtubengebiete*, Math. Annalen, **127** (1954), 406–423.
4. ———, *Complex convexity*, Trans. Amer. Math. Soc., **82** (1956), 17–51.
5. ———, *Holomorphic functionals and complex convexity in Banach spaces*, Pacific J. Math., **7** (1957), 811–831. (Errata in vol. 7 at the end).
6. ———, *Construction of the envelopes of holomorphy of arbitrary domains*, Revista Mat. Hisp. Amer. **17** (1957), 1–26.
7. E. Hille and R. S. Phillips, *Functional analysis and semi-groups*, Amer. Math. Soc. Publ., Revised Ed., New York, 1957.
8. S. Hitotumatu, *Note on the envelope of regularity of a tube domain*, Proc. Japan. Acad. **26** (1950), 21–25.
9. H. Soeder, *Beiträge zur Funktionentheorie in Banachschen Räumen*, Schriftenreihe Math. Inst. Univ. Münster, no. 9, 1956.

MINIMAL SUPERADDITIVE EXTENSIONS OF SUPERADDITIVE FUNCTIONS

ANDREW BRUCKNER

Introduction. A real valued function f is said to be superadditive on an inverval $I = [0, a]$ if it satisfies the inequality $f(x + y) \geqq f(x) + f(y)$ whenever x, y and $x + y$ are in I. Such functions have been studied in detail by E. Hille and R. Phillips [1] and R. A. Rosenbaum [2]. In this paper we show that any superadditive function f on I has a minimal superadditive extension F to the non-negative real line E, and then proceed to show that F inherits much of its behavior from the behavior of f. We deal primarily with superadditive functions which are continuous and non-negative.

A simple example of a superadditive function on $[0, a]$ is furnished by a convex function f with $f(0) \leqq 0$. Also, if f is convex and $f(0) = 0$, then it is easy to verify that its minimal superadditive extension F is given by

$$F(x) = nf(a) + f(x - na)$$

for $na \leqq x < (n + 1)a$. In general, the minimal superadditive extension F is not easily computed. In the sequel we shall discuss two methods for obtaining F. For brevity we shall use the notation f^*F to mean "F is the minimal superadditive extension of f".

1. The decomposition method. DEFINITION. Let $x \in E$. The numbers $x^1, \cdots, x^n$ are said to form an *a-partition* for x if $x^1 + \cdots + x^n = x$ and for each $i = 1, \cdots, n$ we have $0 \leqq x^i \leqq a$.

THEOREM 1. *Let f be a superadditive function on $I = [0, a]$. Then the function F defined on E by the equation*

$$F(x) = \sup \Sigma f(u^i) ,$$

the supremum being taken over all a-partitions of x, is the minimal superadditive extension of f.

Proof. We will show that F is superadditive. The minimality of F will then follow from the fact that any superadditive extension $\hat{f}$ of f must satisfy $\hat{f}(x) \geqq \Sigma f(x^i)$ for all $x \in E$ and all a-partitions $x^1, \cdots, x^n$ of x. Let $x, y \in E$, $\varepsilon > 0$. Choose a-partitions $x^1, \cdots, x^m$ and $y^1, \cdots, y^n$ for

Received November 6, 1959. This paper is part of the author's doctoral thesis, and the author is indebted to Professor John Green for his guidance in its preparation. Thanks are also due the National Science Foundation for their support.

x and y respectively such that $f(x^1) + \cdots + f(x^m) \geqq F(x) - \varepsilon/2$ and $f(y^1) + \cdots + f(y^n) \geqq F(y) - \varepsilon/2$. Then the numbers $x^1, \cdots, x^n, y^1, \cdots, y^n$ form an a-partition for $x + y$ and we have

$$F(x + y) \geqq f(x^1) + \cdots + f(x^m) + f(y^1) + \cdots + f(y^n)$$
$$\geqq F(x) + F(y) - \varepsilon .$$

Suppose we have an approximation for $F(x)$: that is, a number $\varepsilon > 0$ and an a-partition $x^1, \cdots, x^n$ for x such that $F(x) - \Sigma f(x^i) < \varepsilon$. If among the members of this a-partition there are two, say x^j and x^k such that $u = x^j + x^k \leqq a$, then since $f(u) \geqq f(x^j) + f(x^k)$, we have

$$F(x) - [f(u) + \sum_{i \neq j,k} f(x^i)] \leqq F(x) - \sum_1^n f(x^i) < \varepsilon .$$

In other words, replacing two numbers used in the approximation by their sum $u \leqq a$ yields an approximation at least as good as the original. It follows that if x satisfies the inequality $(M - 2)a/2 \leqq x \leqq (M - 1)a/2$, where M is a positive integer, then there exist arbitrarily good approximations for $F(x)$ using only M terms in the a-partition. If f is continuous, then a simple compactness argument results in the following theorem:

THEOREM 2. *Let f be a continuous superadditive function on $[0, a]$, and let F be its minimal superadditive extension. Let x satisfy the inequality $(M - 2)a/2 \leqq x \leqq (M - 1)a/2$. Then $\exists$ an a-partition $x^1, \cdots, x^M$ for x such that*

$$\Sigma f(x^i) = F(x) .$$

Such an a-partition for x will be called a *decomposition* of x, for which we shall use the notation $\langle x \rangle$ whenever convenient. We will denote by $N(x)$ a number so large that for any continuous superadditive function on $[0, a]$, $\exists$ a decomposition $\langle x \rangle$ of x with at most $N(x)$ members. It follows from the above that we can always let $N(x) = 2x/a + 2$, for example.

Henceforth we shall be concerned primarily with continuous non-negative superadditive functions for which we shall use the abbreviation *csa*. It is readily verified that such functions are non-decreasing and vanish at the origin.

2. Combinations of extensions. One might expect that if the members of a family f of *csa* functions are combined in a linear fashion to give another *csa* function h, then combining the members of the family $\tilde{f}$ of minimal superadditive extensions of functions in f in the same way would give rise to a function H which is the minimal superadditive extension of h. However this is not always the case. Consider, for example, the functions f and g defined on $[0, 3]$ as follows: $f(0) = 0$, $f(1) = 0, f(2) = 0, f(3) = 1$ and $g(0) = 0$, $g(1) = 0$, $g(2) = 2$, $g(3) = 3$, f and g linear on $[n, n+1]$, $n = 0, 1, 2$. Simple computations show that whereas $(F + G)(4) = 5$ and $FG(4) = 4$, the minimal superadditive extensions of $f + g$ and fg take on the values 4 and 3 respectively at $x = 4$. The minimal superadditive extension of a sum (product) of superadditive functions is thus not necessarily the sum (product) of the minimal superadditive extensions. However, some processes do commute with taking minimal superadditive extensions.

THEOREM 3. *Let $\{f_n\}$ be a sequence of csa functions converging to the continuous function f on $I = [0, a]$. Let $f_n{}^*F_n$. Then f is csa and $f^* \lim_{n\to\infty} F_n$.*

Proof. That f is superadditive and non-negative is clear. Since for each positive integer n the function f_n is non-decreasing, the convergence of $\{f_n\}$ to f is uniform on I. Given $\varepsilon > 0$ and $x \in E$, let M be such that $n \geqq M \Rightarrow \max_{t\in I} |f_n(t) - f(t)| < \varepsilon/N(x)$ where $N(x)$ is a number chosen as in § 1. Let $k > M$ and let $\langle x^k\rangle \equiv x_k^1, \cdots, x_k^{N(x)}$ and $\langle x\rangle \equiv x^1, \cdots, x^{N(x)}$ be decompositions for x relative to F_k and F respectively. We have

$$F(x) = \sum_{i=1}^{N(x)} f(x^i) \geqq \sum_{i=1}^{N(x)} f(x_k^i)$$

and

$$F_k(x) = \sum_{i=1}^{N(x)} f_k(x_k^i) \geqq \sum_{i=1}^{N(x)} f_k(x^i) \;.$$

It follows from these two inequalities that

$$F(x) - F_k(x) \,| < \varepsilon \;,$$

for $n \geqq M$.

3. **Behavior of the minimal superadditive extension.** It seems reasonable to expect that the minimal superadditive extension F of a *csa* function f will enjoy many of the properties of f. To a certain extent this is true and we are able to predict much about the behavior of F by examining the behavior of f.

THEOREM 4. *Let f be csa on $[0, a]$. If f^*F, then F is csa on E.*

Proof. Clearly F is non-negative. To prove that F is continuous let $\varepsilon > 0$ and choose $\delta < a/2 \ni$ if $u, v \leqq a$ and $|u - v| < \delta$ then $|f(u) - f(v)| < \varepsilon$. Now let x and y be points of E for which $|y - x| < \delta$, say $y = x + h$. Let $\langle y \rangle = y^1, \cdots, y^N$ be a decomposition for y with, say, $y^1, \geqq a/2$. We have

$$F(y) = \sum_1^N f(y_i) \text{ and } F(x) \geqq \sum_2^N f(y^i) + f(y^1 - h) .$$

Hence $0 \leqq F(y) - F(x) \leqq f(y^1) - f(y^1 - h) < \varepsilon$.

In a similar manner one can establish the following theorem, which is stated without proof.

THEOREM 5. *Let f be csa on $[0, a]$. If f^*F, then the following statements hold*:

(a) *If f satisfies a Lipschitz condition with coefficient M, then so does F*;

(b) *If $\langle y \rangle = y^1, \cdots, y^M$ is a decomposition for y and f is differentiable at y^i and y^j, then $f'(y^i) = f'(y^j)$. If, in addition, F is differentiable at y, then $F'(y) = f'(y^i)$.*

One might expect that the differentiability of f on $[0, a]$ would imply the differentiability of F, except possibly at integral multiples of a. Although this turns out not to be the case, we do have the following theorem:

THEOREM 6. *Let f be a csa function on the interval $[0, a]$, with f' continuous on $(0, a)$. For x not an integral multiple of a, let X be the set of points of $[0, a]$ which can be used in a decomposition for x. Then F has a right hand derivative $F_+(x)$ and a left hand derivative $F_-(x)$ at x with*

$$F_+(x) = \sup_{t \in X} f'(t) \equiv S$$

and

$$F_-(x) = \inf_{t \in X} f'(t) \equiv I .$$

Proof. We will prove only the upper equality. The lower can be proved in a similar manner. It suffices to show $D^+F(x) = D_+F(x) = S$ where D^+F and D_+F are the upper and lower right hand derivatives of F. Suppose $\exists \varepsilon > 0 \ni D^+F(x) > S + 2\varepsilon$. Then a sequence $\{h_i\}$ of numbers approaching 0 such that

$$F(x) < F(x + h_i) - (S + \varepsilon)h_i \tag{1}$$

for $i = 1, 2, \cdots$. For each positive integer i, let $(u^i, v^i, \cdots, w^i)$ be a decomposition for $x + h_i$. Without loss of generality, we assume that the sequence $(u^i, v^i, \cdots, w^i)$ converges to, say, $(u, v, \cdots, w)$; otherwise we consider a convergent subsequence. Since x is not an integral multiple of a, one of the numbers $u, v, \cdots, w$ is not equal to 0 or a. Denote such a one by u. From (1) we have

(2) $$F(x) < f(u^i) + f(v^i) + \cdots + f(w^i) - (S + \varepsilon)h_i .$$

Choose $N_1 \ni i > N_1$ implies that

(3) $$f(u^i) < f(u^i - h_i) + [f'(u^i - h_i) + \varepsilon/2]h_i .$$

That N_1 can be so chosen follows from the continuity of f'. In fact, let δ be such that $|u - v| < \delta \Rightarrow |f'(u) - f'(v)| < \varepsilon/4$. Now choose N_1 such that $i > N_1 \Rightarrow u - \delta < u^i - h_i < u^i < u + \delta$. If $y \in [u^i - h_i, u^i]$, with $i > N_1$, then $f'(u^i - h_i) + \varepsilon/2 > f'(y)$. Hence (3) is a valid inequality. For $i > N_1$ we have from (2) and (3),

(4) $$F(x) < f(u^i - h_i) + f(v^i) + \cdots + f(w^i) + [f'(u^i - h_i) - (S + \varepsilon/2)]h_i.$$

Now the sequence $(u^i - h_i, v^i, \cdots, w^i)$ converges to $(u, v, \cdots, w)$ and $u + v + \cdots + w = x$. Thus, since

$$f(u^i) + f(v^i) + \cdots + f(w^i) = F(x + h_i) \geq F(x) ,$$

and F is a superadditive function, we have

$$f(u) + f(v) + \cdots + f(w) = F(x)$$

and $u \in X$. Therefore $f'(u) \leqq S$. By the continuity of f', $\lim_{i\to\infty} f'(u^i - h_i) = f'(u)$. Hence $\exists$ a positive number N_2 such that $i > N_2 \Rightarrow f'(u^i - h_i) < S + \varepsilon/2$. Let $i = \max(N_1, N_2)$. For this i we have from (4),

$$F(x) < f(u^i - h_i) + f(v^i) + \cdots + f(w^i).$$

This is impossible, for $u^i - h_i + v^i + \cdots + w^i = x$ for each $i = 1, 2, \cdots$ and F is superadditive. We have shown $D^+F(x) \leqq S$.

It remains to show $D_+F(x) \geqq S$. Let $\varepsilon > 0$, and let $(u, v, \cdots, w)$ be a decomposition for x such that $u \neq a$, and $f'(u) > S - \varepsilon/4$. Choose $\delta > 0 \ni h < \delta \Rightarrow f(u + h) > f(u) + (S - \varepsilon/2)h_i$. For $h < \delta$,

$$F(x + h) \geq f(u + h) + f(v) + \cdots + f(w) > F(x) + (S - \varepsilon/2)h .$$

The first and third members of this inequality give

$$\frac{F(x + h) - F(x)}{h} > S + \varepsilon/2 .$$

Since ε was arbitrary, $D_+F(x) \geqq S$, and the proof of the theorem is complete.

We now proceed to obtain a linear upper bound for F. If f is *csa* on $[0, a]$, then the function g defined by $g(x) = f(x)/x$ is continuous on $[0, a]$ and satisfies $g(nx) \geqq g(x)$, $n = 1, 2, \cdots$, whenever $nx \leqq a$. It follows that g attains a maximum at some point of $(0, a]$.

THEOREM 7. *Let f be csa on $[0, a]$, f^*F, and let g be defined as* *above. Let t be a point of $(0, a]$ at which g attains its maximum M. Then*

(a) $F(x)/x \le M$ *for all* $x > 0$,

(b) $F(x)/x = M$ *if* x *is an integral multiple of* t,

(c) $\lim_{x\to\infty} F(x)/x = M$,

(d) $\max_{x\in[0,a]} [Mx - f(x)] = \max_{x\in E} [Mx - F(x)]$,

(e) $\lim_{x\to\infty} [F(x) - Mx] = 0$ *if* f *is differentiable at* t.

Proof. The proofs of (a), (b), (c) and (d) are straightforward and will be omitted. Let us then turn to (e). For each $x \in E$, write x in the form $x = nt + y$, where n is an integer and $0 \leqq y < t$. Define a function F^* by $F^*(nt + y) = nf(t + y/n)$, $n = 1, 2, \cdots$. Clearly $F^*(x) \leqq F(x) \leqq Mx$ for all $x \in E$. We will show that $\lim_{x\to\infty} [Mx - F^*(x)] = 0$. By the definition of F^* we have

$$Mx - F^*(x) = M(nt + y) - nf(t + y/n) .$$

Noting that $f(t) = Mt$, we see that the right hand member of this last equation can be written in the form

$$y\left[M - \frac{f(t + y/n) - f(t)}{y/n}\right] \tag{1}$$

Now let $x \to \infty$. Then y is bounded between 0 and t and $n \to \infty$. The expression (1) approaches 0, since $f'(t) = M$.

We observe that the function F^* of the preceding theorem is asymptotic to F with $F^* \leqq F$. Hence $F(x)$ is bounded between $F^*(x)$ and Mx, two functions which are easy to calculate, and whose difference is small when x is large.

4. The polygonal method. The minimal superadditive extension of a *csa* function may also be obtained as the limit of a sequence of polygonal functions. A function p is said to be *polygonal* if p is continuous and piecewise linear. The point $x \in [0, a]$ is called a *vertex* of p if $(x, p(x))$ is a vertex of the polygon forming the graph of p.

THEOREM 8. *Let p be polygonal on $[0, a]$ with equally spaced vertices. Then p is superadditive if and only if p is superadditive on its vertices.*

Proof. If p is superadditive, then p is clearly superadditive on its vertices. To prove the converse consider the function g defined on the set

$$D \equiv \{(x, y): 0 \leqq x, y \leqq a \text{ and } x + y \leqq a\}$$

by the equation $g(x, y) = p(x + y) - p(x) - p(y)$. It is easy to verify that g is planar on any triangle T of the form

$$T = \{(x, y)\colon u_1 \leqq x \leqq u_2;\ v_1 \leqq y \leqq v_2,\ x + y \leqq (\text{or} \geqq)\ u_2 + v_2\}\ ,$$

where (u_1, v_1) and (u_2, v_2) are pairs of successive vertices of p. Hence g attains its minimum on T at one of the points (u_i, v_i) and therefore its minimum on D at a point (u, v) where both u and v are vertices of p. Thus, if g is anywhere negative then g is negative at a point whose two coordinates are vertices of p. This proves the theorem.

Now let p be a polygonal function on $[0, a]$ with vertices at $0, v, 2v, \cdots, mv = a$. We define a function P on E as follows:

$$P(x) = p(x) \quad \text{for } x \leqq a$$
$$P(Mv) = \max_{K=1,\cdots,M-1} [P(kv) + P(Mv - kv)] \quad M \text{ an integer } \geqq m + 1$$

and

$$P \text{ linear on } [Mv, (M + 1)v],\ M = m, m + 1, \cdots .$$

P will be called the function associated with p. It is easy to see that if p is *csa*, then P is *csa*.

DEFINITION. A sequence $\{p_n\}$ of functions defined on $[0, a]$ is called a p-sequence if

(i) each p_n is a polygonal function

(ii) the vertices of p_n are $Ka/2^n$, $\qquad K = 0, 1, \cdots, 2^n$

(iii) $P_n(Ka/2^m) = p_m(Ka/2^m)$ if $m \leqq n$.

In terms of this concept we have

THEOREM 9. *Let $\{p_n\}$ be a p-sequence coverging to the csa function f on $[0, a]$. For each positive integer n let P_n be the function associated with p_n. Then, if f^*F, $\{P_n\}$ converges to F on E.*

Proof. It suffices to show that P_n approaches F on $[0, 2a]$. Let $F^*(x) = \overline{\lim}_{n\to\infty} P_n(x)$. It is easy to check that F^* is superadditive. Let V_k be the set of vertices of P_k in $[a, 2a]$, and let $V = \bigcup_1^\infty V_k$. If $v \in V$, then $\lim_{n\to\infty} P_n(v)$ exists since the sequence $\{P_n(v)\}$ is ultimately non-decreasing and $P_n(v) \leqq F(v)$ for all n. We have $\lim_{n\to\infty} P_n(v) \leqq F(v)$. But since F^* is superadditive, we have $F^* \geqq F$. Hence $F^* = F$ on V. By standard arguments involving the continuity of F, the density of V in $[a, 2a]$, and the monotonicity of each P_n and F^*, it follows that $F \equiv F^*$ and that $F^* = \lim_{n\to\infty} P_n(x)$.

5. Superadditive functions in n-dimensions. It turns out that many of the results obtained in one dimension have their analogues in n-dimensions. The interval $I \equiv [0, a]$ is replaced by a fundamental region R defined by the inequalities $0 \leqq x_i \leqq a_i$, $i = 1, \cdots, n$, where the a_i are arbitrary positive numbers. The decomposition method works, just as it does on the line, and we can prove with little difficulty that to any superadditive function f on R there corresponds a minimal superadditive extension F to $E_n^+ \equiv \{(x_1, \cdots, x_n): 0 \leqq x_i, i = 1, \cdots, \cdots, n\}$. We can also prove a theorem corresponding to Theorem 5, the derivatives here being directional derivatives. In Theorem 7 a certain line $l(x) = Mx$ played an important role. In n-dimensions, for each direction θ we have a plane P_θ which plays the role of l in some direction, and when the function P, defined on the fundamental region R by the equation

$$P(z) = \inf_{\theta} P_\theta(z) ,$$

is extended to E_n^+ by homogeneity it is the minimal concave superadditive function which bounds F from above. It can be proved, at least in E_2^+, that

$$n \max_{z \in R} [P(z) - f(z)] \geqq \max_{z \in E_n^+} [P(z) - F(z)] .$$

Bibliography

1. E. Hille and R. Phillips, *Functional analysis and semi-groups*, Amer. Math. Soc. Colloquim Publications vol. **XXXI** ch. 7 pp. 237-255.

2. R. A. Rosenbaum, *Subadditive functions*, Duke Math J., **17** (1950), p 227-247.

University of California,
Santa Barbara

ON EXPANSIVE HOMEOMORPHISMS

B. F. BRYANT

1. Introduction. A homeomorphism ϕ of a compact metric space X onto X is said to be *expansive* provided there exists $d > 0$ such that if $x, y \in X$ with $x \neq y$, then there exists an integer n such that $\rho(x\phi^n, y\phi^n) > d$ (see [1] and [3]). The question arises as to the possibility of extending the results concerning expansive homeomorphisms to compact uniform spaces. The extension is possible, although trivial in light of the corollary to Theorem 1.

In §§ 3 and 4 the setting is a compact metric space X. Theorem 2 is stronger than Theorem 10.36 of [1] in that we do not require X to be self-dense. Also, the lemmas of which Theorem 2 is a consequence are perhaps of some interest in themselves. In § 4 we show that if X is self-dense, then for each $x \in X$ and each $\varepsilon > 0$ there exists $y \in U(\varepsilon, x)$ such that x and y are not doubly asymptotic.

2. A homeomorphism ϕ of a compact uniform space $(X, \mathscr{U})$ onto $(X, \mathscr{U})$ is said to be expansive provided there exists $U \in \mathscr{U}$ such that $U \neq \Delta$ (the diagonal) and if $x, y \in X$ with $x \neq y$, then there exists an integer n such that $(x\phi^n, y\phi^n) \notin \bar{U}$. For uniform spaces we use the notation of [2], but following Weil [4] we suppose $(X, \mathscr{U})$ is Hausdorff; i.e., $\bigcap\{U: U \in \mathscr{U}\} = \Delta$. We also suppose that each $U \in \mathscr{U}$ is symmetric.

THEOREM 1. *Let $(X, \mathscr{U})$ be a compact uniform space which is not metrizable and let ϕ be a homeomorphism of X onto X. If $U \in \mathscr{U}$, then there exist $x, y \in X$ with $x \neq y$ such that $(x\phi^n, y\phi^n) \in U$ for each integer n. (Compare with Theorem* 10.30 *of* [1].)

Proof. Select $V \in \mathscr{U}$ such that $V \circ V \circ V \subset U$ and $\bar{V} \subset U$ (see [2], p. 180). Since ϕ^n, for each integer n, is uniformly continuous, we may choose $U_1 \in \mathscr{U}$ with $U_1 \subset V$ such that $(p, q) \in U_1$ implies $(p\phi^k, q\phi^k) \in V$ for $k = \pm 1$. For $i > 1$, choose $U_i \in \mathscr{U}$ with $U_i \subset U_{i-1}$ such that $(p, q) \in U_i$ implies $(p\phi^k, q\phi^k) \in V$ for $k = \pm i$. Since $(X, \mathscr{U})$ is not metrizable, the countable set $\{U_i \mid i = 1, 2, \cdots\}$ is not a base for the uniformity $\mathscr{U}$([4], p. 16). Thus there exists $W \in \mathscr{U}$ with $W \subset U$ such that $i \geqq 1$ implies $U_i \cap \text{comp } W \neq 0$. Choose, for each i,$(x_i, y_i) \in U_i \cap \text{comp } W$. Since $X \times X$ is a compact Hausdorff space, there exists (x, y) such that each neighborhood of (x, y) contains (x_i, y_i) for an infinite number of positive integers i. Let n be an arbitrary positive integer, then there exists $m > n$ such that $(x_m, y_m) \in U_n(x) \times U_n(y)$. Hence $(x, x_m) \in U_n$ and $(y, y_m) \in U_n$;

Received October 3, 1959, and in revised form, February 1, 1960.

therefore $(x\phi^k, x_m\phi^k) \in V$ and $(y\phi^k, y_m\phi^k) \in V$ for $k = \pm n$. Also $(x_m, y_m) \in U_m \subset U_n$ so that $(x_m\phi^k, y_m\phi^k) \in V$ for $k = \pm n$. Hence $(x\phi^k, y\phi^k) \in V \circ V \circ V \subset U$ for $k = \pm n$. Each $(x_i, y_i) \in U_i \subset V$ and $\bar{V} \subset U$; hence $(x, y) \in U$. Finally, $x \neq y$. For otherwise we could choose $S \in \mathscr{U}$ such that $S \circ S \subset W$; then $(x_k, y_k) \in S(x) \times S(x)$ for some k, and hence $(x, x_k) \in S$, $(x, y_k) \in S$ so that $(x_k, y_k) \in W$. This completes the proof.

An immediate consequent of the theorem is the following

COROLLARY. *If $(X, \mathscr{U})$ is a compact uniform space on which it is possible to define an expansive homeomorphism, then $(X, \mathscr{U})$ is metrizable.*

3. The author is indebted to the referee for suggesting the arrangement of the material in this section. In the original version, Lemma 2 had a slightly stronger hypothesis and Lemma 3 was essentially contained in the proof of Theorem 2. In this section we suppose that X is an infinite compact metric space and (with the exception of Lemma 3) that ϕ is an expansive homeomorphism (with expansive constant d) of X onto X.

LEMMA 1. *If $x \neq y$ and if there is an integer N such that $n > N$ $\{n < N\}$ implies $\rho(x\phi^n, y\phi^n) \leqq d$, then x and y are positively $\{$negatively$\}$ asymptotic under ϕ.*

Proof. If x and y are not positively asymptotic under ϕ, then there exist $\varepsilon > 0$ and positive integers $n_1 < n_2 < \cdots$ such that $\rho(x\phi^{n_i}, y\phi^{n_i}) \geqq \varepsilon$ with $\lim_{i \to +\infty} x\phi^{n_i} = u$ and $\lim_{i \to +\infty} y\phi^{n_i} = v$. Obviously $u \neq v$. Let m be an arbitrary integer. For all i sufficiently large $n_i + m > N$; hence $\rho(x\phi^{n_i+m}, y\phi^{n_i+m}) \leqq d$. Since $\lim_{i \to +\infty} x\phi^{n_i+m} = u\phi^m$ and $\lim_{i \to +\infty} y\phi^{n_i+m} = v\phi^m$, it is clear that $\rho(u\phi^m, v\phi^m) \leqq d$ for each integer m. This contradicts the hypothesis that ϕ is expansive. The alternative statement may be proved by a similar argument.

LEMMA 2. *If $\omega(x)\{\alpha(x)\}$ contains a periodic point p and $\omega(x)\{\alpha(x)\}$ is not identical with the orbit of p, then there exist w and z in $\omega(x)$ $\{\alpha(x)\}$ such that w and p are positively asymptotic and z and p are negatively asymptotic.*

Proof. Suppose p is of period k. There exist positive integers $n_1 < n_2 < \cdots$ such that $\lim_{i \to +\infty} x\phi^{n_i} = p$. Let k_i be the smallest non-negative integer such that $n_i + k_i$ is a multiple of k. Since $0 \leqq k_i < k$, there exists m such that $k_i = m$ for an infinite number of integers i. Thus there are positive integers $m_1 < m_2 < \cdots$ such that

$$\lim_{i\to+\infty} x\phi^{m_i+m} = \lim_{i\to+\infty} x\phi^{kj_i} = p\phi^m .$$

Denote ϕ^k by θ (with expansive constant d_1) and $p\phi^m$ by q (see [1], p. 86). Thus $\lim_{i\to+\infty} x\theta^{j_i} = q$ and $q\theta = q$. We can assume that $\rho(x\theta^{j_i}, q) < d_1$ for each i.

The points x and q are not positively asymptotic under θ, since otherwise $\omega(x)$ under ϕ would consist of the k points in the orbit of p. Hence, by Lemma 1, there exist arbitrarily large integers r such that $\rho(x\theta^r, q) > d_1$. Therefore we can assume that $s_1 < s_2 \cdots$ are positive integers where s_i is the smallest positive integer such that $\rho(x\theta^{j_i+s_i}, q) > d_1$ and $\lim_{i\to+\infty} x\theta^{j_i+s_i} = u \in \omega(x)$. Let $-a$ be an arbitrary negative integer, then for all i sufficiently large $0 < s_i - a < s_i$. Hence $\rho(x\theta^{j_i+s_i-a}, q) \leqq d_1$, and therefore $\rho(u\theta^{-a}, q) \leqq d_1$ for each negative integer $-a$. Thus, by Lemma 1, u is negatively asymptotic to q under θ and hence under ϕ([1], p. 85). We can assume $j_i < j_i + s_i < j_{i+1}$ and hence that there exist positive integers $t_2 < t_3 < \cdots$ where t_i is the smallest positive integer such that $\rho(x\theta^{j_i-t_i}, q) > d_1$ and $\lim_{i\to+\infty} x\theta^{j_i-t_i} = v \in \omega(x)$. By an argument similar to the above, v is positively asymptotic to q under ϕ. Since $\alpha(x)$ under ϕ coincides with $\omega(x)$ under ϕ^{-1}, this completes the proof.

In the following lemma we do not require ϕ to be expansive.

LEMMA 3. *If x is not periodic and $\omega(x)\{\alpha(x)\}$ is the orbit of a periodic point p, then there exists a point q in the orbit of p such that q and x are positively $\{$negatively$\}$ asymptotic.*

Proof. Let $p \in \omega(x)$ and, as in the first paragraph of the proof of Lemma 2, select positive integers $j_1 < j_2 < \cdots$ such that $\lim_{i\to+\infty} x\theta^{j_i} = q = p\phi^m$ and $q\theta = q$, $\theta = \phi^k$. If x and q are not positively asymptotic under θ, then there exists a positive constant α and a sequence $n_1 < n_2 < \cdots$ of integers such that $\rho(x\theta^{n_i}, q) > \alpha$. Let $\varepsilon > 0$ and choose $\beta > 0$ such that $\beta < \varepsilon$, $\beta < \alpha$, and $\rho(z, w) \leqq \beta$ implies $\rho(z\theta, w\theta) < \varepsilon$. We can assume that $\rho(x\theta^{j_i}, q) < \beta$. Let s_i be the smallest positive integer such that $\rho(x\theta^{j_i+s_i}, q) > \beta$. Then for each i, $\beta < \rho(x\theta^{j_i+s_i}, q) < \varepsilon$. But the sequence $\{x\theta^{j_i+s_i}\}$ has a convergent subsequence. Let s be the limit of such a convergent subsequence, then $s \neq q$, $s \in \omega(x)$ and $\rho(s, q) \leqq \varepsilon$. Thus $\omega(x)$ is not finite, contrary to hypothesis. It follows that x and q are positively asymptotic under θ, and hence under ϕ.

Similarly, if $\alpha(x)$ is the orbit of a periodic point p, then there exists a point q in the orbit of p such that q and x are negatively asymptotic under ϕ.

THEOREM 2. *There exist $a, b, c, d \in X$ such that a and b are positively asymptotic under ϕ and c and d are negatively asymptotic under ϕ.*

Proof. There exists a minimal set $N \subset X$ ([1], p. 15). If N is infinite, then N is self-dense and the conclusion follows from Theorem 10.36 of [1]. Henceforth, suppose each minimal set in X is finite and thus is a periodic orbit.

Since X is compact and infinite, there exists a non-isolated point r. If r is not periodic, let $r = p$. If r is periodic, then there exists $x \neq r$ such that x and r are asymptotic ([1], p. 87). But then x is not periodic and we let $x = p$.

There exists a minimal set $N \subset \omega(p)$, and a minimal set $M \subset \alpha(p)$. Both N and M are periodic orbits. If $N \neq \omega(p)$ or $M \neq \alpha(p)$ the conclusion of the theorem follows from Lemma 2. If $N = \omega(p)$ and $M = \alpha(p)$, the conclusion of the theorem follows from Lemma 3.

4. In addition to the standing hypothesis of § 3 we require X to be self-dense.

LEMMA 4. *If $y \in U(\varepsilon, x)$ implies that each neighborhood of y contains z such that $\rho(y\phi^n, z\phi^n) > d/2$ for some positive {negative} n, then there exists $w \in U(\varepsilon, x)$ such that w and x are not positively {negatively} asymptotic.*

Proof. Let $0 < \alpha < \varepsilon$, then there exist $x_1 \in U(\alpha, x)$ and a positive integer n_1 such that $\rho(x_1\phi^{n_1}, x\phi^{n_1}) > d/2$. Suppose x_1 and x are positively asymptotic (otherwise the lemma holds); hence there exists $m_1 > n_1$ such that $n > m_1$ implies $\rho(x_1\phi^n, x\phi^n) < d/8$. Choose $\alpha_1 > 0$ such that $U(\alpha_1, x_1) \subset U(\alpha, x)$ and $\rho(p, q) < \alpha_1$ implies $\rho(p\phi^n, p\phi^n) < d/8$ for $0 \leqq n \leqq m_1$. For $i > 1$ we select x_i, n_i, m_i, and $\alpha_i > 0$ such that $x_i \in U(\alpha_{i-1}, x_{i-1})$, $n_i > m_{i-1}$, $\rho(x_i\phi^{n_i}, x_{i-1}\phi^{n_i}) > d/2$, $m_i > n_i$, $n > m_i$ implies $\rho(x_i\phi^n, x\phi^n) < d/8$, $U(\alpha_i, x_i) \subset U(\alpha_{i-1}, x_{i-1})$, and $\rho(p, q) < \alpha_i$ implies $\rho(p\phi^n, q\phi^n) < d/8$ for $0 \leqq n \leqq m_i$. We can suppose $\lim_{i \to +\infty} x_i = w \in \overline{U(\alpha, x)} \subset U(\varepsilon, x)$ and $w \neq x$. If $i > 1$, then $n_i > m_{i-1}$ and hence $\rho(x_{i-1}\phi^{n_i}, x\phi^{n_i}) < d/8$. But $\rho(x_i\phi^{n_i}, x_{i-1}\phi^{n_i}) > d/2$, and the triangle inequality implies $\rho(x_i\phi^{n_i}, x\phi^{n_i}) > 3d/8$. If $j > i$, then $x_j \in U(\alpha_i, x_i)$ and, since $m_i > n_i$, $\rho(x_j\phi^{n_i}, x_i\phi^{n_i}) < d/8$. Therefore $\rho(x_j\phi^{n_i}, x\phi^{n_i}) > d/4$ for $j \geqq i$. If $i > 1$ is fixed, then $\rho(x_j\phi^{n_i}, w\phi^{n_i})$ is arbitrarily small for j sufficiently large. Hence $\rho(x\phi^{n_i}, w\phi^{n_i}) \geqq d/4$. Since $\{n_i\}$ is an increasing sequence of positive integers, w and x are not positively asymptotic. This proof establishes the alternative statement by using ϕ^{-1} rather than ϕ.

THEOREM 3. *For each $x \in X$ and each $\varepsilon > 0$ there exists $y \in U(\varepsilon, x)$ such that x and y are not doubly asymptotic.*

Proof. Suppose there exist $x \in X$ and $\varepsilon > 0$ such that $z \in U(\varepsilon, x)$ implies x and z are positively asymptotic. Suppose $\varepsilon < d/2$, then, by the above lemma, there exist $y \in U(\varepsilon, x)$ and $\alpha > 0$ such that $U(\alpha, y) \subset U(\varepsilon, x)$ and $t \in U(\alpha, y)$ implies that $\rho(t\phi^n, y\phi^n) \leqq d/2$ for $n \geqq 0$. Therefore $u, v \in U(\alpha, y)$ implies $\rho(u\phi^n, v\phi^n) \leqq d$ for $n \geqq 0$. Thus, since ϕ is expansive, $u, v \in U(\alpha, y)$ implies $\rho(u\phi^n, v\phi^n) > d$ for some negative n. By the alternative form of the lemma above, there exists $w \in U(\alpha, y)$ such that w and y are not negatively asymptotic. Therefore either w and x are not negatively asymptotic or y and x are not negatively asymptotic, which establishes the theorem.

If X is an infinite minimal set, then a stronger statement can be made. Since X is pointwise almost periodic under ϕ([1], p. 31), $\varepsilon > 0$ implies $\rho(x, x\phi^n) < \varepsilon$ for some $n \neq 0$. It is easy to show that x and $x\phi^n$ are neither positively nor negatively asymptotic.

If X is not self-dense, then, as shown by the following example, each pair of distinct points may be both positively and negatively asymptotic. Let X consist of the real numbers $0, 1/n \{n = \pm 1, \pm 2, \cdots\}$, and let

$$x\phi = \begin{cases} 0 \quad \text{if} \quad x = 0\,. \\ 1/(n+1) \quad \text{if} \quad x = 1/n \quad \text{and} \quad n \neq -1\,. \\ 1 \quad \text{if} \quad x = -1. \end{cases}$$

References

1. W. H. Gottschalk and G. A. Hedlund, *Topological dynamics*, Amer. Math. Soc. Colloquium Publications, vol. 36, Providence, 1955.

2. J. L. Kelley, *General topology*, D. Van Nostrand Co., Inc., New York, 1955.

3. W. R. Utz, *Unstable homeomorphisms*, Proc. Amer. Math. Soc. **1** (1950), 769–774.

4. A. Weil, *Sur les espaces a structure uniform et sur la topologie générale*, Actualités Sci. Ind. 551, Paris, 1937.

Vanderbilt University

ON COMPLETE AND INDEPENDENT SETS OF OPERATIONS IN FINITE ALGEBRAS

JEAN W. BUTLER

In [4] Post obtained a variety of results about truth functions in 2-valued sentential calculus. He studied sets of truth functions which could be used as primitive notions for various systems of 2-valued logics. In particular, he was interested in complete sets of truth functions, i.e., sets having the property that every truth function with an arbitrary finite number of arguments is definable in terms of the truth functions belonging to the set. Among other results Post established a computable criterion for a set of truth functions to be complete. Using this criterion he showed that there is a finite upper bound for the number of elements in any complete and independent set of primitive notions for the 2-valued sentential calculus (and that actually the number 4 is the least upper bound). Alfred Tarski has asked to what extent these results can be extended to n-valued sentential calculus, for any finite n. It will be seen from this note that those results of Post concerning complete sets of truth functions can actually be extended. On the other hand it has been shown recently by A. Ehrenfeucht that the result concerning arbitrary sets of functions cannot be extended.

Both the results of Post and those of this note can be formulated in terms of truth functions of the 2-valued (n-valued) sentential calculus or in terms of finitary operations in arbitrary 2 element (n element) algebras. We choose the second alternative since the many-valued sentential calculi have a rather restricted significance in logic and mathematics.

Thus we shall concern ourselves with finitary operations under which a given set A with n elements is closed. For simplicity we restrict our attention to the case when A is the set N of all natural numbers less than n. This restriction implies no loss of generality, since all the results can be extended by isomorphism to any finite set with n elements. For convenience we will identify N with n, as is frequently done in modern set theory.

For any given natural number k, let n^k be the set of all k-termed sequences $x = \langle x_0, x_1, \cdots, x_{k-1}\rangle$ with terms in n. Denote by $F_{n,k}$ the set of all k-ary operations on and to elements of n, i.e., of all function

Received December 15, 1959. This is the text of a talk given by the author at the Summer Institute of Symbolic Logic held in 1957 at Cornell University; it appeared in condensed form in "Summaries of talks presented at the Summer Institute of Symbolic Logic in 1957 at Cornell University" (mimeographed), 1957, pp. 78–80. The results were first communicated to the American Mathematical Society in [**1**]. The author thanks Professor Alfred Tarski for his guidance and encouragement.

on n^k to n. Let $F_n = \bigcup_{k<\infty} F_{n,k}$, i.e., F_n is the set of all finitary operations on and to elements of n.

For any subset X of F_n we will denote by $\bar{X}$ the smallest set Y which includes X and satisfies the following conditions:

(i) if $f \in Y$ and h is obtained from f by exchanging or identifying two arguments, then $h \in Y$;

(ii) if $f \in Y$, $g \in Y$, and h is obtained from f by replacing an argument by g, then $h \in Y$. A function f is said to be *generated* by a set X if $f \in \bar{X}$. The set X is called *closed* if $\bar{X} = X$, it is called *complete* if $\bar{X} = F_n$, it is called *independent* if there is no proper subset X' of X for which $\bar{X}' = \bar{X}$. A set Y is called a *basis* for a set X if $Y \subseteq X$ and $\bar{Y} = \bar{X}$. A function $f \in F_{n,k}$ is called *reducible* or reducible to first order if its values depend on at most one argument; i.e., if there is an $i < k$ and an $h \in F_{n,1}$ such that for every sequence $x \in n^k$ we have $f(x) = h(x_i)$. Hence f is not reducible if and only if for every $q < k$ there are $x, y \in n^k$ with $x_q = y_q$ and $f(x) \neq f(y)$. We will denote by $\mathscr{R}(f)$ the range of f. We single out two functions in F_n. $\mathbf{V}_n$ is the function of two arguments defined by the formula:

$$x \mathbf{V}_n y = \max(x, y) .$$

$\sim_n$ is the function of one argument defined by:

$$\sim_n x = x + 1 (\mathrm{mod}\, n) .$$

In the following few lemmas we will establish some properties of the notions just defined:

LEMMA 1. *If $f \in F_{n,k}$, $n \geqq 3$, f not reducible and $\mathscr{R}(f) = n$, then $\{f\} \cup F_{n,1}$ generates a function $g \in F_{n,2}$, g not reducible and $\mathscr{R}(g) = n$.*

Proof. We first establish that there exist $q < k$ and $u, v \in n^k$ such that $f(u) \neq f(v)$ and $u_i = v_i$ for all $i \neq q$, $i < n$. Since $\mathscr{R}(f)$ contains more than one element there exist $a, b \in n^k$ such that $f(a) \neq f(b)$. There are $k+1$ sequences $c^{(0)}, c^{(1)}, \cdots, c^{(k)} \in n^k$ with $c^{(0)} = a$, $c^{(k)} = b$, and such that for any $i < k$, $c^{(i+1)}$ is obtained from $c^{(i)}$ by replacing $c_i^{(i)}$ by $c_i^{(k)}$. Hence $c^{(i)}$ and $c_i^{(i+1)}$ differ only in the ith coordinate, moreover $c_i^{(j)} = a_i$ for $i \geqq j$ and $c_i^{(j)} = b_i$ for $i < j$. Since $f(a) = f(c^{(0)}) = f(c_0^{(0)}, c_1^{(0)}, \cdots, c_{k-1}^{(0)})$, $f(b) = f(c^{(k)})$ and $f(a) \neq f(b)$ it cannot be the case that $f(c^{(i)}) = f(c^{(i+1)})$ for all $i < k$. Therefore there is some $q < k$ such that $f(c^{(q)}) \neq f(c^{(q+1)})$ and $c_i^{(q)} = c_i^{(q+1)}$ for all $i \neq q$, $i < k$. Take $c^{(q)}$ for u and $c^{(q+1)}$ for v.

Since $\mathscr{R}(f) = n$, we can choose n sequences $y^{(0)}, y^{(1)}, \cdots, y^{(n-1)} \in n^k$ such that $f(y^{(0)}) = f(u)$, $f(y^{(1)}) = f(v)$, and each value in n is taken on by f for some $y^{(i)}$. There also exist w, $z \in n^k$ for which $w_q = z_q$ and

$f(w) \neq f(z)$, since if this were not the case f would depend only on its qth argument and thus be reducible.

There are two possible cases: (i) there exist $w, z \in n^k$, with $w_q = z_q$, and $f(w) \neq f(z)$, $f(w) \neq f(u)$, $f(w) \neq f(v)$; (ii) for every w, $z \in n^k$ with $w_q = z_q$ and $f(w) \neq f(z)$ neither $f(w)$ nor $f(z)$ is different from both $f(u)$ and $f(v)$.

If case (i) holds there exist $w, z \in n^k$ for which $w_q = z_q$, and $f(w) \neq f(z)$, $f(w) \neq f(u)$, $f(w) \neq f(v)$. With no loss of generality we may assume $f(y^{(2)}) = f(w)$. We define k functions $h_0, h_1, \cdots, h_{k-1} \in F_{n,1}$ separately for h_q and for h_j, $j \neq q$,

$$h_q(x) = \begin{cases} u_q & x = 0 \\ v_q & x = 1 \\ z_q & x = 2 \\ y_q^{(x)} & x > 2 \end{cases} \qquad h_j(x) = \begin{cases} u_j & x = 0 \\ z_j & x = 1 \\ w_j & x = 2 \\ y_j^{(x)} & x > 2 . \end{cases}$$

We define $g \in F_{n,2}$ as follows:

$$(1) \qquad g(x, y) = f(h_0(x), h_1(x), \cdots, h_{q-1}(x), h_q(y), h_{q+1}(x), \cdots, h_{k-1}(x)) .$$

Notice that y appears only in the qth coordinate of f. $\mathscr{R}(g) = n$ since $g(0, 0) = f(u) = f(y^{(0)})$, $g(0, 1) = f(v) = f(y^{(1)})$, $g(2, 2) = f(w) = f(y^{(2)})$, and $g(i, i) = f(y^{(i)})$ for $i > 2$. Moreover g is not reducible, since $g(0, 0) \neq g(0, 1)$ and $g(1, 2) \neq g(2, 2)$.

If case (ii) holds we take for h_q the identity function in $F_{n,1}$ and using any $w, z \in n^k$ for which $w_q = z_q$ and $f(w) \neq f(z)$ we choose $k - 1$ functions $h_j \in F_{n,1}$ for $j \neq q$, satisfying

$$h_j(x) = \begin{cases} u_j & x = 0 \\ w_j & x = 1 \\ z_j & x = 2 \end{cases}$$

and then define g by (1). Now for any $s \in n^k$ with $s_q = y_q^{(i)}$, $2 < i \leqq n$, condition (ii) guarantees that $f(s) = f(y^{(i)})$ since $f(y^{(i)}) \neq f(u)$ and $f(y^{(i)}) \neq f(v)$. Therefore $g(0, u_q) = f(u) = f(y^{(0)})$, $g(1, v_q) = f(v) = f(y^{(1)})$, $g(m, y_q^{(i)}) = f(y^{(i)})$ for $m < n$ and $2 \leqq i < n$. Hence $\mathscr{R}(g) = n$. The function g is not reducible, since $g(0, u_q) \neq g(0, v_q)$ and $g(1, w_q) \neq g(2, w_q)$.

LEMMA 2. *If $f \in F_{n,2}$ is not reducible and $\mathscr{R}(f)$ has p elements $p \geqq 3$, then there exist $i, j, k, l < n$ such that among the function values $f(i, k), f(i, l), f(j, k), f(j, l)$ at least three distinct values are represented.*

Proof. There are two possibilities;

(i) in the table of f all value in $\mathscr{R}(f)$ are taken on across some row, i.e., there is an $i < n$ such that the set of all values $f(i, j)$ with

$j < n$ coincides with $\mathscr{R}(f)$;

(ii) in no row are all values in $\mathscr{R}(f)$ taken on.

If (i) holds, since f is not reducible there must be $j, l < n$ for which $f(i, l) \neq f(j, l)$. Since $p \geqq 3$ and all values in $\mathscr{R}(f)$ are taken on in the ith row there must be a $k < n$ such that $f(i, k)$ is distinct from both $f(i, l)$ and $f(j, l)$.

If (ii) holds, since f is not reducible there is some non-constant row, i.e., some i, j', j'' such that $f(i, j') \neq f(i, j'')$. However, by assumption there is a $w \in \mathscr{R}(f)$ which does not appear in this row, i.e., for all $x < p$, $f(i, x) \neq w$. Since $w \in \mathscr{R}(f)$ there are $j, l < n$ for which $f(j, l) = w$. Hence $f(i, l) \neq f(j, l)$. Since w does not appear in the ith row and the ith row is not constant there is some $k < n$ such that the value $f(i, k)$ is different from both $f(i, l)$ and $f(j, l)$.

LEMMA 3. *If $f \in F_{n,2}$ is not reducible and $\mathscr{R}(f)$ has exactly p elements, $3 \leqq p \leqq n$, then there exist two functions $h_1, h_2 \in F_{n,1}$ with $\mathscr{R}(h_1)$, $\mathscr{R}(h_2)$ each consisting of at most $p - 1$ elements, and such that for every $x \in \mathscr{R}(f)$ we have $f(h_1(x), h_2(x)) = x$.*

Proof. By Lemma 2 we can find $i, j, k, l < n$ such that $f(i, k)$, $f(i, l)$, $f(j, k)$, $f(j, l)$ represent at least three distinct values. Assume $f(i, k) = u$, $f(i, l) = v$, $f(j, k) = w$ are all different. Functions $h_1, h_2 \in F_{n,1}$ can be found such that

$$\begin{aligned} h_1(u) &= i \qquad & h_2(u) &= k \\ h_1(v) &= i \qquad & h_2(v) &= l \\ h_1(w) &= j \qquad & h_2(w) &= k \end{aligned}$$

and

$$h_1(x) = i\ , \quad h_2(x) = k \quad \text{for} \quad x \notin \mathscr{R}(f)$$
$$f(h_1(x), h_2(x)) = x \quad \text{for} \quad x \in \mathscr{R}(f) \sim \{u, v, w\}\ .$$

Clearly, h_1 and h_2 satisfy the requirements of the Lemma. The proof in the other cases is analogous.

LEMMA 4. *If $f \in F_{n,2}$ and $2 \leqq p \leqq n$, and there exist $i, j, k < n$ such that for all $y < p$*

$$f(i, y) = y \quad \textit{and} \quad f(j, y) = k$$

then f together with the functions in $F_{n,1}$ generate a function $g \in F_{n,2}$ such that $g(x, y) = x \mathbf{V}_n y$ for $x, y < p$.

Proof. We shall prove, by induction on p, a slightly weakened form of the theorem, replacing the condition $i, j, k < n$ by $i, j, k < p$. The theorem as stated then follows, since $\{f\} \cup F_{n,1}$ generates a function satisfying the strengthened hypothesis.

For $p = 2$, since $i, j < p$ the function f must agree on $\{0, 1\}$ with one of the following four tables:

$$\text{(i)}\ \begin{array}{c|cc} & 0 & 1 \\ \hline 0 & 0 & 1 \\ 1 & 1 & 1 \end{array} \qquad \text{(ii)}\ \begin{array}{c|cc} & 0 & 1 \\ \hline 0 & 0 & 1 \\ 1 & 0 & 0 \end{array} \qquad \text{(iii)}\ \begin{array}{c|cc} & 0 & 1 \\ \hline 0 & 0 & 0 \\ 1 & 0 & 1 \end{array} \qquad \text{(iv)}\ \begin{array}{c|cc} & 0 & 1 \\ \hline 0 & 1 & 1 \\ 1 & 0 & 1 \end{array} .$$

In case (i) f itself may be taken for g. In the other cases using any homomorphism $h \in F_{n,1}$ which exchanges 0 and 1, g may be constructed as (ii) $h(f(x, h(y)))$, (iii) $h(f(h(x), h(y)))$, (iv) $f(h(x), y)$.

Assume the theorem is true for $p - 1$. From $\{f\} \cup F_{n,1}$ we can construct a function satisfying the induction hypothesis for $p - 1$. Thus we can generate a function g' such that

$$g'(x, y) = x \bigvee_n y \quad \text{for} \quad x, y < p - 1 .$$

Now choose functions $h_1, h_2 \in F_{n,1}$ such that

$$h_1(x) = \begin{cases} i & \text{if} \quad x < p - 1 \\ j & \text{if} \quad x = p - 1 \end{cases} \qquad h_2(x) = \begin{cases} p - 1 & \text{if} \quad x = k \\ k & \text{if} \quad x = p - 1 \\ x & \text{otherwise} \end{cases}$$

and construct $f' \in F_{n,2}$ as follows:

$$f'(x, y) = h_2(f(h_1(x), h_2(y))) .$$

It can be seen that

$$f'(x, y) = \begin{cases} y & \text{if} \quad x < p - 1 \quad \text{and} \quad y < p \\ p - 1 & \text{if} \quad x = p - 1 \quad \text{and} \quad y < p . \end{cases}$$

Now we define $g \in F_{n,2}$:

$$g(x, y) = f'(f'(x, y), g'(x, y)) .$$

For $x, y < p - 1$, $g(x, y) = g'(x, y) = x \bigvee_n y$; for $x = p - 1$, $y < p$, $g(x, y) = f'(p - 1, g'(p - 1, y)) = p - 1$; for $x < p - 1$ and $y = p - 1$, $g(x, y) = f'(p - 1, g'(x, p - 1)) = p - 1$. Therefore g agrees with $\bigvee_n$ for $x, y < p$, which estalishes the Lemma.

LEMMA 5. *If $f \in F_{n,2}$ is not reducible and $\mathscr{R}(f) = p$, $3 \leqq p \leqq n$, then f together with the functions in $F_{n,1}$ generate a function $g \in F_{n,2}$ such that $g(x, y) = x \bigvee_n y$ for all $x, y < p$.*

Proof. The proof is by induction on p. For $p = 3$, using the i, j, k, l of Lemma 3 and appropriate homomorphisms from $F_{n,1}$ we can generate a function $h \in F_{n,2}$ such that the values of h on $\{0, 1\}$ agree with one of the two tables:

$$\text{(i)}\quad \begin{array}{c|cc} & 0 & 1 \\ \hline 0 & 0 & 1 \\ 1 & 2 & 2 \end{array} \qquad\qquad \text{(ii)}\quad \begin{array}{c|cc} & 0 & 1 \\ \hline 0 & 0 & 1 \\ 1 & 2 & 0 \end{array} .$$

We then choose five functions $h_1, h_2, h_3, h_4, h_5 \in F_{n,1}$ such that

$$\begin{array}{lllll} h_1(0) = 0 & h_2(0) = 0 & h_3(0) = 0 & h_4(0) = 0 & h_5(0) = 0 \\ h_1(1) = 0 & h_2(1) = 1 & h_3(1) = 0 & h_4(1) = 1 & h_5(1) = 2 \\ h_1(2) = 1 & h_2(2) = 1 & h_3(2) = 0 & h_4(2) = 0 & h_5(2) = 2 . \end{array}$$

In case (i) g can be constructed as

$$g(x, y) = h(h(x, h_1(y)), h(x, h_2(y))) .$$

To construct g in case (ii) we define $g', g'' \in F_{n,2}$

$$\begin{aligned} g'(x, y) &= h(h(h_3(x), h_1(y)), h_2(h(h_1(x), h_4(y)))) \\ g''(x, y) &= h_4(g'(x, h_5(y))) \end{aligned}$$

and then

$$g(x, y) = g'(g''(x, y), g'(x, y)) .$$

Assuming the theorem true for $p-1$, we prove it for p, $3 < p \leqq n$. First we construct a function f'' satisfying the induction hypothesis for $p-1$. To do this we apply Lemma 2, taking an i, j, k, l such that $f(i, k)$, $f(i, l)$, $f(j, k)$, $f(j, l)$ represent at least 3 distinct values, u, v, w. Since $p > 3$ there is a value $z \in \mathscr{R}(f) \sim \{u, v, w\}$. Let $h \in F_{n,1}$ such that

$$h(x) = \begin{cases} u & x = z \\ x & \text{otherwise.} \end{cases}$$

Then the function $h(f(x, y))$ is not reducible and has $p-1$ elements in its range. The application of an appropriate isomorphism from $F_{n,2}$ will produce a function $f'' \in F_{n,2}$ with $\mathscr{R}(f'') = p - 1$ and f'' not reducible. Then by the induction assumption we can generate a function $g'' \in F_{n,2}$ such that

$$g''(x, y) = x \vee_n y \quad \text{for} \quad x, y < p - 1 .$$

Next by Lemma 3 there exist $h_1, h_2 \in F_{n,1}$ with $\mathscr{R}(h_1)$, $\mathscr{R}(h_2)$ each consisting of at most $p-1$ elements and such that

$$f(h_1(x), h_2(x)) = x$$

for $x < p$.

Let $h_3, h_4 \in F_{n,1}$ be isomorphisms such that

$$h_3(x) < p - 1 \quad \text{for} \quad x \in \mathscr{R}(h_1)$$
$$h_4(x) < p - 1 \quad \text{for} \quad x \in \mathscr{R}(h_2) ,$$

and define $f' \in F_{n,2}$, $h_5, h_6 \in F_{n,1}$:

$$h_5(x) = h_3(h_1(x)) , \qquad x < n$$
$$h_6(x) = h_4(h_2(x)) , \qquad x < n$$
$$f'(x, y) = f(h_3^{-1}(x), h_4^{-1}(y)) , \qquad x, y < n .$$

Then

$$f'(h_5(x), h_6(x)) = x \quad \text{for} \quad x < p$$

and $\mathscr{R}(h_5), \mathscr{R}(h_6) \subseteqq p - 1$.

We can now define a function $g' \in F_{n,2}$ as follows:

$$g'(x, y) = f'(g''(x, h_1(y)), g''(x, h_2(y))) .$$

Then

$$g'(0, y) = f'(h_5(y), h_6(y)) = y \quad \text{for} \quad y < p$$

and

$$g'(p-2, y) = f'(p-2, p-2) \quad \text{for} \quad y < p .$$

Therefore by Lemma 4 we can generate a function $g \in F_{n,2}$ such that $g(x, y) = x \mathbf{V}_n y$ for $x, y < p$.

LEMMA 6. *If* $f \in F_n$, $n \geqq 3$, f *is not reducible, and* $\mathscr{R}(f) = n$ *then* $F_{n,1} \cup \{f\}$ *is complete.*

Proof. By definition $\sim_n \in F_{n,1}$. By Lemma 1 there is a $g \in \overline{F_{n,1} \cup \{f\}} \cap F_{n,2}$ such that $\mathscr{R}(g) = n$ and g is not reducible. Using Lemma 5 with $p = n$ we see that $\mathbf{V}_n \in \overline{F_{n,1} \cup \{f\}}$. It is known[1] that the set $\{\mathbf{V}_n, \sim_n\}$ is complete. Clearly, if $X \subseteqq \bar{Y}$ and X is complete then Y is complete. Therefore $F_{n,1} \cup \{f\}$ is complete.

In [4] Post established a necessary and sufficient condition for a set $X \subseteqq F_2$ to be complete. In order to extend this result to $n > 2$ we use his method. This consists in constructing a finite family $\mathscr{M}_n$ of proper closed subsets of F_n satisfying the condition that every proper closed subset of F_n is included in some set of the family. The existence of such a finite family of maximal sets is an important property of the lattice of all closed subsets of F_n.

By our definition $F_{n,1}$ is closed. Moreover $F_{n,1}$ is finite, containing

[1] Post [3].

exactly n^n elements; therefore the family of all closed subsets of $F_{n,1}$ is finite. For each closed $S \subseteqq F_{n,1}$, $n \geqq 3$, we define a set $M(S)$ as follows:

(i) if $S = F_{n,1}$ then $M(S)$ is the set of all functions $f \in F_n$ such that either f is reducible or $\mathscr{R}(f)$ is a proper subset of n;

(ii) if $S \neq F_{n,1}$, then $M(S)$ is the set of all functions $f \in F_n$ satisfying the following condition: if in f we replace zero or more arguments by functions in S and then identify all arguments, the resulting function is in S.

Finally, we take as $\mathscr{M}_n$, $n \leqq 3$, the family of all sets $M(S)$ where S is any closed subset of $F_{n,1}$.[2] That this family $\mathscr{M}_n$ actually has the property mentioned above is seen from the following.

LEMMA 7. *Let S is a closed subset of $F_{n,1}$, ($n \geqq 3$). Then*

(i) *$M(S)$ is closed.*

(ii) *$M(S)$ is a proper subset of F_n.*

(iii) *$M(S) \cap F_{n,1} = S$.*

(iv) *If Y is a proper closed subset of F_n with $Y \cap F_{n,1} = S$ then $Y \subseteqq M(S)$.*

Proof. We establish Lemma 7 first for the case $S \neq F_{n,1}$. $M(S)$ is closed, since the defining property for $M(S)$ is preserved under exchange or identification of variables and also under substitution. Since S is a proper subset of $F_{n,1}$, $M(S)$ does not contain all functions of $F_{n,1}$. Therefore $M(S)$ is a proper subset of F_n. $S \subseteqq M(S) \cap F_{n,1} \subseteqq S$, hence $M(S) \cap F_{n,1} = S$. The fourth property can be verified directly from the definition of $M(S)$: Let Y be any proper closed subset of F_n with $Y \cap F_{n,1} = S$, $f \in Y$, and h a function obtained by replacing zero or more arguments of f by functions in S and then identifying all arguments. Since $S \subseteqq Y$ and Y is closed, $h \in Y$. But $Y \cap F_{n,1} \subseteqq S$ and $h \in F_{n,1}$, so $h \in S$. Thus every function $f \in Y$ is in $M(S)$. Therefore $Y \subseteqq M(S)$.

We turn now to the case $S = F_{n,1}$. That $M(F_{n,1})$ is closed follows from its definition: Both reducibility and range different from n are preserved under exchange or identification of variables. Let $f, g \in M(F_{n,1})$, h a function obtained by replacing an argument of f by the function g. If $\mathscr{R}(g) \neq n$ then $\mathscr{R}(h) \neq n$ and $h \in M(F_{n,1})$. If g is reducible, either h is reducible or $h = f$, so $h \in M(F_{n,1})$. Clearly $M(F_{n,1})$ is a proper subset of F_n since there exist functions in F_n with full range n which are not reducible. $M(F_{n,1}) \cap F_{n,1} = F_{n,1}$, since every function in $F_{n,1}$ is reducible. The proof of the fourth property follows from Lemma 6:

[2] The corresponding family for $n = 2$ contains nine elements since $F_{2,1}$ has exactly nine closed subsets. In [4] Post defined these nine sets individually: $C_4, R_2, R_3, R_9, C_2, C_3, D_3, A_1, L_1$. Our definition of $M(S)$ is directly applicable to the eight proper closed subsets of $F_{2,1}$. However, it is of interest to note that in the case $S = F_{n,1}$ the structure of $M(S)$ is essentially different for $n > 2$.

Let Y be any proper closed subset of F_n with $Y \cap F_{n,1} = F_{n,1}$. Clearly, Y cannot be complete. By Lemma 6, since $F_{n,1} \subseteqq Y$, Y cannot contain any function f, with $\mathscr{R}(f) = n$, which is not reducible. Hence $Y \subseteqq M(F_{n,1})$. This completes the proof.

Thus we see that the family $\mathscr{M}_n$ of all sets $M(S)$ where S is any closed subset of $F_{n,1}$ consists of finitely many proper closed subsets of F_n. Moreover, if X is any proper closed subset of F_n by property (iv) of Lemma 7 $X \subseteqq M(X \cap F_{n,1}) \in \mathscr{M}_n$, since $X \cap F_{n,1}$ is a closed subset of $F_{n,1}$.

We now state the main result of the note:

THEOREM 1. *A necessary and sufficient condition for a set $X \subseteqq F_n$, $n \geqq 3$, to be complete is that for every closed subset S of $F_{1,n}$ there is an $f \in X$ such that $f \notin M(S)$.*[3]

The proof of this theorem follows directly from Lemma 7. If there were any closed subset S of $F_{1,n}$ such that $X \subseteqq M(S)$ then $\bar{X} \subseteqq M(S)$ since $M(S)$ is closed and hence X would not be complete. On the other hand if for every closed subset S of $F_{1,n}$ there is an $f \in X \sim M(S)$ then $\bar{X} \nsubseteqq M(S)$ for any closed subset S of $F_{1,n}$. By Lemma 7, $\bar{X}$ cannot be a proper subset of F_n. Therefore X must be complete.

COROLLARY 1. *A set $X \subseteqq F_n$ is complete if and only if $F_{n,1} \subseteqq \bar{X}$, $n \geqq 3$, and there is an $f \in X$ such that $\mathscr{R}(f) = n$, and f is not reducible.*

Proof. If X is complete then $F_{n,1} \subseteqq \bar{X}$; and by Theorem 1, $X \nsubseteqq M(F_{n,1})$. Therefore there is an $f \in X \sim M(F_{n,1})$; i.e., $f \in X$, $\mathscr{R}(f) = n$, and f not reducible. On the other hand if $F_{n,1} \subseteqq \bar{X}$, then X is not included in $M(S)$ for any proper closed subset S of $F_{n,1}$. If, in addition, there is an $f \in X$ with $\mathscr{R}(f) = n$ and f not reducible, then $X \nsubseteqq M(F_{n,1})$. Therefore by Theorem 1, X must be complete.

We now state two further results, Theorems 2 and 3,[4] which follow easily from Theorem 1;

THEOREM 2. *There exist finite decision procedures to determine*

[3] An analogous result for $n = 2$ with the family $\mathscr{M}_n$ replaced by the set $\{D_3, C_2, C_3, A_1, L_1\}$ was obtained by Post in [**4**]. It may be noted that our theorem can be sharpened to include this result by adding the restriction: S contains the identity function and at least one other element.

[4] Yablonskiĭ in [**5**] states without proof Theorem 2, which he attributes to A. V. Kuznecov. He also attributes to Kuznecov another result which he states (again without proof) as follows: Every complete set (included in F_n) contains a finite complete subset (i.e., a finite basis). In this form the result is rather obvious and follows directly from the results of Post [**3**]; compare the first part of the proof of Corollary 2. The subsequent remarks of Yablonskiĭ make it likely, however, that Kuznecov obtained a stronger result established here as Corollary 2.

whether or not any finite subset of F_n, $n \geqq 3$, is complete and if complete whether independent.[5]

This theorem depends essentially on the computable character of our definition of the sets $M(S)$. This means that for any $f \in F_n$ and any closed subset S of $F_{n,1}$ we can tell by a finite procedure whether or not $f \in M(S)$. Therefore if X is a finite set, Theorem 1 provides a finite method for determining whether or not X is complete. This part of the proof can also be based on Corollary 1. The only thing to be shown is that all functions belonging to $F_{n,1} \cap \bar{X}$ can be obtained by means of a well determined finite procedure.

If X is complete then X is independent if and only if no proper subset of X is complete. For a finite complete X, therefore, we can determine in finitely many steps whether or not X is independent.

THEOREM 3. *For any natural number n, $n \geqq 2$, there is a natural number p such that every complete and independent subset of F_n has at most p elements.*[6]

For $n = 2$ this theorem was proved by Post in [4]. For $n \geqq 3$ it can be derived directly from Theorem 1. Let p be the number of elements in the family $\mathscr{M}_n$ of all sets $M(S)$ for S any proper closed subset of $F_{n,1}$. By Theorem 1, any set which contains an $f \notin M(S)$ for each $M(S)$ in $\mathscr{M}_n$ is complete. Thus any complete set with more than p elements would contain a proper subset which is complete.

COROLLARY 2. *For any number n, $n \geqq 2$, there is a natural number p such that every complete set included in F_n has a finite basis with at most p element.*

If X $F_n \subseteqq$ is complete, then $\bigvee_n, \sim_n \in \bar{X}$. Hence $\{\bigvee_n, \sim_n\}$ can be generated by a finite subset Y of X. Since $\{\bigvee_n, \sim_n\}$ is complete Y must be complete. Let Z be any complete independent subset of Y. Z is a finite basis of X and by Theorem 3, Z has at most p elements.

By modifying the proof of this result (and in fact making the argument independent of Theorem 1) A. Tarski has obtained the following generalization of Theorem 3.

THEOREM 4. *For any closed set $X \subseteqq F_n$ which has a finite basis there is a natural number q such that every independent basis of X has at most q elements.*

The method of proof is similar to the proof of Theorem 3. We replace the family $\mathscr{M}_n$ by a finite family $\mathscr{L}_X = \{L_0, L_1, \cdots, L_{q-1}\}$ of closed proper subsets of X with the property that for any closed proper

[5] By Post's results in [4], this theorem is also valid for $n = 2$ since the conditions defining the sets D_3, C_2, C_3, A_1 and L_1 are finitely computable.

[6] For $n = 3$, Yablonskiĭ in [5] found $p = 18$.

subset Y of X there is a set L_i, $i < q$, in $\mathscr{L}_X$ such that $Y \subseteqq L_i$. $\mathscr{L}_X$ is constructed as follows. Let B be any finite basis of X. Since B is finite there is a natural number k such that $B \subseteqq \bigcup_{i<k} F_{i,n}$. For each A satisfying

$$A \subseteqq X \cap \bigcup_{i<k} F_{n,i}\,, \quad A \neq X\,, \quad \text{and} \quad \bar{A} \cap \bigcup_{i<k} F_{n,i} = A$$

we define the set $L(A)$ in the same manner as the sets $M(S)$ were defined for $S \neq F_{n,1}$. The proof that the sets $L(A)$ are closed proper sets with the required property in the lattice all closed subsets of X is entirely analogous to the proof of Lemma 7.

For complete sets the upper bound q of Theorem 4 is much larger than the p of Theorem 3.

For $n = 2$, Post in [4] computed an upper bound $p = 5$ in Theorem 3, and then showed that actually 4 was the least upper bound. He also proved that every closed subset of F_2 has a finite basis. Therefore, two further questions arise for $n \geqq 3$:

(1) does every closed system of functions in F_n have a finite basis;

(2) (proposed by A. Tarski) is there any finite procedure to determine the least upper bound for the number of elements in any independent basis of the complete system F_n.

The solutions of these two problems have been communicated to me by A. Ehrenfeucht. He has shown that the solution of problem (2) is positive, while that of problem (1) is negative. Ehrenfeucht exhibits a very simple closed subset of F_n, $n \geqq 3$, which has no finite basis.

(*Added in proof.*) It has been communicated to me by Professor C. C. Chang that Lemma 5 was obtained by Jerzy Slupecki in "A criterion of fullness of many-valued systems of propositional logic", Comptes Rendues des séances de la Société des Sciences et des lettres de Varsovie 33, 1939, Classe III, pp 102–109. Slupecki proves the following extension of Lemma 5: *If* $F_{n,1} \subseteqq X$ *then* X *is complete if and only if there is an* $f \in X$, $f \in F_{n,2}$, f *not reducible and* $\mathscr{R}(f) = n$. Note that Lemma 6 and Corollary 1 extend this result by using Lemmas 1 and 7 to remove the condition $f \in F_{n,2}$, which is necessary for the main results of this note.

BIBLIOGRAPHY

1. Jean W. Butler, *On operations in finite algebras*, Bull. Amer. Math. Soc., **62** (1956), Abstract 552.
2. A. Ehrenfeucht, review of [**5**], Journal of Symbolic Logic 20 (1955), 175.
3. E. L. Post, *Introduction to a general theory of elementary propositions*, American Journal of Mathematics, **43** (1921), 163-185.
4. E. L. Post, *The two-valued interative systems of mathematical logic*, Princeton Annals of Math. Studies, **5** (1941).
5. S. V. Yablonskiĭ (S. V. Âblonskij), *On functional completeness in the three-valued calculus*, Doklady Akademii Nauk SSSR, **95** (1954), 1153-1155.

SEATTLE UNIVERSITY AND
UNIVERSITY OF ARIZONA

AN APPROXIMATION THEOREM FOR THE POISSON BINOMIAL DISTRIBUTION

LUCIEN LE CAM

1. Introduction. Let x_j; $j = 1, 2, \cdots$ be independent random variables such that $\text{Prob}\,(X_j = 1) = 1 - \text{Prob}\,(X_j = 0) = p_j$. Let $Q = \mathscr{L}(\Sigma X_j)$ be the distribution of their sum. This kind of distribution is often referred to as a Poisson binomial distribution. For any finite measure μ on the real line let $\|\mu\|$ be the norm defined by

$$\|\mu\| = \sup_f \left\{ \left| \int f d\mu \right| \right\} .$$

the supremum being taken over all measurable functions f such that $|f| \leqq 1$. Let $\lambda = \Sigma p_j$, let $\Sigma p_j^2 = \lambda \varpi$ and let $\alpha = \sup_j p_j$. Finally let P be the Poisson distribution whose expectation is equal to λ.

The purpose of the present paper is to show that there exist absolute constants D_1 and D_2 such that $\|Q - P\| \leqq D_1 \alpha$ for all values of the p_j's and $\|Q - P\| \leqq D_2 \varpi$ if $4\alpha \leqq 1$.

The constant D_1 is not larger than 9 and the constant D_2 is not larger than 16.

Such a result can be considered a generalization of a theorem of Yu. V. Prohorov [9] according to which such constants exist when all the probabilities p_j are equal.

The norm $\|Q - P\|$ is always larger than the maximum distance $\rho(P, Q)$ between the cumulative distributions. For this distance ρ a very general theorem of A. N. Kolmogorov [6] implies that $\rho(P, Q)$ is at most of order $\alpha^{1/5}$. The improvement obtained here is made possible by the smaller scope of our assumptions.

The method of proof used in the present paper is not quite elementary, since it uses both operator theoretic methods and characteristic functions. The relevant concepts are described in § 2.

A completely elementary approach, described in [4] leads to bounds of the order of $3\alpha^{1/3}$ for the distance ρ. Unfortunately, the elementary method does not seem to be able to provide the more precise result of the present paper.

The developments given here were prompted by discussions with J. H. Hodges, Jr. in connection with the writing of [4].

2. Measures as operators. Let $\{\mathfrak{H}, \mathfrak{A}\}$ be a measurable Abelian group, that is, an Abelian group on which a σ-field $\mathfrak{A}$ has been selected

Received December 9, 1959. The author is a research fellow of the A.P. Sloan Foundation.

in such a way that the map $(x, y) \to x + y$ from $\mathfrak{X} \times \mathfrak{X}$ to $\mathfrak{X}$ is measurable for the σ-fields $\mathfrak{A} \times \mathfrak{A}$ and $\mathfrak{A}$.

Let $\mathscr{B}$ denote the set of bounded measurable numerical functions on $\{\mathfrak{X}, \mathfrak{A}\}$. A finite signed measure μ on $\mathfrak{A}$ defines an operator, also denoted μ, from $\mathscr{B}$ to itself. To the function $f \in \mathscr{B}$ the operator μ makes correspond the element μf whose value at the point x is $(\mu f)(x) = \int f(x + \xi)\mu(d\xi)$. Linear combinations of two operators are defined by the equality

$$(\alpha\mu + \beta\nu)f = \alpha(\mu f) + \beta(\nu f) .$$

The product of two operators will be defined by composition: $(\mu\nu)f = \mu(\nu f)$. In other words,

$$[(\mu\nu)f](x) = \int \mu(dy) \int f(x + \xi + y)\nu(d\xi) .$$

It follows from Fubini's theorem that $\mu\nu = \nu\mu$. The product $\mu\nu$ corresponds to the convolution of the two measures.

For any element f of $\mathscr{B}$ let $|f|$ be the norm $|f| = \sup |f(x)|$. Define the operator norm $\|\mu\|$ by

$$\|\mu\| = \sup\{|\mu f|;\ |f| \leqq 1\} .$$

The norm $\|\mu\|$ is equal to the total mass of μ considered as a measure. It is an immediate consequence of the operator representation of $\mu\nu$ that $\|\mu\nu\| \leqq \|\mu\| \, \|\nu\|$.

Let $\mathfrak{M}$ be the system of operators obtained from all the finite signed measures. What precedes can be summarized by saying that $\mathfrak{M}$ is a normed commutative algebra having for identity the operator I which is the probability measure whose mass is entirely concentrated at the point $x = 0$. It is not difficult to show that $\mathfrak{M}$ is complete for the norm, so that $\mathfrak{M}$ is in fact a real commutative Banach algebra.

Let φ be a complex-valued function of a complex variable z. Suppose that for $|z| < a$, the function φ has a convergent power series expansion. It is then possible to define $\varphi(A)$ for every $A \in \mathfrak{M}$ such that $\|A\| < a$ by simple formal substitution in the power series expansion of φ.

The entity $\varphi(A)$ is then of the form $\varphi(A) = B + iC$ where both B and C belong to $\mathfrak{M}$. Other possible definitions can be found in [3], [2], [8]. If $\hat{\mu}$ is the Fourier transform $\hat{\mu}(t) = \int e^{itx}\,\mu(dx)$ of the measure μ then $\varphi(\mu)$ is the measure where the Fourier transform is $\varphi(\hat{\mu})$.

In most cases of statistical interest, the space $\mathfrak{X}$ is either the real line, or the additive group of integers, or the circle, or a Euclidean space. In those circumstances, as well as in the case where $\mathfrak{X}$ is an arbitrary Abelian locally compact group, we may replace $\mathscr{B}$ by the space of continuous functions which tend to zero at infinity without affecting any of the above properties.

Let M be an arbitrary finite positive measure on $\mathfrak{X}$. Then $\exp(M) = e^M = I + M + \cdots + (1/k)!\, M^k + \cdots$. It follows that $\exp[M - \|M\|I] = \exp[-\|M\|]\exp(M)$ is always a probability measure.

If a random variable X is equal to the origin of $\mathfrak{X}$ with probability $(1-p)$ the distribution $\mathscr{L}(X)$ can be written $\mathscr{L}(X) = I + p(M-1)$ where M is a probability measure.

The following theorem, essentially due to Khintchin [5] and Doeblin [1] is concerned with the distribution Q of a sum ΣX_j of independent variables having distributions $G_j = I + p_j(M_j - I)$ where M_j is a probability measure. The product $\prod_j G_j$ is always convergent when $\lambda = \sum_j p_j$ is finite. Conversely finiteness of λ is necessary to the convergence of $\prod_j G_j$ when $\mathfrak{X}$ is the additive group of integers. More generally, suppose that $\mathfrak{X}$ is the real line and that there exists an $\varepsilon > 0$ such that $\lambda_\varepsilon = \Sigma p_j M_j\{[-\varepsilon, \varepsilon]^c\} = \infty$. Then $\prod_j G_j$ cannot be convergent. This follows for instance from a result of Paul Lévy [7] according to which any interval containing the sum ΣX_j with probability $\alpha > 0$ must have a length of the order of $\varepsilon\sqrt{\lambda_\varepsilon}$.

A refinement of Paul Lévy's theorem can be found in [6], Lemma 1. However, the finiteness of λ is not generally necessary to the convergence of $\prod_j G_j$. This is quite obvious if $\mathfrak{X}$ is the circle and G_1 is the Haar measure of the circle, but the condition is not even necessary on the line.

THEOREM 1. *Let X_j; $j = 1, 2, \cdots$ be independent random variables taking their values in the measurable Abelian group $\mathfrak{X}$. Assume that $\mathscr{L}(X_j) = I + p_j(M_j - I)$ where M_j is a probability measure and assume that $\lambda = \Sigma p_j < \infty$. Let $p_j = \lambda c_j$, let $\varpi = \Sigma c_j p_j$ and finally let $M = \Sigma c_j M_j$. Then*

$$\|Q - P\| \leqq 2\lambda\varpi$$

for $P = \exp[\lambda(M - I)]$.

Proof. The proof is essentially the same as the proof of Theorem 1 in [4], given there in terms of random variables. In terms of operators one can proceed as follows.

Let $F_j = \exp p_j(M_j - I)$ and let $R_1 = \prod_{j \geqq 2} G_j$. For $k > 1$ let $R_k = (\prod_{j \leqq k-1} F_j)(\prod_{j \geqq k+1} G_j)$. Then $R_k F_k = R_{k+1} G_{k+1}$ so that

$$\prod_j G_j - \prod_j F_j = \sum_j R_j (G_j - F_j) .$$

Since R_j is a probability measure, this implies

$$\| \prod_j G_j - \prod_j F_j \| \geqq \sum_j \| G_j - F_j \| .$$

The difference $F_j - G_j$ can be written

$$F_j - G_j = [e^{-p_j} - (1 - p_j)]I + p_j(e^{-p_j} - 1)M_j + \sum_{k=2}^{\infty} \frac{e^{-p_j}}{k!} p_j^k M_j^k .$$

Hence $\| F_j - G_j \| \leqq 2p_j(1 - e^{-p_j}) \leqq 2p_j^2$.

Noting that $\prod_j F_j = \exp[\lambda(M - I)]$, this proves the desired result.

REMARK. The literature does not seem to contain any reference to the fact that Theorem 1 can be proved as in [4] and coupled with Lindeberg's proof of the normal approximation theorem to obtain a completely elementary proof of the general Central Limit theorem.

3. Sums of indicator variables and binomial distributions. In all the subsequent sections of this paper $\mathfrak{X}$ will be the additive group of integers and $\{X_j; 1, 2, \cdots\}$ will be a family of independent random variables such that $\text{Prob}(X_j = 1) = 1 - \text{Prob}(X_j = 0) = P_j$. The distribution $\mathscr{L}(X_j)$ can then be written either as $I + p_j \varDelta$ or $(1 - p_j)I + p_j H$ where $\varDelta$ is the difference operator $\varDelta = H - I$ and H is the probability measure whose mass is entirely concentrated at the point $x = 1$. The Poisson distribution whose expectation is λ can be written $P = \exp(\lambda\varDelta)$.

Letting $\lambda c_j = p_j$ and $\varpi = \Sigma c_j p_j$, Theorem 1 implies that if $Q = \mathscr{L}(\Sigma X_j)$ then the following inequality holds.

PROPOSITION 1. $\| Q - \exp(\lambda\varDelta) \| \leqq 2\lambda\varpi$.

From now on we shall assume that $\lambda < \infty$ *and that* $\alpha = \sup p_j$ *does not exceed* 1/4.

It may be expected that Q would be approximable by a binomial distribution much more closely than by a Poisson distribution. Letting $\lambda = \nu\varpi$, a binomial distribution with ν trials and probability of success ϖ can be written

$$B = (I + \varpi\varDelta)^\nu = (1 - \varpi)^\nu (I + \rho H)^\nu$$

with $\rho = \varpi/1 - \varpi$, at least when ν is an integer. If ν is not an integer the expression

$$B = (1 - \varpi)^\nu \left\{ I + \binom{\nu}{1}\rho H + \cdots + \binom{\nu}{k}\rho^k H^k + \cdots \right\}$$

where

$$\binom{\nu}{k} = \frac{1}{k!}\nu(\nu - 1)\cdots(\nu - k + 1) = \frac{\Gamma(\nu + 1)}{k!\,\Gamma(\nu - k + 1)}$$

still possesses a precise meaning as long as $\rho < 1$. However, B is not a probability measure even though $\int 1 dB = 1$. Let n be the integer such that $(n-1) < \nu \leqq n$. The coefficients $\binom{\nu}{k}$ of order $k = (n+1), (n+2)\cdots$ are alternately positive and negative.

Let $S = (1-\varpi)^{\nu} \sum_{k=n+1}^{\infty} \binom{\nu}{k} \rho^k H^k$. The norm of S is equal to

$$\| S \| = (1-\varpi)^{\nu} \sum_{k=n+1}^{\infty} \left| \binom{\nu}{k} \right| \rho^k = (1-\varpi) \left| \sum_{k=n+1}^{\infty} \binom{\nu}{k} (-\rho)^k \right| .$$

The term inside the absolute value symbol is simply the remainder of the expansion of $(1-\rho)^{\nu}$. By Taylor's formula $\| S \|$ is equal to the absolute value of

$$\frac{1}{n!} \nu(\nu-1) \cdots (\nu-n)(1-\varpi)^{\nu}(1-\rho)^{\nu} \int_0^{\rho/1-\rho} (-1)^n t^n (1+t)^{\nu-n-1} dt .$$

Therefore, since $n-1 < \nu < n$

$$\begin{aligned} \| S \| &\leqq (1-\varpi)^{\nu}(1-\rho)^{\nu} \int_0^{\rho/1-\rho} t^n (1+t)^{-1} dt \\ &\leqq \frac{1}{n+1} (1-\varpi)^{\nu}(1-\rho)^{\nu} \left(\frac{\rho}{1-\rho} \right)^{n+1} \\ &= \frac{\varpi^{n+1}}{n+1} (1-2\varpi)^{\nu-n-1} \leqq \frac{4}{\nu+1} \varpi^{\nu+1} . \end{aligned}$$

In the cases considered here $\nu = (\Sigma p_j)^2 (\Sigma p_j^2)^{-1}$ is always larger than or equal to unity. In all cases where ν is large and ϖ is small $\| S \|$ will be rather negligible.

Note that $\lambda = \nu\varpi = \int x dB$ and $\nu\varpi(1-\varpi) = \int (x-\lambda)^2 dB$. However, this last quantity may not be treated as a variance, since B possesses negative terms.

In spite of this it will be convenient to bound the remainder term

$$S(m) = (1-\varpi)^{\nu} \sum_{k=m+1}^{\infty} \binom{\nu}{k} \rho^k H^k$$

for large values of m, by Chebyshev's inequality. Assuming $\lambda < m \leqq n$ the terms $(1-\varpi)^{\nu} \binom{\nu}{k} \rho^k$ are smaller than $(1-\varpi)^{\nu-n}$ $(1-\varpi)^n \binom{n}{k} \rho^k$. Therefore

$$\| S(m) \| \leqq \frac{4\varpi^{\nu+1}}{\nu+1} + (1-\varpi)^{\nu-n} \sum_{k=m+1}^{n} (1-\varpi)^n \binom{n}{k} \rho^k .$$

Finally, by Chebyshev's inequality applied to the binomial $[1 + \varpi\Delta]^n$, one obtains

$$\| S(m) \| \leqq \frac{4\varpi^{\nu+1}}{\nu+1} + (1-\varpi)^{\nu-n} \frac{n\varpi(1-\varpi)}{[m+1-n\varpi]^2} .$$

In particular, if $m \leqq 2n\varpi < m+1$

$$\| S(m) \| \leqq \frac{4\varpi^{\nu+1}}{\nu+1} + \frac{(1-\varpi)^{1-(n-\nu)}}{n\varpi}$$

$$\leqq [4\varpi^{\nu+2}+1]\frac{1}{\lambda} .$$

To show that Q can be approximated by the Poisson distribution P in the cases where λ is too large for Proposition 1 to have any significance, we shall first show that Q can be approximated by B and then show that B is very close to P. The argument will be divided into three parts according to the values of λ and $\lambda\alpha^2$ for $\alpha^2 = \Sigma c_j(p_j - \varpi)^2$. If λ is large but $\lambda\alpha^2$ is small, bounds will be obtained through operator theoretic methods. If λ is so large that $\lambda\alpha^2$ becomes large, bounds will be obtained through computations on characteristic functions.

4. Approximations by binomial distributions. *In this section, it will be assumed throughout that $\lambda \geqq 3$ and that $\alpha \leqq 1/4$.*

For the distributions Q and B defined in the preceding section we can write

$$\begin{aligned}\log Q - \log B &= \sum_j \log(I + p_j\Delta) - \nu \log(I + \varpi\Delta) \\ &= \lambda \sum_j c_j \left\{ \frac{1}{p_j} \log(I + p_j\Delta) - \frac{1}{\varpi}\log(I+\varpi\Delta)\right\} \\ &= \lambda \sum_{k=2}^{\infty} \frac{(-1)^k}{k+1} \beta_k \Delta^{k+1} = \lambda\Delta M ,\end{aligned}$$

with

$$M = \sum_{k=2}^{\infty} \frac{(-1)^k}{k+1} \beta_k \Delta^k$$

and $\beta_k = \sum_j c_j p_j^k - \varpi^k \geqq 0$.

Since $(-1)^k\Delta^k = \sum_{s=0}^{\infty} \binom{k}{s} (-1)^s H^s$, the measure M assigns negative masses to the odd positive integers and positive masses to the even nonnegative integers.

The norm of M is precisely equal to

$$\| M \| = \sum_{k=2}^{\infty} \frac{\beta_k 2^k}{k+1} = -\sum_j c_j \left\{ \frac{1}{2p_j} \log(1-2p_j) - \frac{1}{2\varpi}\log(1-2\varpi)\right\} .$$

Letting $u = 2\varpi$ and $v_j = 2(p_j - \varpi)$ this can also be written

$$\| M \| = \int_0^1 \Sigma c_j \left\{ \left[\frac{1}{1 - t(u + v_j)} - \frac{1}{1 - tu} \right] \right\} dt .$$

Since $\Sigma c_j v_j = 0$ and $\Sigma c_j v_j^2 = 4a^2$ while

$$[1 - t(u + v_j)]^{-1} - (1 - tu)^{-1} = (1 - tu)^2 \{1 + (tv_j)[1 - t(u + v_j)]^{-1}\} tv_j$$

one can write

$$\begin{aligned} \| M \| &= \int_0^1 \left\{ \Sigma c_j \frac{v_j^2}{[1 - t(u + v_j)]} \right\} \frac{t^2}{(1 - tu)^2} dt \\ &\leqq \frac{4a^2}{1 - 2\alpha} \int_0^1 \frac{t^2}{(1 - tu)^2} dt \\ &\leqq \frac{4a^2}{3(1 - 2\alpha)} \left\{ 1 + \frac{3\varpi}{(1 - 2\varpi)^2} \right\} . \end{aligned}$$

Hence $\| M \| = ha^2$ with

$$h \leqq \frac{4}{3(1-2\alpha)} \left\{ 1 + \frac{3\varpi}{(1-2\varpi)^2} \right\} .$$

One can also write $M = \Delta M_1 = \Delta^2 M_2$ with $\| M \| = 2 \| M_1 \| = 4 \| M_2 \|$.

It results from these equalities that

$$Q = B \exp [\lambda \Delta M] .$$

For every measure μ, Taylor's formula gives

$$e^{\mu} = I + \mu \int_0^1 e^{\xi \mu} d\xi .$$

Hence

$$\begin{aligned} Q - B &= \lambda \Delta B M \int_0^1 e^{\xi \lambda \Delta M} d\xi \\ &= \lambda \Delta^2 B M_1 \int_0^1 e^{\xi \lambda \Delta M} d\xi . \end{aligned}$$

Finally

$$\| Q - B \| \leqq \lambda \| M \| \; \| \Delta B \| e^{2h\lambda a^2}$$

and

$$\| Q - B \| \leqq \frac{1}{2} \lambda \| M \| \; \| \Delta^2 B \| e^{2h\lambda a^2} .$$

One can also note that there exist probability measures F and G such that if $\varepsilon = \| M \|$ then

$$Q \exp [\lambda \varepsilon (F - I)] = B \exp [\lambda \varepsilon (G - I)] .$$

According to the foregoing expressions, to obtain bounds on $\| Q - B \|$ it will be sufficient to evaluate $\| \Delta B \|$ and $\| \Delta^2 B \|$.

Let $f(x) = \binom{\nu}{x} \varpi^x (1 - \varpi)^{\nu - x}$ and consider only values x such that $x \leqq n - 1$. In this range f achieves its maximum at a value x such that $\lambda + \varpi - 1 < x \leqq \lambda + \varpi$. It follows that $(\Delta f)(x')$ is positive for $x' \leqq x$ and negative for $x' > x$. Finally

$$\| \Delta B \| \leqq 2f(x) + \| S \| \, .$$

Let $x = \nu\xi$. An application of Stirling's formula leads to the inequality

$$\log f(x) \leqq -\frac{1}{2} \log [2\pi\nu\xi(1 - \xi)]$$

$$f(x) \leqq \frac{\theta}{\sqrt{\lambda}}$$

with

$$\theta = \frac{1}{\sqrt{2\pi}} \left[\frac{\xi}{\varpi} (1 - \xi) \right]^{-1/2} .$$

Since $\varpi(1 + 1/\nu) - 1/\nu < \xi \leqq \varpi(1 + 1/\nu)$ the quantity $\xi/\varpi(1 - \xi)$ is larger than

$$\left[1 + \frac{1}{\nu} - \frac{1}{\lambda}\right]\left[1 - \varpi\left(1 + \frac{1}{\nu}\right)\right] = \left[1 - \frac{1 - \varpi}{\lambda}\right]\left[1 - \varpi\left(1 + \frac{1}{\nu}\right)\right]$$

$$\geqq \left(1 - \frac{1}{\lambda}\right)\left[1 - \varpi\left(1 + \frac{\varpi}{\lambda}\right)\right] \geqq \frac{2}{3}\left(1 - \frac{13}{48}\right) .$$

Consequently,

$$\theta \leqq \left(\frac{72}{70\pi}\right)^{1/2}$$

and

$$\| \Delta B \| \leqq \frac{2\theta}{\sqrt{\lambda}} + \frac{4\varpi^{\nu+2}}{\lambda} .$$

Thus, we have shown the validity of the following proposition.

PROPOSITION 2. *Let $\lambda \geqq 3$ and $\alpha \leqq 1/4$, then*

$$\| Q - B \| \leqq 2ha^2\sqrt{\lambda} \exp(2h\lambda a^2)\left\{\theta + \frac{4\varpi^{\nu+2}}{\sqrt{\lambda}}\right\}$$

with

$$h \leqq \frac{4}{3}\left(\frac{1}{1-2\alpha}\right)\left[1+\frac{3\varpi}{(1-2\varpi)^2}\right] \leqq \frac{32}{3}$$

and

$$\theta \leqq \left(\frac{36}{35\pi}\right)^{1/2} \leqq \frac{1}{\sqrt{3}} .$$

A computation using the fact that $\Delta M = \Delta^2 M_1$ and the bounds for $\| \Delta^2 B \|$ can be carried out as follows.

Let $u = x + 1 - \nu\varpi$ and let $f(u)$ be the probability of $x = \nu\varpi + u - 1$ for the binomial B. Let $\delta^{-1} = \nu\varpi(1-\varpi)$ and let $\beta = \varpi\delta$ and $\gamma = (1-\varpi)\delta$. Then

$$\frac{f(u+1)}{f(u)} = \frac{1-\beta(u-1)}{1+\gamma u} .$$

The second differences of the function f for $x \leqq n$ are equal to some positive quantity multiplied by

$$g(u) = u^2 - (2\varpi - 1)u - (\nu + 2)\varpi(1-\varpi) .$$

Let r_1 and r_2, $r_1 < r_2$ be the roots of this polynomial. The second differences $(\Delta^2 f)(u)$ are negative for $u \in (r_1, r_2)$ and positive otherwise. Letting $\varphi(u) = (\Delta f)(u)$ it follows that

$$\| \Delta^2 B \| \leqq \varphi(u_1) + | \varphi(u_2) - \varphi(u_1 - 1) | + \varphi(n - \lambda + 1) - \varphi(u_2 - 1) + \frac{8}{\nu+1}\varpi^{\nu+1} .$$

The values u_i are determined by the condition that the corresponding x values, say x_1 and x_2, are respectively the largest integer not exceeding $r_1 + \lambda$ and the smallest integer as large as $r_2 + \lambda$. The roots r_1 and r_2 are given by the expression

$$r = (\varpi - 1/2) \pm \left[(\nu+1)\omega(1-\omega) + \frac{1}{4}\right]^{1/2} .$$

If $\lambda \geqq 3$ the value u_1 is negative while $u_2 - 1$ is positive.

In this case

$$\begin{aligned}\varphi(u_1) &\leqq f(u_1+1)\left[1 - \frac{1+\gamma u_1}{1-\beta(u_1-1)}\right] \\ &\leqq f(u_1+1)\delta[| u_1 | + \varpi] \\ &\leqq \frac{\theta}{\sqrt{\lambda}} || u_1 | + \varpi | \frac{1}{\lambda(1-\varpi)} .\end{aligned}$$

Similarly,

$$| \varphi(u_2) | \leqq f(u_2)\left[\frac{1-\beta(u_2-1)}{1+\gamma u_2} - 1\right]$$

$$\leqq \frac{\theta}{\lambda\sqrt{\lambda}} [| u_2 | + \varpi]\left(\frac{1}{1-\varpi}\right).$$

Note that $| u_1 - 1 | \leqq 1 + 1/2 + \sqrt{\nu\varpi(1-\varpi)} + 1/6 \leqq 5/3 + \sqrt{\lambda(1-\varpi)}$. Hence

$$\varphi(u_1 - 1) \leqq \frac{\theta}{\lambda}\left\{\frac{1}{(1-\varpi)\sqrt{\lambda}}\left[\frac{5}{\lambda} + \sqrt{\lambda(1-\varpi)}\right] + \varpi\right\}$$

$$\leqq \frac{9\theta}{4\lambda}.$$

The other terms can be bounded in a similar manner giving

$$|| \Delta^2 B || \leqq 9\frac{\theta}{\lambda} + \frac{16}{\lambda}\varpi^{\nu+2} \leqq \frac{5.4}{\lambda}.$$

Finally the following result holds.

PROPOSITION 3. *If* $\lambda \geqq 3$ *and* $\alpha \leqq 1/4$ *then*

$$|| Q - B || \leqq (2.7)h \exp [2h\lambda\alpha^2]\alpha^2$$

with $h \leqq 32/3$.

It is possible to obtain bounds on the third difference $|| \Delta^3 B ||$ by similar procedures. The algebra becomes somewhat more cumbersome. Nevertheless, it is not difficult to see that bounds of the type

$$|| Q - B || \leqq C\frac{\log \lambda}{\sqrt{\lambda}} \exp [2\lambda\alpha^2 h]\alpha^2$$

can be obtained in this manner.

The bounds given in Propositions 2 and 3 will be of value if $\lambda\alpha^2$ is small. When λ is so large that $\lambda\alpha^2$ is large, better inequalities than the preceding may be obtained through the use of Fourier transforms.

Let $\hat{\mu}$ be the Fourier transform of the measure μ. For instance $\hat{Q}(t) = \int e^{itx} Q(dx)$. Note the following inequalities.

First

$$| 1 + p(e^{it} - 1) |^2 = 1 - 2p(1-p)(1 - \cos t).$$

Hence, if $| t | \geqq \pi/2$

$$|1 + p(e^{it} - 1)|^2 \leqq 1 - 2p(1 - p)\frac{2}{\pi}|t| .$$

If $|t| \leqq \pi/2$ then

$$1 - \cos t = \frac{t^2}{2}\left[1 - \frac{t^4}{12}\cos \xi t\right]$$

with $|\xi| \leqq 1$.

Consequently, for $|t| \leqq \pi/2$

$$|1 + p(e^{it} - 1)|^2 \leqq 1 - 2p(1 - p)\frac{t^2}{2}\left(\frac{48 - \pi^2}{48}\right)$$

and for $|t| \leqq \pi/4$

$$|1 + p(e^{it} - 1)|^2 \leqq 1 - 2p(1 - p)\frac{t^2}{2}\left(\frac{192 - \pi^2}{192}\right) .$$

It follows that $|\hat{B}(t)| \leqq 1$ and

(1) For $\pi/2 \leqq |t| \leqq \pi$
$\max\{|\hat{B}(t)|, |\hat{Q}(t)|\} \leqq \exp -\{\lambda(1 - \varpi)(2/\pi)|t|\}$.

(2) For $\pi/4 \leqq |t| \leqq \pi/2$
$\max\{|\hat{B}(t)|, |\hat{Q}(t)|\} \leqq \exp[-(b^2/2)\lambda t^2]$

with $b^2 = (1 - \varpi) - \pi^2/48$.

(3) For $|t| \leqq \pi/4$
$\max\{|\hat{B}(t)|, |\hat{Q}(t)|\} \leqq \exp[-(\beta^2/2)\lambda t^2]$

with $\beta^2 = (1 - \varpi)(1 - \pi^2/192)$.

In addition, for $|t| \leqq \pi/4$ and for $z = e^{it} - 1$ one can write

$$\log \hat{Q} - \log \hat{B} = \lambda \Sigma c_j\left[\frac{1}{p_j}\log(1 + p_j z) - \frac{1}{\varpi}\log(1 + \varpi z)\right]$$

$$= -\lambda z^3\int_0^1 \frac{\xi^2}{(1 + \xi\varpi z)^2}\left[\sum_j \frac{c_j\delta_j^2}{1 + \xi p_j z}\right]d\xi$$

with $c_j = p_j/\lambda$ and $\delta_j = p_j - \varpi$.

This gives

$$|\log \hat{Q} - \log \hat{B}| \leqq \frac{1}{3}\lambda a^2|z|^3\psi(z)$$

where

$$\psi(z) = \sup_{|t| \leqq \pi/4}\ \sup_j \left|\int_0^1 \frac{3\xi^2}{[1 + \xi\omega z]^2}\ \frac{1}{(1 + \xi p_j z)}\right| d\xi .$$

Since
$$| 1 + \xi\varpi z |^2 = | (1 - \xi\varpi) + \xi\varpi e^{it} |^2$$
$$= 1 - 2\xi\varpi(1 - \xi\varpi)(1 - \cos t)$$

one has

$$| 1 + \xi\varpi z |^2 \geqq 1 - (2 - \sqrt{2})\frac{\pi}{4}\varpi .$$

Finally

$$\psi(z) \leqq \frac{1}{\sqrt{1 - \frac{\alpha}{2}}} \frac{1}{\left(1 - \frac{\varpi}{2}\right)} .$$

Hence

$$| \log \hat{Q} - \log \hat{B} | \leqq K^2 \lambda a^2 | t |^3$$

with

$$K^2 \leqq \frac{1}{3}\left(1 - \frac{\varpi}{2}\right)^{-1}\left(1 - \frac{\alpha}{2}\right)^{-1/2} .$$

It follows that, for $| t | \leqq \pi/4$ one can write

$$| \hat{Q}(t) - \hat{B}(t) | \leqq | \hat{B}(t) | \lambda a^2 K^2 | t |^3 \exp [\lambda a K^2 | t |^3]$$
$$\leqq \lambda a^2 K^2 | t |^3 \exp \left[-\frac{1}{2} \lambda \gamma^2 t^2 \right]$$

with $\gamma^2 = \beta^2 - a^2 K^2 \pi/4 \geqq 0$.

Let $V = (Q - B)$. The individual terms of V are given by the formula

$$V(k) = \frac{1}{2\pi} \int_{-\pi}^{+\pi} e^{-ikt} \hat{V}(t) dt .$$

Applying to this formula the above inequalities one obtains:

$$2\pi | V(k) | \leqq 2\lambda a^2 K^2 \int_0^\infty t^3 \sup \left[-\frac{1}{2} \lambda \gamma^2 t^2 \right] dt$$
$$+ 4 \int_{\pi/4}^\infty \exp \left[-\lambda b^2 \frac{t^2}{2} \right] dt$$
$$+ 4 \int_{\pi/2}^\infty \exp \left[-\lambda (1 - \varpi) \frac{2}{\pi} \right] dt .$$

Therefore,

$$2\pi | V(k) | \leqq \frac{4K^2 a^2}{\lambda \gamma^4} + \frac{16}{\lambda \pi b^2} \exp \left[-\frac{\lambda b^2 \pi^2}{32} \right] + \frac{2\pi}{\lambda(1 - \varpi)} \exp [-(1 - \varpi)\lambda] .$$

Noting that $xe^{-x} \leqq e^{-1}$ for $x \geqq 0$, this gives

$$2\pi\lambda \mid V(k) \mid \leqq \frac{4K^2a^2}{\gamma^4} + \left\{\frac{16\times 32}{\pi^3 cb^4} + \frac{2\pi}{(1-\varpi)^2 e}\right\}\frac{1}{\lambda} .$$

Let m be an integer such that $m \leqq 2n\varpi < m+1$ with $n-1 < \nu \leqq n$. The sum of the first m terms of $|V(k)|$ is inferior to

$$\frac{1}{\pi}\left\{\frac{4K^2a^2}{\gamma^2} + \frac{16\times 32}{\lambda\pi^3 eb^4} + \frac{2\pi}{\lambda(1-\varpi)^2 e}\right\}\left(1+\frac{1}{\nu}\right) .$$

From this and Chebyshev's inequality it follows that

$$\| Q - B \| \leqq \frac{1}{\pi}\left(1+\frac{1}{\nu}\right)\left\{\frac{4K^2a^2}{\gamma^4} + \frac{16\times 32}{\lambda\pi^3 eb^4} + \frac{2\pi}{\lambda(1-\varpi)^2 e}\right\} + \frac{(1-\varpi)}{\lambda} + \frac{1}{\lambda}[1 + 4\varpi^{\nu+2}] .$$

As a summary, one can state the following.

PROPOSITION 4. *Assume* $\lambda \geqq 3$ *and* $\alpha \leqq 1/4$. *Then, there exist constants* C_1 *and* C_2 *such that*

$$\| Q - B \| \leqq C_1 a^2 + C_2\lambda^{-1} .$$

5. Approximation of the binomial by a Poisson distribution. A theorem of Yu. V. Prohorov [9] states that the binomial $B = [I + \varpi\Delta]^2$ and the Poisson $P = \exp(\lambda\Delta)$ differ little. Explicitly, there is a constant C_0 such that $\| P - B \| \leqq C_0\varpi$.

Prohorov's result is proved in [9] only for integer values of ν. For this reason we shall give here a complete proof which happens to be somewhat simpler than Prohorov's original argument. This proof leads to an evaluation of the constant C_0 which may not be the best available but will serve our purposes.

Let $R(x)$ be the ratio of the binomial probability $B[\{x\}]$ to the Poisson probability $P[\{x\}]$

$$R(x) = \nu(\nu-1)\cdots(\nu-x+1)\varpi^x(1-\varpi)^{\nu-x}e^{\lambda}\lambda^{-2} .$$

Let us restrict ourselves to the interval $0 \leqq x \leqq n$. Since

$$\frac{R(x+1)}{R(x)} = \frac{\nu - x}{\nu(1-\varpi)}$$

the ratio R achieves in this interval a maximum at the point x such that $x - 1 \leqq \lambda < x$.

For this particular value of x, Stirling's formula leads to the inequality

$$\log R(x) \leqq -\frac{1}{2}\log(1-\xi)$$

with

$$\varpi < \xi \leqq \varpi\left(1+\frac{1}{\lambda}\right).$$

Finally for $\lambda \geqq 3$ and $4\varpi \leqq 1$,

$$\begin{aligned} R(x) &\leqq \frac{1}{\sqrt{1-\xi}} \leqq 1 + \frac{\xi}{2\sqrt{1-\xi}} \\ &\leqq 1 + \frac{1}{2}\varpi\left(1+\frac{1}{\lambda}\right)\left[1-\varpi\left(1+\frac{1}{\lambda}\right)\right]^{-1/2} \\ &\leqq 1 + \left(\frac{2}{3}\right)^{1/2}\varpi . \end{aligned}$$

Let f be a nonnegative function such that $0 \leqq f \leqq 1$. The above inequalities imply that

$$\begin{aligned} \int f dB &\leqq \frac{4\varpi^{\nu+1}}{\nu+1} + \int_{x\leqq n} R(x)f(x)P(dx) \\ &\leqq \frac{4\varpi^{\nu+1}}{\nu+1} + \left(\frac{2}{3}\right)^{1/2}\varpi\int f(x)P(dx) + \int f(x)P(dx) \\ &\leqq \int f(x)P(dx) + \varpi\left\{\left(\frac{2}{3}\right)^{1/2}\frac{4\varpi^{\nu}}{\nu+1}\right\}. \end{aligned}$$

Similarly,

$$\int(1-f)dB = 1 - \int f dB \leqq \int(1-f)dP + \varpi\left[\left(\frac{2}{3}\right)^{1/2} + \frac{4\varpi^{\nu}}{\nu+1}\right].$$

Consequently:

PROPOSITION 5. *If* $\lambda \geqq 3$ *and* $4\varpi \leqq 1$, *then*

$$\begin{aligned} \| B - P \| &\leqq 2\varpi\left[\left(\frac{2}{3}\right)^{1/2} + \frac{4\varpi^{\nu}}{\nu+1}\right] \\ &\leqq [1.64]\varpi . \end{aligned}$$

Collecting the inequalities established in the preceding sections one obtains the following statement.

THEOREM 2. *Let* $\{X_j; j = 1, 2, \cdots\}$ *be a family of independent random variables. Assume that* $\mathscr{L}(X_j) = I + p_j\Delta$ *and that* $\lambda = \Sigma p_i$ *is* *finite. Let* $p_j = \lambda c_j$ *and* $\varpi = \Sigma c_j p_j$ *and* $\alpha = \sup_j p_j$. *Denote by* Q *the distribution* $Q = \mathscr{L}(\Sigma X_j)$ *and* P *the Poisson distribution* $P = \exp(\lambda\Delta)$.

There exist constants D_1 *and* D_2 *such that*

(1) *For all values of the* p_j *one has*

$$\| P - Q \| \leqq 2\lambda\varpi$$

and

$$\| P - Q \| \leqq D_1\alpha \ .$$

(2) *If* $4\alpha \leqq 1$ *then*

$$\| P - Q \| \leqq D_2\varpi \ .$$

The constant D_1 *is inferior to* 9 *and the constant* D_2 *is inferior to* 16.

Proof. The proof of Theorem 2 consists essentially of an evaluation of the constants involved in the bounds given by Propositions 2, 3 and 4. To these propositions one must add the following remarks.

The quantity $a^2 = \Sigma c_j(p_j - \varpi)^2$ can be written

$$a^2 = \Sigma c_j\left(p_j - \frac{\alpha}{2}\right)^2 - \left(\frac{\alpha}{2} - \varpi\right)^2 .$$

Hence

$$a^2 \leqq \alpha\varpi\left(1 - \frac{\varpi}{2}\right) \leqq \left(\frac{\alpha}{2}\right)^2 .$$

In particular $a^2 \leqq \alpha\varpi$ and $a \leqq \alpha/2 \leqq 1/8$ for $\alpha \leqq 1/4$. The bound $\| Q - P \| \leqq D_1\alpha$ is operative only when $D\alpha \leqq 2$. It is therefore sufficient to prove that $\| Q - P \| \leqq D_1\alpha$ for $\alpha \leqq 2D_1^{-1}$ and $2\lambda \geqq D_1$. A constant D_1 can then be obtained through application of Proposition 2 for $\lambda a^2 \leqq y^2$ and Proposition 4 for $\lambda a^2 \geqq y^2$, the quantity y^2 being adjusted to give the best value available.

Similarly, the second inequality can be proved by use of Propositions 3 and 4, assuming $2\lambda \geqq 16$ and $\varpi \leqq 1/8$.

Note that the constants 9 and 16 are certainly much too large. For very small values of α or ϖ one can obtain much better values of D_1 and D_2.

Statement 2 of Theorem 2 implies that the approximation by a Poisson distribution will be good even though a few of the probabilities P_j may be close to the bound $\alpha \leqq 1/4$. This will happen provided only that these large values contribute relatively little to the value of λ, the bulk of λ being due to very small values of the p_j.

6. Concluding remarks.

REMARK 1. It would be highly desirable for the applications to lower the values of the coefficients D_1 and D_2 to a more reasonable level. When α is fixed, this can be achieved for D_2 by restricting the range of values of ϖ to which the inequalities apply. For instance, taking $4\alpha = 1$ but $\varpi = 10^{-2}$, the coefficient D_2 can be taken approximately equal to 8. Such a value being still too large one may inquire whether there is a lower bound to the acceptable values of D_2.

In this connection the following remarks may be of interest. When λ becomes very large the distance $(1/\varpi)\,\|\, Q - B \,\|$ becomes rapidly negligible. This can be seen for instance by using the inequalities which led to Proposition 4 and the bounds in $\alpha^2 \log \lambda/\sqrt{\lambda}$ obtained through the use of third differences.

The main contribution to $(1/\varpi)\,\|\, - P \,\|$ is then attributable to the difference between the binomial B and the Poisson measure P.

Prohorov's theorem implies that $(1/\varpi)\,\|\, B - P \,\|$ cannot be much smaller than (.483). Therefore, one cannot expect to obtain a result of the type $\|\, Q - P \,\| \leqq D_2 \varpi$ where D_2 would be substantially smaller than 1/2.

REMARK 2. The result of Theorem 1 cannot be materially improved unless one is willing to restrict further the measures M_j or the group $\mathfrak{X}$.

A slight modification of the proof given here leads to the inequality

$$\|\, Q - P \,\| \leqq 2\Big[1 - \prod_j (1 - \beta_j)\Big],$$

where β_j is taken equal to $p_j(1 - e^{-p_j})$. The bound so obtained is actually reached for certain choices of the measures M_j. An example of this can be constructed when $\mathfrak{X}$ is the real line. It is sufficient to take M_j to be the probability measure giving all its mass to a point x_j and select the values $\{x_j; j = 1, 2, \cdots\}$ to be rationally independent. For any fixed $\varepsilon > 0$ one may find values $p_j < \varepsilon$ such that $2[1 - \prod(1 - \beta_j)] > 2 - \varepsilon$ and such that $\lambda = \sum_j p_j$ be finite.

REFERENCES

1. W. Doeblin, *Sur les sommes d'un grand nombre de variables aléatoires indépendantes*, Bull. des Sciences Mathématiques, **53**, Paris (1939), 23–32.
2. Nelson Dunford and Jacob T. Schwartz, *Linear operators, Part I. General theory*, Interscience Publishers, New York, 1958.
3. Einar Hille and Ralph S. Phillips, *Functional analysis and semi groups*, Amer. Math. Soc. Coll. Publ. 31, Providence, R. I. 1957.
4. J. L. Hodges, Jr. and Lucien Le Cam, *The Poisson approximation to the Poisson*

binomial distribution, to be published in Ann. Math. Stat.

5. A. Khintchine, *Asymptotische Gesetze der Wahrscheinlichkeitsrechnung*, Ergebnisse der Mathematik und ihrer grenzgebiete, Julius Springer, Berlin, 1933.

6. A. N. Kolmogorov, *Deux théorèmes asymptotiques pour les sommes de variables aléatoires* (Russian, French summary), Teoriia Veroiatnosteii, **1** (4), Moscow (1956), 426–436.

7. Paul Lévy, *Théorie de l'addition des variables aléatoires*, Gauthier-Villars, Paris, 1937.

8. M. A. Naimark, *Normed rings*, Moscow, 1956.

9. Yu. V. Prohorov, *Asymptotic behavior of the binomial distribution* (Russian), Uspekhii Matematicheskiikh Nauk, **8** (3), Moscow (1953), 135–142.

INVOLUTIONS ON LOCALLY COMPACT RINGS

PAUL CIVIN

By a proper involution * on a ring R we mean a mapping $x \rightarrow x^*$ defined on R with the following properties:

(i) $(x+y)^* = x^* + y^*$,

(ii) $(xy)^* = y^*x^*$,

(iii) $(x^*)^* = x$ and

(iv) $xx^* = 0$ if and only if $x = 0$. If (iv) is not assumed, the mapping is simply termed an involution. If F is a field with an involution $\sharp$ and R is an algebra over F, we say that an involution on R is an algebra involution if in addition to (i)-(iv) above the following holds:

(v) $(\alpha x)^* = \alpha^\sharp x^*$ for all $x \in R$ and $\alpha \in F$.

We are concerned principally with involutions on two types of locally compact semi-simple rings, namely those which are compact or connected. The main result is that involutions on such rings are automatically continuous. As a byproduct we determine the form of any proper involution on a total matric ring R over a division ring. If in addition R is topological and the division ring admits only continuous involutions, then we note that R has only continuous involutions.

LEMMA *Let D be a division ring with center Z. Let R be a total matric ring over D. Any ring involution * on R induces an involution $\sharp$ on Z, and * is an algebra involution on R with respect to the involution $\sharp$ on Z.*

Direct calculation shows that the center of R consists of the totality of elements of the form αI where $\alpha \in Z$ and I is the identity of R. Suppose x is in the center of R and $y \in R$, then $x^*y = (y^*x)^* = (xy^*)^* = yx^*$, so x^* is in the center of R. Since $I^* = I$ is immediate, it follows that for any $\alpha \in Z$, there is a $\beta \in Z$ such that $(\alpha I)^* = \beta I$. Denote β by $\alpha^\sharp$. It is clear that $\sharp$ is an involution on Z. Moreover, if $\alpha \in Z$ and $x \in R$, $(\alpha x)^* = [(\alpha I)x]^* = x^*\alpha^\sharp I = \alpha^\sharp x$, so * is an algebra involution on R with respect to the involution $\sharp$ on Z.

THEOREM 2. *Let R be a total matric ring over D, where D is a division ring with center Z. Let * be a proper ring involution on R, and let $\sharp$ be the induced involution on Z. Then there exist a set of matrix units $\{g_{ij}\}$ in R such that $g_{ii}^* = g_{ii}$ and a set of non-zero elements γ_i of Z such that $\gamma_i^\sharp = \gamma_i$ such that the involution * has the following form: If $x = \sum \alpha_{ij}e_{ij}$, with $\alpha_{ij} \in D$, then $x^* = \sum \gamma_j^{-1}\alpha_{ij}\gamma_i e_{ji}$.*

Received August 31, 1959, and in revised form January 28, 1960. This research was supported in part by the National Science Foundation, under grant NSF-G5865.

Let e_{ij}, $i, j = 1, \cdots, n$ be a set of matric units for R. The right ideal $e_{11}R$ is minimal, so by a theorem of Rickart [7] there is a unique idempotent $u_1 \in e_{11}R$ such that $u_1^* = u_1 \neq 0$. Let $L_1 = Ru_1$, and $L_k = Re_{1k} = Re_{kk}$, $k = 2, \cdots, n$. The L_k are minimal left ideals so by the Rickart theorem there are unique idempotents $u_k \in L_k$ such that $u_k^* = u_k \neq 0$, $k = 1, \cdots, n$.

We denote by $[A, B, \cdots, C]$ the smallest left ideal containing $A, B, \cdots, C$. The linear independence of u_1 and the e_{1k}, $k = 2, \cdots, n$ implies that $L_k \not\subset [L_1, \cdots, L_{k-1}]$ for $1 < k \leqq n$. It is readily verified that $R = [L_1, \cdots, L_n]$.

Let $g_1 = u_1$ and suppose that $g_1, \cdots, g_{k-1}$ have been defined so that $g_j = g_j^2 = g_j^* \neq 0$, $g_j \in [L_1, \cdots, L_j]$ and $g_i g_j = 0$ for $i \neq j$, $i, j = 1, 2, \cdots, k-1$. We next show that g_k may be defined with the corresponding properties.

Let $v = u_k - \sum_{j=1}^{k-1} u_k g_j$. Since $L_k \not\subset [L_1, \cdots, L_{k-1}]$, $u_k \notin [L_1, \cdots, L_{k-1}]$ and thus $v \neq 0$. Since $L_k = Ru_k$ is a minimal left ideal $u_k R u_k$ is a division ring with unit u_k. The propriety of the involution then yields $vv^* \neq 0$. Since $vv^* \in u_k R u_k$, there is an element $s \in u_k R u_k$ such that $s(vv^*) = (vv^*)s = u_k$. If we apply the involution to the prior relation $(vv^*)s^* = s^*(vv^*) = u_k$, and the uniqueness of inverses in a division ring yields $s = s^*$.

It is claimed that $g_k = v^*sv$ has the desired properties. Since $vg_kv^* = vv^*svv^* = u_kvv^* = vv^* \neq 0$, it follows that $g_k \neq 0$. Clearly $g_k = g_k^*$ and $g_k^2 = v^*svv^*sv = v^*u_ksv = v^*sv = g_k$. If $i = 1, \cdots, k-1$, $g_iv^* = g_i(u_k - \sum_{j=1}^{k-1} g_j u_k) = 0$ by the inductive hypothesis, thus $g_ig_k = g_iv^*sv = 0$. By applying the involution we obtain $g_kg_i = 0$. The induction is thus complete and we may suppose that $g_1, \cdots, g_n$ have been defined.

Clearly $[g_1] = [L_1]$. Suppose that for $1 < k \leqq n$, $[g_1, \cdots, g_{k-1}] = [L_1, \cdots, L_{k-1}]$. The defining property for g_k yields $[g_1, \cdots, g_k] \subset [L_1, \cdots, L_k] = [[g_1, \cdots, g_{k-1}], L_k]$. Thus $g_k = x_1g_1 + \cdots + x_{k-1}g_{k-1} + x_ke_{1k}$. Right multiplication of the last relation by g_k shows that $x_ke_{1k} \neq 0$. Since L_k is a minimal left ideal, there is a $z \in R$ such that $zx_ke_{1k} = e_{1k}$. This may be expressed as $z[g_k - x_1g_1 \cdots - x_{k-1}g_{k-1}] = e_{1k}$. Thus $L_k \subset [g_1, \cdots, g_k]$ and hence $[g_1, \cdots, g_k] = [L_1, \cdots, L_k]$ for $k = 1, \cdots, n$. In particular $R = [g_1, \cdots, g_n]$.

The spaces Rg_k must be irreducible over R, otherwise we would have R decomposed into sums of irreducible R-spaces of different lengths. Thus the ideals Rg_k are minimal. Furthermore if we denote the unit element of R by e, we have $e = y_1g_1 + \cdots + y_ng_n$. Right multiplication by g_j shows that $g_j = y_jg_j$ and thus $e = g_1 + \cdots + g_n$.

The form of an idempotent in $e_{11}R$ and Re_{kk}, $k = 2, \cdots, n$, together with the fact that $\lambda e_{ij} = e_{ij}\lambda$ yields $\lambda u_k = u_k\lambda = u_k\lambda u_k$, $k = 1, \cdots, n$ for any $\lambda \in D$. The inductive method of defining g_k then permits one to deduce that $\lambda g_k = g_k\lambda = g_k\lambda g_k$. For suppose that $\lambda g_j = g_j\lambda$ for $j = 1, \cdots, k-1$. From the way in which v and v^* were defined $\lambda v = v\lambda$ and $\lambda v^* = v^*\lambda$. Since $\lambda g_k = v^*\lambda s v = v^* u_k \lambda s u_k v$, and $g_k\lambda = v^* s\lambda v = v^* u_k s\lambda u_k v$, it is sufficient if we show that $u_k\lambda s u_k = u_k s\lambda u_k$ for all $\lambda \in D$. But $(u_k s\lambda u_k)(vv^*) = svv^*\lambda = u_k\lambda = \lambda u_k = \lambda svv^* = u_k\lambda s u_k(vv^*)$. Since $u_k R u_k$ is a division ring, $u_k s\lambda u_k = u_k\lambda s u_k$ as desired. Hence $\lambda g_k = g_k\lambda$ for all $\lambda \in D$ and $k = 1, \cdots, n$.

Since $(0) \neq Rg_iR$ is a two sided ideal of R, $Rg_iRg_k = Rg_k \neq (0)$, and thus $g_iRg_k \neq (0)$. Suppose $i < k$, and $g_irg_k \neq 0$. Then, by the propriety of the involution, $0 \neq (g_irg_k)(g_irg_k)^* = g_irg_kr^*g_i$. Since the left ideal Rg_i is minimal, g_iRg_i is a division ring, and there exists $t \in R$ such that $(g_itg_i)(g_irg_kr^*g_i) = g_i$. If we take adjoints of the expressions in the preceding equation, we see that $g_itg_i = g_it^*g_i$. Let $g_{ik} = g_itg_irg_k$ and $g_{ki} = g_kr^*g_i$. Then $g_{ik}g_{ki} = g_i$, and consequently $(g_{ik}g_{ki})(g_{ik}g_{ki}) = g_i$, so $0 \neq g_{ki}g_{ik} \in g_kRg_k$, which is a division ring. Also $g_{ki}g_{ik}$ is idempotent so $g_{ki}g_{ik} = g_k$. Finally if we define $g_{ii} = g_i$, we obtain a set of matrix units $\{g_{ij}\}$ for R such that $g_{ii}^* = g_{ii}$. The form of the involution * on R is then an immediate consequence of a theorem of Jacobson and Rickart [2].

We are now in a position in which we may discuss the continuity of involutions.

THEOREM 3. *Let D be a topological division ring such that any involution on D is continuous. If R is a total matric ring over D, then any proper ring involution on R is continuous.*

The result is immediate by virtue of the representation of the involution given in Theorem 2, together with the fact that convergence in R, when it is regarded as a finite dimensional vector space, involves [1] convergence of the coefficients of the representation in terms of a given basis.

We turn now to locally compact semi-simple rings which are either connected or compact. The first item needed concerns their topological algebraic structure.

LEMMA. (a) *A compact semi-simple ring is the topological direct sum of total matric algebras over finite fields.*

(b) *A locally compact connected semi-simple ring is the topological direct sum of a finite number of total matric rings over locally compact division rings.*

Statement (a) is immediate from Theorem 16 of Kaplansky [4]. In the second statement, the semi-simplicity allows the use of Theorem 2 of Kaplansky [5], which shows that the ring is the direct sum of a semi-simple algebra over the reals with a unit and a totally disconnected ring. Since the decomposition is the Peirce decomposition relative to the algebra unit, it is easily seen that one has a topological direct sum. The connectedness then forces the second summand to be zero. The conclusion of the lemmas then follows from Theorem 10 of [5].

It might further be noted that the division rings involved must be connected. Consequently, since the only connected locally compact division rings are the reals, the complexes and the quaternions [3], [6], these are the only rings involved in the conclusion of (*b*).

LEMMA 5. *If * is a proper involution on a direct sum of total matric rings over division rings, then each matric ring is invariant under *. Thus * restricted to an individual matric ring is a proper involution on that ring.*

Let R be the direct sum of rings R_j. Let e° be the unit of a summand R°. Say $e^\circ = e_1 + \cdots + e_n$ is the decomposition of e° in terms of the vector units of R°. The right ideal $e_iR = e_iR^\circ$ is a minimal right ideal of R. Hence, by the theorem of Rickart used previously, there exists a unique idempotent f_k in e_iR such that $0 \neq f_i = f_i^*$. Thus $e_i = f_ie_i$ and $e_i^* = e_i^*f_i$. Consequently if $x \in R^\circ$, $x = e_1x + \cdots + e_nx = f_1e_1x + \cdots + f_ne_nx$, and $x^* = x^*e_1^*f_1 + \cdots + x^*e_n^*f_n$ is in R°.

We are now in a position to establish the continuity of proper involutions on the class of semisimple rings under discussion.

THEOREM 6. *If R is a semi-simple locally compact ring which is either compact or connected then any proper involution * on R is continuous.*

In view of Lemmas 4 and 5, it is sufficient to prove the continuity of * on an individual matric ring. Thus the proof is complete for the compact ring. For the connected ring, all we need note is that the only involutions on the reals, complexes and quaternions are automatically continuous. Hence Theorem 3 applies and the proof is complete.

REFERENCES

1. N. Bourbaki, *Eléments de mathématique, Livre V, Espaces vectoriels topologiques*, Herman et Cie, Act. Sci. et Ind. 1189, 1229, Paris 1953, 1955.

2. N. Jacobson and C. E. Rickart, *Homomorphisms of Jordan rings of self-adjoint elements*, Trans. Amer. Math. Soc., **72** (1952), 310-322.

3. N. Jacobson and O. Taussky, *Locally compact rings*, Proc. Nat. Acad., Sci. U.S.A., **21** (1935), 106-108.

4. I. Kaplansky, *Topological rings*, Amer. J. Math., **69** (1947), 153-183.

5. ———, *Locally compact rings*, Amer. J. Math., **70** (1948), 447-459.

6. Y. Otobe, *Note on locally compact simple rings*, Proc. Imp. Acad. Tokyo, **20** (1944), 283.

7. C. E. Rickart, *Representation of certain Bancach algebras*, Duke Math. J., **18** (1951), 27-39.

UNIVERSITY OF OREGON

NORMAL EXTENSIONS OF FORMALLY NORMAL OPERATORS

EARL A. CODDINGTON

1. Introduction. Let $\mathfrak{H}$ be a Hilbert space. If T is any operator in $\mathfrak{H}$ its domain will be denoted by $\mathfrak{D}(T)$, its null space by $\mathfrak{N}(T)$. A *formally normal* operator N in $\mathfrak{H}$ is a densely defined closed operator such that $\mathfrak{D}(N) \subset \mathfrak{D}(N^*)$, and $\|Nf\| = \|N^*f\|$ for all $f \in \mathfrak{D}(N)$. Intimately associated with such an N is the operator $\bar{N}$ which is the restriction of N^* to $\mathfrak{D}(N)$. The operator N is formally normal if and only if $\bar{N}$ is. A *normal operator* N in $\mathfrak{H}$ is a formally normal operator for which $\mathfrak{D}(N) = \mathfrak{D}(N^*)$; in this case $\bar{N} = N^*$. A densely defined closed operator N is normal if and only if $N^*N = NN^*$.[1]

Let N be formally normal in $\mathfrak{H}$. Since $\bar{N} \subset N^*$ we have $N \subset \bar{N}^*$, where $\bar{N}^* = (\bar{N})^*$. Thus we see that a closed symmetric operator is a formally normal operator such that $N = \bar{N}$, and a self-adjoint operator is a normal operator such that $N = \bar{N}$ $(= N^*)$. If a closed symmetric operator has a normal extension in $\mathfrak{H}$, this extension is self-adjoint. It is known that a closed symmetric operator may not have a self-adjoint extension in $\mathfrak{H}$. Necessary and sufficient conditions for such extensions were given by von Neumann.[2] However, until recently, conditions under which a formally normal operator N can be extended to a normal one in $\mathfrak{H}$ were known only for certain special cases.[3,4] Kilpi[5] considered the problem in terms of the real and imaginary parts of N. It is the purpose of this note to characterize the normal extensions of N in a manner similar to the von Neumann solution for the symmetric case.

If N_1 is a normal extension of a formally normal operator N in $\mathfrak{H}$, then it is easy to see that $N \subset N_1 \subset \bar{N}^*$, and $\bar{N} \subset N_1^* \subset N^*$. In Theorem 1 we describe $\mathfrak{D}(\bar{N}^*)$ and $\mathfrak{D}(N^*)$ for any two operators $N, \bar{N}$ satisfying $N \subset \bar{N}^*$, $\bar{N} \subset N^*$. With the aid of this result a characterization of the normal extensions N_1 of a formally normal N in $\mathfrak{H}$ is given in Theorem 2. It is indicated in Theorem 3 how the domains of normal extensions

Received January 13, 1960. This work was supported in part by the National Science Foundation and the Office of Naval Research.

[1] See, e.g., B. v. Sz. Nagy, *Spektraldarstellung linearer Transformationen des Hilbertschen Raumes*, Ergeb. Math., **5** (1942), 33.

[2] Ibid; p. 39.

[3] Y. Kilpi, "Über lineare normale Transformationen im Hilbertschen Raum", Annales Academiae Scientiarum Fennicae, Series A-I, No. **154** (1953).

[4] R. H. Davis, "Singular normal differential operators", Technical Report No. 10, Department of Mathematics, University of California, Berkeley, Calif., (1955).

[5] Y. Kilpi, "Über das komplexe Momentenproblem", Annales Academiae Scientiarum Fennicae, Series A-I, No. **236** (1957).

can be described by abstract boundary conditions.

I would like to thank Ralph Phillips for instructive conversations during this work.

2. Domains.

THEOREM 1. *Let N, $\bar{N}$ be two closed densely defined operators in a Hilbert space $\mathfrak{H}$ such that $N \subset \bar{N}^*$, $\bar{N} \subset N^*$. Then*

$$\mathfrak{D}(\bar{N}^*) = \mathfrak{D}(N) + \mathfrak{M}, \qquad \mathfrak{D}(N^*) = \mathfrak{D}(\bar{N}) + \bar{\mathfrak{M}},$$

where $\mathfrak{M} = \mathfrak{N}(I + N^*\bar{N}^*)$, $\bar{\mathfrak{M}} = \mathfrak{N}(I + \bar{N}^*N^*)$. Here I is the identity operator, and the sums are direct sums.

Proof. Let $N, \bar{N}$ be any two closed densely defined operators in $\mathfrak{H}$ such that $N \subset \bar{N}^*$, $\bar{N} \subset N^*$. Then $(Nf, g) = (f, \bar{N}g)$ for all $f \in \mathfrak{D}(N)$, $g \in \mathfrak{D}(\bar{N})$. Define an operator $\mathscr{N}$ in the Hilbert space $\mathfrak{H}_2 = \mathfrak{H} \oplus \mathfrak{H}$ with domain $\mathfrak{D}(\mathscr{N})$ the set of all $\hat{f} = \{f_1, f_2\}$ with $f_1 \in \mathfrak{D}(N)$, $f_2 \in \mathfrak{D}(\bar{N})$, and such that $\mathscr{N}\hat{f} = \{\bar{N}f_2, Nf_1\}$. Then $\mathscr{N}$ is closed symmetric. Indeed $\mathfrak{D}(\mathscr{N})$ is dense in $\mathfrak{H} \oplus \mathfrak{H}$, and, if $\hat{f} = \{f_1, f_2\}$, $\hat{g} = \{g_1, g_2\}$ are in $\mathfrak{D}(\mathscr{N})$, we have

$$(\mathscr{N}\hat{f}, \hat{g}) = (\bar{N}f_2, g_1) + (Nf_1, g_2) = (f_1, \bar{N}g_2) + (f_2, Ng_1) = (\hat{f}, \mathscr{N}\hat{g}).$$

Since N and $\bar{N}$ are closed, so is $\mathscr{N}$. The adjoint $\mathscr{N}^*$ of $\mathscr{N}$ has domain $\mathfrak{D}(\mathscr{N}^*)$ the set of all $\hat{g} = \{g_1, g_2\}$ such that $g_1 \in \mathfrak{D}(\bar{N}^*)$, $g_2 \in \mathfrak{D}(N^*)$; and $\mathscr{N}^*\hat{g} = \{N^*g_2, \bar{N}^*g_1\}$.

We now show that the defect spaces of $\mathscr{N}$, namely,

$$\mathfrak{E}(+i) = \{\hat{\phi} \in \mathfrak{D}(\mathscr{N}^*) : \mathscr{N}^*\hat{\phi} = i\hat{\phi}\},$$
$$\mathfrak{E}(-i) = \{\hat{\psi} \in \mathfrak{D}(\mathscr{N}^*) : \mathscr{N}^*\hat{\psi} = -i\hat{\psi}\},$$

have the same dimension. We have $\hat{\phi} = \{\phi_1, \phi_2\} \in \mathfrak{E}(+i)$ if and only if $\phi_1 \in \mathfrak{D}(\bar{N}^*)$, $\phi_2 \in \mathfrak{D}(N^*)$, $N^*\phi_2 = i\phi_1$, $\bar{N}^*\phi_1 = i\phi_2$. The latter is true if and only if $N^*(-\phi_2) = -i\phi_1$, $\bar{N}^*\phi_1 = -i(-\phi_2)$. Thus we see that the unitary map $\mathscr{U}$ of $\mathfrak{H}_2$ onto itself given by $\mathscr{U}\{f_1, f_2\} = \{f_1, -f_2\}$ carries $\mathfrak{E}(-i)$ onto $\mathfrak{E}(+i)$ in an isometric way. This proves $\dim \mathfrak{E}(+i) = \dim \mathfrak{E}(-i)$.

We note that $\{\phi_1, \phi_2\} \in \mathfrak{E}(+i)$ if and only if $\phi_1 \in \mathfrak{D}(N^*\bar{N}^*)$, $(I + N^*\bar{N}^*)\phi_1 = 0$, and $\phi_2 = -i\bar{N}^*\phi_1$. Alternatively $\{\phi_1, \phi_2\} \in \mathfrak{E}(+i)$ if and only of $\phi_2 \in \mathfrak{D}(\bar{N}^*N^*)$, $(I + \bar{N}^*N^*)\phi_2 = 0$, and $\phi_1 = -iN^*\phi_2$. Thus we see that the algebraic dimensions of the spaces $\mathfrak{M} = \mathfrak{N}(I + N^*\bar{N}^*)$, $\bar{\mathfrak{M}} = \mathfrak{N}(I + \bar{N}^*N^*)$, $\mathfrak{E}(+i)$, and $\mathfrak{E}(-i)$ are all the same. Further it is easy to see that $\bar{N}^*$ maps $\mathfrak{M}$ one-to-one onto $\bar{\mathfrak{M}}$, the inverse mapping being $-N^*$ restricted to $\bar{\mathfrak{M}}$.

Since $\dim \mathfrak{E}(+i) = \dim \mathfrak{E}(-i)$ the operator $\mathscr{N}$ has self-adjoint extensions in $\mathfrak{H}_2$. They are in a one-to-one correspondence with the isometries of $\mathfrak{E}(-i)$ onto $\mathfrak{E}(+i)$. If $\mathscr{S}$ is a self-adjoint extension of $\mathscr{N}$ there is a unique isometry $\mathscr{V}$ of $\mathfrak{E}(-i)$ onto $\mathfrak{E}(+i)$ such that $\mathfrak{D}(\mathscr{S}) = \mathfrak{D}(\mathscr{N}) + (\mathscr{I} - \mathscr{V})\mathfrak{E}(-i)$, where $\mathscr{I}$ is the identity operator on $\mathfrak{H}_2$. Let us consider that self-adjoint extension $\mathscr{S}$ of $\mathscr{N}$ determined in this way by the isometry $-\mathscr{U}$ restricted to $\mathfrak{E}(-i)$. Then we have $\hat{h} \in \mathfrak{D}(\mathscr{S})$ if and only if $\hat{h} = \hat{f} + \hat{\psi} + \mathscr{U}\hat{\psi}$, for some $\hat{f} \in \mathfrak{D}(\mathscr{N})$, $\hat{\psi} \in \mathfrak{E}(-i)$. If $\hat{h} = \{h_1, h_2\}$, $\hat{f} = \{f_1, f_2\}$, $\hat{\psi} = \{\psi_1, \psi_2\}$, this means $h_1 = f_1 + 2\psi_1$, $h_2 = f_2$, where $f_1 \in \mathfrak{D}(N)$, $\psi_1 \in \mathfrak{M}$, $f_2 \in \mathfrak{D}(\bar{N})$. Thus $\mathfrak{D}(\mathscr{S})$ is the set of all $\{h_1, h_2\}$ with $h_1 \in \mathfrak{D}(N) + \mathfrak{M}$, $h_2 \in \mathfrak{D}(\bar{N})$. Now the operator $\mathscr{S}_1$ with domain all $\{h_1, h_2\}$ with $h_1 \in \mathfrak{D}(\bar{N}^*)$, $h_2 \in \mathfrak{D}(\bar{N})$, and such that $\mathscr{S}_1\{h_1, h_2\} = \{\bar{N}h_2, \bar{N}^*h_1\}$, is readily seen to be a self-adjoint operator in $\mathfrak{H}_2$ satisfying $\mathscr{N} \subset \mathscr{S} \subset \mathscr{S}_1 \subset N^*$. Hence $\mathscr{S} = \mathscr{S}_1$, and we see that $\mathfrak{D}(\bar{N}^*) = \mathfrak{D}(N) + \mathfrak{M}$. The sum is a direct one, for if $f \in \mathfrak{D}(N) \cap \mathfrak{M}$, $0 = (I + N^*\bar{N}^*)f = f + N^*Nf$ implying $0 = (f + N^*Nf, f) = \|f\|^2 + \|Nf\|^2$, or $f = 0$.

A similar argument shows that the self-adjoint extension $\mathscr{S}$ of $\mathscr{N}$ determined by the isometry $\mathscr{V}$ equal to $\mathscr{U}$ restricted to $\mathfrak{E}(-i)$ has domain the set of all $\{h_1, h_2\}$ with $h_1 \in \mathfrak{D}(N)$, $h_2 \in \mathfrak{D}(\bar{N}) + \bar{\mathfrak{M}}$. This operator is equal to the self-adjoint extension of $\mathscr{N}$ having domain the set of all $\{h_1, h_2\}$ with $h_1 \in \mathfrak{D}(N)$, $h_2 \in \mathfrak{D}(N^*)$, implying that $\mathfrak{D}(N^*) = \mathfrak{D}(\bar{N}) + \bar{\mathfrak{M}}$, a direct sum. This completes the proof of Theorem 1.

Note added in proof. The results of Theorem 1 can be obtained more directly, although some of the discussion given in the proof above is required for our proof of Theorem 2. Let $\mathfrak{G}(T)$ denote the graph of an operator T. If A, B are any two closed operators with dense domain, and $A \subset B$, then it is easy to see that $\mathfrak{G}(B) \ominus \mathfrak{G}(A)$ is the set of all $\{u, Bu\} \in \mathfrak{G}(B)$ such that $u \in \mathfrak{N}(I + A^*B)$. Since

$$\mathfrak{G}(B) = \mathfrak{G}(A) \oplus [\mathfrak{G}(B) \ominus \mathfrak{G}(A)],$$

we have $\mathfrak{D}(B) = \mathfrak{D}(A) + \mathfrak{N}(I + A^*B)$, a direct sum. This implies Theorem 1.

3. Normal extensions.

THEOREM 2. *If N_1 is a normal extension of a formally normal operator N in a Hilbert space $\mathfrak{H}$, then there exists a unique linear map W of $\mathfrak{M}$ onto itself satisfying*

(i) $W^2 = I$,

(ii) $\|\phi\|^2 + \|\bar{N}^*\phi\|^2 = \|W\phi\|^2 + \|\bar{N}^*W\phi\|^2$, $(\phi \in \mathfrak{M})$,

(iii) $(I - W)\mathfrak{M} = \bar{N}^*(I + W)\mathfrak{M}$,

(iv) $\|\bar{N}^*(I - W)\phi\| = \|N^*(I - W)\phi\|$, $(\phi \in \mathfrak{M})$.

In terms of W we have

$$\mathfrak{D}(N_1) = \mathfrak{D}(N) + (I - W)\mathfrak{M}, \quad N_1 f = \bar{N}^* f, \qquad (f \in \mathfrak{D}(N_1)). \tag{1}$$

Conversely, if W is any linear map of $\mathfrak{M}$ onto $\mathfrak{M}$ satisfying (i)—(iv) *above, then the operator N_1 defined by* (1) *is a normal extension of N in $\mathfrak{H}$.*

REMARKS. Condition (i) implies that $P_1 = (1/2)(I + W)$ and $P_2 = (1/2)(I - W)$ are projections (not necessarily orthogonal) in $\mathfrak{M}$, and $\mathfrak{M}$ is the direct sum of $\mathfrak{M}_1 = P_1\mathfrak{M}$ and $\mathfrak{M}_2 = P_2\mathfrak{M}$. If $\phi \in \mathfrak{M}$, then $\phi \in \mathfrak{M}_1$ if and only if $W\phi = \phi$, and $\phi \in \mathfrak{M}_2$ if and only if $W\phi = -\phi$.

Condition (ii) implies that if $\phi, \phi' \in \mathfrak{M}$ then

$$(\phi, \phi') + (\bar{N}^*\phi, \bar{N}^*\phi') = (W\phi, W\phi') + (\bar{N}^* W\phi, \bar{N}^* W\phi') .$$

If $\phi \in \mathfrak{M}_1$, $\phi' \in \mathfrak{M}_2$ we see that $(\phi, \phi') + (\bar{N}^*\phi, \bar{N}^*\phi') = 0$, which means that the graph of $\bar{N}^*$ restricted to $\mathfrak{M}_1$ is orthogonal to the graph of $\bar{N}^*$ restricted to $\mathfrak{M}_2$.

Since $\bar{N}^*$ is one-to-one from $\mathfrak{M}$ onto $\bar{\mathfrak{M}}$, condition (iii) implies that $\mathfrak{M}_2 = \bar{N}^*\mathfrak{M}_1 \subset \mathfrak{M} \cap \bar{\mathfrak{M}}$, and $\mathfrak{M}_2$ has the same algebraic dimension as $\mathfrak{M}_1$. In particular the dimension of $\mathfrak{M}$ must be even.

Proof of Theorem 2. Let N_1 be a normal extension of the formally normal operator N in $\mathfrak{H}$. Then we have $N \subset N_1 \subset \bar{N}^*$, $\bar{N} \subset N_1^* \subset N^*$. Let the operator $\mathscr{N}_1$ in $\mathfrak{H}_2$ be defined with domain all $\{h_1, h_2\}$ such that $h_1 \in \mathfrak{D}(N_1)$, $h_2 \in \mathfrak{D}(N_1^*)$, and so that $\mathscr{N}_1\{h_1, h_2\} = \{N_1^*h_2, N_1h_1\}$. Then it is easily seen that $\mathscr{N}_1$ is a self-adjoint extension of the operator $\mathscr{N}$ defined in the proof of Theorem 1.

Let $\mathscr{N}_1$ be *any* self-adjoint extension of $\mathscr{N}$, and let $\mathscr{V}$ be the unique isometry of $\mathfrak{E}(-i)$ onto $\mathfrak{E}(+i)$ such that $\mathfrak{D}(\mathscr{N}_1) = \mathfrak{D}(\mathscr{N}) + (\mathscr{I} - \mathscr{V})\mathfrak{E}(-i)$. Then we may write $\mathscr{V} = \mathscr{W}\mathscr{U}$, where $\mathscr{U}$ is the isometry defined on $\mathfrak{E}(-i)$ to $\mathfrak{E}(+i)$ by $\mathscr{U}\{\psi_1, \psi_2\} = \{\psi_1, -\psi_2\}$, and $\mathscr{W}$ is a unitary map of $\mathfrak{E}(+i)$ onto itself. For $\{\phi_1, \phi_2\} \in \mathfrak{E}(+i)$ let $\mathscr{W}\{\phi_1, \phi_2\} = \{\chi_1, \chi_2\}$. Then $\phi_1, \chi_1 \in \mathfrak{M}$ and $\phi_2 = -i\bar{N}^*\phi_1$, $\chi_2 = -i\bar{N}^*\chi_1$. Define the map W of $\mathfrak{M}$ into $\mathfrak{M}$ by $W\phi_1 = \chi_1$. Then W is linear, and since $\mathscr{W}$ is unitary, W is onto, and

$$\| \{\phi, -i\bar{N}^*\phi\} \|^2 = \| \{W\phi, -i\bar{N}^* W\phi\} \|^2 , \qquad (\phi \in \mathfrak{M}) ,$$

or

$$\| \phi \|^2 + \| \bar{N}^*\phi \|^2 = \| W\phi \|^2 + \| \bar{N}^* W\phi \|^2 , \qquad (\phi \in \mathfrak{M}) . \tag{2}$$

Conversely, suppose W is a linear map of $\mathfrak{M}$ onto $\mathfrak{M}$ satisfying (2). Then for $\hat{\phi} = \{\phi, -i\bar{N}^*\phi\} \in \mathfrak{E}(+i)$ define $\mathscr{W}\hat{\phi} = \{W\phi, -i\bar{N}^* W\phi\}$. Then $\mathscr{W}$ maps $\mathfrak{E}(+i)$ onto $\mathfrak{E}(+i)$ and (2) implies that $\mathscr{W}$ is unitary. Thus we see that the self-adjoint extensions $\mathscr{N}_1$ of $\mathscr{N}$ are in a one-to-one correspondence with the linear maps W of $\mathfrak{M}$ onto $\mathfrak{M}$ satisfying (2). We have $\hat{h} = \{h_1, h_2\} \in \mathfrak{D}(\mathscr{N}_1)$ if and only if $\hat{h}$ can be represented in the form $\hat{h} = \hat{f} + (\mathscr{I} - \mathscr{W}\mathscr{U})\hat{\psi}$, where $\hat{f} = \{f_1, f_2\} \in \mathfrak{D}(\mathscr{N})$, $\hat{\psi} = \{\phi, i\bar{N}^*\phi\} \in \mathfrak{E}(-i)$. This means $h_1 = f_1 + (I - W)\phi$, $h_2 = f_2 + i\bar{N}^*(I + W)\phi$, where $f_1 \in \mathfrak{D}(\mathscr{N})$, $f_2 \in \mathfrak{D}(\bar{N})$, $\phi \in \mathfrak{M}$.

The self-adjoint extension $\mathscr{N}_1$ arising from the normal extension N_1 of N has the property that if $\hat{h} = \{h_1, h_2\} \in \mathfrak{D}(\mathscr{N}_1)$ then so does $\mathscr{P}_1\hat{h} = \{h_1, 0\}$. It will now be shown that a self-adjoint extension $\mathscr{N}_1$ of $\mathscr{N}$ has this property if and only if the W corresponding to $\mathscr{N}_1$ satisfies $W^2 = I$. First suppose $\mathscr{P}_1\hat{h} \in \mathfrak{D}(\mathscr{N}_1)$ for all $\hat{h} \in \mathfrak{D}(\mathscr{N}_1)$. Letting $h_1 = f_1 + (I - W)\phi$, $h_2 = f_2 + i\bar{N}^*(I + W)\phi$ as above, we see that this implies that there exist elements $f_1' \in \mathfrak{D}(N)$, $f_2' \in \mathfrak{D}(\bar{N})$, $\phi' \in \mathfrak{M}$, such that

$$\begin{aligned} f_1 + (I - W)\phi &= f_1' + (I - W)\phi' , \\ 0 &= f_2' + i\bar{N}^*(I + W)\phi' . \end{aligned}$$

Since $\mathfrak{D}(N) + \mathfrak{M}$ and $\mathfrak{D}(\bar{N}) + \bar{\mathfrak{M}}$ are direct sums these equations imply that $f_1 = f_1'$, $(I - W)\phi = (I - W)\phi'$, $f_2' = 0$, and $\bar{N}^*(I + W)\phi' = 0$. The last equation implies $(I + W)\phi' = 0$ since $\bar{N}^*$ is one-to-one from $\mathfrak{M}$ to $\bar{\mathfrak{M}}$. Thus we have

$$\begin{aligned} \phi' + W\phi' &= 0 , \\ \phi' - W\phi' &= \phi - W\phi , \end{aligned} \tag{3}$$

from which results $2\phi' = (I - W)\phi$. Returning to the first equation in (3) we obtain $(I + W)(I - W)\phi = (I - W^2)\phi = 0$ for all $\phi \in \mathfrak{M}$, showing that $W^2 = I$. Conversely, suppose $W^2 = I$ on $\mathfrak{M}$. Then if $\hat{h} = \{h_1, h_2\} \in \mathfrak{D}(\mathscr{N}_1)$, $h_1 = f_1 + (I - W)\phi$, $h_2 = f_2 + i\bar{N}^*(I + W)\phi$, define $\phi' = (1/2)(I - W)\phi$. Then equations (3) will be valid, implying that

$$\begin{aligned} f_1 + (I - W)\phi &= f_1 + (I - W)\phi' , \\ 0 &= 0 + i\bar{N}^*(I + W)\phi' , \end{aligned}$$

which shows that $\mathscr{P}_1\hat{h} = \{h_1, 0\} \in \mathfrak{D}(\mathscr{N}_1)$.

If $\mathscr{N}_1$ is any self-adjoint extension of $\mathscr{N}$ for which $W^2 = I$, then $\mathfrak{D}(\mathscr{N}_1)$ consists of those $\{h_1, h_2\}$ such that $h_1 = f_1 + (I - W)\phi$, $h_2 = f_2 + i\bar{N}^*(I + W)\phi'$, for some $f_1 \in \mathfrak{D}(N)$, $f_2 \in \mathfrak{D}(\bar{N})$, and $\phi, \phi' \in \mathfrak{M}$. The point is that ϕ and ϕ' need not now be the same element. Indeed, if h_1, h_2 have such representations let $\phi'' = (1/2)(I - W)\phi + (1/2)(I + W)\phi'$. Then $(I - W)\phi = (I - W)\phi''$, and $(I + W)\phi' = (I + W)\phi''$, which implies that $\{h_1, h_2\} \in \mathfrak{D}(\mathscr{N}_1)$. For such an $\mathscr{N}_1$ define N_1 to be the operator in $\mathfrak{H}$ with $\mathfrak{D}(N_1) = \mathfrak{D}(N) + (I - W)\mathfrak{M}$, and $N_1 h_1 = \bar{N}^* h_1$ for $h_1 \in \mathfrak{D}(N_1)$. Similarly define N_2 on $\mathfrak{D}(N_2) = \mathfrak{D}(\bar{N}) + \bar{N}^*(I + W)\mathfrak{M}$ by $N_2 h_2 = N^* h_2$ for $h_2 \in \mathfrak{D}(N_2)$. In terms of N_1 and N_2 we have $\{h_1, h_2\} \in \mathfrak{D}(\mathscr{N}_1)$ if and only if $h_1 \in \mathfrak{D}(N_1)$, $h_2 \in \mathfrak{D}(N_2)$, and $\mathscr{N}_1\{h_1, h_2\} = \{N_2 h_2, N_1 h_1\}$. A short computation shows that $\mathfrak{D}(\mathscr{N}_1^*)$ is the set of all $\{g_1, g_2\}$ such that $g_1 \in \mathfrak{D}(N_2^*)$,

$g_2 \in \mathfrak{D}(N_1^*)$, and $\mathscr{N}_1^*\{g_1, g_2\} = \{N_1^* g_2, N_2^* g_1\}$. But since $\mathscr{N}_1 = \mathscr{N}_1^*$ we obtain $N_2 = N_1^*$. Hence $\mathfrak{D}(\mathscr{N}_1)$ consists of all $\{h_1, h_2\}$ with $h_1 \in \mathfrak{D}(N_1)$, $h_2 \in \mathfrak{D}(N_1^*)$, and $\mathscr{N}_1\{h_1, h_2\} = \{N_1^* h_2, N_1 h_1\}$. Here

$$\mathfrak{D}(N_1) = \mathfrak{D}(N) + (I - W)\mathfrak{M}\,, \qquad \mathfrak{D}(N_1^*) = \mathfrak{D}(\bar{N}) + \bar{N}^*(I + W)\mathfrak{M}\,, \tag{4}$$

and $N \subset N_1 \subset \bar{N}^*$, $\bar{N} \subset N_1^* \subset N^*$. Thus any self-adjoint extension $\mathscr{N}_1$ of $\mathscr{N}$ having the property that $W^2 = I$ determines a unique operator N_1 in $\mathfrak{H}$ as above, which is easily seen to be closed. In particular, if N_1 is a normal extension of N, then the equalities (4) hold.

It remains to characterize those $\mathscr{N}_1$ such that $W^2 = I$ for which N_1 is normal, that is $\mathfrak{D}(N_1) = \mathfrak{D}(N_1^*)$ and $\| N_1 h \| = \| N_1^* h \|$, $h \in \mathfrak{D}(N_1)$. We claim that this is true if and only if

$$(I - W)\mathfrak{M} = \bar{N}^*(I + W)\mathfrak{M}\,, \tag{5}$$

and

$$\| \bar{N}^*(I - W)\phi \| = \| N^*(I - W)\phi \|\,, \qquad (\phi \in \mathfrak{M})\,. \tag{6}$$

If (5) is valid then (4) implies that $\mathfrak{D}(N_1) = \mathfrak{D}(N_1^*)$, since $\mathfrak{D}(N) = \mathfrak{D}(\bar{N})$. Let $h \in \mathfrak{D}(N_1)$, $h = f + (I - W)\phi$, $f \in \mathfrak{D}(N)$, $\phi \in \mathfrak{M}$. Then $(I - W)\phi \in \mathfrak{M} \cap \bar{\mathfrak{M}}$, and we have $N_1 h = Nf + \bar{N}^*(I - W)\phi$, $N_1^* h = \bar{N}f + N^*(I - W)\phi$. Thus

$$\| N_1 h \|^2 = \| Nf \|^2 + (Nf, \bar{N}^*(I - W)\phi) + (\bar{N}^*(I - W)\phi, Nf) + \| \bar{N}^*(I - W)\phi \|^2\,,$$

and

$$\| N_1^* h \|^2 = \| \bar{N}f \|^2 + (\bar{N}f, N^*(I - W)\phi) + (N^*(I - W)\phi, \bar{N}f) + \| N^*(I - W)\phi \|^2\,.$$

Since N is formally normal $\| Nf \| = \| \bar{N}f \|$. Moreover $\bar{N}^*(I - W)\phi \in \bar{\mathfrak{M}}$ implies that $(Nf, \bar{N}^*(I - W)\phi) = (f, N^*\bar{N}^*(I - W)\phi) = -(f, (I - W)\phi)$, and similarly $(\bar{N}f, N^*(I - W)\phi) = -(f, (I - W)\phi)$. Using (6) we see that $\| N_1 h \| = \| N_1^* h \|$ for all $h \in \mathfrak{D}(N_1)$, proving that N_1 is normal.

Conversely, suppose N_1 is normal. Then (6) is clearly valid, for $(I - W)\phi \in \mathfrak{D}(N_1)$ by (4). Suppose $h \in \mathfrak{D}(N_1) = \mathfrak{D}(N_1^*)$ and $h = f + (I - W)\phi = f' + \bar{N}^*(I + W)\phi'$ with $f, f' \in \mathfrak{D}(N)$, $\phi, \phi' \in \mathfrak{M}$. We show that $f = f'$ and $(I - W)\phi = \bar{N}^*(I + W)\phi'$. Applying this to $f = 0$ we obtain $(I - W)\mathfrak{M} \subset \bar{N}^*(I + W)\mathfrak{M}$, and with $f' = 0$ we get $\bar{N}^*(I + W)\mathfrak{M} \subset (I - W)\mathfrak{M}$, proving (5). Now for any $g \in \mathfrak{D}(N)$ we have $(N_1 h, N_1 g) = (N_1^* h, N_1^* g)$, or

$$(Nf, Ng) + (\bar{N}^*(I - W)\phi, Ng) = (\bar{N}f', \bar{N}g) - ((I + W)\phi', \bar{N}g)\,.$$

Since $(\bar{N}f', \bar{N}g) = (Nf', Ng)$ and $(\bar{N}^*(I - W)\phi, Ng) = -((I - W)\phi, g)$, this yields

$$(Nf, Ng) - ((I - W)\phi, g) = (Nf', Ng) - (\bar{N}^*(I + W)\phi', g) ,$$

or

$$(N(f - f'), Ng) + (\bar{N}^*(I + W)\phi' - (I - W)\phi, g) = 0 .$$

But $\bar{N}^*(I + W)\phi' - (I - W)\phi = f - f'$, and hence

$$(N(f - f'), Ng) + (f - f', g) = 0$$

for all $g \in \mathfrak{D}(N)$. Letting $g = f - f'$ we obtain $f = f'$ as desired. This completes the proof of Theorem 2.

4. Abstract boundary conditions. For $u \in \mathfrak{D}(\bar{N}^*)$, $v \in \mathfrak{D}(N^*)$ define $\langle uv \rangle = (\bar{N}^*u, v) - (u, N^*v)$.

THEOREM 3. *If N_1 is a normal extension of the formally normal operator N such that $\mathfrak{D}(N_1) = \mathfrak{D}(N) + (I - W)\mathfrak{M}$, then $\mathfrak{D}(N_1)$ may be described as the set of all $u \in \mathfrak{D}(\bar{N}^*)$ satisfying $\langle u\alpha \rangle = 0$ for all $\alpha \in (I - W)\mathfrak{M}$.*[6]

REMARK. For differential operators the conditions $\langle u\alpha \rangle = 0$ become boundary conditions. They are self-adjoint ones, that is, $\langle \alpha\alpha' \rangle = 0$ for all $\alpha, \alpha' \in (I - W)\mathfrak{M}$. Indeed $\alpha, \alpha' \in \mathfrak{D}(N_1) = \mathfrak{D}(N_1^*)$ and for any $\alpha \in \mathfrak{D}(N_1)$, $\alpha' \in \mathfrak{D}(N_1^*)$ we have $(\bar{N}^*\alpha, \alpha') = (N_1\alpha, \alpha') = (\alpha, N_1^*\alpha') = (\alpha, N^*\alpha')$.

Proof of Theorem 3. If $u \in \mathfrak{D}(N_1)$, $\alpha \in (I - W)\mathfrak{M} \subset \mathfrak{D}(N_1^*)$, the above argument shows that $\langle u\alpha \rangle = 0$. Conversely suppose $u \in \mathfrak{D}(\bar{N}^*)$ and $\langle u\alpha \rangle = 0$ for all $\alpha \in (I - W)\mathfrak{M}$. Let $u = f + (I - W)\phi + (I + W)\phi$, where $f \in \mathfrak{D}(N)$, $\phi \in \mathfrak{M}$. We note that $\langle \ \ \rangle$ is linear in the first spot, and $f + (I - W)\phi \in \mathfrak{D}(N_1)$. Thus $\langle (I + W)\phi\, \alpha \rangle = 0$ for all $\alpha \in (I - W)\mathfrak{M}$. Let $\alpha = \bar{N}^*(I + W)\phi \in (I - W)\mathfrak{M}$, since $(I - W)\mathfrak{M} = \bar{N}^*(I + W)\mathfrak{M}$. Then

$$0 = \langle (I + W)\phi\, \bar{N}^*(I + W)\phi \rangle = (\bar{N}^*(I + W)\phi, \bar{N}^*(I + W)\phi) + ((I + W)\phi, (I + W)\phi) ,$$

which proves that $(I + W)\phi = 0$, and hence $u \in \mathfrak{D}(N_1)$ as desired.

UNIVERSITY OF CALIFORNIA,
LOS ANGELES

[6] A result similar to Theorem 3 appears in the report by Davis (4) for the case when $\dim (\mathfrak{D}(\bar{N}^*)/\mathfrak{D}(N)) < \infty$.

SOME CLASSES OF EQUIVALENT GAUSSIAN PROCESSES ON AN INTERVAL

JACOB FELDMAN

1. Introduction. Let T be an index set, R, S real-valued nonnegative definite functions of two variables in T, and m, n real-valued functions on T. Let Ω be the set of all real-valued functions on T, and $\mathscr{S}$ the Borel field of cylinder sets. There are then unique measures μ, ν on $\mathscr{S}$ such that the functions x_t on Ω defined by $x_t(\omega) = \omega(t)$ form Gaussian stochastic processes, with means respectively m and n, and covariances respectively R and S. It is shown in [2] that μ and ν are either mutually absolutely continuous or totally singular, and a necessary and sufficient condition for equivalence is given.

Suppose now that T is a subset of the real line, and $R(s, t) = r(s - t)$, $S(s, t) = s(s - t)$, where r and s are continuous nonnegative-definite functions, and hence can be written as inverse Fourier transforms of finite measures $d\rho, d\sigma$. Thus, using respectively the measures μ and ν on $\Omega, x_t - m(t)$ and $x_t - n(t)$ are the restrictions to T of stationary Gaussian processes on the real line. For simplicity, only the case $m = n = 0$ will be considered.

When T is the entire real line, then it is easy to see, by looking at $d\rho$ and $d\sigma$, exactly when $\mu \sim \nu$, as is essentially known (see [3]). The precise conditions are:

a. ρ and σ must have *identical* non-atomic parts.

b. Their points of positive mass be the same, and if the masses are a_i and b_i at x_i, then $\sum\{(a_i/b_i) - 1\}^2$ must be finite.

Now suppose T is a finite interval. The problem of determining from knowledge of ρ and σ whether μ and ν are equivalent becomes much more difficult. We here discuss only a certain class of cases. Because of stationarity, one need only consider an interval symmetric about 0. Continuity of r and s implies that the Gaussian process is continuous with probability one at any given point, so that it makes no difference whether the interval is open or closed. There is no essential loss of generality, then, in considering only the closed interval $[-\pi, \pi]$. The following facts will then be proven:

THEOREM. *Let $d\rho(x) = \{dx/(1 + x^2)^u\}$, where u is an integer $\geqq 1$, and let σ be some other finite nonnegative measure on the real line. Write $\tau = \sigma - \rho$. The following conditions are necessary and sufficient that the Gaussian processes induced on $[-\pi, \pi]$ by the Fourier trans-*

Received December 21, 1959, Research partly supported by Contract NONR-222 (60).

forms of ρ *and* σ *have equivalent measures on path space:*

(a) *if* k_n *is a sequence of* C_∞ *functions with support in* $]-\pi, \pi[$ *and* K_n *is the Fourier transform of* k_n, *then* $\int |K_n|^2 d\sigma \to 0$ *implies* $\int |K_n|^2 d\rho \to 0$.

(b) *The Fourier transform (in the sense of Schwartz distributions) of* $(1 + x^2)^u d\tau(x)$ *agrees on* $]-2\pi, 2\pi[$ *with a function* ψ *such that*

$$\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} |\psi(s-t)|^2 ds\,dt < \infty .$$

REMARK 1. It will be seen that sufficiency still holds if (a) is weakened to:

(a′) $\int |K_n|^2 d\sigma \to 0$ *and* $K_n \to K$ *in* $\mathscr{L}_2(\rho)$ *implies that* $K = 0$ *on some set of positive* ρ*-measure.*

REMARK 2. As a consequence of Remark 1, it is clear that if σ has a component which is absolutely continuous with respect to ρ, then Condition (a) automatically satisfied.

Retaining the notation of the theorem:

COROLLARY 1. *If* $d\sigma = \Phi d\rho$, *where* Φ *is a function such that* $\Phi - 1$ *is a finite linear combination of functions in various* $L_a(-\infty, \infty)$ *classes,* $1 \leqq a \leqq 2$, *then the Gaussian processes induced by* ρ *and* σ *have equivalent measures on path space.*

One direction of the following corollary was proven by *D.* Slepian in [5], using techniques of G. Baxter in [1]:

COROLLARY 2. *If* A_j *and* B_j *are polynomials, with degrees respectively* a_j *and* b_j, $j = 1, 2$, *and* $b_j > a_j$, *then the Gaussian processes whose spectral measures are* $|A_j(x)/B_j(x)|^2\, dx$ *have equivalent measures on path space if and only if*

(a) $b_1 - a_1 = b_2 - a_2$

(b) *the ratio of the leading coefficients of* A_1 *and* B_1 *has the same absolute value as the ratio of the leading coefficients of* A_2 *and* B_2.

The author wishes to thank J. F. Trèves for several useful discussions about distributions.

2. Some preliminaries on functions of exponential type. First, some notation. Functions will be complex-valued functions of a real variable, unless otherwise stated. $\hat{F}$ will mean the Fourier transform of F (in various degrees of generalization, depending on context), and $\check{F}$ the conjugate Fourier transform. sup (f) will mean the points where $f \neq 0$. $\mathscr{E}_a = \{F \mid F$ extends to an entire function of exponential type $\leqq a\pi\}$. $\mathscr{H}_a = \mathscr{E}_a \cap \mathscr{L}_2(-\infty, \infty)$, or, by the Payley-Wiener theorem,

$$= \{\hat{f} \mid f \in \mathscr{L}_2(-\infty, \infty), \sup(f) \subset [-a\pi, a\pi]\} .$$
$$\hat{\mathscr{D}}_a = \{f \mid f \in \mathscr{C}_\infty, \overline{\sup(f)} \subset]-a\pi, a\pi[\ \}, \quad \mathscr{D}_a = \{\check{f} \mid f \in \hat{\mathscr{D}}_a\} .$$

u will be a fixed integer $\geqq 1$, and $p(x) = (i + x)^u$. ρ is the measure $d\rho(x) = \{1/|p(x)|^2\}dx$. $\mathscr{K}$ will denote the completion of $\mathscr{D}_1$ in the inner product $\langle F, G\rangle = \int F\bar{G}d\rho$.

Naturally, $\mathscr{K}$ really consists of equivalence classes of functions; but it will turn out that there is a continuous, in fact entire, member in each class. H_1 will denote a fixed function of $\mathscr{D}_1$ such that $h_1 = \hat{H}_1$ is nonnegative and has integral 1. For $a > 0$, $h_a(s)$ will be $(1/a)h_1(s/a)$, $H_a(x) = H_1(ax)$, so that $h_a = \hat{H}_a$, and $H_a \in \mathscr{D}_a$. Then H_a vanishes faster than any polynomial, $|H_a(x)| \leqq 1$ for all x, and $\lim_{a\to 0} H_a(x) = 1$ uniformly on any finite interval.

LEMMA 1. *If* $F \in \mathscr{E}_1$ *and* $\int |F|^2 d\rho < \infty$, *then* $F \in \mathscr{K}$.

Proof. If $(1/2) < c < 1$, then

$$\left(\int |F(cx) - F(x)|^2 d\rho(x)\right)^{1/2} \leqq \left(\int_{-b}^{b} |F(cx) - F(x)|^2 d\rho(x)\right)^{1/2}$$
$$+ \left(\int_{|x|>b} |F(cx)|^2 d\rho(x)\right)^{1/2} + \left(\int_{|x|>b} |F(x)|^2 d\rho(x)\right)^{1/2} .$$

Now,

$$\int_{|x|>b} |F(cx)|^2 d\rho(x) = \frac{1}{c}\int_{|x|>bc} |F(x)|^2 \frac{1}{\left|p\left(\frac{x}{c}\right)^2\right|} dx$$
$$\leqq 2\int_{|x|<(b/2)} |F(x)|^2 \frac{1}{|p(x)|^2} dx .$$

Choosing b large, and then choosing c close enough to 1 to make $|F(cx)-F(x)|$ small on $[-b, b]$, we see that it suffices to show that the function $G\colon x \to F(cx)$ is in $\mathscr{K}$. Notice that $G \in \mathscr{E}_1$, as $c < 1$.

H_aG is square-integrable, since H_a vanishes faster than $(1/|p|^2)$. So H_aG is in $\mathscr{H}_{a+c}$, its Fourier transform being some g' in $\mathscr{L}_2(-\infty, \infty)$ with support in $[-(a+c)\pi, (a+c)\pi]$. Thus $h_a * g' \in \mathscr{D}_{2a+c}$, and $H_a^2 G \in \mathscr{D}_{2a+c}$. Choosing a small causes $H_a^2 G$ to be in $\mathscr{D}_1$, and simultaneously causes $\int |H_a^2 G - G|^2 d\rho$ to get small. This proves the lemma.

Let $\mathscr{H} = \{pF \mid F \in \mathscr{H}_1\}$, and $\mathscr{D} = \{pF \mid F \in \mathscr{D}_1\}$. Lemma 1 tells us $\mathscr{H} \subset \mathscr{K}$.

LEMMA 2. $\mathscr{H}$ *is precisely the closure of* $\mathscr{D}$ *in* $\mathscr{K}$.

Proof. First, we see that $\mathscr{H}$ is closed. If $F_n \in \mathscr{H}_1$ and

$$\int | pF_n - G |^2 d\rho \to 0, \text{ then } \int | F_n(x) - F_m(x) |^2 dx \to 0 .$$

Since $\mathscr{H}_1$ is complete, there is some $F \in \mathscr{H}_1$ with $\int | F_n(x) - F(x) |^2 dx \to 0$. So some subsequence of the pF_n converges almost everywhere to pF. Thus $pF = G$ almost everywhere.

To approximate elements pF in $\mathscr{H}$ by elements in $\mathscr{D}$, just approximate F in $\mathscr{L}_2(-\infty, \infty)$ by elements in $\mathscr{D}_1$, using the technique of Lemma 1.

LEMMA 3. *$\mathscr{K} \ominus \mathscr{H}$ is precisely the finite-dimensional space $\mathscr{L}$ of functions of the form $x \to e^{ix\pi} q(i - x)$, where q is a polynomial of degree $\leqq u - 1$.*

Proof. Suppose $F \in \mathscr{K} \ominus \mathscr{H}$. Then $\int F\overline{pG} d\rho = 0$ for all G in $\mathscr{D}_1$, i.e. $\int \{F(x)/p(x)\} \overline{G(x)}\, dx = 0$ for all G in $\mathscr{D}_1$. Now, (F/p) is in $\mathscr{L}_2(-\infty, \infty)$, so it has a Fourier transform k which is likewise square-integrable, and, by Plancherel's theorem, $\int k(s)\overline{g(s)} ds = 0$ for all g in $\hat{\mathscr{D}}_1$. So k vanishes in $]-\pi, \pi[$.

Since $F \in \mathscr{K}$, F can be approximated in $\mathscr{K}$ by functions F_n in $\mathscr{D}_1$. Each F_n is in $\mathscr{D}_{a_n}$ for some $a_n < 1$, since $\overline{\sup(F_n)} \subset]-\pi, \pi[$, and hence $\subset]-a_n\pi, a_n\pi[$ for some $a_n < 1$. Let k_n be the Fourier transform of F_n/p. Then $k_n \to k$ in $\mathscr{L}_2(-\infty, \infty)$, and k_n is in the domain of the $\mathscr{L}_2$-differential operator $p(-iD) = i^u(I - D)^u$. So $p(-iD)k_n = f_n$, where f_n is the Fourier transform of F_n. Since f_n vanishes outside some $[-a_n\pi, a_n\pi]$, $a_n < 1$, k_n must be of the form $\sum_j a_i^{(n)} s^j e^s$ in $]-\infty, -\pi[$ and $\sum_j b_j^{(n)} s^j e^s$ in $]\pi, \infty[$, where j ranges between 0 and $u - 1$. Since k_n is in $\mathscr{L}_2(-\infty, \infty)$, the $b_j^{(n)}$ are zero, and, letting φ be the indicator of $]-\infty, -\pi[$, we get $\varphi k_n = \varphi \sum_j a_j^{(n)} s^j e^s$. This converges in $\mathscr{L}_2(-\infty, \infty)$, so the limit is of the form $\varphi \sum_j a_j s^j e^s$. Then $k_n \to 0$ in $]\pi, \infty[$, 0 in $[-\pi, \pi]$, and $\sum_j a_j s^j e^s$ in $]-\infty, -\pi[$, so $k = \varphi \sum_j a_j s^j e^s$. F/p is then a linear combination of terms like $\int_{-\infty}^{-\pi} e^{-ixs} s^j e^s ds$, $0 \leqq j \leqq u - 1$, which is a linear combination of terms like $e^{ix\pi}(i + x)^{-j}$, $1 \leqq j \leqq u$. Multiplying by p gives the result.

Combining information from lemmas 1, 2, 3 we get a description of $\mathscr{K}$:

PROPOSITION. $\mathscr{K}$ is the orthogonal direct sum of $\mathscr{H}$ and $\mathscr{L}$.

LEMMA 4. $\mathscr{D} = \mathscr{H} \cap \mathscr{D}_1$.

Proof. $\mathscr{D} \subset \mathscr{H}$, by definition, since $\mathscr{D}_1 \subset \mathscr{H}_1$. Also $\mathscr{D} \subset \mathscr{D}_1$, since $\mathscr{D}_1$ is closed under multiplication by polynomials (because $\hat{\mathscr{D}}_1$ is closed under differentiation). So $\mathscr{D} \subset \mathscr{H} \cap \mathscr{D}_1$, and it remains to show $\mathscr{D} \supset \mathscr{H} \cap \mathscr{D}_1$.

Suppose $G \in \mathscr{H}$. Then G is a $\langle\, , \rangle$ limit of elements G_n in $\mathscr{D}$, by Lemma 2. G_n then has the form pF_n, F_n in $\mathscr{D}_1$. Thus F_n is an $\mathscr{L}_2(-\infty, \infty)$ Cauchy sequence, hence has a limit F. Then $pF = G$.

Suppose G is also in $\hat{\mathscr{D}}_1$. Then G is infinitely differentiable. Since $\hat{G} = p\hat{F} = p(-iD)\hat{F}$, we conclude that $\hat{F}$ is infinitely differentiable. Now it must be shown that $\hat{F}$ vanishes outsides some interval $[-a\pi, a\pi]$, $0 < a < 1$. But $\hat{G} = p(-iD)\hat{F}$ vanishes outside such on interval, so $\hat{F}$ is analytic outside $[-a\pi, a\pi]$. Also, $\hat{F}$ vanishes outside $[-\pi, \pi]$, since each $\hat{F}_n$ has support in $]-\pi, \pi[$. Therefore, $\hat{F}$ vanishes outside $[-a\pi, a\pi]$. So $\hat{F}$ is in $\mathscr{D}_1$, and F is in $\mathscr{D}_1$.

LEMMA 5. *$\mathscr{D}_1/\mathscr{D}$ is finite dimensional.*

Proof. $\mathscr{D}_1/\mathscr{D} = \mathscr{D}_1/\mathscr{D}_1 \cap \mathscr{H} \approx (\mathscr{D}_1 + \mathscr{H})/\mathscr{H} \subset \mathscr{K}/\mathscr{H} \approx \mathscr{L}$.

3. **Proof of theorem.** In [2] it is shown that a necessary and sufficient condition for equivalence of μ and ν is that there be an *equivalence operator* from the closed linear span of $\{x_t \mid t \in T\}$ in $\mathscr{L}_2(\mu)$ to their closed linear span in $\mathscr{L}_2(\nu)$, sending the μ-equivalence class of x_t to the ν-equivalence class of x_t for each $t \in T$. (An equivalence operator, as defined in [2], is a linear homeomorphism H between two Hilbert spaces such that $I - H^*H$ is Hilbert Schmidt). Actually, we shall want the condition in *complex* $\mathscr{L}_2$, while the proof in [2] is for real $\mathscr{L}_2$; however, the transition from the one to the other is immediate.

Under this condition, H would map $\int_{-\pi}^{\pi} f(x_t)dt$ as an $\mathscr{L}_2(\mu)$— valued integral to $\int_{-\pi}^{\pi} f(t)x_t dt$ as an $\mathscr{L}_2(\nu)$-valued integral, for each $f \in \hat{\mathscr{D}}_1$; and conversely, if H had this effect on all such $\int_{-\pi}^{\pi} f(t)x_t dt$, then by choosing a sequence of f approximating a delta function, one could verify that H sent the equivalence class of x_t in $\mathscr{L}_2(\mu)$ to the equivalence class of x_t in $\mathscr{L}_2(\nu)$. Therefore, putting inner products $(\, ,)$ and $(\, ,)^\cdot$ on $\hat{\mathscr{D}}_1$ by the rules

$$(f, g) = \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \gamma(s-t)f(s)\overline{g(t)}dsdt\ ,$$

$$(f, g)^\cdot = \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \rho(s-t)f(s)\overline{g(t)}dsdt\ ,$$

and noting that $(f, g) = \int\left(\int_{-\pi}^{\pi} f(s)x_s ds\right)\left(\int_{-\pi}^{\pi} g(t)x_t dt\right)d\mu$ and

$$(f, g)^\cdot = \int\left(\int_{-\pi}^{\pi} f(s)x_s ds\right)\left(\int_{-\pi}^{\pi} g(t)x_t dt\right)d\nu,$$

it follows that a necessary and sufficient condition for the equivalence of μ and ν is the existence of an equivalence operator from the $(\, ,)$ completion of $\hat{\mathscr{D}}_1$ to its $(\ ,\ )^{\cdot}$ completion, and sending the $(\ ,\ )$-equivalence class of f to its $(\ ,\ )^{\cdot}$-equivalence class.

Now let $\langle F, G\rangle^{\cdot} = \int F\bar{G}d\sigma$, where F and G are in $\mathscr{D}_1$ (and hence continuous and bounded, so that the integral exists). Let $\dot{\mathscr{K}}$ be the closure of $\mathscr{D}_1$ in $\mathscr{L}_2(\sigma)$. Let J be the map assigning to F in $\mathscr{D}_1$ its equivalence class in $\dot{\mathscr{K}}$. Since $\langle F, G\rangle = (\hat{F}, \hat{G})$, and $\langle F, G\rangle^{\cdot} = (\hat{F}, \hat{G})^{\cdot}$, the necessary and sufficient condition for the equivalence of μ and ν in the theorem is that J be the restriction to $\mathscr{D}_1$ of an equivalence map from $\mathscr{K}$ to $\dot{\mathscr{K}}$.

To prove sufficiency of the conditions in the theorem, suppose first that $|p(x)|^2\ d\tau(x)$ has a generalized Fourier transform (see [4]) which agrees on $]-2\pi, 2\pi[$ with a function ψ such that $\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} |\psi(s-t)|^2\, dsdt = a^2 < \infty$. We extend ψ by making it 0 outside $]-2\pi, 2\pi[$.

Lemma 6. *If* $F \in \mathscr{D}$, *then* $\langle F, F\rangle^{\cdot} \leqq (1 + a)\langle F, F\rangle$.

Proof. Write $F=pG$, $G \in \mathscr{D}_1$. Then $\int |F|^2 d\sigma = \int |F|^2 d\rho + \int |G|^2 |p|^2 d\tau$. Now, $\hat{G}$ is in $\hat{\mathscr{D}}_1$, so $\hat{G} * \hat{G}$ is infinitely differentiable with support in $]-2\pi, 2\pi[$. Then, by Schwartz's definition of generalized Fourier transform, we get $\int |G|^2 |p|^2 d\tau = \int_{-2\pi}^{2\pi} \hat{G}*\hat{\bar{G}}(s)\psi(s)\, ds = \int_{-2\pi}^{2\pi}\int_{a(s)}^{b(s)} \hat{G}(s-t)\hat{\bar{G}}(1-t)$ $\psi(s)dtds$, where $a(s) = \max(-\pi, s-\pi)$ and $b(s) = \min(\pi, s+\pi)$. Letting $s - t = s'$, and $t = -t'$ gives $\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \hat{G}(s')\hat{G}(t')\psi(s'-t')ds'dt'$, whose absolute value, by the Schwartz inequality, is

$$\leqq \left[\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} |\hat{G}(s)\overline{\hat{G}(t)}|^2\, dsdt \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} |\psi(s-t)|^2 dsdt\right]^{1/2}$$
$$= \left(\int_{-\pi}^{\pi} |\hat{G}(s)|^2\, ds\right)a = \left(\int |F|^2 d\rho\right)a\,.$$

Pick a complete orthonormal set (c.o.n.s.) $f_1, f_2, \cdots$ for $\mathscr{L}_2(-\pi, \pi)$ out of the dense subset $\hat{\mathscr{D}}_1$. Let $F_n = \check{f}_n$, and $G_n = pF_n$. Then the G_n form a c.o.n.s. for $\mathscr{H}$ (in $\langle\ ,\ \rangle$) consisting of elements of $\mathscr{D}$, because the F_n are a c.o.n.s. for $\mathscr{H}_1$ consisting of elements of $\mathscr{D}_1$.

Lemma 7. $\sum_{n,m=1}^{\infty} |\langle G_n, G_m\rangle - \langle G_n, G_m\rangle^{\cdot}|^2 = a^2$.

Proof. $\int G_n(x)\overline{G_m(x)}d\tau(x) = \int_{-2\pi}^{2\pi} \hat{F}_n * \hat{F}_m(s)\psi(s)ds$. By using a change of variable as in the previous lemma, this equals $\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} f_n(s)\overline{f_m(t)}\,\psi(s-t)dsdt$. But the functions $(s, t) \rightarrow \overline{(f_n s)} f_m(t)$ form a c.o.n.s. in $\mathscr{L}_2([-\pi, \pi]\times[-\pi, \pi])$, so that $\sum_{n,m=1}^{\infty}\left|\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} f_n(s)\overline{f_m(t)}\psi(s-t)dsdt\right|^2$ is exactly $\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\left|\psi(s-t)\right|^2 dsdt$.

Now consider the map J from $\mathscr{D}_1$ to $\dot{\mathscr{K}}$. *Lemma* 6 implies that its restriction to $\mathscr{D}$ is bounded, and, since $\mathscr{D}_1/\mathscr{D}$ is finite-dimensional (Lemma 5), J is bounded as an operator from $\mathscr{D}_1$ to $\dot{\mathscr{K}}$ (a finite-dimensional extension of a bounded operator is bounded, as is readily seen). So J extends uniquely to a bounded operator A from $\mathscr{K}$ to $\dot{\mathscr{K}}$.

LEMMA 8. *$I—A^*A$ is a Hilbert-Schmidt operator.*

Proof. Complete the o.n.s. $G_1, G_2, \cdots$ by adding to it a c.o.n.s. $G_0, G_{-1}, \cdots, G_{1-u}$ in $\mathscr{L}$. Then, letting $k = u - 1$,

$$\begin{aligned}
&\textstyle\sum_{n,m=-k}^{\infty} |\langle(I - A^*A)G_n, G_m\rangle|^2 \\
&= \textstyle\sum_{n,m=1}^{\infty} |\langle(I - A^*A)G_n, G_m\rangle|^2 \\
&\quad + \textstyle\sum_{n=-k}^{0}\sum_{m=-k}^{\infty} |\langle(I - A^*A)G_n, G_m\rangle|^2 \\
&\quad + \textstyle\sum_{n=-k}^{\infty}\sum_{m=-k}^{0} |\langle G_n, (I - A^*A)G_m\rangle|^2 \\
&= \textstyle\sum_{n,m=1}^{\infty} |\langle G_n, G_m\rangle - \langle AG_n, AG_m\rangle^{\cdot}|^2 \\
&\quad + 2\textstyle\sum_{n=-k}^{0} |\langle I - A^*A)G_n, (I - A^*A)G_n\rangle|^2 ,
\end{aligned}$$

using Parseval's equality. But $\langle AG_n, AG_m\rangle^{\cdot} = \langle G_n, G_m\rangle^{\cdot}$ for $n, m > 0$, since such G_n are in $\mathscr{D}_1$, so that the sum is exactly

$$a^2 + 2\textstyle\sum_{n=-k}^{0}\langle(I - A^*A)G_n, (I - A^*A)G_n\rangle .$$

In order to complete the proof, it must be shown that A is a homeomorphism from $\mathscr{K}$ onto $\dot{\mathscr{K}}$. Since $I - A^*A$ is completely continuous, it will suffice to show

(1) that the range of A is dense in $\dot{\mathscr{K}}$.

(2) that A sends no nonzero element to zero.

(1) is clear, since the range of A contains the range of J, which is dense by the very definition of $\dot{\mathscr{K}}$.

We now make use of (a), or rather of the weaker (a′), to prove (2). Suppose, in fact, that $A(K)$ is zero in $\dot{\mathscr{K}}$ for some K in $\mathscr{K}$. Let K_n be a sequence of members of $\mathscr{D}_1$ converging to K in $\mathscr{K}$. Then K_n converges to zero in $\dot{K}$, since $A(K_n) = J(K_n)$. Then, by (a′), $K = 0$ on a set of positive ρ measure. But the Proposition of the previous section tells us that K is analytic. Thus $K = 0$.

To show the necessity of condition (a), suppose J has an extension to an equivalence operator from $\mathscr{K}$ to $\dot{\mathscr{K}}$, which we call A. Then (a) is immediate from the fact that A is continuously invertible.

Since $I - A^*A$ is an equivalence operator, $\sum_{n,m=1}^{\infty} |\langle G_n, G_m\rangle - \langle G_n G_m\rangle^{\cdot}|^2 < \infty$, where $G_1, G_2, \cdots$ is the c.o.n.s. in $\mathscr{D}$ for $\mathscr{H}$ previously constructed. Define an operator Z on $\mathscr{L}_2([-\pi, \pi]\times[-\pi, \pi])$ as follows: let $f_{n,m}(s, t) = f_n(s)f_m(t)$, where $G_n = p\check{f}_n$. For $Q = \sum_{n,m} a_{n,m} f_{n,m}$, Let $Z(Q) = \sum_{n,m} a_{n,m}(\langle G_n, G_m\rangle - \langle G_n, G_m\rangle^{\cdot})$. Then

$$|Z(Q)|^2 \leqq \textstyle\sum_{n,m} |a_{n,m}|^2 \sum_{n,m} |\langle G_n, G_m\rangle - \langle G_n, G_m\rangle^{\cdot}|^2 .$$

So $Z(Q)$ has the form $\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} Q(s,t)\Psi(s,t)\,ds\,dt$ for some Ψ such that $\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} |\Psi(s,t)|^2 ds\,dt < \infty$. In particular, consider $f, g \in \mathscr{D}_1$, and let $f = \sum_n a_n f_n$, $g = \sum b_m f_m$. Let $Q(s,t) = f(s)\overline{g(t)}$. Then $Z(Q) = \sum_{n,m} a_n b_m (\langle G_n, G_m\rangle - \langle G_n, G_m\rangle^{\cdot}) = \sum_{n,m} a_n \bar{b}_m \int (pF_n)(\overline{pF_m})\,dt = \int \check{f}\bar{\check{g}}\,|p|^2\,dt$.

Let $0 < r < 2\pi$, and let f, g have the closure of their supports in in $]-\pi+r, \pi[$. Let $f'(s) = f(s+r)$, $g'(s) = g(s+r)$. Then f', g' are in $\mathscr{D}_1$, and their inverse Fourier transforms satisfy $\check{f}'(x) = e^{irx}\check{f}(x) = e^{irx}\check{g}(x)$. Then

$$\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} f'(s)\overline{g'(t)}\Psi(s,t)\,ds\,dt = \int \check{f}'\bar{\check{g}}'\,|p|^2 dt$$

$$= \int \check{f}\bar{\check{g}}\,|p|^2 dt = \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} f(s)\overline{g(t)}\Psi(s,t)\,ds\,dt\,.$$

But

$$\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} f(s+r)\overline{g(t+r)}\Psi(s,t)\,ds\,dt = \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} f(s)\overline{g(t)}\,\psi(s-r, t-r)\,ds\,dt\,.$$

in view of the restrictions on the support of f and g. Since this holds for all such f, g, the equality $\Psi(s-r, t-r) = \Psi(s,t)$ holds for almost all (s,t) for which $s, t, s-r, t-r$ are in $]-\pi, \pi[$ (r being fixed). Thus, $\{(r,s,t) \mid s, t, s-r,$ are in $]-\pi, \pi[$ and $\Psi(s-r, t-r) \neq \Psi(s,t)\}$ has measure zero.

Applying Fubini's theorem, we get: for almost all pairs s, t in $]-\pi, \pi[$ the set $\{r \mid s-r, t-r$ lie in $]-\pi, \pi[$ and $\Psi(s-r, t-r) \neq \Psi(s,r)\}$ has measure 0. Denote by Δ the exceptional set of pairs (s,t).

Now let Γ_s be the line of slope 1 which passes through $(s, -s)$, where $-\pi < s < \pi$. Let Γ be the set of s for which $\Gamma_s \cap \Delta$ is *not* a set of measure 0. Then Γ has measure 0, again by Fubini's theorem, and by rotation-invariance of Lebesgue measure. If s is in $]-\pi, \pi[$ but not in Γ, then almost all points on that portion of L_s which lies in $]-\pi, \pi[\times]-\pi, \pi[$ assign to Ψ a common value; thus, if the function Ψ' is defined on $]-\pi, \pi[$ by $\Psi'(s,t) = \int_{a(s,t)}^{b(s,t)} \Psi(s-r, t-r)\,dr$, where $a(s,t) = \max(s-\pi, t-\pi)$ and $b(s,t) = \min(s+\pi, t+\pi)$, then, for (s,t) on Γ_r, $\Psi'(s,t)$ has this common value. Thus, for almost all r, $\Psi'(s,t) = \Psi(s,t)$ for almost all (in linear measure) points (s,t) with $-\pi < s, t < \pi$ and s, t on Γ_r. Then $\Psi'(s,t)$ is equal almost everywhere to $\Psi(s,t)$. Now set $\psi(r) = \Psi(-r/2, r/2)$, $-2\pi < r < 2\pi$.

Then

$$\Psi'(s,t) = \Psi'(s-(s+t)/2, t-(s+t)/2)$$

$$= \Psi'(-(t-s)/2, (t-s)/2) = \psi(t-s),$$

for s, t in $]-\pi, \pi[$. This completes the proof.

Corollary 1 is just a consequence of the fact (proven in [4]) that if

Φ is as in the statement, then $(\bar{\Phi}-1)dx$ has a generalized Fourier transform which is a function φ square-summable in any finite interval, so that

$$\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\left|\varphi(s-t)\right|^2 ds\,dt \leqq \left|\int_{-\pi}^{\pi}\int_{-2\pi}^{2\pi}\right|\varphi(r)\Big|^2 dr \leqq 2\pi\int_{-2\pi}^{2\pi}\left|\varphi(r)\right|^2 dr\,.$$

To prove corollary 2: let c_j be the absolute value of the ratio of the leading terms of A_j and B_j, and let $u_j = b_j - a_j = \deg(B_j) - \deg(A_j)$. It is clear in general that equivalence of the Gaussian processes induced by given covariances is unaffected if both covariances are multiplied by the same constant. Thus, we find that the process whose spectral measure is

$$\left|\frac{A_j(x)}{B_j(x)}\right|^2 dx$$

has measure on path space equivalent to that whose spectral measure is

$$\frac{c_j}{(1+x^2)^{u_j}}dx,$$

because the quotient of

$$\left|\frac{A_j(x)}{B_j(x)}\right|^2 \quad \text{by} \quad \frac{c_j}{(1+x^2)^{u_i}}$$

is of the form: 1 plus a function in $\mathscr{L}_2(-\infty,\infty)$. So the problem is reduced to whether or not the processes with spectral measures

$$\frac{1}{(1+x^2)^{u_1}}\,dx \text{ and } \frac{c_2c_1^{-1}}{(1+x^2)^{u_2}}\,dx$$

are equivalent. The criterion is that

$$\left(1-\frac{c_2c_1^{-1}}{(1+x^2)^{u_2-u_1}}\right)dx$$

have a generalized Fourier transform which agrees with a function on $]-2\pi, 2\pi[$ having certain properties. But this generalized Fourier transform is explicitly calculated (see [4]), and is of the required form when and only when $c_2 = c_1$ and $u_2 = u_1$.

BIBLIOGRAPHY

1. G. Baxter, *A strong limit theorem for Gaussian processes*, Proc. Amer. Math. Soc., (1956).
2. J. Feldman, *Equivalence and perpendicularity of Gaussian Processes*, Pacific J. Math., **8**, No. 4, (1958).

3. U. Grenander, *Stochastic processes and statistical inference,* Arkiv furti Matemak, Band 1, nr. 17.

4. L. Schwartz, *Théorie des distributions*: Hermann and C^{ie}, Paris, 1951.

5. D. Slepian, *Some comments on the detections of Gaussian signals in Gaussian noise*: I. R. E. Transactions on Information Theory, Vol. I–4, June, 1958.

UNIVERSITY OF CALIFORNIA, BERKELEY

WEAK AND STRONG CONVERGENCE FOR MARKOV PROCESSES

S. R. FOGUEL

1. Introduction. Let (Ω, Σ, P) be a probability space and $x_t(\omega)$ a Markov process defined on it. For every Borel set on the real line $P_t(\omega, A)$ is the conditional probability that $x_t \in A$ given x_0. The purpose of this paper is to study the limiting behavior, of the family of functions, $p_t(\omega, A)$, for $t \to \infty$ and A fixed.

In § 3 we investigate conditions for the weak convergence, in the sense of $L_2(\Omega, \Sigma, P)$, of $p_t(\omega, A)$. The classical result on Markov processes, as presented in [2] p. 353, is generalized to functions $x_t(\omega)$ with nondiscrete ranges. Under the additional assumption of existence of finite stationary measures.

It should be noted that

$$p_{ij}^{(n)} = \frac{(p_n(\omega, \{j\}), \chi_{x_0} = i)}{P(x_0 = i)}$$

where the parenthesis stand for scalar product and $\chi_{x_0} = i$ is the characteristic function of the set $x_0(\omega) = i$. Thus weak convergence of $p_n(\omega, \{j\})$ implies ordinary convergence of $p_{ij}^{(n)}$.

In § 4 the strong convergence in $L_2(\Omega, \Sigma, P)$ is studied. Our results are similiar to Theorem 11 of [4] though the exact relation between the two theories is not clear to us.

The paper deals with real processes and L_2 is the real Hilbert space.

Throughout the paper a weak form of the definition of Markov processes is used. We do not assume any of the regularity properties which are usually imposed.

2. Notation and general background. Let $x_t(\omega)$ be a set of measurable functions, defined on Ω, where t runs over $[0, \infty)$ or the positive integers. This set of functions, will be called a Markov process if whenever $t_1 < t_2 < t_3$ then *conditional probability that* $x_{t_3} \in A$ *given* x_{t_1} *and* x_{t_2}, *is equal to the conditional probability that* $x_{t_3} \in A$ *given* x_{t_2}.

In order to simplify this condition let us observe the following:

If Σ_1 *is a sub* σ *algebra of* Σ *and* $f \in L_2(\Omega, \Sigma, P)$ *then the conditional expectation of* f *with respect to* Σ_1 *is equal* a.e. *to* $E_1 f$ *where* E_1 *is the self adjoint projection on the subspace of* L_2 *generated by characteristic functions of sets in* Σ_1.

With the Markov process, $x_t(\omega)$, associate a collection of subspaces,

Received November, 23, 1959, and in revised from March 2, 1960.

B_t of $L_2(\Omega, \Sigma, P)$, where B_t is the closed subspace spanned by characteristic functions of sets of the form $x_t^{-1}(A)$, A a Borel set on the line. Let E_t be the self adjoint projection on B_t.

THEOREM 2.1. *If the set of functions $x_t(\omega)$ is a Markov process, then*

$$E_{t_1}E_{t_2}E_{t_3} = E_{t_1}E_{t_3} \quad \textit{for} \quad t_1 < t_2 < t_3 . \tag{2.1}$$

Proof. Let $t_1 < t_2 < t_3$. If $g \in B_{t_3}$ then $g - E_{t_2}g$ is orthogonal to B_{t_1}. Thus

$$E_{t_1}(E_{t_3} - E_{t_2}E_{t_3}) = 0 .$$

DEFINITION. *A Collection of spaces $B_t \subset L_2(\Omega)$, is a Markov class if equation* 2.1 *holds.*

From the above definition follows:

THEOREM 2.2. *Let B_t be a Markov class. If $f \in B_{t_1} \cap B_{t_2}$ and $t_1 < t < t_2$ then $f \in B_t$.*

Proof. If $f = E_{t_1}f = E_{t_2}f$ then

$$\begin{aligned} \| E_t f \|^2 &= (E_t f, f) = (E_t E_{t_2} f, E_{t_1} f) = (E_{t_1} E_t E_{t_2} f, f) \\ &= \| (E_{t_1} E_{t_2} f, f) = \| f \|^2 . \end{aligned}$$

Thus $f = E_t f \in B_t$.

DEFINITION. *A Markov process is called stationary if*

$$P(x_{t_1+\alpha} \in A_1 \cap x_{t_2+\alpha} \in A_2) = P(x_{t_1} \in A_1 \cap x_{t_2} \in A_2) . \tag{2.2}$$

In particular for a stationary Markov process

$$P(x_t \in A) = P(x_0 \in A) . \tag{2.3}$$

Let T_t be the transformation from B_0 to B_t defined for characteristic functions in B_0 by

$$T_t \chi_{x_0 \in A} = \chi_{x_t \in A} . \tag{2.4}$$

LEMMA 2.4. *Let $x_t(\omega)$ be a stationary Markov process. The transformation T_t can be extended in a unique way to all of B_0 such that*

(a) $$\| T_t x \| = \| x \| \quad \textit{if} \quad x \in B_0$$

(b) $$T_t B_0 = B_t$$

(c) $$(T_{t_1+\alpha} x, T_{t_2+\alpha} y) = (T_{t_1} x, T_{t_2} y)$$

for every $x \in B_0$, $y \in B_0$ *and* $\alpha > 0$.

Proof. In order to consider T_t as a transformation in B_0 we have to show that:

If A_1 and A_2 are two Borel sets and $\chi_{x_0^{-1}(A_1)}$, $\chi_{x_0^{-1}(A_2)}$ differ by a set of measure zero, then

$$\chi_{x_t^{-1}}(\omega)_{(A_1)} = \chi_{x_t^{-1}(A_2)}(\omega) \quad \text{a.e.}$$

Now by assumption

$$\| \chi_{x_0^{-1}(A_1)} \| = \| \chi_{x_0^{-1}(A_2)} \| = \| \chi_{x_0^{-1}(A_1 \cap A_2)} \| .$$

But by 2.3

$$\| \chi_{x_t^{-1}(A_1)} \| = \| \chi_{x_t^{-1}(A_2)} \| = \| \chi x_{x_t^{-1}(A_1 \cap A_2)} \|$$

which means

$$\chi_{x_t^{-1}(A_1)} = \chi_{x_t^{-1}(A_2)} \quad \text{a.e.}$$

Let us extend T_t to linear combinations of characteristic functions by additivity. If conditions a and c are satisfied for this dense set, we will be able to extend T_t by continuity to all of B_0 and T_t will satisfy a, b and c. It is enough to show that the extension of T_t to linear combinations is unique. For then c follows from 2.2, and a holds because every linear combination of characteristic functions in B_0, can be written with disjoint characteristic functions. Let us assume, then, that there exists numbers a_i and Borel sets A_i such that

$$\Sigma a_i \chi_{x_0^{-1}(A_i)} = 0 \quad \text{but} \quad \Sigma a_i \chi_{x_t^{-1}(A_i)} \neq 0 .$$

Thus there are k integers $i_1, \cdots, i_k$ with

$$\chi_{x_t^{-1}(B \cap A_i)} = 0 \quad \text{a.e.,} \ i \neq i_j$$

where

$$B = \bigcap_{j=1}^{k} A_{i_j}, \ P(x_t^{-1}(B)) > 0$$

and

$$\sum_{i=1}^{k} a_{i_i} \neq 0 .$$

But then, by 2.3,

$$\chi_{x_0^{-1}(B \cap A_i)} = 0 \quad \text{a.e.}$$

if $i \neq i_j$ and for $\omega \in x_0^{-1}(B)$

$$\Sigma a_i \chi_{x_0^{-1}(A_i)}(\omega) = \sum_{j=1}^{k} a_{i_j} \neq 0 \ .$$

This contradicts our assumption for

$$P(x_0^{-1}(B)) = P(x_t^{-1}(B)) \neq 0 \ .$$

REMARK. From a follows that T_t preserves inner products.

DEFINITION. *A Markov class is called stationary if there exist transformations T_t from B_0 to B_t satisfying a, b and c of Lemma 2.4.*

In the rest of the paper we will use the notation

$$\chi_{t,A} = \chi_{x_t^{-1}(A)}$$

3. Weak convergence. The main tool in this section is:

LEMMA 3.1. *Let B_t be a stationary Markov class. If $\bigcap_{n=0}^{\infty} B_n = 0$ then*

$$\text{weak lim } T_n x_0 = 0$$

for every $x_0 \in B_0$.

For the proof we need the following.

LEMMA 3.2. *Let B_t be a stationary Markow class, and $\bigcap_{n=0}^{\infty} B_n = 0$. If for some subsequence n_i, of the integers,*

$$\text{weak lim } T_{n_i} x_0 = x \neq 0$$

then

$$x = E_0 x + \sum_{n=1}^{\infty} (E_n - E_{n-1}) x$$

and the terms of the sum are mutually orthogonal.

Proof. Let $n < m$ then

$$(*) \qquad E_n E_m x = \text{weak} \lim_{i \to \infty} E_n E_m T_{n_i} x_0 = \text{weak} \lim_{i \to \infty} E_n T_{n_i} x_0 = E_n x$$

by Equation 2.1 Thus

$$(**) \qquad E_n(E_m x - E_{m-1} x) = E_n x - E_n x = 0 \ .$$

Now

$$\| E_N x \|^2 = \| E_0 x + \sum_{n=1}^{N} (E_n - E_{n-1}) x \|^2 = \| E_0 x \|^2 + \sum_{n=1}^{N} \| (E_n - E_{n-1}) x \|^2$$

hence the sum converges. Let

$$y = E_0 x + \sum_{n=1}^{\infty} (E_n - E_{n-1}) x \, .$$

If $z = E_n z \in B_n$ then by (**)

$$(y, z) = (E_n y, z) = (E_n x, z) = (x, z) \, .$$

Also if z is orthogonal to all the spaces B_n then

$$(y, z) = (x, z) = 0 \, .$$

Thus $y = x$.

LEMMA 3.3. *Under the same conditions, there exists a subsequence n'_i, of n_i, such that if $z_n \in B_0$ is defined by*

$$T_n z_n = E_n x / \| x \| \tag{***}$$

Then

$$\text{weak lim } z_{n'_i} = 0 \, .$$

Proof. Let $z_{n'_i}$ converges weakly to z. Such subsequence exists because a Hilbert space is weakly sequentially compact. Now $z \in B_0$, we shall prove that $z \in B_k$, for all k, and thus $z = 0$. Now, by equations (***) and 2.2

$$(T_k z_{n+k}, z_n) = (T_{n+k} z_{n+k}, T_n z_n) = (E_{n+k} x / \| x \|, E_n x / \| x \|) \xrightarrow[n \to \infty]{} 1 \, .$$

Hence

$$\| T_k z_{n+k} - z_n \|^2 \leqq 2 - 2(T_k z_{n+k}, z_n) \to 0 \, .$$

If $u \in L_2(\Omega)$ then

$$(T_k z_{n'_i+k}, u) = ((T_k z_{n'_i+k} - z_{n'_i}), u) + (z_{n'_i}, u) \to (z, u)$$

or

$$\text{weak lim } T_k z_{n'_i+k} = z$$

and by Hahn Banach Theorem $z \in B_k$.

Proof of Lemma 3.1. It is enough to show that for any subsequence n_i, there exists a subsequence n'_i, of n_i, such that

$$\text{weak lim } T_{n'_i} x_0 = 0 \, .$$

We may assume that $T_{n_i}x_0$ converges weakly to x. Let n_i' be chosen by Lemma 4.3. Then

$$\begin{aligned} 0 = \lim_{i\to\infty} (z_{n_i'}, x_0) &= \lim_{i\to\infty} (T_{n_i'}z_{n_i'}, T_{n_i'}x_0) \\ &= \lim_{i\to\infty} (E_{n_i'}x/\| x \|, T_{n_i'}x_0) = \| x \| \end{aligned}$$

For $E_{n_i}x$ tends strongly to x, by Lemma 3.2, and by assumption $T_{n_i}x_0$ converges weakly to x.

COROLLARY. *Let x_t be a stationary Markov process. If $\bigcap_{n=0}^{\infty} B_n = \{1\}$ then*

$$\text{weak lim } \chi_{n,A} = \| \chi_{0,A} \|^2 1 \ .$$

Proof. The Markov class $B_t - \{1\}$ satisfies the conditions of Lemma 3.1, hence

$$\text{weak lim } \chi_{n,A} - \| \chi_{n,A} \|^2 1 = 0 \ .$$

In the rest of this section let x_t be a given stationary Markov process. Let

$$C_0 = \bigcap_{n=0}^{\infty} B_n \ .$$

By Theorem 2.2

$$C_0 = \bigcap_{i=0}^{\infty} B_{t_i}$$

wherever $t_0 = 0$ and $t_i \to \infty$. Let

$$C_m = \bigcap_{n=m}^{\infty} B_n \quad \text{and} \quad D_m = B_m - C_m \ .$$

REMARK. $\{1\}$ stands for the space of constants. Also if B and C are subspaces $B - C$ is the orthogonal complement of C in B.

LEMMA 3.4. *For every integer n*

$$T_n C_0 = C_n \ , \quad T_n D_0 = D_n$$

and

$$C_n \subset C_{n+1} \ .$$

Proof. Let $x = T_m x_0$. The vector x belongs to C_m, if and only if, for every integer k there exists a vector $x_k \in B_0$ such that

$$x = T_{m+k} x_k \ .$$

But then

$$\|x\|^2 = (T_{m+k}x_k, T_m x_0) = (T_k x_k, x_0)$$

and $\|x_0\| = \|x\| = \|T_k x_k\|$. Hence $x_0 = T_k x_k$ and $x_0 \in B_k$ for all k: $x_0 \in C_0$. Now $y \in D_m$ if and only if $y = T_m y_0$ and

$$(y, x) = 0 \quad \text{if} \quad x \in C_m .$$

This is equivalent to

$$(T_m y_0, T_m x_0) = 0 \quad \text{if} \quad x_0 \in C_0, \quad \text{or} \quad (y_0, x_0) = 0 .$$

Thus $y \in D_m$ if and only if $y_0 \in D_0$.

LEMMA 3.5. *Both C_m and D_m are stationary Markov classes.*

Proof. The class C_m is Markov because $C_m \subset C_{m+1}$. Now let F_m be the projection on C_m and G_m the projection on D_m. Then

$$G_m = E_m(I - F_m) .$$

If $n \geqq m$ then $E_n F_m = F_m$ hence E_n and $I - F_m$ commute. Let $m_1 < m_2 < m_3$ then

$$\begin{aligned} G_{m_1}G_{m_2}G_{m_3} &= E_{m_1}(I - F_{m_1})E_{m_2}(I - F_{m_2})E_{m_3}(I - F_{m_3}) \\ &= E_{m_1}E_{m_2}E_{m_3}(I - F_{m_1})(I - F_{m_2})(I - F_{m_3}) \\ &= E_{m_1}E_{m_3}(I - F_{m_1})(I - F_{m_3}) = G_{m_1}G_{m_3} . \end{aligned}$$

We used Equation 2.1 and the fact that $I - F_m$ decreases with m.

THEOREM 3.6. *If $x \in D_0$ then $T_n x$ tends weakly to zero.*

Proof. The Markov class D_m satisfies the conditions of Theorem 3.1 for

$$\bigcap_{n=0}^{\infty} D_m \subset D_0 \cap \bigcap_{n=0}^{\infty} B_n = 0 .$$

It remains to study the monotone stationary Markov class C_m. Define

$$C_{-m} = T_m^{-1} C_0, \quad H = \bigcap_{m=1}^{\infty} C_{-m} .$$

REMARK. If C_0 is finite dimensional then $C_0 \subset C_m$ and both have same dimension:

$$C_0 = C_m \quad \text{and} \quad H = C_0 .$$

THEOREM 3.7. *If $x \in C_0$ is orthogonal to H then*

$$\text{weak}\lim_{n\to\infty} T_n x = 0$$

Proof. If $m > k$ then $C_{-m} \subset C_{-k}$: if $x \in C_{-m}$ then $T_m x \in C_0$. Let $y_0 \in C_0$ and $T_{m-k} y_0 = T_m x$ then

$$\| T_m x \|^2 = (T_m x, T_{m-k} y_0) = (T_k x, y_0)$$

Thus $y_0 = T_k x \in C_0$.

Now if F_{-m} is the projection of C_0 on C_{-m} then for each $x \in C_0$ $F_{-m}x$ converges to the projection of x on H (See [3] p. 266). Thus

$$x = \lim (I - F_{-m})x$$

or x is the limit of vectors orthogonal to C_{-m}.

Let us prove that

$$\text{weak}\lim_{n\to\infty} T_n x = 0$$

if x is orthogonal to C_{-m}, and because this is a dense set the theorem will follow.

The vector x is orthogonal to C_{-m}, and hence to C_{-m-p} for all p. Now

$$(T_{rm+d}x, T_d x) = (T_{rm}x, x)$$

but $x \in C_0$ and for some $y_0 \in C_0$, $x = T_{rm} y_0$ thus

$$(T_{rm+d}x, T_d x) = (T_{rm}x, T_{rm}y_0) = (x, y_0) = 0$$

for $y_0 \in C_{-rm}$. Thus the m sequences

$$\{T_{rm+d}x\} d = 0, 1, \cdots, m-1$$

consist of mutually orthogonal elements and thus converge weakly to zero.

It remains to study T on H.

THEOREM 3.8. *On the space H, T is a unitary operator and $T_n = T^n$.*

Proof. If $x \in H$ then $T_n x \in C_0$ for all n and it is possible to take $T_m(T_n x)$. But then

$$(T_{n+m}x, T_n(T_m x)) = \| T_m x \|^2$$

thus $T_{n+m}x = T_n(T_m x)$, or $T_n x = T^n x$. Thus if $y = Tx \in C_0$ then $T_n y = T_{n+1} x \in C_0$ and $y \in H$.

In order to show that T is unitary we have to show that it is onto. Let $x \in H$ then for some $x_0 \in C_0$ $Tx_0 = x$. But then $T_n x_0 = T_{n-1}x \in C_0$ and $x_0 \in H$.

In general the powers of a unitary operator do not converge. However the operator T has some special properties.

LEMMA 3.9. *If* $f \in L_2(\Omega)$ *and* $f \in H$ *then* $\chi_{f^{-1}(A)} \in H$ *for every Borel set* A.

Proof. In order to prove this we have to go back to the definitions of H and T. Now, if $f \in B_n$ and A is a Borel set, then $f^{-1}(A) = x_n^{-1}(A_n)$ for some A_n and thus $\chi_{f^{-1}(A)} \in B_n$. Thus $f \in C_0$ implies that $\chi_{f^{-1}(A)} \in C_0$. But $f \in H$ so $T_n f \in H$. The Lemma will be proved if we show that

$$T_n \chi_{f^{-1}(A)} = \chi_{(T_n f)^{-1}(A)} \quad \text{a.e.}$$

If $M \leqq f \leqq N$ then $M \leqq T_n f \leqq N$, thus it is enough to prove the above equation under the assumption that A is a bounded set and f a bounded function. If f is bounded (hence $T_n f$ is bounded also) it defines a self adjoint operator on $L_2(\Omega)$,: the multiplication operator. Thus as an operator

$$f = \int \lambda \chi_{f^{-1}}(d\lambda)$$

$$T_n f = \int \lambda T_n \chi_{f^{-1}(d\lambda)} = \int_{(T_n f)^{-1}(d\lambda)} .$$

Now T_n transforms characteristic functions to characteristic functions and $T_n \chi_{f^{-1}(A)}$, $\chi_{(T_n f)^{-1}(A)}$ are both the spectral measure of $T_n f$. Thus

$$T_n \chi_{f^{-1}(A)} = \chi_{(T_n f)^{-1}(A)} . \quad \text{a.e.}$$

This lemma shows that H is generated by characteristic functions. Let us study the limits of $T_n x$ when x is a characteristic function.

LEMMA 3.10. *Let* H *be generated by a countable number of disjoint characteristic functions* χ_i. *For a given* χ_i *there is an integer* m: $T_m \chi_i = \chi_i$ *and then*

$$T_{rm+d} \chi_i = T_d \chi_i .$$

Proof. For every n $T_n \chi_i$ is a characteristic function, hence either $T_n \chi_i = \chi_i$ or

$$(T_n \chi_i, \chi_i) = 0 .$$

If $(T_n \chi_i, \chi_i) = 0$ for all n then $(T_m \chi_i, T_n \chi_i) = (T_{m-n} \chi_i, \chi_i) = 0$ thus there exist infinitely many disjoint sets of equal measure which is impossible.

Now if for some m, $T_m \chi_i = \chi_i$, let m be the smallest integer that this happens. Then

$$T_{rm+d}\chi_i = T^d T^{rm}\chi_i = T^d\chi_i = T_d\chi_i \ .$$

THEOREM 3.11. *Let* x_t *be a stationary Markov process. If* H *is generated by a countable collection of disjoint characteristic functions* $\{\chi_i\}$ *then for every* $y \in B_0$ *such that* $(y, \chi_i) \neq 0$ *for finitely many* i's *(*y *has a "finite" support), there exists an integer* m *such that the* m *sequences*

$$\{T_{km+d}y\} \qquad d = 1, 2\cdots, m$$

converge weakly.

Proof. From Theorems 3.6 and 3.7 it follows that

$$\text{weak lim } T_n(y - \Sigma(y, \chi_i)\|\ \chi_i\|^{-2}\chi_i) = 0 \ .$$

Let $\chi_{i_1}, \chi_{i_2}, \cdots, \chi_{i_n}$ be those functions for which $(y, \chi_i) \neq 0$. Now $T^{m_j}\chi_{i_j} = \chi_{i_j}$. Choose m to be the product of this m_j. Thus

$$T_{km+d}\chi_{i_j} = T^d\chi_{i_j} \ .$$

Hence

$$\text{(3.1)} \qquad \underset{k\to\infty}{\text{weak lim}}\ T_{km+d}y = \underset{k\to\infty}{\text{weak lim}}\ T_{km+d}\Sigma(y, \chi_i)\,\|\chi_i\|^{-2}\chi_i$$
$$= \Sigma(y, \chi_i)\,\|\chi_i\|^{-2}T^d\chi_i.$$

COROLLARY 1. *Equation* 3.1 *holds if the function* x_0 *has countable range.*

This is a classical theorem see [2] p. 353.

COROLLARY 2. *If there exists a finite measure* φ*, on the line, such that, for some* $\varepsilon > 0$, $\varphi(A) \leqq \varepsilon$ *implies that*

$$E_0\chi_{n,A} \neq \chi_{n,A}$$

for some n*, then the space* H *is generated by a finite number of disjoint characteristic functions. Thus an integer* m *exists, such that Equation* 3.1 *holds for all* $y \in B_0$.

Proof. Let k be an integer greater or equal to $\varphi(\Omega)\varepsilon$. If $\chi_{0,A_i} \in H$ $i = 1, \cdots, k$ where the A_i are disjoint then

$$\varphi(\Omega) \geqq \Sigma\varphi(A_i) \geqq \min\,(\varphi(A_i))k$$

or $\varphi(A_{i_0}) \leqq \varphi(\Omega)/k \leqq \varepsilon$ for some i_0. But then, for some n, $\chi_{n,A_{i_0}} \notin H$ hence

$$\chi_{0,A_{i_0}} \notin H .$$

Thus there are at most $k-1$ disjoint characteristic functions that generate H.

REMARK. This last corollary is similiar to Doeblin's condition as given in [1] page 192.

4. **Strong convergence.** Throughout this section we assume:

4.1. *There exists a real number* $t_0 > 0$ *such that the space* $B_0 \cap B_{t_0}$ *is finite dimensional and there is a positive angle between* $B_{t_0} - B_0 \cap B_{t_0}$ *and* $B_0 - B_0 \cap B_{t_0}$.

Two subspaces, B^* and B^{**}, are said to have a positive angle between them if

$$\sup \{(b^*, b^{**}) \,|\, \| b^* \| = \| b^{**} \| = 1 \text{ and } b^* \in B^*, b^{**} \in B^{**}\} < 1 .$$

CONDITION 4.1. Is equivalent to each of the following.

(a) The point 1 is not in the essential spectrum of $E_0 E_{t_0} E_0$(or $E_{t_0} E_0 E_{t_0}$).

(b) The operator $E_0 E_{t_0} E_0$ (or $E_{t_0} E_0 E_{t_0}$) is quasi compact.

(c) The operator $E_0 E_{t_0} E_0$(or $E_{t_0} E_0 E_{t_0}$) is a sum of a compact operator and an operator of norm less than 1.

(d) The norm of E_0 restricted to $B_{t_0} - B_0 \cap B_{t_0}$ is less than one.

LEMMA 4.1. *If* $t > t_0$ *then Condition* 4.1 *is satisfied when* B_{t_0} is replaced by B_t.

Proof. Let us use the form given in c for 4.1. Now

$$E_t E_0 E_t = E_t (E_{t_0} E_0 E_{t_0}) E_t$$

by Equation 2.1, hence it is a sum of a compact and an operator of norm less than 1.

Now from Theorem 2.2 it follows that $B_0 \cap B_t$ decreases with t. Let t_1 be such that

$$\dim (B_0 \cap B_{t_1}) \leqq \dim (B_0 \cap B_t) \text{ for all } t.$$

It is easy to see that $B_0 \cap B_{t_1}$ is generated by a finite number of disjoint characteristic functions. Let them be $\chi_1, \cdots, \chi_k$, thus

$$B_0 \cap B_{t_1} = B_0 \cap B_t = \operatorname{span} \{\chi_1, \cdots, \chi_k\} \qquad t > t_1 .$$

because by Theorem 2.2

$$B_0 \cap B_{t_1} \supset B_0 \cap B_t$$

and they have the same dimension.

LEMMA 4.2. *If $t > 0$ then*

$$T_t(B_0 \cap B_{t_1}) = B_0 \cap B_{t_1}$$

and

$$T_t(B_0 - B_0 \cap B_{t_1}) = B_t - B_0 \cap B_{t_1} = B_t - B_0 \cap B_t .$$

Proof. A vector $x \in B_0 \cap B_{t_1}$, if and only if, $x \in B_0$ and $x = T_{t_1}y$ for some $y \in B_0$. But then

$$(T_t x, T_{t+t_1}y) = (x, T_{t_1}y) = \| x \|^2 = \| T_t x \|^2$$

or

$$T_t x = T_{t+t_1}y: \quad T_t x \in B_t \cap B_{t+t_1} .$$

Thus

$$T_t(B_0 \cap B_{t_1}) = B_t \cap B_{t+t_1} \supset B_0 \cap B_{t+t_1} = B_0 \cap B_{t_1}$$

by Theorem 2.2 and the remark above. This shows that

$$T_t(B_0 \cap B_{t_1}) = B_0 \cap B_{t_1} .$$

Let $x \in B_0$ be orthogonal to $B_0 \cap B_{t_1}$. If $y \in B_0 \cap B_{t_1}$, then $y = T_t z$ where $z \in B_0 \cap B_{t_1}$. Thus

$$(T_t x, y) = (T_t x, T_t z) = (x, z) = 0 .$$

THEOREM 4.3. *Let $x \in B_0$ and let $c =$ norm of E_0 restricted to $B_{t_1} - B_0 \cap B_{t_1}$.*
Then $c < 1$ and

$$\| E_0 T_t x - \sum_{i=1}^{k}(x, \chi_i) \| \chi_i \|^{-2} T_t \chi_i \| \leqq c^n \| x \| \tag{4.2}$$

where n is an integer such that $nt_1 < t$.

Proof. The vector $x - \sum_{i=1}^{k}(x, \chi_i) \| \chi_i \|^{-2}\chi_i$ is orthogonal to $B_0 \cap B_{t_1}$ and hence so is

$$y = T_t x - \Sigma(x, \chi_i) \| \chi_i \|^{-2} T_t \chi_i .$$

Thus

$$\| E_0 y \| = \| E_0 E_t y \| = \| E_0 E_{t_1} E_{2t_1} \cdots E_{nt_1} y \| .$$

Now the norm E_{jt_1} restricted to $B_{(j+1)t_1} - B_0 \cap B_{t_1}$ is equal to c hence

$$\| E_0 y \| \leqq c^n \| y \| \leqq c_n \| x \| .$$

It becomes now interesting to study $T_t \chi_i$.

THEOREM 4.4. *For each given t there is a permutation of the integer $1, 2, \cdots, k, \pi_t$, such that*

$$T_t\chi_i = \chi_{\pi_t i} .$$

Also there exists an integer m such that

$$T_{mt}\chi_i = \chi_{(\pi_t i)^m} = \chi_i \quad \text{for all } i .$$

Proof. Let us use induction on k. Let $\chi_{i_1}, \chi_{i_2}, \cdots, \chi_{i_j}$ be a subset of χ_i, $i = 1, \cdots, k$, with minimum norm: $\| \chi_{i_r} \| \leqq \| \chi_i \|$. Then $T_t\chi_{i_r}$ is a characteristic function in $B_0 \cap B_{t_1}$ with norm smaller or equal to the norm of $\chi_1, \chi_2, \cdots, \chi_k$: $T_t\chi_{i_r} \in (\chi_{i_1}, \cdots, \chi_{i_j})$.

This shows that T_t maps the set $(\chi_{i_1}, \cdots, \chi_{i_k})$ into, therefore onto, itself. If χ_i is not in this set then $T_t\chi_i$ will be also, orthogonal to χ_{i_r}. In the remaining set there are less than k functions and by induction the first part of the theorem is proved. The second part is an easy result on permutations.

The last two theorems include the classical result on Markov processes with a finite number of states. There might be a connection to Theorem 11 of [4].

If $\dim B_0 \cap B_{t_1} = 1$ then

$$\| T_t x - (x, 1)1 \| \leqq c^n \| x \|$$

where $nt_1 < t$ and 1 is χ_Ω. This is a similiar to the case of independent functions. Let us conclude this section by studying this case. Thus let B_1 and B_2 be two subspaces of $L_2(\Omega)$ generated by characteristic functions χ_A and $\chi_{A'}$, where A and A' belong to some σ subalgebras of Σ. The cosine of the angle between $B_1 - \{1\}$ and $B_2 - \{1\}$, c, is given by

$$(*) \qquad c = \sup\{(\Sigma a_i\chi_{A_i}, \Sigma a_i'\chi A_i') \,|\, 1 = \Sigma a_i^2 P(A_i) = \Sigma a_i^2 P(A_i')$$

and

$$\Sigma a_i P(A_i) = \Sigma a_i' P(A_i') = 0\} .$$

THEOREM 4.5. *The number c is smaller than*

1. $\sup | (P(A \cap A') - P(A)P(A'))P(A \cap A')^{-1} | = c_1 .$

2. $\sup | (P(A \cap A') - P(A)P(A')(P(A)P(A'))^{-1} | = c_2 .$

Where A and A' belong to the σ subalgebras generating B_1 and B_2 respectively.

Proof. Let us show that $c \leqq c_1$, the other inequality is proved in a similiar way. Now let a_i, a_i', A_i and A_i' satisfy the conditions of equation (*). Then

$$\sum_{i,j} a_i a'_j P(A_i \cap A_j) = \sum_{i,j} a_i a'_j (P(A_i \cap A'_j) - P(A_i)P(A'_j)) + \sum_{i,j} a_i a_j P(A_i)P(A_j) .$$

The second term is equal to zero. Thus

$$\begin{aligned} | \Sigma a_i b'_j P(A_j \cap A'_j) | &\leqq c_1 \sum_{i,j} | a_i a'_j | P(A_i \cap A'_j) \\ &\leqq c_1 \Big(\sum_{i,j} a_i^2 P(A_i \cap A'_j) \Big)^{1/2} \Big(\sum_{i,j} a'^2_j P(A_i \cap A'_j) \Big)^{1/2} \\ &= c_1 \Big(\sum_i a_i^2 P(A_i) \Big)^{1/2} \Big(\sum_j a'^2_j P(A'_j) \Big)^{1/2} = c_1 . \end{aligned}$$

A more convenient form of the conditions of Lemma 3.2 is

1. c_1 is the largest number for which

$$(1 + c_1)^{-1} \leqq P(A \cap A')(P(A)P(A'))^{-1} \leqq (1 - c_1)^{-1} .$$

2. c_2 is the largest number for which

$$1 - c_2 \leqq P(A \cap A')(P(A)P(A'))^{-1} \leqq 1 + c_2 .$$

BIBLIOGRAPHY

1. J. L. Doob, *Stochastic Processes*, Wiley, New York, 1953.
2. W. Feller, *An Introduction to Probability Theory*, Wiley New York, 1957.
3. F. Riesz, and B. Sz. Nagy, *Lecons D'analyse Fonctionelle*, Budapest, 1953.
4. K. Yosida, and S. Kakutani, *Operator theoretical treatment of Markoff's process and mean ergodic theorem*, Ann. Math. **42**, (1941) 188-228.

UNIVERSITY OF CALIFORNIA
BERKELEY CALIFORNIA

SOME ZERO SUM TWO-PERSON GAMES WITH MOVES IN THE UNIT INTERVAL

MARTIN FOX

Introduction. Consider the following zero sum two person game. The players alternately choose points $t_i \in [0, 1]$ for $i = 1, 2, \cdots, n$, the choice being made by player I if i is odd and by player II if i is even. After the ith move the player who is to make the $(i+1)$st move observes the value of $\phi_i(t_1, t_2, \cdots, t_i)$ where ϕ_i is some function on the i-dimensional closed unit cube to some set A_i. The payoff is $f(t_1, t_2, \cdots, t_n)$ where f is a continuous, real-valued function.

If all the ϕ_i are constant we have the case of no information. Ville [1] showed that in this case such a game has a value. At the other extreme, if the ϕ_i are all one-to-one we have the case of perfect information so the game has a value.

The purpose of the present paper is to show that, in general, games of the form introduced in the first paragraph do not have values and to consider two cases in which they do. The counter-examples to be presented will be compared with Ville's classical example of a game on the unit square which has no value.

It is shown in § 2 that the games considered always have values when $n = 2$.

An example of a game with no value is presented in § 3. In this example $n = 3$ and the ϕ_i take only a finite number of values.

In § 4 it is shown that the additional hypothesis of continuity of the ϕ_i is not sufficient to guarantee existence of a value. In that example $n = 4$. The case $n = 3$ with continuous ϕ_i remains unsolved.

Section 5 deals with a special case for which n is arbitrary and yet the game has a value. In this case the ϕ_i each take only a finite number of values and each is constant on sets which are finite unions of i-dimensional generalized intervals.

1. Preliminary remarks. In this section the notation to be used in this paper will be introduced. This will be facilitated by the introduction of the normal forms of the games under consideration.

A pure strategy for player I is a vector $x = (x_1, x_2, \cdots, x_{[(n+1)/2]})$ where $x_1 \in [0, 1]$ and the x_i for $i = 2, 3, \cdots, [(n+1)/2]$ are functions on A_{2i-2} to $[0, 1]$. If moves $t_1, t_2, \cdots, t_{2i-2}$ have been made, then the ith move made by player I (the $(2i-1)$st move in the game) will be $x_i(\phi_{2i-2}(t_1, t_2, \cdots, t_{2i-2}))$. His first move will be x_1.

Received November 16, 1959. Excerpt from dissertation for the degree of Doctor of Philosophy, University of California, 1959. Partial support was given by Office of Ordnance Research, U.S. Army under Contract DA-04-200-ORD-171, Task Order 3.

A pure strategy for player II is a vector $y = (y_1, y_2, \cdots, y_{[n/2]})$ where each y_i is a function on A_{2i-1} to $[0, 1]$. If moves $t_1, t_2, \cdots, t_{2i-1}$ have been made, then the ith move made by player II (the $(2i)$th move in the game) will be $y_i(\phi_{2i-1}(t_1, t_2, \cdots, t_{2i-1}))$.

When player I uses the pure strategy x and player II uses the pure strategy y let $t_i(x, y)$ be the ith move made in the game. The t_i are defined recursively as follows:

$$\begin{aligned} t_1(x, y) &= x_1 ; \\ t_{2i}(x, y) &= y_i(\phi_{2i-1}(t_1(x, y), t_2(x, y), \cdots, t_{2i-1}(x, y))) \\ &\qquad \text{for } i = 1, 2, \cdots, [n/2] ; \\ t_{2i-1}(x, y) &= x_i(\phi_{2i-2}(t_1(x, y), t_2(x, y), \cdots, t_{2i-2}(x, y))) \\ &\qquad \text{for } i = 2, 3, \cdots, [(n + 1)/2] . \end{aligned}$$

The payoff function is given by $M(x,y) = f(t_1(x,y), t_2(x,y), \cdots, t_n(x,y))$. The payoff as a function of mixed strategies will also be denoted by M.

In our case, since the moves are points in $[0, 1]$, the strategy spaces X and Y are products, usually infinite dimensional, each coordinate space being $[0, 1]$. Hence, the choice of a strategy by player I is equivalent to the choice of a distribution function F on X. It will be convenient to let the space P of mixed strategies for player I be the family of all distribution functions on X which assign probability 1 to a finite subset of X. The same will be done for Q, the space of mixed strategies for player II.

If H is a distribution function on the real line and S is any subset of the real line which is Borel measurable, we will let HS be the probability assigned to S by H.

For $F \in P$ we let $F_{i,\alpha}$ denote the marginal distribution function of the coordinate of player I's strategy which corresponds to his ith move when $\phi_{2i-2} = \alpha$. Similar notation will be used for $G \in Q$.

2. The case $n = 2$. In this section it will be shown that any game $\mathscr{G}$ of the type given in the introduction for which $n = 2$ has a value. It is not even necessary to assume that ϕ_1 is a measurable function.

For any $\alpha \in A_1$ let $\mathscr{G}(\alpha) = (\phi_1^{-1}(\alpha), [0, 1], M_\alpha)$ where M_α is f restricted to $\phi_1^{-1}(\alpha) \times [0, 1]$. It follows by the proof used for Ville's minimax theorem that each $\mathscr{G}(\alpha)$ has a value $v(\alpha)$. Let

$$v = \sup_{\alpha \in A_1} v(\alpha) .$$

Fix $\varepsilon > 0$ and let α^* be such that $v(\alpha^*) > v - \varepsilon$. For each $\alpha \in A_1$ let $F^{(\alpha)}$ and $G^{(\alpha)}$ be ε-good strategies for players I and II, respectively, in $\mathscr{G}(\alpha)$. The distribution function $F^{(\alpha)}$ assigns probability 1 to a finite subset of $\phi_1^{-1}(\alpha)$. Since $F^{(\alpha^*)}$ is a distribution function on $[0, 1]$ which is the strategy space for player I in $\mathscr{G}$, it can also be used as a strategy in $\mathscr{G}$. Let y be any pure strategy for player II in $\mathscr{G}$. Since $y_1(\alpha^*) \in [0, 1]$, it follows that $y_1(\alpha^*)$ is a pure strategy for player II in $\mathscr{G}(\alpha^*)$. Hence,

$$\begin{aligned} M(F^{(\alpha^*)}, y) &= \int_{\phi_1^{-1}(\alpha^*)} f(t, y_1(\alpha^*))F^{(\alpha^*)}(dt) \\ &= \int M_{\alpha^*}(t, y_1(\alpha^*))F^{(\alpha^*)}(dt) \\ &= M_{\alpha^*}(F^{(\alpha^*)}, y_1(\alpha^*)) \\ &> v(\alpha^*) - \varepsilon > v - 2\varepsilon . \end{aligned}$$

Let G be any strategy for player II in $\mathscr{G}$ such that $G_{1,\alpha} = G^{(\alpha)}$ for all $\alpha \in A_1$. Let x be any pure strategy for player I in $\mathscr{G}$. For some $\alpha \in A_1$ it must be true that $x \in \phi_1^{-1}(\alpha)$ so that x is also a pure strategy for player I in $\mathscr{G}(\alpha)$. Then,

$$\begin{aligned} M(x, G) &= \int f(x, t)G_{1,\alpha}(dt) \\ &= \int M_\alpha(x, t)G^{(\alpha)}(dt) \\ &= M_\alpha(x, G^{(\alpha)}) \\ &< v(\alpha) + \varepsilon \leqq v + \varepsilon . \end{aligned}$$

From the two inequalities obtained above it follows that the value of $\mathscr{G}$ is v.

3. A counter-example for $n = 3$. In this section the counter-example for $n = 3$ will be given. The functions ϕ_i $(i = 1, 2)$ each take only a finite number of values. The similarity of this example to Ville's example will be discussed.

For this example let

$$\phi_1(t_1) \equiv 0 ;$$

$$\phi_2(t_1, t_2) = \begin{cases} -1 \text{ if } t_1 = 0 \text{ or } 0 < \min(t_2, 1 - t_2) \leqq t_1 ; \\ t_2 \text{ if } t_2 = 0 \text{ or } 1 \text{ and } t_1 \neq 0 ; \\ 2 \text{ if } 0 < t_1 < t_2 \leqq \dfrac{1}{2} \\ 3 \text{ if } 0 < t_1 < 1 - t_2 < \dfrac{1}{2} \end{cases}$$

$$f(t_1, t_2, t_3) = -| t_3 - t_2 |.$$

Let F be any strategy for player I. Fix $\varepsilon > 0$ and let $\delta \in (0, \varepsilon)$ be sufficiently small so that $F_1(0, \delta) < \varepsilon$. Let $G\{\delta\} = G\{1 - \delta\} = 1/2$. Then,

$$M(F, G) \leqq -\frac{1}{2}(F_1[\delta, 1] + F_1\{0\})\Big[\int | t_3 - \delta | F_{2,-1}(dt_3)$$

$$+ \int | t_3 - (1 - \delta) | F_{2,-1}(dt_3)\Big]$$

$$< -\frac{1}{2}(1 - \varepsilon)\Big[\Big(\frac{1}{2} - \delta\Big) + \Big(1 - \delta - \frac{1}{2}\Big)\Big] < -\frac{1}{2} + \frac{3}{2}\varepsilon$$

so that

$$\sup_F \inf_G M(F, G) \leqq -\frac{1}{2}.$$

Let G be any strategy for player II. Fix $\varepsilon > 0$ and let $x_1 \in (0, 1/2)$ be sufficiently small so that $G(0, x_1] + G[1 - x_1, 1) < \varepsilon$.
Let

$$x_2(\alpha) = \begin{cases} \frac{1}{2} \text{ if } \alpha = -1 ; \\ \alpha \text{ if } \alpha = 0 \text{ or } 1 ; \\ \frac{1}{4} \text{ if } \alpha = 2 ; \\ \frac{3}{4} \text{ if } \alpha = 3 . \end{cases}$$

Let $x = (x_1, x_2)$ so that x is a pure strategy for player I. Then,

$$M(G, x) \geqq -\int_{(0,x_1]}\Big(\frac{1}{2} - t_2\Big)G(dt_2) - \int_{[1-x_1,1)}\Big(t_2 - \frac{1}{2}\Big)G(dt_2)$$

$$-\int_{[0,1/2]}\Big|\frac{1}{4} - t_2\Big|G(dt_2) - \int_{(1/2,1]}\Big|\frac{3}{4} - t_2\Big|G(dt_2)$$

$$> -\varepsilon - \frac{1}{4}$$

so that

$$\inf_G \sup_F M(F, G) \geqq -\frac{1}{4}$$

and the game has no value.

In Ville's example the payoff function is such as to force each player to attempt to choose a point closer to 1 than does his opponent without actually choosing 1. It is impossible for either player to guarantee he will achieve this with any preassigned positive probability no matter what pure strategy his opponent may use. In the example just presented a similar situation arises on the first two moves. In Ville's example the competition to choose a point close to the endpoint is a direct competition over payoff. In the present example this competition is over the information player I will receive, which, of course, helps determine the payoff. If on his first move player I chooses a point closer to 0 (but not 0) than the choice of his opponent is to both 0 and 1, then he will obtain more accurate information about the location of his opponent's choice than would be the case otherwise. Player II is prevented from choosing an endpoint since to do so would be to give his opponent perfect information.

4. A counter-example with continuous ϕ_i. In this section a counter-example will be presented in which the functions ϕ_i are all continuous. In this example $n = 4$. Again a comparison will be made with Ville's example.

Let

$$\phi_1(t_1) \equiv 0 \,;$$

$$\phi_2(t_1, t_2) = t_1(1 - t_1)t_2 \,;$$

$$\phi_3(t_1, t_2, t_3) = \begin{cases} 0 \text{ if } \min(t_1, 1 - t_1) \leqq t_2 \leqq \max(t_1, 1 - t_1) \,; \\ t_2(1 - t_2)(t_1 - t_2)\left| t_1 - \dfrac{1}{2} \right| \text{ if } t_2 < t_1 < \dfrac{1}{2} \\ \qquad\qquad\qquad\qquad \text{or } \dfrac{1}{2} < t_1 < t_2 \,; \\ t_2(1 - t_2)[t_1 - (1 - t_2)]\left| t_1 - \dfrac{1}{2} \right| \text{ if } \dfrac{1}{2} \leqq t_1 < 1 - t_2 \\ \qquad\qquad\qquad\qquad \text{or } 1 - t_2 < t_1 \leqq \dfrac{1}{2} \,; \end{cases}$$

$$f(t_1, t_2, t_3, t_4) = | t_1 - t_4 | - 10 \, | t_2 - t_3 | \,.$$

Assume $t_2 \neq 0$ or 1. Then, $\phi_3(t_1, t_2, t_3) > 0$ for $\min(t_2, 1 - t_2) < t_1 < 1/2$ while $\phi_3(t_1, t_2, t_3) < 0$ for $1/2 < t_1 < \max(t_2, 1 - t_2)$. On the other hand, $\phi_3(t_1, t_2, t_3) = 0$ otherwise.

Let F be any strategy for player I. Fix $\varepsilon > 0$ and let $\delta \in (0, \varepsilon)$ be sufficiently small so that $F_1(0, \delta] + F_1[1 - \delta, 1) < \varepsilon$. Let

$$y_2(\alpha) = \begin{cases} \dfrac{1}{2} & \text{if } \alpha = 0 \,; \\ \dfrac{1}{4} & \text{if } \alpha > 0 \,; \\ \dfrac{3}{4} & \text{if } \alpha < 0 \,. \end{cases}$$

Let G assign probability 1/2 to each of the pure strategies (δ, y_2) and $(1 - \delta, y_2)$. Then,

$$\begin{aligned}
M(F, G) \leqq & \int_{[0,\delta]}\left(\frac{1}{2} - t_1\right)F_1(dt_1) + \int_{[1-\delta,1]}\left(t_1 - \frac{1}{2}\right)F_1(dt_1) \\
& + \int_{(\delta,1/2)}\left|t_1 - \frac{1}{4}\right|F_1(dt_1) + \int_{(1/2,1-\delta)}\left|t_1 - \frac{3}{4}\right|F_1(dt_1) \\
& - 10[F_1\{0\} + F_1\{1\}]\left[\frac{1}{2}\int|\delta - t_3|\,F_{2,0}(dt_3) \right.\\
& \left. + \frac{1}{2}\int|1 - \delta - t_3|\,F_{2,0}(dt_3)\right] \\
< & \frac{1}{2}[F_1\{0\} + F_1\{1\}] + \frac{1}{2}\varepsilon + \frac{1}{4}[1 - \varepsilon - F_1\{0\} - F_1\{1\}] \\
& -5[F_1\{0\} + F_1\{1\}]\left[\left(\frac{1}{2} - \delta\right) + \left(1 - \delta - \frac{1}{2}\right)\right] \\
& = \frac{1}{4} + \frac{1}{4}\varepsilon - [F_1\{0\} + F_1\{1\}]\left[5(1 - 2\delta) - \frac{1}{4}\right] \\
< & \frac{1}{4} + 11\varepsilon
\end{aligned}$$

so that $\sup_F \inf_G M(F, G) \leqq 1/4$.

Let G be any strategy for player II. Fix $\varepsilon > 0$ and let $\delta \in (0, \varepsilon) \cap (0, 1/2)$ be sufficiently small so that $G_{1,0}(0, \delta) + G_{1,0}(1 - \delta, 1) < \varepsilon$. Let $x_2(\alpha) = \alpha/[\delta(1 - \delta)]$ and let F assign probability 1/2 to each of the pure strategies (δ, x_2) and $(1 - \delta, x_2)$. When player I uses the strategy F the value of the nonpositive term in f will always be zero. Thus,

$$\begin{aligned}
M(F, G) \geqq & \left[1 - G_{1,0}(0, \delta) - G_{1,0}(1 - \delta, 1)\right] \\
& \times \left[\frac{1}{2}\int|\delta - t_4|\,G_{2,0}(dt_4) + \frac{1}{2}\int|1 - \delta - t_4|\,G_{2,0}(dt_4)\right] \\
> & \frac{1}{2}(1 - \varepsilon)\left[\left(\frac{1}{2} - \delta\right) + \left(1 - \delta - \frac{1}{2}\right)\right] \\
> & \frac{1}{2} - \frac{3}{2}\varepsilon
\end{aligned}$$

so that $\inf_G \sup_F M(F, G) \geqq 1/2$ and the game has no value.

Here again the primary competition between the players is to make their first moves as close to the endpoints as possible without actually choosing the endpoints. If player I is successful in choosing a point t_1 at least as close to one of the endpoints as is player II's choice, then player II will have less information about t_1 than would be the case otherwise. Player I is prevented from choosing an endpoint by the fact that if he does so he will get no information about his opponent's first move so that he cannot guarantee that he can keep the negative term close to zero. Player II is prevented from choosing an endpoint by the fact that when he does so the function ϕ_3 will take the value zero no matter what his opponent does so that he will have no information about player I's first move.

5. The case of information sets which are unions of generalized intervals. The case to be considered here is that in which each ϕ_i takes only a finite number of values and each is constant only on sets which are finite unions of i-dimensional generalized intervals. This is the only case considered in this paper in which n remains arbitrary.

Let the values of ϕ_i be $1, 2, \cdots, m_i$. Let $P_j\phi_i^{-1}(k)$ be the projection on the jth coordinate of $\phi_i^{-1}(k)$ where $j = 1, 2, \cdots, i$. The interval $[0, 1]$ can be subdivided into disjoint sets $B_{j1}, B_{j2}, \cdots, B_{jl_j}$ such that for each B_{jl} there exist $i_1, i_2, \cdots, i_r$ and $k_1, k_2, \cdots, k_u$, all integers, such that $t \in B_{jl}$ if, and only if, $t \in P_j\phi_i^{-1}(k)$ whenever $i \in \{i_1, i_2, \cdots, i_r\}$ and $k \in \{k_1, k_2, \cdots, k_u\}$ while $t \notin P_j\phi_i^{-1}(k)$ otherwise. Suppose j is even so that player II makes the jth move. Let $y = (y_1, y_2, \cdots, y_{[n/2]})$ and $y' = (y'_1, y'_2, \cdots, y'^{[n/2]})$ be any strategies for player II such that $y_i = y'_i$ for $i \neq j/2$ and if $y_{j/2}(k) \in B_{jl}$, then $y'_{j/2}(k) \in B_{jl}$. For any pure strategy x for player I we have $t_i(x, y) = t_i(x, y')$ for $i = 1, 2, \cdots, j-1$ since for these values of i player II's moves are unchanged. If $t_j(x, y) \in B_{jl}$, then $t_j(x, y') \in B_{jl}$. Hence,

$$\phi_j(t_1(x, y), t_2(x, y), \cdots, t_j(x, y)) = \phi_j(t_1(x, y'), t_2(x, y'), \cdots, t_j(x, y'))$$

so that $t_{j+1}(x, y) = t_{j+1}(x, y')$. Suppose that $t_i(x, y) = t_i(x, y')$ for $i = j+1, j+2, \cdots, i_0$. Then, $\phi_{i_0}(t_1(x, y), t_2(x, y), \cdots, t_{i_0}(x, y)) = \phi_{i_0}(t_1(x, y'), t_2(x, y'), \cdots, t_{i_0}(x, y'))$ so that $t_{i_0+1}(x, y) = t_{i_0+1}(x, y')$. Thus, $t_i(x, y) = t_i(x, y')$ for all $i \neq j$.

For each $j = 1, 2, \cdots, n-1$ fix $\delta_j > 0$ and select points $t_{j1}, t_{j2}, \cdots, t_{jv_j}$ such that for any $t_j \in B_{jl}$ there exists $t_{jv} \in B_{jl}$ such that for any $t_1, t_2, \cdots, t_{j-1}, t_{j+1}, \cdots, t_n$ we have

$$\begin{aligned} | f(t_1, t_2, \cdots, t_{j-1}, t_j, t_{j+1}, \cdots, t_n) \\ - f(t_1, t_2, \cdots, t_{j-1}, t_{jv}, t_{j+1}, \cdots, t_n) | < \delta_j . \end{aligned}$$

Select the t_{jv} in such a way that as $\delta_j \downarrow$ the set of all the t_{jv} increases monotonically.

Let the game $\mathscr{G}(\delta_1, \delta_2, \cdots, \delta_i) = (X(\delta_1, \delta_2, \cdots, \delta_i), Y(\delta_1, \delta_2, \cdots, \delta_i), M_{\delta_1, \delta_2, \cdots, \delta_i})$ be our original game with the jth move for $j = 1, 2, \cdots, i$ restricted to $t_{j1}, t_{j2}, \cdots, t_{jv_j}$. In $\mathscr{G}(\delta_1, \delta_2, \cdots, \delta_{n-1})$ the player who makes the $(n-1)$st move has only a finite number of strategies so that $\mathscr{G}(\delta_1, \delta_2, \cdots, \delta_{n-1})$ has a value (see Wald [2]).

Suppose $\mathscr{G}(\delta_1, \delta_2, \cdots, \delta_{i-1}, \delta_i)$ has a value for all $\delta_i > 0$. It follows, by a proof similar to Ville's, that $\mathscr{G}(\delta_1, \delta_2, \cdots, \delta_{i-1})$ has a value. Thus, by induction, $\mathscr{G}$ will also have a value.

Acknowledgment. The author wishes to express his gratitude to Professor David Blackwell for suggesting the problem treated here and for his continued interest in its solution.

References

1. J. Ville, *Sur la théorie générale des jeux où intervient l'habilité des joueurs,* note in *Traité du Calcul des Probabilites et de ses Applications,* by E. Borel, Tome IV, Fascicule II, Applications aux jeux de hasard, (1938), 105–113.

2. A. Wald, *Generalization of a theorem by von Neumann concerning zero sum two person games,* Ann. Math. **46** (1945), 281–286.

Michigan State University

SINGULARITIES OF THREE-DIMENSIONAL HARMONIC FUNCTIONS

R. P. GILBERT

Introduction. Recently G. Szegö [9] and Z. Nehari [8] have obtained some interesting results connecting the singularities of axially symmetric harmonic functions with those of analytic functions. In this paper we shall show that a similar connection also exists between the singularities of a three-dimensional harmonic function and a function of two complex variables. We may do this by considering the Whittaker-Bergman operator [10] [1] $B_3(f, \mathscr{L}, X_0)$ which transforms functions of two complex variables $f(t, u)$, into harmonic functions of three variables.

$$H(X) = B_3(f, \mathscr{L}, X_0),\ B_3(f, \mathscr{L}, X_0) = \frac{1}{2\pi i}\int_{\mathscr{L}} f(t, u)\frac{du}{u}$$

$$t = \left[-(x - iy)\frac{u}{2} + z + (x + iy)\frac{u^{-1}}{2}\right],$$

$$|X - X_0| < \varepsilon, \qquad X \equiv (x, y, z),\ X_0 \equiv (x_0, y_0, z_0)\ ,$$

where $\mathscr{L}$ is a closed differentiable arc[1] in the u-plane, and $\varepsilon > 0$ is sufficiently small. We may see how this operator maps the functions $f(t, u)$ into harmonic functions by considering the homogeneous polynomials of degree n in x, y, z, which are defined by

$$t^n = \left\{-(x-iy)\frac{u}{2} + z + (x + iy)\frac{u^{-1}}{2}\right\}^n = \sum_{m=-n}^{+n} h_{n,m}(x, y, z)u^{-m}\ .$$

The $h_{n,m}(x, y, z)$ are linearly independent polynomials, which form a complete system [4]. Now, any harmonic function regular in a neighborhood of the origin $|X| < \varepsilon$, may be expanded into a series

$$H(X) = H(x, y, z) = \textstyle\sum_{n=0}^{\infty} \sum_{l=-n}^{+n} a_{n,l} h_{n,l}(x, y, z),$$

which converges inside the smallest sphere on whose surface there is a singularity of $H(X)$.

From the definition of the harmonic polynomials we see that

$$\frac{1}{2\pi i}\int_{\mathscr{L}} t^n u^m \frac{du}{u} = h_{n,m}(x, y, z)\ ,$$

where $\mathscr{L}$ is, say, the unit circle. In spherical coordinates this result may be recognized as one of Heine's [7] integral representations for the

Received January 19, 1960. This work has been submitted to Carnegie Institute of Technology in partial fulfillment of the requirement for the degree of Doctor of Philosophy.

[1] We shall usually consider $\mathscr{L}$ to be closed; however there is nothing preventing us from considering open arcs also.

associated Legendre functions.[2]

It follows then that if $H(X)$ is regular for $|X| < \varepsilon$ it may be generated by an integral operator

$$H(X) = \frac{1}{2\pi i}\int_{\mathscr{L}} f(t, u)\frac{du}{u} ,$$

where

$$f(t, u) = \sum_{n=0}^{\infty} \sum_{m=-n}^{+n} a_{nm} t^n u^m .$$

The harmonic functions which are regular at infinity, $|X| > 1/\varepsilon$, are of the form

$$H^{\infty}(X) = \frac{1}{r} H\left(\frac{x}{r^2}, \frac{y}{r^2}, \frac{z}{r^2}\right) ,$$

and may also be generated by the Whittaker operator; however, in this case we use the functions

$$G(t, u) = \frac{1}{t}\sum_{n=0}^{\infty} \sum_{m=-n}^{+n} a_{nm} t^{-n} u^m .$$

How the functions $G(t, u)$ transform may be seen from Heine's other representation

$$\begin{aligned}\frac{1}{2\pi i}\int_{\mathscr{L}} t^{-n}u^m \frac{du}{tu} &= h^{\infty}_{n,m}(x, y, z)\\ &= \frac{(n-m)!\,(n+m)!}{(n!)^2 2^n}\,\frac{1}{r} h_{n,m}\left(\frac{x}{r^2}, \frac{y}{r^2}, \frac{z}{r^2}\right)\\ &\frac{(n-m)!}{n!}(-i)^m r^{-n-1} P_n^m(\cos\theta) e^{im\varphi} ,\end{aligned}$$

where, as before, $\mathscr{L}$ is the unit circle.

Occasionally it is convenient to continue the arguments x, y, z to complex values in order to study the behavior of $H(X)$. For instance, if we introduce, as a particular continuation, the complex spherical coordinates

$$r = +(x^2 + y^2 + z^2)^{1/2} .$$

[2] By introducing spherical coordinates

$$\begin{aligned} x &= r \sin\theta \cos\varphi ,\\ y &= r \sin\theta \sin\varphi ,\\ z &= r\cos\theta ,\end{aligned}$$

the polynomials may be written in the form $h_{n,m}(x, y, z) = (n!/(n+m!)) r^n P_n^m(\cos\theta) e^{im\varphi}$

Integrals of terms $t^n u^m$, where $|m| > |n| > 0$, vanish; consequently, we may restrict ourselves to just those functions where $|m| \leqq |n|$.

$$\zeta = +\left(\frac{x+iy}{x-iy}\right)^{1/2},$$

$$\xi = \frac{z}{r},$$

which reduce to $\zeta = e^{i\varphi}$, $\xi = \cos\theta$, for real x, y, z, we may obtain an inverse Whittaker operator.

LEMMA. *Let $V(r, \cos\theta, e^{i\varphi})$, be a harmonic function regular at infinity*; i.e.

$$V(r, \cos\theta, e^{i\phi}) \equiv H^{\infty}(X) = \frac{1}{2\pi i}\int_{\mathscr{L}} G(t, u)\frac{du}{u},$$

where

$$G(t, u) = \left(\sum_{n=0}^{\infty}\sum_{m=-n}^{+n} a_{nm}t^{-n-1}u^m\right),$$

and $\mathscr{L}$ is the unit circle.

Then $G(s, u)$ may be generated by the integral transform

$$G(s, u) = \frac{1}{4\pi i}\int_{-1}^{+1}\left[\int_{\Gamma}\frac{r(s+t)}{(s-t)^2}V(\gamma, \xi, \zeta)\frac{d\zeta}{\zeta}\right]d\xi.$$

The integration path in the ξ-plane is the linear segment $-1 \leqq \xi \leqq 1$, the path in ζ-plane is the unit circle.

Proof. Let us define

$$\frac{1}{r}K\left(\frac{r}{t}, \xi, \frac{iu}{\zeta}\right) \equiv \sum_{n=0}^{\infty}\sum_{m=-n}^{+n}(2n+1)\frac{n!}{(n+m)!}\left(\frac{r}{t}\right)^n P_n^m(\xi)\left(\frac{iu}{\zeta}\right)^m;$$

it follows then, directly from the orthogonality relation

$$\int_{-1}^{+1}P_n^m(\xi)P_s^m(\xi)d\xi = \delta_{ns}\frac{2}{2n+1}\frac{(n+m)!}{(n-m)!},$$

that

$$s^{-n}u^m = \frac{1}{4\pi i}\int_{-1}^{+1}\left[\int_{\Gamma}K\left(\frac{r}{s}, \xi, \frac{iu}{\zeta}\right)\left(\frac{\zeta}{i}\right)^m\frac{d\zeta}{\zeta}\right]\frac{(n-m)!}{n!}r^{-n-1}P_n^m(\xi)d\xi$$

(where the integration paths are those mentioned in the hypothesis). Recalling the generating function for the spherical harmonics

$$\sum_{n=0}^{\infty}t^n = \sum_{n=0}^{\infty}\sum_{m=-n}^{+n}\frac{n!}{(n+m)!}r^nP_n^m(\xi)\left(\frac{\zeta}{iu}\right)^m$$

we see that K may be formally summed to

$$K\left(\frac{r}{s}, \xi, \frac{iu}{\zeta}\right) = \left(1 - 2s\frac{\partial}{\partial s}\right) r \sum_{n=0}^{\infty} \left(\frac{t}{s}\right)^n = rs\frac{s+t}{(s-t)^2} ,$$

providing $(|t/s|) < 1$. In this case, K is an analytic function of t, and hence also analytic in r, ξ, and ζ. The harmonic functions $H^{\infty}(x, y, z)$, which are regular at infinity, have a Taylor series expansion of the form $\sum_{j,k,l=0}^{\infty} A_{jkl} x^{-j} y^{-k} z^{-l}$.

If this series converges for $x^2 + y^2 + z^2 > (1/\varepsilon^2)$, then the series

$$\sum_{j,k,l=0}^{\infty} A_{jkl}(x_1 + ix_2)^{-j}(y_1 + iy_2)^{-k}(z_1 + iz_2)^{-l} ,$$

if rewritten in the form

$$\sum_{\substack{a,b,c \\ r,s,t}} B_{abcrst} x_1^{-a} x_2^{-r} y_1^{-b} y_2^{-s} z_1^{-c} z_2^{-t} ,$$

will converge for $x_1^2 + y_1^2 + z_1^2 > (2/\varepsilon^2)$, and $x_2^2 + y_2^2 + z_2^2 > (2/\varepsilon^2)$. Hence, $H^{\infty}(x, y, z)$ is an analytic function of the complex variable x, y, z, in some neighborhood of infinity. The harmonic function $V(r, \xi, \zeta)$ obtained by replacing x, y, z in $H^{\infty}(x, y, z)$ by

$$x = \frac{r}{2}(\zeta + \zeta^{-1})\sqrt{1 - \xi^2} ,$$

$$y = \frac{r}{2i}(\zeta - \zeta^{-1})\sqrt{1 - \xi^2} ,$$

$$z = r\xi ,$$

consequently is an analytic function of r, ξ, ζ, except of course at $\xi = \pm 1$, and $\zeta = 0$.

It may be concluded, therefore, that the integrals involved in our representation for $G(s, u)$ are Cauchy-integrals, since the integrand is a single-valued analytic function of ξ and ζ.

II. Singularities of harmonic functions generated by the Whittaker-Bergman operator. Bergman [2] has considered a special class of harmonic functions generated by the Whittaker operator and has given a simple procedure for finding their singularities. He does this as follows:

Suppose that $(1/u)f(t, u)$ has the form $P(t, u)/Q(t, u)$, where P and Q are polynomials in t and u. In order to study the harmonic function

$$H(X) = B_3(f, \mathscr{L}, X_0) = \frac{1}{2\pi i}\int_{\mathscr{L}} f(t, u)\frac{du}{u} ,$$

we consider the singularity manifold of P/Q, i.e.

$$Z^3 = E\left\{Q\left[-(x-iy)\frac{u}{2} + z + (x+iy)(2u)^{-1}, u\right] = 0\right\}.$$

The manifold Z^3 may also be written in the form

$$Z^3 = E\{u = \phi_\nu(X), \nu = 1, 2, 3, \cdots n\},$$

where the $\phi_\nu(X)$ are algebraic functions of x, y, z, and the degree of u in Q is n. At every point (x, y, z), except those which satisfy the equation

$$\prod_{K \neq s} [\phi_K(X) - \phi_s(X)] = 0,$$

there are n distinct branches of $Z^3_\nu = E\{u = \phi_\nu(X), \nu = 1, 2, 3, \cdots, n\}$, of Z^3. We choose the contours $\mathscr{L}_\nu$, $v = 1, 2, 3, \cdots, n$, so that one and only one $u = \phi_\nu(X)$ lies inside $\mathscr{L}_\nu$. It follows from the residue theorem that

$$H_\nu(X) = \frac{1}{2\pi i}\int_{\mathscr{L}_\nu} \frac{P(t, u)}{Q(t, u)} du,$$

where $H_\nu(X)$ is the corresponding branch of

$$H(X) = \frac{P[-(x-iy)u/2 + z + (x-iy)(2u)^{-1}, u]}{\partial\{Q[-(x-iy)u/2 + z + (x-iy)(2u)^{-1}, u]\}/\partial u},$$

with

$$Q\left[-(x-iy)\frac{u}{2} + z + (x-iy)(2u)^{-1}, u\right] = 0.$$

We notice that $H(X)$ becomes singular for those values of (x, y, z) which satisfy the equations

$$Q\left[-(x-iy)\frac{u}{2} + z + (x-iy)(2u)^{-1}, u\right] = 0,$$

$$\partial\left\{Q\left[-(x-iy)\frac{u}{2} + z + (x-iy)(2u)^{-1}, u\right]\right\}/\partial u = 0.$$

We shall now show that Bergman's result does not depend on the fact that $(1/u)f(t, u)$ is an algebraic function, but holds under more general conditions. The only restriction we will impose is that the singularities of $(1/u)f(t, u)$ can be written in the implicit form $S(x, y, z, u) = 0$.

THEOREM 1. *If $Z^3 = E\{S(x, y, z, u) = 0\}$ is an implicit representation of the singularities of*

$$\frac{1}{u}f(t, u), \textit{ then } H(X) = \frac{1}{2\pi i}\int_{\mathscr{L}} f(t, u)\frac{du}{u} ,$$

(where $\mathscr{L}$ is the unit circle) *is regular at* $X = (x, y, z)$, *providing this point does not lie simultaneously on the two surfaces*

$$S(x, y, z, u) = 0 ,$$

and

$$\frac{\partial}{\partial u}S(x, y, z, u) = 0 .$$

Proof. The proof of Theorem 1 will be based on a modified form of an idea employed by Hadamard in the proof of his theorem on the multiplication of singularities [8] [5]. The integral representation of $H(X)$ is valid for all points (x, y, z) which can be reached from an initial point by continuation along a curve $\Gamma(X)$ (in three dimensional real-space, R^3), provided no point of $\Gamma(X)$ corresponds to a singularity of $(1/u)f(t, u)$ on the integration path. This initial domain of definition of $H(X)$ can now be enlarged by continuously deforming the integration path provided, again, that in this process of deformation the integration path at no time crosses a singularity of $(1/u)f(t, u)$. Accordingly, we may now write $H(X)$ as

$$H(X) = \frac{1}{2\pi i}\int_{\mathscr{L}'} f(t, u)\frac{du}{u} ,$$

where $\mathscr{L}'$ is now a new integration path obtained by observing the above precautions.

Since t is dependent on $X = (x, y, z)$, the singularities of the integral move in the u-plane as we continue $H(X)$ along $\Gamma(X)$. Now, as long as we can avoid crossing such a singularity by deforming the contour $\mathscr{L}'$ we are still able to continue $H(X)$. Let us assume we have been able to continue $H(X)$ to the point $X_1 \equiv (x_1, y_1, z_1)$, and let us consider the singularities of the integral for $X = X_1$. The singularities of $(1/u)f(t, u)$ are those values of u satisfying $S(x_1, y_1, z_1, u) = 0$. From Taylor's theorem we may describe the local properties of S about some point $u = \alpha$, for which $S = 0$, by

$$S(x_1, y_1, z_1, u) = (u - \alpha)\frac{\partial}{\partial u}S(x_1, y_1, z_1, \alpha) + \frac{(u-\alpha)^2}{2!}\frac{\partial^2 S(x_1, y_1, z_1, \alpha)}{\partial u^2} \cdots .$$

Unless $\partial S/\partial u = 0$ at $u = \alpha$, in a neighborhood of $u = \alpha$ we may approximate S by

$$S(x_1, y_1, z_1, u) \cong (u - \alpha)\frac{\partial S(x_1, y_1, z_1, \alpha)}{\partial u} .$$

Therefore in some neighborhood of $u = \alpha$, say $|u - \alpha| < \varepsilon$, S does not vanish save at $u = \alpha$. Clearly, then, by deforming $\mathscr{L}'$ we can avoid crossing $u = \alpha$, or any other point $u = \beta$ for which $S(x_1, y_1, z_1, \beta) = 0$, if we follow an arc of the circle $|u - \alpha| = \varepsilon/2$ about $u = \alpha$. This completes our proof.

Using the language of real geometry we may say that unless we are in the neighborhood of the envelope $\mathscr{E}(x, y, z) = 0$ to $S(x, y, z, u) = 0$ (in which case there are an infinite number of such surfaces tangent to $\mathscr{E}(x, y, z) = 0$ we may avoid crossing these singularities by deforming $\mathscr{L}'$.

THEOREM 2. *Let $r = \Phi(\xi, \zeta)$ be a representation of the singularities of $V(r, \xi, \zeta) \equiv H^{\infty}(X)$, $X \in C^3$. The function of two complex variables*

$$G(s, u) = \frac{1}{4\pi i}\int_{-1}^{+1}\left[\oint_{\zeta}\frac{s + t}{(s - t)^2} r V(r, \xi, \zeta)\frac{d\zeta}{\zeta}\right]d\xi ,$$

is then regular at (s, u) providing (s, u) does not lie on the "envelope" of the two parameter family

$$\psi(s, u \,|\, \xi, \zeta) \equiv \Phi(\xi, \zeta)\left[\xi + \frac{i}{2}\sqrt{1 - \xi^2}\left(\frac{u}{\zeta} + \frac{\zeta}{u}\right)\right] - s = 0 .$$

Proof. The proof of this theorem closely parallels the one for Theorem 1. As before, we consider the analytic continuation of $G(s, u)$ along an arc $\Gamma^4(s^{-1}, u)$, beginning at $s^{-1} = 0$, $u = 1$. The integral representation of

$$G(s, u) = \frac{1}{4\pi i}\int_{-1}^{+1}\left[\oint_{\zeta}\frac{s + t}{(s - t)^2} r V(r, \xi, \zeta)\frac{d\zeta}{\zeta}\right]d\xi$$

will remain the same if either integration path (in ξ or ζ planes) is continuously deformed in such a manner so that at no time they cross a singularity of the integrand. Therefore, we may write $G(s, u)$ as

$$G(s, u) = \frac{1}{4\pi i}\int_{-1\mathscr{L}_{\xi}}^{+1}\left[\int_{\mathscr{L}_{\zeta}}\frac{s + t}{(s - t)^2} r V(r, \xi, \zeta)\frac{d\zeta}{\zeta}\right]d\xi$$

where $\mathscr{L}_{\xi}$ and $\mathscr{L}_{\zeta}$ are new integration paths obtained by observing the above precautions. Now, the kernel in our integral representation is singular whenever

$$t - s = r\left[\xi + \frac{i}{2}\sqrt{1 - \xi^2}\left(\frac{u}{\zeta} + \frac{\zeta}{u}\right)\right] - s = 0 ,$$

and the harmonic function is singular for $\Phi(\xi, \zeta) - r = 0$. We notice a significant difference in these two singularity manifolds; as $G(s, u)$ is continued along $\Gamma^4(s^{-1}, u)$ the singularities of the kernel move in the

ξ, ζ-planes, while those of the harmonic function remain fixed. By using the Hadamard idea we realize that we may always avoid an advancing singularity by deforming one of our contours with the possible exception occuring when the two manifolds coincide. Therefore, unless $r = \Phi(\xi, \zeta)$ as a function of ξ, and ζ also satisfies $t - s = 0, G(s, u)$ must be regular. This leads us to consider the two parameter family,

$$\psi(s, u \mid \xi, \zeta) \equiv \Phi(\xi, \zeta)\left[\xi + \frac{i}{2}\sqrt{1-\xi^2}\left(\frac{u}{\zeta} + \frac{\zeta}{u}\right)\right] - s = 0 ,$$

as the only possible singularities of $G(s, u)$.

Let us assume that we have been able to continue $G(s, u)$ to (s_0, u_0) and let us consider those values of ξ, ζ satisfying $\psi(s_0, u_0 \mid \xi, \zeta) = 0$. These values are singularities of the integrand which must be investigated to determine whether they are avoidable by deforming the paths of integration. Let $\xi = \alpha$, and $\zeta = \beta$ be singularities which may cross either $\mathscr{L}_\xi$ or $\mathscr{L}_\zeta$ respectively if $G(s, u)$ is continued further along $\Gamma^4(s^{-1}, u)$. In a bicylindrical neighborhood $|\xi - \alpha| < \varepsilon_1, |\zeta - \beta| < \varepsilon_2$, we may expand $\psi(s_0, u_0 \mid \xi, \zeta)$ in a double Taylor series as

$$\psi(s_0, u_0 \mid \xi, \zeta) = (\xi - \alpha)\frac{\partial}{\partial \xi}\psi(s_0, u_0 \mid \alpha, \beta) + (\zeta - \beta)\frac{\partial}{\partial \zeta}\psi(s_0, u_0 \mid \alpha, \beta)$$
$$+ \frac{1}{2}\left[(\xi - \alpha)^2\frac{\partial^2\psi}{\partial \xi^2} + 2(\xi - \alpha)(\zeta - \beta)\frac{\partial^2\psi}{\partial\xi\partial\zeta} + (\zeta - \beta)^2\frac{\partial^2\psi}{\partial\zeta^2}\right] + \cdots .$$

Now, unless the first variation of $\psi(s_0, u_0 \mid \xi, \zeta)$ vanishes at (α, β), ψ may be approximated as

$$\psi(s_0, u_0 \mid \xi, \zeta) \cong (\xi - \alpha)\frac{\partial}{\partial \xi}\psi(s_0, u_0 \mid \alpha, \beta) + (\zeta - \beta)\frac{\partial}{\partial \zeta}\psi(s_0, u_0 \mid \alpha, \beta) .$$

In this case it is always possible to choose a secant to the circle $|\xi - \alpha| = \varepsilon_1/2$ not passing through $\xi = \alpha$, and a secant to the circle $|\zeta - \beta| = \varepsilon_2/2$ not passing through $\zeta = \beta$, such that $\psi(s_0, u_0 \mid \xi, \zeta) \neq 0$ on those portions of the secants inside the respective circles. It follows that, in this case, we may deform the paths $\mathscr{L}_\xi$, and $\mathscr{L}_\zeta$ so that they follow the secants about the singular point (α, β) and thereby continue $G(s, u)$ still further. The only possible singularities of $G(s, u)$ are therefore those values of s and u satisfying simultaneously

$$\psi(s, u \mid \xi, \zeta) = 0$$

and

$$\frac{\partial\psi}{\partial\xi} + \frac{\partial\psi}{\partial\zeta}\pi'(\xi) = 0 ,$$

where $\zeta = \pi(\xi)$ is an arbitrary relationship between ξ and ζ. This completes our proof.

We notice here, that a particular class of singularities of $G(s, u)$ may occur for s and u satisfying simultaneously

$$\psi(s, u \mid \xi, \zeta) = 0 ,$$

$$\frac{\partial \psi}{\partial \xi} = 0 ,$$

and

$$\frac{\partial \psi}{\partial \zeta} = 0 .$$

We have reduced the problem of locating the singularities of $G(s,u)$ to obtaining the envelope of a three parameter family of complex surfaces

$$\psi(s, u \mid r, \xi, \zeta) = 0 ,$$

where the parameters r, ξ, ζ are subject to the condition

$$A(r, \xi, \zeta) = 0 .$$

It was most natural, because of the Cauchy integrals involved, to consider ξ and ζ as independent parameters, and r the dependent parameter. However, unless we are in the neighborhood of a "singular point" of $A = 0$, it is no longer necessary to make this distinction.

For a point (s, u) to lie on the envelope $E(s, u) = 0$, the first variation,

$$\delta\psi = \frac{\partial \psi}{\partial r}\delta r + \frac{\partial \psi}{\partial \xi}\delta\xi + \frac{\partial \psi}{\partial \zeta}\delta\zeta ,$$

must vanish. If we proceed as before, and consider r dependent, we obtain

$$\left(\frac{\partial A}{\partial r}\frac{\partial \psi}{\partial \xi} - \frac{\partial A}{\partial \xi}\frac{\partial \psi}{\partial r}\right)\delta\xi + \left(\frac{\partial A}{\partial r}\frac{\partial \psi}{\partial \zeta} - \frac{\partial A}{\partial r}\frac{\partial \psi}{\partial r}\right)\delta\zeta = 0 ,$$

which implies that an arbitrary functional relationship exists between ξ and ζ, or more generally a relationship $B(r, \xi, \zeta) = 0$, such that

$$\left(\frac{\partial B}{\partial r}\frac{\partial \psi}{\partial \xi} - \frac{\partial B}{\partial \xi}\frac{\partial \psi}{\partial r}\right)\delta\xi + \left(\frac{\partial B}{\partial r}\frac{\partial \psi}{\partial \zeta} - \frac{\partial B}{\partial \zeta}\frac{\partial \psi}{\partial r}\right)\delta\zeta = 0 ,$$

where

$$\begin{vmatrix} \dfrac{\partial A}{\partial r}\dfrac{\partial \psi}{\partial \xi} - \dfrac{\partial A}{\partial \xi}\dfrac{\partial \psi}{\partial r} & \dfrac{\partial A}{\partial r}\dfrac{\partial \psi}{\partial \zeta} - \dfrac{\partial A}{\partial \zeta}\dfrac{\partial \psi}{\partial r} \\ \dfrac{\partial B}{\partial r}\dfrac{\partial \psi}{\partial \xi} - \dfrac{\partial B}{\partial \xi}\dfrac{\partial \psi}{\partial r} & \dfrac{\partial B}{\partial r}\dfrac{\partial \psi}{\partial \zeta} - \dfrac{\partial B}{\partial \zeta}\dfrac{\partial \psi}{\partial r} \end{vmatrix} = 0 \,.$$

Let us consider the envelope of $\psi(s, u \mid r, \xi, \zeta) = 0$ [subject to $A(r, \xi, \zeta) = 0$] under the transformation of parameters

$$\begin{aligned} r &= +\,(x^2 + y^2 + z^2)^{1/2} \\ \xi &= z/ +\,(x^2 + y^2 + z^2)^{1/2} \,, \\ \zeta &= +\left(\frac{x + iy}{x - iy}\right)^{1/2} . \end{aligned}$$

We realize that, for $X = (x, y, z) \supset R^3$, the Jacobian cannot vanish and hence the transformation is one-to-one. However, as may be confirmed by direct computation

$$\frac{\partial(r, \xi, \zeta)}{\partial(x, y, z)} \neq 0, \text{ for all } X \in C^3 \,,$$

which are a finite distance from the origin.

Under this transformation our family of complex surfaces becomes

$$\{\psi(s, u \mid r, \xi, \zeta) = 0\} \rightarrow \{\chi(s, u \mid x, y, z) = 0\} \,,$$

with the auxiliary condition

$$\{A(r, \xi, \zeta) = 0\} \rightarrow \{P(x, y, z) = 0\} \,.$$

Now, for a point (s, u) to lie on the envelope to $\chi = 0$, the first variation must vanish, i.e.

$$\begin{aligned} \delta\chi &= \frac{\partial \chi}{\partial x}\delta x + \frac{\partial \chi}{\partial y}\delta y + \frac{\partial \chi}{\partial z}\delta z = 0 \\ &= \left(\frac{\partial \psi}{\partial r}\frac{\partial r}{\partial x} + \frac{\partial \psi}{\partial \xi}\frac{\partial \xi}{\partial x} + \frac{\partial \psi}{\partial \zeta}\frac{\partial \zeta}{\partial x}\right)\delta x \\ &\quad + \left(\frac{\partial \psi}{\partial r}\frac{\partial r}{\partial y} + \frac{\partial \psi}{\partial \xi}\frac{\partial \xi}{\partial y} + \frac{\partial \psi}{\partial \zeta}\frac{\partial \zeta}{\partial y}\right)\delta y \\ &\quad + \left(\frac{\partial \psi}{\partial r}\frac{\partial r}{\partial z} + \frac{\partial \psi}{\partial \xi}\frac{\partial \xi}{\partial z} + \frac{\partial \psi}{\partial \zeta}\frac{\partial \zeta}{\partial z}\right)\delta z \\ &= \frac{\partial \psi}{\partial r}\left(\frac{\partial r}{\partial x}\delta x + \frac{\partial r}{\partial y}\delta y + \frac{\partial r}{\partial z}\delta z\right) \\ &\quad + \frac{\partial \psi}{\partial \xi}\left(\frac{\partial \xi}{\partial x}\delta x + \frac{\partial \xi}{\partial y}\delta y + \frac{\partial \xi}{\partial z}\delta z\right) \\ &\quad + \frac{\partial \psi}{\partial \zeta}\left(\frac{\partial \zeta}{\partial y}\delta x + \frac{\partial \zeta}{\partial y}\delta y + \frac{\partial \zeta}{\partial z}\delta z\right) = 0 \,. \end{aligned}$$

From our auxiliary condition we have

$$\left[\frac{\partial r}{\partial x}\delta x + \frac{\partial r}{\partial y}\delta y + \frac{\partial r}{\partial z}\delta z\right]$$
$$= -\frac{\partial A}{\partial \xi}\left[\frac{\partial \xi}{\partial x}\delta x + \frac{\partial \xi}{\partial y}\delta y + \frac{\partial \xi}{\partial z}\delta z\right]\Big/\frac{\partial A}{\partial r}$$
$$-\frac{\partial A}{\partial \zeta}\left[\frac{\partial \zeta}{\partial x}\delta x + \frac{\partial \zeta}{\partial y}\delta y + \frac{\partial \zeta}{\partial z}\delta z\right]\Big/\frac{\partial A}{\partial r},$$

which, together with $\delta\chi = 0$, yields

$$\left[\frac{\partial A}{\partial r}\frac{\partial \psi}{\partial \xi} - \frac{\partial A}{\partial \xi}\frac{\partial \psi}{\partial r}\right][\delta\xi(x, y, z)] + \left[\frac{\partial A}{\partial r}\frac{\partial \psi}{\partial \zeta} - \frac{\partial A}{\partial \zeta}\frac{\partial \psi}{\partial r}\right][\delta\zeta(x, y, z)] = 0 .$$

We conclude that under a one-to-one, continuous transformation of parameters the envelope is invariant and we have the following corollary to Theorem 2. Let $F(x, y, z) = 0$ be a representation of the singularities of $H^\infty(X)$, $X \in C^3$. Then the function $G(s, u)$, which generates $H^\infty(X)$ under the Whittaker operator, can only have singularities on the envelope, $E(s, u) = 0$, to the family of complex surfaces

$$\chi(s, u\,|\,x, y, z) \equiv \left[-(x - iy)\frac{u}{2} + z + (x + iy)\frac{1}{2u}\right] - s = 0 ,$$

where the parameters (x, y, z) are subject to the auxiliary condition $F(x, y, z) = 0$.

To illustrate the use of Theorem 1, we consider the case where $(1/u)f(t, u)$ has the particular form

$$\frac{1}{u}f(t, u) = F\left[t^{-1}\left(u - \frac{1}{u}\right)\right];$$

$F(x)$ is an arbitrary function of x singular at $x = \beta$. This choice of $(1/u)f(t, u)$ generates an $H(X)$ having a simple type of singularity. Since the singularities of $(1/u)f(t, u)$ satisfy $u - (1/u) = t\beta$, we represent the singularity manifold as

$$S(x, y, z, u) = -u[\beta(x - iy) + 2] + 2\beta z + \frac{1}{u}[\beta(x + iy) + 2] .$$

Eliminating u between $S = 0$, and $\partial S/\partial u = 0$, we obtain the locus $(x + 2/\beta)^2 + y^2 + z^2 = 0$, for the singularities of $H(X)$.

When β is real this reduces to a point singularity in R^3. However, if β is complex the singularities in R^3 are given by

$$x = -\frac{2}{|\beta|^2}\Re e\beta ,$$

$$y^2 + z^2 = \frac{4}{|\beta|^4}(\Im\mathfrak{m}\beta)^2 .$$

We note that these are only the possible singularities of $H(X)$. To find the actual singularities we make use of our inverse Whittaker operator to find which of the possible singularities of $H(X)$ correspond to singularities of $(1/u)f(t, u)$.

Let us consider the locus of

$$\left(x + \frac{2}{\beta}\right)^2 + y^2 + z^2 = 0 \text{ in } R^3, \text{ that is}$$

$$x = -\frac{2}{|\beta|^2}\Re\mathfrak{e}\beta ,$$

and

$$y^2 + z^2 = \frac{4}{|\beta|^4}(\Im\mathfrak{m}\beta)^2 .$$

If we wish to find which singularities of $(1/u)f(t, u)$ correspond to this real locus, we eliminate two parameters from χ and consider the first variation with respect to the remaining parameter. Doing this,

$$\chi = -\frac{x}{2}\left(u - \frac{1}{u}\right) + \frac{iy}{2}\left(u + \frac{1}{u}\right) + z - s = 0, \text{ becomes}$$

$$\chi = 2\frac{\Re\mathfrak{e}\beta}{|\beta|^2}\left(u - \frac{1}{u}\right) \pm \frac{i}{2}\sqrt{\frac{4}{|\beta|^4}(\Im\mathfrak{m}\bar{\beta})^2 - z^2}\left(u + \frac{1}{u}\right) + z - s = 0 .$$

The first variation is then

$$\frac{\partial\chi}{\partial z} = \frac{\pm\frac{i}{2}(-z)\left(u + \frac{1}{u}\right)}{\sqrt{\frac{4}{|\beta|^2}(\Im\mathfrak{m}\bar{\beta})^2 - z^2}} + 1$$

Eliminating z, between χ and $\partial\chi/\partial z$ yields

$$\pm\mp i(\Im\mathfrak{m}\beta)\left(u + \frac{1}{u}\right)^2 + (\Re\mathfrak{e}\beta)\left(u - \frac{1}{u}\right) \pm 4i(\Im\mathfrak{m}\beta)$$
$$= s|\beta|^2\left(u - \frac{1}{u}\right) .$$

By choosing suitable signs this is recognized readily as

$$\left(u - \frac{1}{u}\right) = \beta s .$$

REMARK. In concluding we note, that as in the case of harmonic functions regular at the origin, a connection will exist between the coefficients of the series development for $f(t, u)$ and the singularities of $H(X)$[3]. Hence, it would be of interest to investigate whether a relation exists between singularities as predicted by Theorem 1 of this paper, and the corresponding coefficients of the series development for $f(t, u)$. Such an investigation should lead to a classification of harmonic functions in terms of their pole-like singularities in three-dimensional complex space.

REFERENCES

1. S. Bergman, *Zur Theories der algebraischen Potential funktionen des drei-dimensional Raumes*, Math. Ann., **99** (1928), 629-659.
2. ———, *Essential singularities of solutions of a class of linear partial differential equations in three variables*, J. Rational Mech. and Anal, **3** (1954), 539-560.
3. ———, *Integral operators in the theory of linear partial differential equations*, Ergebnisse der Mathematik und ihrer Grenzgebiete, New Series, **23**, (1960).
4. R. Courant and D. Hilbert, *Methods of mathematical physics*, **1**, (1953).
5. J. Hadamard, *Théorème sur les séries entières*, Acta Math., **22** (1898), 55-64.
6. E. Kreyszig, *On regular and singular harmonic functions of three variables*, Archive Rational Mech. and Anal., to appear.
7. P. M. Morse and H. Feshbach, *Methods of theoretical physics*, part II, 1953.
8. Z. Nehari, *On the singularities of Legendre expansions*, J. Rational Mech. and Anal., **5** (1956), 987-992.
9. G. Szegö, *On the singularities of zonal harmonic expansions*, J. Rational Mech. and Anal., **3** (1954).
10. E. T. Whittaker and G. N. Watson, *A course of modern analysis*, 1920.

MICHIGAN STATE UNIVERSITY

3 References to Integral Operators in the Theory of Linear Partial Differential Equations, see Bergman [**3**] and Kreyszig [**6**].

PARTITIONS OF MASS-DISTRIBUTIONS AND OF CONVEX BODIES BY HYPERPLANES

B. GRÜNBAUM

1. Introduction. The following results are well-known (Neumann [7]; Eggleston [3], [4, p. 125–126], [5, p. 118]; Newman [8]:

(A) For any mass-distribution in the plane, such that the total mass contained in every half-plane is finite and depends continuously on the position of the half-plane, there exists a point P such that each half-plane which contains P, contains at least 1/3 of the total mass.

(B) For any convex body K in the plane there exists a point P such that for each half-plane H containing P the area of $H \cap K$ is at least 4/9 of the area of K.

The main object of the present note is to generalize (A) and (B) to higher-dimensional Euclidean spaces.

In the following m shall denote a fixed (non-negative) finite measure on the ring of subsets of E^n generated by the closed half-spaces in E^n. (For the terminology and results on measures see, e.g., Halmos [6].)

For a real λ, $0 \leqq \lambda \leqq 1/2$, we define $\mathscr{C}(m, \lambda)$ as the subset of E^n consisting of those points $P \in E^n$ which satisfy the condition: For any closed half-space $H \subset E^n$, with $P \in H$, the relation $m(H) \geqq \lambda \cdot m(E^n)$ holds.

Obviously, $\mathscr{C}(m, \lambda)$ is a compact, convex (possibly empty) set.

Using the notation of $\mathscr{C}(m, \lambda)$, Theorem (A) may be extended as follows:

THEOREM 1. *$\mathscr{C}(m, 1/(n+1)) \neq \phi$ for any measure m in E^n.*

Let $V(S)$ denote the volume (n-dimensional Lebesgue measure) of the set $S \subset E^n$. For any convex body $K \subset E^n$, we denote by m_K the measure (defined for all Lebesgue measurable subsets S of E^n) obtained by taking $m_K(S) = V(S \cap K)$. We denote $\mathscr{C}(m_K, \lambda)$ by $\mathscr{C}(K, \lambda)$.

Theorem (B) may now be generalized as follows:

THEOREM 2. *If K is any convex body in E^n then*

$$\mathscr{C}\left(K, \left(\frac{n}{n+1}\right)^n\right) \neq \phi .$$

We shall prove Theorems 1 and 2 in the following two sections.

Received January 22, 1960. This research was supported by the United States Air Force through the Air Force Office of Scientific Research of the Air Research and Development Command, under contract No. AF49(638)-253. Reproduction in whole or in part is permitted for any purpose of the United States Government.

The last section contains remarks and comments.

2. Proof of Theorem 1.[1] If v is a unit vector (in E^n) and α is a real number, let $H(v, \alpha)$ be the closed half-space

$$H(v, \alpha) = \{x \in E^n; (x, v) \leqq \alpha\}.$$

Let $\alpha(v)$ be defined by

$$\alpha(v) = \min \left\{\alpha; m(H(v, \alpha)) \geqq \frac{n}{n+1} m(E^n)\right\},$$

(the minimum is attained since $m(H(v, \alpha))$ is continuous to the right as a function of α). Let $H(v) = H(v, \alpha(v))$ and

$$H^*(v) = \{x \in E^n; (x, v) \geqq \alpha(v)\}.$$

(Without loss of generality we shall in the sequel assume $m(E^n) = 1$.) Obviously,

$$\mathscr{C}\left(m \frac{1}{(n+1)}\right) \supset \bigcap_v H(v);$$

hence, if $\bigcap_v H(v) \neq \phi$ the proof is complete. On the other hand, if $\bigcap_v H(v) = \phi$, we shall show that

$$C\left(m \frac{1}{(n+1)}\right) \neq \phi$$

in the following way. The half-spaces $H(v)$ are closed convex sets, and it is easily seen that a finite number of them may be found such that their intersection is compact. By Helly's theorem on intersections of convex sets (see, e.g., Rademacher-Schoenberg [9]) the assumption $\bigcap_v H(v) = \phi$ implies the existence of an $n+1$ membered family of unit vectors v_i, $0 \leqq i \leqq n$, such that $\bigcap_{i=0}^n H(v_i) = \phi$. Using an inductive argument it is easily seen that we may assume that every n of the vectors v_i are linearly independent. Therefore (denoting $H_i = H(v_i)$ and $H_i^* = H_i^*(v_i)$) the set $S = \bigcap_{i=0}^n H_i^*$ is a non-degenerate simplex whose faces are contained in the hyperplanes $H_i \cap H_i^*$, $0 \leqq i \leqq n$. By the definition of $\alpha(v)$ we have $m(H_i^*) \geqq 1/(n+1)$ and $m(\text{Int } H_i^*) \leqq 1/(n+1)$ for all i. Therefore $m(H_j \cap \text{Int } H_i^*) \leqq 1/(n+1)$, and thus $m(H_j \cap H_i) \geqq (n-1)/(n+1)$ for all $i \neq j$. Now, since $\bigcap_{i=0}^n H_i = \phi$, we have

$$\frac{n}{n+1} \geqq m(H_i) \geqq m\left[H_i \cap \left(\bigcup_{j \neq i} H_j\right)\right] \geqq \frac{1}{n-1} \sum_{\substack{0 \leqq j \leqq n \\ j \neq i}} m(H_i \cap H_j)$$
$$\geqq \frac{1}{n-1} \cdot n \cdot \frac{n-1}{n+1} = \frac{n}{n+1}.$$

[1] The author is indebted to Professor B. M. Stewart for the correction of an error in the original proof.

Thus, for all i, equality signs hold throughout. In particular,

$$m\left(\bigcap_{\substack{0\leq j\leq n\\ j\neq i}} H_j\right) = \frac{1}{n+1}$$

for all i (i.e., the support of m is contained in the "vertex-regions" of the simplex $S = \bigcap_i H_i^*$), and it is immediately verified that

$$\mathscr{C}\left(m; \frac{1}{(n+1)}\right) \supset S \neq \phi .$$

This ends the proof of Theorem 1.

3. **Proof of Theorem 2.** Let G_k denote the centroid of the convex body $K \subset E^n$. We shall prove Theorem 2 by establishing the stronger statement $G_K \in \mathscr{C}(K, \alpha_n)$, where $\alpha_n = (n/(n+1))^n$. Assuming, to the contrary, that $G_k \notin \mathscr{C}(K, \alpha_n)$, there exists a hyperplane L containing G_K such that the volume of the part of K contained in one of the half-spaces determined by L is less than $\alpha_n \cdot V(K)$. We shall obtain a contradiction from this assumption.

Let G_K be the origin of an orthogonal system of coordinates $(x_1, \cdots, x_n)$ of E^n, such that L is the hyperplane determined by $x_1 = 0$.

Let H^+ be the half-space $\{(x_1, \cdots, x_n); x_1 \geqq 0\}$ and H^- the other closed half-space determined by L. We may assume that $V(K \cap H^-) < \alpha_n \cdot V(K)$. For any set $S \subset E^n$ we shall use the notations $S^- = S \cap H^-$ and $S^+ = S \cap H^+$. Let $\hat{K}$ be the set obtained from K by spherical symmetrization ("Schwarzsche Abrundung", Bonnesen-Fenchel [1, p. 71]; "Schwarz rotation process", Eggleston [5, p. 100]) with respect to the x_1-axis (i.e., $\hat{K}$ is the union of the $(n-1)$-dimensional spheres obtained by taking in each hyperplane $L_t = \{(x_1, \cdots, x_n); x_1 = t\}$ an $(n-1)$-dimensional sphere with center $(t, 0, \cdots, 0)$ and $(n-1)$-dimensional volume equal to that of $K \cap L_t$). It is well known that $\hat{K}$ is a convex body, and obviously $V(\hat{K}^-) = V(K^-)$, $V(\hat{K}^+) = V(K^+)$ and $G_{\hat{K}} = G_K$. Therefore $V(\hat{K}^-) < \alpha_n \cdot V(\hat{K})$ and $G_{\hat{K}} \notin \mathscr{C}(\hat{K}, \alpha_n)$. Let C^- denote the (orthogonal) hypercone with base $\hat{K} \cap L$ and vertex $(c, 0, \cdots, 0) \in H^-$, where c is chosen in such a way that $V(C^-) = V(\hat{K}^-)$. Let C be the hypercone obtained by extending C^- (along its generators) into H^+ in such a way that $V(C^+) = V(\hat{K}^+)$. With C thus defined, it is easily verified that the x_1-coordinate of G_{C^-} (resp. G_{C^+}) is not greater than that of $G_{\hat{K}^-}$ (resp. $G_{\hat{K}^+}$). Therefore, $G_C \in H^-$, and thus the hyperplane L^*, parallel to L and passing through G_C, divides C into two parts in such a way that the part contained in H^- has a volume smaller than $\alpha_n \cdot V(C)$. But by a simple computation we find (since the centroid of a hypercone divides its height in the ratio $1:n$) that the volume in question equals $\alpha_n \cdot V(C)$. The contradiction reached proves the theorem.

4. **Remarks.** (i) It is very easy to find examples which show that the bounds in Theorems 1 and 2 are the best possible. From the proofs given, it is also easy to deduce that if $\mathscr{C}(K, \alpha_n + \varepsilon) = \phi$ for all $\varepsilon > 0$ then K is a simplex, and that $\mathscr{C}(m, 1/(n + 1) + \varepsilon) = \phi$ for all $\varepsilon > 0$ only if the support of m is contained in the "vertex-regions" of some (possibly degenerate) simplex, and all the "vertex-regions" have the same measure.

(ii) The proof of Theorem 1 may be somewhat simplified if the measure m is assumed to be continuous (as in Theorem (A)). The advantage of the more general form is that it includes, e.g., measures generated by finite point-sets, surface-area etc.

(iii) The following obvious corollary of Theorem 2 is interesting because of its independence on the dimension:

For any convex body $K \subset E^n$ we have

$$G_K \in \mathscr{C}(K, e^{-1}) = C(K, 0.3678\cdots) .$$

(iv) It would be interesting to find the analogue of Theorem 2 obtained by substituting the $(n - 1)$-dimensional surface area $A(K)$ for the volume $V(K)$ of $K \subset E^n$. The problem seems to be unsolved even for $n = 2$.

(v) It is easily proved that for any continuous mass-distribution in the plane there exists a pair of orthogonal lines such that each "quadrant" determined by them contains 1/4 of the total mass. The analogous statement is not true for n mutually orthogonal hyperplanes in E^n; does it become true if the condition of orthogonality is omitted?

(vi) It is well known (Buck and Buck [2]) that for any continuous mass-distribution in the plane there exist three concurrent straight lines such that each of the six "wedges" determined by them contains 1/6 of the total mass. Does this fact generalize to E^n when the three lines are replaced by $n + 1$ hyperplanes with a common $(n - 2)$-dimensional intersection?

Added in proof. After submitting the present note for publication, the following facts came to our attention:

(i) Theorems (A) and B are proved, and Theorem 1 suggested, in I. M. Jaglom—W. G. Boltjanski, Konvexe Figuren, Berlin, 1956, pp. 16, 18, 27, 104–106, 116, 135–136 (this is a translation of the Russian original, which appeared in 1951); Theorem (b) is there attributed (without references) to A. Winternitz.

(ii) A proof of Theorem 1 (using Brouwer's fixed-point theorem), together with some related results, was given in B. J. Birch, On 3N points in a plane, Proc. Cambridge Philos. Soc., 55 (1959), 289–293.

(iii) A proof of Theorem 2, very similar to the one given in the present paper, was found independently by P. C. Hammer; it is contained in a paper "Volumes cut from convex bodies by planes", submitted to "Mathematika".

(iv) The relation $\mathscr{C}\left(m, \frac{1}{2}\right) \neq \phi$ (resp. $\mathscr{C}\left(\mathrm{K}, \frac{1}{2}\right) \neq \phi$ holds for any distribution of masses (resp. convex body) with a center of symmetry. Obviously, $\mathscr{C}\left(m, \frac{1}{2}\right) \neq \phi$ is possible also for mass-distributions without a center. The conjecture (trivial for the plane) that $\mathscr{C}\left(K, \frac{1}{2}\right) \neq \phi$ characterizes centrally symmetric convex bodies was first established Professor F. J. Dyson; it is hoped that a proof will be published soon.

(v) Results generalizing Theorem 1 were established by R. Rado in the paper, "A theorem on general measure", J. London Math. Soc., 21 (1946), 291–300. Rado's proof also uses Helley's theorem, but in a fashion different from the one used in the present paper.

References

1. T. Bonnesen and W. Fenchel, *Theorie der konvexen Körper*, Springer, 1934.
2. R. C. Buck and E. F. Buck, *Equipartition of convex sets,* Mathematics Magazine, **22** (1949), 195–198.
3. H. G. Eggleston, *Some properties of triangles as extremal convex curves,* Journ. London Math. Soc., **28** (1953), 32–36.
4. ———, *Problems in Euclidean space: Application of Convexity,* Pergamon Press 1957.
5. ———, *Convexity,* Cambridge Tracts in Math. and Math. Phys., No. 47, Cambridge 1958.
6. P. Halmos, *Measure theory,* Van Nostrand 1950.
7. B. H. Neumann, *On an invariant of plane regions and mass distributions,* Journ. London Math. Soc., **20** (1945), 226–237.
8. D. J. Newman, *Partitioning of areas by straight lines,* Abstract 548–108. Notices Amer. Math. Soc., **5** (1958), 510.
9. H. Rademacher and I. J. Schoenberg, *Helly's theorems on convex domains and Tchebycheff's approximation problem,* Canad. Journ. Math., **2** (1950), 245–256.

Institute for Advanced Study

REGULAR COVERING SURFACES OF RIEMANN SURFACES

SIDNEY M. HARMON

Introduction. The homotopy and homology groups of a given arcwise connected surface are topological invariants. A smooth covering surface F^* is a locally-topological equivalent of its base surface F. Consequently, it is natural that the fundamental and homology groups of F^*, $T(F^*)$ and $H(F^*)$ respectively, should be related to those of F, $T(F)$ and $H(F)$ respectively. In this paper the term homology is always used for the 1-dimensional case. The cover transformations of a covering surface F^* are topological self-mappings such that corresponding points have the same projection on F. These cover transformations form a group which we will denote by $\Gamma(F^*)$. The homology properties of F should influence $\Gamma(F^*)$ by means of the composition of the self-topological mapping and the locally-topological mapping $F^* \to F$.

Section 1 considers the general class of smooth covering surfaces on which there exists a continuation along every arc of the base surface. We refer to such a covering surface as a regular covering surface F^*. A number of results are collected and put into the form in which they are needed to derive the main theorems. The class $\{F^*\}$ is shown to form a complete lattice. Next there is shown a one to one correspondence between all subgroups $N_i \subset T(F)$, such that N_i contains the commutator subgroup N_c of $T(F)$, and the set of all subgroups $H_i \subset H(F)$. This correspondence leads to isomorphisms which relate the associated subgroups.

Section 2 considers a special class of regular covering surfaces $\{F_h^*\}$ in which F_h^* is characterized by the properties that it corresponds to a normal subgroup of $T(F)$ and $\Gamma(F_h^*)$ is Abelian. In our notation these covering surfaces form the class of homology covering surfaces (cf. Kerékjárto [5]). An equivalent characterization of the property that F^* corresponds to a normal subgroup is the assumption that above any closed curve on F there never lie two curves on F^* one of which is closed and the other open. There are derived here for $\{F_h^*\}$ an isomorphism and correspondence theorem which relates subgroups $\Gamma_i \subset \Gamma(F_h^*)$ to quotient groups of $H(F)$ and $T(F)$. The class $\{F_h^*\}$ is shown to form a complete and modular lattice. If the base surface F is an orientable or non-orientable closed surface, with covering surface F_h^*, the rank of $\Gamma(F_h^*)$ is determined in terms of the genus of F and the

Received December 3, 1958. This paper contains a part of the author's doctoral thesis which was completed at the University of California at Los Angeles, together with some extensions. The author is indebted to Professor Leo Sario for his guidance in its preparation.

rank of an associated subgroup $H_i \subset H(F)$.

Section 3 considers the Schottky covering surface F_S^* of a closed orientable surface. We denote the boundary of the conformal equivalent of F_S^* in the plane by E_S. There is obtained here a criterion for the vanishing of the linear measure of E_S.

We will refer to functions defined on a Riemann surface by an abbreviated notation as follows: Green's functions by G, nonconstant analytic functions with finite Dirichlet integral by AD and non-constant analytic bounded functions by AB. We denote the class of Riemann surfaces on which there does not exist any G, AD and AB functions respectively by O_G, O_{AD} and O_{AB}. If W is an open (non-compact) Riemann surface we are led to the problem of studying it from the following comparative viewpoint (Sario [13]). Suppose that P is a property of all closed (compact) Riemann surfaces, determine open Riemann surfaces which possess the same property. Recently Mori [8] established a connection between homology covering surfaces and the classes O_G, O_{AD} and O_{AB}.

Section 4 applies the results of the previous three sections to the classification of Riemann surfaces. It considers regular covering surfaces of a closed Riemann surface F of genus p. We refer to the covering surface of F which corresponds to $N_c \subset T(F)$ as the commutator covering surface F_c^*. It is shown that the results obtained in [8] for homology covering surfaces F_h^* with respect to O_{AD}, O_G and O_{AB} may be applied to any regular covering surface F_i^* which is weaker than F_c^*. In the case of O_{AB} this yields for F_i^* a criterion in terms of the generators of quotient groups of $T(F)$ and $H(F)$. A generalization of Painlevé's problem for an open Riemann surface is proved, and there is also obtained a criterion based on vanishing linear measure of a plane point set which determines that a Schottky covering surface is in O_{AB}.

1. Regular Covering Surfaces.

1.1. Definitions. *A surface* is a connected Hausdorff space on which there exists an open covering by sets which are homeomorphic with open sets of the 2-dimensional Euclidean space.

A surface F^* is a *smooth covering surface* of a base surface F if there exists a mapping $f: F^* \to F$ such that for every $p^* \in F^*$ a neighborhood V^* of p^* is mapped topologically onto a neighborhood V of $p = f(p^*) \in F$.

F^* is a *regular covering surface* of F if it is smooth and if every arc γ on F can be continued along γ from any point over the initial point of γ. [2] (The term "unramified and unbounded" also appears in the literature instead of the term "regular" used here.)

1.2. FUNDAMENTAL GROUP. The results in this subsection are needed for the later treatment and may be found or are implied in the literature; and the development closely parallels that of Ahlfors and Sario [2]. The following result is well-known.

LEMMA 1. *Let* $\{\gamma\}$ *be the homotopic classes of those curves from* 0 *on* F *which have a closed continuation* $\{\gamma^*\}$ *from* $0^* \in F^*$. *Then* $D = \{\gamma\}$ *is a subgroup of the fundamental group* $T(F)$ *with origin at* 0.

Let the notation (F^*, f) and F represent a regular covering surface F^* of F with topological mapping $f: F^* \to F$ and homotopic classes originating at 0^* where $f(0^*) = 0$. We will identify (F_1^*, f_1) and (F_2^*, f_2) if there exists a topological mapping $\phi: F_1^* \to F_2^*$ such that $f_1 = f_2 \circ \phi$ and $\phi(0_1^*) = 0_2^*$. It is clear that this identification is defined by means of an equivalence relation.

The proofs of the following proposition and of the subsequent Lemmas 2 through 4 may be obtained from reference [2] or [9].

PROPOSITION 1. *The mapping* ϕ *in the identification of* (F_1^*, f_1) *and* (F_2^*, f_2) *with* $\phi(0_1^*) = 0_2^*$ *is uniquely determined.*

With the foregoing identification, we obtain

LEMMA 2. *There exists a one to one correspondence between identified pairs* (F^*, f) *and the subgroups* D *of* $T(F)$. *Two pairs can be represented by means of the same* (F^*, f) *if and only if the corresponding subgroups are conjugate.*

LEMMA 3. *The fundamental group* $T(F^*)$ *of* (F^*, f) *is isomorphic with the corresponding subgroup* D *of* $T(F)$.

If (F_2^*, f) covers F_1^* and (F_1^*, f_1) covers F, then it is clear that $(F_2^*, f_1 \circ f)$ covers F where $f_1 \circ f(0_2^*) = 0$. If two pairs (F_2^*, f_2) and (F_1^*, f_1) cover F, we say that the former is *stronger* than the latter if and only if there exists an f such that (F_2^*, f) covers F_1^* and $f_2 = f_1 \circ f$. This relation is clearly transitive.

Let D_1 and D_2 be the subgroups of $T(F)$ which correspond respectively to (F_1^*, f_1) and (F_2^*, f_2), then we have

LEMMA 4. *The pair* (F_2^*, f_2) *is stronger than* (F_1^*, f_1) *if and only if* $D_2 \subset D_1$.

1.3. COMPLETE LATTICE THEOREM. By means of Lemmas 2 and 4, we obtain an ordering of the regular covering surfaces according to relative strength which is isomorphic with the ordering of the corresponding subgroups of $T(F)$ by inclusion.

Let $\{D_a\}$ with a in the index set A be a finite or infinite subset of a lattice L. Then L is *complete* if for all $\{D_a\} \subset L$, there exists in L a least upper bound $\bigcup_{a \in A} D_a$ and a greatest lower bound $\bigcap_{a \in A} D_a$.

THEOREM 1. *The system of regular covering surfaces of F is a complete lattice.*

Proof. The system of subgroups $\{D_a\}$ of $T(F)$ with $a \in A$ is partially ordered by inclusion. Also the union of any number of subgroups $\{D_{a_i}\}$ for $a_i \in A$ is a subgroup $\bigcup_{a_i \in A} D_{a_i}$ which is the least upper bound for $\{D_{a_i}\}$. Similarly, the intersection of any number of subgroups $\{D_{a_i}\}$ is a subgroup $\bigcup_{a_i \in A} D_{a_i}$ which is the greatest lower bound for $\{D_{a_i}\}$. Consequently the system of subgroups $\{D_a\}$ is a complete lattice. Because of the isomorphy obtained from Lemmas 2 and 4, the corresponding regular covering surfaces form a complete lattice.

It can be shown that any complete lattice has a zero and a universal element. The weakest covering surface of F corresponds to $T(F)$ and is F itself or (F^*, e), where e is the identity; the strongest covering surface corresponds to the unit element of $T(F)$ and is the universal covering surface of F.

1.4. RELATIONS BETWEEN FUNDAMENTAL AND HOMOLOGY GROUPS. The commutator subgroup of $T(F)$ will be denoted by N_c. The covering surface F_c^* which corresponds to N_c will be referred to as the commutator covering surface. (Uberlagerungsfläche der Integralflunktionen, Weyl [17])

LEMMA 5. (Nevanlinna [9; 61–63]) *There exists a homomorphism from the elements of $T(F)$ onto the elements of $H(F)$ for which the kernel is the commutator subgroup.*

If θ is a homomorphism from T to H with kernel K, the fundamental theorem for group homomorphisms yields the isomorphism $T/K \cong H$. A second fundamental theorem for group homomorphisms may be stated in the following form (Kurosh [6]).

LEMMA 6. *Let $\theta : F \rightarrow H$ be a homomorphism with kernel K. Then*

(i) *There is a one to one correspondence between subgroups N_i of T such that $T \supset N_i \supset K$ and all subgroups H_i of H. In this correspondence H_i consists of all images of elements of N_i and N_i consists of all inverse images of elements of H_i.*

(ii) *If N_i is normal in T then H_i is normal in H and conversely.*

(iii) *If N_i and K are normal in T then $T/N_i \cong (T/K)/(N_i/K)$.*

THEOREM 2. *Let $\{N_i\}$ be the set of all subgroups such that $T(F) \supset N_i \supset N_c$ and let $\{H_i\}$ be the set of all subgroups $H_i \subset H(F)$. Then*

(i) *There exists a one to one correspondence between* $\{N_i\}$ *and* $\{H_i\}$. *In this correspondence* H_i *consists of all images of elements of* N_i *and* N_i *consists of all inverse images of elements of* H_i.

(ii) $N_i/N_c \cong H_i$.

Proof. To prove the first part, we use the homomorphism of Lemma 5 $\theta : T(F) \to H(F)$ with kernal N_c. Part (i) of the theorem is then an immediate consequence of Lemma 6 (i).

To obtain the isomorphism (ii) we note that N_c is normal in N_i and that the restricted homomorphism $\theta : N_i \to H_i$ is onto. We apply the fundamental theorem for group homomorphisms which yields the required isomorphism.

If in Theorem 2 we set $N_i = T(F)$, we obtain $T(F)/N_c \cong H(F)$ as a special case.

1.5. Relations Between the Fundamental Group and the Group of Cover Transformations.

Definition. A *cover transformation* of a regular covering surface (F^*, f) is a topological self-mapping ϕ such that, for every $p^* \in F^*$, $\phi(p^*)$ and p^* have the same projection.

The totality of cover transformations on F^* clearly form a group. We will denote this group by $\Gamma(F^*)$.

In the sequel, unless otherwise indicated, D or D_i will refer to the subgroup of $T(F)$ which corresponds to the covering surface F^* or F_i^* respectively, according to the specifications of Lemma 2. We note that $\Gamma(F^*)$ and the normalizer of D are unaffected by the choice of 0 and 0*.

Lemma 7. [9; 83] *Let* M *be the normalizer of* D *in* $T(F)$. *Then there exists a homomorphism* $\phi : M \to \Gamma(F^*)$ *with the kernel* D.

Theorem 3. *Let* $\{D_i\}$ *be the set of all subgroups* D_i *such that* $M \supset D_i \supset D$ *and let* $\{\Gamma_i\}$ *be the set of all subgroups of* $\Gamma(F^*)$. *Then*

(i) *There exists a one to one correspondence between* $\{D_i\}$ *and* $\{\Gamma_i\}$. *In this correspondence* Γ_i *consists of all images of elements of* D_i, *and* D_i *consists of all inverse images of elements of* Γ_i.

(ii) $\Gamma_i \cong D_i/D$.

Proof. We use the homomorphism ϕ of Lemma 7 with kernel D. Part (i) of the theorem is then an immediate consequence of Lemma 6 (i). To obtain the isomorphism (ii), we note that D is the kernel of ϕ and D is normal in M and, therefore, normal in $D_i \subset M$. By (i), ϕ maps D_i onto Γ_i. The restriction of ϕ to D_i in conjunction with the fundamental theorem for group homomorphisms yields the required isomorphism.

If in Theorem 3 we set $D_i = M$, we find from part (i) of the theorem that M is mapped onto $\Gamma(F^*)$. Consequently, we obtain from (ii),

$$\Gamma(F^*) \cong M/D\,, \tag{1}$$

as a special case.

COROLLARY. *If D is normal in $T(F)$, then the one to one correspondence and isomorphism specified in Theorem 3 holds for all subgroups D_i, such that $T(F) \supset D_i \supset D$.*

Proof. If D is normal in $T(F)$, then the normalizer of D is $T(F)$. We replace M in Theorem 3 by $T(F)$ and obtain the required result.

A special case of the corollary is obtained if in Theorem 3 (ii) we set $D_i = T(F)$. We then find that

$$\Gamma(F^*) \cong T(F)/D\,. \tag{2}$$

2. Homology Covering Surfaces.

2.1. DEFINITIONS AND BASIC RESULT. A regular covering surface of F is *normal* if it corresponds to a normal subgroup of $T(F)$ [2]. (The term "unramified, unbounded and regular" also appears in the literature instead of the single term "normal" used here.)

PROPOSITION 2. (Seifert-Threlfall [16; 196]) *If (F^*, f) is a normal covering surface of F, then there exists a unique cover transformation which carries any given point $p^* \in (F^*, f)$ into a prescribed point p_1^* with the same projection.*

A regular covering surface is referred to as a *commutative covering surface* if its group of cover transformations is Abelian.

A *homology covering surface* is a covering surface which is simultaneously normal and commutative.

2.2. CRITERION THEOREM.

THEOREM 4. *A regular covering surface F_i^* is a homology covering surface of F if and only if it is weaker than the commutator covering surface F_c^*, or equivalently, if and only if $N_i \supset N_c$, where F_i^* and F_c^* correspond respectively to the subgroups N_i and N_c of $T(F)$.*

Proof. To prove the sufficiency of the condition, we first consider F_c^* which corresponds to N_c which is clearly normal in $T(F)$. By the isomorphism (2), we obtain $\Gamma(F_c^*) \cong T(F)/N_c$. The latter quotient group is Abelian; for if $a, b \in T(F)$, $ab(ba)^{-1} = aba^{-1}b^{-1} \in N_c$; hence $N_c ab = N_c ba$.

By hypothesis, F_i^* is weaker than F_c^*; consequently by Lemma 4, $N_i \supset N_c$. From the fact that $T(F)/N_c$ is Abelian and $N_i \supset N_c$ in conjunction with Lemma 6 (ii), it follows that any subgroup N_i which contains N_c is normal. We conclude from Lemma 6 (iii) that $T(F)/N_i$ is Abelian. The latter quotient group is isomorphic to $\Gamma(F_i^*)$ by the special case (2). We conclude that F_i^* is simultaneously a normal and commutative covering surface and therefore a homology covering surface.

Conversely, we suppose that F_i^* is a homology covering surface. From the special case (2), we obtain $\Gamma(F_i^*) \cong T(F)/N_i$. By hypothesis, the left member of the isomorphism is Abelian; consequently $T(F)/N_i$ is Abelian. Because of the commutativity of $T(F)/N_i$ and the normality of N_i, we obtain for $a, b \in T(F)$, $N_i aba^{-1}b^{-1} = N_i$; therefore, $N_i \supset N_c$. We conclude, by Lemma 4, that F_i^* is weaker than F_c^*.

The last statement of the theorem is an immediate consequence of Lemma 4.

2.3. Isomorphism and Correspondence Theorem.

THEOREM 5. *Let $\{F_{hi}^*\}$ be the set of all homology covering surfaces of F under the identification of Lemma 2, and let $\{N_{hi}\}$ be the set of all corresponding subgroups of $T(F)$ under the isomorphy of Lemma 3; such that $T(F_{hi}^*) \cong N_{hi}$. Let $\{H_i\}$ be the set of all subgroups of $H(F)$ under the correspondence indicated in Theorem 2, such that $N_{hi}/N_c \cong H_i$. Then*

(i) $\Gamma(F_{hi}^*) \cong H(F)/H_i \cong T(F)/N_{hi} \cong [T(F)/N_c]/(N_{hi}/N_c)$.

(ii) *There exists a one-to-one correspondence between the identified sets $\{F_{hi}^*\}$ and the sets $\{N_{hi}\}$ and $\{H_i\}$.*

Proof. To derive the first and second isomorphisms of (i), we note that because of the commutativity of the homology groups, H_i is normal in $H(F)$. We consider the composite mapping $\phi \circ \theta$,

$$\phi \circ \theta[T(F)] = \phi[H(F)] = H(F)/H_i \ ,$$
$$\phi \circ \theta[a \in T(F)] = \phi(a') = H_i a' \ .$$

This mapping is composed of the homomorphism θ of Lemma 5 and the natural homomorphism ϕ; consequently the composition is a homomorphism. The kernel of $\phi \circ \theta$ consists of all $a \in T(F)$ such that $H_i a' = H_i$. We note that by Theorem 4, $N_{hi} \supset N_c$; hence Theorem 2 (i) is applicable. From the specifications in Theorem 2 (i) for $\theta : N_{hi} \to H_i$, we find that the kernel of $\phi \circ \theta$ is precisely N_{hi}. The fundamental theorem of group homomorphism, together with the special case (2), now yield

$$\Gamma(F_i^*) \cong T(F)/N_{hi} \cong H(F)/H_i \ .$$

To derive the third isomorphism of (i), we note that N_{ht} and N_c are normal in $T(F)$. Hence an application of the fundamental theorem for group homomorphisms yields the result.

For the proof of (ii), we note that by Theorem 4, any homology covering surface F^*_{hi} satisfies $N_{hi} \supset N_c$. Hence Theorem 2 is applicable. We apply Theorem 2 (i) to obtain a one-to-one correspondence between $\{N_{hi}\}$ and $\{H_i\}$. The one-to-one correspondence obtained is carried through $\{N_{hi}\}$ to $\{F^*_{hi}\}$, under the postulated identification, by means of Lemma 2. This completes the proof of the theorem.

2.4. COMPLETE AND MODULAR LATTICE THEOREM. A lattice is called *modular* (Dedekind structure) if it satisfies the following weak form of the distributive law:

If $a \supset b$, then $a \cap (b \cup c) = (a \cap b) \cup (a \cap c)$.

LEMMA 8. (Kurosh [7]). *The lattice of normal subgroups of any group is modular.*

THEOREM 6. *The system of homology covering surfaces* $\{F^*_{hi}\}$ *of* F *is a complete and modular lattice.*

Proof. Let $\{N_{hi}\}$ correspond to the collection $\{F^*_{hi}\}$. In the course of the proof of Theorem 1, it was shown that the system af subgroups $\{D_i\}$ of $T(F)$ is a complete lattice. $\{N_{hi}\}$ is therefore a subset of a complete lattice. From the definition of a homology covering surface and from Theorem 4 every N_{hi} is normal in $T(F)$ and $N_{hi} \supset N_c$. The union or intersection of any number of normal subgroups of $\{N_{hi}\}$ is a normal subgroup containing N_c. Consequently, $\{N_{hi}\}$ is a sublattice and a complete lattice. By the normality of N_{hi} and Lemma 8, $\{N_{hi}\}$ is also a modular lattice. We conclude from Theorem 5 (ii) and Lemma 4 that $\{F^*_{hi}\}$ is a complete and modular lattice.

2.5. RANK OF THE GROUP OF COVER TRANSFORMATIONS. We consider the rank of the group of cover transformations for homology covering surfaces for which the base surface F is closed. In this case, $T(F)$ and $H(F)$ are finitely generated. We have

LEMMA 9. (Seifert-Threlfall [16; 145]). *Let* F *be a closed surface of genus* p. *If* F *is orientable,* $H(F)$ *is a free Abelian group of* $2p$ *generators; if* F *is nonorientable,* $H(F)$ *is the direct product of a free Abelian group of* $p-1$ *generators and a group of order* 2.

Because the homology group of a closed surface is finitely generated. it always has a finite rank.

The following lemma is fundamental in the theory of Abelian groups.

LEMMA 10. *Let H be an Abelian group of finite rank r, and let H_i be a subgroup of H. Then H_i and H/H_i are also of finite rank and $r(H) = r(H_i) + r(H/H_i)$.*

THEOREM 7. *Let F be a closed surface of genus p, and let $\{F^*_{hi}\}$ be the class of homology covering surfaces of F such that*

$$T(F^*_{hi}) \cong N_{hi} \subset T(F) , \qquad N_{hi}/N_c \cong H_i \subset H(F) .$$

If F is orientable, then

$$r[\Gamma(F^*_{hi})] = 2p - r(H_i)$$

and

$$0 \leqq r[\Gamma(F^*_{hi})] \leqq 2p .$$

If F is nonorientable, then

$$r[\Gamma(F^*_{hi})] = p - 1 - r(H_i)$$

and

$$0 \leqq r[\Gamma(F^*_{hi})] \leqq p - 1 .$$

In either case, $r[\Gamma(F^*_{hi})]$ assumes all integral values in the indicated ranges.

Proof. We note that the rank of a free Abelian group is equal to the number of its generators, that the rank of an Abelian group in which all elements have finite order is zero, and that the rank of an Abelian group equals the sum of the ranks of the factors in the direct product decomposition of the group. Consequently, it follows from Lemma 9, that if F is orientable, $r[H(F)] = 2p$, and that if F is nonorientable, $r[H(F)] = p - 1$. By use of Theorem 5 (i) and Lemma 10, and by substituting for $r[H(F)]$ the values just deduced we find that if F is orientable

$$r[\Gamma(F^*_{hi})] = 2p - r(H_i) ,$$

and that if F is nonorientable,

$$r[\Gamma(F^*_{hi})] = p - 1 - r(H_i) .$$

Because H_i is a subgroup of $H(F)$

$$0 \leqq r(H_i) \leqq r[H(F)] .$$

For each integer n such that $0 \leqq n \leqq r[H(F)]$, there exists a subgroup H_i which is generated by n linearly independent elements; therefore $r(H_i) = n$. We conclude that if F is orientable,

$$0 \leqq r[\Gamma(F^*_{hi})] \leqq r[H(F)] = 2p ,$$

and that if F is nonorientable,

$$0 \leqq r[\Gamma(F^*_{hi})] \leqq r[H(F)] = p - 1 .$$

In both cases $r[(F^*_{hi})]$ assumes all integral values in the indicated ranges.

In connection with Theorem 7 it is of interest to note that the quantities $2p$ and $p-1$ are the 1-dimensional Betti numbers for a closed orientable and a closed nonorientable surface respectively.

3. Schottky Covering Surface of a Riemann Surface.

3.1. DEFINITIONS FOR RIEMANN SURFACES. We shall define a Riemann surface topologically as a Hausdorff space with certain restrictive properties.

DEFINITION. A *Riemann surface* F is a surface together with a collection of local homeomorphisms $\{h\}$ from open sets of F onto open sets of the complex plane which satisfy the following conditions.

(i) The totality of domains of $\{h\}$ form a covering of F.

(ii) The images of every nonnull common domain of h_i and $h_j \in \{h\}$ are directly conformally equivalent in the complex plane through the composite homeomorphism $h_i \circ h_j^{-1}$.

We denote the domain of $h_i \in \{h\}$ by Δ_i. If $p \in \Delta_i$, then $z = h_i(p)$ is uniquely determined. Because of condition (ii), the conformally invariant properties of F are independent of the choice of $h_i \in \{h\}$. Consequently in considering such properties we may regard z in the complex plane as a local variable instead of $p \in F$. In this paper we shall be concerned exclusively with conformally invariant properties of F; therefore we will resort to the local variable notation z whenever it is convenient.

DEFINITION. A complex-valued function f is *analytic* on F if and only if $f \circ h_i^{-1}$ is analytic on $h_i\ (\Delta_i)$ for every $h_i \in \{h\}$ with domain Δ_i.

DEFINITION. A real-valued function u is *harmonic* on F if $u \circ h_i^{-1}$ is harmonic on $h_i\ (\Delta_i)$ for every $h_i \in \{h\}$ with domain Δ_i.

The Riemann surface as defined here is an orientable surface because the composite mapping $h_i \circ h_j^{-1}$ is directly conformal and consequently sense-preserving. It can be shown that the Riemann surface is topologically a countable space.

3.2. BASIC CONSIDERATIONS. In this section the base surface F is assumed to be a closed Riemann surface of finite genus p. By suitably cutting F, we can obtain a planar region F'_{0S}, such that an infinite number of copies of F'_{0S}, when put together under special identifications of their boundaries, will generate the Schottky covering surface F_S^* of F. The surface F_S^* is a planar, open Riemann surface. We will study the boundary of the conformal equivalent of F_S^* in the complex plane by means of a Schottky group.

3.3. GENERATORS OF SCHOTTKY GROUP. The conformal equivalent of the initial copy F'_{0S} is an infinite region R_0, where R_0 is bounded by $2p$ disjoint circles Q_i, Q'_i $(i = 1, 2, \cdots, p)$, lying in the finite plane. We will refer to this set of circles which bound R_0 as $\{Q_0\}$. The p pairs of circles $\{Q_0\}$ correspond to a system of p hyperbolic or loxodromic linear transformations which generate a group of linear transformations G called the *Schottky group* (Schottky [15]). The group G can be shown to be denumerably infinite and is properly discontinuous up to a set of discrete points E_S, called the *singular set of the Schottky group.* The transforms of R_0 converge for $p > 1$ to a nondenumerable discrete set of points E_S which is the boundary of the conformal image of F_S^* in the plane.

A set has *zero linear measure* if it can be covered by a sequence of disks $\{K_i\}$ with radii $\{r_i\}$ such that $\sum r_i$ is arbitrarily small. We will denote the linear measure of the singular set of the Schottky group by $m(E_S)$.

We consider a configuration of the bounding circles $\{Q_0\}$ corresponding to a Schottky group G, in order to obtain a criterion for the vanishing of $m(E_S)$.

Let the $2p$ circles $\{Q_0\}$ be paired in such a manner that a set of p hyperbolic or loxodromic linear transformations $S_1, \cdots, S_p$ operate on the extended complex plane and yield

$$S_1Q_1 = Q'_1\,, \quad S_2Q_2 = Q'_2\,, \quad \cdots, \quad S_pQ_p = Q'_p\,, \tag{3}$$

with the exterior of each Q_i mapped into the interior of Q'_i. The set of such generators will be designated as $\{S_0\}$. A general form for the transformation S_i is

$$S_i = \frac{az + b}{cz + d}\,. \tag{4}$$

S_i and other linear transformations in the sequel will be normalized by the condition $ad - bc = 1$. The circles $\{Q_0\}$ have the general equations

$$Q_i : |z - q_i| = r_i; \quad Q'_i : |z - q'_i| = r'_i\,. \tag{5}$$

A general normalized transformation of $\{S_0\}$ corresponding to the circles (5) may be written as

$$S_i(z) = \frac{\dfrac{q_i' z}{\sqrt{r_i r_i'}} - \dfrac{q_i q_i' + r_i r_i'}{\sqrt{r_i r_i'}}}{\dfrac{z}{\sqrt{r_i r_i'}} - \dfrac{q_i}{\sqrt{r_i r_i'}}} \tag{6}$$

in which $q_i + r_i e^{i\theta}$ transforms into $q_i' - r_i' e^{i(2\pi - \theta)}$.

The set $\{S_0\}$ corresponding to the form (6) will generate a Schottky group.

Let ξ_1 and ξ_2 denote the fixed points of a generator S_i, where ξ_1 and ξ_2 are finite. Then

$$\xi_1, \xi_2 = \frac{a - d \pm \sqrt{(a+d)^2 - 4}}{2c} . \tag{7}$$

Since $S_i(\infty) = q_i'$, $S_i(z) = z'$ may be expressed in terms of a cross-ratio as

$$\frac{z' - \xi_1}{z' - \xi_2} = K \frac{z - \xi_1}{z - \xi_2} ,$$

where K is a multiplier such that

$$K = \frac{q_i' - \xi_1}{q_i' - \xi_2} , \qquad K \neq 1 .$$

By simplification, this reduces to

$$K = \left(\frac{q_i - q_i' + \sqrt{(q_i' - q_i)^2 - 4 r_i r_i'}}{2\sqrt{r_i r_i'}} \right)^2 . \tag{8}$$

We note that K is independent of z and that the fixed points are independent of the power of z. Consequently, $S_i^n(z) = z^{(n)}$ may be expressed as

$$\frac{z^{(n)} - \xi_1}{z^{(n)} - \xi_2} = K^n \frac{z - \xi_1}{z - \xi_2} .$$

This yields

$$S_i^n(z) = \frac{(K^n \xi_2 - \xi_1) z - \xi_1 \xi_2 (K^n - 1)}{(K^n - 1) z + \xi_2 - K^n \xi_1} .$$

To normalize $S_i^{(n)}(z)$ we divide through by

$$K^{n/2}(\xi_1 - \xi_2) = D'$$

and obtain

$$S_i^n(z) = \frac{\left(\dfrac{K^n \xi_2 - \xi_1}{D'} \right) z - \dfrac{\xi_1 \xi_2 (K^n - 1)}{D'}}{\dfrac{(K^n - 1) z}{D'} + \dfrac{\xi_2 - K^n \xi_1}{D'}} . \tag{9}$$

3.4. ISOMETRIC CIRCLES. Because of the mappings $S_i Q_i = Q_i'$ in the generation of the Schottky group, there is a particular convenience in utilizing the concept of isometric circle under a linear transformation (e.g., Ford [4]).

Let S be a linear transformation expressed in the general form (4). Then length and area are unaltered in magnitude in the neighborhood of a point z if and only if $|cz + d| = 1$. The locus of such points for $c \neq 0$ may be written as the circle $|z + d/c| = 1/|c|$, with center $-d/c$ and radius $1/|c|$.

DEFINITION. Let S be the linear transformation

$$S(z) = (az + b)/(cz + d) .$$

Then the circle

$$I : |cz + d| = 1 , \qquad c \neq 0 ,$$

which is the complete locus of points in the neighborhood of which length and area are unaltered in magnitude by S is called the *isometric circle* of S.

LEMMA 11. *Let the linear transformations S have I as its isometric circle, and let $S(I) = I'$. Then S^{-1} has I' as its isometric circle.*

Proof. By definition S carries I into a circle I' without alteration of lengths in the neighborhood of any point of I. Consequently S^{-1} carries $I' = S(I)$ back to I without alteration of lengths. By the uniqueness of I', we conclude that I' is the isometric circle of S^{-1}.

LEMMA 12. (Ford [4]). *Let I and I' be the isometric circles of S and S^{-1} respectively and let L be the perpendicular bisector of the line joining the centers of I and I'. If S is a hyperbolic, elliptic or parabolic linear transformation, S is equivalent to the composition of an inversion in I followed by a reflection in the line L; if S is loxodromic, there is in addition a rotation about the center of I' through the angle $-2\arg(a + d)$.*

THEOREM 8. *Let S be a linear transformation. Suppose that S and S^{-1} have the isometric circles I and I' respectively. Then for every n*

(i) *The circles $S^{-n}(I)$ and $S^{n}(I')$ are equal in magnitude and $S^{-n}(I) \subset I$, $S^{n}(I') \subset I'$.*

(ii) *$S^{-n}(I)$ is the isometric circle of S^{2n+1}.*

(iii) *The radii of the circles $S^{-n}(I)$ and $S^{n}(I')$ are each equal to*

$1/|c|$, *where c is the coefficient in the general expression for a linear transformation corresponding to* S^{2n+1}.

Proof. Because S and S^{-1} have the respective isometric circles I and I', we conclude from Lemma 11 that $S(I) = I'$. Let L be the perpendicular bisector of the line joining the centers of I and I'. We first consider the case where S is nonloxodromic. Then by Lemma 12, $S^{-n}(I)$ is obtained by successive compositions of an inversion in I' followed by a reflection in L, and $S^{n}(I')$ is obtained by successive compositions of an inversion in I followed by a reflection in L. We note that for all linear transformations the size of the circle is influenced only by the inversion. The circles $S^{-k}(I)$ and $S^{k}(I')$ are symmetrical with respect to L for all $k < n$. Because of the symmetry of the inversion with respect to the equal circles I' and I, we conclude that $S^{-n}(I)$ and $S^{n}(I')$ are equal. Further, from the geometrical interpretation of S^{n} and S^{-n} as expressed by Lemma 12, it follows that $S^{-n}(I) \subset I$ and $S^{n}(I') \subset I'$.

If S is loxodromic, there is in addition, in the foregoing compositions a rotation. For $S^{-k}(I)$ and $k < n$ the required rotation is $-2k \arg[-(a + d)] = -2k\pi - 2k \arg(a + d)$ about the center of I, and for $S^{k}(I')$ the required rotation is $-2k \arg(a+d)$ about the center of I'. The circles $S^{-k}(I)$ and $S^{k}(I')$ are therefore symmetrical with respect to the intersection of L and the line joining the centers of I' and I. This symmetry yields equal circles in the successive inversions with respect to the circles I' and I. We conclude again that $S^{-n}(I)$ and $S^{n}(I)$ are equal and that $S^{-n}(I) \subset I$ and $S^{n}(I') \subset I'$. This completes the proof of part (i).

To prove part (ii) we consider $S^{2n+1} \circ S^{-n}(I)$. The first n operations by S transform $S^{-n}(I)$ to I. The inversions associated with these transformation are all in I and are of the type $S^{-(n-j)}(I)$ inverts to $S^{n-j-1}(I')$, where $j = 0, 1, \cdots, n-1$. The $n+1$st operation transforms I to I' and involves the identity inversion, i.e., I inverts to I. The last n operations by S transform I' to $S^{n}(I')$. The inversions associated with these transformations are all in I and are of the type $S^{n-j-1}(I')$ inverts to $S^{-(n-j)}(I)$. The latter n inversions are thus inverses of the aforementioned n inversions. Hence the resulting inversions associated with S^{2n+1} preserve infinitesimal lengths on $S^{-n}(I)$. The reflection and rotation components of S^{2n+1} clearly preserve infinitesimal lengths. Therefore $S^{-n}(I)$ is the isometric circle of S^{2n+1}.

Part (iii) of the theorem is a consequence of the fact that an isometric circle may be written in the form $|z + d/c| = 1/|c|$.

We collect here some results on the inversion of one circle into another circle which will be needed subsequently. In the sequel, the circles Q_1 and Q_2 are always disjoint. If a circle Q_1 is inverted into a circle Q_2, we will designate the image circle by Q_{12} and a corresponding subscript notation will be used for the radii r and centers q of the respective circles.

Let Q_1 and Q_2 be given by

$$Q_1 : |z - q_1| = r_1 ,$$
$$Q_2 : |z - q_2| = r_2 .$$

We denote by l the line which passes through the centers of Q_1, Q_2 and we take the points $\alpha, \beta \in Q_1 \cap l$. Suppose that Q_1 is inverted into Q_2 with α, β transforming into α_i, β_i respectively. Then

$$|\alpha - q_2|\cdot|\alpha_i - q_2| = r_2^2 ,$$
$$|\beta - q_2|\cdot|\beta_i - q_2| = r_2^2 .$$

We denote the distance between q_1 and q_2 by e and obtain

$$r_{12} = \frac{r_1 r_2^2}{|\alpha - q_2|\cdot|\beta - q_2|} = \frac{r_1 r_2^2}{e^2 - r_1^2} , \tag{11}$$

$$|q_{12} - q_2| = \frac{e r_2^2}{e^2 - r_1^2} . \tag{12}$$

LEMMA 13. *Let Q_1, Q_2 be disjoint circles with centers q_1, q_2 and radii r_1, r_2 respectively. Then*

(i) r_{12} increases with increasing r_1 and fixed e and also with decreasing e and fixed r_1.

(ii) If Q_1 is enlarged to $Q_{1'}$ in such a manner that $Q_1 \subset Q_{1'}$ and $Q_{1'}$ is disjoint from Q_2, then $r_{1'2} > r_{12}$ and $r_{21'} > r_{21}$.

(iii) q_{12} lies on the line joining q_1 and q_2, and $|q_2 - q_{12}|$ decreases with increasing e.

Proof. To prove (i) we note that because Q_1 and Q_2 are disjoint, $e > r_1$. The result then follows from equation (11).

For the proof of (ii), we denote the line passing through q_1 and q_2 by l. It is sufficient to consider the case in which the center $q_{1'}$ lies on l and one of the two points in $Q_1 \cap l$ is fixed during the enlargement of Q_1. We use the first equation in (11) to find the total derivative with respect to $r_{1'}$. We obtain the result that if $Q_{1'}$ is inverted into Q_2, $dr_{1'2}/dr_{1'} > 0$; and if Q_2 is inverted into $Q_{1'}$, $dr_{21'}/dr_{1'} > 0$. Because $r_{1'}$ is steadily nondecreasing, we conclude that $r_{1'2} > r_{12}$ and $r_{21'} > r_{21}$.

The first part of (iii) follows from elementary geometrical considerations of inversions. The second part of (iii) is obtained by differentiating, in equation (12), $|q_2 - q_{12}|$ with respect to e and noting that the derivative is negative.

3.5. CRITERION FOR VANISHING LINEAR MEASURE. In the sequence of circles which bound the successive generations of mapped regions in the conformal mapping of F_S^*, the size of the circles is influenced only by the inversions associated with the elements of G in the Schottky group. Suppose we enlarge any circle $\theta_i \in \{Q_0\}$ to $Q_{i'}$ in such a manner that $Q_i \subset Q_{i'}$ and $Q_{i'}$ are disjoint from all other circles in $\{Q_0\}$. Then by repeated applications of Lemma 13 (i) and (ii) it follows that in the limit $m(E_S')$ for the new configuration will be greater than $m(E_S)$. Consequently, for establishing a criterion for the vanishing of $m(E_S)$, we may modify the configuration $\{Q_0\}$ to one in which all circles Q_i are of equal unit size, subject to the conditions just mentioned. We will refer to this modified configuration of $\{Q_0\}$ as $\{Q_0\}_A$.

We consider the configuration $\{Q_0\}_A$. Let e_i be the distance between the centers of the pair Q_i and Q_i' $(i = 1, 2, \cdots, p)$, and let ${}_id_j$ be the distance between the centers of two arbitrary circles Q_i and $Q_j \in \{Q_0\}_A$. We denote by e the minimum e_i and by d the minimum ${}_id_j$. If

$$\text{(B)} \qquad d \geqq e \ ,$$

we will say that $\{Q_0\}_A$ satisfies condition (B) and denote the configuration by $\{Q_0\}_{AB}$. The modified configuration $\{Q_0\}_{AB}$ will have a corresponding group of hyperbolic or loxodromic linear transformations G' which is associated with the Schottky group G corresponding to $\{Q_0\}$. In the sequel, we will use the same notation for the circles in $\{Q_0\}_{AB}$ and for the generators of G' as used previously for those in $\{Q_0\}$ and in $\{S_0\}$ respectively.

THEOREM 9. *Let G be a Schottky group with p generators. Suppose that there exists a configuration $\{Q_0\}_{AB}$ which is associated with G. Then the linear measure of the singular set of G vanishes if*

$$\text{(C)} \qquad p < \frac{e}{4}(e + \sqrt{e^2 - 4}) \ .$$

Proof. Because of equations (3) and (6) in subsection 3.3 and because Q_i and $Q_{i'}$ are equal for all i, Q_i and $Q_{i'}$ are the isometric circles of the hyperbolic or loxodromic linear transformations S_i and S_i^{-1} respectively. Consequently, an arbitrary element of the group generated by $\{S_0\}$ is by Lemma 12 equivalent to a succession of compositions. Each of these compositions is an inversion in one of the circles $\{Q_0\}_{AB}$ followed by a reflection in the perpendicular bisector of the line joining the center of this circle to the center of its paired circle and a rotation about the center of some Q_i. We note that in the compositions, the size of the image circles is influenced only by the inversions.

Let Q_1 and Q_1' with centers at q_1 and q_1' respectively be that pair of circles in $\{Q_0\}_{AB}$ which has the minimum distance e between their centers, and let S_1 be the corresponding generator. We may take q_1 at the origin and q_1' to be positive and real. Thus

$$Q_1 : |z| = 1 ; \qquad Q_1' : |z - q_1'| = 1 . \tag{13}$$

With this choice, we find from equation (6) that $a + d$ is real and $|a + d| > 2$; hence S_1 is hyperbolic.

By hypothesis, the distance between q_1 and q_1' is smallest for the circles Q_1 and Q_1' in comparison with any other two circles in $\{Q_0\}_{AB}$; also, all of the circles in $\{Q_0\}_{AB}$ are equal. Consequently, we conclude from Lemma 13 (i) that the circle $S_1(Q_1') \subset Q_1'$ has the maximal radius for all circles of the first generation. We denote by q_{S_1} the center of $S_1(Q_1')$. By noting that S_1 is hyperbolic, it follows from Lemma 13 (iii) together with simple geometrical considerations that the distance between q_{S_1} and q_1 is minimal in comparison with the distance between the center of any other circles $S_i(Q_j) \subset Q_i'$ of the first generation and the center of any circle in $\{Q_0\}_{AB}$ exterior to Q_i'. Consequently, if we apply Lemma 13 (i) again, we find that $S_1^2(Q_1') \subset Q_1'$ has the maximal radius for all circles of the second generation.

Another application of Lemma 13 (iii) shows that the distance between $q_{S_1^2}$ and q_1 is minimal in comparison with the distance between the center of any other circle $S_1 \circ S_k(Q_j) \subset Q_1'$ of the second generation and the center of any circle in $\{Q_0\}_{AB}$ exterior to Q_1'. Similarly we obtain a corresponding result for the nth generation. We conclude by induction that the circle $S_1^n(Q_1') \subset Q_1'$ has the maximal radius for all circles of the nth generation for all n.

Let r_n denote the radius of $S_1^n(Q_1') \subset Q_1'$. We note that S_1 and S_1^{-1} have the isometric circles Q_1 and Q_1' respectively. Consequently Theorem 8 (iii) is applicable and we obtain

$$r_n = 1/|c| ,$$

where c refers to the coefficient of the linear transformation corresponding to S_1^{2n+1}. By utilizing this equation and equations (9), (13), (6), (7) and (8) and replacing q_1' by e, we obtain

$$\begin{aligned} r_n &= \left| \frac{\xi_1 - \xi_2}{K^{\frac{2n+1}{2}} - K^{\frac{-(2n+1)}{2}}} \right| \\ &= \frac{\sqrt{e^2-4}}{\left| \left(\frac{-e+\sqrt{e^2-4}}{2}\right)^{2n+1} - \left(\frac{-e+\sqrt{e^2-4}}{2}\right)^{-(2n+1)} \right|} \\ &= \frac{\sqrt{e^2-4}}{\left(\frac{-e+\sqrt{e^2-4}}{2}\right)^{2n+1} + \left(\frac{e+\sqrt{e^2-4}}{2}\right)^{2n+1}} . \end{aligned}$$

The total number of circles in the nth generation is $2p(2p-1)^n$. We denote the total length of these circles by L_n. Then

$$L_n \leqq \frac{4\pi p(2p-1)^n\sqrt{e^2-4}}{\left(\frac{-e+\sqrt{e^2-4}}{2}\right)^{2n+1}+\left(\frac{e+\sqrt{e^2-4}}{2}\right)^{2n+1}}.$$

We find that

$$\lim_{n\to\infty} L_n = 0$$

if

$$p < \frac{(e+\sqrt{e^2-4})^2+4}{8} = \frac{e(e+\sqrt{e^2-4})}{4}.$$

Because $m(E_S) \leqq \lim_{n\to\infty} L_n$, this is the required criterion.

COROLLARY. *Suppose that the Schottky covering surface* F_S^* *corresponds to a Schottky group* G *with* p *generators. Let* G *be associated with a configuration* $\{Q_0\}_{AB}$ *which satisfies Condition C of Theorem 9. Then the boundary of the conformal equivalent of* F_S^* *in the plane has zero linear measure.*

Proof. By definition the boundary of the conformal equivalent of F_S^* in the plane is the singular set of G. The conclusion then follows immediately from Theorem 9.

4. Classification of Riemann Surfaces.

4.1. EXHAUSTIONS AND HARMONIC MODULI. An arc is *analytic* if it is the conformal image of a closed interval in the complex plane.

By virtue of the countability of a Riemann surface there always exists on such a surface an exhaustion which may be described as follows.

DEFINITION. A nested sequence $\{w_n\}$ of compact regions is an exhaustion of an open Riemann surface W if

(i) W_n is interior to W_{n+1}.

(ii) The boundary β_n of W_n consists of a finite number of closed disjoint piecewise analytic curves.

(iii) Each complement $W_n - W_{n-1}$ consists of a finite number of disjoint noncompact regions.

(iv) $\bigcup_{n\to\infty} W_n = W$.

For every n $(n = 0, 1, \cdots)$, the complement $W_n - W_{n-1}$ consists of a finite number $k(n)$ of disjoint subregions E_{ni} $(i = 1, 2, \cdots, k(n))$ of finite genus. The boundary of E_{ni} consists of two or more closed disjoint piecewise analytic curves which are subsets of β_{n-1} and β_n. We denote the intersections of the boundary of E_{ni} with β_{n-1} and β_n, by β_{ni} and β'_{ni} respectively. There exists on E_{ni} a unique harmonic function u_{ni} which is continuous on the closure of E_{ni}, vanishes on β_{ni} and is constantly equal to unity on β'_{ni}. The function u_{ni} is called the harmonic measure of β'_{ni} with respect to E_{ni}.

If E_{ni} is planar and β_{ni} and β'_{ni} each consist of one component, then E_{ni} is doubly connected. In this case, the function $U = e^{u_{ni}+iu^*{ni}}$ maps E_{ni} conformally onto an annulus, where u^*_{ni} represents the conjugate harmonic function of u_{ni}.

Let E_{ni} $(i = 1, 2, \cdots, k(n) < \infty,\ n = 0, 1, \cdots)$ be a collection of doubly-connected subregions of the open Riemann surface W, which may be represented as annuli and which satisfy the following conditions:

(i) Each annulus E_{ni} is bounded by two closed, disjoint and piecewise analytic curves β_{ni} and β'_{ni}.

(ii) Any two of the annuli have no points in common.

(iii) The complementary set of $\bigcup_{i=1}^{k(n)} E_{ni}$ with respect to W has precisely one compact component W_n.

(iv) W_n is bounded by the $k(n)$ curves and contains the annuli $E_{n'i}$ with $n' < n$.

We define the *harmonic modulus* μ_{ni} of E_{ni} as

$$\mu_{ni} = 2\pi \Big/ \int_{\beta_{ni}} du^*_{ni} \ .$$

4.2. General Concept. The classification problem will be studied from the viewpoint of Sario [13] which classifies open Riemann surfaces according to their possession or nonpossession of a given property P shared by all closed Riemann surfaces. If W has the property P, we say that W has a removable boundary with respect to P. Thus the behavior of the open surface with respect to P is the same as if it were closed, that is, had no boundary. We will consider three properties shared by all closed Riemann surfaces, namely, they possess no G, AD or AB functions.

4.3. The Class 0_G. The Green's function $g(z, \zeta)$ of a relatively compact Jordan region R is defined as the unique harmonic function on R which possesses the singularity $-\log|z-\zeta|$ at a point $\zeta \in R$ and which vanishes continuously on the boundary β of R.

In order to generalize this definition to an arbitrary open Riemann surface, we will require the well-known *Harnack's Principle* which we state in the following form [2].

Lemma 14. *Suppose that a family $\mathscr{U}$ of harmonic functions on a*

Riemann surface W satisfies the following condition.

To any $u_i, u_j \in \mathscr{U}$ there exists a $u_k \in \mathscr{U}$ with $u_k \geqq \max(u_i, u_j)$ on W. Then the function

$$U(z) = \sup_{u_i \in \mathscr{U}} u_i(z)$$

is either harmonic or constantly equal to ∞.

We consider an open Riemann surface W and an exhaustion of the type described in subsection 4.1. If W_n is one of the compact elements of the sequence $\{W_n\}$ in the exhaustion, its Green's function $g_n(z, \zeta)$ has the usual interpretation. By the maximum principle $g_n(z, \zeta)$ is a monotone increasing sequence of harmonic functions on W. Consequently by Harnack's principle, the sequence has a limiting function $g(z, \zeta)$ on W which is either harmonic with the exclusion of the pole $-\log|z-\zeta|$ or else is identically infinite. In the first case we define $g(z, \zeta)$ to be Green's function for W with a pole at ζ. It can be shown that if the Green's function g exists it is the smallest positive harmonic function with the singularity $-\log|z-\zeta|$. Also it satisfies the equality $\inf g = 0$. If a harmonic function with the same singularity as g tends to 0 as z approaches the boundary of W, then it is identical with g. We conclude that the Green's function is independent of the exhaustion.

LEMMA 15. Mori [8]. *Let F_h^* be a homology covering surface of a closed Riemann surface F and let $r[\Gamma(F_h^*)]$ be the rank of the group of cover transformations of F_h^*. Then $F_h^* \in 0_G$ if and only if $r[\Gamma(F_h^*)] \leqq 2$.*

THEOREM 10. *Let F_i^* be a regular covering surface of a closed Riemann surface F such that F_i^* is weaker than the commutator covering surface of F, or equivalently*

$$T(F_i^*) \cong N_i \subset T(F)\,, \qquad N_i \supset N_c\,, \qquad N_i/N_c \cong H_i \subset H(F)\,.$$

Then $F_i^* \in 0_G$ if and only if

$$\text{(i)} \qquad r[\Gamma(F_i^*)] = r\left[\frac{T(F)}{N_i}\right] = r\left[\frac{H(F)}{H_i}\right] = r\left[\frac{T(F)/N_c}{N_i/N_c}\right] \leqq 2$$

or equivalently

$$\text{(ii)} \qquad 2p-2 \leqq r(H_i) \leqq 2p\,.$$

Proof. To prove (i) we note that by Theorem 4, F_i^* is a homology covering surface. The conclusion then follows from Lemma 15 and Theorem 5 (i).

To prove (ii), we note that F_i^* is a homology covering surface and F is orientable. Consequently, Theorem 7 for the orientable case is applicable. We obtain

$$r[\Gamma(F_i^*)] = 2p - r(H_i) ,$$

in which

$$0 \leqq r(H_i) \leqq r[H(F)] .$$

If F_i^* satisfies (i),

$$r[\Gamma(F_i^*)] \leqq 2 .$$

Therefore,

$$2p - 2 \leqq r(H_i) \leqq r[H(F)] .$$

Conversely, we suppose that F_i^* satisfies (ii). Then

$$r[\Gamma(F_i^*)] = 2p - r(H_i) \leqq 2p - (2p - 2) = 2 .$$

Hence, (i) and (ii) are equivalent.

4.4. THE CLASS 0_{AD}. If $f(z)$ is an analytic function on a Riemann surface W, the Euclidean area of the image W' is given by the Dirichlet integral

$$D(f) = \iint_W |f'(z)|^2 dxdy ,$$

where $z = x + iy$ is the local variable. It follows that the existence on W of an AD function implies the existence of a conformal equivalent of W with finite Euclidean area. For simply-connected regions, the possibility of conformal equivalence with a finite or infinite disk is precisely the classical type problem. Hence the classification according to 0_{AD} is a generalization to arbitrary Riemann surfaces of this classical problem.

LEMMA 16. Mori [8]. *If F_h^* is a homology covering surface of a closed Riemann surface, then $F_h^* \in 0_{AD}$.*

THEOREM 11. *If F_i^* is a regular covering surface of a closed Riemann surface F such that F_i^* is weaker than the commutator covering surface of F, or equivalently,*

$$T(F_i^*) \cong N_i \subset T(F) , \qquad N_i \supset N_c ,$$

then

$$F_i^* \in 0_{AD} .$$

Proof By Theorem 4, F_i^* is a homology covering surface. The result is then a consequence of Lemma 16.

4.5. THE CLASS 0_{AB}. If we consider an AB function $f(z)$ defined in a region W, of the extended complex plane, which is complementary to a finite set of isolated points $\{p_i\}$, it is well known from the classical theory that the singularities $\{p_i\}$ can be removed by appropriately defining $f(z)$ at the points p_i. Painlevé [10] generalized this concept by investigating the analytic continuation of AB functions across arbitrary point set boundaries of regions in the extended complex plane. This is the classical Painlevé's problem.

The connection of the classification according to 0_{AB} with Painlevé's problem is shown by the following lemma.

LEMMA 17. [10], [1]. *Suppose E is a compact set in the extended plane and W is its complement. Let G be a relatively compact region in the plane with analytic boundary α and $E \subset G$. If $G_0 = G - E$, then every AB function, defined in G_0, possesses an analytic continuation to all of G, if and only if $W \in 0_{AB}$.*

Proof. Suppose that $W \in 0_{AB}$. Let $F(z) \in AB$ be defined in G_0. By the compactness of E we can enclose the points of E in a finite number of piecewise analytic closed curves $\{C_i\}$. We apply Cauchy's integral formula to the region contained in G but exterior to $\{C_i\}$. Then we can write

$$f(z) = f_1(z) + f_2(z) ,$$

where $f_1(z)$ is analytic in G, and $f_2(z)$ is analytic in the region exterior to $\{C_i\}$. We have for $f_2(z)$,

$$|f_2(z)| \leqq Ml/\rho ,$$

where M is the supremum of $f(z)$, l is a finite length and $\rho > 0$. Consequently $f_2(z)$ is an AB function in W. Because $W \in 0_{AB}$, $f_2(z)$ is constant. Consequently $f_1(z)$ + constant is an analytic continuation of $f(z)$ across E.

Conversely, we suppose that the analytical continuation across E is possible for every AB function defined in G_0. If $f(z)$ is an AB function on W, then the analytic continuation of $f(z)$ across E is an AB function in the extended plane. Therefore $f(z)$ must reduce to a constant. Hence we conclude that $W \in 0_{AB}$.

The lemma just proved shows that Painlevé's problem is the special case of the classification according to 0_{AB}, where the surface is restricted to plane regions.

The following lemma is implicit in the works of Painlevé [10].

LEMMA 18. *Let E be a compact set in the extended plane and let W be the complement of E. If the linear measure of E is zero, then* $W \in 0_{AB}$.

The following is a generalization of Lemma 18.

THEOREM 12. *Let W be an open Riemann surface with boundary β. Suppose that there exists a planar neighborhood N of β such that the relative boundary of N is a single contour α. If the boundary of the conformal equivalent of N in the plane has zero linear measure, then* $W \in 0_{AB}$.

Proof. N is planar by hypothesis; therefore it can be mapped conformally onto a region N' of a disk $K: |z| < 1$. In this mapping β appears as a closed point set E interior to K. The linear measure of E vanishes by hypothesis; therefore by Lemma 18, $W \in 0_{AB}$.

If W is of finite genus p with boundary β, then the postulated planar neighborhood of β in Theorem 12 is assured. For in this case, we can find a compact region $W_0 \subset W$, with genus p, bounded by a single contour α, with α lying entirely in W. The complement $N = W - W_0$ is then a planar neighborhood of β and has a single contour α as its relative boundary. The following corollary is then an immediate consequence of Theorem 12.

COROLLARY. *If W is of finite genus and if the linear measure of β vanishes under the conformal mapping of N in the plane, then* $W \in 0_{AB}$.

THEOREM 13. *Let F be a closed Riemann surface of finite genus p. Suppose that there exists for the Schottky covering surface F_S^* of F a modified configuration $\{Q_0\}_{AB}$, in the sense of subsection* 3.5 *such that $p < (e/4)(e + \sqrt{e^2 - 4})$. Then* $F_S^* \in 0_{AB}$.

Proof. By the corollary to Theorem 9, the boundary of the conformal equivalent of F_S^* in the plane has zero linear measure. We note that F_S^* is an open Riemann surface of zero genus. The conclusion then follows from the corollary to Theorem 12.

We consider an open Riemann surface W on which the domains of the homeomorphism $h_i \in \{h\}$ are denoted by Δ_i. Let $\lambda(z)$ be a continuous and positive (except for isolated points) function on each domain Δ_i of W. If two domains Δ_j and Δ_k overlap, let $\lambda(z)$ satisfy the covariance relation

$$\lambda(z_j) = \lambda(z_k)\left|\frac{dz_k}{dz_j}\right|$$

at corresponding points z_j and z_k in $\Delta_j \cap \Delta_k$. We further require that all points in W have an infinite distance from the ideal boundary of W. We say that the differential

$$ds = \lambda(z)\,|\,dz\,|$$

defines a *conformal metric* on W, if it satisfies all the conditions just indicated.

Suppose that a conformal metric is defined on W. We fix a point 0 in W and let D_ρ be the domain formed by those points whose distance from 0 is less than ρ, where $0 < \rho < \infty$. For $\rho < \infty$, we assume that the domains are compact and that they generate W as $\rho \to \infty$. Each domain D_ρ is bounded by β_ρ, where β_ρ consists of a finite number $k(\rho)$ of closed disjoint piecewise analytic curves, $\beta_{\rho 1}, \beta_{\rho 2}, \cdots, \beta_{\rho[k(\rho)]}$. Let

$$l_i = \int_{\beta\rho_i} ds\ , \qquad i = 1, 2, \cdots, k(\rho)\ ,$$

$$\Lambda(\rho) = \max_i \int_{\beta\rho_i} ds\ ,$$

$$K(N) = \max_{\rho' \leq \rho} k(\rho')\ .$$

Then we have

LEMMA 19. (Pfluger [11]). *If*

$$\overline{\lim_{N\to\infty}}\left[4\pi\int_0^N \frac{d\rho}{\Lambda(\rho)} - \log K(N)\right] = \infty$$

on W, *then* $W \in 0_{AB}$.

In [8], Mori states without proof a modification of Lemma 19 which does not involve the assumption of a conformal metric on the surface. For the modified version of the lemma, we assume an exhaustion of W and obtain as in subsection 4.1 the corresponding collection of annuli $\{E_{ni}\}$. We set

$$\mu_n = \min_i \mu_{ni} = 2\pi\Big/\int_{\beta_{ni}} du^*_{ni}\ ,$$

$$K(N) = \max_{n \leq N} k(n)\ .$$

Then we prove

LEMMA 20. *If*

$$\overline{\lim_{N\to\infty}}\left\{\sum_{n=j}^{N}\mu_n - \frac{1}{2}\log K(N)\right\} = \infty\ ,$$

then

$$W \in 0_{AB}$$

Proof. We consider the postulated exhaustion of W and the corresponding annuli $\{E_{ni}\}$. Let E_{ni} be one such annulus which is bounded by β_{ni} and β^*_{ni} and let $u_{ni}(z)$ be the harmonic measure of β^*_{ni} with respect to E_{ni}. By the maximum principle, $0 < u_{ni}(z) < 1$ in E_{ni}. We define the function $u_{ni}(z)$ to be the distance of the point z from β_{ni}. Then the function $|\operatorname{grad} u_{ni}(z)|$ defines a conformal metric on the annulus E_{ni}, for which

$$ds = |\operatorname{grad} u_{ni}(z)|\,|dz|\ .$$

Let β_{ρ_i} denote the set of points on E_{ni} which have the distance ρ from β_{ni}. Then

$$l_i = \int_{\beta_{\rho_i}} ds = \int_{\beta_{\rho_i}} \frac{\partial u_{ni}}{\partial n}\,|dz| = \int_{\beta_{\rho_i}} du^*_{ni} = \frac{2\pi}{\mu_{ni}}$$

where ∂n is normal to ds.

The result then follows from Lemma 19.

In [8], Mori utilized Lemma 20 to prove

LEMMA 21. *Let F^*_h be a homology covering surface of a closed Riemann surface F. Suppose that the group of cover transformation $\Gamma(F^*_h)$ has the system of $2p$ generators C_{2i-1}, C_{2i} $(i = 1, 2, \cdots, p)$. If there exists for each i a relation of the form*

$$\gamma_{2i-1}C_{2i-1} + \gamma_{2i}C_{2i} = 0$$

where γ_{2i-1} and γ_{2i} are integers and do not vanish simultaneously, then $F^*_h \in 0_{AB}$.

Let F be a closed Riemann surface of genus p. Suppose that F is cut along p disjoint nondividing cycles to produce a planar surface F'_0. Following Royden [12], we shall refer to a regular covering surface F^* of F as a covering surface of type S, if it consists of a finite or infinite number of copies of F'_0.

COROLLARY. [8]. *A homology covering surface F^*_h of type S of a closed Riemann surface F is in 0_{AB}.*

Proof. Let the $2p$ nondividing cycles C_{2i-1}, C_{2i} $(i = 1, 2, \cdots, p)$ correspond to the $2p$ generators of $\Gamma(F^*_h)$. If we cut F along the nondividing cycles C_{2i-1} $(i = 1, 2, \cdots, p)$, then the cycles C_{2i-1} correspond to the identity element in $\Gamma(F^*_h)$. Hence we may take $\gamma_{2i-1} = 1$ and $\gamma_{2i} = 0$ and obtain

$$\gamma_{2i-1}C_{2i-1} + \gamma_{2i}C_{2i} = 0 \, , \qquad (i = 1, 2, \cdots, p) \, .$$

The conclusion then follows from Lemma 21.

THEOREM 14. *Let F_i^* be a regular covering surface of a closed Riemann surface F of genus p such that F_i^* is weaker than the commutator covering surface of F, or equivalently,*

$$T(F_i^*) \cong N_i \subset T(F) \, , \quad N_i \supset N_c \, , \quad (N_i/N_c) \cong H_i \subset H(F) \, .$$

Suppose that

(i) *$\Gamma(F_i^*)$ has the $2p$ generators C_{2i-1}, C_{2i} $(i = 1, 2, \cdots, p)$ such that C_{2i-1}, C_{2i} correspond respectively to a_{2i-1}, a_{2i} under the isomorphisms of Theorem* 5 (i). *If there exists for each $i = 1, 2, \cdots, p$ a relation of the form*

$$\gamma_{2i-1}a_{2i-1} + \gamma_{2i}a_{2i} = 0$$

where γ_{2i-1} and γ_{2i} are integers and do not vanish simultaneously, and a_{2i-1}, a_{2i} $(i = 1, 2, \cdots, p)$ refer to the $2p$ generators of the Abelian groups

$$\frac{H(F)}{H_i} \, , \quad \frac{T(F)}{N_i} \quad \text{and} \quad \frac{T(F)/N_c}{N_i/N_c} \, ,$$

or

(ii) F_i^* is of type S,

then $F_i^* \in 0_{AB}$.

Proof. By Theorem 4, F_i^* is a homology covering surface. The conclusion then follows from Lemma 21 and its corollary in conjunction with Theorem 5 (i).

REFERENCES

1. Lars Ahlfors and Arne Beurling, *Conformal invariants and function-theoretic null-sets*, Acta Math., **83** (1950), 101-129.
2· Lars Ahlfors and Leo Sario, *Riemann Surfaces*, Princeton University Press, Princeton 1960.
3. R. Courant, *Geometrische Funktionentheorie.* Hurwitz-Courant: *Funktionentheorie*, Springer, Berlin, (1925), 476-484.
4. Lester R. Ford, *Automorphic Functions*, McGraw-Hill, New York, (1929), 23.
5. B. V. Kerékjárto, *Vorlesungen Uber Topologie*, Springer, Berlin, (1923), 182-184.
6. A. G. Kurosh, *The Theory of Groups*, Chelsea, New York, **1** (1955), 71-74.
7. ———, *The Theory of Groups*, Chelsea, New York, **2** (1955), 92.
8. A. Mori, *A note on unramified Abelian covering surfaces of a closed Riemann surface*, J. Math. Soc. Japan, **6** (1954), 162-176.
9. Rolf Nevanlinna, *Uniformisierung*, Die Grundlehren der Mathematischen Wissenschaften in Einzeldarstellungen mit besonderer Berücksichtigung der Anwendungsgebiete, Bd. LXIV, Springer-Verlag, Berlin-Gottingen-Heidelberg, (1953), 64-83.

10. P. Painlevé, *Sur les lignes singulières des fonctions analytiques*, Ann. Fac. Sci. Toulouse, **2** (1888).
11. Albert Pfluger, *Sur l'existence de fonctions non constantes, analytiques, uniformes et bornées sur une surface de Riemann Ouverte*, C. R. Acad. Sci. Paris, **230** (1950), 166-168.
12. H. L. Royden, *Harmonic functions on open Riemann surfaces*, Trans. Amer. Math. Soc., **73** (1952), 40-94.
13. Leo Sario, *Uber Riemannsche Flachen mit hebbarem Rand*, Ann. Acad. Sci. Fenn. Ser. A. I. Math.-Phys., **50** (1948).
14. ———, *Modular criteria on Riemann surfaces*, Duke Math. J., **20** (1953), 279-286.
15. F. Schottky, *Ueber eine specielle Function, welche bei einer bestimmten linearen Transformation ihres Argumentes unverändert bleibt*, Crelle's J., **101** (1887), 227.
16. H. Seifert and W. Threlfall, *Lehrbuch der Topologie*, Teubner, Leipzig und Berlin, 1934.
17. H. Weyl, *Die Idee der Riemannschen Flache*, Teubner, Leipzig und Berlin, 1913.

FRESNO STATE COLLEGE

THE MULTIPLICATIVE SEMIGROUP OF INTEGERS MODULO m

EDWIN HEWITT AND H. S. ZUCKERMAN

1. Introduction. Throughout this paper, m denotes a fixed integer >1. The set of all residue classes modulo m is denoted by S_m. For an integer x, $[x]$ denotes the residue class containing x. Under the usual multiplication $[x]\cdot[y] = [xy]$, S_m is a semigroup. The subgroup of S_m consisting of all residue classes $[x]$ such that $(x, m) = 1$ is denoted by G_m.

We write $m = \prod_{j=1}^{r} p_j^{\alpha_j}$, where the p_j are distinct primes and the α_j are positive integers. Following the usual conventions, we take void products to be 1 and void sums to be 0.

In 2.6–2.11 of [2], the structure of finite commutative semigroups is discussed. In § 2, we work out this structure for S_m. In § 3, we give a construction based on [2], 3.2 and 3.3, for all of the semicharacters of S_m. In § 4, we prove that if χ is a semicharacter of S_m assuming a value different from 0 and 1, then $\sum_{[x]\in S_m} \chi([x]) = 0$. In § 5, we compute $\chi([x])$ explicitly in terms of the integer x, for an arbitrary semicharacter χ of S_m. In § 6, we discuss the structure of the semigroup of all semicharacters of S_m.

Our interest in S_m arose from seeing the interesting paper [4] of Parízek and Schwarz. Some of their results appear in somewhat different form in § 2. Other writers ([1], [5], [6], [7]) have also dealt with S_m from various points of view. In particular, a number of the results of § 2 appear in [6] and in more detail in [7]. We have also benefitted from conversations with R. S. Pierce.

2. The structure of S_m. Let G be any finite commutative semigroup, and let a denote an idempotent of G. The sets $T_a = \{x : x \in G, x^m = a$ for some positive integer $m\}$ are pairwise disjoint subsemigroups of G whose union is G. The set $U_a = \{x : x \in T_a, x^l = x$ for some positive integer $l\}$ is a subgroup of G and is the largest subgroup of G that contains a. For a complete discussion, see [2], 2.6-2.11. In the present section, we identify the idempotents a of S_m and the sets T_a and U_a. We first prove a lemma.

2.1 LEMMA. *Let x be any non-zero integer, written in the form*

$$\prod_{j=1}^{r} p_j^{\beta_j}\cdot a, \qquad \beta_j \geqq 0\ ,\ (a, m) = 1\ .$$

Received January 9, 1960. The authors gratefully acknowledge financial support from the National Science Foundation, under Grant NSF—G 5439.

Then there is an integer c prime to m such that

$$x \equiv \prod_{j=1}^{r} p_j^{\lambda_j} \cdot c \pmod{m},$$

where $\lambda_j = \min(\alpha_j, \beta_j)$ $(j = 1, \cdots, r)$. *If*

$$x \equiv \prod_{j=1}^{r} p_j^{\mu_j} \cdot d \pmod{m},$$

where $0 \leqq \mu_j \leqq \alpha_j$ $(j=1, \cdots, r)$ *and* $(d, m) = 1$, *then* $\mu_j = \lambda_j$ $(j=1, \cdots, r)$. *However, it may happen that* $d \not\equiv c \pmod{m}$.

Proof. Let $b = \prod_{\substack{j \\ \alpha_j = \beta_j}} p_j$. Then we have

$$\begin{aligned} x + bm &= p_1^{\beta_1} \cdots p_r^{\beta_r} a + p_1^{\alpha_1} \cdots p_r^{\alpha_r} b \\ &= \prod_{j=1}^{r} p_j^{\min(\alpha_j, \beta_j)} \cdot (Aa + B), \end{aligned}$$

where

$$A = \prod_{j=1}^{r} p_j^{\max(0, (\beta_j - \alpha_j))}$$

and

$$B = \prod_{j=1}^{r} p_j^{\max(0, (\alpha_j - \beta_j))} \cdot b .$$

Then it is easy to see that $(Aa + B, m) = 1$, so that

$$x \equiv \prod_{j=1}^{r} p_j^{\min(\alpha_j, \beta_j)} \cdot c \pmod{m},$$

where $c = Aa + B$ is prime to m. The last two statements of the lemma are also easily checked.

2.2 THEOREM. *Consider the* 2^r *sequences* $\{\delta_1, \cdots, \delta_r\}$, *where* $\delta_j = 0$ *or* α_j $(j = 1, \cdots, r)$. *Corresponding to each such sequence, there is exactly one idempotent of the semigroup* S_m, *and different sequences give different idempotents. The idempotent corresponding to* $\{\delta_1, \cdots, \delta_r\}$ *can be written as*

$$\left[\prod_{j=1}^{r} p_j^{\delta_j} \cdot d\right],$$

where d is any solution of the congruence

$$\prod_{j=1}^{r} p_j^{\delta_j} \cdot d \equiv 1 \left(\operatorname{mod} \prod_{j=1}^{r} p_j^{\alpha_j - \delta_j}\right).$$

Proof. An element $[x]$ of S_m is idempotent if and only if $x^2 \equiv x \pmod{m}$. If x is written as in 2.1, this congruence becomes $\prod_{j=1}^r p_j^{2\lambda_j} \cdot c^2 \equiv \prod_{j=1}^r p_j^{\lambda_j} c \pmod{m}$, which is equivalent to

$$(1) \qquad \prod_{j=1}^{r} p_j^{\lambda_j} \cdot c \equiv 1 \left(\bmod \prod_{j=1}^{r} p_j^{\alpha_j - \lambda_j} \right).$$

The congruence (1) has a solution c if and only if $\prod_{j=1}^r p_j^{\lambda_j}$ is relatively prime to $\prod_{j=1}^r p_j^{\alpha_j-\lambda_j}$, that is, if and only if $\lambda_j = 0$ or α_j $(j = 1, \cdots, r)$. If c_0 is a solution of (1), then all solutions of (1) are given by

$$c = c_0 + y \prod_{j=1}^{r} p_j^{\alpha_j - \lambda_j},$$

where y is an integer. Plainly

$$\left[\prod_{j=1}^{r} p_j^{\lambda_j} c \right] = \left[\prod_{j=1}^{r} p_j^{\lambda_j} c_0 \right]$$

for all such c.

We have thus proved the existence of a unique idempotent

$$\left[\prod_{j=1}^{r} p_j^{\delta_j} \cdot d \right]$$

corresponding to a sequence $\{\delta_1, \cdots, \delta_r\}$, where $\delta_j = 0$ or α_j $(j = 1, \cdots, r)$. If $\{\delta_1, \cdots, \delta_r\}$ and $\{\delta_1', \cdots, \delta_r'\}$ are distinct such sequences, the corresponding idempotents are distinct by 2.1.

2.21 COROLLARY. *Let*

$$\left[\prod_{j=1}^{r} p_j^{\delta_j} \cdot d \right]$$

and

$$\left[\prod_{j=1}^{r} p_j^{\delta_j'} \cdot d' \right]$$

be idempotents in S_m, *written as in* 2.2. *Then their product is the idempotent*

$$\left[\prod_{j=1}^{r} p_j^{\max(\delta_j, \delta_j')} \cdot d'' \right],$$

as in Theorem 2.2.

This follows directly from 2.1 and the obvious fact that products of idempotents are idempotent.

We next determine the sets T_a and U_a defined above.

2.3 THEOREM. *Let*

$$[x] = \left[\prod_{j=1}^{r} p_j^{\lambda_j} c\right]$$

be any element of S_m, *where* $0 \leqq \lambda_j \leqq \alpha_j$ $(j = 1, \cdots, r)$ *and* $(c, m) = 1$. *Then* $[x] \in T_a$, *where the idempotent*

$$a = \left[\prod_{\substack{1 \leqq j \leqq r \\ \lambda_j > 0}} p_j^{\alpha_j} \cdot d\right],$$

and d *is as in* 2.2.

Proof. The idempotent a such that $[x] \in T_a$ has the property that $[x]^{nk} = a$ for some positive integer k and all integers $n \geqq$ some fixed positive integer n_0 (see [2], 2.6.2). For $n = n_0 \cdot \max(\alpha_1, \cdots, \alpha_r)$, 2.1 implies that

$$a = [x]^{nk} = [x^{nk}] = \left[\prod_{j=1}^{r} p_j^{nk\lambda_j} \cdot c^{nk}\right] = \left[\prod_{j=1}^{r} p_j^{\min(nk\lambda_j, \alpha_j)} \cdot d'\right] = \left[\prod_{j=1}^{r} p_j^{\delta_j} \cdot d\right],$$

where $\delta_j = 0$ if $\lambda_j = 0$ and $\delta_j = \alpha_j$ if $\lambda_j > 0$, and d' and d are relatively prime to m.

2.4 THEOREM. *Let*

$$a = \left[\prod_{j=1}^{r} p_j^{\delta_j} \cdot d\right]$$

be any idempotent of S_m, *written as in* 2.2. *The group* U_a *consists of all elements of* S_m *of the form*

$$\left[\prod_{j=1}^{r} p_j^{\delta_j} \cdot c\right]$$

where $(c, m) = 1$.

Proof. Let $[x] \in U_a$. Then for some integers $l > 1$ and $k \geqq 1$ and all integers $n \geqq n_0$, we have $[x]^l = [x]$ and $[x]^{nk} = a$. This implies that $[x] = [x]^{nk+l}$. Writing x as in 2.1 and using 2.1, we now have

$$\prod_{j=1}^{r} p_j^{\lambda_j} \cdot c \equiv \prod_{j=1}^{r} p_j^{\lambda_j(nk+l)} c^{nk+l} \equiv \prod_{\substack{1 \leqq j \leqq r \\ \lambda_j > 0}} p_j^{\alpha_j} \cdot h \pmod{m},$$

provided that n is sufficiently large; here $(h, m) = 1$. From 2.1 we infer that $\lambda_j = 0$ or α_j $(j = 1, \cdots, r)$. Since $[x] \in U_a \subset T_a$, 2.3 now implies that $\lambda_j = \delta_j$ $(j = 1, \cdots, r)$.

Now let $x = \prod_{j=1}^{r} p_j^{\delta_j} \cdot c$, where $(c, m) = 1$. Then 2.3 shows that $[x] \in T_a$. To prove that $[x] \in U_a$, we need to find an integer $l > 1$ such that $[x]^l = [x]$. This is equivalent to finding an l such that

$$\left(\prod_{j=1}^{r} p_j^{\delta_j}\cdot c\right)^l \equiv \prod_{j=1}^{r} p_j^{\delta_j}\cdot c \pmod{m}\ ,$$

and this congruence is equivalent to the congruence

$$\left(\prod_{j=1}^{r} p_j^{\delta_j}\cdot c\right)^{l-1} \equiv 1 \left(\bmod\ \prod_{j=1}^{r} p_j^{\alpha_j-\delta_j}\right).$$

Since

$$\prod_{j=1}^{r} p_j^{\delta_j}\cdot c$$

is relatively prime to the modulus, such an l exists.

We now identify the groups U_a.

2.5 THEOREM. *Let*

$$a = \left[\prod_{j=1}^{r} p_j^{\delta_j}\cdot d\right]$$

be any idempotent of S_m, *written as in* 2.2. *Let*

$$A = \prod_{j=1}^{r} p_j^{\alpha_j-\delta_j}\ .$$

The group U_a *is isomorphic to the group* G_A.

Proof. For every integer x, let $[x]'$ be the residue class modulo A to which x belongs. For $[x]\in S_m$, let $\tau([x]) = [x]'$. Plainly τ is single-valued and is a homomorphism of S_m onto S_A. We need only show that τ is one-to-one on U_a. If $(c, m) = (c^*, m) = 1$ and

$$\tau\left(\left[\prod_{j=1}^{r} p_j^{\delta_j}\cdot c\right]\right) = \tau\left(\left[\prod_{j=1}^{r} p_j^{\delta_j}\cdot c^*\right]\right),$$

then

$$\prod_{j=1}^{r} p_j^{\delta_j}\cdot c \equiv \prod_{j=1}^{r} p_j^{\delta_j}\cdot c^* \pmod{A}\ ,$$

which implies that $c \equiv c^* \pmod{A}$, because $\left(\prod_{j=1}^{r} p_j^{\delta_j}, A\right) = 1$. Since $\prod_{j=1}^{r} p_j^{\delta_j}\cdot A = m$, we can multiply the last congruence by $\prod_{j=1}^{r} p_j^{\delta_j}$ to obtain

$$\prod_{j=1}^{r} p_j^{\delta_j}\cdot c \equiv \prod_{j=1}^{r} p_j^{\delta_j}\cdot c^* \pmod{m}\ .$$

3. A construction of the semicharacters of S_m. A semicharacter of S_m is a complex-valued multiplicative function defined on S_m that is not identically zero. The set X_m of all semicharacters of S_m forms a semigroup under pointwise multiplication, since [1] is the unit of S_m

and $\chi([1]) = 1$ for all $\chi \in X_m$. In this section, we apply the construction of [2], 3.2 and 3.3, to obtain the semicharacters of S_m. In § 5, we will give a second construction of the semicharacters of S_m, more explicit than the present one, and independent of [2]. This construction will enable us to identify X_m as a semigroup (§ 6).

Theorems 3.2 and 3.3 of [2] give a description of all semicharacters of S_m in terms of the groups U_a. Let χ_a be any character of the group U_a. We extend χ_a to a function on all of S_m in the following way:

$$(1)\quad \chi([x]) = \begin{cases} 0 \text{ if } ab \neq a \text{ for the idempotent } b \text{ such that } [x] \in T_b; \\ \chi_a([x]a) \text{ if } ab = a \text{ for the idempotent } b \text{ such that } [x] \in T_b. \end{cases}$$

The set of all such functions χ is the set X_m.

3.1 THEOREM. *The semigroup X_m has exactly*

$$\prod_{j=1}^{r} (1 + p_j^{\alpha_j} - p_j^{\alpha_j - 1})$$

elements.

Proof. For each idempotent $a = [p_1^{\delta_1} \cdots p_r^{\delta_r} c]$ as in 2.2, (1) yields as many distinct semicharacters of S_m as there are characters of the group U_a. The group U_a has just as many characters as elements. By 2.5, U_a consists of

$$\varphi\left(\prod_{j=1}^{r} p_j^{\alpha_j - \delta_j}\right) = \prod_{\substack{1 \leq j \leq r \\ \delta_j = 0}} \{p_j^{\alpha_j - 1}(p_j - 1)\}$$

elements. Also, distinct idempotents a and b of S_m yield distinct semicharacters of S_m under the definition (1). Therefore the number of elements in X_m is

$$(2)\qquad \sum_{\delta} \varphi\left(\prod_{j=1}^{r} p_j^{\alpha_j - \delta_j}\right) = \sum_{\delta} \varphi\left(\prod_{\substack{1 \leq j \leq r \\ \delta_j = 0}} p_j^{\alpha_j}\right) = \sum_{\delta} \left(\prod_{\substack{1 \leq j \leq r \\ \delta_j = 0}} \varphi(p_j^{\alpha_j})\right)$$
$$= \prod_{j=1}^{r} (1 + \varphi(p_j^{\alpha_j})) = \prod_{j=1}^{r} (1 + p_j^{\alpha_j} - p_j^{\alpha_j - 1}) .$$

The sums in (2) are taken over all sequences $\{\delta_1, \cdots, \delta_r\}$ where each δ_j is 0 or α_j.

3.2 THEOREM. *Let χ be a semicharacter of S_m as given in* (1) *with the idempotent $a = [p_1^{\delta_1} \cdots p_r^{\delta_r} d]$, and let χ' be a semicharacter with the idempotent $a = [p_1^{\delta_1'} \cdots p_r^{\delta_r'} d']$. Then the semicharacter $\chi\chi'$ is given by* (1) *with the idempotent $a'' = [p_1^{\min(\delta_1, \delta_1')} \cdots p_r^{\min(\delta_r, \delta_r')} d]$.*

This theorem follows at once from 2.21 and the definition (1).

We now prove two facts needed in § 4.

3.3 THEOREM. *Let χ be a semicharacter of S_m that assumes somewhere a value different from* 0 *and* 1. *Then χ assumes a value different from* 1 *somewhere on G_m.*

Proof. Definition (1) implies that the character χ_a of U_a assumes a value different from 1. It is also easy to see that $G_m = U_{[1]}$. For $[x] \in G_m$, definition (1) implies that $\chi([x]) = \chi_a(a[x])$. We need therefore only show that the mapping $[x] \to a[x]$ carries G_m onto U_a.

Write $a = [p_1^{\delta_1} \cdots p_r^{\delta_r} d]$. Every element of U_a can be written as $[p_1^{\delta_1} \cdots p_r^{\delta_r} c]$ where $(c, m) = 1$, by 2.4. We must produce an $[x] \in G_m$ such that $a[x] = [p_1^{\delta_1} \cdots p_r^{\delta_r} c]$. That is, we must produce an integer x such that

$$\prod_{j=1}^{r} p_j^{\delta_j} \cdot dx \equiv \prod_{j=1}^{r} p_j^{\delta_j} \cdot c \pmod{m} \tag{3}$$

and $(x, m) = 1$. The congruence (3) is equivalent to

$$dx \equiv c \left(\operatorname{mod} \prod_{j=1}^{r} p_j^{\alpha_j - \delta_j} \right). \tag{4}$$

Since d is relatively prime to the modulus in (4), the congruence (4) has a solution x_0. We determine x as a number

$$x_0 + l \prod_{j=1}^{r} p_j^{\alpha_j - \delta_j} ,$$

where l is an integer for which

$$x_0 + l \prod_{j=1}^{r} p_j^{\alpha_j - \delta_j} \equiv 1 \left(\operatorname{mod} \prod_{j=1}^{r} p_j^{\delta_j} \right).$$

Clearly

$$x = x_0 + l \prod_{j=1}^{r} p_j^{\alpha_j - \delta_j}$$

satisfies (3) and the condition $(x, m) = 1$.

3.4. Let $\{\lambda_1, \cdots, \lambda_r\}$ be a sequence of integers such that $0 \leqq \lambda_j \leqq \alpha_j$ $(j = 1, \cdots, r)$, and consider the set $V(\lambda_1, \cdots, \lambda_r)$ of all $[p_1^{\lambda_1} \cdots p_r^{\lambda_r} x] \in S_m$ with $(x, m) = 1$. It is easy to see that this set is contained in T_a, where a is the idempotent

$$\left[\prod_{\substack{1 \leqq j \leqq r \\ \lambda_j > 0}} p_j^{\alpha_j} \cdot d \right].$$

3.5 THEOREM. *Given $\lambda_1, \cdots, \lambda_r$, there is a positive integer k such that the mapping $[x] \to [p_1^{\lambda_1} \cdots p_r^{\lambda_r} x]$ of G_m onto $V(\lambda_1, \cdots, \lambda_r)$ is exactly k to one.*

Proof. Let u be any integer such that $(u, m) = 1$, and let $[x_1]$, $\cdots$, $[x_{k_u}]$ be the distinct elements of G_m such that $[p_1^{\lambda_1} \cdots p_r^{\lambda_r} x_j] = [p_1^{\lambda_1} \cdots p_r^{\lambda_r} u]$. That is,

$$p_1^{\lambda_1} \cdots p_r^{\lambda_r} x_j \equiv p_1^{\lambda_1} \cdots p_r^{\lambda_r} u \pmod{m} \quad (j = 1, \cdots, k_u) .$$

Let u^* be any solution of $uu^* \equiv 1 \pmod{m}$. If $(v, m) = 1$, then we have

$$p_1^{\lambda_1} \cdots p_r^{\lambda_r} u^* v x_j \equiv p_1^{\lambda_1} \cdots p_r^{\lambda_r} v \pmod{m} .$$

Since $(u^* v x_j, m) = 1$ $(j = 1, \cdots, k_u)$ and the elements $[u^* v x_1], \cdots, [u^* v x_{k_u}]$ are distinct in G_m, it follows that $k_u \leqq k_v$. Similarly, we have $k_v \leqq k_u$.

4. A property of semicharacters of S_m. It is well known and obvious that if H is a finite group and χ is a character of H, then $\sum_{x \in H} \chi(x) = 0$ or $o(H)$ according as $\chi \neq 1$ or $\chi = 1$. This result does not hold in general for finite commutative semigroups. As a simple example, consider the cyclic finite semigroup $T = \{x, x^2, \cdots, x^l, \cdots, x^{l+k-1}\}$, where $x^{l+k} = x^l$, and l and $l + k$ are the first pair of positive integers $m, n, m < n$, for which $x^m = x^n$. The following facts are easy to show, and follow from the general theory in [2]. The subset $\{x^l, x^{l+1}, \cdots, x^{l+k-1}\}$ is the largest subgroup of T. Its unit is the element x^{uk}, where the integer u is defined by $l \leqq uk < l + k$. The general semicharacter of T is the function χ whose value at x^h is $\exp(2\pi i h j/k)$, where $j = 0$, $1, \cdots, k - 1$. For $j = 1, 2, \cdots, k - 1$, the sum $\sum_{h=1}^{k+l-1} \chi(x^h)$ is equal to

$$\frac{1 - \exp\left(\dfrac{2\pi i(k + l)j}{k}\right)}{1 - \exp\left(\dfrac{2\pi i j}{k}\right)} ,$$

which is 0 if and only if $k/(k, l)$ divides j. Hence the sum of a semicharacter assuming values different from 0 and 1 need not be 0.

Curiously enough, the above-mentioned property of groups holds for the semigroup S_m.

4.1 THEOREM. *Let χ be a semicharacter of S_m that assumes somewhere a value different from 0 and 1. Then* $\sum_{[x] \in S_m} \chi([x]) = 0$.

Proof. It is obvious from 2.1 that the sets $V(\lambda_1, \cdots, \lambda_r)$ of 3.4 are pairwise disjoint and that their union is S_m. We therefore need only show that $\sum_{[x] \in V(\lambda_1, \cdots, \lambda_r)} \chi([x]) = 0$ for all $\{\lambda_1, \cdots, \lambda_r\}$. By 3.3, χ assumes a value different from 1 somewhere on the group G_m, so that $\sum_{[x] \in G_m} \chi([x]) = 0$. (Note that χ on G_m is a character of the group G_m.) Thus we have $0 = \sum_{[x] \in G_m} \chi([p_1^{\lambda_1} \cdots p_r^{\lambda_r}])\chi([x]) = \sum_{[x] \in G_m} \chi([p_1^{\lambda_1} \cdots p_r^{\lambda_r} x]) = k \sum \chi([y])$, where $[y]$ runs through $V(\lambda_1, \cdots, \lambda_r)$.

5. A second construction of semicharacters of S_m. In this section, we compute explicitly all of the semicharacters of S_m. The case m even is a little different from the case m odd. When m is even, we will take $p_1 = 2$. To compute the semicharacters of S_m, we need to examine the structure of S_m in more detail than was done in § 3. For this purpose, we fix once and for all the following numbers.

5.1 DEFINITION. *For $j = 1, \cdots, r$, let*
$g_j =$ *a primitive root modulo* $p_j^{\alpha_j}$ *if* p_j *is odd*;
$g_1 = 5$ *if* $p_1 = 2$;
$h_j = g_j + y_j p_j^{\alpha_j}$ *where* y_j *is such that* $h_j \equiv 1 \pmod{m/p_j^{\alpha_j}}$;
$h_0 = -1 + y_0 p_1^{\alpha_1}$ *where* y_0 *is such that* $h_0 \equiv 1 \pmod{m/p_1^{\alpha_1}}$;
$q_j = p_j + z_j p_j^{\alpha_j}$ *where* z_j *is such that* $q_j \equiv 1 \pmod{m/p_j^{\alpha_j}}$;
For $j = 1, \cdots, r$, $l = 1, \cdots, r$, $j \neq l$, and p_l odd, let k_{jl} be a positive integer such that $p_j \equiv g_l^{k_{jl}} \pmod{p_l^{\alpha_l}}$.
For $j = 2, \cdots, r$ and $p_1 = 2$ let
k_{j1} *be a positive integer such that* $p_j \equiv (-1)^{(p_j-1)/2} g_1^{k_{j1}} \pmod{p_1^{\alpha_1}}$.

Plainly $y_0, y_1, \cdots, y_r$ and $z_1, \cdots, z_r$ exist. For p_l odd, the integers k_{jl} exist because g_l is a primitive root modulo $p_l^{\alpha_l}$. For $p_1 = 2$, the integers k_{j1} exist for $\alpha_1 \geqq 3$ by [3], p. 82, Satz 126. For $\alpha_1 = 1$ or 2, k_{j1} can be any positive integer.

5.2. Let x be any integer $\neq 0$. Then $x = \prod_{j=1}^{r} p_j^{\beta_j(x)} \cdot a(x)$, where $\beta_j(x) \geqq 0$ and $(a(x), m) = 1$. Plainly the numbers $\beta_j = \beta_j(x)$ and $a = a(x)$ are uniquely determined by x. For $j = 1, \cdots, r$ and p_j odd, let $e_j = e_j(x)$ be any positive integer such that

$$a(x) \equiv g_j^{e_j(x)} \pmod{p_j^{\alpha_j}} .$$

The number $e_j(x)$ is uniquely determined modulo $\varphi(p_j^{\alpha_j})$. For $p_1 = 2$, let

$e_1 = e_1(x)$ be any positive integer such that

$$a(x) \equiv (-1)^{(a(x)-1)/2} g_1^{e_1(x)} \pmod{p_1^{\alpha_1}} .$$

For $\alpha_1 \geqq 3$, $e_1(x)$ exists and is uniquely determined modulo $p_1^{\alpha_1 - 2}$ (see [3], p. 82, Satz 126). For $\alpha_1 = 1$ or 2, $e_1(x)$ can be any positive integer.

If m is even, let

$$(1_e) \qquad A(x) = \left(\prod_{j=2}^{r} h_0^{(p_j-1)\beta_j/2}\right)\left(\prod_{l=1}^{r} \prod_{\substack{j=1 \\ j \neq 1}}^{r} h_l^{\beta_j k_{jl}}\right)\left(\prod_{j=1}^{r} q_j^{\beta_j}\right) h_0^{(a-1)/2} \left(\prod_{j=1}^{r} h_j^{e_j}\right) .$$

If m is odd, let

$$(1_o) \qquad A(x) = \left(\prod_{l=1}^{r} \prod_{\substack{j=1 \\ j \neq 1}}^{r} h_l^{\beta_j k_{jl}}\right)\left(\prod_{j=1}^{r} q_j^{\beta_j}\right)\left(\prod_{j=1}^{r} h_j^{e_j}\right) .$$

If m is even, it is easy to see from 5.1 that

$$(2)\qquad A(x) \equiv \left(\prod_{j=2}^{r} (-1)^{(p_j-1)\beta_j/2}\right)\left(\prod_{j=2}^{r} g_1^{\beta_j k_{j1}}\right) p_1^{\beta_1}(-1)^{(a-1)/2} g_1^{e_1} \pmod{p_1^{\alpha_1}}$$
$$\equiv \left(\prod_{j=2}^{r} (-1)^{(p_j-1)/2} g_1^{k_{j1}}\right)^{\beta_j} p_1^{\beta_1}(-1)^{(a-1)/2} g_1^{e_1}$$
$$\equiv \prod_{j=2}^{r} p_j^{\beta_j} \cdot p_1^{\beta_1} a \equiv x \pmod{p_1^{\alpha_1}},$$

and, if $n = 2, \cdots, r$,

$$A(x) \equiv \prod_{\substack{j=1\\ j\neq n}}^{r} g_n^{\beta_j k_{jn}} \cdot p_n^{\beta_n} g_n^{e_n} \equiv \prod_{\substack{j=1\\ j\neq n}}^{r} p_j^{\beta_j} \cdot p_n^{\beta_n} a \equiv x \pmod{p_n^{\alpha_n}}.$$

Therefore $A(x) \equiv x \pmod{m}$ if m is even.

If m is odd, then for $n = 1, \cdots, r$, we have

$$A(x) \equiv \prod_{\substack{j=1\\ j\neq n}}^{r} g_n^{\beta_j k_{jn}} \cdot p_n^{\beta_n} g_n^{e_n} \equiv \prod_{\substack{j=1\\ j\neq n}}^{r} p_j^{\beta_j} \cdot p_n^{\beta_n} a \equiv x \pmod{p_n^{\alpha_n}}.$$

Therefore $A(x) \equiv x \pmod{m}$ if m is even or odd.

5.3. Suppose that χ is any semicharacter of S_m. Let ψ be the function defined for all integers x by the relation $\psi(x) = \chi([x])$. Then ψ is obviously a semicharacter of the integers under multiplication, and $\psi(x) = \psi(y)$ if $x \equiv y \pmod{m}$. We will construct the semicharacters of S_m by finding all of the functions ψ with these properties. As 5.2 shows, ψ is determined by its values on $h_0, h_1, \cdots, h_r$ and $q_1, \cdots, q_r$. We now set down relations involving the h's and q's which restrict the values that ψ can assume on these integers.

5.4. If p_j is odd, then

$$h_j^{\varphi(p_j^{\alpha_j})} \equiv 1 \pmod{p_j^{\alpha_j}}, \quad h_j^{\varphi(p_j^{\alpha_j})} \equiv 1 \left(\bmod\ \frac{m}{p_j^{\alpha_j}}\right);$$

hence

$$h_j^{\varphi(p_j^{\alpha_j})} \equiv 1 \pmod{m}.$$

Also,

$$h_0^2 \equiv 1 \pmod{p_1^{\alpha_1}}, \qquad h_0^2 \equiv 1 \left(\bmod\ \frac{m}{p_1^{\alpha_1}}\right);$$

hence $h_0^2 \equiv 1 \pmod{m}$.

If $p_1 = 2$ and $\alpha_1 = 1$, then $h_0 \equiv 1 \pmod{2}$, $h_0 \equiv 1 \pmod{m/2}$; hence $h_0 \equiv 1 \pmod{m}$.

If $p_1 = 2$ and $\alpha_1 = 1$ or 2, then
$h_1 \equiv 5 \equiv 1 \pmod{p_1^{\alpha_1}}$, $h_1 \equiv 1 \pmod{m/p_1^{\alpha_1}}$; hence $h_1 \equiv 1 \pmod{m}$.
If $p_1 = 2$ and $\alpha_1 \geqq 3$, then
$h_1^{2^{\alpha_1-2}} \equiv 1 \pmod{p_1^{\alpha_1}}$, $h_1^{2^{\alpha_1-2}} \equiv 1 \pmod{m/p_1^{\alpha_1}}$; hence $h_1^{2^{\alpha_1-2}} \equiv 1 \pmod{m}$.
(The first congruence on the line above is proved in [3], p. 81, Satz 125.)

For $j = 1, \cdots, r$, we have

$$q_j^{\alpha_j} \equiv 0\,, \qquad q_j^{\alpha_j}h_j \equiv 0\,, \qquad q_j^{\alpha_j+1} \equiv 0 \pmod{p_j^{\alpha_j}}\,,$$

$$q_j^{\alpha_j} \equiv 1\,, \qquad q_j^{\alpha_j}h_j \equiv 1\,, \qquad q_j^{\alpha_j+1} \equiv 1 \left(\bmod \frac{m}{p_j^{\alpha_j}}\right).$$

Therefore we have

$$q_j^{\alpha_j} \equiv q_j^{\alpha_j}h_j \equiv q_j^{\alpha_j+1} \pmod{m}\,.$$

Also, if $p_1 = 2$, we have

$$q_1^{\alpha_1} \equiv 0\,, \qquad q_1^{\alpha_1}h_0 \equiv 0 \pmod{p_1^{\alpha_1}}\,,$$

$$q_1^{\alpha_1} \equiv 1\,, \qquad q_1^{\alpha_1}h_0 \equiv 1 \left(\bmod \frac{m}{p_1^{\alpha_1}}\right).$$

Therefore we have

$$q_1^{\alpha_1} \equiv q_1^{\alpha_1}h_0 \pmod{m}\,.$$

5.5 If ψ is to be a function on the integers such that $\psi(x) = \chi([x])$ for some semicharacter χ of S_m, then the choices of the values of ψ at the h's and q's are restricted by the congruences modulo m derived in 5.4. Thus, since $\chi([1]) = 1$, we have

$\psi(h_j)^{\varphi(p_j^{\alpha_j})} = 1$ if p_j is odd;
$\psi(h_0) = \pm 1$, and $\psi(h_0) = 1$ if $\alpha_1 = 1$ and $p_1 = 2$;
$\psi(h_1) = 1$ if $p_1 = 2$ and $\alpha_1 = 1$ or 2;
$\psi(h_1)^{2^{\alpha_1-2}} = 1$ if $p_1 = 2$ and $\alpha_1 \geqq 3$.

Also we have

$$\psi(q_j)^{\alpha_j} = \psi(q_j)^{\alpha_j}\psi(h_j) = \psi(q_j)^{\alpha_j+1} \text{ for } j = 1, \cdots, r\,.$$

If $p_1 = 2$, we have

$$\psi(q_1)^{\alpha_1} = \psi(q_1)^{\alpha_1}\psi(h_0)\,.$$

The last two equalities give us:

$$\psi(q_j) \neq 0 \text{ implies } \psi(h_j) = \psi(q_j) = 1;$$

and

$$\psi(q_1) \neq 0 \text{ implies } \psi(h_0) = 1 \text{ if } p_1 = 2.$$

5.6. To construct our functions ψ, we now choose numbers ω_0, $\omega_1, \cdots, \omega_r$ and $\mu_1, \cdots, \mu_r$ which are to be $\psi(h_0)$, $\psi(h_1), \cdots, \psi(h_r)$ and $\psi(q_1), \cdots, \psi(q_r)$. The relations in 5.5 show that we must take these numbers such that:

$\omega_j^{\varphi(p_j^{\alpha_j})} = 1$ if $j = 1, \cdots, r$ and p_j is odd;
$\omega_0 = \pm 1$; $\omega_0 = 1$ if $p_1 = 2$ and $\alpha_1 = 1$, or if m is odd[1];
$\omega_1 = 1$ if $p_1 = 2$ and $\alpha_1 = 1$ or 2;
$\omega_1^{2^{\alpha_1-2}} = 1$ if $p_1 = 2$ and $\alpha_1 \geqq 3$;
$\mu_j = 0$ or 1 if $j = 1, \cdots, r$;
$\omega_j = 1$ if $\mu_j = 1, j = 1, \cdots, r$;
$\omega_0 = 1$ if $p_1 = 2$ and $\mu_1 = 1$.

Formulas (1_e) and (1_0) of 5.2 now require us to define $\psi(x)$ for non-zero integers x as follows:

$$(3_e) \quad \psi(x) = \left(\prod_{j=2}^{r} \omega_0^{(p_j-1)\beta_j(x)/2}\right)\left(\prod_{l=1}^{r}\prod_{\substack{j=1\\j\neq l}}^{r} \omega_l^{\beta_j(x)k_{jl}}\right)\left(\prod_{j=1}^{r} \mu_j^{\beta_j(x)}\right) \cdot \omega_0^{(a(x)-1)/2}\left(\prod_{j=1}^{r}\omega_j^{e_j(x)}\right) \text{ if } m \text{ is even}^2;$$

$$(3_0) \quad \psi(x) = \left(\prod_{l=1}^{r}\prod_{\substack{j=1\\j\neq l}}^{r} \omega_l^{\beta_j(x)k_{jl}}\right)\left(\prod_{j=1}^{r} \mu_j^{\beta_j(x)}\right)\left(\prod_{j=1}^{r}\omega_j^{e_j(x)}\right) \text{ if } m \text{ is odd.}$$

Finally, we define $\psi(0) = \psi(m)$.

The q's, h's, and k's appearing in (1) and (3) were fixed once and for all in terms of m. The ω's and μ's are at our disposal and serve to define ψ. The β's are determined uniquely from x; but the e's are not. As noted in 5.2, e_j is determined modulo $\varphi(p_j^{\alpha_j})$ if p_j is odd, and e_1 is determined modulo $p_1^{\alpha_1-2}$ if $p_1 \doteq 2$ and $\alpha_1 \geqq 3$. Since $\omega_j^{\varphi(p_j^{\alpha_j})} = 1$ if p_j is odd, $\omega_1^{2^{\alpha_1-2}} = 1$ if $p_1 = 2$ and $\alpha_1 \geqq 3$, and $\omega_1 = 1$ if $p_1 = 2$ and $\alpha_1 \leqq 2$, we see that ψ is uniquely defined by the formulas (3_e) and (3_0).

5.7. We now prove that $\psi(xy) = \psi(x)\psi(y)$. Since ψ is obviously bounded and not identically zero, this will show that ψ is a semicharacter.

Suppose first that $x \neq 0, y \neq 0$. Then we have

$$x = \prod_{j=1}^{r} p_j^{\beta_j(x)}\cdot a(x)\,, \quad y = \prod_{j=1}^{r} p_j^{\beta_j(y)}\cdot a(y)\,, \quad xy = \prod_{j=1}^{r} p_j^{\beta_j(x)+\beta_j(y)}\cdot a(x)a(y)\,.$$

[1] We take $\omega_0 = 1$ when m is odd merely as a matter of convenience. Actually, as will shortly be apparent, ω_0 does not appear in the definition of ψ if m is odd.

[2] We take $0^0 = 1$.

Therefore $a(xy) = a(x)a(y)$ and $\beta_j(xy) = \beta_j(x) + \beta_j(y)$ for $j = 1, \cdots, r$. Also we have

$$g_j^{e_j(xy)} \equiv a(xy) \equiv a(x)a(y) \equiv g_j^{e_j(x)} g_j^{e_j(y)} \equiv g_j^{e_j(x)+e_j(y)} \pmod{p_j^{\alpha_j}}$$

if p_j is odd. Since g_j is a primitive root modulo $p_j^{\alpha_j}$ and $\omega_j^{\varphi(p_j^{\alpha_j})} = 1$, it follows that $e_j(xy) \equiv e_j(x) + e_j(y) \pmod{\varphi(p_j^{\alpha_j})}$ and $\omega_j^{e_j(xy)} = \omega_j^{e_j(x)}\omega_j^{e_j(y)}$ if p_j is odd $(j = 1, \cdots, r)$. If $p_1 = 2$, then $a(x)$ and $a(y)$ are odd, and plainly

$$\frac{a(xy) - 1}{2} \equiv \frac{a(x) - 1}{2} + \frac{a(y) - 1}{2} \pmod{2} .$$

Therefore we have

$$\omega_0^{(a(xy)-1)/2} = \omega_0^{(a(x)-1)/2}\omega_0^{(a(y)-1)/2}$$

for both admissible values of ω_0. Furthermore,

$$\begin{aligned}(-1)^{(a(xy)-1)/2} g_1^{e_1(xy)} &\equiv a(x)a(y) \\ &\equiv (-1)^{(a(x)-1)/2} g_1^{e_1(x)} (-1)^{(a(y)-1)/2} g_1^{e_1(y)} \pmod{p_1^{\alpha_1}} ,\end{aligned}$$

if $p_1 = 2$. Therefore we have

$$g_1^{e_1(xy)} \equiv g_1^{e_1(x)+e_1(y)} \pmod{p_1^{\alpha_1}} ,$$

if $p_1 = 2$.
Hence, if $\alpha_1 \geqq 3$ and $p_1 = 2$, we have $e_1(xy) \equiv e_1(x) + e_1(y) \pmod{p_1^{\alpha_1-2}}$, as follows from [3], p. 82, Satz 126 (recall that $g_1 = 5, p_1 = 2$). Hence

$$\omega_1^{e_1(xy)} = \omega_1^{e_1(x)}\omega_1^{e_1(y)} \qquad \text{if } \alpha_1 \geqq 3,\ p_1 = 2 .$$

The last equality also holds if $\alpha_1 \leqq 2$ and $p_1 = 2$, since $\omega_1 = 1$ in this case.

The foregoing computations, together with (3), now show that $\psi(xy) = \psi(x)\psi(y)$ if $xy \neq 0$.

We next show that $\psi(xy) = \psi(x)\psi(y)$ if $xy = 0$. We compute $\psi(m)$. Since $\beta_j(m) = \alpha_j > 0$ for $j = 1, \cdots, r$, we have

$$\prod_{j=1}^{r} \mu_j^{\beta_j(m)} = \begin{cases} 1 \text{ if } \mu_1 = \cdots = \mu_r = 1 , \\ 0 \text{ otherwise.} \end{cases}$$

If $\mu_1 = \cdots = \mu_r = 1$, then by 5.6, we have $\omega_0 = \omega_1 = \cdots = \omega_r = 1$, so that $\psi(x) = 1$ for all x. In this case, we have $\psi(xy) = \psi(x)\psi(y)$ for all x and y. If some $\mu_j = 0$, then $\psi(m) = 0$, and hence $\psi(0) = 0$. In this case, $\psi(xy) = \psi(x)\psi(y)$ if $xy = 0$.

5.8. We now prove that $\psi(x) = \psi(y)$ if $x \equiv y \pmod{m}$. Suppose first that $xy \neq 0$ and $x \equiv y \pmod{m}$. Then

$$\prod_{j=1}^{r} p_j^{\beta_j(x)} \cdot a(x) \equiv \prod_{j=1}^{r} p_j^{\beta_j(y)} \cdot a(y) \pmod{m} .$$

From this, we see that $\beta_j(x) > 0$ if and only if $\beta_j(y) > 0$. If, for some j, we have $\beta_j(x) > 0$ and $\mu_j = 0$, then $\beta_j(y) > 0$ and $\psi(x) = 0 = \psi(y)$.

Now we can suppose that $\mu_j = 1$ for all j such that $\beta_j(x) > 0$. Then $\omega_j = 1$ if $\beta_j(x) > 0$ ($j = 1, \cdots, r$) and $\omega_0 = 1$ if $\beta_1(x) > 0$. If m is odd, or if m is even and $\beta_1(x) > 0$, we have

$$(4) \qquad \psi(x) = \Bigg(\prod_{\substack{l=1 \\ \beta_l(x)=0}}^{r} \prod_{\substack{j=1 \\ j\neq l}}^{r} \omega_l^{\beta_j(x)k_{jl}} \Bigg) \Bigg(\prod_{\substack{j=1 \\ \beta_j(x)=0}}^{r} \omega_j^{e_j(x)} \Bigg) ,$$

$$(5) \qquad \psi(y) = \Bigg(\prod_{\substack{l=1 \\ \beta_l(x)=0}}^{r} \prod_{\substack{j=1 \\ j\neq l}}^{r} \omega_l^{\beta_j(y)k_{jl}} \Bigg) \Bigg(\prod_{\substack{j=1 \\ \beta_j(x)=0}}^{r} \omega_j^{e_j(y)} \Bigg) .$$

If m is even and $\beta_1(x) = 0$, we have

$$(6) \quad \psi(x) = \Bigg(\prod_{j=2}^{r} \omega_0^{(p_j-1)\beta_j(x)/2} \Bigg) \Bigg(\prod_{\substack{l=1 \\ \beta_l(x)=0}}^{r} \prod_{\substack{j=1 \\ \beta_j(x)>0}}^{r} \omega_l^{\beta_j(x)k_{jl}} \Bigg) \omega_0^{(a(x)-1)/2} \Bigg(\prod_{\substack{j=1 \\ \beta_j(x)=0}}^{r} \omega_j^{e_j(x)} \Bigg) ,$$

$$(7) \quad \psi(y) = \Bigg(\prod_{j=2}^{r} \omega_0^{(p_j-1)\beta_j(y)/2} \Bigg) \Bigg(\prod_{\substack{l=1 \\ \beta_l(x)=0}}^{r} \prod_{\substack{j=1 \\ \beta_j(x)>0}}^{r} \omega_l^{\beta_j(y)k_{jl}} \Bigg) \omega_0^{(a(y)-1)/2} \Bigg(\prod_{\substack{j=1 \\ \beta_j(x)=0}}^{r} \omega_j^{e_j(y)} \Bigg) .$$

Since $x \equiv y \pmod{m}$, we see from 5.2 that $A(x) \equiv A(y) \pmod{m}$ and hence

$$(8) \qquad A(x) \equiv A(y) \pmod{p_n^{\alpha_n}} \text{ for } n = 1, \cdots, r .$$

The congruence

$$(9) \qquad A(x) \equiv \prod_{\substack{j=1 \\ j\neq n}}^{r} h_n^{\beta_j(x)k_{jn}} \cdot q_n^{\beta_n(x)} h_n^{e_n(x)} \pmod{p_n^{\alpha_n}}$$

holds if p_n is odd. To verify this, use (1_e) and (1_0) together with 5.1. Notice that for $n = 1$, we use only (1_0).

The congruences (8) and (9), together with the fact that $\beta_n(x) = 0$ if and only if $\beta_n(y) = 0$, now show that

$$\prod_{\substack{j=1 \\ j\neq n}}^{r} h_n^{\beta_j(x)k_{jl}} \cdot h_n^{e_n(x)} \equiv \prod_{\substack{j=1 \\ j\neq n}}^{r} h_n^{\beta_j(y)k_{jn}} \cdot h_n^{e_n(y)} \pmod{p_n^{\alpha_n}}$$

if p_n is odd and $\beta_n(x) = 0$. This implies that

$$\sum_{\substack{j=1 \\ j\neq n}}^{r} \beta_j(x)k_{jn} + e_n(x) \equiv \sum_{\substack{j=1 \\ j\neq n}}^{r} \beta_j(y)k_{jn} + e_n(y) \pmod{\varphi(p_n^{\alpha_n})} ,$$

and

$$(10) \qquad \prod_{\substack{j=1 \\ j\neq n}}^{r} \omega_n^{\beta_j(x)k_{jn}} \cdot \omega_n^{e_n(y)} = \prod_{\substack{j=1 \\ j\neq n}}^{r} \omega_n^{\beta_j(y)k_{jn}} \cdot \omega_n^{e_n(y)} ,$$

if p_n is odd and $\beta_n(x) = 0$.

Similarly, if $p_1 = 2$ and $\beta_1(x) = 0$, in which case $g_1 = 5$, (2) implies that

$$A(x) \equiv \left(\prod_{j=2}^{r} (-1)^{(p_j-1)\beta_j(x)/2}\right)\left(\prod_{j=2}^{r} 5^{\beta_j(x)k_{j1}}\right)(-1)^{(a(x)-1)/2}5^{e_1(x)} \pmod{2^{\alpha_1}} . \tag{11}$$

The congruences (8) and (11), together with the fact that $\beta_1(y) = 0$, now show that

$$(-1)^{\sum_{j=2}^{r}\frac{1}{2}(p_j-1)\beta_j(x)+\frac{1}{2}(a(x)-1)} 5^{\sum_{j=2}^{r}\beta_j(x)k_{j1}+e_1(x)} \equiv$$
$$\equiv (-1)^{\sum_{j=2}^{r}\frac{1}{2}(p_j-1)\beta_j(y)+\frac{1}{2}(a(y)-1)} 5^{\sum_{j=2}^{r}\beta_j(y)+e_1(y)} \pmod{2^{\alpha_1}}$$

From this congruence, we find that

$$\sum_{j=2}^{r}\frac{1}{2}(p_j-1)\beta_j(x)+\frac{1}{2}(a(x)-1) \equiv$$
$$\sum_{j=2}^{r}\frac{1}{2}(p_j-1)\beta_j(y)+\frac{1}{2}(a(y)-1) \pmod 2$$

if $\alpha_1 \geqq 2$, and

$$\sum_{j=2}^{r}\beta_j(x)k_{j1}+e_1(x) \equiv \sum_{j=2}^{r}\beta_j(y)k_{j1}+e_1(y) \pmod{2^{\alpha_1-2}}$$

if $\alpha_1 \geqq 3$. Since $\omega_0 = 1$ if $\alpha_1 = 1$ and $\omega_1 = 1$ if $\alpha_1 = 1$ or 2, we now have

$$\prod_{j=2}^{r}\omega_0{}^{(p_j-1)\beta_j(x)/2}\cdot\omega_0{}^{(a(x)-1)/2} = \prod_{j=2}^{r}\omega_0{}^{(p_j-1)\beta_j(y)/2}\cdot\omega_0{}^{(a(y)-1)/2} \tag{12}$$

if $\alpha_1 \geqq 1$, and

$$\prod_{j=2}^{r}\omega_1^{\beta_j(x)k_{j1}}\cdot\omega_1^{e_1(x)} = \prod_{j=2}^{r}\omega_1^{\beta_j(y)k_{j1}}\cdot\omega_1^{e_1(y)} \tag{13}$$

if $\alpha_1 \geqq 1$. Multiplying (10) over the relevant values of n, we have

$$\left(\prod_{\substack{n=1\\ \beta_n(x)=0\\ p_n>2}}^{r}\prod_{\substack{j=1\\ j\neq n}}^{r}\omega_n^{\beta_j(x)k_{jn}}\right)\left(\prod_{\substack{n=1\\ \beta_n(x)=0\\ p_n>2}}^{r}\omega_n^{e_n(x)}\right) = \left(\prod_{\substack{n=1\\ \beta_n(x)=0\\ p_n>2}}^{r}\prod_{\substack{j=1\\ j\neq n}}^{r}\omega_n^{\beta_j(y)k_{jn}}\right)\left(\prod_{\substack{n=1\\ \beta_n(x)=0\\ p_n>2}}^{r}\omega_n^{e_n(y)}\right) . \tag{14}$$

If m is odd, or if m is even and $\beta_1(x) > 0$, (14), (4), and (5) show that $\psi(x) = \psi(y)$. If m is even and $\beta_1(x) = 0$, we multiply (12), (13), and (14) together. Comparing the result with (6) and (7), we find that $\psi(x) = \psi(y)$ in this case also.

We have therefore proved that $\psi(x) = \psi(y)$ if $x \equiv y \pmod m$ and $xy \neq 0$. If $x \equiv 0 \pmod m$ and $x \neq 0$, then $\psi(x) = \psi(m)$. Since $\psi(0) = \psi(m)$ by definition, the proof is complete.

5.9. The foregoing construction of the functions ψ, and from these the semicharacters χ of S_m, $\chi([x]) = \psi(x)$, clearly gives us all of the semicharacters of S_m. As the ω's and μ's of 5.6 run through all admissible values, each semicharacter χ appears exactly once. We could show this by exhibiting, for each pair ψ and ψ', a number x such that $\psi(x) \neq \psi'(x)$. Rather than do this, we prefer to count the ψ's and compare their number with the number obtained in 3.1.

For p_j odd, the number of possible values of ω_j is $\varphi(p_j^{\alpha_j})$ if $\mu_j = 0$ and 1 if $\mu_j = 1$. Hence this number is $\varphi(p_j^{\alpha_j(1-\mu_j)})$. For $p_1 = 2$, there are several cases to consider ($\mu_1 = 0$ or 1, $\alpha_1 = 1$, $\alpha_1 = 2$, $\alpha_1 \geqq 3$). In each case, it is easy to see that the number of admissible pairs $\{\omega_0, \omega_1\}$ is $\varphi(2^{\alpha_1(1-\mu_1)})$. Thus, for each sequence $\{\mu_1, \cdots, \mu_r\}$, the total number of sequences $\{\omega_0, \omega_1, \cdots, \omega_r\}$ is equal to

$$\prod_{j=1}^{r} \varphi(p_j^{\alpha_j(1-\mu_j)}) .$$

Summing this number over all possible $\{\mu_1, \cdots, \mu_r\}$, we obtain $\prod_{j=1}^{r} (1 + p_j^{\alpha_j} - p_j^{\alpha_j - 1})$, as in Theorem 3.1.

6. The structure of X_m.

6.1. Let χ and χ' be any semicharacters of S_m, and let $(\mu_1, \cdots, \mu_r; \omega_0, \omega_1, \cdots, \omega_r)$ and $(\mu_1', \cdots, \mu_r'; \omega_0', \omega_1', \cdots, \omega_r')$ be the parameters as in 5.6 that determine χ and χ', respectively. The product $\chi\chi'$ then has as its parameters

$$(1) \qquad (\mu_1\mu_1', \cdots, \mu_r\mu_r'; \ \omega_0\omega_0', \omega_1\omega_1', \cdots, \omega_r\omega_r') .$$

Thus, all of the χ's in X_m for which the μ's are a fixed sequence of 0's and 1's form a group, plainly the direct product of cyclic groups, one corresponding to each zero value of μ. These are maximal subgroups of X_m, and X_m is the union of these subgroups. The multiplication rule (1) shows clearly how elements of different subgroups are multiplied. The rule (1) shows also that X_m resembles a direct product of groups and $\{0, 1\}$ semigroups. It fails to be one because of the condition in 5.6 that $\mu_j = 1$ implies $\omega_j = 1$.

6.2. The characters modulo m of number theory (see [3], p. 83) are of course among the semicharacters that we have computed. They are exactly those for which $\mu_1 = \mu_2 = \cdots = \mu_r = 0$. In the description of § 3, they are the semicharacters that are characters on the group G_m and are 0 elsewhere on S_m.

6.3. We can also map X_m into S_m, and represent X_m as a subset of S_m with a new definition of multiplication. Let χ be in X_m and let

χ have parameters $(\mu_1, \cdots, \mu_r; \omega_0, \omega_1, \cdots, \omega_r)$. For m odd and $j = 0, 1, \cdots, r$ or m even and $j = 0, 2, 3, \cdots, r$, let w_j be any integer such that $\omega_j = \exp(2\pi i w_j/\varphi(p_j^{\alpha_j}))$. For m even and $\alpha_1 = 1$ or 2, let $w_1 = 0$; for m even and $\alpha_1 \geqq 3$, let w_1 be any integer such that $\omega_1 = \exp(2\pi i w_1/2^{\alpha_1-2})$.

We now define the mapping

$$\chi \rightarrow \tau(\chi) = \left[h_0^{w_0(1-\mu_1)} \prod_{j=1}^{r} (h_j^{w_j(1-\mu_j)} q_j^{\alpha_j\mu_j})\right], \tag{2}$$

which carries X_m into S_m. Evidently τ is single-valued.

6.4 THEOREM. *The mapping τ is one-to-one.*

Proof. Suppose that χ and χ' are semicharacters of S_m with parameters as in 6.1. Suppose that $\tau(\chi) = \tau(\chi')$, that is,

$$h_0^{w_0(1-\mu_1)} \prod_{j=1}^{r} (h_j^{w_j(1-\mu_j)} q_j^{\alpha_j\mu_j}) \equiv h_0^{w_0'(1-\mu_1')} \prod_{j=1}^{r} (h_j^{w_j'(1-\mu_j')} q_j^{\alpha_j\mu_j'}) \pmod{m} . \tag{3}$$

This congruence, along with 5.1, implies that

$$h_l^{w_l(1-\mu_l)} p_l^{\alpha_l\mu_l} \equiv h_l^{w_l'(1-\mu_l')} p_l^{\alpha_l\mu_l'} \pmod{p_l^{\alpha_l}}$$

for $l = 1, \cdots, r$ and p_l odd. Since $(h_l, p_l) = 1$, and μ_l and μ_l' are 0 or 1, it is obvious that $\mu_l = \mu_l'$. If $\mu_l = \mu_l' = 1$, then from 5.6, we have $\omega_l = \omega_l' = 1$. If $\mu_l = \mu_l' = 0$, then $h_l^{w_l} \equiv h_l^{w_l'} \pmod{p_l^{\alpha_l}}$, so that $w_l \equiv w_l' \pmod{\varphi(p_l^{\alpha_l})}$ and hence $\omega_l = \omega_l'$.

If $p_1 = 2$, (2) implies that

$$h_0^{w_0(1-\mu_1)} h_1^{w_1(1-\mu_1)} p_1^{\alpha_1\mu_1} \equiv h_0^{w_0'(1-\mu_1')} h_1^{w_1'(1-\mu_1')} p_1^{\alpha_1\mu_1'} \pmod{p_1^{\alpha_1}} . \tag{4}$$

Again, we have $\mu_1 = \mu_1'$. If $\mu_1 = \mu_1' = 1$, then 5.6 states that $\omega_0 = \omega_0' = \omega_1 = \omega_1' = 1$. If $\alpha_1 = 1$, then $\omega_0 = \omega_0' = 1$, also by 5.6. If $\alpha_1 = 2$ and $\mu_1 = \mu_1' = 0$, then (3), along with 5.1, shows that $(-1)^{w_0} \equiv (-1)^{w_0'} \pmod{4}$, and hence $\omega_0 = \omega_0'$. If $\alpha_1 \geqq 3$ and $\mu_1 = \mu_1' = 0$, then we have $(-1)^{w_0}5^{w_1} \equiv (-1)^{w_0'}5^{w_1'} \pmod{2^{\alpha_1}}$. Once again, [3], p. 82, Satz 126 shows that $(-1)^{w_0} = (-1)^{w_0'}$ and that $w_1 \equiv w_1' \pmod{2^{\alpha_1-2}}$. Hence $\omega_0 = \omega_0'$ and $\omega_1 = \omega_1'$. Therefore τ is one-to-one.

6.5. The set $\tau(X_m)$ consists of all the elements $[p_1^{\delta_1} \cdots p_r^{\delta_r} a]$ of S_m for which $\delta_j = 0$ or α_j, and $(a, m) = 1$. It is evident from (2) that $\tau(X_m)$ is contained in the set $\{[p_1^{\delta_1} \cdots p_r^{\delta_r} a]\}$. The reverse inclusion is established by a routine examination of cases, which we omit.

6.6. The mapping τ plainly defines a new multiplication in $\tau(X_m)$: $\tau(\chi)^*\tau(\chi') = \tau(\chi')$. Every residue class $\tau(\chi)$ contains a number

$$x = h_0^{w_0(1-\mu_1)} \prod_{j=1}^{r} (h_j^{w_j(1-\mu_j)} q_j^{\alpha_j\mu_j}) .$$

If x' is another number of this form, then it can be shown that $[x]^*[x']$ is equal to $[xx'/\prod q_j^{\alpha_j}]$, where the product $\prod q_j^{\alpha_j}$ is taken over all j, $j = 1, \cdots, r$, for which $p_j \mid xx'$. We omit the details.

LITERATURE

1. Eckford Cohen, *A finite analogue of the Goldbach problem*, Proc. Amer. Math. Soc. **5** (1954), 478–483.

2. Edwin Hewitt and H. S. Zuckerman, *Finite dimensional convolution algebras*, Acta Math. **93** (1955), 67–119.

3. Edmund Landau, *Vorlesungen über Zahlentheorie*, Band I. S. Hirzel Verlag, Leipzig, 1927.

4. B. Parízek, and Š. Schwarz, *O multiplikatívnej pologrupe zvyškových tried* (mod *m*), Mat.-Fyz. Časopis Slov. Akad. Ved **8** (1958), 136–150.

5. E. T. Parker, *On multiplicative semigroups of residue classes*, Proc. Amer. Math. Soc. **5** (1954) 612–616.

6. H. S. Vandiver and Milo W. Weaver, *Introduction to arithmetic factorization and congruences from the standpoint of abstract algebra*, Herbert Ellsworth Slaught Memorial Papers, no. 7, 1958. Math. Assoc. of America.

7. Milo W. Weaver, *Cosets in a semi-group*, Math. Mag. **25** (1952), 125–136.

THE UNIVERSITY OF WASHINGTON

RELATION OF A DIRECT LIMIT GROUP TO ASSOCIATED VECTOR GROUPS

PAUL D. HILL

A set M with a binary, transitive relation $<$ is said to be directed if for each pair a, b in M, there is a c in M such that $a < c, b < c$. Let $\{G_a\}_{a \in M}$ be a collection of groups indexed by a directed set $M = \{a, b, \cdots ; <\}$, and for each $a < b$ in M let h_b^a be a homomorphism of G_a into G_b. The homomorphisms are assumed to satisfy the relations

(i) $h_c^b h_b^a = h_c^a$ if $a < b < c$

and

(ii) if $a < a$, then h_a^a is the identity.

We call such a system a direct system of groups and define a direct limit group of this system in the following manner. Two elements $g_a \in G_a$ and $\bar{g}_b \in g_b$ are said to be equivalent if there is a $c > a, b$ such that $h_c^a(g_a) = h_c^b(\bar{g}_b)$. Let g_a^* denote the collection of elements which are equivalent to g_a. Now given any two equivalence classes g_a^* and $\bar{g}_b^*$, there exists a c and elements $g_c, \bar{g}_c$ in G_c such that $g_a^* = g_c^*$ and $\bar{g}_b^* = \bar{g}_c^*$. We define $g_a^* \cdot \bar{g}_b^* = (g_c \bar{g}_c)^*$. This multiplication is a well defined binary operation on the set, G^*, of equivalence classes. And it may be shown that G^* is a (multiplicative) group, which we define to be the direct limit group of the given system.

Let $G = \prod G_a$ be the restricted direct product of the given groups G_a, and consider the groups G_a as subgroups of G. An element in G of the form $g_a^{-1} h_b^a(g_a)$ is called a relation. Let H be the subgroup generated by the relations of G. Note that the inverse of a relation is a relation. By a "last" element of M we mean an element b such that $a < b$ for all a in M. If M contains no last element, it is immediate that given $a_1, a_2, \cdots, a_k$ in M, there exists a $b \in M$ with the property $a_i < b, a_i \neq b$ for $i = 1, 2, \cdots, k$.

LEMMA 1. *If M contains no last element, the commutator group K of G is contained in H.*

Proof. Let $x = g_{a_1} g_{a_2} \cdots g_{a_k}$ and $y = \bar{g}_{b_1} \bar{g}_{b_2} \cdots \bar{g}_{b_j}$ be arbitrary elements of G, where $a_m = a_n$ or $b_m = b_n$ implies that $m = n$. First choose a with the property that $a_i < a, a_i \neq a$, and $b_i \neq a$ for all i. Then choose b such that $b_i < b, b_i \neq b, a_i \neq b$, and $a \neq b$. We have

$$xyx^{-1}y^{-1} = \prod_{i=1}^{k} g_{a_i} \prod_{i=1}^{j} \bar{g}_{b_i} \prod_{i=k}^{1} g_{a_i}^{-1} \prod_{i=j}^{1} \bar{g}_{b_i}^{-1}$$

Received September 8, 1959, and in revised form September 11, 1959.

$$= \prod_{i=1}^{k} g_{a_i} \prod_{i=1}^{k} h_a^{a_i}(g_{a_i}^{-1}) \prod_{i=1}^{j} \bar{g}_{b_i} \prod_{i=1}^{j} h_b^{b_i}(\bar{g}_{b_i}^{-1})$$
$$= \prod_{i=k}^{1} g_{a_i}^{-1} \prod_{i=k}^{1} h_a^{a_i}(g_{a_i}) \prod_{i=j}^{1} \bar{g}_{b_i}^{-1} \prod_{i=j}^{1} h_b^{b_i}(\bar{g}_{b_i})$$
$$= \prod_{i=1}^{k} g_{a_i} h_a^{a_i}(g_{a_i}^{-1}) \prod_{i=1}^{i} \bar{g}_{b_i} h_b^{b_i}(\bar{g}_{b_i}^{-1}) \prod_{i=k}^{1} g_{a_i}^{-1} h_a^{a_i}(g_{a_i}) \prod_{i=j}^{1} \bar{g}_{b_i}^{-1}(\bar{g}_{b_i}) .$$

Thus $xyx^{-1}y^{-1} \in H$. Since H is a group, the lemma follows.

COROLLARY 1. *If M contains no last element, then H is a normal subgroup of G.*

The following example shows that H may not be normal in G if M contains a last element.

EXAMPLE. Suppose that M is $\{1, 2; \leqq\}$. Let G_2 be the symmetric group on the set $\{1, 2, 3\}$, and let G_1 be the subgroup of G_2 of those elements fixing 3. Define h_2^1 to be the identity isomorphism of G_1 into G_2. Then H, a cyclic group of order 2, is generated by $((1, 2), (1, 2))$. It is, therefore, not normal in G.

LEMMA 2. *If g_a in G_a is in H, then there exists a b such that $h_b^a(g_a) \in K_b$, the commutator group of G_b.*

Proof. In general, if x_a in G_a is the product $x_{a_1}x_{a_2} \cdots x_{a_n}$ where $x_{a_i} \in G_{a_i}$ and if $b > a, a_i$ for $i = 1, 2, \cdots, n$, then $h_b^a(x_a)$ can be written as the product of the elements $h_b^{a_1}(x_{a_1}), h_b^{a_2}(x_{a_2}), \cdots, h_b^{a_n}(x_{a_n})$ in some order. This fact is easily proved by induction on n. If $n > 1$, by the induction hypothesis we may as well assume that the factors x_{a_i} are nontrivial. Thus two of the factors must be contained in a single group G_{a_i}. And the product $x_{a_1}x_{a_2} \cdots x_{a_n}$ can be contracted to a product of the same form with one less factor by taking one of the new factors to be the product of two of the old and letting the other factors remain unchanged (except, possibly, for the order in which they appear).

Since g_a is in H, it can be written in the form $\prod_{i=1}^{k} g_{a_i}^{-1} h_{b_i}^{a_i}(g_{a_i})$. Choose b such that $b > a, b_i$ for $i = 1, 2, \cdots, k$. Then

$$h_b^a(g_a)K_b = \prod_{i=1}^{k} h_b^{a_i}(g_{a_i}^{-1}) h_b^{b_i} h_{b_i}^{a_i}(g_{a_i}) K_b$$
$$= \prod_{i=1}^{k} h_b^{a_i}(g_{a_i}^{-1}) h_b^{a_i}(g_{a_i}) K_b = 1_b K_b = K_b ,$$

which proves the lemma.

THEOREM 1. *If H is a normal subgroup of G, then G/H is a homomorphism image of G^*, where the kernel of the homomorphism is contained in the commutator subgroup, K^**

REMARKS. The theorem is well known [1] in case the groups G_a are abelian. In this case H is necessarily normal and $K^* = 1$ is the identity. Thus $G^* \cong G/H$, and we have two equivalent definitions for the direct limit.

Proof of theorem. Let f be the mapping of G^* into G/H, defined by: $g_a^* \to g_a H$. In order to show that f is single-valued, let $g_a^* = \bar{g}_b^*$. There exists $c > a, b$ such that $h_c^a(g_a) = h_c^b(\bar{g}_b)$. Thus $g_a^{-1}\bar{g}_b = g_a^{-1}h_c^a(g_a)h_c^b(\bar{g}_b)^{-1}\bar{g}_b$. Since $h_c^b(\bar{g}_b)^{-1}\bar{g}_b = \bar{g}_b h_c^b(\bar{g}_b^{-1})$, we have $g_a^{-1}\bar{g}_b \in H$, which implies that f is independent of the representative of g_a^*. The multiplicative property of f is immediate. We next show that f is onto. Let $gH \in G/H$ and let $g = g_{a_1}g_{a_2}\cdots g_{a_k}$, where the a_i's are distinct. Choose b such that $a_i < b$ for $i = 1, 2, \cdots, k$. If $a_i = b$ for some i, we may as well assume that $i = k$ since the g_{a_i}'s commute. For each i, $g_{a_i}^{-1}h_b^{a_i}(g_{a_i}) \in H$. Thus

$$\bar{g} = \prod_{i=k}^{1} g_{a_i}^{-1}h_b^{a_i}(g_{a_i}) = \prod_{i=k}^{1} g_{a_i}^{-1} \prod_{i=k}^{1} h_b^{a_i}(g_{a_i})$$

is in H, which implies that $g\bar{g}H = gH$. But $g\bar{g} = \prod_{i=k}^{1} h_b^{a_i}(g_{a_i})$ is in G_b. Hence $f((g\bar{g})^*) = gH$, and f is onto. Since $g_a \in K_a$, the commutator group of G_a, implies that $g_a^* \in K^*$, it follows from Lemma 2 that the kernel of f is contained in K^*.

THEOREM 2. *If M contains no last element, then $G^*/K^* \cong G/H$.*

Proof. By Corollary 1, H is normal in G. Thus by Theorem 1, we need only show that the kernel of f is the whole commutator group, K^*. However, if g_a^* is a commutator of G^*, then there exist a b and a commutator $\bar{g}_b \in K_b$ such that $g_a^* = \bar{g}_b^*$. Since $K_b \subseteqq K$, by Lemma 1 $\bar{g}_b \in H$. Thus $f(g_a^*) = H$, and the theorem folllows.

The limit group G^* is abelian if and only if for every a in M and for every commutator g_a of the group G_a, there exists a $b > a$ (depending on g_a) such that $h_b^a(g_a) = 1_b$. Also, under this condition the commutator subgroups, K_a, of the groups G_a are contained in H, and H is normal in G since the conjugate of a generator of H transformed by a general element of G

$$\begin{aligned}\{x_c\}_{c\in M}g_a^{-1}h_b^a(g_a)\{x_c\}_{c\in M}^{-1} &= x_a x_b g_a^{-1}h_b^a(g_a)x_a^{-1}x_b^{-1}\\ &= x_a g_a^{-1}x_a^{-1}g_a\cdot g_a^{-1}h_b^a(g_a)\cdot h_b^a(g_a)^{-1}x_b h_b^a(g_a)x_b^{-1}\end{aligned}$$

remains in H.

COROLLARY 2. *If the limit group G^* is abelian, then $G^* \cong G/H$. Moreover, the converse holds if M contains no last element.*

A directed set $M = \{a, b, \cdots; <\}$ is said to be completely directed if for every a in M all but a finite number of b's in M satisfy the relation $a < b$. In particular, the positive integers are completely directed by $<$.

Letting $G' = \prod' G_a$ be the complete direct product of the given groups G_a, we have

LEMMA 3. *If M is completely directed and has no last element, then G^* is contained (in the sense of isomorphism) in the factor group G'/G.*

Proof. Define a mapping h of G^* into G'/G by: $g_a^* \to \{x_b\}_{b \in M} G$, where $x_a = g_a$ and $a < b$ implies that $x_b = h_b^a(g_a)$. The coordinate x_b may be chosen as an arbitrary element of G_b if b fails to satisfy $a \leqq b$. It may be shown that h is a homomorphism with trivial kernel, which proves the lemma.

Letting F be the inverse image of $h(G^*)$ under the natural homomorphism of G' onto G'/G, we observe

COROLLARY 3. *Let M satisfy the conditions of Lemma 3, and let G^* be abelian. Then in the chain*

$$G' \supsetneqq F \supsetneqq G \supsetneqq H \supsetneqq 1$$

we have $F/G \cong G/H \cong G^*$.

REFERENCE

1. S. Lefschetz, *Algebraic Topology*, New York, American Mathematical Society Colloquium Publications, 1942.

AUBURN UNIVERSITY

COMMUTATOR GROUPS OF MONOMIAL GROUPS

C. V. HOLMES

This paper is a study of the commutator groups of certain generalized permutation groups called complete monomial groups. In [2] Ore has shown that every element of the infinite permutation group is itsself a commutator of this group. Here it is shown that every element of the infinite complete monomial group is the product of at most two commutators of the infinite complete monomial group. The commutator subgroup of the infinite complete monomial group is itself, as is the case in the infinite symmetric group, [2]. The derived series is determined for a wide class of monomial groups.

Let H be an arbitrary group, and S a set of order B, $B \geqq d, d=\aleph_0$. Then one obtains a monomial group after the manner described in [1]. A monomial substitution over H is a linear transformation mapping each element x of S in a one-to-one manner onto some element of S multiplied by an element h of H, the multiplication being formal. The element h is termed a factor of the substitution. If substitution u maps x_i into $h_j x_j$, while substitution v maps x_j into $h_t x_t$, then the substitution uv maps x_i into $h_j h_t x_t$. A substitution all of whose factor are the identity e of H is called a permutation and the set of all permutations is a subgroup which is isomorphic to the symmetric group on B objects. A substitution which maps each element of S into itself multiplied by an element of H is called a multiplication. The set of all multiplications form a subgroup which is the strong direct product of groups H_α, each H_α isomorphic to H. Hereafter monomial substitutions which are permutations will be denoted by s, while those that are multiplications will be denoted by v. The monomial group whose elements are the monomial substitutions, restricted by the definitions of C and D as given below, will be denoted by $\sum(H; B, C, D)$, where the symbols in the name are to be interpreted as follows, H the given arbitrary group, B the order of the given set S, C a cardinal number such that the number of non-identity factors of any substitution of the group is less than C, D a cardinal number such that the number of elements of S being mapped into elements of S distinct from themselves by any substitution of the group is less than D. In the event $C = D = B^+, B^+$ the successor of B, the resulting monomial group is termed the complete monamial group generated by the given group H and the given set S. $S(B, M)$, $d \leqq M \leqq D$, will denote the subgroup of permutations which map fewer than M elements of S onto elements of S distinct from themselves, while $V(B, N)$,

Received October 25, 1958, and in revised form August 27, 1959.

$d \leqq N \leqq C$, will denote the subgroup of multiplications which have fewer than N nonidentity factors. In particular $S(B, d)$ denotes the subgroup of finite permutations and $V(B, d)$ the subgroup consisting of those multiplications which have finitely many nonidentity factors. The concept of alternating as associated with permutation groups may be extended in an obvious manner to monomial groups. $A(B, d)$ will denote the alternating subgroup of the permutation group $S(B, d)$, while $\sum_A(H; B, d, d)$ will denote the alternating subgroup of the monomial group $\sum(H; B, d, d)$. Any substitution may be written as the product of a multiplication and a permutation. Hence we may write $\sum(H; B, C, D) = V(B, C) \cup S(B, D)$, where $\cup$ here and throughout will mean group generated by the set. G' will be used to denote the commutator subgroup of the group G.

THEOREM 1. *The commutator subgroup* $V'(B, C), d \leqq C \leqq B^+$, *of* $V(B, C)$ *is the set of all elements*

$$v' = (h'_1, h'_2, h'_3, \cdots), h'_i \in H' ,$$

where there exists an integer N *such that each* h'_i *is the product of* N *or fewer commutators of* H.

Proof. The theorem follows from the fact that $V(B, C)$ is the strong direct product, each of whose summands is isomorphic to H, together with the remark following the lemma page 308 of [2].

THEOREM 2. *The commutator subgroup* $S'(B, C), d < C \leqq B^+$, *of* $S(B, C)$ *is* $S(B, C)$. *The commutator subgroup* $S'(B, d)$ *of* $S(B, d)$ *is* $A(B, d)$.

The proof is contained in [2].

THEOREM 3. *The commutator subgroup* $\sum'(H; B, d, d)$ *of* $\sum(H; B, d, d)$ *is* $A(B, d) \cup V^+(B, d)$ *where* $V^+(B, d)$ *is the set of all elements of* $V(B, d)$ *whose product of factors is a member of* H'.

Proof. By reason of Theorem 2 we have
$\sum'(H; B, d, d) \supset A(B, d)$, and that
$\sum'(H; B, d, d) \supset V^+(B, d)$

will now be demonstrated.

If h_i is the only nonidentity factor of the multiplication v_i, then the commutator $v_i s v_i^{-1} s$, where $s = (x_i, x_j)$, is a multiplication whose only nonidentity factors are h_i and h_i^{-1}. It then follows that any multiplication v of $V^+(B, d)$ with n nonidentity factors can be written as the product of $n + 1$ multiplications, n of which are of the type of the commutator described above, and the remaining member having as its only nonidentity factor the product of the factors of v. But the first n members of the product belong to $\sum'(H; B, \mathrm{d}, \mathrm{d})$, while the other member of the product is an element of $V'(B, d)$, by reason of Theorem 1, and hence

$$\sum'(H; B, d, d) \supset V^+(B, d), \text{ since } V'(B, d) \subset \sum'(H; B, d, d).$$

Then

$$\sum'(H; B, d, d) \supset V^+(B, d) \cup A(B, d).$$

Since G/G' is abelian for any group G, and G' is the smallest group for which this is true, to demonstrate that

$$\sum(H; B, d, d)/V^+(B, d) \cup A(B, d)$$

is abelian will imply that

$$\sum'(H; B, d, d) \subset V^+(B, d) \cup A(B, d) ,$$

and the conclusion of the theorem will follow.

That $V^+(B, d) \supset V'(B, d)$ follows from the definition of $V^+(B, d)$, and hence $V(B, d)/V^+(B, d)$ is abelian. Therefore any two multiplications commute mod $V^+(B, d) \cup A(B, d)$. Since $A(B, d)$ consists of all even permutations there are but two cosets of $A(B, d)$ in $S(B, d)$, namely, $A(B, d)$ and $(x_1, x_2)A(B, d)$. Thus any element of the factor group $\sum(H; B, d, d)/V^+(B, d) \cup A(B, d)$ has one of the forms

$$v[V^+(B, d) \cup A(B, d)]$$

or

$$v(x_1, x_2)[V^+(B, d) \cup A(B, d)] , \quad v \in V(B, d) .$$

But $v(x_1, x_2)v^{-1}(x_1, x_2)$ is the commutator $(h_1h_2^{-1}, h_2h_1^{-1}, e, \cdots)$ which belongs to $V^+(B, d)$. That is, (x_1, x_2) and v commute mod $[V^+(B, d) \cup A(B, d)]$, and hence $\sum(H; B, d, d)/V^+(B, d) \cup A(B, d)$ is abelian, which implies $\sum'(H; B, d, d) \subset V^+(B, d) \cup A(B, d)$, and we have

$$\sum'(H; B, d, d) = V^+(B, d) \cup A(B, d) .$$

The following theorem asserts that the derived series for $\sum(H; B, d, d)$ consists of but two distinct terms.

THEOREM 4. *The commutator subgroup* $\sum''(H; B, d, d)$ *of* $\sum'(H; B, d, d)$ *is* $\sum'(H; B, d, d)$.

Proof. $A(B, d) = A'(B, d)$, as was demonstrated in Theorem 7 of [2], and hence $\sum''(H; B, d, d)$ contains $A(B, d)$.

Consider elements v_1 and v_2 of $\sum'(H; B, d, d)$, where the factors of v_1 are all e except the first two and they are inverses of one another, and the factors of v_2 are all e except the first and third and they are inverses of one another. The commutator $v_1v_2v_1^{-1}v_2^{-1}$, which is an element of $\sum''(H; B, d, d)$, has as its first factor a commutator of H and all other factors e. It then follows that any element of $V'(B, d)$ is the product of elements of $\sum''(H; B, d, d)$ and hence is an element of $\sum''(H; B, d, d)$. That is $\sum''(H; B, d, d) \supset V'(B, d)$. Then one can in the manner described in the first part of Theorem 3 write any element v of $V^+(B, d)$ as the product of $n+1$ elements, each member of the product being an element of $\sum''(H; B, d, d)$. That is $\sum''(H; B, d, d)$ contains $V^+(B, d)$, and hence $\sum''(H; B, d, d)$ contains $V^+(B, d) \cup A(B, d) = \sum'(H; B, d, d)$.

THEOREM 5. *The commutator subgroup*

$$\textstyle\sum'_A(H; B, d, d) \text{ of } \sum_A(H; B, d, d) \text{ is } V^+(B, d) \cup A(B, d) .$$

This theorem together with Theorem 3 states that $\sum(H; B, d, d)$ has for its commutator subgroup $\sum_A(H; B, d, d)$. This is the analogue for monomial groups of the result Ore obtains for permutation groups in [2], and as stated in the second part of Theorem 2.

Proof. We have

$$\textstyle\sum'(H; B, d, d) \subset \sum_A(H; B, d, d) \subset \sum(H; B, d, d) ,$$

hence,

$$\textstyle\sum''(H; B, d, d) \subset \sum'_A(H; B, d, d) \subset \sum'(H; B, d, d).$$

Then by reason of Theorem 4,

$$\textstyle\sum'(H; B, d, d) = \sum''(H; B, d, d) = V^+(B, d) \cup A(B, d) .$$

Hence $\sum'_A(H; B, d, d) = V^+(B, d) \cup A(B, d)$.

THEOREM 6. *The commutator subgroup* $\sum'(H; B, C, D)$, $d < C \leqq D \leqq B^+$, *of* $\sum(H; B, C, D)$ *is* $\sum(H; B, C, D)$.

This theorem is also an analogue of a result Ore obtains in [2] for permutation groups as stated in the first part of Theorem 2.

Proof. It is shown in [2] that the commutator subgroup $S'(B, D)$ of $S(B, D)$ is $S(B, D)$. Hence $\sum'(H; B, C, D)$ contains $S(B, D)$. The conclusion of the theorem will then follow if it can be demonstrated that $\sum'(H; B, C, D) \supset V(B, C)$. Let

$$s = (\cdots, x_{-1}, x_0, x_1, \cdots)$$

and

$$v = (\cdots, h_{-1}, h_0, h_1, \cdots)$$

be elements of $\sum(H; B, C, D)$. Then the commutator $svs^{-1}v^{-1}$ an element of $\sum'(H; B, C, D)$ has the form

$$(\cdots, h_0h_{-1}^{-1}, h_1h_0^{-1}, h_2h_1^{-1}, \cdots) .$$

Let

$$v_c = (\cdots, c_{-1}, c_0, c_1, \cdots)$$

be an arbitrary element of $V(B, C)$, and consider the following set of equations.

$$\cdots, h_0h_{-1}^{-1} = c_{-1}, h_1h_0^{-1} = c_0, h_2h_1^{-1} = c_1, \cdots$$

This set of equations has solutions,

$$h_0 = c_{-1}, h_{-1} = e, h_n = c_{n-1}h_{n-1}, h_{-n} = \left[\prod_{i=2}^{n} c_{-i} \right]^{-1} .$$

Then if the factors of v be represented in terms of the factors of v_c as indicated above, we see that

$$svs^{-1}v^{-1} = v_c \in \sum'(H; B, C, D) ,$$

and hence $\sum'(H; B, C, D)$ contains $V(B, C)$, and therefore

$$\sum(H; B, C, D) = \sum'(H; B, C, D) .$$

COROLLARY 1. *Any element u of $\sum(H; B, C, D)$, $d < C \leqq D \leqq B^+$, is the product of at most two commutators.*

Proof. Every element of $S(B, D)$ is a commutator of $S(B, D)$, as was shown in [2]. Every element of $V(B, C)$ is a commutator of $\sum(H; B, C, D)$, as was shown in Theorem 6. Therefore any element of $\sum(H; B, C, D)$ which is either a multiplication or a permutation is a commutator. But every element of $\sum(H; B, C, D)$ maybe written as the product of a multiplication and a permutation and consequently may be written as the product of two commutators.

To see that the assertion that every element of $\sum(H; B, C, D)$ is the product of at most two commutators is the strongest possible, suppose every element of $\sum(H; B, C, D)$ is a commutator. Let

$$u \in \sum(H; B, d, d) \subset \sum(H; B, C, D) .$$

Then $u = u_1u_2u_1^{-1}u_2^{-1}$, u_1 and u_2 elements of $\sum(H; B, C, D)$. But since u belongs to $\sum(H; B, d, d)$ we can choose a u_1^* and u_2^* in $\sum(H; B, d, d)$ by causing u_1 and u_2 to become the map of x_i into ex_i except for those maps which yield the permutation and nonidentity factors of u. It then follows that u is an element of $\sum'(H; B, d, d)$, and hence $\sum(H; B, d, d) = \sum'(H; B, d, d)$. But this is a contradiction to Theorem 3.

References

1. Ore, Oystein, *Theory of monomial groups*, Trans, Amer. Math. Soc. **51** (1942).
2. Ore, Oystein, *Some remarks on commutators*, Proc. Amer. Math. Soc. **2** (1951).

University of Kansas and
San Diego State College

THE NONEXISTENCE OF EXPANSIVE HOMEOMORPHISMS ON A CLOSED 2-CELL

J. F. JAKOBSEN AND W. R. UTZ

1. Introduction. If X is a metric space with metric ρ and $T(X)=X$ is a self-homeomorphism of X, then T is said to be expansive[1] provided there exists a $\delta > 0$ depending only upon X and T such that corresponding to each distinct pair $x, y \in X$ there exists an integer $n(x, y)$ for which $\rho(T^n(x), T^n(y)) > \delta$. W. H. Gottschalk [2] has asked if the n-cell can carry an expansive homeomorphism. B. F. Bryant [1] obtained a partial answer to this question when he essentially showed that there are no expansive self-homeomorphisms of a closed 1-cell, that is, of an arc. In this paper we show that there are no expansive self-homeomorphisms of a closed 2-cell and, in the final section, point out an error in a paper of R. F. Williams. The authors wish to acknowledge the referee's assistance in condensing the paper.

Throughout the paper, X will denote a metric space with metric ρ and $T(X) = X$ will denote a self-homeomorphism of X. The set $0(x)= \bigcup_{n\in I}\{T^n(x)\}$, where I denotes the integers, is called the orbit of x under T. A set $M\subset X$ is said to be minimal under T if, and only if, M is non-vacuous and M is the closure of the orbit of each of its points. If $x, y \in X$, then $0(x)$ and $0(y)$ are said to be positively (negatively) asymptotic if corresponding to $\varepsilon > 0$, there exists an integer N such that

$$\rho(T^n(x), T^n(y)) < \varepsilon \text{ for all } n > N(n < N).$$

If $0(x)$ and $0(y)$ are both positively and negatively asymptotic, then the orbits are said to be doubly asymptotic.

2. Self-homeomorphisms of the 2-cell. In this section we show with the aid of results of van Kampen that there is no expansive self-homeomorphism of a circle, and from this obtain the same result for a simple closed curve and a closed 2-cell.

THEOREM. *If T is a homeomorphism of a closed 2-cell onto itself, then T is not expansive.*

Proof. If there is an expansive homeomorphism, T, of a closed 2-cell onto itself then, since the boundary of the 2-cell is invariant under

Received August 24, 1959, and in revised form March 15, 1960. This research was supported by the United States Air Force through the Air Force Office of Scientific Research and Development Command under Contract No. AF 18 (600) 1108.

[1] In most of the literature cited, the term "unstable" is used in place of "expansive".

T, T must be expansive on the simple closed curve forming the boundary of the 2-cell. Since T is an expansive self-homeomorphism of a simple closed curve, there must be an expansive self-homeomorphism of a circle since it is known [1] that if T is an expansive self-homeomorphism of a metric space X and $g(X) = Y$ is a homeomorphism onto the metric space Y such that g^{-1} is uniformly continuous, then gTg^{-1} is an expansive self-homeomorphism of Y.

Hereafter we assume that T is an expansive self-homeomorphism of a circle, C. We first show that T cannot have a periodic point. If T has at least two distinct periodic points on C, then for some integer m, $T^m = \Phi$ has at least two fixed points on C and it is easy to see that either Φ or Φ^2 leaves an arc invariant. Powers of an expansive homeomorphism are expansive [3] and hence either Φ or Φ^2 is an expansive self-homeomorphism of an arc in violation of the cited result of Bryant.

If T has exactly one periodic point on C, then the point must be fixed under T and the orbit of every other point is doubly asymptotic to the fixed point. There are uncountably many such orbits contrary to the fact that when X is compact and T is an expansive self-homeomorphism of X, then the number of distinct orbits doubly asymptotic to any fixed point is at most countably infinite.

Since we have shown that T has no periodic point on C, C is either a minimal set under T, or [4] there is a minimal set which is a Cantor set and which consists of the common cluster points of orbits. In the first instance T is topologically equivalent to a rotation and is therefore not expansive. In the second instance, a component, A, of the complement of the minimal set is chosen. Now, $T^n(A)$ is an open arc and its diameter goes to zero with increasing or decreasing n. Taking two distinct points of A which are sufficiently close, they remain close for all n by virtue of the continuity of T. This contradicts the hypothesis that T is expansive and the theorem is proved.

3. An example of Williams. R. F. Williams [5] has given two examples of non-degenerate continua and self-homeomorphisms of them which are said to be expansive. One example, where the continuum is the inverse limit space of the unit circle in the complex plane under the bonding map $g(z) = z^2$ and with the shift homeomorphism, is expansive. The other example contains an error which we now explain.

Using the notation of Williams' example, let

$$a = \frac{10^n - 1}{10^n}, \qquad b = \frac{10^n + 1}{10^n}$$

and consider the points

$$x = (a, a/2, a/2^2, a/2^3, \cdots),$$
$$y = (a, b/2, b/2^2, b/2^3, \cdots)$$

for an arbitrary but fixed positive integer n. It is not difficult to see that the maximum value of $\rho(f^j(x), f^j(y))$ occurs for $j = -1$. Since

$$\rho(f^{-1}(x), f^{-1}(y)) = 1/10^n(1 + 1/2^2 + 1/2^4 + \cdots)$$

this maximum can be made arbitrarily small by taking n sufficiently large. Thus the homeomorphism f is not expansive.

The failure of this example suggests seeking another continuous function on $[0, 1]$ such that the shift homeomorphism of the inverse limit space onto itself is expansive. However, such an example is impossible. The authors can prove that the shift homeomorphism on the inverse limit space of any continuous transformation of an arc onto itself cannot be expansive. The proof of the theorem is long and will not be given here.

REFERENCES

1. B. F. Bryant, *Unstable self-homeomorphisms of a compact space*, Vanderbilt University Thesis, 1954.

2. W. H. Gottschalk, *Minimal sets: an introduction to topological dynamics*, Bull. Amer. Math. Soc., **64** (1958), 336-351.

3. W. R. Utz, *Unstable homeomorphisms*, Proc. Amer. Math. Soc., **1** (1950), 769-774.

4. E. R. van Kampen, *The topological transformations of a simple closed curve into itself*, Amer. J. Math., **57** (1935), 142-152.

5. R. F. Williams, *A note on unstable homeomorphisms*, Proc. Amer. Math. Soc., **6** (1955), 308-309.

STATE UNIVERSITY OF IOWA
AND
UNIVERSITY OF MISSOURI

MULTIPLICATION ON CLASSES OF PSUEDO-ANALYTIC FUNCTIONS

JOHN JEWETT

Lipman Bers [1, 2] has formulated a theory of solutions of linear elliptic partial differential equations in terms of classes of psuedo-analytic functions on a plane domain D. The theory for each class of psuedo-analytic functions is based on the notion of a generating pair of Holder continuous complex valued functions F and G defined on D and satisfying $\mathfrak{Im}\,[\overline{F(z)}G(z)] > 0$ in D.

If w is any function defined on D, then there exist two real valued functions ϕ and ψ such that w can be written uniquely as

$$w(z) = \phi(z)F(z) + \psi(z)G(z) \,. \tag{1}$$

A function w defined on D is said to be (F, G)-psuedo-analytic (of the first kind) if a certain generalized derivative exists or equivalently if the equations

$$\begin{aligned} \phi_x F_1 - \phi_y F_2 + \psi_x G_1 - \psi_y G_2 &= 0 \\ \phi_y F_1 + \phi_x F_2 + \psi_y G_1 + \psi_x G_2 &= 0 \end{aligned} \tag{2}$$

are satisfied in D, where the subscripts x and y refer to partial derivatives with respect to x and y and the subscripts 1 and 2 refer to the real and imaginary parts of the functions F and G. If $F = 1$ and $G = i$, these equations reduce to the Cauchy-Riemann equations.

Given a generating pair (F, G) let B denote the class of all functions which are (F, G)-psuedo-analytic. If $F = 1$ and $G = i$, then B is the class of analytic functions on D, which will be referred to in this paper as A.

Any B has many of the properties of the ring of analytic functions. In particular very close analogues of the identity theorem, the Cauchy theorem, the Cauchy integral formula, the standard convergence theorems, and power series expansions have been proved.

With each class B is associated a class B' of psuedo-analytic functions of the second kind. This association is made by a mapping η of B into B' defined by

$$\eta(\phi F + \psi G) = \phi + i\psi \,.$$

On the class A of analytic functions this mapping is clearly the identity.

Each class B is a vector space with the usual definition of addition

Received October 13, 1959.

of functions and multiplication by scalars and η is a vector space isomorphism of B onto B'. The class A is a ring under the usual pointwise multiplication of functions. Since the classes of psuedo analytic functions each bear such marked resemblances to the class A of analytic functions, the question arises as to whether there exist for other classes appropriate generalizations of the ordinary multiplications of function. We shall prove that if such a multiplication bears a certain slight resemblance to the point wise multiplication, then B is multiplicatively isomorphic to A under the mapping η and conversely.

We denote the ordinary multiplication of functions by juxtaposition. Let m denote any mapping from $B \times B$ to the set of all functions from D to the plane. In particular let m_p be the mapping defined as follows: if $w = \phi F + \psi G$ and $w' = \phi' F + \psi' G$, let

$$m_p(w, w') = (\phi\phi' - \psi\psi')F + (\phi\psi' + \psi'\phi)G .$$

THEOREM. *Let B be a system of psuedo-analytic functions on the plane domain D and let m be a multiplication on B* (any mapping from $B \times B$ to the set of all functions from D to the plane). *Let m be associative and bilinear with respect to addition in B. Then a necessary and sufficient condition for the mapping η to be a multiplicative isomorphism of B onto the ring A of all analytic functions on D is that there exists a non-constant w in B such that $m(w, G) = m_p(w, G)$ and $m_p(w, G) \in B$.*

The proof of this theorem will be preceded by a lemma.

LEMMA. *Suppose that for all w and w' in B, $m(w, w') = m_p(w, w')$. Then the mapping η defined above is an isomorphism of B onto the ring A of analytic functions on D if and only if $F_1 = G_2$ and $F_2 = -G_1$.*

Proof of Lemma. A simple calculation shows that η is an isomorphism of B onto B' if and only if $m = m_p$. So the condition concerning isomorphism in the lemma is that $B' = A$.

By adding and subtracting terms involving ψ the system (2) is seen to be equivalent to

$$\begin{aligned} &F_1(\phi_x - \phi_y) - F_2(\phi_y + \psi_x) + \psi_y(F_1 - G_2) + \psi_x(F_2 + G_1) = 0 \\ &F_1(\phi_y + \psi_x) + F_2(\phi_x - \psi_y) + \psi_y(F_2 + G_1) - \psi_x(F_1 - G_2) = 0 . \end{aligned} \tag{3}$$

First suppose $F_1 = G_2$ and $F_2 = -G_1$. Then this system becomes

$$\begin{aligned} &F_1(\phi_x - \psi_y) - F_2(\phi_y + \psi_x) = 0 \\ &F_1(\phi_y + \psi_x) + F_2(\phi_x - \psi_y) = 0 . \end{aligned} \tag{4}$$

It is clear that if $w^* = \phi + i\psi$ is analytic, then $\eta^{-1}(w^*)$ satisfies the system (4). Therefore $A \subset B'$

Suppose then that B' contains a function $w = \phi + i\psi$ which at some point z of D does not satisfy the Cauchy-Riemann equations. For this point the system (4) is a system of homogeneous algebraic equations with a non zero determinant whose value is

$$[\phi_x(z) - \psi_y(z)]^2 + [\phi_y(z) + \psi_y(z)]^2 .$$

Hence the only solution at z is the zero solution and thus

$$\mathfrak{Im}\,[\overline{F(z)}G(z)] = F_1^2(z) + F_2^2(z) = 0$$

which contradicts the definition of generating pair. Thus $B' \subset A$ and we have proved that $A = B'$.

Conversely suppose that η is an isomorphism onto A so that $A = B'$. Let $w^* = \phi + i\psi$ be a non constant analytic function in B'. Then the system (3) becomes for this w^*

$$(5) \qquad \begin{aligned} \psi_y(F_1 - G_2) + \psi_x(F_2 + G_1) &= 0 \\ \psi_y(F_2 + G_1) - \psi_x(F_1 - G_2) &= 0 . \end{aligned}$$

If for some z the equations $F_1(z) = G_2(z)$ and $F_2(z) = -G_1(z)$ do not both hold, then the determinant of this system is non-zero at z and hence by continuity of F and G the determinant is non-zero in some neighborhood of z and hence $\psi_x = \psi_y = 0$ on this neighborhood. By the identity theorem for harmonic functions ψ_x and ψ_y must then be zero everywhere so that ψ is constant. A similar argument demonstrates the constancy of ϕ so that w^* is constant contrary to assumption. This completes the proof of the lemma.

Proof of Theorem. Suppose first that η is a multiplicative isomorphism of B onto the ring A of analytic functions in D. Then as before m is identically equal to m_p so that for $w = \phi F + \psi G \in B$ we have $m(w, G) = -\psi F + \phi G$. Substituting this function for $\phi F + \psi G$ in the system (1) yields that $m(w, G)$ is in B if and only if

$$(6) \qquad \begin{aligned} -\psi_x F_1 + \psi_y F_2 + \phi_x G_1 - \phi_y G_2 &= 0 \\ -\psi_y F_1 - \psi_x F_2 + \phi_y G_1 + \phi_x G_2 &= 0 . \end{aligned}$$

By the lemma $F_1 = G_2$ and $F_2 = -G_1$. Using this to substitute for the G's in the system (6) we obtain the system (4) and this system must be satisfied because $\phi + i\psi$ is analytic. Thus if w is in B then so is $m(w, G)$ and the condition of the theorem is necessary.

Conversely suppose that there exists a non-constant w in B such that $m(w, G) = m_p(w, G) = -\psi F + \phi G$ and this function is in B. Then ϕ and ψ satisfy both (1) and (6) and since w is non-constant there must exist a z such that this system of four equations has a non-zero solution, i.e., the determinant of this system must be zero.

The determinant of this system is

$$[(F_1 - G_1)^2 + (F_2 + G_1)^2][(F_1 + G_2)^2 + (F_2 - G_1)^2] . \tag{7}$$

Now $\mathfrak{Im}(\bar{F}G) = F_1G_2 - F_2G_1$ which must be everywhere positive since F and G form a generating pair. If the second factor of (7) is zero, then it follows that

$$\mathfrak{Im}(\bar{F}G) = -F_1^2 - F_2^2 < 0 .$$

Hence the first factor must be zero and the lemma implies that η is an isomorphism of B onto A.

BIBLIOGRAPHY

1. Lipman Bers, *Theory of Psuedo-analytic functions*, Mimeographed Lecture Notes, New York University, 1953.

2. ———, *Local Theory of Psuedo-analytic functions*, Lectures on Functions of a Complex Variable, Ann Arbor, Michigan, 1955, pp. 213–244.

UNIVERSITY OF GEORGIA

ANALYTIC AUTOMORPHISMS OF BOUNDED SYMMETRIC COMPLEX DOMAINS

HELMUT KLINGEN

In a former paper [2] I determined the full group of one-to-one analytic mappings of a bounded symmetric Cartan domain [1]. Those investigations were incomplete, because it was impossible to treat the second Cartan-type of $n(n-1)/2$ complex dimensions for odd n by this method. The present note is devoted to a new shorter proof of the former result (n even), which furthermore covers the remaining case of odd n.

Take the complex $n(n-1)/2$-dimensional space of skew symmetric n-rowed matrices Z. The irreducible bounded symmetric Cartan space in question is the set $\mathcal{E}_n$ of those matrices Z, for which

$$I + Z\bar{Z} > 0 \, , \quad Z' = -Z,$$

is positive definite. Here I is the n by n unit matrix. Obviously $\mathcal{E}_2$ is the unit circle. It is easy to see that analytic automorphisms of $\mathcal{E}_n$ are described by the group ϕ of the mappings

$$W = (AZ + B)(-\bar{B}Z + \bar{A})^{-1} \, , \tag{1}$$

where the n-rowed matrices A, B fulfill

$$M^*KM = K \quad \text{with} \quad M = \begin{pmatrix} A & B \\ -\bar{B} & \bar{A} \end{pmatrix}, \quad K = \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix}.$$

Here M^* denotes the conjugate transpose of M. For $n = 4$

$$W = \tilde{Z}$$

is a further analytic automorphism, where $\tilde{Z}$ arises from Z by interchanging the elements z_{14} and z_{23},

$$\tilde{Z} = \begin{pmatrix} 0 & z_{12} & z_{13} & z_{23} \\ -z_{12} & 0 & z_{14} & z_{24} \\ -z_{13} & -z_{14} & 0 & z_{34} \\ -z_{23} & -z_{24} & -z_{34} & 0 \end{pmatrix}.$$

For $W\bar{W}$ and $\tilde{Z}\bar{\tilde{Z}}$ have the same characteristic roots. But this mapping is not contained in ϕ, since $CZ = \tilde{Z}D$ cannot be satisfied identically in Z by non-singular constant matrices C, D. On the other hand the following theorem holds.

Received October 13, 1959.

THEOREM. *Each analytic automorphism of $\mathcal{E}_n$ can be written as $W = f(Z)$ or $W = f(\tilde{Z})$ (only for $n = 4$) with $f \in \phi$.*

Therefore the group ϕ is already the full group of analytic automorphisms for $n \neq 4$. Only in the exceptional case $n = 4$ there are the further mappings $W = f(\tilde{Z})$, which together with ϕ form the full group of analytic automorphisms. The proof of this theorem consists of two parts. The first analytic part is a reproduction of my former proof [2], which will be given here again for completeness, the second part is of algebraic character.

The group ϕ acts transitively on $\mathcal{E}_n$. For take an arbitrary point Z_1 of $\mathcal{E}_n$, choose the matrix A such that

$$A(I + Z_1\bar{Z}_1)A^* = I$$

and define $B = -AZ_1$. Then (1) maps Z into 0. Therefore it is sufficient to investigate the stability group of the zero matrix.

First we show that each analytic one-to-one mapping $W = W(Z)$ of $\mathcal{E}_n$ with the fixed point 0 is linear. For an arbitrary point $Z_1 \in \mathcal{E}_n$ let $r_1, \cdots, r_n, 0 \leqq r_1 \leqq \cdots \leqq r_n < 1$, be the characteristic roots of $Z_1 Z_1^*$. Then also tZ_1 belongs to $\mathcal{E}_n$, if t is a complex number with $t\bar{t}r_n < 1$. Consequently there exists a power series expansion

$$W(tZ_1) = \sum_{k=1}^{\infty} t^k W_k(Z_1) , \qquad t\bar{t}r_n < 1 . \tag{2}$$

The elements of the skew-symmetric matrices $W_k(Z_1)$ are homogeneous polynomials of degree k in the independent elements of Z_1. Because of $I + W(tZ_1)\bar{W}(tZ_1) > 0$ for $\bar{t}t = 1$, one obtains from (2)

$$\frac{1}{2\pi i}\int_{t\bar{t}=1} (I + W(tZ_1)\bar{W}(tZ_1))\frac{dt}{t} = I + \sum_{k=1}^{\infty} W_k(Z_1)\bar{W}_k(Z_1) > 0 \tag{3}$$

and in particular $I + \bar{W}_1(Z_1)W_1(Z_1) > 0$. Therefore the linear function $W_1(Z)$ is an analytic mapping of $\mathcal{E}_n$ into itself. Its determinant D is at the same time the Jacobian of the function $W(Z)$ with respect to Z. By interchanging Z and W it can be assumed $D\bar{D} \geqq 1$. Consequently $W(Z)$ is an analytic automorphism of $\mathcal{E}_n$ and even maps the boundary onto itself. Take now in particular

$$Z_1 = U'PU , \quad P = [(0), p_1 F, \cdots, p_m F] , \quad F = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \tag{4}$$

with an unitary matrix $U, m = [n/2]$. P shall be the matrix, which is built up by the two-rowed blocks $p_1 F, \cdots, p_m F$ and possibly by the element 0 along the main diagonal. Z_1 belongs to the interior of $\mathcal{E}_n$, if $-1 < p_k < 1$ $(k = 1, \cdots, m)$, and to the boundary, if $-1 \leqq p_k \leqq 1$ $(k =$

$1, \cdots, m)$ and $p_k = \pm 1$ for at least one k. Now $|I + W_1(Z_1)\bar{W}_1|$ is a polynomial in $p_1, \cdots, p_m$ of total degree $4m$ and on the other hand (see [2], Lemma 4) the square of a polynomial. As $|I + W_1(Z_1)\bar{W}_1|$ vanishes on the boundary of $\mathcal{E}_n$, this polynomial is divisible by

$$|I + Z_1\bar{Z}_1| = \prod_{k=1}^{m} (1 - p_k^2)^2 .$$

Because the constant terms and the degrees of both polynomials are equal, one obtains

$$(5) \qquad |I + W_1(Z_1)\bar{W}_1| = |I + Z_1\bar{Z}_1|$$

even identically in Z_1; for each skew-symmetric matrix Z_1 permits a representation (4) (see [2], Lemma 3). On account of (5) and the linearity of W_1 the matrices $W_1\bar{W}_1$ and $Z\bar{Z}$ always have the same characteristic roots and this implies

$$(6) \qquad W_1(Z) = U'ZU$$

with unitary U, which for the present still depends on Z.

Put now

$$Z = uX, \quad X = U_1', [e^{i\zeta_1}F, \cdots, e^{i\zeta_r}F, (0)]U_1, \quad 0 \leqq u \leqq 1,$$

with real variables $\zeta_1, \cdots, \zeta_r$. Then $Z \in \mathcal{E}_n$ and by (6)

$$W_1W_1^* = u^2U'U_1'\begin{pmatrix} I^{(n-1)} & 0 \\ 0 & (0) \end{pmatrix}\bar{U}_1\bar{U}$$

for all u between 0 and 1. Because of (3) one obtains

$$\bar{U}_1\bar{U}(I + W_1\bar{W}_1 + W_k\bar{W}_k)U'U_1' > 0 \qquad (k = 2, 3, \cdots) .$$

If u tends to 1, one gets

$$\begin{pmatrix} 0 & 0 \\ 0 & (1) \end{pmatrix} + \bar{U}_1\bar{U}W_k\bar{W}_kU'U_1' > 0 ,$$

hence $W_k(X) = 0$. As W_k is a polynomial, $W_k(Z)$ even vanishes identically in Z. Therefore the stability group of $\mathcal{E}_n$ is linear.

The investigation of $W = W_1(Z)$ is now a purely algebraic problem. The representation (6) shows that rank $W =$ rank Z and beyond this the equality of the characteristic roots of $W\bar{W}$ and $Z\bar{Z}$. These properties will be used in order to determine $W(Z)$ explicitly. We have to prove

$$(7) \qquad W(Z) = U'ZU \quad \text{or} \quad W(Z) = U'\tilde{Z}U$$

with unitary constant U, where the second type only occurs for $n = 4$. The proof of this fact will be given by induction. The assertion (7) is trivial for the unit circle ($n = 2$). Let us assume its correctness for $2, 3, \cdots, n-1$ and consider $\mathcal{E}_n$. Write the linear mapping $W(Z)$ of $\mathcal{E}_n$ onto itself as

$$W = \sum_{k<l} z_{kl} A_{kl}$$

with constant skew-symmetric n by n matrices A_{kl}. Because of the equality of the characteristic roots of WW^* and ZZ^* the hermitian matrix $A_{kl}A_{kl}^*$ has $1, 1, 0, \cdots, 0$ as characteristic roots. Therefore after unitary transformation of W we can assume $A_{12} = E_{12}$, where in general E_{kl} denotes the skew-symmetric matrix the elements of which are all zero besides the element in the kth row and lth column and the element in the lth row and kth column, which are 1 respectively -1. Since $\operatorname{tr}(A_{12}\bar{A}_{kl}) = 0$ for $(k, l) \neq (1, 2)$, one obtains

$$A_{kl} = \begin{pmatrix} 0^{(2)} & * \\ * & * \end{pmatrix} \qquad (k, l) \neq (1, 2)\,.$$

$A_{12} = E_{12}$ does not change, if W is transformed by

$$\begin{pmatrix} U^{(2)} & 0 \\ 0 & V \end{pmatrix}$$

with unitary $U, V, |U| = 1$. Therefore

$$A_{13} = \begin{pmatrix} 0^{(2)} & B \\ -B' & C \end{pmatrix}, \quad B = \begin{pmatrix} b_1 & 0 & \\ 0 & b_2 & 0 \end{pmatrix}$$

can be assumed. From rank W = rank Z identically in Z one obtains possibly after unitary transformation $A_{13} = E_{13}$.

For $A_{14} = (a_{kl})$ we get two possibilities. First the equation $\operatorname{tr}(A_{12}\bar{A}_{14}) = \operatorname{tr}(A_{13}\bar{A}_{14}) = 0$ implies $a_{12} = a_{13} = 0$. After unitary transformation all the elements of the first row besides a_{14} are zero. Then take only the elements z_{12}, z_{13}, z_{14} of Z distinct from zero; from rank W = rank $Z = 2$ one sees

$$A_{14} = E_{14} \quad \text{or} \quad A_{14} = E_{23}\,.$$

By a similar consideration $A_{1\nu}$ turns out to be $E_{1\nu}$ or E_{23}. But actually for $\nu > 4$ the second possibility $A_{1\nu} = E_{23}$ may not occur. For $A_{14} = A_{1\nu} = E_{23}$ is impossible because of $\operatorname{tr}(A_{14}\bar{A}_{1\nu}) = 0$. If $A_{14} = E_{14}, A_{1\nu} = E_{23}$, choose only the elements $z_{1\nu}, z_{14} \neq 0$, then rank $Z = 2$ but rank $W = 4$. Therefore $A_{1\nu} = E_{1\nu}$ $(\nu \neq 4)$, $A_{14} = E_{14}$ or E_{23}. Furthermore $A_{14} = E_{23}$ may only happen if $n = 4$. For assume $A_{14} = E_{23}, A_{15} = E_{15}$ and take only the elements $z_{14}, z_{15} \neq 0$. This implies rank $Z = 2$ but rank $W = 4$.

Let us summarize our results. After a suitable unitary transformation W can be written as

$$W = \begin{pmatrix} 0 & z' \\ -z & L(QZ_0) \end{pmatrix}, \quad Z = \begin{pmatrix} 0 & z' \\ -z & Z_0 \end{pmatrix},$$

besides the exceptional case $n = 4$, $A_{14} = E_{23}$. Now $L(Z_0)$ is an analytic automorphism of $\mathcal{E}_{n-1}$ with the fixed point 0. For $n = 3$ we know $L(Z_1) = e^{i\zeta}Z_1$ with a real constant ζ. Therefore $W = U'ZU$ with a constant unitary matrix U, which is the theorem for $n = 3$. For $n > 5$ the induction hypothesis shows

$$W = \begin{pmatrix} 0 & z'U' \\ -Uz & Z_0 \end{pmatrix}$$

with constant unitary U. From the equality

$$\text{rank}\, W = \text{rank}\, Z$$

U turns out to be a diagonal matrix. Finally consider the sum of the two-rowed principal minors of $W\bar{W}$ and $Z\bar{Z}$. These two quantities are equal identically in Z because of the fact that $W\bar{W}$ and $Z\bar{Z}$ have the same characteristic roots. By this identity one obtains $U = aI$ with a complex number a of absolute value 1, which again proves our theorem.

There still remain the cases $n = 4$ and 5. For $n = 4$, $A_{14} = E_{14}$ we can use the reasoning above. Let $A_{14} = E_{23}$; since

$$\text{tr}\,(A_{1\nu}\bar{A}_{23}) = \text{tr}\,(A_{1\nu}\bar{A}_{24}) = \text{tr}\,(A_{1\nu}\bar{A}_{34}) = 0 \qquad (\nu = 2, 3, 4)$$

W only differs from $\tilde{Z}$ in the last row, where a linear combination of z_{23}, z_{24}, z_{34} appears. The identity between the ranks of Z and W shows $w_{14} = a_1 z_{23}$, $w_{24} = a_2 z_{24}$, $w_{34} = a_3 z_{34}$. Now it is easy to compute the sum of the two-rowed principal minors of $W\bar{W}$ and $Z\bar{Z}$. This computation shows again the assertion for $n = 4$.

For $n = 5$ we know by the induction hypothesis

$$L(Z_0) = U'Z_0U \quad \text{or} \quad L(Z_0) = U'\tilde{Z}_0U$$

with constant unitary U. The first case can be treated as above. In the second case one obtains

$$W = \begin{pmatrix} 0 & z'U' \\ -Uz & Z_0 \end{pmatrix}.$$

Choose once only $z_{14}, z_{24} \neq 0$, then only $z_{14}, z_{34}, z_{45} \neq 0$. In any case rank $Z = 2$, hence rank $W = 2$. But this implies that all the elements of the third column of U vanish, which is a contradiction to the unitary character of U. This final remark completes the proof.

References

1. E. Cartan, *Sur les domaines bornés homogènes de l'espace de n variables complexes,* Oeuvres, partie 1, 1259–1304.
2. H. Klingen, *Diskontinuierliche Gruppen in symmetrischen Räumen,* Mathematische Annalen **129** (1955), 470–488.

Mathematisches Institut der Universität
Göttingen, Germany

ORDERED SEMIGROUPS IN PARTIALLY ORDERED SEMIGROUPS

R. J. KOCH

In this note we establish a local version of the following result: a locally compact connected partially ordered non-degenerate semigroup S with unit contains a non-degenerate linearly ordered local subsemigroup (containing the unit). This is an extension of a result of Gleason [2; 664] who proved a similar theorem under the additional hypotheses that

(1) S is a semigroup with right invariant uniform structure and

(2) for any compact neighborhood U of the identity there are nets $\{x_i\}$ in S and $\{n_i\}$ integers such that $x_i \to e$ and $x_i^{n_i} \notin U$. A consequence of our theorem is the fact that a nondegenerate compact connected partially ordered semigroup with unit contains a standard thread joining the unit to the minimal ideal.

By a local semigroup S we mean a Hausdorff space with an open subset U and a multiplication $m: U \times U \to S$ which is continuous and associative insofar as is meaningful. A unit is an (unique, if it exists) element u of U satisfying $ux = xu = x$ for all $x \in U$. A local subsemigroup of S is a subset L containing the unit such that for some open set V about the unit, $(V \cap L)^2 \subset L$. We say that the local semigroup S is partially ordered if the relation $\leqq$ defined by $a \leqq b$ if and only if $a = bc$ is reflexive and antisymmetric. In case S is a semigroup, S is partially ordered if and only if each principal right ideal has a unique generator, i.e. (assuming a unit) that $aS = bS$ implies $a = b$. In this case, $\leqq$ is also transitive.

Closure is denoted by $*$, the null set by $\square$, the boundary of V by $F(V)$, and the complement of B in A by $A\backslash B$.

As in [4] we use the following topolopy for the space $\mathscr{S}(X)$ of non-empty closed subsets of the space X: for open sets U and V of X, let $N(U, V) = \{A \mid A \in \mathscr{S}(X), A \subset U, A \cap V \neq \square\}$; take $\{N(U, V) \mid U, V \text{ open}\}$ for a sub-basis for the open sets of $\mathscr{S}(X)$. It is easy to see that if X is compact Hausdorff, so is $\mathscr{S}(X)$.

THEOREM 1. *Let S be a locally compact partially ordered local semigroup with unit u, and let U_0 be a non-degenerate open connected set about u with U_0^6 defined. Then S contains a non-degenerate compact connected linearly ordered local sub-semigroup L with $u \in L \subset U_0$.*

Received December 7, 1959. This work was supported by the United States Air Force through the Air Force Office of Scientific Research, Air Research and Development Command, under Contract AF 18 (603)-89

Proof. Let U_1 be an open set containing u with U_1^* compact and $U_1^* \subset U_0$. Define $\leqq$ on $U_1^* \times U_1^*$ by: $a \leqq b$ if and only if $a = bc$ for some $c \in U_1^*$. From the compactness of U_1^* it is easily seen that Graph ($\leqq$) is closed in $U_1^* \times U_1^*$. We show first that $\leqq$ is transitive on some neighborhood of u. Let U_2 be an open set about u with $U_2^2 \subset U_1$. We claim there is an open set U containing u, $U \subset U_2$, such that if $a, b \in U^*$ with $a = bc$ for some $c \in U_1^*$, then $c \in U_2$. If this is false, then for any open set U with $u \in U \subset U_2$, there are elements a and b of U^* with $a = bc$ for some $c \in U_1^* \backslash U_2$. Hence there are nets a_α and b_α converging to u with $a_\alpha = b_\alpha \cdot c_\alpha$ where $c_\alpha \in U_1^* \backslash U_2$. It follows that c_α must also converge to u, a contradiction. Since $U_2^2 \subset U_1$ it follows that $\leqq$ is transitive on U^*. Also the restriction of $\leqq$ on $U^* \times U^*$ is closed and hence U^* is locally convex [6]. We show next that there exists an open set V_1 with $u \in V_1 \subset U$ such that $e^2 = e \in V_1$ implies $eU_0e \neq e$. Suppose the contrary; we can then find a net of idempotents $e_\alpha \to u$ with $e_\alpha U_0 e_\alpha = e_\alpha$. Let $x \in U_0$; then $e_\alpha = e_\alpha x e_\alpha$ converges to $uxu = x$, so that $x = u$ and U_0 is degenerate, a contradiction. Let V be a convex open set with $u \in V \subset V^* \subset (V^*)^2 \subset V_1$. Then $e^2 = e \in V$ implies $eU_0e \neq e$.

Let $\mathscr{C}$ denote the collection of all closed chains C in U^* with $u \in C$, $C \cap S\backslash V \neq \Box$, and $(C \cap V)^2 \subset C$. Note that $\mathscr{C} \neq \Box$, for if $a \in F(V)$, then the elements u and a constitute an element of $\mathscr{C}$.

(i) *$\mathscr{C}$ is closed in $\mathscr{S}(U^*)$.* We will show that $\mathscr{C}$ is an intersection of closed set. Since the collection of all closed chains which contain u and meet $S\backslash V$ is closed [4], it remains to show that the collection of closed chains C satisfying $(C \cap V)^2 \subset C$ is closed. Suppose A is a closed chain with $(A \cap V)^2 \not\subset A$; then there are elements a and b of $A \cap V$ with $ab \in S\backslash A$. Hence there exist open sets U_a, U_b, and W containing a, b, and A respectively, with $U_a \cdot U_b \cap W = \Box$. Now $N(W, U_a) \cap N(W, U_b)$ is an open set about A, and contains no chain C with $(C \cap V)^2 \subset C$. This establishes (i).

As in [4], we define $L(x) = \{y \mid y \leqq x\}$, $M(x) = \{y \mid x \leqq y\}$, and $(x, y) = \{z \mid x < z < y\}$. Let δ be an open cover of U^*, and define a subset M_δ of $\mathscr{S}(U^*)$ by: $C \in M_\delta$ if and only if C is a closed chain in U^*, and for any x and y in C with $x < y$ and $(x, y) \cap C = \Box$, there exists $D \in \delta$ such that D^* meets both $L(x) \cap C$.

(ii) *$M_\delta \cap \mathscr{C} \neq \Box$ for any open cover δ of U^*.* Let δ be an open cover of U^*, and let $\mathscr{D}$ be the collection of all closed chains C with $u \in C \subset U$, $C \in M_\delta$, and $(V \cap C)^2 \subset C$. Let τ be a maximal tower in $\mathscr{D}$, and let $T = U\tau$. Then T^* is a closed chain, $u \in T^* \subset U^*$, and $(V \cap T^*)^2 \subset T^*$. As in [4], $T^* \in M_\delta$, and it remains to show that $T^* \in \mathscr{C}$, i.e., that $T^* \cap S\backslash V \neq \Box$. Suppose $T^* \subset V$; (note then that $T = T^*$) then since $(T \cap V)^2 \subset T$, T is a compact chain and a semigroup. Let $e = \inf T$. Since $e^2 \leqq e$ and $e^2 \in T$ we have $e^2 = e$. We show next that e is a zero for T. Let $y \in T$, then $ey \in T$ and $ey \leqq e$, so $ey = e$ and e is a left zero for T. Hence the minimal ideal K of T consists of left zeros for T [1]. Let $f \in K$; then $e \leqq f$ so there exists $c \in U_1^*$ with $e = fc$. Therefore $f = fe = e$, and e is the unique left zero, and hence a zero for T. Let $W \in \delta$ with $e \in W$. If $eU_0e \cap W \cap V$ contains an idempotent $g \neq e$, then $T \cup g$ is a semigroup: for if $x \in T$ then $xg = x(eg) = eg = g$ and $gx = (ge)x = g(ex) = ge = g$. Also $T \cup g$ is a chain, so by the maximality of τ, $T = T \cup g$, a contradiction.

Hence we may assume that $eU_0e \cap W \cap V$ has a unique idempotent e. Since $\leqq$ is antisymmetric, the maximal subgroup of S containing e is e. Also eU_0e is a local semigroup with unit e, $eU_0e \neq e$, and e is not isolated in eU_0e which is the continuous image of U_0 and hence connected. Hence [5; 122] there is a non-degenerate one parameter local semigroup A with $e \in A \subset eU_0e \cap W \cap V$; let $a \in A$ with $a \neq e$ and $a^2 \in A$. Define $a^0 = e$ and let $B_k = \bigcup_{n=0}^{k} a^n[a, e]$, $B_\infty = \bigcup_{n=0}^{\infty} a^n[a, e]$ where $[a, e]$ denotes the sub-arc of A from a to e. We assume temporarily that all products involved in forming B_k and B_∞ are defined. Each of the sets $a^n[a, e]$ is a compact connected chain (hence an arc) with minimal element a^{n+1} and maximal element a^n. Hence B_k is a compact connected chain from a^{k+1} to e. Also B_∞ is a connected chain, hence B_∞^* is a closed connected chain. Using the easily established commutativity of B_k and B_∞^* it follows that for $x \in T$ and $b \in B_k$ (or B_∞^*) then $xb = x(eb) = (xe)b = eb = b$, and similarly $bx = b$. Hence $[(T \cup B_k^2) \cap V]^2 \subset T \cup (B_k^2 \cap V)^2$ and similarly with B_k replaced by B_∞^*. We distinguish two cases:

Case 1: For some $k \geqq 0$, $a^{k+1} \in V$ and $a^{k+2} \notin V$. Then since V is convex, $a^0, a, \cdots, a^{k+1}$ are in V and all products involved in forming B_k are defined, so that $B_k \subset V$ and $B_{k+1} \not\subset V$. We show first that $B_k^2 \cap V \subset B_k$. Let $z \in B_k^2 \cap V$; then $z = xy$ with $x, y \in B_k$, so $x = a^n x'$ and $y = a^m y'$ with x' and y' in $[a, e]$. Hence $xy = a^{m+n}x'y'$. If $x'y' \in A$, then since $z \in V$ it follows that $m + n \leqq k$. If $x'y' \notin A$, then $x'y' = at$ for some $t \in A$, so $xy = a^{m+n+1}t$ and $m + n + 1 \leqq k$. In either case, then, $z \in B_k$. Note that $(T \cup B_k)^2 \in M_\delta$ since B_k^2 is a connected chain. Also $[(T \cup B_k^2) \cap V]^2 \subset T \cup (B_k^2 \cap V)^2 \subset T \cup B_k^2$, so that $T \cup B_k^2 \in \mathscr{D}$. This contradicts the maximality of τ.

Case 2: $a^k \in V$ for each $k \geqq 0$. Using the convexity of V we see that all products involved in forming B_∞ are defined, and $B_\infty = B_\infty^2 \subset V$, hence $B_\infty^* = B_\infty^{*2}$. Since B_∞^* is a connected chain, it follows that $T \cup B_\infty^* \in M_\delta$. Also $[(T \cup B_\infty^*) \cap V]^2 \subset T \cup B_\infty^*$, so that $T \cup B_\infty^* \in \mathscr{D}$, a contradiction to the maximality of τ. The proof of (ii) is now complete.

(iii) *$M_\delta \cap \mathscr{C}$ is closed for each finite open cover δ of U^*.*

This proof is similar to that in [4], and is omitted.

For any finite open cover δ of U^*, let $P_\delta = M_\delta \cap \mathscr{C}$. The collection of sets $\{P_\delta\}$ is a descending family, so $\bigcap P_\delta \neq \square$. If $C \in \bigcap P_\delta$, then as shown in [4], C is an arc. Clearly C is a local semigroup, and the proof is complete.

In what follows, a *standard thread* is a compact connected semigroup irreducibly connected between a zero and a unit. The structure of standard threads is known [5; 130]. The example in [4] shows that a compact connected semigroup with zero and unit need not contain a standard thread joining the zero to the unit. The problem of finding standard threads joining zero to unit has an affirmative solution in case either

(1) S is compact, connected, and one-dimensional [3], or

(2) S is compact, connected, and each element is idempotent [4].
A third solution is given by the following corollary.

COROLLARY 1. *If S is a non-degenerate compact connected partially ordered semigroup with unit u, then the minimal ideal K consists of left zeros for S, K consists of the set of minimal elements, and some elements of K can be joined by a standard thread to the unit.*

Proof. Note that Graph ($\leqq$) is closed since S is compact. Let G be a compact group in S, with unit e. Since $x^2 \leqq x$ for each $x \in S$, then for $x \in G$ we have $e \geqq x \geqq x^2 \geqq \cdots$, and $\{x^n\}$ clusters at an idempotent, which must be e. We conclude that $x = e$, and hence that each compact group in S is trivial. From this fact it is clear that K is proper, for otherwise $K = S$ would be a compact group [1]. From the fact that $aS = bS$ implies $a = b$ we conclude that each minimal right ideal is a single element, hence each element of K is a left zero for S [1]. Since a minimal element x of S is characterized by the equality $xS = x$, it is clear that K consists of the set of minimal elements of S, and hence that $S \backslash K$ is convex. In the proof of the Theorem, we take $S = U_0 = U_1 = U_2 = U$, and $V = S \backslash K$. Hence there is a compact connected linearly ordered local semigroup L containing u, with $L \cap S \backslash V \neq \square$. Since the elements of K are minimal it follows that L is a semigroup, and hence a standard thread.

REFERENCES

1. A. H. Clifford, *Semigroups containing minimal ideals*, Amer. J. Math. **70** (1948), 521-526.

2. A. M. Gleason, *Arcs in locally compact groups*, Proc. Nat. Acad. Science **36** (1950), 663-667.

3. R. P. Hunter, *On the semigroup structure of continua*, Trans. Amer. Math. Society. **93** (1959), 356-368.

4. R. J. Koch, *Arcs in partially ordered spaces*, Pacific J. Math. **9** (1959), 723-728.

5. P. S. Mostert and A. L. Shields, *On the structure of semigroups on a compact manifold with boundary*, Amer. Math **65** (1957), 117-143.

6. L. Nachbin, *Sur les spaces topologiques ordonnés* C. R. Acad. Sci. Paris **226** (1948),

LOUISIANA STATE UNIVERSITY

ON A COMMUTATOR RESULT OF TAUSSKY AND ZASSENHAUS

MARVIN MARCUS[1] AND N. A. KHAN[2]

1. Introduction and results. Let M_n denote the set of n-square matrices over a field F. For A, B in M_n let $[A, B] = AB - BA'$, where A' is the transpose of A and define inductively

$$[A, B]_k = [A, [A, B]_{k-1}] \,. \tag{1.1}$$

If $P^{-1}JP = A$, then

$$[A, X] = [P^{-1}JP, X] = P^{-1}[J, PXP'](P^{-1})' \,,$$

and similarly

$$[A, X]_k = P^{-1}[J, PXP']_k(P^{-1})' \,. \tag{1.2}$$

Now for a fixed A let T be the linear map of M_n into itself defined by

$$T(Y) = [A, Y] \tag{1.3}$$

and (1.1) implies that

$$T^k(Y) = [A, Y]_k \,.$$

In a recent paper [1], Taussky and Zassenhaus showed that *A is nonderogatory if and only if any nonsingular X in the null space of T is symmetric.* In this note we investigate the structure of the null space of both T and T^2 for arbitrary A.

Enlarge the field F to include λ_i, $i = 1, \cdots, p$, the distinct eigenvalues of A, and let $(x - \lambda_i)^{e_{ij}}$, $j = 1, \cdots, n_i$, $e_{i1} > \cdots > e_{in_i}$, $i = 1, \cdots, p$ be the distinct elementary divisors of A where $(x - \lambda_i)^{e_{ij}}$ appears with multiplicity r_{ij}. Set $m_i = \sum_{j=1}^{n_i} r_{ij}e_{ij}$, the algebraic multiplicity of λ_i. *Let $\eta(T)$ denote the null space of T, $\sigma(T)$ denote the subspace of symmetric matrices in $\eta(T)$, and $\gamma(T)$ denote the subspace of skew-symmetric matrices in $\eta(T)$.* We show that

$$\dim \eta(T) = \sum_{i=1}^{p}\left[\sum_{j=1}^{n_i}\left(r_{ij}^2 e_{ij} + 2r_{ij}\sum_{k=j+1}^{n_i} r_{ik}e_{ik}\right)\right], \tag{1.4}$$

$$\dim \sigma(T) = \frac{1}{2}\sum_{i=1}^{p}\left[\sum_{j=1}^{n_i}\left\{r_{ij}(r_{ij}+1)e_{ij} + 2r_{ij}\sum_{k=j+1}^{n_i} r_{ik}e_{ik}\right\}\right], \tag{1.5}$$

Received December 17, 1959. The work of this author was supported by U. S. National Science Foundation Grant, NSF-G5416. The second author is a Postdoctorate Fellow of the National Research Council of Canada. The authors are grateful to Professor O. Taussky for her helpful suggestions.

$$\dim \eta(T^2) = \sum_{i=1}^{p}\left[\sum_{j=1}^{n_i}\left\{r_{ij}^2(2e_{ij}-1) + 4r_{ij}\sum_{k=j+1}^{n_i} r_{ik}e_{ik}\right\}\right], \tag{1.6}$$

$$\dim \sigma(T^2) = \frac{1}{2}\sum_{i=1}^{p}\left[\sum_{j=1}^{n_i}\left\{r_{ij}^2(2e_{ij}-1) + r_{ij} + 4r_{ij}\sum_{k=j+1}^{n_i} r_{ik}e_{ik}\right\}\right]. \tag{1.7}$$

In case A is nonderogatory, $n_i = 1$, $r_{ij} = 1$, $i=1, \cdots, p$ and (1.4) and (1.5) reduce to

$$\dim \eta(T) = n = \dim \sigma(T) .$$

Thus every matrix X satisfying

$$AX = XA' \tag{1.8}$$

where A is non-derogatory is symmetric, the result in [1]. Moreover, if every matrix X satisfying (1.8) is symmetric then $\dim \eta(T) = \dim \sigma(T)$. Using the formulas (1.4) and (1.5) we see that this condition implies that

$$\sum_{i=1}^{p}\sum_{j=1}^{n_i}(r_{ij}^2 - r_{ij})e_{ij} + 2\sum_{i=1}^{p} r_{ij}\sum_{k=j+1}^{n_i} r_{ik}e_{ik} = 0 .$$

Now since r_{ij}, e_{ij} and n_i are all positive integers we conclude that $r_{ij} = 1$, $j = 1, \cdots, n_i$ and $n_i = 1$. That is, there is only one elementary divisor corresponding to each eigenvalue. Hence, *if every matrix X satisfying* (1.8) *is symmetric then A is non-derogatory*, a result also found in [1].

We also show in this case that $\eta(T)$ *consists of matrices of the form PXP' where P is fixed* (depending on A) *and X is persymmetric*, (i.e. all the entries of X on each line perpendicular to the main diagonal are equal).

We next note that $\eta(T) = \sigma(T) + \gamma(T)$ (direct) and $\eta(T^2) = \sigma(T^2) + \gamma(T^2)$ (direct). The first statement is easy to show; we indicate the brief proof of the second statement:

Since $X = \dfrac{X + X'}{2} + \dfrac{X - X'}{2}$, if $X \in \eta(T^2)$, then

$$\begin{aligned} T^2(X + X') &= [A, [A, X + X']] \\ &= [A, [A, X] + [A, X']] \\ &= [A, [A, X]] + [A, [A, X']] \\ &= T^2(X) - [A, [A, X]'] \\ &= [A, [A, X]]' \\ &= (T^2(X))' = 0 . \end{aligned}$$

Similarly, $T^2(X - X') = 0$. Thus any $X \in \eta(T^2)$ is expressible uniquely as a sum of two elements, one in $\sigma(T^2)$ and the other in $\gamma(T^2)$. Hence

(1.9)
$$\dim \gamma(T) = \dim \eta(T) - \dim \sigma(T)$$
$$= \frac{1}{2} \sum_{i=1}^{p} \left[\sum_{j=1}^{n_i} \left\{ r_{ij}(r_{ij} - 1)e_{ij} + 2r_{ij} \sum_{k=j+1}^{n_i} r_{ik}e_{ik} \right\} \right],$$

(1.10)
$$\dim \gamma(T^2) = \dim \eta(T^2) - \dim \sigma(T^2)$$
$$= \frac{1}{2} \sum_{i=1}^{p} \left[\sum_{j=1}^{n_i} \left\{ r_{ij}^2(2e_{ij} - 1) - r_{ij} + 4r_{ij} \sum_{k=j+1}^{n_i} r_{ik}e_{ik} \right\} \right].$$

In case A is non-derogatory, (1.6), (1.7) and (1.10) reduce to

$$\dim \eta(T^2) = 2n - p ,$$

$$\dim \sigma(T^2) = n ,$$

$$\dim \gamma(T^2) = n - p .$$

We thus conclude that *unless all the eigenvalues of A are distinct* ($p = n$) *there exist skew-symmetric matrices X satisfying*

(1.11)
$$A^2X - 2AXA' + X(A')^2 = 0 .$$

If $p = n$, and A is non-derogatory

$$\dim \eta(T^2) = n = \dim \sigma(T^2)$$

and any matrix X satisfying (1.11) *is symmetric.*

On the other hand suppose

$$\dim \eta(T^2) = \dim \sigma(T^2) .$$

From (1.6) and (1.7) we conclude that

$$\sum_{i=1}^{p} \left[\sum_{j=1}^{n_i} \left\{ r_{ij}^2(2e_{ij} - 1) - r_{ij} + 4r_{ij} \sum_{k=j+1}^{n_i} r_{ik}e_{ik} \right\} \right] = 0 .$$

Hence $n_i = 1$, $r_{ij} = 1$, $e_{ik} = 1$ and we conclude that $p = n$. That is, *if every matrix X satisfying* (1.11) *is symmetric then the eigenvalues of A are distinct.*

We show finally (Theorem 2) that *if A is an n-square matrix with p distinct eigenvalues then both* $\dim \gamma(T)$ *and* $\dim \gamma(T^2)$ *are at most* $\frac{1}{2}(n - p)(n - p + 1)$. *Moreover, for each p this bound is best possible.*

Thus *if there exists a skew-symmetric solution of* (1.8) *or* (1.11), *then A has multiple eigenvalues, without the assumption that A is non-derogatory.*

II. *Proofs.* Let $E_{ij} \in M_n$ be the matrix with 1 in position i, j and 0 elsewhere. With respect to this basis, ordered lexicographically, it may be checked that T has the matrix representaion

$$T = I \otimes A - A \otimes I \tag{2.1}$$

where $\otimes$ indicates Kronecker product.

From (1.2) we may take A to be in Jordan canonical form J, since $[A, X]_k = 0$ if and only if $[J, PXP']_k = 0$ and PXP' is symmetric if and only if X is. We write

$$J = \sum_{s=1}^{p}{}^{\cdot} J_s \tag{2.2}$$

where

$$J_s = \lambda_s I_{m_s} + \sum_{t=1}^{n_s}{}^{\cdot} \sum_{1}^{r_{st}}{}^{\cdot} U_{e_{st}} ; \tag{2.3}$$

$\sum^{\cdot}$ indicates direct sum, I_t is a t-square identity matrix, U_t is t-square auxiliary unit matrix (i.e. 1 in the superdiagonal and 0 elsewhere) and $\sum_1^{r_{ji}}{}^{\cdot}\, U_{e_{st}}$ is the direct sum of $U_{e_{st}}$ with itself r_{ij} times.

By a routine computation we see that

$$T^k(Y) = 0$$

if and only if

$$\sum_{\alpha=0}^{k} \binom{k}{\alpha} (-1)^{\alpha} J_s^{k-\alpha} Y_{st} (J_t')^{\alpha} = 0 , \qquad s, t = 1, \cdots, p , \tag{2.4}$$

where $Y = (Y_{st})$, $s, t = 1, \cdots, p$ is a partitioning of Y conformal with the partitioning of J given by (2.2).

For $s \neq t$, it is clear that the matrix representation of (2.4),

$$(I_{m_t} \otimes J_s - J_t \otimes I_{m_s})^k$$

has the single nonzero eigenvalue $(\lambda_s - \lambda_t)^k$ and thus $Y_{st} = 0$. Hence we need only consider the equation (2.4) for $s = t$. We may again partition Y_{ss} conformally with J_s in (2.3). We are thus led to consider the null space of the mapping

$$(I_{e_{si}} \otimes U_{e_{sj}} - U_{e_{si}} \otimes I_{e_{sj}})^k . \tag{2.5}$$

LEMMA 1. *Let* $T = I_m \otimes U_n - U_m \otimes I_n$. *Then*

$$\dim \eta(T) = \min(m, n) , \tag{2.6}$$

$$\dim \eta(T^2) = \begin{cases} 2 \min(m, n) , & \textit{if } m \neq n \\ 2n - 1, & \textit{if } m = n . \end{cases} \tag{2.7}$$

Proof. Suppose $n \leq m$ and that $T(X) = 0$. Let $x_1, \cdots, x_m$ be the column n-vectors of X. Then we have

$$U_n x_j - x_{j+1} = 0 \,, \quad j = 1, 2, \cdots, m-1 \,, \qquad (2.8)$$
$$U_n x_m = 0 \,.$$

For $r = 1, 2, \cdots, n-1$ consider the $(r - j + 1)$ coordinate of (2.8) for $j = 1, \cdots, r$ and we conclude that

$$x_{r+1,1} = x_{r,2} = \cdots = x_{1,r+1} = c_{r+1} \,.$$

Next consider the $(n - j + 1)$ coordinate of (2.8) for $j = 1, \cdots, n$ to obtain

$$0 = x_{n2} = x_{n-1,3} = \cdots = x_{1,n+1} \,.$$

Similarly we see that the remaining elements of X are zero. Hence we find that the jth column of the $n \times m$ matrix X is the transpose of the n-vector

$$[c_j, c_{j+1}, \cdots, c_n, 0, \cdots, 0]$$

for $j = 1, 2, \cdots, n$. The other $m - n$ columns are zero.

In case $n \geqq m$, it is easy to check that the jth row of X is the m-vector

$$[c_j, c_{j+1}, \cdots, c_m, 0, \cdots, 0]$$

for $j = 1, 2, \cdots, m$. The other $n - m$ rows are zero.

This establishes (2.6). To prove (2.7) let $T^2(X) = 0$ and $x_1, x_2, \cdots, x_m$ be the column n-vectors of X. Let us consider the following cases:

(i) $m = n$.

We have

$$U_n^2 x_n = 0, \;\; U_n^2 x_{n-1} = 2U_n x_n$$

and

$$U_n^2 x_j - 2U_n x_{j+1} + x_{j+2} = 0, \; j = 1, 2, \cdots, n-2 \,.$$

Solving these equations recursively we find that the lst, 2nd and jth rows of X are respectively

$$[x_{11}, x_{12}, \cdots, x_{1,n-2}, x_{1,n-1}, x_{1n}] \,,$$

$$[x_{21}, x_{22}, \cdots, x_{2,n-2}, x_{2,n-1}, 0]$$

and

$$(j-1)[x_{2,j-1}, x_{2,j}, \cdots, x_{2,n-1}, 0, \cdots, 0]$$
$$- (j-2)[x_{1,j}, x_{1,j+1}, \cdots, x_{1,n}, 0, \cdots, 0] \,,$$

for $j = 3, 4, \cdots, n$.

The number of arbitrary parameters in X is $2n - 1$.

(ii) $n < m$.

Here we have the following equations:

$$U_n^2 x_j - 2U_n x_{j+1} + x_{j+2} = 0,\ j = 1, 2, \cdots, m-2 \tag{2.9}$$
$$U_n^2 x_{m-1} - 2U_n x_m = 0$$
$$U_n^2 x_m = 0$$

and by solving recursively again we find that the 1st, 2nd and jth rows of X are respectively the m-vectors

$$[x_{11}, \cdots, x_{1,n-1}, x_{1,n}, nx_{n,2}, 0, \cdots, 0]\ ,$$

$$[x_{21}, \cdots, x_{2,n-1}, (n-1)x_{n,2}, 0, 0, \cdots, 0]$$

and

$$[(j-1)x_{2,j-1}, \cdots, (j-1)x_{2,n-1}, (n-j+1)x_{n,2}, 0, \cdots, 0] \\ - (j-2)[x_{1,j}, \cdots, x_{1,n}, 0, 0, \cdots, 0]$$

for $j = 3, 4, \cdots, n$.

In case $n > m$, by similar computation we find that the 1st, 2nd and jth rows of X are respectively

$$[x_{11}, \cdots, x_{1,m-2}, x_{1,m-1}, x_{1m}]\ ,$$

$$[x_{21}, \cdots, x_{2,m-2}, x_{2,m-1}, x_{2m}]$$

and

$$(j-1)[x_{2,j-1}, \cdots, x_{2,m-1}, x_{2m}, 0, \cdots, 0] \\ - (j-2)[x_{1,j}, \cdots, x_{1,m}, 0, 0, \cdots, 0]$$

for $j = 3, 4, \cdots, m+1$. The remaining $n - m - 1$ rows are zero.

From case (ii), we observe that the number of parameters in X is $2 \min(m, n)$.

We now state and prove the following

LEMMA 2. *Let A be an n-square matrix with the single eigenvalue λ and let $(x - \lambda)^{n_i}$ be an elementary divisor of A of multiplicity r_i, $i = 1, \cdots, p$, $n_1 > \cdots > n_p$. Then the most general matrix X satisfying* (1.11) *has*

$$\sum_{i=1}^{p} \left[r_i^2(2n_i - 1) + 4r_i \sum_{j=i+1}^{p} r_j e_j \right] \tag{2.10}$$

arbitrary parameters.

Moreover if X is symmetric it contains

$$\frac{1}{2}\sum_{i=1}^{p}\left[r_i^2(2n_i - 1) + r_i + 4r_i \sum_{j=i+1}^{p} r_j n_j\right] \tag{2.11}$$

parameters.

Proof. Without any loss of generality we can assume that

$$A = \sum_{i=1}^{p}{}^{\cdot}\sum_{j=1}^{r_i}{}^{\cdot} U_i \tag{2.12}$$

where $\sum^{\cdot} U_i$ indicates the direct sum of U_i with itself r_i times. We partition X conformally with A in (2.12) and observe that the equation

$$U_i^2 X_{ij} - 2U_i X_{ij} U_j' + X_{ij}(U_j')^2 = 0$$

determines the structure of any block X_{ij} in the partitioning of X.

From case (i) of Lemma 1, we conclude that any block X_{ij} corresponding to equal U_i's contains $2n_i - 1$ arbitrary parameters and there are r_i^2 such blocks. Also from case (ii) any block in X that corresponds to U_i and U_j, $i < j$, contains $2n_j$ arbitrary parameters. Hence the total number of parameters in X is given by (2.10).

In order to find the number of parameters in a symmetric X we first consider a diagonal block. Its structure has been discussed in Lemma 1, case (i). We observe that if this matrix is symmetric, the number of parameters in it reduces from $2n_i - 1$ to n_i.

Then we consider two symmetrically placed off-diagonal blocks X_{ij} and X_{ji} of orders $n_i \times n_j$ and $n_j \times n_i$ respectively. If X is to be symmetric then by equating the terms of X_{ij} and X_{ji} which are symmetrically placed about the main diagonal of X, the number of arbitrary parameters in X_{ij} and X_{ji} reduces from $2(2n_j)$ to $2n_j$. If X_{ij} and X_{ji} are of order $n_i \times n_i$ then the number of parameters reduces from $2(2n_i - 1)$ to $2n_i - 1$.

We are now in a position to sum the number of parameters in X if it is symmetric and satisfies (1.11). There are r_i blocks in the main diagonal, each of order n_i, $i = 1, \cdots, p$. The number of parameters in each of these blocks is n_i. There are $r_i(r_i - 1)/2$ other square blocks of order n_i. Each of them contains $(2n_i - 1)$ parameters. Thus

$$\frac{1}{2}\sum_{i=1}^{p}\{r_i^2(2n_i - 1) + r_i\}$$

is the number of parameters in all those blocks of X which are square. Since any block of order $n_i \times n_j$ where $n_i > n_j$ contains $2n_j$ parameters, and since we are considering X to be symmetric, we conclude that the total number of arbitrary parameters in X is given by (2.11).

We can similarly prove the following

LEMMA 3. *Let A be the matrix given in Lemma 2. Then the most* *general matrix X satisfying* (1.8) *has*

$$\sum_{i=1}^{p}\left(r_i^2 n_i + 2r_i \sum_{j=i+1}^{p} r_j n_j\right)$$

arbitrary parameters.

Moreover if X is symmetric, it contains

$$\frac{1}{2}\sum_{i=1}^{p}\left[r_i(r_i+1)n_i + 2r_i \sum_{j=i+1}^{p} r_j n_j\right]$$

parameters.

We now state and prove the following

THEOREM 1. *Let A be an n-square matrix with distinct eigenvalues $\lambda_1, \cdots, \lambda_p$ and let $(x-\lambda_i)^{e_{ij}}$, $j = 1, \cdots, n_i$, $e_{i1} > \cdots > e_{in_i}$ be the elementary divisors of A corresponding to λ_i, where each $(x-\lambda_i)^{e_{ij}}$ has been repeated r_{ij} times. Then* (1.4), (1.5), (1.6) *and* (1.7) *hold.*

Proof. It was pointed out earlier that if $Y = (Y_{rs})$, $r, s = 1, \cdots, p$ is the partitioning of Y conformal with the partitioning of J in (2.2), then all the off-diagonal blocks are zero. Hence we have simply to find the number of parameters in Y_{ii}, $i = 1, \cdots, p$.

As proved in Lemma 2, the number of parameters in Y_{ii} is

$$\sum_{j=1}^{n_i}\left[r_{ij}^2(2e_{ij}-1) + 4r_{ij}\sum_{k=j+1}^{n_i} r_{ik}e_{ik}\right].$$

Summing the above with respect to i we obtain the formula (1.6). In case Y is symmetric, the number of parameters in Y_{ii} is

$$\frac{1}{2}\sum_{j=1}^{n_i}\left[r_{ij}^2(2e_{ij}-1) + r_{ij} + 4r_{ij}\sum_{k=j+1}^{n_i} r_{ik}e_{ik}\right].$$

Summing the above on i we obtain (1.7).

Similarly, we can make use of Lemma 3 in proving (1.4) and (1.5).

We now prove

THEOREM 2. *Let A be as given in Theorem* 1. *Then the maximum number of linearly independent skew-symmetric matrices satisfying* (1.8) *or* (1.11) *is*

$$\frac{1}{2}(n-p)(n-p+1).$$

Proof. In order to prove our result for $\dim \gamma(T^2)$, let $m_i = \sum_{j=1}^{n_i} r_{ij}e_{ij}$ and consider

$$\begin{aligned} m_i^2 - m_i - \sum_{j=1}^{n_i}\left[r_{ij}^2(2e_{ij}-1) - r_{ij} + 4r_{ij}\sum_{k=j+1}^{n_i} r_{ik}e_{ik}\right] \\ = \sum_{j=1}^{n_i}\left[r_{ij}^2e_{ij}^2 + 2r_{ij}e_{ij}\sum_{k=j+1}^{n_i} r_{ik}e_{ik} - r_{ij}e_{ij}\right] \\ - \sum_{j=1}^{n_i}\left[r_{ij}^2(2e_{ij}-1) - r_{ij} + 4r_{ij}\sum_{k=j+1}^{n_i} r_{ik}e_{ik}\right] \\ = \sum_{j=1}^{n_i}\left[r_{ij}^2(e_{ij}-1)^2 - r_{ij}(e_{ij}-1) + 2r_{ij}(e_{ij}-2)\sum_{k=j+1}^{n_i} r_{ik}e_{ik}\right]. \end{aligned}$$

Now, it is clear that $r_{ij}^2(e_{ij}-1) \geqq r_{ij}(e_{ij}-1)$. The last term in the above expression will be negative only when $e_{ij}=1$. But we know that $e_{i1} > e_{i2} > \cdots > e_{in_i}$, so that e_{ij} will be 1 only for $j = n_i$. In that case $\sum_{k=j+1}^{n_i}$ does not appear, and we have

$$\frac{1}{2}\sum_{j=1}^{n_i}\left[r_{ij}^2(2e_{ij}-1) - r_{ij} + 4r_{ij}\sum_{k=j+1}^{n_i} r_{ik}e_{ik}\right] \leqq \frac{1}{2}(m_i^2 - m_i)\ .$$

This holds for $i = 1, \cdots, p$.

To determine a bound on $\gamma(T)$, consider

$$\begin{aligned} m_i^2 - m_i - \sum_{j=1}^{n_i}\left[r_{ij}(r_{ij}-1)e_{ij} + 2r_{ij}\sum_{k=j+1}^{n_i} r_{ik}e_{ik}\right] \\ = \sum_{j=1}^{n_i}\left[r_{ij}^2e_{ij}(e_{ij}-1) + 2r_{ij}(e_{ij}-1)\sum_{k=j+1}^{n_i} r_{ik}e_{ik}\right] \\ \geqq 0, \text{ since } e_{ij} \geqq 1. \end{aligned}$$

Thus we have

$$\frac{1}{2}\sum_{j=1}^{n_i}\left[r_{ij}(r_{ij}-1)e_{ij} + 2r_{ij}\sum_{k=j+1}^{n_i} r_{ik}e_{ik}\right] \leqq \frac{1}{2}(m_i^2 - m_i)\ .$$

It may be observed that the upper bound is attained for $r_{i1} = m_i$, $e_{i1} = 1$ and the remaining e's and r's all zero.

We have thus proved that

$$\dim \gamma(T^2) \leqq \frac{1}{2}\sum_{i=1}^{p}(m_i^2 - m_i)$$

and

$$\dim \gamma(T) \leqq \frac{1}{2}\sum_{i=1}^{p}(m_i^2 - m_i)\ ,$$

where m_i is the multiplicity of the eigenvalue λ_i of A.

Now we have to maximize $\sum_{i=1}^{p}(m_i^2 - m_i)$ under the condition that

$m_1 + \cdots + m_p = n$, the order of A. Note that

$$m_i^2 - m_i = (m_i - 1)^2 + (m_i - 1)$$

and each $m_i - 1 \geqq 0$. Hence, we have

$$\sum_{i=1}^{p} (m_i - 1)^2 \leqq \left[\sum_{i=1}^{p} (m_i - 1)\right]^2 = (n - p)^2 .$$

Thus the maximum value of both $\dim \gamma(T^2)$ and $\dim \gamma(T)$ is

$$\frac{1}{2}[(n - p)^2 + (n - p)] .$$

The bounds are achieved when $m_1 = \cdots = m_{p-1} = 1$ and $m_p = n - p + 1$.

Reference

1. O. Taussky and H. Zassenhaus, *On the similarity transformation between a matrix and its transpose.* Pacific J. Math. **9** (1959), 893–896.

The University of British Columbia, Vancouver, Canada
and
Muslim University, Aligarh, India

UNARY ALGEBRAS

JOHN G. MARICA AND STEVEN J. BRYANT

This paper is concerned with algebraic systems composed of a nonempty set A and a single unary operation on A; i.e., a function on A into A, usually denoted by $'$. Such a system is called a unary algebra.

Our main objective is to give a proof of the following theorem: If A and B are two finite unary algebras and A^2 is isomorphic to B^2, then A is isomorphic to B.[1] In this statement, A^2 means the Cartesian product of the algebra A with itself. In addition to this result, we prove some basic structure theorems for unary algebras, and cancellation theorems for some classes of unary algebras. In § 7 we list some counter examples which indicate limitations on the generalization of these results to infinite algebras.

Most of the definitions and theorems were suggested by the graphs of unary algebras and should be easily understood in this context. The graphs are obtained by joining each element to its "prime" or "successor" by a directed line segment. Theorems, equations, definitions, etc. are numbered consecutively in each section.

1. Notation and general theorems. We shall not distinguish notationally between an algebra and the set of elements of the algebra, and unless it is essential to do otherwise, we shall use $'$ to denote the operation in all of the algebras discussed. In general, upper case letters will denote algebras, lower case letters will denote elements. Brackets and parentheses are used in several senses, but for p, q integers (p, q) and $[p, q]$ will always denote the g.c.d. and l.c.m., respectively, of p and q.

1.1. DEFINITION. If $A \subset B, A \neq 0, B$ a unary algebra, and A is closed under $'$, i.e., $A' \subset A$, then A will be called a subalgebra of B.

1.2. LEMMA. *If F is a family of subalgebras of A, then $\bigcup F$ is a subalgebra of A, and if $\bigcap F \neq \phi$, then $\bigcap F$ is a subalgebra of A.*

The proof is immediate.

1.3. DEFINITION. If A and B are unary algebras and $A \cap B = \phi$ then $A \cup B$ is the algebra formed from $A \cup B$ by applying the operation of A to elements of A and the operation in B to elements of B.

Received April 7, 1959, in revised form February 2, 1960. To be presented before the Amer. Math. Soc., Monterey, California, April 18, 1959

[1] This problem was mentioned by Tarski in a course taken by one of the authors during 1951.

1.4. DEFINITION. $A \times B$ is the Cartesian product of A and B, i.e., the algebra formed from the Cartesian product of the sets A and B by defining the operation componentwise.

1.5. THEOREM. $A \times (B \cup C) \sim A \times B \cup A \times C$.

The proof is immediate.

1.6. DEFINITION. If $x \in A$ we define $x^0 = x$, and for $n > 0, x^n = (x^{n-1})'$. Thus, $x^1 = x', x^2 = (x')'$, etc.

1.7. DEFINITION. x is called a cyclic element of A if $x^a = x$ for some $n > 0$.

1.8. DEFINITION. $A/k = \{x : x \in A$ and x^k is cyclic$\}$. In particular, $A/0$ is the set of all cyclic elements of A.

It is easily seen that unless A/k is empty it is a subalgebra of A.

1.9. DEFINITION. If x, y are in A then we say x is connected to y if and only if for some n, m $x^n = y^m$.

This relation is an equivalence relation (in fact, a congruence relation) and we have:

1.10. DEFINITION. The equivalence classes with respect to the relation of 1.9 are called the components of A, the class to which x belongs being written $C(x)$.

If an algebra has only one component we call it connected. The components are disjoint subalgebras and an algebra is completely characterized by the set of its components. Formally, we have

1.11. THEOREM. *If A, B are unary algebras then $A \sim B$ if and only if the components of A are pairwise isomorphic to the components of B.*

The proof is obtained by defining the isomorphism $f: A \sim B$ as the union of the isomorphisms on the components; and conversely, given f, f restricted to each of the components of A yields an isomorphism onto a component of B. In general, if a sequence of algebras is pairwise isomorphic with another sequence, in some order, we shall write $\{A_i\} \sim \{B_j\}$.

Suppose that A and B are unary algebras, $x \in A$ and $y \in B$, $f: A \sim B$, (that is, f is an isomorphism of A onto B) and $f(x) = y$. If x^r is cyclic, and hence for some p $x^{r+p} = x^r$, then y satisfies the same equation. It follows from this that the image of A/k under f must be B/k, the result holding for all k.

If C and D are unary algebras, $z \in C, w \in D, z^{r+p} = z^r$, and $w^{r+t} = w^r$, then in $C \times D$, with $q = [p, t]$ we have $(z, w)^{r+q} = (z, w)^r$. We collect these results in

1.12. THEOREM. *If $f\colon A \sim B$ then $f\colon A/k \sim B/k$. Moreover, for any C, D $C/k \times D/k \sim (C \times D)/k$.*

In particular, this shows that isomorphisms send the cyclic part of one algebra onto the cyclic part of the image. This fact will be used frequently in the remainder of the paper

2. The cyclic case. Let us now restrict our attention to finite algebras and, in fact, to those finite algebras in which every element is cyclic. We call these cyclic algebras, and a component of a cyclic algebra, a cycle.

It is evident that a cyclic is characterized by the number of its elements, one with p elements being called a p-cycle. It is also clear that a cyclic algebra is determined up to isomorphism by the number of p-cycles for $p = 1, 2, \cdots$. We now want to show directly that for cyclic algebras $A^2 \sim B^2$ implies $A \sim B$.

If x is cyclic then by Definition 1.7 $x^n = x$ for some n, let us write $o(x)$ for the smallest such n and call this the order of x. With A and B cyclic let a_i, b_i, c_i, and d_i be the number of elements of order i in A, B, A^2 and B^2 respectively. Then Lemma 2.1, which follows, is evident, and Lemma 2.2 is quickly obtained by counting, using the fact that $o(x) = r$ and $o(y) = s$ implies $o(x, y) = [r, s]$.

2.1. LEMMA. *$A \sim B$ if and only if for each i $a_i = b_i$.*

2.2. LEMMA. *$c_n = 2 \sum a_i a_j + a_n^2$, in which the sum is extended over all pairs (i, j) with $i < j$ and $[i, j] = n$.*

Suppose now that $A^2 \sim B^2$ and hence for each i, $c_i = d_i$. If A is not isomorphic to B then there is a smallest n for which $a_n \neq b_n$. We have always:

$$c_n = 2 \sum_{\substack{i<j \\ [i,j]=n}} a_i a_j + a_n^2 = 2 \sum_{\substack{i<j<k \\ [i,j]=n}} a_i a_j + 2 \sum_{\substack{i \mid n \\ i \neq n}} a_i a_n + a_n^2$$

and similarly for d_n, since $j = n$ and $[i, j] = n$ implies $i \mid n$. From this we obtain

$$2 \sum_{\substack{i \mid n \\ i=n}} a_i a_n + a_n^2 = 2 \sum_{\substack{i \mid n \\ i \neq n}} b_i b_n + b_n^2$$

since $a_i = b_i$ for any $i < n$. Hence

$$a_n[2 \sum_{\substack{i \mid n \\ 1 \neq n}} a_i] + a_n^2 = b_n[2 \sum_{\substack{i \mid n \\ i \neq n}} b_i] + b_n^2 .$$

But the expressions in brackets are the same since $i < n$ and from this it follows readily that $a_n = b_n$ which is a contradiction. We have shown

2.3. THEOREM. *If A and B are finite cyclic algebras and $A^2 \sim B^2$ then $A \sim B$.*

We do not have cancellation even for finite cyclic algebras (see § 7).

3. Ordering. In § 2 we have seen that the cyclic part of an algebra is well behaved, with respect to the square root problem, and we now turn to the noncyclic part. Consider the class of unary algebras defined by:

3.1. DEFINITION. A unary algebra A will be called basic if it is connected, has exactly one cyclic element, and for each $k \geqq 0$ A/k is finite. $\boldsymbol{B}$ denotes the class of basic algebra.

Notice that in a basic algebra there is an idempotent element, namely the single cyclic element. The rest of this section is devoted to the ordering of $\boldsymbol{B}$ in a useful way; the procedure is somewhat complicated and we have several preliminary definitions.

Suppose $A \in \boldsymbol{B}$ and $x \in A$, let $P(x)$ be the set of all elements of A which precede x; i.e.,

3.2. DEFINITION. $P(x) = \{y : y \in A$ and for some $n, y^n = x\}$.

This set of elements can be turned into a basic algebra by changing the definition of x', setting $x' = x$, and leaving everything else unchanged. The resulting algebra will also be called $P(x)$, and if x is the cyclic element of A, $P(x) = A$.

If $x \in A, A \in \boldsymbol{B}$, and a is the cyclic element of A then in view of the connectivity of A we may make

3.3 DEFINITION. $\deg(x) =$ the smallest integer n such that $x^n = a$.

Notice that for $A \in \boldsymbol{B}, A/k$ is the subalgebra of A consisting of elements with degree less than or equal to k.

3.4. DEFINITION. If A is finite $h(A) = \max\{\deg(x) : x \in A\}$.

3.5. DEFINITION. For $A \in \boldsymbol{B}$, the width of $A = w(a) =$ the number of elements of degree 1.

3.6. DEFINITION. $[A] = \{P(x) : \deg(x) = 1$ and $x \in A$.

$[A]$ is a collection of basic algebras and as mentioned after 1.11 we shall write $[A] \sim [B]$ when the elements of these sets are pairwise isomorphic.

3.7. THEOREM. *$A \sim B$ if and only if $[A] \sim [B]$.*

The proof of this should offer no difficulty since the members of $[A]$ are disjoint.

If A and B are in $\boldsymbol{B}$ then $A/0 \sim B/0$ since each has only a single element. The proof of the following theorem is included in § 8, but the theorem is probably not surprising.

3.8. THEOREM. *If A and B are in $\boldsymbol{B}$ and for each $k \geqq 0$ $A/k \sim B/k$ then $A \sim B$.*

In view of 3.7 we may make the following

3.9. DEFINITION. If A is not isomorphic to B, $e(A, B)$ is the largest integer for which $A/e \sim B/e$.

If A is not isomorphic to B and $e(A,B) = 0$ then $w(A) < w(B)$ or conversely; we order A, B accordingly. If $e(A, B) = 1$ then $w(A) = w(B)$ and $[A]$ and $[B]$ have the same length, but $[A/2] \not\sim [B/2]$. Each member of these sets is an algebra of height $\leqq 1$ and the collection of algebras of height $\leqq 1$ is ordered as above. We may then arrange the collections $[A/2]$, $[B/2]$ in nondecreasing order and compare them lexicographically. Continuing this process yields an ordering of $\boldsymbol{B}$.

The following lemma, together with 3.12 and 3.13 is devoted to a precise statement and proof of the preceding remarks. In the lemma, A and B are in $\boldsymbol{B}$ and the members of $[A]$ and $[B]$ will be assumed ordered by a relation R. We write $[A][R][B]$ to mean that $[A]$ is length-R-lexicographically less than $[B]$ in the following sense:

(i) $[A]$ is shorter than $[B]$, or

(ii) length $[A]$ = length $[B]$ and $[A]$ is lexicographically less that $[B]$ when both are regarded as nondecreasing sequences relative to R (i.e., the members of $[A]$ are indexed so that $A_i \sim A_{i+1}$ or $A_i R A_{i+1}$ for $A_i \in [A]$).

To simplify matters we write = instead of $\sim$.

3.10. LEMMA. *Let R_k, $k \geqq 0$, be a relation satisfying:*

(i) *$(A, B) \in R_k$ implies $e(A, B) \leqq k$.*

(ii) *If $e(A, B) \leqq k$ then either $(A, B) \in R_k$ or $(B, A) \in R_k$ and not both.*

(iii) *$(A, B) \in R_k$ and $(B, C) \in R_k$ implies $(A, C) \in R_k$.*

(iv) *If $e(A, B) \leqq k$, $(A, B) \in R_k$ if and only if*

$$[A/e(A, B) + 1][R_k][B/e(A, B) + 1] .$$

Then there is a unique relation R_{k+1} satisfying the same conditions (with k replaced by $k + 1$) and containing R_k.

Proof. We show first that such an extension is unique. Let (A, B) be in R_{k+1}; then $e(A, B) \leqq k + 1$, and by (iv),

$$[A/e(A, B) + 1][R_{k+1}][B/e(A, B) + 1] .$$

If A_i is in $[A/e(A, B) + 1]$ then $h(A_i) \leqq k + 1$, so if $A_i \neq A_j$, $e(A_i, A_j) \leqq k$ and similarly for $e(B_i, B_j)$ and $e(A_i, B_j)$. This means that $[R_k]$ can be applied. But R_{k+1} contains R_k hence $[R_{k+1}]$ and $[R_k]$ must agree for these sequences and it follows that R_{k+1} is unique.

We now define R_{k+1} by:

3.11. DEFINITION. $R_{k+1} = \{(A, B) : e(A, B) = k + 1 \text{ and } [A/e(A, B)+1] [R_k][B/e(A, B) + 1)\} \cup R_k$.

Properties (i) and (ii) are easily checked. In order to check (iv), suppose that $e(A, B) \leqq k + 1$. If $(A, B) \in R_{k+1}$ then

$$[A/e(A, B) + 1][R_k][B/e(A, B) + 1] ,$$

but $[R_{k+1}]$ agrees with $[R_k]$ whenever both are defined. Conversely, if $e(A, B) \leqq k + 1$ then $[R_{k+1}]$ agrees with $[R_k]$ and the definition implies that $(A, B) \in R_{k+1}$. Finally, to prove (iii) take $e = e(A, B) < e(B, C)$ and hence $< k + 1$. Then $A/e = B/e = C/e$ while $[A/e + 1][R_k][B/e + 1] = [C/e + 1]$ and $(A, C) \in R_k \subseteqq R_{k+1}$. A similar proof is obtained if $e(A, C) > e(B, C)$ or $e(A, C) = e(B, C) < k + 1$. If $e(A, C) = e(B, C) = k + 1$ then $(A, C) \in R_{k+1}$ because of the transitivity of $[R_k]$. This completes the proof of the lemma.

3.12. DEFINITION. $R_0 = \{(A, B) : e(A, B) = 0 \text{ and } w(A) < w(B)\}$.

It is readily seen that R_0 satisfies the conditions of 3.10 For each $k > 0$ let R_{k+1} be the extension of R_k given in 3.10 and let $R = \bigcup R_i$, $k \geqq 0$.

3.13. DEFINITION. $A \leqq B$ if and only if ARB or $A = B$.

It can be shown, using § 3.8, that $\leqq$ is a simple ordering of $\boldsymbol{B}$ (strictly speaking, of the isomorphism types of members of $\boldsymbol{B}$). Two properties of $<$ obtained from sections 3.10–3.13 which we shall use are

3.14. If $w(A) < w(B)$ then $A < B$.

3.15. $A < B$ if and only if $A/e(A, B) + 1 < B/e(A, B) + 1$.

4. Dot product.

4.1. DEFINITION. If A and B are basic algebras then $A \cdot B$ is defined to be the subalgebra of $A \times B$ consisting of all pairs (x, y) for which $\deg(x)$ (in A) $= \deg(y)$ (in B).

The following facts are easily derived.

4.2. $A/k \cdot B/k \doteq (A \cdot B)/k$ (see 1.12).

4.3. $h(A \cdot B) = \min[h(A), h(B)]$.

4.4. $w(A \cdot B) = w(A)w(B)$.

The main theorem on the dot product is the following one and in it the properties of lexicographic order are used without mention. The order between the algebras is the one given in 3.13.

4.5. THEOREM. (i) *If* $A < B$ *and* $h(C) \leqq e(A, B)$ *then* $A \cdot C = B \cdot C$.
(ii) *If* $A < B$ *and* $h(C) > e(A, B)$ *then* $A \cdot C < B \cdot C$.

Proof. (i) is easy, for in this case $A \cdot C = (A/h(C)) \cdot C = (B/h(C)) \cdot C = B \cdot C$. (ii) is proved by induction on $e = e(A, B)$. If $e = 0$ then $A < B$ if and only if $w(A) < w(B)$, but $w(A \cdot C) = w(A)w(C) < w(B)w(C) = w(B \cdot C)$ and hence $A \cdot C < B \cdot C$. Now take $e > 0$, assuming (ii) for smaller values of e. If $A < B$ then $[A/e+1] < [B/e+1]$ and $[A/e] = [B/e]$. Let $[C/e+1] = \{C_1, \cdots, C_p\}$; $h(C_i) \leqq e$, the same holding for the members of $[A/e+1]$ and $[B/e+1]$. If $h(C_i) \leqq e - 1$ then $[A/e+1] \cdot C_i = [B/e+1 \cdot C_i$. If $h(C_i) = e$, then let $A_m < B_m$ be the first pair in which $[A/e+1]$ and $[B/e+1]$ differ; then $A_m \cdot C_i < B_m \cdot C_i$ by the inductive hypothesis, while for $j < m$, $A_j \cdot C_i = B_j \cdot C_i$. Thus $[A/e+1] \cdot C_i < [B/e+1] \cdot C_i$ lexicographically. For any C_i in $[C/e+1]$ then, either $[A/e+1 \cdot C_i = [B/e+1] \cdot C_i$ or $[A/e+1] \cdot C_i < [B/e+1] \cdot C_i$. But $[A/e+1] \cdot [C/e+1]$ is just the ordered union of $[A/e+1] \cdot C_i$, $C_i \in [C/e+1]$, and at least one strict inequality must hold. Hence, $[A/e+1] \cdot [C/e+1] < [B/e+1] \cdot [C/e+1]$ lexicographically, and $A/e+1 \cdot C/e+1 < B/e+1 \cdot C/e+1$, while $A/e \cdot C/e = B/e \cdot C/e (= B \cdot C/e)$ and $A \cdot C < B \cdot C$, by definition of $<$.

4.6. COROLLARY. *If* A, B, C *are infinite and* $A < B$ *then* $A \cdot C < B \cdot C$.

4.7. COROLLARY. *If* A, B, C *are basic algebras and* $A \leqq B$ *then* $A \cdot C \leqq B \cdot C$.

Up to isomorphism, the collection of infinite basic unary algebras with $\cdot$ forms a commutative semigroup which by 4.6 is ordered. This is the semigroup to which we apply the following lemma.

4.8. LEMMA. *If* $\langle S, \cdot, \leqq \rangle$ *if an ordered semigroup*[2], S^* denotes the set of all finite nonempty, nondecreasing sequences in S; for $\{x_i\}$ and $\{y_j\}$ in S^*, $\{x_i\} * \{y_j\}$ is defined as the nondecreasing sequence formed from $\{x_i \cdot y_j\}$; and $\leqq^*$ is the length lexicographic order; then $\langle S^*, *, \leqq^* \rangle$ is an ordered semigroup.

Proof. Suppose $x = \{x_i\}$, $y = \{y_j\}$, $z = \{z_m\}$ are in S^* and $x <^* y$. If length $x <$ length y then length $xz <$ length yz and $xz <^* yz$. If length $x =$ length y there is a $t > 0$ with $x_i = y_i$ for any $i < t$ and $x_t < y_t$. Let $\underline{x} = \{x_i : i < t\}$, $\bar{x} = \{x_i : i \geqq t\}$ and similarly for $\underline{y}$ and $\bar{y}$. The smallest elements in $\bar{y} * z$ and $\bar{y} *$ are formed from x_t, y_t, and z_1.

[2] We use ordered in the sense of Clifford [**2**], i.e., $a < b$ implies $a \cdot c < b \cdot c$.

Therefore, $\bar{x} * z < \bar{y} * z$. Now $\underset{=}{x} = \underset{=}{y}$, $\underset{=}{x} * z = \underset{=}{y} * z$ and we have $x * z = (\underset{=}{x} \cup \bar{x}) * z < (\underset{=}{y} \cup \bar{y}) * z = y * z$. The reason is: inserting equal sequences in two ordered sequences cannot change their order.

5. **Unraveled algebras.** Let A be a finite unary algebra, x a cyclic element of A and $X = C(x)$ (1.10). We associate with x an infinite basic algebra which we think of as "X unraveled backwards, starting at x", and call $W(x)$.

5.1. DEFINITION. $W(x) = \langle X \times I, * \rangle$ with $I = \{0, 1, \cdots\}$

$$(x, 0)^* = (x, 0)$$
$$(x, k)^* = (x', k - 1) \text{ for } k > 0$$
$$(y, k)^* = (y', k) \text{ for } y \neq x .$$

It is not difficult to see that $W(x)$ is a basic algebra, the only cyclic element being $(x, 0)$. $(y, m) \in W(x)$ we still use $\deg(y, m)$ as in 3.3. If x is any cyclic element in a connected algebra A and $y \in A$ then for some n, $y^n = x$; in this context we need

5.2. DEFINITION. $\deg_x(y) = m$ if and only if m is the least nonnegative integer for which $y^m = x$. If y is not in $C(x)$ then $\deg_x y$ is not defined.

If the cyclic part of a connected algebra is a p-cycle we say the algebra is p-cyclic. Lemma 5.3 follows immediately from the definitions.

5.3. *If* $(y, m) \in W(x)$ *and* $C(x)$ *is* p-*cyclic, then* $\deg(y, m) = \deg_x(y) + mp$.

5.4. LEMMA. *If* a *and* x *are in* A, b *and* y *are in* B, $C(a)$ *is* p-*cyclic, and* $C(b)$ *is* q-*cyclic; then in* $A \times B$; $(x, y) \in C(a, b)$ *if and only if* $\deg_a x = \deg_b y \bmod (p, q)$.

Proof. $(x, y) \in C(a, b)$ if and only if for some m, $(x, y)^m = (a, b)$, which is equivalent to $x^m = a$ and $y^m = b$. Such an m exists if and only if there are nonnegative integers r, s with

5.5. $m = \deg_a x + rp = \deg_b y + sq$.

The necessary and sufficient condition for the existence of r, s is that $\deg_a x = \deg_b y \bmod (p, q)$. This completes the proof.

If $(x, y) \in C(a, b)$ and $m = \deg_{(a,b)}(x, y)$ then the integers satisfying 5.5 are unique and will be denoted by

5.6. $r_0 = (\deg_{(a,b)}(x, y) - \deg_a x)/p$, $s_0 = (\deg_{(a,b)}(x, y) - \deg_b y)/q$.

A result on which the rest of the development depends is

5.7. THEOREM. *If* A, B *are finite unary algebras,* $a \in A/0$ *and* $b \in B/0$*; then* $W(a) \cdot W(b) \sim W(a, b)$ *in* $A \times B$.

Proof. Let $C(a)$ be p-cyclic and $C(b)$ be q-cyclic. Take an element $((x, k), (y, m))$ in $W(a) \cdot W(b)$ and using the definition of $\cdot$, and 5.3 obtain:

5.8. $\deg_a x + kp = \deg(x, k) = \deg(y, m) = \deg_b y + mq$.

We have then $\deg_a x \equiv \deg_b y \bmod (p, q)$ and can apply 5.4; yielding: $(x, y) \in C(a, b)$ and

5.9. $\deg_{(a,b)}(x, y) = \deg_a x + r_0 p = \deg_b y + s_0 q$.

Subtracting equation 5.9 from equation 5.8 and dividing by $[p, q]$, it is easily seen that the result is an integer h,

5.10.
$$h = h((x, k), (y, m)) = (\deg(x, k) - \deg_{(a,b)}(x, y))/[p, q]$$
$$= (\deg(y, m) - \deg_{(a,b)}(x, y))/[p, q].$$

Clearly, h is a well defined function and we can now define a function f which we shall show is the required isomorphism,

5.11. $f((x, k), (y, m)) = ((x, y), h)$.

To see that f is one-to-one and onto one need only take $((x, y), h)$ in $W(a, b)$ and solve equation 5.10 for k and m, using 5.8, the solution being unique. It remains to be shown that f commutes with the operations involved. Using $*$ for the operations in the W algebras, and recalling that on $W(a) \cdot W(b)$, $*$ is defined componentwise, let $z = ((x, k), (y, m))$. We want to show that $f(z^*) = [f(z)]^*$ and we need to consider three cases:

(1) $x \neq a$;

(2) $x = a, y = b$, neither k nor $m = 0$; and

(3) $x = a, y = b, k = m = 0$. No other cases are needed, for if $(y, m) = (b, 0)$ then $(x, k) = (a, 0)$, otherwise z would not be in $W(a) \cdot W(b)$. All other possible cases are taken care of by symmetry.

Case 1. $[f(z)]^* = [(x, y), h]^* = [(x', y'), h]$ since $(x, y) \neq (a, b)$, $z^* = [(x', k), (y', m)]$ the value of m being unimportant, and $f(z^*) = [(x', y'), h]$ in which h is calculated from 5.10; $\underline{h} = (\deg(x', k) - \deg_{(a,b)}(x', y'))/[p, q]$. But $\deg(x', k) = \deg(x, k) - 1$ and $\deg_{(a,b)}(x', y') = \deg_{(a,b)}(x, y) - 1$, hence $\underline{h} = h$ and this case is complete.

Case 2. $[f(z)]^* = [(a, b), h]^* = [(a', b'), h - 1]$ and $z^* = [(a', k - 1), (b', m - 1)]$, $f(z^*) = [(a', b'), \underline{h}]$. We calculate h and $\underline{h}$:

$$h[(a, k), (b, m)] = (\deg(a, k) - \deg_{(a,b)}(a, b))/[p, q]$$
$$= (\deg_a a + kp - 0)/[p, q] = kp/[p, q]\,.$$

$$\underline{h}[(a', k-1), (b', m-1)] = (\deg_a a' + (k-1)p - [p, q])/[p, q]$$
$$= (p + (k-1)p - [p, q])/[p, q] = h - 1\,.$$

This completes the proof of the theorem, since Case 3 is trivial.

We now associate with any unary algebra the collection of its W algebras by

5.12. DEFINITION. If A is a unary algebra $W(A) = \{W(a) : a \in A/0\}$.

Since $A/0 \times B/0 \sim A \times B/0$ it follows that $W(A \times B) = \{W(a, b) : (a, b\} \in (A \times B)/0\}$ which is naturally pairwise isomorphic to $\{Wa \cdot Wb : a \in A/0, b \in B/0\}$. This last expression we write $W(A) \cdot W(B)$ instead of $*$ as in Lemma 4.8 This proves

5.13. THEOREM. $W(A \times B) \sim W(A) \cdot W(B)$.

5.14. THEOREM. *If A, B are connected, finite, and A is p-cyclic, B is q-cyclic, then $A \sim B$ if and only if $p = q$ and for some $a \in A_0$, $b \in B_0$, $W(a) \sim W(b)$.*

Proof. One of the implications is clear; for the other, let f be an isomorphism $f\colon W(a) \sim W(b)$ for some $a \in A/0, b \in B/0$. We have seen that isomorphisms preserve degree and they certainly preserve number of predecessors. Thus, $f(a, 0) = (b, 0)$ and since $(a, 1)$ may be characterized as the only element of degree p with infinitely many predecessors $f(a, 1)$ must be the corresponding element of $W(b)$. But $p = q$ so that this element is $(b, 1)$. Moreover, $f\colon A \times 0 \sim B \times 0$ since these are the sets of elements which do not precede $(a, 1)$ and $(b, 1)$ respectively. It then follows immediately that the first coordinate of f is an isomorphism of A onto B.

Notice that in the connected case, the existence of an isomorphism between any $W(a)$ and $W(b)$ is sufficient, (with $p = q$), to insure the isomorphism of A and B. If the two sequences $W(A)$ and $W(B)$ have the same number of elements, then we must have $p = q$ and conversely. This yields

5.15. COROLLARY. *If A, B are finite and connected then $A \sim B$ if and only if $W(A) \sim W(B)$.*

6. Cancellation. We can now apply the preceding results to the cancellation problem for finite unary algebras.

It is readily seen that in any ordered semigroup either of $x \cdot x = y \cdot y$ or $x \cdot z = y \cdot z$ implies $x = y$. The system of infinite basic algebras with $\cdot$ and $\leqq$ is (up to isomorphism) such a semigroup. This system includes W algebras and we apply Lemma 4.8 to the system of finite sequences of W algebras. From these considerations we obtain for finite unary algebras A, B, C, 6.1. LEMMA. *$W(A)\cdot W(A)\sim W(B)\cdot W(B)$ or $W(A)\cdot W(C)\sim W(B)\cdot W(C)$ implies $W(A)\sim W(B)$.*

6.2. LEMMA. *If A, B, and C are finite unary algebras and $A^2\sim B^2$ or $A\times C\sim B\times C$, then $W(A)$ is pairwise isomorphic to $W(B)$.*

Proof. From 5.13 $W(A)\cdot W(A)\sim W(A^2)\sim W(B^2)\sim W(B)\cdot W(B)$, and similarly for $A\times C\sim B\times C$. The lemma follows by applying 6.1.

6.3. THEOREM. *If A, B, C, are connected finite unary algebras and $A^2\sim B^2$ or $A\times C\sim B\times C$, then $A\sim B$.*

Proof. The theorem follows from 6.1 and 5.15.

If an algebra is not connected let us say that it is *pure* if all the components are p-cyclic for some fixed p and use the term p-cyclic for pure algebras as well as connected algebras.

6.4. THEOREM. *If A, B, and C are pure finite unary algebras and $A^2\sim B^2$; or $A\times C\sim B\times C$ and $A/0\sim B/0$, then $A\sim B$.*

Proof. From 6.2 we obtain $W(A)\sim W(B)$. Since we have unique square roots for cyclic algebras (2.3) and from the hypothesis in the other case we see that the cyclic structure of both A and B is the same. They are both pure and consequently there is an integer p such that all of the cycles of A and B are p-cycles. Given $W(A)$ and $W(B)$ and p we can put each of the elements of $W(A)$ and $W(B)$ back together again—see 5.14 and its proof. This will yield the components of A and B each repeated p times, hence A and B must be isomorphic.

If A is a finite algebra we can write $A=A_1\cup A_2\cup\cdots\cup A_n$ in which each A_i is a pure subalgebra (a collection of components) and the decomposition is maximal in the sense that $A_i\cup A_j$ is not pure. This decomposition is clearly unique up to order and the integer n is called the length of A. We can now complete the solution of the square root problem with

6.5. THEOREM. *If A and B are finite unary algebras and $A^2\sim B^2$ then $A\sim B$.*

The proof is by induction on n, the length of A (and in view of 2.3, also the length of B). For $n=1$ the conclusion is part of § 6.4. For $n>1$, assuming the result for smaller n, we know from 2.3 that the cyclic structure of A and B is the same. Hence, in the decomposition into pure subalgebras, $A=A_1\cup A_2\cup\cdots\cup A_n$, $B=B_1\cup B_2\cup\cdots\cup B_n$, we may assume that both A_i and B_i consist of p_i-cyclic algebras with $A_i/0\sim B_i/0$, and $p_1<p_2<\cdots<p_n$.

It is not possible for A_i to be isomorphic to B_i for all $i\neq j$ and

$A_j \not\sim B_j$. For we know that $W(A) \sim W(B)$ and if the above condition were to hold then there would be a one-to-one correspondence between the W algebras obtained from A_i $i \neq j$ and B_i $i \neq j$ and hence between those obtained from A_j and B_j. But this and the fact that A_j and B_j have the same cyclic structure implies that $A_j \sim B_j$.

Hence, if $A \not\sim B$ then for some smallest $j < n$ $A_j \not\sim B_j$. If I is the set of indices for which $p_i, i \in I$ divides p_j; let $P = \bigcup \{A_{p_i}: i \in I\}$, $Q = \bigcup \{B_{p_i}: i \in I\}$ and define A'', B'' so that $A = P \cup A''$, $B = Q \cup B''$. Then $A^2 = P^2 \cup PA'' \cup A''P \cup A''^2$. But for integers m, p, q; $[p, q]$ divides m if and only if p divides m and q divides m. Hence the components of P^2 are those which are k-cyclic with k dividing p_j. From this it follows that when the isomorphism given between A^2 and B^2 is restricted to P^2, it maps P^2 isomorphically onto Q^2. Hence $P \sim Q$ since P and Q are shorter than A and B. This implies that $A_j \sim B_j$ which is a contradiction and completes the proof.

7. Summary. We have seen that all finite unary algebras have unique (if any) square roots, and that in some cases $A \times C \sim B \times C$ implies $A \sim B$. This last implication does not hold in general for finite algebras. The simplest example of its failure is: A a 2-cycle and B two 1-cycles. In this case $A \times A \sim B \times A$ and $A \not\sim B$.[3] It is easy to see that if A is a k-cycle and B is any collection of p_i-cycles with $p_i \mid k$ and $\sum p_i = k$ then the same situation obtains.

In view of 6.2 it would be reasonable to conjecture that if the cyclic structure of A and B is the same then $A \times C \sim B \times C$ implies $A \sim B$. Whenever there is only one way of putting the W algebras back together, as in the pure case when the size of the cyclic parts is determined, we can obtain cancellation.

In the infinite case, it is known that an algebra need not have a unique square root, a simple example being: A and B free unary algebras with k and l generators, $k \neq l$. Then $A^2 \sim B^2$ but $A \not\sim B$. It seems likely that the results of this paper could be generalized to suitable classes of infinite algebras such as basic algebras, locally finite connected algebras, or algebras satisfying some kind of descending chain condition. We have not as yet attempted such generalizations.

8. Proof of Theorem 3.8.

THEOREM. *If A, B are basic algebras with $A/k \sim B/k$ for all $k \geqq 0$, then $A \sim B$.*

Proof. If A, B are finite the theorem is trivial; we assume that they are infinite. For each k let I_k be the set of isomorphisms of A/k

[3] This example is attributed to B. Jónsson by Birkhoff [**1**], p. 96, ex. 4.

onto B/k. I_k is not empty, and since A_k is finite, so is I_k. If $m > k$ and $\phi \in I^m$ then $(\phi)A/k(\phi$ restricted to $A/k) \in I_k$; hence, some members of I_k must be the restrictions of infinitely many isomorphisms of greatet degree, and in fact of arbitrarily great degree. Let E_k be the subser of I_k consisting of isomorphisms of this type. If $\phi \in E_k$ there is a member of E_{k+1}, ψ, with $(\psi)A/k = \phi$, let E_ϕ be the subset of E_{k+1} satisfying this condition. By the axiom of choice there is a function f which selects for each ϕ in E_k an $f(\phi)$ in E_ϕ.

We now define ϕ_0: $A/0 \sim B/0$ by $\phi_0(a) = b$ (this is the only member of I_0); and for $k > 0$ $\phi_k = f(\phi_{k-1})$. We show that $\phi = \bigcup \phi_k$ is an isomorphism, $\phi : A \sim B$. In the sequence $\phi_0, \phi_1, \cdots$ each ϕ is the restriction of ϕ_{i+1} to A/i so that ϕ is a function. Since $\bigcup A/i$ and each A/i is the domain of ϕ_i the domain of ϕ is A. If $x \in A$ then $x \in A/i$ for some i $\phi(x') = \phi_i(x') = \phi_i(x)' = \phi(x)'$. It is equally easy to see that ϕ is one-to-one and onto B, and consequently is an isomorphism.

References

1. Garrett Birkhoff, *Lattice Theory*, Amer. Math. Soc. Colloquium Publications, vol. 25, rev. ed., 1948.

2. A. H. Clifford, *Totally ordered commutative semigroups*, Bull. Amer. Math. Soc., **64** (1958), 305-316.

3. B. Jónsson and A. Tarski, *Direct decompositions of finite algebraic systems*, Notre Dame Mathematical Lectures, No. 5, (1947), 64 pp.

Fresno State College

ON UNIVALENCE OF A CONTINUED FRACTION

E. P. MERKES AND W. T. SCOTT

1. Introduction. For a fixed positive integer α let K_α denote the class of functions $f(z)$ which are regular at $z = 0$ and which have C-fraction expansions of the form

$$(1.1) \qquad f(z) \sim \frac{z}{1+} \frac{a_1 z^\alpha}{1+} \frac{a_2 z^\alpha}{1+} \cdots \frac{a_n z^\alpha}{+\ 1+} \cdots, \; |a_n| \leqq 1/4 .$$

From an elementary convergence theorem for continued fractions [4, p.42], it follows that each function of the class K_α is regular for $|z| < 1$. This and the one-to-one correspondence between C-fractions and power series [4, p. 400] permit a replacement of the correspondence symbol in (1.1) by equality for $|z| < 1$.

The purpose of this paper is to determine for K_α the radius of univalence, $U(\alpha)$, and bounds for the starlike radius, $S(\alpha)$, and the radius of convexity, $C(\alpha)$. In the case of S-fractions it was shown by Thale [3] that $U(1) \geqq 12\sqrt{2}-16$ and Perron [2] established the fact that actual equality holds. This result is a special case of Theorem 2.1 whose proof employs value region techniques similar to those used by Thale and Perron. Moreover, the result $S(1) \geqq 8/9$ in [3] is improved in Theorem 4.2.

The developments in this depend on the following value region theorem which is an immediate consequence of a result of Paydon and Wall [1]:

THEOREM 1.1. *If* $f(z) \in K_\alpha$ *and* $|z|^\alpha = \rho^\alpha \leqq 4r(1-r)$, $0 \leqq r \leqq 1/2$, *then*

$$(1.2) \qquad \left| \frac{f(z)}{z} - \frac{1}{1-r^2} \right| \leqq \frac{r}{1-r^2} .$$

Moreover, for $z = \sqrt[\alpha]{4r(1-r)}\, e^{im\pi/\alpha}$, $(m = 1, 2, \cdots, \alpha)$, *there is a value of* $f(z)/z$ *on the boundary of the disc* (1.2) *if and only if there exists a* φ, $0 \leqq \varphi < 2\pi$, *such that* $f(z) \equiv f(z; \varphi)$, *where*

$$(1.3) \qquad f(z;\varphi) = \frac{z}{1+} \frac{\frac{1}{4} e^{i\varphi} z^\alpha}{1\ +} \frac{\frac{1}{4} z^\alpha}{1+} \cdots \frac{\frac{1}{4} z^\alpha}{+\ 1+} \cdots .$$

2. Determination of $U(\alpha)$. For $f(z) \in K_\alpha$ and for a fixed positive integer n put

Received April 27, 1959, and in revised form July 31, 1959.

$$f_{0,n}(z) = z\,,$$
$$f_{p+1,n}(z) = \frac{z}{1 + a_{n-p}z^{\alpha-1}f_{p,n}(z)}\,, \qquad (p = 0, 1, \cdots, n-1)\,, \tag{2.1}$$

where the numbers a_j are the coefficients in the C-fraction expansion (1.1) of $f(z)$. It is easily seen that $f_{n,n}(z)$ is the approximant of (1.1) of order $n + 1$, and that $f_{p,n}(z) \in K_\alpha$ for each p.

For non-negative integers s, t, and for non-zero numbers z_1, z_2, (2.1) may be used to show that

$$\begin{aligned} &z_1^s z_2^t f_{p+1,n}(z_1) - z_1^t z_2^s f_{p+1,n}(z_2) \\ &= \frac{f_{p+1,n}(z_1)f_{p+1,n}(z_2)}{z_1 z_2}\{z_1^{s+1}z_2^t - z_1^t z_2^{s+1} - a_{n-p}[z_1^{t+\alpha-1}z_2^{s+1}f_{p,n}(z_1) \\ &\quad - z_1^{s+1}z_2^{t+\alpha-1}f_{p,n}(z_2)]\}\,, \qquad (p = 0, 1, \cdots, n-1)\,. \end{aligned} \tag{2.2}$$

This identity plays a fundamental role in the proof of the following theorem.

THEOREM 2.1. *The radius of univalence of K_α is given by*

$$U(2) = 2\sqrt{2}/3\,,$$
$$[U(\alpha)]^\alpha = \left[\frac{6\sqrt{\alpha^2 - 2\alpha + 9} - 2(\alpha + 7)}{(\alpha - 2)^2}\right], \qquad (\alpha = 1, 3, 4, \cdots)\,. \tag{2.3}$$

There is no larger region, containing the disc $|z| < U(\alpha)$, in which all functions of K_α are univalent.

Proof. For $f(z) \in K_\alpha$ and for a fixed positive odd integer $n=2m+1$ it follows from (2.2) that

$$\begin{aligned} &f_{n,n}(z_1) - f_{n,n}(z_2) \\ &= \frac{f_{n,n}(z_1)f_{n,n}(z_2)}{z_1 z_2}\{z_1 - z_2 - a_1[z_1^{\alpha-1}z_2 f_{n-1,n}(z_1) - z_1 z_2^{\alpha-1}f_{n-1,n}(z_2)]\}\,. \end{aligned} \tag{2.4}$$

Repeated application of (2.2) yields

$$\begin{aligned} &a_1[z_1^{\alpha-1}z_2 f_{n-1,n}(z_1) - z_1 z_2^{\alpha-1}f_{n-1,n}(z_2)] \\ &= \sum_{j=1}^{m+1}(z_1 z_2)^{(j-1)\alpha+1}(z_1^{\alpha-1} - z_2^{\alpha-1})\prod_{p=1}^{2j-1}a_p\frac{f_{n-p,n}(z_1)f_{n-p,n}(z_2)}{z_1 z_2} \\ &\quad - \sum_{j=1}^{m}(z_1 z_2)^{j\alpha}(z_1 - z_2)\prod_{p=1}^{2j}a_p\frac{f_{n-p,n}(z_1)f_{n-p,n}(z_2)}{z_1 z_2}\,. \end{aligned} \tag{2.5}$$

For z_1 and z_2 in the disc $|z| < 1$, r can be chosen with $0 < r < 1/2$ such that $|z_i|^\alpha \leqq 4r(1-r)$, $(i = 1, 2)$, and by Theorem 1.1, $|f_{p,n}(z_i)/z_i| \leqq 1/(1-r)$, $(i = 1, 2; p = 0, 1, \cdots, n)$. When the triangle inequality is applied to the right member of (2.5) and the indicated bounds are used, there results

$$\begin{aligned}
&| a_1 | \, | z_1^{\alpha-1} z_2 f_{n-1,n}(z_1) - z_1 z_2^{\alpha-1} f_{n-1,n}(z_2) | \\
&\leqq | z - z_2 | \left[\sum_{j=1}^{m+1} (\alpha - 1)\left(\frac{r}{1-r}\right)^{2j-1} + \sum_{j=1}^{m} \left(\frac{r}{1-r}\right)^{2j} \right] \\
&< | z_1 - z_2 | \frac{r}{1-2r}[\alpha - 1 - (\alpha - 2)r] \, .
\end{aligned}$$

This inequality and (2.4) give

$$\begin{aligned}
&| f_{n,n}(z_1) - f_{n,n}(z_2) | \\
&\geqq \frac{| f_{n,n}(z_1) f_{n,n}(z_2) |}{| z_1 z_2 |} | z_1 - z_2 | \left\{ 1 - \frac{r[\alpha - 1 - (\alpha - 2)r]}{1-2r} \right\} \, . \qquad (2.6)
\end{aligned}$$

Since Theorem 1.1 shows that neither of the factors $| f_{n,n}(z_i)/z_i |$, $(i=1, 2)$, is zero, it follows from (2.6) that $f_{n,n}(z_1) \neq f_{n,n}(z_2)$ for $z_1 \neq z_2$ if r is such that $1 - 2r > r[\alpha - 1 - (\alpha - 2)r]$. This is equivalent to the condition $r < r_0(\alpha)$ where

$$r_0(2) = 1/3$$

$$r_0(\alpha) = \frac{\alpha + 1 - \sqrt{\alpha^2 - 2\alpha + 9}}{2(\alpha - 2)} \, , \qquad (\alpha = 1, 3, 4, \cdots) \, ,$$

and it is easily seen that $f_{2m+1,2m+1}(z)$ is univalent for $| z |^\alpha < [U(\alpha)]^\alpha = 4r_0(\alpha)[1 - r_0(\alpha)]$.

If the function $f(z)$ has a non-terminating C-fraction (1.1), the univalence of $f(z)$ for $| z | < U(\alpha)$ is an immediate consequence of the fact that $f(z)$ is the uniform limit of its sequence of even approximants, $f_{2m+1,2m+1}(z)$, for $| z | \leqq \rho < 1$. The case where $f(z)$ has a C-fraction expansion (1.1) terminating with an odd number of partial quotients may be reduced to the previously considered case for even approximants by adding a partial quotient, $a_{2m} z^\alpha / 1$ with $a_{2m} = 0$, and noting that $f_{2m-1,2m-1}(z) = f_{2m,2m}(z)$ in this case.

In order to complete the proof that the radius of univalence of K_α is the value $U(\alpha)$ given in (2.3), it suffices to exhibit a function of K_α which is not univalent in $| z | < \rho$ for any $\rho > U(\alpha)$. Such a function is the function $f(z, \pi)$ of (1.3), that is,

$$f(z, \pi) = \frac{2z}{3 - \sqrt{1 + z^\alpha}} \, ,$$

where the branch of the radical with positive real part for $| z | < 1$ is used. This function is not univalent at the points $e^{im\pi/\alpha} U(\alpha)$, $(m = 1, 2, \cdots, \alpha)$, where its derivative vanishes.

The final statement in Theorem 2.1 may be verified by applying to the function $f(z, \pi)$ the observation that, for every real θ, $e^{-i\theta} f(e^{i\theta} z) \in K_\alpha$

whenever $f(z) \in K_\alpha$.

3. A covering theorem. The value region inequality (1.2) can be rewritten as

$$\left|\frac{f(z)}{z} - \frac{4}{2+\rho^\alpha+2\sqrt{1-\rho^\alpha}}\right| \leqq \frac{2(1-\sqrt{1-\rho^\alpha})}{2+\rho^\alpha+2\sqrt{1-\rho^\alpha}}, \tag{3.1}$$

where $|z| = \rho$ and $f(z) \in K_\alpha$. Thus for $|z| = \rho$ the following inequalities, which provide a means of comparison between K_α and various classes of univalent functions, are obtained:

$$\frac{2}{3-\sqrt{1-\rho^\alpha}} \leqq \Re e\left\{\frac{f(z)}{z}\right\} \leqq \frac{2}{1+\sqrt{1-\rho^\alpha}}, \tag{3.2}$$

$$\left|\Im m\frac{f(z)}{z}\right| \leqq \frac{2(1-\sqrt{1-\rho^\alpha})}{2+\rho^\alpha+2\sqrt{1-\rho^\alpha}}, \tag{3.3}$$

$$\frac{2\rho}{3-\sqrt{1-\rho^\alpha}} \leqq |f(z)| \leqq \frac{2(1-\sqrt{1-\rho^\alpha})}{\rho^{\alpha-1}}, \tag{3.4}$$

$$\left|\arg\frac{f(z)}{z}\right| \leqq \arcsin\frac{1-\sqrt{1-\rho^\alpha}}{2}. \tag{3.5}$$

Each of the inequalities (3.2)-(3.5) is sharp. This fact follows at once from Theorem 1.1 since equality in any one of (3.2)-(3.5) depends on the attainment by $f(z)/z$ of a suitable boundary value for the disc (3.1) or (1.2).

The following theorem is an immediate consequence of (3.4) and Theorem 2.1:

THEOREM 3.1. *If $f(z) \in K_\alpha$, then the image of $|z| < U(\alpha)$ by $w = f(z)$ contains the disc*

$$|w| < \frac{2U(\alpha)}{3-\sqrt{1-[U(\alpha)]^\alpha}}, \tag{3.6}$$

and is contained in the disc

$$|w| < 2\frac{1-\sqrt{1-[U(\alpha)]^\alpha}}{[U(\alpha)]^{\alpha-1}}. \tag{3.7}$$

These results are sharp.

4. A lower bound for $S(\alpha)$. An upper bound for $S(\alpha)$, the starlike radius for the class K_α, is evidently the value $U(\alpha)$ determined in § 2. In this section a lower bound for $S(\alpha)$ is found by determining a number

$\rho_1(\alpha)$ such that every function of K_α is starlike in the disc $|z| < \rho_1(\alpha)$.

LEMMA 4.1. *If* $f(z) \in K_\alpha$ *and* $|a| \leqq 1/4$, *then*

$$w(z) = -\frac{az^{\alpha-1}f(z)}{1 + az^{\alpha-1}f(z)} \tag{4.1}$$

satisfies

$$\left| w - \frac{r^2}{1-r^2} \right| \leqq \frac{r}{1-r^2} \tag{4.2}$$

whenever $|z|^\alpha \leqq 4r(1-r)$, $0 \leqq r \leqq 1/2$.

Proof. The lemma is obvious when $a = 0$. For $0 < |a| \leqq 1/4$, (4.1) yields

$$\frac{f(z)}{z} = \frac{1}{az^\alpha} \cdot \frac{-w(z)}{1 + w(z)} ,$$

and the desired result is easily obtained by applying the inequality $|f(z)/z| \leqq 1/(1-r)$, which is a consequence of Theorem 1.1.

LEMMA 4.2. *If* α *is a positive integer and if for fixed* r, $0<r<1/2$, c *and* d *are numbers such that*

$$0 \leqq c \leqq \frac{1 + (\alpha-2)r^2}{1-2r^2} , \quad 0 < d = \frac{1 + (\alpha-2)r}{1-2r} - c , \tag{4.3}$$

then $\sigma = 1$ *satisfies*

$$|\sigma - c| \leqq d . \tag{4.4}$$

Moreover, if w *is a parameter satisfying* (4.2) *and if* σ_0 *satisfies* (4.4), *then* σ_1 *satisfies* (4.4) *where*

$$\sigma_1 = 1 + w(\sigma_0 + \alpha - 1) . \tag{4.5}$$

Proof. It is obvious that $1 - c \leqq d$ holds for all r, $0 < r < 1/2$, and that $-d \leqq 1 - c$ holds provided

$$c \leqq \frac{2 + (\alpha - 4)r}{2(1-2r)} .$$

The fact that $\sigma = 1$ satisfies (4.4) may be verified by noting that the upper bound of c in this last inequality exceeds the upper bound on c in (4.3) for all r, $0 < r < 1/2$.

The proof of the second statement is obtained by using (4.2), (4.3), (4.4), (4.5), and the triangle inequality to show that

$$\begin{aligned}|\sigma_1 - c| &\leqq \left|1 - c + \frac{(c+\alpha-1)r^2}{1-r^2}\right| \\ &\quad + (c+\alpha-1)\left|w - \frac{r^2}{1-r^2}\right| + |w||\sigma_0 - c| \\ &\leqq \frac{1+(\alpha-2)r^2-(1-2r^2)c}{1-r^2} + \frac{(c+\alpha-1)r}{1-r^2} + \frac{rd}{1-r^2} = d\,.\end{aligned}$$

LEMMA 4.3. *If* (4.3) *holds for* $0 < r < 1/2$, *there is a value of* c *satisfying* $c \geqq d$ *if and only if* $0 < r \leqq r_1(\alpha)$, *where* $r_1(\alpha)$ *is the smallest positive root of*

$$1 - (\alpha+2)r + 2(\alpha-1)r^2 - 2(\alpha-2)r^3 = 0\,. \tag{4.6}$$

Proof. By (4.3) the inequality $c \geqq d$ holds if and only if

$$\frac{1+(\alpha-2)r^2}{1-2r^2} \geqq \frac{1+(\alpha-2)r}{2(1-2r)}\,,$$

which is equivalent to the statement that the left member of (4.6) is nonnegative. Clearly $r_1(\alpha) < 1/2$.

THEOREM 4.1. *If* $f(z) \in K_\alpha$ *and* c, d *satisfy* (4.3), *where* $|z|^\alpha = \rho^\alpha \leqq 4r(1-r)$, *then*

$$\left|z\frac{f'(z)}{f(z)} - c\right| \leqq d\,. \tag{4.7}$$

Proof. For the functions $f_{p,n}(z)$ of (2.1) put

$$\sigma_{p,n} = z\frac{f'_{p,n}}{f_{p,n}}\,, \quad w_{p,n} = -\frac{a_{n-p}z^{\alpha-1}f_{p+1,n}}{1 + a_{n-p}z^{\alpha-1}f_{p+1,n}}\,,$$

and note by differentiation that $\sigma_{p+1,n} = 1 + w_{p,n}(\sigma_{p,n} + \alpha - 1)$. For $|z| = \rho$ inductive application of Lemmas 4.1 and 4.2 shows that (4.7) holds for $f_{n,n}$, and the validity of (4.7) in this case for $|z| \leqq \rho$ follows from the maximum property for harmonic functions. Inasmuch as $f_{n,n}$ is the $(n+1)$th approximant of (1.1) the theorem holds for functions of K_α having terminating C-fraction expansions. The validity of the theorem in the case of non-terminating C-fractions (1.1) is an immediate consequence of the uniform convergence of $f_{n,n}$ to f on any closed subset of $|z| < 1$.

THEOREM 4.2. *The starlike radius of* K_α *satisfies* $S(\alpha) \geqq \rho_1(\alpha)$ *where*

$[\rho_1(\alpha)]^\alpha = 4r_1(\alpha)[1 - r_1(\alpha)]$ *and where* $r_1(\alpha)$ *is the smallest positive root of* (4.6).

Proof. For $r \leqq r_1(\alpha)$ Lemma 4.3 shows that Theorem 4.1 can be applied to any function $f(z) \in K_\alpha$ with $c \geqq d$, and hence that

$$\Re z \frac{f'(z)}{f(z)} \geqq 0, \quad |z| \leqq \rho_1(\alpha) .$$

Since this inequality insures that $f(z)$ is starlike for $|z| < \rho_1(\alpha)$ the proof is complete.

In particular, $r_1(1) = (\sqrt{3} - 1)/2$ and $S(1) \geqq 4\sqrt{3} - 6$ which improves the lower bound of 8/9 obtained for $S(1)$ in [3].

5. A lower bound for $C(\alpha)$. It is clear that $S(\alpha)$ and $U(\alpha)$ are upper bounds for $C(\alpha)$, the radius of convexity of K_α. In this section a lower bound for $C(\alpha)$ is found by determining a number $\rho_2(\alpha)$ such that every function of K_α is convex for $|z| < \rho_2(\alpha)$.

LEMMA 5.1. *Let* α *denote a positive integer and let* $r_2(\alpha)$ *be the smallest positive root of the equation*:

$$\begin{aligned}(5.1)\qquad 1 - (\alpha^2 + 2\alpha + 6)r + 6(\alpha^2 + \alpha + 2)r^2 - 4(3\alpha^2 + 2)r^3 \\ + 12(\alpha - 1)\alpha r^4 - 4\alpha(\alpha - 2)r^5 = 0 .\end{aligned}$$

If for fixed r, $0 < r \leqq r_2(\alpha)$, σ_0 *and* σ_1 *are numbers which satisfy*

$$(5.2)\qquad |\sigma_0 - c| \leqq d, \ |\sigma_1 - c| \leqq d ,$$

where

$$(5.3)\qquad \frac{1 + (\alpha - 2)r}{2(1 - 2r)} \leqq c \leqq \frac{1 + (\alpha - 2)r^2}{1 - 2r^2}, \quad d = \frac{1 + (\alpha - 2)r}{1 - 2r} - c ,$$

and if

$$(5.4)\qquad \gamma_1 = 2(\sigma_1 - 1) + \frac{\sigma_1 - 1}{\sigma_1}\left[\gamma_0 \frac{\sigma_0}{\sigma_0 + \alpha - 1} + (\alpha - 1)\frac{2\sigma_0 + \alpha - 2}{\sigma_0 + \alpha - 1}\right],$$

then $|\gamma_0| \leqq 1$ *implies* $|\gamma_1| \leqq 1$.

Proof. For $0 < r < r_1(\alpha)$, where $r_1(\alpha)$ is as determined in Theorem 4.2, $0 < d < c$ and

$$c^2 - d^2 - c \leqq -\frac{\alpha r^2[(\alpha - 1) - 2(\alpha - 2)r + 2(\alpha - 2)r^2]}{(1 - 2r)^2(1 - 2r^2)} \leqq 0 .$$

Thus by (5.2)

$$\left|\frac{\sigma_1 - 1}{\sigma_1} - \frac{c^2 - d^2 - c}{c^2 - d^2}\right| \leqq \frac{d}{c^2 - d^2}$$

and it follows that

$$\left|\frac{\sigma_1 - 1}{\sigma_1}\right| \leqq \frac{1}{c - d} - 1 .$$

Similarly, (5.2) can be used to show that

$$\left|\frac{\sigma_0}{\sigma_0 + \alpha - 1}\right| \leqq \frac{c + d}{c + d + \alpha - 1} ,$$

$$\left|(\alpha - 1)\frac{2\sigma_0 + \alpha - 2}{\sigma_0 + \alpha - 1}\right| \leqq (\alpha - 1)\frac{2(c + d) + \alpha - 2}{c + d + \alpha - 1} .$$

For $|\gamma_0| \leqq 1$ application to (5.4) of the triangle inequality, (5.2) and the bounds determined above lead to the inequality

$$(5.5) \quad |\gamma_1| \leqq 2(c + d - 1) + \left[\frac{1}{c - d} - 1\right]\frac{(2\alpha - 1)(c + d) + (\alpha - 1)(\alpha - 2)}{c + d + \alpha - 1} .$$

The desired inequality, $|\gamma_1| \leqq 1$, will hold for those values of $r < r_1(\alpha)$ for which the right member of (5.5) does not exceed 1, or equivalently, for which

$$(5.6) \quad c - d \geqq \frac{(2\alpha - 1)(c + d) + (\alpha - 1)(\alpha - 2)}{(2\alpha - 1)(c + d) + (\alpha - 1)(\alpha - 2) + [3 - 2(c + d)][c + d + \alpha - 1]} = D .$$

Since $2c = (c + d) + (c - d)$, (5.3) shows that the existence of a value of c satisfying (5.6) is insured for all $r < r_1(\alpha)$ for which

$$(5.7) \qquad 2\frac{1 + (\alpha - 2)r^2}{1 - 2r^2} \geqq (c + d) + D .$$

This last inequality is equivalent to the requirement that the polynomial in the left member of (5.1) be non-negative.

The proof of the lemma will be completed by establishing the existence of a smallest positive zero, $r_2(\alpha)$ of (5.1) for which $r_2(\alpha) < r_1(\alpha)$. Since the equation (4.7) determining $r_1(\alpha)$ is equivalent to

$$2\frac{1 + (\alpha - 2)r^2}{1 - 2r^2} = c + d ,$$

and since $D > 0$ for $r = r_1(\alpha)$, it follows that (5.7) fails to hold for $r = r_1(\alpha)$. The desired conclusion about $r_2(\alpha)$ is then easily obtained by noting that (5.7) holds with strict inequality for $r = 0$.

THOREM 5.1. *The radius of convexity of K_α satisfies*

$$[C(\alpha)]^\alpha \geqq 4r_2(\alpha)[1 - r_2(\alpha)] = [\rho_2(\alpha)]^\alpha \tag{5.8}$$

where $r_2(\alpha)$ is the smallest positive root of (5.1)

Proof. For the functions $f_{p,n}(z)$ of (2.1) put

$$\sigma_{p,n} = z\frac{f'_{p,n}}{f_{p,n}}, \qquad \gamma_{p,n} = z\frac{f''_{q,n}}{f'_{p,n}}.$$

It is easily verified from (2.1) that

$$\gamma_{p+1} = 2(\sigma_{p+1} - 1) + \frac{\sigma_{p+1} - 1}{\sigma_{p+1}}\left[\frac{\gamma_p\sigma_p}{\sigma_p + \alpha - 1} + (\alpha - 1)\frac{2\sigma_p + \alpha - 2}{\sigma_p + \alpha - 1}\right]$$

where the subscript n has been omitted. Theorem 4.1 and the fact that $\gamma_{0,n} = 0$ show that the hypotheses of Lemma 5.1 are satisfied, and inductive application of the lemma yields $|\gamma_{n,n}| \leqq 1$. It follows that

$$\Re e[1 + \gamma_{n,n}] \geqq 0\ , \quad |z| \leqq \rho_2(\alpha)\ ,$$

which insures the convexity of the $(n + 1)$th approximant of any C-fraction (1.1) for $|z| < \rho_2(\alpha)$, and the proof of the theorem may be completed, as in Theorem 4.1, by reference to uniform convergence.

It is found that $\rho_2(1) > .641$. An upper bound for $C(\alpha)$ can be obtained by finding for the function $f(z, \pi)$ of (1.3) the zeros of $zf''(z, \pi) + f'(z, \pi)$ with smallest modulus. For $\alpha = 1$ this smallest modulus is approximately .707.

REFERENCES

1. J. F. Paydon and H. S. Wall, *The continued fraction as a sequence of linear transformations,* Duke Math. J., **9** (1942), 360–372.

2. O. Perron, *Uber ein Schlichtheitsschranke von James S. Thale,* S.-B. Math.-Nat. Kl. Bayer. Akad. Wiss. 1956.

3. J. S. Thale, *Univalence of continued fractions and Stieltjes transforms,* Proc. Amer. Math. Soc., **7** (1956), 232–244.

4. H. S. Wall, *Analytic theory of continued fractions,* Van Nostrand, New York, 1948.

DE PAUL UNIVERSITY
NORTHWESTERN UNIVERSITY

ASYMPTOTIC PROPERTIES OF DERIVATIVES OF STATIONARY MEASURES

SHU-TEH C. MOY

1. Introduction. Let X be a non-empty set and $\mathscr{S}$ be a σ-algebra of subsets of X. Consider the infinite product space $\Omega = \prod_{n=-\infty}^{\infty} X_n$ where $X_n = X$ for $n = 0, \pm 1, \pm 2, \cdots$ and the infinite product σ-algebra $\mathscr{F} = \prod_{n=-\infty}^{\infty} \mathscr{S}_n$ where $\mathscr{S}_n = \mathscr{S}$ for $n = 0, \pm 1, \pm 2, \cdots$. Elements of Ω are bilateral infinite sequences $\{\cdots, x_{-1}, x_0, x_1, \cdots\}$ with $x_n \in X$. Let us denote the elements of Ω by w. If $w = \{\cdots, x_{-1}, x_0, x_1, \cdots\}$ x_n is called the nth coordinate of w and shall be considered as a function on Ω to X. Let T be the shift transformation on Ω to Ω: the nth coordinate of Tw is equal to the $n + 1$th coordinate of w. For any function g on Ω, Tg is the function defined by $Tg(w) = g(Tw)$ so that $Tx_n = x_{n+1}$ for any integer n. We shall consider two probability measures μ, ν defined on $\mathscr{F}$. For $n = 1, 2, \cdots$ let $\Omega_n = \prod_{i=1}^{n} X_i$ where $X_i = X, i = 1, 2 \cdots, n$ and $\mathscr{F}_n = \prod_{i=1}^{n} \mathscr{S}_i$ where $\mathscr{S}_i = \mathscr{S}, i = 1, 2, \cdots, n$. Then $\Omega_1 = X$ and $\mathscr{F}_1 = \mathscr{S}$. Let $\mathscr{F}_{m\,n}, m \leqq n, n = 0, \pm 1, \pm 2, \cdots$, be the σ-algebra of subsets of Ω consisting of sets of the form

$$[w = \{\cdots, x_{-1}, x_0, x_1 \cdots\}: (x_m, x_{m+1}, \cdots, x_n) \in E]$$

Where $E \in \mathscr{F}_{n-m+1}$. Then $\mathscr{F}_{m\,n} \subset \mathscr{F}_{m\,n+1} \subset \mathscr{F}$. Let $\mu_{m\,n}, \nu_{m\,n}$ be the contractions of μ, ν, respectively to $\mathscr{F}_{m\,n}$. If $\nu_{m\,n}$ is absolutely continuous with respect to $\mu_{m\,n}$, the derivative of $\nu_{m\,n}$ with respect to $\mu_{m\,n}$ is a function of $x_m, \cdots, x_n$ and shall be designated by $f_{m\,n}(x_m, \cdots, x_n)$. Since $f_{m\,n}(x_m, \cdots, x_n)$ is positive with ν-probability one $1/f_{m\,n}(x_m, \cdots x_n)$ is well defined with ν-probability one. We shall let the function $1/f_{m\,n}(x_m, \cdots, x_n)$ take on the value 0 when $f_{m\,n}(x_m, \cdots, x_n) \leqq 0$. Thus $1/f_{m\,n}(x_m, \cdots, x_n)$ is well defined everywhere. In fact $1/f_{m\,n}(x_m, \cdots, x_n)$ is the derivative of $\nu_{m\,n}$-continuous part of $\mu_{m\,n}$ with respect to $\nu_{m\,n}$. According to the celebrated theorem of E. S. Anderson and B. Jessen [1] and J. L. Doob ([2]), pp. 343) $1/f_{m\,n}(x_m, \cdots, x_n)$ converges with ν-probability one as $n \to \infty$. If we assume that μ, ν are stationary, i.e., μ, ν are T invariant, more precise results may be expected. A fundamental theorem of Information Theory, first proved by C. Shannon for stationary Markovian measures [5] and later generalized to any stationary measure by B. McMillan [4], may be considered as a theorem of this sort. In their theorem X is assumed to be a finite set. In this paper we shall first treat Markovian stationary measures μ, ν with X being

Received September 2, 1959, and in revised form January 18, 1960. The author is very much indebted to the referee for his suggestions which have been of substantial benefit to this paper.

any set, finite or infinite, and $\mathscr{S}$ any σ-algebra of subsets of X. It will be proved that $n^{-1}\log f_{m\,n}(x_m, \cdots, x_n)$ converges as $n \to \infty$ with ν-probability one and also in $L_1(\nu)$ under some integrability conditions. The case that ν is only stationary is also treated. Similar convergence theorem is proved under the assumption that X is countable.

2. Asymptotic properties of derivatives of a Markovian measure with stationary transition probabilities with respect to another such measure.

Let X, $\mathscr{S}$, Ω, $\mathscr{F}$, Ω_n, $\mathscr{F}_n$, $\mathscr{F}_{m\,n}$, $\mu_{m\,n}$, $\nu_{m\,n}$ $f_{m\,n}(x_m, \cdots, x_n)$ be as in § 1. x_n, $n = 0, \pm 1, \pm 2, \cdots$, are considered as functions or random variables on Ω to X. Notations for conditional prababilities and conditional expectations relative to one or several random variables will be as in [2], chapter 1, § 7. Since we have two probability measures we shall use subscripts μ, ν to indicate conditional probabilities and conditional expectations taken under measures μ, ν respectively. In this section μ, ν *are assumed to be Markovian* i.e., for any $A \in \mathscr{S}$, $m < n$, $n = 0 \pm 1, \pm 2, \cdots$,

(1) $P_\mu[x_n \in A \mid x_m, \cdots, x_{n-1}] = P_\mu[x_n \in A \mid x_{n-1}]$ with μ-probability one and

(2) $P_\nu[x_n \in A \mid x_m, \cdots, x_{n-1}] = P_\nu[x_n \in A \mid x_{n-1}]$ with ν-probability one. For any set $E \subset \Omega$ let I_E be the real valued function on Ω defined by

$$\begin{aligned} I_E(w) &= 1 \text{ if } w \in E \\ &= 0 \text{ if } w \notin E . \end{aligned}$$

LEMMA 1. *If $\nu_{n-1\,n}$ is absolutely continuous with respect to $\mu_{n-1\,n}$ then for any $A \in \mathscr{S}$*

$$(3) \qquad \begin{aligned} &P_\nu[x_n \in A \mid x_{n-1}] f_{n-1\,n-1}(x_{n-1}) \\ &\quad = E_\mu[I_{(x_n \in A)} f_{n-1\,n}(x_{n-1}, x_n) \mid x_{n-1}] \text{ with } \mu\text{-probability one.} \end{aligned}$$

Proof. For any $A, B \in \mathscr{S}$

$$\begin{aligned} &\nu[x_n \in A, x_{n-1} \in B] \\ &\quad = \int_{[x_{n-1} \in B]} P_\nu[x_n \in A \mid x_{n-1}] d\nu \\ &\quad = \int_{[x_{n-1} \in B]} P_\nu[x_n \in A \mid x_{n-1}] f_{n-1\,n-1}(x_{n-1}) d\mu . \end{aligned}$$

On the other hand

$$\begin{aligned} &\nu[x_n \in A, x_{n-1} \in B] \\ &\quad = \int_{[x_{n-1} \in B]} I_{x_n \in A} f_{n-1\,n}(x_{n-1}, x_n) \mid x_{n-1}) d\mu \end{aligned}$$

$$= \int_{[x_{n-1} \in B]} E_\mu[I_{x_n \in A} f_{n-1\,n}(x_{n-1}, x_n) \mid x_{n-1}] d\mu \,.$$

Hence for any $B \in \mathscr{S}$

$$\int_{[x_{n-1} \in B]} P_\nu[x_n \in A \mid x_{n-1}] f_{n-1\,n-1}(x_{n-1}) d\mu$$
$$= \int_{[x_{n-1} \in B]} E_\mu[I_{x_n \in A} f_{n-1\,n}(x_{n-1}, x_n) \mid x_{n-1}] d\mu \,,$$

therefore (3) is true with μ-probability one. Dividing both sides of (3) by $f_{n-1\,n-1}(x_{n-1})$ we then have

$$P_\nu[x_n \in A \mid x_{n-1}] = \frac{E_\mu[I_{x_n \in A} f_{n-1\,n}(x_{n-1}, x_n) \mid x_{n-1}]}{f_{n-1\,n-1}(x_{n-1})} \,. \tag{4}$$

With μ-probability one on the set $[f_{n-1\,n-1}(x_{n-1}) > 0]$. Since $\nu[f_{n-1\,n-1}(x_{n-1}) > 0] = 1$, (4) is true with ν-probability one.

THEOREM 1. *If $\nu_{n-1\,n}$ is absolutely continuous with respect to $\mu_{n-1\,n}$ for $n = 0, \pm 1, \pm 2, \cdots$ then $\nu_{m\,n}$ is absolutely continuous with respect to $\mu_{m\,n}$ for $n = 0, \pm 1, \pm 2, \cdots$ and $m \leqq n$ with*

$$f_{m\,n}(x_m, \cdots, x_n) = f_{m\,m+1}(x_m, x_{m+1}) \frac{f_{m+1\,m+2}(x_{m+1}, x_{m+2})}{f_{m+1\,m+1}(x_{m+1})} \cdots \frac{f_{n-1\,n}(x_{n-1}, x_n)}{f_{n-1\,n-1}(x_{n-1})} \tag{5}$$

with μ-probability one.

Proof. We shall prove the theorem for the case that $m = 1, n = 2, 3, \cdots$. The proof for the general case that m is any integer is similar. Since $\nu_{1\,2}$ is absolutely continuous with respect to $\mu_{1\,2}$ by hypothesis, (5) is trivially true for $m = 1, n = 2$. Suppose $\nu_{1\,k}$ $(k \geqq 2)$ is absolutely continuous with respect to $\mu_{1\,k}$ and $f_{1\,k}(x_1, \cdots, x_k)$ is given by (5) with μ-probability one. For any $A \in \mathscr{S}, B \in \mathscr{F}_k$

$$\nu[x_{k+1} \in A, (x_1, \cdots, x_k) \in B]$$
$$= \int_{[(x_1, \cdots, x_k) \in B]} P_\nu[x_{k+1} \in A \mid x_1, \cdots, x_k] d\nu \,.$$

Since ν is Markovian and by (4)

$$\nu[x_{k+1} \in A, (x_1, \cdots, x_k) \in B]$$
$$= \int_{[(x_1, \cdots, x_k) \in B]} P_\nu[x_{k+1} \in A \mid x_k] d\nu$$

$$= \int_{[(x_1, \cdots, x_k) \in B]} \frac{E_\mu[I_{x_{k+1} \in A} f_{k\,k+1}(x_k, x_{k+1}) \mid x_k]}{f_{k\,k}(x_k)} d\nu$$

$$= \int_{[(x_1, \cdots, x_k) \in B]} \frac{E_\mu[I_{x_{k+1} \in A} f_{k\,k+1}(x_k, x_{k+1}) \mid x_k]}{f_{k\,k}(x_k)} f_{1\,k}(x_1, \cdots, x_k) d\mu .$$

Since μ is Markovian

$$E_\mu[I_{x_{k+1} \in A} f_{k\,k+1}(x_k, x_{k+1}) \mid x_k] = E_\mu[I_{x_{k+1} \in A} f_{k\,k+1}(x_k, x_{k+1}) \mid x_1, \cdots, x_k]$$

with μ-probability one. Hence

$$\nu[x_{k+1} \in A, (x_1, \cdots, x_k) \in B]$$
$$= \int_{(x_1, \cdots, x_k) \in B} E_\mu\left[I_{x_{k+1} \in A} \frac{f_{k\,k+1}(x_k, x_{k+1})}{f_{k\,k}(x_k)} f_{1\,k}(x_1, \cdots, x_k) \mid x_1, \cdots, x_k \right] d\mu$$
$$= \int_{(x_1, \cdots, x_k) \in B} I_{x_{n+1} \in A} f_{1\,k}(x_1, \cdots, x_k) \frac{f_{k\,k+1}(x_k, x_{k+1})}{f_{k\,k}(x_k)} d\mu .$$

Hence

$$\nu[x_{k+1} \in A, (x_1, \cdots, x_k) \in B]$$
$$= \int_{[x_{k+1} \in A, (x_1, \cdots, x_k) \in B]} f_{1\,k}(x_1, \cdots, x_k) \frac{f_{k\,k+1}(x_k, x_{k+1})}{f_{k\,k}(x_k)} d\mu$$

for any $A \in \mathscr{S}, B \in \mathscr{F}_k$. Hence for any $E \in \mathscr{F}_{1\,k+1}$

$$\nu(E) = \int_E f_{1\,k}(x_1, \cdots, x_k) \frac{f_{k\,k+1}(x_k, x_{k+1})}{f_{k\,k}(x_k)} d\mu ,$$

Therefore $\nu_{1\,k+1}$ is absolutely continuous with respect to $\mu_{1\,k+1}$ and

$$(6) \qquad f_{1\,k+1}(x_1, \cdots, x_{k+1}) = f_{1\,k}(x_1, \cdots, x_k) \frac{f_{k\,k+1}(x_k, x_{k+1})}{f_{k\,k}(x_k)}$$

with μ-probability one. (6) together with the supposition that (5) holds true for $m = 1, n = k$ implies that (5) holds true for $m = 1, n = k + 1$. Thus the theorem for the case that $m = 1$ is proved.

Any Markovian probability measure on $\mathscr{F}$ is said to have *stationary transition probabilities* if E being a set of probability one implies that TE, $T^{-1}E$ are also of probability one and for any $A \in \mathscr{S}$ and any n

$$P[x_{n+1} \in A \mid x_n] = TP[x_n \in A \mid x_{n-1}]$$

with probability one. Thus for a Markovian probability measure with stationary transition probabilities we have for any pair of integers m, n and any $A \in \mathscr{S}$

(7) $P[x_n \in A \mid x_{n-1}] = T^{n-m}P[x_m \in A \mid x_{m-1}]$ with probability one and

(8) $E[g(x_{n-1}, x_n) \mid x_{n-1}] = T^{n-m}E[g(x_{m-1}, x_m) \mid x_{m-1}]$ wtih probability one for any real valued $\mathscr{F}_2$-measurable function g on Ω_2.

THEOREM 2. *Let both μ, ν have stationary transition probabilities. If $\nu_{n\,n}$ is absolutely continuous with respect to $\mu_{n\,n}$ for $n = 0, \pm 1, \pm 2, \cdots$ and $\nu_{1\,2}$ is absolutely continuous with respect to $\mu_{1\,2}$ then $\nu_{m\,n}$ is absolutely continuous with respect to $\mu_{m\,n}$ for $m \leqq n, n = 0, \pm 1, \pm 2, \cdots$ and*

$$f_{m\,n}(x_m, \cdots, x_n) = f_{m\,m}(x_m)\frac{f_{1\,2}(x_m, x_{m+1})}{f_{1\,1}(x_m)} \cdots \cdots \frac{f_{1\,2}(x_{n-1}, x_n)}{f_{1\,1}(x_{n-1})} \tag{9}$$

with μ-probability one.

Proof. By Lemma 1, for any $A \in \mathscr{S}$

$$P_\nu[x_2 \in A \mid x_1] = \frac{E_\mu[I_{x_2 \in A} f_{1\,2}(x_1, x_2) \mid x_1]}{f_{1\,1}(x_1)} \tag{10}$$

with ν-probability one. For any $A, B \in \mathscr{S}$

$$\begin{aligned}
&\nu[x_n \in A, x_{n-1} \in B] \\
&= \int_{[x_{n-1} \in B]} P_\nu[x_n \in A \mid x_{n-1}] d\nu \\
&= \int_{[x_{n-1} \in B]} T^{n-2} P_\nu[x_2 \in A \mid x_1] d\nu \\
&= \int_{[x_{n-1} \in B]} \{T^{n-2} P_\nu[x_2 \in A \mid x_1]\} f_{n-1\,n-1}(x_{n-1}) d\mu\,.
\end{aligned}$$

Hence by (10) and (8)

$$\begin{aligned}
&\nu[x_n \in A, x_{n-1} \in B] \\
&= \int_{[x_{n-1} \in B]} T^{n-2} \left\{ \frac{E_\mu[I_{x_2 \in A} f_{1\,2}(x_1, x_2) \mid x_1]}{f_{1\,1}(x_1)} \right\} f_{n-1\,n-1}(x_{n-1}) d\mu \\
&= \int_{[x_{n-1} \in B]} \frac{E_\mu[I_{x_n \in A} f_{1\,2}(x_{n-1}, x_n) \mid x_{n-1}]}{f_{1\,1}(x_{n-1})} f_{n-1\,n-1}(x_{n-1}) d\mu \\
&= \int_{[x_{n-1} \in B]} I_{x_n \in A} f_{n-1\,n-1}(x_{n-1}) \frac{f_{1\,2}(x_{n-1}, x_n)}{f_{1\,1}(x_{n-1})} d\mu \\
&= \int_{[x_n \in A, x_{n-1} \in B]} f_{n-1\,n-1}(x_{n-1}) \frac{f_{1\,2}(x_{n-1}, x_n)}{f_{1\,1}(x_{n-1})} d\mu\,.
\end{aligned}$$

Thus for any $E \in \mathscr{F}_{n-1\,n}$

$$\nu(E) = \int_E f_{n-1\,n-1}(x_{n-1}) \frac{f_{1\,2}(x_{n-1}, x_n)}{f_{1\,1}(x_{n-1})} d\mu . \tag{11}$$

Hence for any integer n, $\nu_{n-1\,n}$ is absolutely continuous with respect to $\mu_{n-1\,n}$ and Theorem 1 is applicable. (11) also implies that

$$f_{n-1\,n}(x_{n-1}, x_n) = f_{n-1\,n-1}(x_{n-1}) \frac{f_{1\,2}(x_{n-1}, x_n)}{f_{1\,1}(x_{n-1})} \tag{12}$$

with μ-probability one. Hence

$$\frac{f_{n-1\,n}(x_{n-1}, x_n)}{f_{n-1\,n-1}(x_{n-1})} = \frac{f_{1\,2}(x_{n-1}, x_n)}{f_{1\,1}(x_{n-1})} \tag{13}$$

with μ-probability one on the set $[f_{n-1\,n-1}(x_{n-1}) > 0]$. However, except that w belongs to a set of μ-probability $0, n > 1, f_{n-1\,n-1}(x_{n-1}(w)) = 0$ imply that $f_{1\,n-1}(x_1(w), \cdots, x_{n-1}(w)) = 0$, hence

$$f_{1\,n-1}(x_1, \cdots, x_{n-1}) \frac{f_{n-1\,n}(x_{n-1}, x_n)}{f_{n-1\,n-1}(x_{n-1})} = f_{1\,n-1}(x_1, \cdots, x_{n-1}) \frac{f_{1\,2}(x_{n-1}, x_n)}{f_{1\,1}(x_{n-1})}$$

with μ-probability one. Thus by (6)

$$f_{1\,n}(x_1, \cdots, x_n) = f_{1\,n-1}(x_1, \cdots, x_{n-1}) \frac{f_{1\,2}(x_{n-1}, x_n)}{f_{1\,1}(x_{n-1})} \tag{14}$$

with μ-probability one. Combining (12) (13) and by induction, if $n > 1$

$$f_{1\,n}(x_1, \cdots, x_n) = f_{1\,1}(x_1) \frac{f_{1\,2}(x_1, x_2)}{f_{1\,1}(x_1)} \cdots \frac{f_{1\,2}(x_{n-1}, x_n)}{f_{1\,1}(x_{n-1})}$$

with μ-probability one. Thus we have proved the theorem for the case that $m = 1$. For the general case the proof is similar.

THEOREM 3. *If μ has stationary transition probabilities and ν is stationary and if*

$$\int |\log f_{m\,m+1}(x_m, x_{m+1})|\, d\nu < \infty \quad \textit{then}$$

$$\int |\log f_{m\,n}(x_m, \cdots, x_n)|\, d\nu < \infty \quad \textit{for } n = m, m+1, m+2, \cdots$$

and $n^{-1} \log f_{m\,n}(x_m, \cdots, x_n)$ converges as $n \to \infty$ with ν-probability one and also in $L_1(\nu)$ to a function g with $\int g\, d\nu = a$ where

$$a = \int [\log f_{1\,2}(x_1, x_2) - \log f_{1\,1}(x_1)] d\nu \geqq 0$$

In particular, if ν is ergodic, $g = a$ with ν-probability one.

Proof. We shall first prove the theorem for the case that $m = 1$. Since for any $A \in \mathscr{S}$

$$\nu[x_1 \in A] = \int_{[x_1 \in A]} f_{1\,1}(x_1) d_\mu = \int_{[x_1 \in A]} f_{1\,2}(x_1, x_2) d\mu ,$$

hence

$$E_\mu[f_{1\,2}(x_1, x_2) \mid x_1] = f_{1\,1}(x_1) .$$

Since $\int \mid \log f_{1\,2}(x_1, x_2) \mid d\nu < \infty$ hence

$$\int \mid f_{1\,2}(x_1, x_2) \log f_{1\,2}(x_1, x_2) \mid d\mu = \int \mid \log f_{1\,2}(x_1, x_2) \mid d\nu < \infty .$$

The real valued function $L(\xi) = \xi \log \xi$ defined for all real $\xi \geqq 0[L(0)$ is taken to be $0]$ is convex. By Jensen's inequality for conditional expectations ([2], pp. 33)

$$E_\mu[L\{f_{1\,2}(x_1 x_2)\} \mid x_1] \geqq L\{f_{1\,1}(x_1)\} . \tag{15}$$

By (15) and the fact that $L(\xi)$ is a function bounded below by a constant, we have

$$\int \mid L\{f_{1\,1}(x_1)\} \mid d\mu = \int \mid \log f_{1\,1}(x_1) \mid d\nu < \infty$$

and

$$\int \log f_{1\,2}(x_1, x_2) d_\nu - \int \log f_{1\,1}(x_1) d_\nu = a \geqq 0 .$$

Now by Theorem 2

$$\log f_{1\,n}(x_1, \cdots, x_n) = \log f_{1\,1}(x_1) + \sum_{i=2}^{n} \{\log f_{1\,2}(x_{i-1}, x_i) - \log f_{1\,1}(x_{i-1})\} .$$

Since ν is stationary, $\log f_{1\,n}(x_1, \cdots, x_n)$ is ν-integrable. Applying the ergodic theorem $n^{-1} \log f_{1\,n}(x_1, \cdots, x_n)$ converges with ν-probability one and also in $L_1(\nu)$ to a function g with

$$\int g d\nu = \int [\log f_{1\,2}(x_1, x_2) - \log f_{1\,1}(x_1)] d\nu = a \geqq 0 .$$

For m being any integer, we only need to mentioned that by (13),

$$\log f_{m\,m+1}(x_m, x_{m+1}) - \log f_{m\,m}(x_m) = \log f_{1\,2}(x_1, x_2) - \log f_{1\,1}(x_1)$$

with ν-probability one and therefore the same conclusion follows with a similar proof.

COROLLARY 1. *Suppose μ, ν satisfy the hypothesis of Theorem 3 for $m = 1$. If ν is ergodic and if there is an $A \in \mathscr{S}$ such that*

$$\nu\{P_\nu[x_2 \in A \mid x_1] \neq P_\mu[x_2 \in A \mid x_1]\} > 0 \tag{16}$$

then ν is singular with respect to μ.

Proof. First we shall show that follows from (16)

$$\mu[f_{1\,1}(x_1) \neq f_{1\,2}(x_1, x_2)] > 0 \,. \tag{17}$$

For, if $f_{1\,1}(x_1) = f_{1\,2}(x_1, x_2)$ with μ-probability one then by Lemma 1 $P_\nu[x_2 \in A \mid x_1]f_{1\,1}(x_1) = P_\mu[x_2 \in A \mid x_1]f_{1\,1}(x_1)$ with μ-probability one. Thus $P_\nu[x_2 \in A \mid x_1] = P_\mu[x_2 \in A \mid x_1]$ with ν-probability one for every $A \in \mathscr{S}$. Now the function $L(\xi) = \xi \log \xi$ is strictly convex, hence it follows from (17) that

$$a = \int [L\{f_{1\,2}(x_1, x_2)\} - L\{f_{1\,1}(x_1)\}]d_\mu > 0 \,.$$

Applying Theorem 3 $f_{1\,n}(x_1, \cdots, x_n) \to \infty$ with ν-probability one as $n \to \infty$. Hence $1/f_n(x_1, \cdots, x_n) \to 0$ with ν-probability one as $n \to \infty$. Let $\mathscr{F}'$ be the σ-algebra generated by $\bigcup_{n=1}^{\infty} \mathscr{F}_{1\,n}$ and μ', ν' be the contractions of μ, ν to $\mathscr{F}'$ respectively. Since $1/f_{1\,n}(x_1, \cdots, x_n)$ is the derivative of $\nu_{1\,n}$-continuous part of $\mu_{1\,n}$ with respect to $\nu_{1\,n}$, $1/f_{1\,n}(x, \cdots, x_n)$ converges with ν-probability one as $n \to \infty$ to the derivative of ν'-continuous part of μ' with respect to ν' ([2], pp. 343). Now $1/f_{1\,n}(x_1, \cdots, x_n)$ converges to 0 with ν-probability one, hence the ν'-continuous part of μ' is 0 and μ', ν' are mutually singular. Hence μ, ν are mutually singular.

3. Extension to k-Markovian measures.

The results of the preceding section can be extended to k-Markovian measures immediately. We shall state the theorems only since the proofs in the preceding section with obvious modifications apply as well.

THEOREM 4. *Let μ, ν be any two k-Markovian measures on $\mathscr{F}$. If $\nu_{n-k\,n}$ is absolutely continuous with respect to $\mu_{n-k,\,n}$ for $n = 0, \pm 1, \pm 2, \cdots$, then $\nu_{m\,n}$ is absolutely continuous with respect to $\mu_{m\,n}$ for $n = 0, \pm 1, \pm 2, \cdots$ and $m \leqq n$ with*

$$f_{m\,n}(x_m, \cdots, x_n) = f_{m\,m+k}(x_m, \cdots, x_{m+k}) \frac{f_{m+1,m+1+k}(x_{m+1}, \cdots, x_{m+1+k})}{f_{m+1,m+k}(x_{m+1}, \cdots, x_{m+k})} \cdots \frac{f_{n-k\,n}(x_{n-k}, \cdots, x_n)}{f_{n-k\,n-1}(x_{n-k}, \cdots, x_{n-1})} \tag{18}$$

with μ-probability one.

THEOREM 5. *Let μ, ν be two k-Markovian measures on $\mathscr{F}$ with stationary transition probabilities. If $\nu_{n-k+1,n}$ is absolutely continuous with respect to $\mu_{n-k+1,n}$ for $n = 0, \pm 1, \pm 2, \cdots$ and $\nu_{1\,k+1}$ is absolutely continuous with respect to $\mu_{1\,k+1}$ then $\nu_{m\,n}$ is absolutely continuous with respect to $\mu_{m\,n}$ for $n = 0, \pm 1, \pm 2, \cdots, m \leqq n$ and*

$$(19) \quad f_{m\,n}(x_m, \cdots, x_n) = f_{m\,m+k-1}(x_m, \cdots, x_{m+k-1}) \frac{f_{1\,k+1}(x_{m+1}, \cdots, x_{m+k+1})}{f_{1,k}(x_{m+1}, \cdots, x_{m+k})} \frac{f_{1\,k+1}(x_{n-k}, \cdots, x_n)}{f_{1\,k}(x_{n-k}, \cdots, {}_{n-1})}$$

with μ-probability one.

THEOREM 6. *Let μ, ν be two k-Markovian measures such that ν is stationary and μ has stationary transition probabilities. If*

$$\int |\log f_{m\,m+k}(x_m, \cdots, x_{m+k})|\, d\nu < \infty$$

then $\int |\log f_{m\,n}(x_m, \cdots, x_n)|\, d\nu < \infty$ for $n = m, m+1, m+2, \cdots$ and $n^{-1} \log f_{m\,n}(x_m, \cdots, x_n)$ converges as $n \to \infty$ with ν-probability one to a function g with $\int g d\nu = a \geqq 0$ where

$$a = \int |\log f_{1\,k+1}(x_1, \cdots, x_{k+1}) - \log f_{1\,k}(x_1, \cdots, x_k)|\, d\nu \geqq 0 .$$

In particular, if ν is ergodic, $g = a$ with ν-probability one.

COROLLARY 2. *Suppose μ, ν satisfy the hypothesis of Theorem 6 for $m = 1$. If ν is ergodic and if there is a set $A \in \mathscr{S}$ such that*

$$(20) \qquad \nu\{[P_\nu[x_{k+1} \in A \mid x_1, \cdots, x_k] \neq \mathrm{P}_\mu[x_{k+1} \in A] \mid x_1, \cdots, x_k]\} > 0$$

Then ν is singular with respect to μ.

4. A generalization of McMillan's theorem. In the setting of this paper, McMillan's Theorem may be stated as the following. Let X be a finite set of K points and $\mathscr{S}$ be the σ-algebra of all subsets of X. Let ν be any stationary probability measure on $\mathscr{F}$ and μ be the measure on $\mathscr{F}$ such that $\mu[X_m = a_0, X_{m+1} = a_1, \cdots, X_n = a_{n-m}] = K^{-(n-m+1)}$ for any intergers m, n and $a_0, a_1 \cdots a_{n-m}$ in X. μ may be described as the equally distributed independent measure on $\mathscr{F}$. Then $n^r f_{1\,n}(x_1, \cdots, x_n)$ converges as $n \to \infty$ in $L_1(\nu)$. In particular, if ν is ergodic, the limit function is equal to $\log K - H$ with ν-probability one where H is the entropy of ν measure [4]. We shall generalize this theorem to the case that X is countable and μ is Markovian with stationary transition probabilities.

THEOREM 7. *Let the totality of elements of X be $a_1, a_2, \cdots$ and ν be a stationary probability measure on $\mathscr{F}$ such that $\int -\log \nu_1(x_1) d\nu < \infty$ where ν_1 is the function defined on X by $\nu_1(a_i) = \nu[x_1 = a_i]$. Let μ be a Markovian measure on $\mathscr{F}$ with stationary transition probabilities. Let $p(a_i, a_j)$ be the value of $P_\mu[x_1 = a_j \mid x_0]$ when $x_0 = a_i$. Let $\nu_{1\,n}$ be absolutely continuous with respect to $\mu_{1\,n}$ for $n = 1, 2, \cdots$. If*

$$\int -\log p(x_1, x_2) d\nu < \infty$$

and $\int |\log f_{1\,1}(x_1)| \, d\nu < \infty$ then $\int |\log f_{1\,n}(x_1, \cdots, x_n)| \, d\nu < \infty$ for $n = 1, 2, \cdots$ and $n^{-1} \log f_{1\,n}(x_1, \cdots, x_n)$ converges as $n \to \infty$ in $L_1(\nu)$. In particular, if ν is ergodic, the limit is equal to a constant with ν-probability one.

Proof. Let

$$\nu_n(a_{i_1}, a_{i_2}, \cdots, a_{i_n}) = \nu[x_1 = a_{i_1}, x_2 = a_{i_2}, \cdots, x_n = a_{i_n}]$$

and

$$\mu_n(a_{i_1}, a_{i_2}, \cdots, a_{i_n}) = \mu[x_1 = a_{i_1}, x_2 = a_{i_2}, \cdots, x_n = a_{i_n}] .$$

Then

$$f_{1\,n}(x_1, \cdots, x_n) = \frac{\nu_n(x_1, \cdots, x_n)}{\mu_n(x_1, \cdots, x_n)}$$

with μ-proba[illegible] e and

$$= a_i \mid x_{n-1}, \cdots, x_1] = \frac{\nu_n(x_1, \cdots, x_{n-1}, a_i)}{\nu_{n-1}(x_1, \cdots, x_{n-1})}$$

with ν-probability one and

$$P_\mu[x_n = a_i \mid x_{n-1}] = \frac{\mu_n(x_1, \cdots, x_{n-1}, a_i)}{\mu_n(x_1, \cdots, x_{n-1})}$$

with μ-probability one. Hence

$$\frac{f_{1\,n}(x_1, \cdots, x_n)}{f_{1\,n-1}(x_1, \cdots, x_{n-1})} = \sum_{i=1}^{\infty} \frac{P_\nu[x_n = a_i \mid x_{n-1}, \cdots, x_1]}{P_\mu[x_n = a_i \mid x_{n-1}]} I_{x_n = a_i}$$

with ν-probability one and

$$\text{(21)} \quad \log \frac{f_{1\,n-1}(x_1, \cdots, x_n)}{f_{1\,n-1}(x_1, \cdots, x_{n-1})} = \sum_{i=1}^{\infty} \log P_\nu[x_n = a_i \mid x_{n-1}, \cdots, x_1] I_{x_n = a_i}$$

$$-\log p(x_{n-1}, x_n)$$

$$= T^n g_n$$

with ν-probability one where

$$g_n = \sum_{i=1}^{\infty} \log P_\nu[x_0 = a_i \mid x_{-1}, \cdots, x_{-(n-1)}] I_{x_0 = a_i} - \log p(x_{-1}, x_0) . \tag{22}$$

We know that $P_\nu[x_0 = a_i \mid x_{-1}, \cdots, x_{-(n-1)}]$ converges with ν-probability one as $n \to \infty$ to $P_\nu[x_0 = a_i \mid x_{-1}, x_{-2}, \cdots]$ by Doob's Martingale Convergence Theorem. Hence $L\{P_\nu[x_0 = a_i \mid x_{-1}, \cdots, x_{-(n-1)}]\}$ converges with ν-probability one to $L\{P_\nu[x_0 = a_i \mid x_{-1}, x_{-2}, \cdots]\}$. But $L(\xi)$ is a bounded function for $0 \leqq \xi \leqq 1$, hence $L\{P_\nu[x_0 = a_i \mid x_{-1}, x_{-(n-1)}\}$ are uniformly bounded with ν-probability one. Hence $L\{P_\nu[x_0 = x_i \mid x_{-1}, \cdots, x_{-(n-1)}]\}$ also converges in $L_1(\nu)$ to $L\{P_\nu[x_0 = a_i \mid x_{-1}, x_{-2}, \cdots]\}$ as $n \to \infty$. Now by Jensen's inequality $\int - L\{P_\nu[x_0 = a_i \mid x_{-1}, \cdots, x_{-(n-1)}]\} d\nu \leqq -L\{P_\nu[x_0 = a_i]\}$. Since

$$\sum_{i=1}^{\infty} - L\{P_\nu[x_0 = a_i]\} = \int - \log \nu_1(x_0) d\nu < \infty$$

$$\sum_{i=1}^{m} - L\{P_\nu[x_0 = a_i \mid x_{-1}, \cdots, x_{-(n-1)}]\}$$

converges in $L_1(\nu)$, as $m \to \infty$, to

$$\sum_{i=1}^{\infty} - L\{P_\nu[x_0 = a_i \mid x_{-1}, \cdots, x_{-(n-1)}]\}$$

uniformly in n. Hence

$$\sum_{i=1}^{\infty} - L\{P_\nu[x_0 = a_i \mid x_{-1}, \cdots, x_{-(n-1)}$$

converges in $L_1(\nu)$ to

$$\sum_{i=1}^{\infty} - L\{P_\nu[x_0 = a_i \mid x_{-1}, x_{-2}, \cdots]\} \text{ as } n \to \infty. \quad \text{Now}$$

$$\int - \sum_{i=1}^{\infty} \log P_\nu[x_0 = a_i \mid x_{-1}, \cdots, x_{-(n-1)}] I_{x_0 = a_i} d\nu$$
$$= \int - \sum_{i=1}^{\infty} L\{P_\nu[x_0 = a_i \mid x_{-1}, \cdots, x_{-(n-1)}]\} d\nu \text{ and}$$

$$\int - \sum_{i=1}^{\infty} \log P_\nu[x_0 = a_i \mid x_{-1}, x_{-2}, \cdots] I_{x_0 = a_i} d\nu$$
$$= \int - \sum_{i=1}^{\infty} L\{P_\nu[x_0 = a_i \mid x_{-1}, x_{-2}, \cdots]\} d\nu, \text{ hence}$$

$$\lim_{n \to \infty} \int - \sum_{i=1}^{\infty} \log P_\nu[x_0 = a_i \mid x_{-1}, \cdots, x_{-(n-1)}] I_{x_0 = a_i} d\nu = \int - \sum_{i=1}^{\infty} \log P_\nu[x_0 = a_i \mid x_{-1}, x_{-2}, \cdots] I_{x_0 = a_i} d\nu . \tag{23}$$

(23) together with the facts that the sequence

$$\left\{-\sum_{i=1}^{\infty}\log P_\nu[x_0 = x_i \mid x_{-1}, \cdots, x_{-(n-1)}]I_{x_0=a_i}\right\}$$

is also convergent with ν-probability one and that the functions

$$-\sum_{i=1}^{\infty}\log P_\nu[x_0 = x_i \mid x_{-1}, \cdots, x_{-(n-1)}]I_{x_0=a_i}$$

are non negative with ν-probability one imply that

$$\sum_{i=1}^{\infty} P_\nu[x_0 = a_i \mid x_{-1}, \cdots, x_{-(n-1)}]I_{x_0=a_i}$$

converges as $n \to \infty$ in $L_1(\nu)$ to

$$\sum_{i=1}^{\infty} P_\nu[x_0 = a_i \mid x_{-1}, x_{-2}, \cdots]I_{x_0=a_i} .$$

Thus we have $\{g_n\}$ to be an $L_1(\nu)$ convergent sequence. Let the limit of the sequence *be* h. Let $\bar{h}$ be the $L_1(\nu)$ limit of $1/n(h + Th + \cdots + T^n h)$ as $n \to \infty$. Now by (21)

$$\log f_{1\,2}(x_1, \cdots, x_n) = \log f_{1\,1}(x_1) + \sum_{i=2}^{n} T^i g_i. \quad \text{Thus}$$

$$\int\left|\frac{1}{n}\log f_{1\,n}(x_1, \cdots, x_n) - \bar{h}\right|d\nu$$

$$\leqq \frac{1}{n}\int|\log f_{1\,1}(x_1)|\,d\nu + \int\left|\frac{1}{n}\left(\sum_{i=2}^{n} T^i g_i - \sum_{i=2}^{n} T^i h\right)\right|d\nu$$

$$+ \int\left|\frac{1}{n}\sum_{i=2}^{n} T^i h - \bar{h}\right|d\nu$$

$$= \frac{1}{n}\int\left|\log f_{1\,1}(x_1)\right|d\nu + \frac{1}{n}\sum_{i=2}^{n}\int|g_i - h|\,d\nu$$

$$+ \int\left|\frac{1}{n}\sum_{i=2}^{n} T^i h - \bar{h}\right|d\nu \to 0 \text{ as } n \to \infty .$$

COROLLARY 3. *Under the hypothesis of Theorem* 7, *if* ν *is ergodic and not Markovian then* ν *is singular to* μ.

Proof. If ν is ergodic then the $L_1(\nu)$ limit, $\bar{h}$, of $\{1/n \log f_{1\,n}(x_1, \cdots, x_n)\}$ is equal with ν probability one to

$$\int\sum_{i=1}^{\infty} L\{P_\nu[x_0 = a_i \mid x_{-1}, x_{-2}, \cdots]\}d\nu - \int\log p(x_{-1}, x_0)d\nu$$

which is greater or equal to

$$\int \sum_{i=1}^{\infty} L\{P_\nu[x_0 = a_i \mid x_{-1}, x_{-2}]\}d\nu - \int \log p(x_{-1}, x_0)d\nu \ .$$

Hence by (21)

$$\begin{aligned}\bar{h} &\geqq \int \sum_{i=1}^{\infty} \log P_\nu[x_0 = a_i \mid x_{-1}, x_{-2}]I_{x_0=a_i}d\nu - \int \log p(x_{-1}, x_0)d\nu \\ &= \int \log f_{1\,3}(x_1, x_2, x_2)d\nu - \int \log f_{1\,2}(x_1, x_2)d\nu \ .\end{aligned}$$

However $\int \log f_{1\,3}(x_1, x_2, x_3)d\nu - \int \log f_{1\,2}(x_1, x_2)d\nu = 0$ if and only if

(24) $$\mu[f_{1\,2}(x_1, x_2) \neq f_{1\,3}(x_1, x_2, x_3)] = 0 \ .$$

(24) implies that

$$P_\nu[x_3 \in A \mid x_1, x_2] = P_\mu[x_3 \in A \mid x_1, x_2]$$

with ν-probability one for any $A \in \mathscr{S}$. This is impossible since μ is Markovian and ν is not. Hence $\bar{h} > 0$ with ν-probability one. Hence $f_{1\,n}(x_1, \cdots, x_n) \to \infty$ with ν probability one and ν is singular to μ by the same argument used in the proof in Corollary 1.

The extensions of Theorem 7 and Corollary 3 to k-Markovian μ is obvious.

References

1. Erik Sparre Anderson and Borge Jessen, *Some limit theorems in an abstract set*, Danske Vid. Selsk. Nat.-Fys. Medd. **22**, no. 14 (1946). **22**, No. 14 (1946).

2. J. L. Doob, *Stochastic Processes*, John Wiley and Sons, Inc., New York.

3. Amiel Feinstein, *Foundations of Information Theory*, McGraw-Hill Inc. New York, Toronto, London.

4. B. McMillan, *The basic theorems of information theory*, Annals of Math. Statistics, **24** (1953), 196-219.

5. C. E. Shannon, *The mathematical theory of communication*, Bell Syst. Techn. Journ. **27** (1948), 379-423, 623-456.

WAYNE STATE UNIVERSITY

CONCERNING BOUNDARY VALUE PROBLEMS[1]

J. W. NEUBERGER

1. Introduction. This paper follows work on integral equations by H. S. Wall [4], [5], J. S. MacNerney [1], [2] and the present author [3]. Some results of these papers are used here to investigate certain boundary value problems.

In § 2, results of Wall and MacNerney are used to study a linear boundary value problem which includes problems of the following kind: Suppose that each of a_{ij}, $i, j = 1, \cdots, n$ is a continuous function, a and b are numbers and each of b_{ij}, c_{ij} and d_i, $i, j = 1, \cdots, n$ is a number. Is there a unique function n-tuple $f_1, \cdots, f_n$ such that

$$f_i' = \sum_{j=1}^{n} a_{ij} f_j \quad \text{and} \quad \sum_{j=1}^{n} [b_{ij} f_j(a) + c_{ij} f_j(b)] = d_i \,, \qquad i = 1, \cdots, n?$$

Section 3 contains some observations concerning a nonlinear boundary value problem which includes the problem of solving a certain system of nonlinear first order differential equations together with a nonlinear boundary condition. An example is given in the final section.

S denotes a normed, complete, abelian group (norms are denoted by $\|\cdot\|$). B denotes the normed, complete, abelian group of all bounded endomorphisms from S to S (the norm of an element T of B is the g.l.b. of the set of all M such that $\|Tx\| \leqq M\|x\|$ for all x in S). B^* denotes the set to which T belongs only if T is a continuous function from S to S. If $[a, b]$ denotes a number interval, then $C_{[a,b]}$ denotes the set to which f belongs only if f is a continuous function from $[a, b]$ to S. The identity function on the numbers is denoted by j.

The reader is referred to [1] for a definition of the integral of a function from a number interval $[a, b]$ to B with respect to a function from $[a, b]$ to B and to [3] for a definition of the integral of a function from $[a, b]$ to S with respect to a function from $[a, b]$ to B^*. [1] and [3] contain existence theorems for these integrals and a discussion of some of their properties.

2. A linear boundary value problem. Suppose that $[a, b]$ is a number interval and F is a continuous function from $[a, b]$ to B which is of bounded variation on $[a, b]$. The following are theorems:

(i) There is a unique continuous function M from $[a, b] \times [a, b]$ to B such that $M(t, u) = I + \int_u^t dF \cdot M(j, u)$ for each of t and u in $[a, b]$. (I denotes the identity element in B)

Received March 25, 1959, and in revised form January 22, 1960.

[1] Presented to the Society in part, August, 1958, and in part, January, 1959.

(ii) $M(t, u)M(u, v) = M(t, v)$ if each of t, u and v is in $[a, b]$.

(iii) If h is a continuous function from $[a, b]$ to S and c is in $[a, b]$, then the only element X of $C_{[a,b]}$ such that $X(t) = h(t) + \int_c^t dF \cdot X$ for each t in $[a, b]$ is given by $X(t) = M(t, c)h(c) + \int_c^t M(t, j)dh$ for each t in $[a, b]$.[2]

THEOREM A. *Suppose that H is a function from $[a, b]$ to B which is of bounded variation on $[a, b]$. A necessary and sufficient condition that there be a unique element Y of $C_{[a,b]}$ such that*

(*) $Y(t) = Y(u) + g(t) - g(u) + \int_a^t dF \cdot Y$ *and* $\int_a^b dH \cdot Y = C$ *for each C in S and each g in $C_{[a,b]}$ is that $\int_a^b dH \cdot M(j, a)$ have an inverse which is from S onto S.*

Proof. Consider first the following lemma. If Y is in $C_{[a,b]}$ and satisfies (*) for each of u and t in $[a, b]$, then

$$\left[\int_a^b dH \cdot M(j, a)\right]Y(a) = C - \int_a^b\left[\int_j^b dH(s) \cdot M(s, j)\right]dg .$$

Suppose Y is in $C_{[a,b]}$ and satisfies (*) for each of u and t in $[a, b]$. By (iii), $Y(t) = M(t, a)Y(a) + \int_a^t M(t, j)dg$ for each t in $[a, b]$ and thus

$$C = \int_a^b dH \cdot Y = \left[\int_a^b dH\cdot M(j, a)\right]Y(a) + \int_a^b dH(s) \cdot \left[\int_a^s M(s, j)dg\right]$$
$$= \left[\int_a^b dH \cdot M(j, a)\right]Y(a) + \int_a^b\left[\int_j^b dH(s) \cdot M(s, j)\right]dg .[3]$$

Hence,

$$\left[\int_a^b dH \cdot M(j, a)\right]Y(a) = C - \int_a^b\left[\int_j^b dH(s) \cdot M(s, j)\right]dg .$$

Denote $\int_a^b dH \cdot M(j, a)$ by Q. Suppose that (*) has a unique solution for each g in $C_{[a,b]}$ and each C in S.

Denote by W a point of S, by g an element of $C_{[a,b]}$,

[2] Certain essential ideas for Theorems (i) and (ii) were given by Wall in [**4**]. In [**5**,] Wall gave these theorems for S an n-dimensional Euclidean space or suitable infinite dimensional space. In [**1**], MacNerney extended Wall's theory in proving these theorems for any normed, linear and complete space. Modifications of MacNerney's proofs to the case of S a normed, complete, abelian group are so slight that the proofs are omitted. Discussion concerning the properties and computation of M can be found in each paper listed as reference to this paper.

[3] A proof that $\int_a^b dH(s) \cdot \left[\int_a^s M(s, j)dg\right] = \int_a^b\left[\int_j^b dH(s) \cdot M(s, j)\right]dg$ which follows closely a similar argument for ordinary integrals, is ommitted.

$$W + \int_a^b \left[\int_j^b dH(s) \cdot M(s, j) \right] dg$$

by C and by X the unique element of $C_{[a,b]}$ satisfying (*) for this g and C. By the above lemma, $QX(a) = C - \int_a^b \left[\int_j^b dH(s) \cdot M(s, j) \right] dg = W$. Thus each point of S is the image of some point of S under Q, that is, Q takes S onto S.

Suppose that Q is not reversible and denote by each of W, U and V a point in S such that $QU = W$, $QV = W$ and $U \neq V$. Denote by Y and Z two elements of $C_{[a,b]}$ such that $Y(t) = U + g(t) - g(a) + \int_a^t dF \cdot Y$ and $Z(t) = V + g(t) - g(a) + \int_a^t dF \cdot Z$ for each t in $[a, b]$. Thus, $Y(t) = Y(u) + g(t) - g(u) + \int_u^t dF \cdot Y$ and $Z(t) = Z(u) + g(t) - g(u) + \int_u^t dF \cdot Z$, for each of u and t in $[a, b]$. Since $Y(a) = U$ and $Z(a) = V$, it follows that $Y \neq Z$. As in the proof of the lemma,

$$\int_a^b dH \cdot Y = QU + \int_a^b \left[\int_j^b dH(s) \cdot M(s, j) \right] dg$$

and

$$\int_a^b dH \cdot Z = QV + \int_a^b \left[\int_j^b dH(s) \cdot M(s, j) \right] dg$$

and so

$$\int_a^b dH \cdot Y = \int_a^b dH \cdot Z \,,$$

which means that there is a boundary value problem of the type (*) which has two solutions, which contradicts the above assumption. Thus if (*) has a unique solution for each g in $C_{[a,b]}$ and each C in S, Q takes S onto S reversibly.

Suppose that Q takes S onto S reversibly. Denote by g an element of $C_{[a,b]}$ and by C a point in S. Denote

$$\left[\int_a^b dH \cdot M(j, a) \right]^{-1} \left\{ C - \int_a^b \left[\int_j^b dH(s) \cdot M(s, j) \right] dg \right\}$$

by U and denote by X the element of $C_{[a,b]}$ such that $X(t) = U + g(t) - g(a) + \int_a^t dH \cdot X$ for each t in $[a, b]$. Noting that $X(t) = X(u) + g(t) - g(u) + \int_u^t dH \cdot X$ and that $X(t) = M(t, a)U + \int_a^t M(t, j) dg$ for each of u and t in $[a, b]$ and substituting for X in $\int_a^b dH \cdot X$, it is seen that $\int_a^b dH \cdot X = C$. Thus X satisfies (*) for this g and C. Suppose Y is in $C_{[a,b]}$ and satisfies (*). Then, by the above lemma,

$$QY(a) = C - \int_a^b \left[\int_j^b dH(s) \cdot M(s, j) \right] dg$$

and so $Y(a) = U$ which means that $Y(t) = U + g(t) - g(u) + \int_a^t dF \cdot Y$ and hence by (iii), $X = Y$. Thus if Q takes S onto S reversibly, there is a unique solution to (*) for each g in $C_{[a,b]}$ and C in S.

THEOREM B. *If* $\int_a^b dH \cdot M(j, a)$ *has a bounded inverse which takes* S *onto* S, *that is, if* $\left[\int_a^b dH \cdot M(j, a)\right]^{-1}$ *is in* B, *then there is a function* R *from* $[a, b]$ *to* B *and a function* K *from* $[a, b] \times [a, b]$ *to* B *such that if* g *is in* $C_{[a,b]}$ *and* C *is in* S, *then the only element* Y *of* $C_{[a,b]}$ *satisfying* (*) *for each of* t *and* u *in* $[a, b]$ *is given by* $Y(t) = R(t)C + \int_a^b K(t, j)dg$ *for each* t *in* $[a, b]$. *Moreover, such a pair of functions* R *and* K *is given by* $R(t) = \left[\int_a^b dH \cdot M(j, t)\right]^{-1}$ *and*

$$K(t, u) = \begin{cases} -\left[\int_a^b dH \cdot M(j, t)\right]^{-1}\int_u^b dH \cdot M(j, u) + M(t, u) & \text{if } a \leqq u \leqq t \\ -\left[\int_a^b dH \cdot M(j, t)\right]^{-1}\int_u^b dH \cdot M(j, u) & \text{if } t \leqq u \leqq b. \end{cases}$$

Proof. Suppose that g is in $C_{[a,b]}$ and C is in S. From Theorem A, (*) has a unique solution Y for this C and g, and from the lemma in the proof of Theorem A,

$$\left[\int_a^b dH \cdot M(j, a)\right]X(a) = C - \int_a^b\left[\int_j^b dH(s) \cdot M(s, j)\right]dg$$

and so

$$X(a) = \left[\int_a^b dH \cdot M(j, a)\right]^{-1}\left\{C - \int_a^b\left[\int_j^b dH(s) \cdot M(s, j)\right]dg\right\}.$$

Using (iii) and the fact that

$$M(t, a)\left[\int_a^b dH \cdot M(j, a)\right]^{-1} = \left[\int_a^b dH \cdot M(j, t)\right]^{-1},$$

$$\begin{aligned} X(t) &= \left[\int_a^b dH \cdot M(j, t)\right]^{-1}C - \int_a^b\left\{\left[\int_a^b dH \cdot M(j, t)\right]^{-1}\int_j^b dH(s) \cdot M(s, j)\right\}dg \\ &\quad + \int_a^t M(t, j)dg \\ &= R(t)C + \int_a^b K(t, j)dg \end{aligned}$$

where R and K are defined as in the statement of the theorem.

3. A nonlinear boundary value problem. Here a problem is considered which includes the one in the preceding section. Essentially, the requirements of § 2 that each of $F(t)$ and $H(t)$ be an element of B for every t in $[a, b]$ and that F and H be of bounded variation are replaced by considerably weaker conditions. Theorem D gives a necessary and sufficient condition for the nonlinear problem considered to have a unique solution. First a fundamental theorem for a certain type of integral equation is given.

THEOREM C. *Suppose that* $[a, b]$ *is a number interval and* F *is a function from* $[a, b]$ *to* B^* *such that if* A *is in* S *and* $r > 0$, *there is a variation function* U *on* $[a, b]$ *and a variation function* V *on* $[a, b]$ *such that*

$$\| [F(p) - F(q)]x \| \leqq U(p, q)$$

and

$$\| [F(p) - F(q)]x - [F(p) - F(q)]y \| \leqq V(p, q) \| x - y \|$$

if each of p and q is in $[a, b]$, $\| A - x \| \leqq r$ and $\| A - y \| \leqq r$. Then, if c is in $[a, b]$, there is a segment Q' containing c such that if Q is the common part of Q' and $[a, b]$, there is only one continuous function Y from Q to S such that $Y(t) = A + \int_c^t dF \cdot Y$ if t is in Q.

This follows from Theorem F of [3].

DEFINITION. Suppose F is a function from $[a, b]$ to B^* and c is in $[a, b]$. If there is a point A in S and an element Y of $C_{[a,b]}$ such that $Y(t) = A + \int_c^t dF \cdot Y$ for each t in $[a, b]$, then the set which contains only each such point A is denoted by $F_{c;[a,b]}$.

LEMMA 4.1. *Suppose that* F *satisfies the hypothesis of Theorem C and for some number* c *in* $[a, b]$ *and that there is a segment* Q' *as in the theorem which has* $[a, b]$ *as subset. Then, for each number* u *in* $[a, b]$, *there is a set* $F_{u;[a,b]}$.

Proof. Given such a number c and segment Q', then $Q = [a, b]$ and there is a point A in S and an element Y of $C_{[a,b]}$ such that $Y(t) = A + \int_c^t dF \cdot Y$ for each t in $[a, b]$. Thus if u is in $[a, b]$, $Y(u) = A + \int_c^u dF \cdot Y$ and $Y(t) = Y(u) + \int_u^t dF \cdot Y$ for each t in $[a, b]$. Thus there is a set $F_{u;[a,b]}$.

DEFINITION. Suppose the hypothesis of Lemma 4.1 holds. M denotes a function from $[a, b] \times [a, b]$ such that if each of t and u is in $[a, b]$, $M(t, u)$ is the function from $F_{u;[a,b]}$ to $F_{t;[a,b]}$ such that if A is in $F_{u;[a,b]}$, $M(t, u)A$ is $Y(t)$ where Y is the element of $C_{[a,b]}$ satisfying $Y(s) = A + \int_u^s dF \cdot Y$ for each s in $[a, b]$.

Lemma 4.2. *Under the hypothesis of Lemma 4.1, $M(s, t)M(t, u) = M(s, u)$ for each of s, t and u in $[a, b]$.*

Proof. Suppose that each of s, t and u is in $[a, b]$ and A is in $F_{u;[a,b]}$. Then, $Y(s) = A + \int_u^s dF \cdot Y$ and $Y(t) = A + \int_u^t dF \cdot Y$ so that $Y(s) = Y(t) + \int_t^s dF \cdot Y$, $Y(t) = M(t, u)A$ and $Y(s) = M(s, u)A$. Therefore, $Y(s) = M(t, u)A + \int_t^s dF \cdot Y$ and $Y(s) = M(s, t)[M(t, u)A] = [M(s, t)M(t, u)]A$. Thus, $M(s, u) = M(s, t)M(t, u)$.

Theorem D. *Suppose that in addition to the hypothesis of Theorem C, it is true that for some c in $[a, b]$, there is a set $F_{c;[a,b]}$. Suppose furthermore that T is a function from $C_{[a,b]}$ to S and that C is in S. The following two statements are equivalent:*

(i) *There is only one element Y of $C_{[v,b]}$ such that*

(**) *$TY = C$ and $Y(t) = Y(u) + \int_u^t dF \cdot Y$ for each of t and u in $[a, b]$.*

(ii) *For some u in $[a, b]$, the function R from $F_{u;[a,b]}$, defined by $RA = T[M(j, u)A]$ for each A in $F_{u;[a,b]}$ takes only one element of $F_{u;[a,b]}$ into C.*

Proof. Suppose that for some u in $[a, b]$, the function R as defined in Theorem D takes only the point U of $F_{u;[a,b]}$ into C. Denote by Y the element of $C_{[a,b]}$ such that $Y(t) = U + \int_u^t dF \cdot Y$ for each t in $[a, b]$. Thus, $Y(t) = Y(s) + \int_s^t dF \cdot Y$ and $Y(t) = M(t, u)U$ for each of t and s in $[a, b]$ and $TY = T[M(j, u)Y(u)] = C$. Suppose X is in $C_{[a,b]}$ and satisfies (**). Then, $X(t) = M(t, s)X(s)$ for each of t and s in $[a, b]$ and so $TX = T[M(j, u)X(u)]$ which means that $R[X(u)] = C$ which in turn implies that $X(u) = U$ and so $X(t) = U + \int_u^t dF \cdot X$ for each t in $[a, b]$. By Theorem B, $X = Y$. Thus the existence of such a u in $[a, b]$ and such a function R implies that (**) has a unique solution.

Suppose that (**) has a unique solution Y which is in $C_{[a,b]}$. Denote by u a number in $[a, b]$. Thus $Y(t) = Y(u) + \int_u^t dF \cdot Y$ and $Y(t) = M(t, u)Y(u)$ for each t in $[a, b]$ and so $TY = T[M(j, u)Y(u)]$. Denote by R the function from $F_{u;[a,b]}$ to S so that $RA = T[M(j, u)A]$ for each A in $F_{u;[a,b]}$. Thus $R[Y(u)] = C$. Suppose that $V \neq Y(u)$ and $RV = C$. Denote by X the element of $C_{[a,b]}$ so that $X(t) = V + \int_u^t dF \cdot X$ for each t in $[a, b]$. $X \neq Y$ as $X(u) \neq Y(u)$. But $X(t) = X(s) + \int_s^t dF \cdot X$ for each of t and s in $[a, b]$ and $TX = [M(j, u)X(u)] = T[M(j, u)V] = RV = C$, a contradiction. Thus there is not such a point V in $F_{u;[a,b]}$ and so the existence of a unique element of $C_{[a,b]}$ satisfying (**) implies the existence of the required function R.

4. An example. Suppose that $[a, b]$ is a number interval, S the number plane, each of p and q a continuous function from $[a, b]$ to a number set such that $p(t) > 0$ for each t in $[a, b]$ and each of a_{ij}, b_{ij} and c_i, $i, j = 1, 2$, a number. The problem of solving

$$(\Delta) \qquad \begin{aligned} &(py')'\, qy = G \\ &a_{11}y(a) + a_{12}p(a)y'(a) + b_{11}y(b) + b_{12}p(b)y'(b) = c_1 \\ &a_{12}y(a) + a_{22}p(a)y'(a) + b_{21}y(b) + b_{22}p(b)y'(b) = c_2 \end{aligned}$$

for each continuous function G from $[a, b]$ to a number set and each ordered number pair (c_1, c_2) is equivalent to the problem of finding a function pair f_1, f_2 each of which is from $[a, b]$ to a number set such that

$$\begin{bmatrix} f_1' \\ f_2' \end{bmatrix} = \begin{bmatrix} 0 & 1/q \\ q & 0 \end{bmatrix} \begin{bmatrix} f_1 \\ f_2 \end{bmatrix} + \begin{bmatrix} 0 \\ G \end{bmatrix}$$

and

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} f_1(a) \\ f_2(a) \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \begin{bmatrix} f_1(b) \\ f_2(b) \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix},$$

i.e., the problem of finding a continuous function f from $[a, b]$ to S such that

$$(\delta) \qquad f = \begin{bmatrix} f_1 \\ f_2 \end{bmatrix},$$

$$f(t) = f(u) + g(t) - g(u) + \int_u^t dF \cdot f$$

and

$$\int_a^b dH \cdot f = A_1 f(a) + A_2 f(b) = C$$

for each of u and t in $[a, b]$ where $g(t) = \begin{bmatrix} 0 \\ G(t) \end{bmatrix}$, $F(t)$ is the linear transformation from S to S associated with

$$\begin{bmatrix} 0 & \int_a^t (1/p)dj \\ \int_a^t q\,dj & 0 \end{bmatrix}$$

for each t in $[a, b]$, each of A_1 and A_2 is a linear transformation from S to S with A_1 associated with $\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$ and A_2 associated with $\begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}$ and H is defined in the following way: $H(a) = N_b$, the transformation which takes each point of S into $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$, $H(u) = A_1$ if $a < u < b$ and $H(b) = A_1 + A_2$. Suppose that M satisfies $M(t, u) = I + \int_u^t dF \cdot M(j, u)$ for each of t and u in $[a, b]$. From § 2, for (δ) to have a unique continuous solution for each g and each C it is necessary and sufficient that $\int_a^b dH \cdot M(j, a) = A_1 + M(b, a)A_2$ have an inverse which is from S onto S_r. Here is $\int_a^b dH \cdot M(j, a)$ has an inverse, it is from S to S and is bounded

Suppose that $\int_a^b dH \cdot M(j, a)$ has an inverse, G is a continuous function from $[a, b]$ to a number set, C is in S and $g = \begin{bmatrix} 0 \\ G \end{bmatrix}$. By Theorem B, there is a function K from $[a, b] \times [a, b]$ to B and a function R from $[a, b]$ to B such that $f(t) = R(t)C + \int_a^b K(t, j)dg$ for each t in $[a, b]$. Denote by each of $R_{ij}, K_{ij}, i, j = 1, 2$ a function from $[a, b]$ to a number set such that if each of t and u is in $[a, b]$, $R(t)$ is associated with

$$\begin{bmatrix} R_{11}(t) & R_{12}(t) \\ R_{21}(t) & R_{22}(t) \end{bmatrix}$$

and $K(t, u)$ is associated with

$$\begin{bmatrix} K_{11}(t, u) & K_{12}(t, u) \\ K_{21}(t, u) & K_{22}(t, u) \end{bmatrix}.$$

Thus, $f_1(t) = R_{11}(t)c_1 + R_{12}(t)c_2 + \int_a^b K_{12}(t, j)dG$ for each t in $[a, b]$ and f_1 is the unique solution to $(\varDelta)$.

References

1. J. S. MacNerney, *Stieltjes integrals in linear spaces*, Ann. of Math., **61** (1955), 354-367.
2. ———, *Continuous products in linear spaces*, J. Elisha Mitchell Sci. Soc., **71** (1955), 185-200.
3. J. W. Neuberger, *Continuous products and nonlinear integral equations*, Pacific J. Math., **8** (1958), 529-549.
4. H. S. Wall, *Concerning continuous continued fractions and certain system of Stieltjes integral equations*, Rend. Circ. Mat. Palermo II (2), (1953) 73-84.
5. ———, *Concerning harmonic matrices*, Arch. Math., **V** (1954), 160-167.

THE ILLINOIS INSTITUTE OF TECHNOLOGY
AND THE UNIVERSITY OF TENNESSEE

INTEGRAL CLOSURE OF DIFFERENTIAL RINGS

EDWARD C. POSNER

We prove that a commutative differentiably simple ring of characteristic zero finitely generated over its field of constants is integrally closed in its field of quotients. (A ring is differentiably simple if it has non-trivial multiplication and has no ideal invariant under a given family of derivations; i.e., has no differential ideals other than (0). The field of constants is the subring of the ring annihilated by each derivation of the family of derivations.) The result of the first sentence is used to obtain a condition that the powers of an element of a function field in one variable form an integral basis. The following results from [1] will be used: A commutative differentiably simple ring of characteristic zero is an integral domain whose ring of constants is a field. Crucial is the following lemma:

LEMMA. *Let F be a field of characteristic zero; $x_1, \cdots, x_n$ be n independent transcendentals over F; $y_1, \cdots, y_q$ be integral over $x_1, \cdots, x_n$; and d an F-derivation of $F[x, y]$ into itself. Then d (or rather its natural extension to $F(x, y)$) sends O_x (the set of elements of $F(x, y)$ integral over $x_1, \cdots, x_n$) into itself.*

Proof. In general any F-derivation of $F(x, y)$ into itself can be written as

$$d = \sum_{i=1}^{n} A_1 \frac{\partial}{\partial x_i} ,$$

A_i elements of $F(x, y)$, $1 \leqq i \leqq n$, Further, d maps $F[x, y]$ into itself if and only if $d(x_i)$ is in $F[x, y]$ for each i and $d(y_j)$ is in $F[x, y]$ for each j. The first set of conditions is equivalent to the condition that A_i be in $F[x, y]$ for each i.

In order to be able to use power series, we assume that F is algebraically closed. For if not, let $\bar{F}$ be its algebraic closure. Let d also be the extension of d to $\bar{F}(x, y)$. Since d sends $\bar{F}[x, y]$ into itself, d send $\bar{O}_x$ into itself, where $\bar{O}_x$ denotes the ring of integral functions of $\bar{F}(x, y)$. A fortiori, d sends O_x into $\bar{O}_x$. But $\bar{O}_x \cap F[x, y] = O_x$ so actually d sends O_x into itself as required.

Let P be a place of $F(x, y)$ over F which has residue field F and which is finnite on $x_1, \cdots, x_n$. We will prove that if g, in $F(x, y)$, is finite at P, $d(g)$ is finite at P. Let a_i denote the residue of x_i at P; then there exist uniformizing parameters $t_1, \cdots, t_n$ at P such that

Received August 13, 1959, and in revised form January 23, 1960.

$x_i - a_i$ is a positive integral power of t_i, say $x_i - a_i = t_i^{p_i}$. Every element B of $F(x, y)$ finite at P has a power series in $t_1, \cdots, t_n$ with coefficients in F. We call the smallest power of t_i occurring in this series the i-order of B at P, and denote it by $\mathrm{ord}_{P,i} B$; the definition of $\mathrm{ord}_{P,i} B$ extends to arbitrary elements B of $F(x, y)$ in an obvious way. Fixing i, we see that if $\mathrm{ord}_{P,i} d(B) \geqq \mathrm{ord}_{P,i} B$ for every B finite at P then $\mathrm{ord}_{P,i} d(B) \geqq 0$ for every such B. Suppose there exists some B finite at P with $\mathrm{ord}_{P,i} d(B) < \mathrm{ord}_{P,i} B$. Then $\alpha_i - p_i < 0$, where $\alpha_i = \mathrm{ord}_{P,i} A_i$, so that $r_i = p_i - \alpha_i > 0$, and $\mathrm{ord}_{P,i} B = r_i + \mathrm{ord}_{P,i} dB$ for every (B) in $F(x, y)$ with $\mathrm{ord}_{P,i} B \neq 0$. Since d maps $F[x, y]$ into itself, the only values which $\mathrm{ord}_{P,i} B$ can have when B is in $F[x, y]$ are integral multiples of r_i, for otherwise some element of $F[x, y]$ would have negative i-order. Since $t_1, \cdots, t_n$ are uniformizing parameters, it follows that $r_i = 1$, for otherwise we could replace t_i by $t_i^{r_i}$. Thus, d drops positive i-orders by 1, so that $\mathrm{ord}_{P,i} d(B) \geqq 0$ for every B finite at P. Since this holds for every i, $d(B)$ is finite at P whenever B is. Since this holds for every P, we conclude that d maps O_x into itself.

THEOREM 1. *Let F be a field of characteristic zero, $A = F[z_1, z_2, \cdots, z_k]$ a commutative finitely generated ring extension of F. Let D be a (finite or infinite) family of derivations of A into itself over F. Let A be differentiably simple under D. Then A is integrally closed in its quotient field K.*

Proof. A is an integral domain by (1). By Noether's Normalization Lemma, we can write $A = F[x_1, \cdots, x_n; y_1, \cdots, y_q]$, with n the transcendence degree of K/F and $y_1, \cdots, y_q$ interal over $x_1, \cdots, x_n$. To prove $A = O_x$, let I denote the conductor of O_x, that is, the set of elements u of $F[x, y]$ such that $u \cdot O_x \subset F[x, y]$; by [3], pp. 271–2, prop. 6, I is a non-zero ideal of $F[x, y]$. To prove I differential under D, let d be in D, h be in I, g be in O_x. Then $h \cdot g$ is in $F[x, y]$, $d(h \cdot g)$ is in $F[x, y]$, $d(h)g + hd(g)$ is in $F[x, y]$. Now $d(g)$ is in O_x by the lemma so $hd(g)$ is in $F[x, y]$ since h is in I. Then $d(h)g$ is in $F[x, y]$, I is differential under D. Then $I = F[x, y]$ so $1 \cdot O_x \subset F[x, y]$, $O_x = F[x, y]$ as promised.

REMARK. D can always be taken to be finite since the derivations of $F[x, y]$ into itself form a finite $F[x, y]$-module.

The converse of Theorem 1 is false, i.e., there are integrally closed finitely generated domains which are not differentiably simple under any family of F-derivations. For example, let $y^2 = x_1^3 + x_2^3$. Then $F[x, y] = O_x$ but is not differentiably simple over F. In fact, the ideal (x_1, x_2, y) of $F[x, y]$ is differential for any derivation, as is easy to see. But when $n = 1$, we do have the converse. (For background material, see [2];

pp. 83–88.)

THEOREM 2. *Let K be a function field in one variable over a field F of characteristic zero, and let x be an element of K transcendental over F. Let O_x denote the set of elements of K integral over x. Then O_x is differentiably simple with field of constants F under a family of two or fewer derivations.*

Proof. First we shall specify the derivations. O_x is a Dedekind ring, i.e., every ideal of O_x is invertible. Let $K = F(x, y)$ with y integral over x and let $f(x, y) = 0$ be the irreducible monic for y. Define d on K by

$$d(g(x, y)) = \frac{\partial g}{\partial x}\frac{\partial f}{\partial y} - \frac{\partial g}{\partial y}\frac{\partial f}{\partial x}\,.$$

This is well-defined, and d sends O_x into itself by the lemma. Let J be the ideal of O_x generated by the values of d of integral elements. J is invertible, so there exist $h_i(x, y)$ in K, $1 \leqq i \leqq q$, such that $h_i d$ sends O_x into itself and such that there exist u_i in O_x, $1 \leqq i \leqq q$, with $\sum_{i=1}^{q} h_i d(u_i) = 1$. ($q$ can be taken to be 2. For J is generated by f_x and f_y, since $d(M(x, y)) = f_y M_x - f_x M_y$ for M in K. q can be taken to be 1 if and only if J is principal, which need not occur.) The family D is $\{h_1 d, \cdots, h_q d\}$. To prove O_x differentiably simple under D, suppose the contrary. As in the preceding and following theorems, F may be assumed to be algebraically closed. If O_x has a non-zero differential ideal, it has a maximal differential ideal I, since O_x has a unit. O_x^2 is not contained in I, so by Theorem 4 of [1], I is prime. But every prime ideal of O_x is maximal; in fact, if w belongs to O_x, there is a λ in F with $w - \lambda$ in I. Since I is differential for D, $h_i d(w) - h_i d(\lambda)$ is in I, $1 \leqq i \leqq q$, $h_i d(w)$ is in I, $1 \leqq i \leqq q$. That is, $h_i d(w)$ is in I for all w in O_x. Then $\sum_{i=1}^{q} h_i d(u_i) = 1$, 1 is in I, $I = O_x$. This contradiction proves that O_x has no differential ideals. Its field of constants is F. For if u is in $F(x, y)$ and $d(u) = 0$ then $(d/dx)(u) = 0$, so that u belongs to F.

THEOREM 3. *Let K, F, x, O_x be as in the hypothesis of Theorem 2. Let R be an order of O_x and let y be an element of K integral over x with irreducible monic f such that $K = F(x, y)$. Then $R = O_x$ if and only if y belongs to R and the ideal J in R generated by f_x and f_y is invertible.*

Proof. If $R = O_x$, then y belongs to R and every ideal in R is invertible. Conversely, suppose that y belongs to R and that J, the ideal generated in R by the values of d, is invertible. (Here d is the same derivation as in Theorem 2.) That is, assume that there exist h_i

in K, $1 \leqq i \leqq q$, with $h_i d$ sending R into itself, and elements v_i in R, $1 \leqq i \leqq q$, with $1 = \sum_{i=1}^{q} h_i d(v_i)$. We shall prove R differentiably simple under $D = \{h_1 d, \cdots, h_q d\}$. It is known that every prime ideal of R is maximal; it fact, if I is a prime ideal of R, and w is an element of R, there is a λ in F with $w - \lambda$ in I. If R has a differential ideal, it has a maximal differential ideal, and one proceeds as in Theorem 2. So R is differentiably simple under D. By Theorem 1, R is integrally closed in K, i.e., $R = O_x$ as required.

As an illustration, let $K = F(x, y)$ with $f(x, y) = y^n - P(x) = 0$, $n \geqq 1$, P a polynomial in x with no repeated roots. Here $R = F[x, y]$. Let us examine the ideal in $F[x, y]$ generated by f_x and f_y, i.e., by $P'(x)$ and y^{n-1}. This ideal contains $y^{n-1}y = y^n = P(x)$ and $p'(x)$. $P(x)$ and $P'(x)$ have no common factor, so there are polynomials $Q(x)$ and $S(x)$ with $QP + SP' = 1$. Then the ideal generated by f_x and f_y is $F[x, y]$ and so is trivially invertible. We conclude $F[x, y] = O_x$.

BIBLIOGRAPHY

1. E. C. Posner, *Differentiably simple rings*, Proc. Amer. Math. Soc., **11** (1960), 337343.

2. B. L. Van der Waerden, *Modern algebra*. Volume II. New York: Frederick Ungar Publishing Company, 1950.

3. A. Weil, *Foundations of algebraic geometry*, New York: American Mathematical Society, 1946.

THE UNIVERSITY OF WISCONSIN
MADISON, WISCONSIN

SOME THEOREMS ON MAPPINGS ONTO

M. REICHBACH

Introduction and summary. Let $F: X \to Y$ be a continuous mapping of a topological space X into a space Y. An important problem is to find conditions[1] under which this mapping is a mapping onto: $F(X) = Y$. In the present paper, the following consideration is used in proving theorems on mappings onto.

Conditions are given under which the image $F(X)$ is closed and open in Y; hence for a connected $Y, F(X) = Y$.

This idea is not new. For instance C. Kuratowski[2] showed that, i, a subgroup G of a topological additive group X has the Baire propertyf then either G is of the first category in X or G is *both open and closed in* X, *so that* $G = X$ *if* X *is connected.*

It was also used by the author in [10], to obtain some generalizations of the Fundamental Theorem of Algebra.

In this paper the results obtained in [10] are generalized to general topological spaces.

In § I the notion of a "polynomial mapping" is introduced. Roughly speaking, a mapping $F: X \to Y$ is called a polynomial mapping if it maps every sequence which does not contain a convergent subsequence onto a sequence which also does not contain a convergent subsequence. It is proved that a polynomial mapping $F: X \to Y$ of a complete space X into a space Y maps sets closed in X onto sets closed in Y.[3]

The role of the disconnection properties in the proofs of theorems on mappings onto is discussed and a generalization of the Fundamental Theorem of Algebra to n-dimensional Euclidean spaces is obtained.

In § II some theorems on mappings onto are proved for the so-called generalized F-spaces and the Fundamental Theorem of Algebra is generalized to such spaces. Finally an application of this generalization to an existence theorem in some class of functional equations is given. For the sake of completeness, many known definitions are recalled.

I, 1. Let X be a space in which convergence satisfying the following two conditions is defined:

(a_0) if $x_n \to x$ and $x_n \to y$ then $x = y$

(a) if $x_n \to x$ and $k_1 < k_2 < \cdots$, then $x_{k_n} \to x$

The set of all convergent sequences $\{x_n\} \subset X$ will be denoted by L. Note that

Received December 21, 1959.

[1] For examples of such conditions see [2], chapter XI.

[2] See [6], p. 38 and [7], p. 81. Also [4], p. 8.

[3] According to the terminology used by Whyburn in [11], we can say that if $F: X \to Y$ is a polynomial mapping, then it is a strongly closed mapping.

(b) if $\{x_n\} \in L$, $k_1 < k_2 < \cdots$ and $x_{k_n} \to x$ then $x_n \to x$
Indeed, since $\{x_n\} \in L$, there exists a point x_0 such that $x_n \to x_0$. Hence by (a), $x_{k_n} \to x_0$ and since $x_{k_n} \to x$ we obtain by (a_0), $x_0 = x$. Thus $x_n \to x$.

In the usual way, we can introduce the notions of a subspace P of X and closedness-and-openness in X (or in P). A set $P \subset X$ is called connected if it is not a union of two non-empty disjoint sets closed in P.

Let C be a set of sequences $\{x_n\} \subset X$ such that $L \subset C$. The set C will be called the set of Cauchy (or fundamental) sequences. If $L = C$, the space is called complete. Note that since the definition of C is quite arbitrary (it is only needed that $L \subset C$) we can put $C = L$ and the space will be a complete space.

I, 2. A mapping $F: X \to Y$ of a space X into a space Y is called continuous if for every sequence $\{x_n\}_{n=1,2,\dots} \subset X$ the condition $x_n \to x$ implies $F(x_n) \to F(x)$[4]. By "mappings" we shall in the sequel understand continuous mappings only. We introduce now the following

DEFINITION 1. A sequence $\{x_n\}_{n=1,2,\dots}$; $x_n \in X$ is called a non-Cauchy sequence or simply a NC sequence if it does not contain a subsequence belonging to C.

If the set of Cauchy sequences is defined as usual, then

(c) in a *finite* dimensional Banach space the set of NC sequences is identical with the set of sequences $\{x_n\}_{n=1,2,\dots}$ with $x_n \to \infty$.

Indeed, if $x_n \to \infty$, then by the completeness of the Banach space, $\{x_n\}$ cannot contain a Cauchy sequence $\{x_{k_n}\}$, since otherwise, there would be $x_{k_n} \to x$ for some x, which is impossible, by $x_n \to \infty$. On the other hand if x_n does not tend to ∞, there exists a bounded subsequence $\{x'_n\}$ of $\{x_n\}$ and since X is finite dimensional, $\{x'_n\}$ contains a convergent subsequence $\{x'_{k_n}\}$, which is a Cauchy sequence.

DEFINITION 2. A mapping $F: X \to Y$ is called a polynomial mapping[5] if the condition

$$\{x_n\} \text{ is a } NC-\text{sequence}, \ x_n \in X$$

implies that

$$\{F(x_n)\} \text{ is a } NC-\text{sequence}, \ F(x_n) \in Y.$$

By (c) we obtain that

(d) Polynomial mappings $F: X \to Y$ of a *finite* dimensional Banach space X into a finite dimensional Banach space Y are identical with

[4] In fact we should denote the convergence relation in Y by a symbol other than "$\to$" used for convergence in X, but the meaning of "$\to$" will always be clear from the text.

[5] The definition of a polynomial mapping was first introduced by the author in **[10]** for metric spaces. The definition introduced here is a generalization of this definition to general topological spaces.

those which map sequences $\{x_m\}_{m=1,2,\ldots}$ tending to ∞ onto sequences $\{F(x_m)\}_{m=1,2,\ldots}$ also tending to ∞.

In particular, a ($\neq$ constant) complex polynomial in the complex plane (2-dimensional Banach space) is a polynomial mapping of the plane into itself. This justifies the notion "polynomial mapping".

We prove now the following

LEMMA 1. If $F: X \to Y$ is a polynomial mapping of a complete space X into a space Y, then for every set A closed in X the image $F(A)$ is closed in Y. (In particular $F(X)$ is closed in Y).

Proof. Let $y_n \in F(A)$ be points belonging to $F(A)$ and let $y_n \to y$. We shall show that $y \in F(A)$. Indeed, there exist points $x_n \in A$ such that $y_n = F(x_n)$. Now $\{x_n\}$ cannot be a NC – sequence since $\{F(x_n)\}$ would also be a NC sequence (F being polynomial mapping) and this is impossible by $L \subset C$ and $y_n \to y$. Therefore, there exists a subsequence $\{x_{k_n}\}$ of $\{x_n\}$ which belongs to C. The space X being complete, there is $x_{k_n} \to x$ for some x and $x \in A$ since A is closed. Thus by the continuity of F, $F(x_{k_n}) \to F(x)$. But $\{F(x_{k_n})\}$ is a subsequence of $\{y_n\}$ and therefore by (b) we have $y_n \to F(x)$. Hence by $y_n \to y$, $F(x) = y$ and by $x \in A$, there is $y \in F(A)$.

DEFINITION 3. A mapping $F: X \to Y$ is said to be open in the point $y_0 \in F(X)$ if there exists an open (in Y) set $U(y_0)$ containing y_0, such that $U(y_0) \subset F(X)$.

Evidently, $F(X)$ is open in Y if and only if $F: X \to Y$ is open in every point $y_0 \in F(X)$.

THEOREM 1. *If $F: X \to Y$ is a polynomial mapping of a complete space X into a connected space Y, which is open in every point $y \in F(X)$, then $F(X) = Y$.* (i.e. $F: X \to Y$ is a mapping onto).

Proof. By Lemma 1, $F(X)$ is closed in Y and by the assumption, $Y - F(X)$ is closed in Y. Hence by the connectedeness of Y there is $F(X) = Y$.

I, 3. We shall now investigate the role of the disconnection properties of subsets of Y in the proofs of theorems on mappings onto. Throughout § I, 3 we shall assume that our spaces satisfy the first countability axiom and thus all the topological relations may be expressed in terms of convergent sequences. The following Lemma is evident.

LEMMA 2. *A mapping $F: X \to Y$ is not open in the point $y \in F(X)$ if and only if $y \in Fr(F(X))$, where $Fr(F(X))$ denotes the boundary of*

$F(X)$ in Y.[6]

We prove now the following

THEOREM 2. *If $F: X \to Y$ is a polynomial mapping of a complete space X into a connected space Y which is open in every point $y \in F(X) - J$, where $J \subset F(X)$ is a set which does not disconnect the space Y and if $F: X \to Y$ is open in at least one point $y_0 \in F(X)$, then $F(X) = Y$.*[7]

Proof. Since $F: X \to Y$ is open in the point $y_0 \in F(X)$ there exists an open (in Y) set $U(y_0) \subset F(X)$. Denote by U the union of all sets open in Y, which are contained in $F(X)$. Evidently $Fr(U) \subset Fr(F(X))$. Now suppose, that there would exist a point $x_0 \in Y - F(X)$ and let x be any point belonging to U. Since the set J does not disconnect Y there exists in $Y - J$ a connected set K containing x_0 and x. But the set $Fr(U)$ disconnect the space Y between x_0 and x.[8] Therefore there exists a point $y \in [Fr(U) - J] \cap K$ and since $Fr(U) \subset Fr(F(X))$ the point $y \in [Fr(FX)) - J] \cap K \subset Fr(F(X)) - J$.

By Lemma 1, $F(X)$ is closed in Y and therefore $y \in F(X)$. But, by assumption $F: X \to Y$ is open in every point $y \in F(X) - J$, which by Lemma 2 contradicts the fact that $y \in Fr(F(X)) - J$. Thus the assumption that there exists a point $x_0 \in Y - F(X)$ leads to a contradiction.

I, 4. We prove now the

First generalization of the Fundamental Theorem of Algebra. Let $F: X \to X$ be a mapping of the n-dimensional Euclidean space X, with $n \geqq 2$ into itself defined by $\eta_i = \eta_i(\xi_1, \cdots, \xi_n)$ $i = 1, 2, \cdots n$, where the real functions η_i and their derivatives $\partial\eta_i/\partial\xi_k$ are continuous in X. If then the Jacobian $D = \partial(\eta_1, \cdots \eta_n)/\partial(\xi_1, \cdots, \xi_n) \neq 0$ in every point $x = x(\xi_1 \cdots \xi_n) \in X - J_0$, where J_0 is a countable set and if the condition $x_m(\xi_1^m, \cdots \xi_n^m) \to \infty$ implies $F(x_m) \to \infty$ for every sequence of points $x_m \in X$, then $F: X \to X$ is a mapping onto: $F(X) = X$.

Proof. Since the condition $x_m \to \infty$ implies $F(x_m) \to \infty$ we obtain by (d) that $F: X \to X$ is a polynomial mapping. Now by the countability of J_0 the set $J = F(J_0)$ is also countable and therefore 0-dimensional.[9] Hence, since the dimension n of X is $\geqq 2$, J does not disconnect X.[10] Further, if $y = F(x)$ is any point of $F(X) - J$, the mapping $F: X \to X$ is open in y, because, by the assumption $D \neq 0$ for points $x \notin J_0$, a neighbourhood (in X) of $y = F(x)$ is covered. Finally, since J_0 is countable there exists at least one point y_0 in which $F: X \to X$ is open. Thus put-

[6] If X is a space and A a subset of X, the boundary $Fr(A) = \bar{A} \cap \overline{X - A}$.

[7] This Theorem was suggested to the author by H. Hanani.

[8] S. [**3**], p. 247, also [**8**], p. 80.

[9] In the sense of Menger-Urysohn

[10] See [**5**], p. 48.

ting $Y = X$ in Theorem 2, we see that all the assumptions of this theorem hold. Therefore $F(X) = X$.

REMARK 1. If $F: X \to X$ is any ($\neq$ constant) complex polynomial defined on the complex plane X, then the mapping $F: X \to X$ is defined by two real functions $\eta_1 = \eta_1(\xi_1, \xi_2)$, $\eta_2 = \eta_2(\xi_1, \xi_2)$, where η_1 and η_2 are respectively, the real and imaginary parts of $F(x) = \eta_1 + i\eta_2$, $x = \xi_1 + i\xi_2 \in X$, $i^2 = -1$. Now for the Jacobian D there is $D = |F'(x)|^2 \neq 0$ except for a finite set of points ($F'(x)$ denotes the derivative of $F(x)$) and therefore, by the above generalization of the Fundamental Theorem of Algebra, $F(X) = X$.

REMARK 2. Our proof of the first generalization of the Fundamental Theorem of Algebra is based on Theorem 2. An essential role in Theorem 2 is played by the assumption that the set J does not disconnect Y. This assumption is satisfied because the dimension of the Euclidean space X is assumed to be $\geqq 2$. (A countable set does not disconnect an Euclidean space with dimension $\geqq 2$). This explains the role, for the Fundamental Theorem of Algebra, of the fact that the dimension of the Euclidean plane is $\geqq 2$.

II, 1. Let now X be a space (in the sense of I, 1) which is simultaneously a linear space (with multiplication by real or complex numbers). We introduce the following

DEFINITION 4. A mapping $F: X \to X$ of a linear space into itself is said to have a lower bounded rate of change in the point $y_0 \in F(X)$ if there exists a point $x_0 \in F^{-1}(y_0)$, a number (may be complex) $\lambda(x_0) \neq 0$ and an open in X set $U(y_0)$ containing y_0, such that for every $y' \in U(y_0)$ the sequence: $x'; Ax'; AAx'; \cdots$ where $Ax = x - \lambda(x_0)(F(x) - y')$ is a Cauchy sequence for some point $x' \in X$ (the point x' depends on y').

LEMMA 3. *If $F: X \to X$ is a mapping of a complete linear space X into itself, having a lower bounded rate of change in the point $y_0 \in F(X)$, then $F: X \to X$ is open in the point y_0.*

Proof. Let x'; x_0, $\lambda(x_0) \neq 0$ and $U(y_0)$ be the points, the number and the open (in X) set defined in the foregoing definition and let $y' \in U(y_0)$ be any point of $U(y_0)$. We shall show that $y' \in F(X)$. Indeed, since the sequence $x^-; Ax'; AAx'; \cdots$ is a Cauchy sequence and X is complete, it has a limit x'_0. Now by the continuity of A, we have $Ax'_0 = x'_0$, i.e. $x'_0 - \lambda(x_0)(F(x'_0) - y') = x'_0$ and hence $F(x'_0) = y'$.

II, 2. Here we shall introduce the notion of a generalized F-space and prove some theorems on mappings onto in these spaces. We begin with the definition of a generalized metric space.

DEFINITION 5. A set X of points is called a generalized metric space with metric ρ if on the Cartesian product $X \times X$ a non-negative real function $\rho(x, y)$ is defined: $0 \leqq \rho(x, y) \leqq \infty$, $x, y \in X$ which satisfies the usual axioms of any metric, i.e. $\rho(x, y) = 0$ if and only if $x = y$, $\rho(x, y) = \rho(y, x)$ and $\rho(x, y) + \rho(y, z) \geqq \rho(x, z)$; x, y, z, belong to X.

Thus the difference between the definitions of a metric space and of the generalized metric space consists in the fact that the function $\rho(x, y)$ may assume the value of ∞.

EXAMPLE 1. Take the set X of all real continuous functions $x(t)$, $-\infty < t < \infty$ and define $\rho(x, y) = \sup_t |x(t) - y(t)|$. Here, X is a generalized metric space, but not a metric space.

Evidently every metric space is also a generalized metric space. In generalized metric spaces we can define convergent sequences by saying that $x_n \rightarrow x$ if $\rho(x_n, x) \rightarrow 0$ and this convergence satisfies (a_0) and (a). Thus every generalized space is a space in the sense of I, 1. If we define, as usual, a Cauchy sequence in X by saying that $\{x_n\}_{n=1,2,\cdots} \in C$ if for every $\varepsilon > 0$ there exists a $N(\varepsilon)$ such that $\rho(x_n, x_m) < \varepsilon$ for $n, m > N(\varepsilon)$ then we have $L \subset C$. The set X in Example 1 is, as is easy to see, a complete space. Let now X be a generalized metric space which is simultaneously a linear space (with multiplication by real or complex numbers) such that the following two conditions hold

(e) $\rho(x, y) = \rho(x - y, 0)$

and

(f) $(hx, 0) \leqq |h| . \rho(x, 0)$ for every number h and every point $x \in X$.

In such a case we shall call X a generalized F-space. For instance, the space X in Example 1 is a generalized F-space. Another example of a generalized F space is the set of all sequences $x = (\xi_1, \xi_2, \cdots)$ of real numbers ξ_i, $i = 1, 2, \cdots$ if we define $\rho(x, y) = \sup_i |\xi_i - \eta_i|$ where $y = (\eta_1, \eta_2, \cdots)$. Note that on the set on which ρ is finite, it satisfies the axioms of the so-called F-spaces.[11] This justifies the notion of generalized F-space.

Property T. Let $F: X \rightarrow X$ be a mapping of a generalized F-space into itself such that for the point $y_0 \in F(X)$ there exists a point $x_0 \in F^{-1}(y_0)$, a spherical region $S(x_0, r)$,[12] a complex function $\lambda(x_0) \neq 0$ and a real function $\alpha = \alpha(x_0)$: $0 < \alpha < 1$ such that for any two points x and y

[11] See [**1**], p. 35.

[12] A spherical region $S(x_0, r)$ with x_0 as centre and r as radius is defined as the set of all points $x \in X$ satisfying $\rho(x_0, x) < r$.

belonging to $S(x_0, r)$ there is

(g) $\rho[x - y - \lambda(x_0)(F(x) - F(y)), 0] \leqq \alpha\rho(x, y)$.

A mapping $F: X \to X$ for which (g) holds is said to have the property T in the point $y_0 \in F(X)$.[13]

We prove now

LEMMA 4. *If $F: X \to X$ is a mapping of a generalized F-space into itself having the property T in the point $y_0 \in F(X)$, then $F: X \to X$ has a lower bounded rate of change in the point $y_0 \in F(X)$.*

Proof. Let $S(x_0, r), \lambda(x_0) \neq 0$ and $\alpha = \alpha(x_0)$ be the spherical region and the functions appearing in the definition of the property T and put $r_0 = [(\varDelta - \alpha)/|\lambda(x_0)|]\cdot r$. It suffices to show that for every $y' \in S(y_0, r_0)$ the sequence $x_0, Ax_0, AAx_0, \cdots$ where $Ax = x - \lambda(x_0)(F(x) - y')$ is a Cauchy sequence:[14]

We have for $x_1 = Ax_0$;

$$\begin{aligned}\rho(x_0, x_1) &= \rho(\lambda(x_0)(F(x_0) - y'), 0) = \rho[\lambda(x_0)(y_0 - y'), 0]\\ &\leqq |\lambda(x_0)|\,\rho(y_0, y') \leq |\lambda(x_0)|\cdot r_0 = (1 - \alpha)r .\end{aligned}$$

Thus $\rho(x_0, x_1) \leqq (1 - \alpha)r$. Now for $x, y \in S(x_0, r)$ we have by (g):

$$\text{(g)};\ \rho(Ax, Ay) = \rho(x - y - \lambda(x_0)(F(x) - F(y)), 0) \leqq \alpha\rho(x, y)$$

and therefore $\rho(Ax_0, Ax_1) \leqq \alpha\rho(x_0, x_1) \leqq (1 - \alpha)\cdot\alpha\cdot r$. Denoting $x_n = Ax_{n-1}$, $n = 1, 2, \cdots$ we obtain by induction $\rho(x_n, x_{n+1}) \leqq (1 - \alpha)\alpha^n r$, hence by $0 < \alpha < 1$ it is easily seen that $x_0, Ax_0, AAx_0, \cdots$ is a Cauchy sequence.

From Lemmas 3 and 4, and from Theorem 2, we obtain the

Second generalization of the Fundamental Theorem of Algebra. If $F: X \to X$ is a polynomial mapping of a complete generalized F-space X into itself, having the property T in every point $y \in F(X) - J$ where $J \subset F(X)$ is a set which does not disconnect the space X, and if there exists at least one point $y_0 \in F(X)$ in which $F: X \to X$ has the property T, then $F(X) = X$.

Proof. By Lemma 4, $F: X \to X$ has a lower bounded rate of change in every point $y \in F(X) - J$; hence by Lemma 3 $F: X \to X$ is open in every point $y \in F(X) - J$ where by assumption J does not disconnect X. Also, by assumption $F: X \to X$ is open in the point $y_0 \in F(X)$. Now, since a generalized F-space is connected (as a linear space) we see, by putting $Y = X$ in Theorem 2, that the assumptions of this theorem hold. Hence $F(X) = X$.

---

[13] For some ideas concerning this definition the author is indebted to D. Tamari.

[14] The following part of the proof is analogous to the proof of Banach's so-called theorem of contraction mappings, see [**9**] p. 47 and remark 2 p. 49.

REMARK 3. We shall now show that the above theorem is in fact a generalization of the Fundamental Theorem of Algebra. Let $F: X \to X$ be a ($\neq$ constant) polynomial, mapping the complex plane X into itself. Since every Banach space is evidently a generalized F-space, the complex plane with the usual metric $\rho(x, y) = |x - y|$ is a generalized F-space. Now take any point $x_0 \in X$ such that the derivative $F'(x_0) \neq 0$ and let $y_0 = F(x_0)$. Put $\lambda(x_0) = 1/(F'(x_0))$ and $\alpha = 1/2$. Since for $x \to x_0$ we have $(F(x_0) - F(x))/(x_0 - x) \to F'(x_0)$ there exists a spherical region $S(x_0, r)$ such that for $x, y \subset S(x_0, r), x \neq y$ there is

$$\left| \frac{F(x) - F(y)}{x - y} \cdot \frac{1}{(F'(x_0)} - 1 \right| \leqq \frac{1}{2}$$

and therefore $|x - y - (1/(F'(x_0))(F(x) - F(y))| \leqq \alpha |x - y|$. But this inequality holds evidently also for $x = y$ and therefore the mapping $F: X \to X$, defined by the complex polynomial $F(x), x \in X$, has the property T in every point $y_0 = F(x_0)$ for which $F'(x_0) \neq 0$. Now the set J of points $y = F(x)$ for which $F'(x) = 0$ is finite and thus it does not disconnect the complex plane X. Hence by the above second generalization of the Fundamental Theorem of Algebra there is $F(X) = X$, i.e., a complex polynomial maps the complex plane onto itself.

II, 3. It is known that

(h) a n-dimensional set does not disconnect the $n + 2$-dimensional Euclidean space.[15]

Thus

LEMMA 5. *A finite dimensional subset J of an infinite dimensional Banach space X does not disconnect X.*

Proof. Let $x_0 \in X - J$ be any fixed point and x any point of $X - J$. Suppose that the dimension of J is n and take any $(n + 2)$-dimensional plane E^{n+2} (homeomorphic with the Euclidean plane E^{n+2}) containing the points x_0 and x: $E^{n+2} \subset X$. Since the set $E^{n+2} \cap J$ is at most n-dimensional it does not disconnect E^{n+2} (by (h)) and therefore there exists a connected set $K \subset E^{n+2} - J \subset X - J$ which contains the points x_0 and x. Thus every point $x \in X - J$ may be joined with the point $x_0 \in X - J$ by a connected set $K \subset X - J$ i.e., the set $X - J$ is connected.

Let now $\| \ \|$ denote the norm in the Banach space X and define $\rho(x, y) = \| x - y \|$.

We prove the following

THEOREM 3. *Let $F: X \to X$ be a polynomial mapping of an infinite*

[15] S. [5], p. 48. The term "dimension" is used in the sense of Menger-Urysohn.

dimensional Banach space X into itself, which maps finite dimensional sets onto finite dimensional sets. If $F: X \to X$ has the property T in every point $y \in F(X - J_0)$ where J_0 is a finite dimensional set, then $F(X) = X$.

Proof. Since J_0 has a finite dimension there exists a point $x_0 \in X - J_0$ and thus the mapping $F: X \to X$ has the property T in the point $y_0 = F(x_0)$. Now, the set J of points in which the mapping does not have the property T is contained in $F(J_0)$ and since $F: X \to X$ maps finite dimensional sets onto finite dimensional sets the set J does not disconnect the space X (by Lemma 5). Hence by the second generalization of the Fundamental Theorem of Algebra there is $F(X) = X$.

Analogously to Lemma 5 it can be proved that

LEMMA 6. *A 0-dimensional set J does not disconnect a n-dimensional Banach space for $n \geqq 2$.*[16]

Hence

THEOREM 4. *Let $F: X \to X$ be a polynomial mapping of a n-dimensional Banach space X into itself, with $n \geqq 2$. If $F: X \to X$ has the property T in every point $y \in F(X - J_0)$ where J_0 is a countable set, then $F(X) = X$.*

Proof. For the proof, it suffices to note that the set $F(J_0)$ is countable and thus, by Lemma 6, does not disconnect the space X. The rest of the proof is analogous to that of Theorem 3 and may be left to the reader.

II, 4. *An application.* Let X be the generalized F-space of all real continuous functions $x(t)$ defined on the real line $-\infty < t < \infty$ with metric $\rho(x, y) = \sup_t |x(t) - y(t)|$ and let $\phi(t, u)$ be a real continuous function defined for $-\infty < t < \infty$, $-\infty < u < \infty$ satisfying the conditions:

(i) There exists a real number $m > 0$ such that for every pair $u_1 \geqq u_2$ of numbers there is $\phi(t, u_1) - \phi(t, u_2) \geqq m(u_1 - u_2)$.

(j) For each function $x_0(t) \in X$ there exist numbers $r > 0$ and M (depending on $x_0(t)$ and r) such that for $x(t)$ and $y(t) \in S(x_0, r)$ there is $|\phi(t, x(t)) - \phi(t, y(t)| \leqq M|x(t) - y(t)|$ for every t: $-\infty < t < \infty$.

Then, the mapping $F(x(t)) = \phi(t, x(t))$ maps X onto X.

Proof. We shall first show that $F: X \to X$ is a polynomial mapping. Indeed, we have by (i) $\rho(F(x_n), F(x_m)) \geqq m \cdot \rho(x_n, x_m)$ for every pair $x_n(t)$ and $x_m(t)$ of functions. Therefore, if the sequence $\{F(x_n)\}$ would contain

[16] This Lemma and Theorem 4 were suggested to the author by H. Hanani.

Cauchy subsequence $\{F(x_{k_n})\}$ the sequence $\{x_{k_n}\}$ would be a Cauchy subsequence of the sequence $\{x_n\}$. Thus $F: X \to X$ maps NC sequences onto NC sequences, i.e., it is a polynomial mapping.

Since our space is complete it suffices, by the second generalization of the Fundamental Theorem of Algebra, to prove that $F: X \to X$ has the property T in every point $y_0 = F(x_0) \in F(X)$. Indeed, take any two points $x(t)$ and $y(t)$ belonging to $S(x_0, r)$. Then for t such that $x(t) \geqq y(t)$ we have by (i) and (j) $m(x(t) - y(t)) \leqq F(x(t)) - F(y(t)) \leqq M(x(t) - y(t))$, hence $-(m/M)[x(t) - y(t)] \geqq -(1/M)\cdot[F(x(t)) - F(y(t))] \geqq -[x(t) - y(t)]$. Thus $(1 - m/M)\cdot[x(t) - y(t)] \geqq x(t) - y(t) - 1/M[F(x(t)) - F(y(t))] \geqq 0$. Therefore for any t such that $x(t) \geqq y(t)$ we have

(k) $(1 - m/M) \vee |x(t) - y(t)| \geqq |x(t) - y(t) - (1/M)[F(x(t)) - F(y(t))]|$.

Analogously, for any t such that $y(t) \geqq x(t)$, (k) holds and therefore (k) holds for every t. Thus assuming that $M > m$ and putting $\lambda = 1/M$ and $\alpha = 1 - m/M$ we see by (k) that $F: X \to X$ has the property T in the point $y_0 = F(x_0)$.

EXAMPLE 2. If $\phi(t, u)$ is a real continuous function defined for $-\infty < t < \infty$ and $-\infty < u < \infty$ having a continuous derivative $\phi_u(t, u)$ such that there exist constants m and $M: M > m > 0$ for which

$$m \leqq \phi_u(t, u) \leqq M$$

for every t and u, then evidently (i) and (j) hold and hence the function $F(x(t)) = \phi(t, x(t))$ maps X onto X. Such a function $\phi(t, u)$ is for instance the function $\phi(t, u) = 2u + 3t + \sin(u + t)$.

REMARK 4. An analogous theorem to the above application was proved in [10] for the space X of all real continuous functions $x(t)$ defined in a *finite* interval $a \leqq t \leqq b$[17].

REFERENCES

1. S. Banach, *Theorie des operations linearies*, Hafner Publ. Co. (1932), New York.

2. S. Eilenberg, and N. Steenrod, *Foundations of algebraic topology*, Princeton University Press, (1952), Princeton, New Jersey.

3. F. Hausdorff, *Grundzüge der Mengenlehre*, (1949), Chelsea Publ. Comp. New York.

4. E. Hille, *Functional analysis and semigroups*, Amer. Math. Soc., Colloquium Publ. Vol. XXXI, (1948), New York.

5. W. Hurewicz, and H. Wallman, *Dimension Theory*, Princeton University Press, (1948), Princeton.

6. C. Kuratowski, *Sur la propriété de Baire dans les groupes métriques*, Studia Math., 4, (1933), 38-40.

7. ———, *Topologie I*, Monografie Matematyczne, (1952), Warszawa.

8. ———, *Topologie II*, Monografie Matemtyczne, (1952), Warszawa.

[17] See [**10**], p. 162, Application of Theorem 4.

9. L. A. Lusternik, and W. J. Sobolew, *Elementy funkcjonalnago analiza,* (1951), Moskva-Leningrad.
10. M. Reichbach, *Generalizations of the fundamental theorem of Algebra,* Bulletin of the Research Council of Israel, 7F, (1958), 155-164.
11. G. T. Whyburn, *Topological analysis,* Princeton University Press, (1958), Princeton, New Jersey.

DEPARTMENT OF MATHEMATICS, TECHNION,
ISRAEL INSTITUTE OF TECHNOLOGY, HAIFA

TWO EXTREMAL PROBLEMS

MARVIN ROSENBLUM AND HAROLD WIDOM

1. Introduction. Let $\mathscr{P}_0$ be the class of all complex trigonometric polynomials P of the form $P_0 + P_1e^{i\phi} + P_2e^{2i\phi} + \cdots$. Let σ and μ be, respectively normalized Lebesgue measure and any finite non-negative Borel measure on the real interval $(-\pi, \pi]$. Suppose $\mu = \mu_A + \mu_S$, with $d\mu_A(\phi) = f(\phi)d\sigma(\phi)$, is the Lebesgue decomposition of μ into absolutely continuous and singular measures. In this note we shall be concerned with two generalizations of the problem Q_0: Find

$$I_0(\mu) = \inf_{P \in \mathscr{P}_0}\left[\int |1 + e^{i\phi}P(e^{i\phi})|^2 \, d\mu(\phi)\right]^{\frac{1}{2}} .$$

Q_0 was solved by Szegö for the case $\mu = \mu_A$ and in general by M. G. Krein and Kolmogorov. They showed that $I_0(\mu) = \exp \frac{1}{2}\int \log f \, d\sigma$ if $\log f$ is integrable and $I_0(\mu) = 0$ otherwise. (See [3], pp. 44, 231.)

We shall consider:

Problem Q_1: Suppose $\int |g|^2 d\mu < \infty$. Find

$$I_1(g, \mu) = \inf_{P \in \mathscr{P}_0}\left[\int |g + P|^2 \, d\mu\right]^{\frac{1}{2}} ,$$

and

Problem Q_2: Suppose $\int |h| \, d\sigma < \infty$. Find

$$I_2(h, \mu) = \sup_{P \in \mathscr{P}_0}\left\{\left|\int Ph \, d\sigma\right| \Big/ \left[\int |P|^2 \, d\mu\right]^{\frac{1}{2}}\right\} .$$

Clearly $I_1(e^{-i\phi}, \mu) = I_0(\mu)$. Also

$$[I_2(1, \mu)]^{-1} = \inf_{P \in \mathscr{P}_0}\left\{\left[\int |P|^2 \, d\mu\right]^{\frac{1}{2}} \Big/ \left|\int P d\sigma\right|\right\} = I_0(\mu) ,$$

so Q_0 is a particularization of both Q_1 and Q_2. There are other special cases of Q_1 and Q_2 that can be found in the work of Szegö [5] and Grenander and Szegö [3]. Of particular interest are the following:

(i) Let $g(\phi) = e^{-i(k+1)\phi}$, where k is a positive integer. Then Q_1 is the problem of linear prediction k units ahead of time ([3], p. 184).

(ii) Let $h(\phi) = 1/(1 - \alpha e^{-i\phi})$, $|\alpha| < 1$. Then

$$I_2(h, \mu) = \sup_{P \in \mathscr{P}_0}\left\{|P(\alpha)| \Big/ \left[\int |P|^2 \, d\mu\right]^{\frac{1}{2}}\right\} .$$

Received October 20, 1959, in revised form February 2, 1960. This work was done while both authors held National Science Foundation postdoctoral fellowships.

See [3], p. 48.

Throughout we shall indulge in the following notational conveniences: We shall write $I_1(g, f)$ and $I_2(h, f)$ for $I_1(g, \mu_A)$ and $I_2(h, \mu_A)$ respectively, and, in certain contexts, consider two functions identical that are equal everywhere except for a set of Lebesgue measure zero.

We have divided this note into six sections. First we indicate an interesting duality between $I_1(e^{-i\phi}g(\phi), f)$ and $I_2(g, 1/f)$ that relates the problems Q_1 and Q_2 under certain restrictive hypotheses. In section three we fashion the theory that will handle Q_1 and Q_2. This is the solution of a Riemann-Hilbert problem (which we call problem Q_3), which is applied in §§ 4, 5 and 6 to Q_1 and Q_2.

2. Duality of I_1 and I_2. This will fall out of the following Banach space lemma:

Let $\mathscr{P}_0$ be a subspace of a Banach space $\mathscr{L}$ and let $\mathscr{P}_0^{\perp}$ be the annihilator of $\mathscr{P}_0$ in the dual space $\mathscr{L}^*$. If $g \in \mathscr{L}$, then

$$\inf \{\| g + P \| : P \in \mathscr{P}_0\} = \sup \{| l(g)| : l \in \mathscr{P}_0^{\perp}, \| l \| \leqq 1\} .$$

For a proof see Bonsall [2].

THEOREM 1. *Suppose f and $1/f$ are in $L^1(-\pi, \pi)$ and $\int | g |^2 f \, d\sigma < \infty$. Then*

$$I_1(e^{-i\phi}g(\phi), f) = I_2(g, 1/f) .$$

Sketch of proof. By the above lemma

$$I_1(e^{-i\phi}g(\phi), f) = \sup \left\{\left|\int e^{-i\phi}g(\phi)h(\phi)f(\phi)d\sigma\right| \Big/ \left[\int | h |^2 f d\sigma\right]^{\frac{1}{2}}\right\} ,$$

where the supremum is taken over all h such that $\int e^{in\phi}h(\phi)f(\phi)d\sigma = 0$ for $n = 0, 1, 2, \cdots$. Through the substitution $e^{-i\phi}hf = P$ if follows that

$$I_1(e^{-i\phi}g(\phi), f) = \sup \left\{\left|\int Pf \, d\sigma\right| \Big/ \left[\int | P |^2 \frac{1}{f} d\sigma\right]^{\frac{1}{2}}\right\} ,$$

where now the supremum is taken over all P such that $\int e^{in\phi}P(\phi)d\sigma = 0$ for $n = 1, 2, \cdots$. It can be shown that it is sufficient merely to consider suprema for $P \in \mathscr{P}_0$, which proves the theorem.

The restrictive condition $1/f \in L^1(-\pi, \pi)$ seems essential to the formulation of the preceding duality relation, but at least this relation indicates that there exist close tie-ins between Q_1 and Q_2. We shall solve a Riemann-Hilbert problem for the unit circle that, when applied to Q_1 and Q_2, solves both.

3. The Riemann-Hilbert problem Q_3. Let f be a non-negative function in $L^1 = L^1(-\pi, \pi)$, and suppose that $\mathscr{P}$ is the closure of $\mathscr{P}_0$ in the Hilbert space $L^2(f)$ of functions square integrable with respect to the measure $fd\sigma$. Thus, for example, $\mathscr{P}$ in $L^2(1) = L^2$ can be identified with the Hardy space H^2. The problem Q_3 is:

Given $k \in L^1$, find functions $P \in \mathscr{P}$ and q satisfying

$$Pf = k + q \,, \quad \text{and} \tag{1}$$

$$\int qe^{-in\phi}\, d\sigma = 0 \,, \qquad n = 0, 1, \cdots . \tag{2}$$

$\left(\text{Note that since } \int |P|^2 f\, d\sigma < \infty, \text{ we have } Pf \in L^1 \text{ and so } q = Pf - k \in L^1.\right)$

We first list some prefactory material. We associate with any non-negative $f \in L^1$ such that $\log f \in L^1$ the analytic functions

$$\begin{aligned} F^+(z) &= \exp \frac{1}{2}\int \frac{e^{i\phi}+z}{e^{i\phi}-z}\log f(\phi)\, d\sigma(\phi),\ |z|<1 \,, \\ F^-(z) &= \exp \frac{1}{2}\int \frac{z+e^{i\phi}}{z-e^{i\phi}}\log f(\phi)\, d\sigma(\phi),\ |z|>1 \,. \end{aligned} \tag{3}$$

F^+ and F^- belong to H^2 and K^2 respectively, and $\overline{F^-(z)} = F^+(1/\bar{z})$ if $|z| > 1$. (A function $F(z)$ is said to belong to K^p if $F(1/z)$ belongs to H^p.) Since the boundary functions in H^2 and K^2 exist in mean square, we can define

$$\begin{aligned} f^+(\phi) &= \lim_{r\to 1-} F^+(re^{i\phi}) \,, \\ f^-(\phi) &= \lim_{r\to 1+} F^-(re^{i\phi}) \,. \end{aligned} \tag{4}$$

These functions satisfy

$$f(\phi) = f^-(\phi)f^+(\phi) = |f^+(\phi)|^2 = |f^-(\phi)|^2 \,. \tag{5}$$

For any non-negative $f \in L^1$ and $\varepsilon > 0$ we define $F_\varepsilon^\pm(z)$, $f_\varepsilon^\pm(\phi)$ by (3) and (4) with f replaced by $f_\varepsilon = f + \varepsilon$. Here we need not assume that $\log f \in L^1$. Note that since $f + \varepsilon \geqq \varepsilon > 0$, we have $1/F_\varepsilon^+ \in H^\infty$ and $1/F_\varepsilon^- \in K^\infty$. Moreover $|f_\varepsilon^+(\phi)|^2 = f(\phi) + \varepsilon$, so $|f_\varepsilon^-(\phi)| = |f_\varepsilon^+(\phi)| \geqq [f(\phi)]^{1/2}$.

Next we define an operator $(\ \ )_+$ as follows. Its domain D consists of all L^1 functions k with Fourier series $\sum_{-\infty}^{\infty} c_n e^{in\phi}$ such that $\sum_0^\infty |c_n|^2 < \infty$, and k_+ is the function with Fourier series $\sum_0^\infty c_n e^{in\phi}$. We define the operator $(\ \ )_-$ by $k_- = k - k_+$. Notice that $k_+ \in H^2$ and $k_- \in K^1$ with $\int k_- d\sigma = 0$.

Our discussion of Q_3 proceeds in the following order. First we prove uniqueness. Then we solve Q_3 in certain special cases (these being sufficient, it will turn out, to handle Q_1), and finally find the solution in the general case.

We are indebted to the referee for the proof of the next lemma.

LEMMA 2. *Q_3 has at most one solution.*

Proof. Suppose $Pf = q$ where $P \in \mathscr{P}$ and q satisfies (2). Then P is orthogonal, in $L^2(f)$, to all exponentials $e^{in\phi}$ $(n \geqq 0)$. Since P belongs to the closed manifold $\mathscr{P}$ spanned by these exponentials we conclude $P = 0$.

One can formally solve Q_3 by means of the usual factorization methods (see [4], for example). Write $f = f^+f^-$, so $Pf = k + q$ implies

$$Pf^+ = \frac{k}{f^-} + \frac{q}{f^-}\ .$$

Applying $(\ \ )_+$ to both sides we obtain $Pf^+ = (k/f^-)_+$, $P = (1/f^+)(k/f^-)_+$. The following theorem justifies this procedure in certain cases.

THEOREM 3. (i) *Suppose* $\log f \in L^1$ *and* $k/f^- \in D$. *Then* Q_3 *has the solution*

$$(6) \qquad P = \frac{1}{f^+}\left(\frac{k}{f^-}\right)_+ \qquad q = -f^-\left(\frac{k}{f^-}\right)_-\ .$$

(ii) *Suppose* $\log f \notin L^1$ *and* $k^2/f \in L^1$. *Then* Q_3 *has the solution*

$$P = \frac{k}{f} \qquad q = 0\ .$$

Proof. (i) Let $\varepsilon > 0$. Since the function f^+ is outer, it follows from a theorem of Beurling [1] that there exists a $P_0 \in \mathscr{P}_0$ such that

$$\int\left|\left(\frac{k}{f^-}\right)_+ - P_0f^+\right|^2 d\sigma < \varepsilon\ .$$

Therefore by (5)

$$\int\left|\frac{1}{f^+}\left(\frac{k}{f^-}\right)_+ - P_0\right|^2 f d\sigma < \varepsilon\ ,$$

so P as defined in (6) belongs to $\mathscr{P}$. Furthermore, with q as defined in (6),

$$Pf - q = f^-\left[\left(\frac{k}{f_-}\right)_+ + \left(\frac{k}{f^-}\right)_-\right] = k\ .$$

It remains to show that $q \in K^1$. Certainly q belongs to $K^{1/2}$ since it is the product of the two K^1 functions $-f^-$ and $(k/f^-)_-$. But since also

$q = Pf - k$, it belongs to L^1. Therefore ([6], p. 163) $q \in K^1$.

(ii) If $\log f \notin L^1$, the space $\mathscr{P}$ is identical with $L^2(f)$ ([3], § 33) and so $k/f \in \mathscr{P}$.

We now give the complete solution of Q_3.

THEOREM 4. (i) *The limit*

$$\lim_{\varepsilon \to 0+} \int \left| (k/f_\varepsilon^-)_+ \right|^2 d\sigma$$

exists either finitely or infinitely.

(ii) *A necessary and sufficient condition that Q_3 have a solution P, q is that the limit be finite.*

(iii) *If the limit is finite then*

$$P = \lim (1/f_\varepsilon^+)(k/f_\varepsilon^-)_+$$

in the space $L^2(f)$, and

$$\int | P |^2 f \, d\sigma = \lim_{\varepsilon \to 0+} \int |(k/f_\varepsilon^-)_+ |^2 \, d\sigma \, .$$

Proof. Assume first that Q_3 has a solution P, q and divide both sides of (1) by f_ε^-. Since $q/f_\varepsilon^- \in K^1$ and $\int q/f_\varepsilon^- d\sigma = 0$ we have $q/f_\varepsilon^- \in D$ and $(q/f_\varepsilon^-)_+ = 0$; also $Pf/f_\varepsilon^- \in L^2 \subset D$. Therefore we can apply $(\ \ )_+$ to both sides, obtaining

$$(Pf/f_\varepsilon^-)_+ = (k/f_\varepsilon^-)_+ \, .$$

Consequently

$$\int |(k/f_\varepsilon^-)_+ |^2 \, d\sigma \leqq \int | Pf/f_\varepsilon^- |^2 \, d\sigma \leqq \int | P |^2 f d\sigma \, , \tag{7}$$

and so

$$\limsup_{\varepsilon \to 0+} \int |(k/f_\varepsilon^-)_+ |^2 \, d\sigma < \infty \, . \tag{8}$$

Conversely suppose that $\{\varepsilon_n\}$ is a sequence of ε's such that $\varepsilon_n \to 0+$ and

$$\int |(k/f_\varepsilon^-)_+ |^2 \, d\sigma = O(1) \text{ for } \varepsilon = \varepsilon_n \, . \tag{9}$$

By Theorem 3(i) there corresponds to each $\varepsilon = \varepsilon_n$ a solution P_ε, q_ε of $(f + \varepsilon)P_\varepsilon = k + q_\varepsilon$. We have

$$\int | P_\varepsilon |^2 f d\sigma \leqq \int | P_\varepsilon |^2 f_\varepsilon d\sigma = \int |(k/f_\varepsilon^-)_+ |^2 \, d\sigma = O(1) \, . \tag{10}$$

Thus there exists a subsequence of $\{\varepsilon_n\}$ such that $\{P_\varepsilon\}$ converges weakly in $L^2(f)$ to an element $P \in \mathscr{P}$. It will follow that P, $Pf - k$ satisfies Q_3 if the L^1 function $q = Pf - k$ satisfies (2). We have for $n = 0, 1, 2, \cdots$

$$\begin{aligned}\int q(\phi)e^{-in\phi}d\sigma &= \int \{P_\varepsilon(\phi)[f(\phi) + \varepsilon] - k(\phi)\}e^{-in\phi}d\sigma \\ &\quad + \int [P(\phi) - P_\varepsilon(\phi)]f(\phi)e^{-in\phi}d\sigma - \varepsilon \int P_\varepsilon(\phi)e^{-in\phi}d\sigma \\ &= J_1 + J_2 + J_3 .\end{aligned}$$

Theorem 3(i) implies that $J_1 = 0$. By the weak convergence of the P_ε we can make J_2 as small as desired by taking ε_n sufficiently small. Finally (10) implies that $\int |\varepsilon^{1/2}P_\varepsilon|^2 d\sigma = O(1)$, so by the Schwarz inequality $|J_3| \leqq \varepsilon^{1/2} \int |\varepsilon^{1/2}P_\varepsilon| d\sigma = O(\varepsilon^{1/2})$ as $\varepsilon_n \to 0$. Thus P, q satisfy Q_3, so (8), holds and (9) is true for *any* sequence $\{\varepsilon_m\}$ of ε's that converge to $0+$. By what we have shown there corresponds to any such sequence $\{\varepsilon_m\}$ a subsequence such that P_ε converges weakly to the unique (Lemma 2) element P. Thus we can consider ε to be a real variable and conclude that P_ε converges weakly in $L^2(f)$ to $P \in \mathscr{P}$ as $\varepsilon \to 0+$ provided that

$$\liminf_{\varepsilon \to 0+} \int |k/f_\varepsilon^-)_+|^2 d\sigma < \infty .$$

We next prove that in fact P_ε converges strongly to P in $L^2(f)$. It suffices to show that $\int |P_\varepsilon|^2 fd\sigma \to \int |P|^2 fd\sigma$. Weak convergence gives

$$\liminf_{\varepsilon \to 0+} \int |P_\varepsilon|^2 fd\sigma \geqq \int |P|^2 fd\sigma .$$

On the other hand, as in (7),

$$\int |P_\varepsilon|^2 fd\sigma \leqq \int |P_\varepsilon|^2 f_\varepsilon d\sigma = \int |(k/f_\varepsilon^-)_+|^2 d\sigma \leqq \int |P|^2 fd\sigma .$$

so

$$\limsup_{\varepsilon \to 0+} \int |P_\varepsilon|^2 fd\sigma \leqq \int |P|^2 fd\sigma .$$

Thus

$$\lim_{\varepsilon \to 0+} \int |P_\varepsilon|^2 fd\sigma$$

exists, and equals

$$\lim_{\varepsilon \to 0+} \int |(k/f_\varepsilon^-)_+|^2 d\sigma = \int |P|^2 fd\sigma .$$

Thus the proof is complete.

4. Solution of Q_1. In Q_1 we wish to find

$$I_1(g, \mu) = \inf_{P \in \mathscr{P}_0} \left[\int | g + P |^2 d\mu \right]^{\frac{1}{2}} ,$$

where g is a given function in $L^2(\mu)$. Since $I_1(g, \mu)$ represents the distance from g to the manifold $\mathscr{P}_0$ in $L^2(\mu)$, there exists a (unique) function P belonging to the closure $\mathscr{P}'$ of $\mathscr{P}_0$ in $L^2(\mu)$ such that

$$I_1(g, \mu) = \left[\int | g + P |^2 d\mu \right]^{\frac{1}{2}} .$$

This function P is such that $g + P$ is orthogonal to $\mathscr{P}_0$, so

$$\int [g(\phi) + P(\phi)] e^{-in\phi} d\mu(\phi) = 0 \qquad n = 0, 1, 2, \cdots .$$

It follows from a theorem of the brothers Riesz ([6], p. 158) that the measure ν given by

$$\nu(E) = \int_E [g(\phi) + P(\phi)] d\mu(\phi)$$

is absolutely continuous with respect to Lebesgue measure. Let F be a Borel set of Lebesgue measure zero such that $\mu_S((-\pi, \pi] - F) = 0$. Then $g + P$ vanishes on F almost everywhere with respect to μ_S, so

$$\int_F | g + P |^2 d\mu_S = 0$$

and

$$\int | g + P |^2 d\mu = \int_{\mathscr{C}} | g + P |^2 d\mu_A = \int | g + P |^2 f d\sigma .$$

Since $\mu \geqq \mu_A$ it follows that $I_1(g, \mu) = I_1(g, f)$, and this common value is attained by the same extremizing function $P \in \mathscr{P}' \subset \mathscr{P}$.

Now,

$$\int [g(\phi) + P(\phi)] e^{-in\phi} f(\phi) d\sigma = 0 \qquad n = 0, 1, \cdots ,$$

so if we set $q = (g + P)f$ we have $Pf = -gf + q$, where $P \in \mathscr{P}$ and q satisfies (2). Since $(gf)^2/f = g^2 f \in L^1$, we can apply Theorem 3 to this situation. The extremizing function

$$P = \begin{cases} -(1/f_+)(gf_+)_+ & \text{if} \quad \log f \in L^1 \\ -g & \text{if} \quad \log f \notin L^1 , \end{cases}$$

and since

$$I_1(g,f) = \left[\int |g+P|^2 f d\sigma\right]^{\frac{1}{2}} = \left[\int |q|^2/f d\sigma\right]^{\frac{1}{2}}$$

we have

$$I_1(g,\mu) = I_1(g,f) = \begin{cases} \left[\int |(gf^+)_-|^2 d\sigma\right]^{\frac{1}{2}} & \text{if } \log f \in L^1 \\ 0 & \text{if } \log f \notin L^1 . \end{cases}$$

5. Solution of Q_2. Given $h \in L^1$, we will evaluate

$$I_2(h,\mu) = \sup_{P \in \mathscr{P}_0} \left\{ \left| \int Ph\, d\sigma \right| \Big/ \left[\int |P|^2 d\mu\right]^{\frac{1}{2}} \right\} .$$

Since $\mu \geqq \mu_A$ it is clear that if $I_2(h,f)$ is finite so is $I_2(h,\mu)$. We shall show that, conversely, if $I_2(h,\mu)$ is finite then so is $I_2(h,f)$ and in fact $I_2(h,f) = I_2(h,\mu)$. So now suppose $I_2(h,\mu) < \infty$. Then the linear functional L on $\mathscr{P}_0$ given by

$$L(P) = \int Ph\, d\sigma$$

is bounded on $L^2(\mu)$. Therefore if $\mathscr{P}'$ denotes the closure of $\mathscr{P}_0$ in $L^2(\mu)$, there is a uniquely determined $Q \in \mathscr{P}'$ such that $L(P) = \int P\bar{Q}\, d\mu$. Then we have

$$\int e^{-in\phi}[Q(\phi)d\mu(\phi) - \bar{h}(\phi)d\sigma(\phi)] = 0 \qquad n = 0, 1, \cdots .$$

We again apply the F. and M. Riesz theorem, and deduce that the measure ν given by

$$\nu(E) = \int_E Q d\mu - \int_E h\, d\sigma$$

is absolutely continuous with respect to Lebesgue measure. Letting F be a Borel set of Lebesgue measure zero such that $\mu_S((-\pi,\pi] - F) = 0$, we see that Q vanishes on F almost everywhere with respect to μ_S. Consequently

$$\int e^{-in\phi}[Q(\phi)f(\phi) - \bar{h}(\phi)]d\sigma(\phi) = 0 \qquad n = 0, 1, \cdots ,$$

so $Qf = \bar{h} + q$, where $Q \in \mathscr{P}' \subset \mathscr{P}$ and q satisfies (2). Thus the linear functional

$$L(P) = \int Ph\, d\sigma = \int P\bar{Q}f\, d\sigma ,$$

$P \in \mathscr{P}_0$, is bounded on $L^2(f)$, so $I_2(h, f)$ is finite and in fact equals $I_2(h, \mu)$. We deduce from Theorem 4 that

$$I_2(h, \mu) = I_2(h, f) = \lim_{\varepsilon \to 0+} \left[\int | (\bar{h}/f_\varepsilon^-)_+ |^2 \, d\sigma \right]^{\frac{1}{2}} ,$$

and Q may be exhibited as an $L^2(f)$ limit in the mean.

6. Some formulae for $I_2(h, \mu)$. We can obtain a simpler formula for $I_2(h, \mu)$ if we assume that $h^2/f \in L^1$ and apply Theorem 3. Then

$$I_2(h, \mu) = \begin{cases} \left[\int | (\bar{h}/f^-)_+ |^2 \, d\sigma \right]^{\frac{1}{2}} = \left[\int |(e^{-i\phi} h(\phi)/f^+(\phi))_- |^2 \, d\sigma(\phi) \right]^{\frac{1}{2}} & \\ & \text{if } \log f \in L^1 , \\ \left[\int | h |^2/f d\sigma \right]^{\frac{1}{2}} & \text{if } \log f \notin L^1 . \end{cases}$$

This, in conjunction with our solution of Q_1, gives the duality discussed in Theorem 1. Note that the hypothesis $1/f \in L^1$ of Theorem 1 implies that $\log f \in L^1$.

Another simple formula for $I_2(h, \mu)$ is available if we know that the Fourier series $\sum_{-\infty}^{\infty} h_n e^{in\phi}$ of h is such that $h_{-n} = O(R_0^{-n})$ as $n \to +\infty$ for some $R_0 > 1$. Then the function $H(z) = \sum_0^\infty h_{-n} z^{-n}$ is analytic in $| z | > 1/R_0$. We have

$$\int | (\bar{h}/f_\varepsilon^-)_+ |^2 \, d\sigma = \int | (e^{-i\phi} h(\phi)/f_\varepsilon^+(\phi))_- |^2 \, d\sigma ,$$

which by the Parseval relation equals

$$\sum_{n=0}^{\infty} \left| \int e^{in\phi} h(\phi) f_\varepsilon^+(\phi) d\sigma \right|^2 = \sum_{n=0}^{\infty} \left| \frac{1}{2\pi} \int_{|z|=1} z^{n+1} H(z) \Big/ F_\varepsilon^+(z) dz \right|^2$$

$$= \sum_{n=0}^{\infty} \left| \frac{1}{2\pi} \int_{|z|=R} z^{n+1} H(z) \Big/ F_\varepsilon^+(z) dz \right|^2 ,$$

where $1/R_0 < R < 1$. Let us also assume that $\log f \in L^1$, so F^+ is well-defined and

$$H(Re^{i\phi})/F_\varepsilon^+(Re^{i\phi}) \longrightarrow H(Re^{i\phi})/F^+(Re^{i\phi})$$

in L^2 as $\varepsilon \to 0+$. It follows that

$$I_2(h, \mu)^2 = \sum_{n=0}^{\infty} \left| \frac{1}{2\pi} \int_{|z|=R} z^{n+1} H(z) \Big/ F^+(z) dz \right|^2 .$$

Now, if we write

$$\frac{1}{F^+(z)} = \sum_{n=0}^{\infty} f_n z^n ,$$

then

$$I_2(h, \mu)^2 = \sum_{n=0}^{\infty} \left| \sum_{m=0}^{\infty} h_{-n-m} f_m \right|^2 .$$

Thus if H is the Hankel matrix $[h_{-n-m}]_{n,m=0}^{\infty}$, and Φ the column vector with components $f_0, f_1, \cdots$, then

$$I_2(h, \mu) = \| H\Phi \| ,$$

where the norm is that of l^2.

For example, let α be such that $|\alpha| < 1$ and consider

$$\sup_{P \in \mathscr{P}} \left\{ | P(\alpha) | \Big/ \left(\int | P |^2 d\mu \right)^{\frac{1}{2}} \right\} .$$

Thus we wish to evaluate $I_2(1/(1 - \alpha e^{-i\phi}), \mu)$. Here $h_{-n} = \alpha^n$, $n = 0, 1, \cdots$, so

$$I_2(h, \mu)^2 = \sum_{n=0}^{\infty} \left| \sum_{m=0}^{\infty} \alpha^{n+m} f_m \right|^2 = 1/[(1 - |\alpha|^2)| F^+(\alpha)|^2] ,$$

as in [2], p. 48.

Bibliography

1. A. Beurling, *On two problems concerning linear transformations in Hilbert space*, Acta Math., **81** (1948), 239-255.
2. F. F. Bonsall, *Dual extremum problems in the theory of functions*, Jour. London Math. Soc., **31** (1956), 1-5-110.
3. U. Grenander and G. Szegö, *Toeplitz forms and their applications*, Berkeley and Los Angeles, 1958.
4. N. I. Muskhelishvili, *Singular integral equations*, Groningen, 1953.
5. Szegö, *Orthogonal polynomials*, A. M. S. colloquium publication, **23** (1939).
6. A. Zygmund, *Trigonometrical series*, New York, 1952.

INSTITUTE FOR ADVANCED STUDY
UNIVERSITY OF VIRGINIA
CORNELL UNIVERSITY

A CLASS OF LINEAR DIFFERENTIAL-DIFFERENCE EQUATIONS

MORTON SLATER* AND HERBERT S. WILF**

I. Introduction. The purpose of this paper is to study the following integral equation:

$$\varphi(x) = \int_x^{x+1} K(y)\varphi(y)dy \tag{1}$$

or the differential-difference equation

$$\varphi'(x) = K(x+1)\varphi(x+1) - K(x)\varphi(x) \tag{1'}$$

with the boundary condition

$$\lim_{x\to\infty} \varphi(x) = 1 . \tag{2}$$

Equations of the type (1), (1') have been investigated in great generality by many authors. In particular, the interested reader is referred to Yates [6], and Cooke [2], for recent developments, and a bibliography of significant earlier work. The equations of the form (1) which we shall consider are related to the class of linear differential-difference equations with asymptotically constant coefficients, a class treated thoroughly by Wright [5], and Bellman [1].

The novelty of the results below arises from the boundary condition (2) which appears not to have been studied before, and which gives reresults of an essentially different character from those of the works cited above. The system (1), (2) is of interest in some problems connected with the theory of neutron slowing down (Placzek [3]).

A further departure from previous work is the fact that no use is made of complex variable methods or the asymptotic characteristic equation of the kernel $K(y)$.

Aside from some fairly obvious theorems concerning uniqueness, boundedness and positivity, our main results are the following:

(a) necessary and sufficient conditions for the existence of a solution of (1), (2); this is achieved by constructing a minorant for the solution.

(b) proof of the existence of $\varphi(-\infty)$ under fairly general conditions.

(c) an application of Fubini's theorem to exhibit a rather surpris-

Received February 4, 1960.

* Nuclear Development Corporation of America, White Plains, N. Y.

** The above research was done while the second author was employed by the Nuclear Development Corporation of America, White Plains, N. Y. At present his address is the University of Illinois, Urbana, Illinois.

ing relation between an integral of the solution over the real axis and its limits at $\pm\infty$. We assume

H 1 $K(x)$ is measurable ,

H 2 $0 < K(x) \leqq 1$, for almost all x ,

H 3 For $x \geqq M$, $K(x)$ increases ,

H 4 $\lim_{x\to\infty} K(x) = 1$,

throughout the paper.

To summarize the results below, we shall give necessary and sufficient conditions for the existence (Theorem 4), uniqueness (Theorem 1), boundedness (Theorem 2), and positivity (Theorem 3) of the solution; a a sufficient condition for its monotonicity (Theorem 5); a proof of the existence of $\varphi(-\infty)$(Theorem 6) and the evaluation of a definite integral involving the solution (Theorem 7).

By "solution" we shall always mean a function $\varphi(x)$ satisfying both (1) and (2). All integrals are to be understood in the sense of Lebesgue.

II. Existence and uniqueness of solutions.

THEOREM 1. *Under* H 1 – H 4, *the solution* $\varphi(x)$, *when it exists, is unique.*

Proof. If the theorem is false, there exists a function $\psi(x)$ not identically zero which satisfies (1) and for which

$$\lim_{x\to\infty} \psi(x) = 0 .$$

Then by the continuity of $\psi(x)$ there exist numbers η and x_0 such that $\eta > 0$, $|\psi(x_0)| = \eta$ and for all $x > x_0$, $|\psi(x)| < \eta$. But then

$$\eta = |\psi(x_0)| \leqq \int_{x_0}^{x_0+1} |\psi(y)|\, dy < \eta$$

a contradiction, which completes the proof.

THEOREM 2. *With* H1 – H4 *we have, for any solution* $\varphi(x)$ *of* (1), (2),

$$|\varphi(x)| \leqq 1 \qquad (-\infty < x < \infty) . \tag{3}$$

Proof. For if $|\varphi(x)| > 1$ for some x, then by (2) and the continuity of $|\varphi(x)|$ there is a $C > 1$ and an x_0 such that $|\varphi(x_0)| = C$, and for all $x > x_0$, $|\varphi(x_0)| < C$. But then

$$|\varphi(x_0)| \leqq \left| \int_{x_0}^{x_0+1} |\varphi(y)|\, dy\right.$$

implies $C < C$, which is a contradiction.

THEOREM 3. *Supposing* H1 – H4, *the solution* $\varphi(x)$ *of* (1) *and* (2), *when it exists, is positive for all* x, *and is non-decreasing for* $x \geqq M$.

Proof. We prove positivity first. If $\varphi(x)$ is not >0 for all x, then by (2) and the continuity of $\varphi(x)$ there is an x_0 such that $\varphi(x_0) = 0$ and for all $x > x_0$, $\varphi(x) > 0$. Then

$$\varphi(x_0) = 0 = \int_{x_0}^{x_0+1} K(y)\varphi(y)dy ,$$

which is a contradiction by H2.

To prove the monotonicity part, we define

$$\psi_0(x) = 1 , \tag{4}$$

and

$$\psi_{n+1}(x) = \int_x^{x+1} K(y)\psi_n(y)dy . \tag{5}$$

Since $0 < K(y) \leqq 1$, $\psi_1(x) \leqq \psi_0(x)$, and since

$$\psi_n(x) - \psi_{n+1}(x) = \int_x^{x+1} K(y)[\psi_{n-1}(y) - \psi_n(y)]dy , \tag{6}$$

we see by induction that $\{\psi_n(x)\}$ is a decreasing sequence. But since $\varphi(x) \leqq 1 = \psi_0(x)$, we see by a second induction that $\psi_n(x) \geqq \varphi(x)$ for all x. Hence the $\psi_n(x)$ decrease to a limit function $\psi(x)$ satisfying (1) by Lebesgue's dominated convergence theorem, and

$$\lim_{x\to\infty} \psi(x) = 1$$

since $\varphi(x) \leqq \psi(x) \leqq 1$. Now $\psi_0(x)$ is non-decreasing for $x \geqq M$, and thus so is $\psi_1(x)$, and again by induction, $\psi_n(x)$ and hence $\psi(x)$. But by Theorem 1, $\psi(x) = \varphi(x)$, which proves the theorem.

LEMMA 1. *Under* H1 – H4 *and*

$$\text{H5:} \quad 1 - K(x) \in \mathscr{L}(M, \infty)$$

there is a function $S(x)$ *such that* $S(x) \geqq 0, S(x)$ *is non-decreasing*, $\lim_{x\to\infty} S(x) = 1$, *and*

$$S(x) \leqq \int_x^{x+1} K(y)S(y)dy \qquad (-\infty < x < \infty) . \tag{7}$$

Proof. Define

$$S(x) = \begin{cases} 0 & x \leqq M \\ C_n & M + \dfrac{n}{2} \leqq x < M + \dfrac{n+1}{2} \end{cases} \qquad (n = 0, 1, 2, \cdots) \tag{8}$$

where the C_n are constants to be determined, and define

$$q_n = \int_{M+(n/2)}^{M+(n+1/2)} K(y)dy \; . \tag{9}$$

Now, requiring that $S(x)$ satisfy (1) at the points $M + (n/2)$ gives

$$C_n q_n + C_{n+1} q_{n+1} = C_n$$

that is

$$C_{n+1} = C_n \left[\frac{1 - q_n}{q_{n+1}} \right] ,$$

and

$$C_{n+1} = \prod_{j=0}^{n} \left[\frac{1 - q_j}{q_{j+1}} \right] C_0 \; . \tag{10}$$

But since

$$\frac{1 - q_j}{q_{j+1}} - 1 = \frac{1 - (q_j + q_{j+1})}{q_{j+1}} \geqq 0$$

we see that the C_n form a non-decreasing sequence. Also

$$\frac{1 - q_j}{q_{j+1}} - 1 \leqq \frac{1 - K\{M + (n/2)\}}{K(M)}$$

since $K(y)$ increases. But then H5 implies that

$$\sum_{n=0}^{\infty} \{1 - K[M + (n/2)]\}$$

converges, and so the limit of the product in (10) exists. We can then choose C_0 so that

$$\lim_{n \to \infty} C_n = 1 \; .$$

It remains to show that (7) is everywhere satisfied. If $x_0 > M$ and $x_0 \neq M + (n/2)$ for any n, let $M + (n_0/2)$ be the largest of the $M + (n/2)$ which is less than x_0. Then

$$\begin{aligned} &\int_{x_0}^{x_0+1} K(y) S(y) dy \\ &\quad = \int_{M+(n_0/2)}^{M+(n_0+2/2)} K\left(y + x_0 - M - \frac{n_0}{2} \right) S\left(y + x_0 - M - \frac{n_0}{2} \right) dy \\ &\quad \geqq \int_{M+(n_0/2)}^{M+(n_0+2/2)} K(y) S(y) dy \\ &\quad = C_{n_0} \\ &\quad = S(x_0) \, , \end{aligned}$$

since K and S are positive and non-decreasing.

We can now prove

THEOREM 4. *Let* H1 – H4 *hold. Then, necessary and sufficient for the existence of a solution of* (1), (2) *is* H5.

Proof. Suppose $\varphi(x)$ exists, then

$$\begin{aligned}\varphi(x) &= \int_x^{x+1} K(y)\varphi(y)dy \\ &= \int_x^{x+1} \varphi(y)dy - \int_x^{x+1} [1 - K(y)]\varphi(y)dy .\end{aligned}$$

Choose ε between 0 and 1 and $x_0 > M$ such that $\varphi(x) > 1 - \varepsilon$ for $x \geqq x_0$. Then

$$(1-\varepsilon)\int_{x_0}^{x_0+1} [1 - K(y)]dy \leqq \varphi(x_0 + 1) - \varphi(x_0)$$

since $\varphi(x)$ is non-decreasing (Theorem 3) for $x \geqq M$. Replacing x_0 by $x_0 + 1$, etc., and adding

$$\int_{x_0}^{\infty} [1 - K(y)]dy \leqq 1 - \varphi(x_0) < \infty .$$

On the other hand, if H5 holds, consider again the $\psi_n(x)$ of (4)–(5). Since $\{\psi_n(x)\}$ is a decreasing sequence, and

$$\psi_{n+1}(x) - S(x) \geqq \int_x^{x+1} K(y)[\psi_n(y) - S(y)]dy$$

we see that $\psi_n(x) \geqq S(x)$ for all n and x. Hence $\psi_n(x)$ decreases to a a limit $\varphi(x)$, satisfying (1), and since

$$1 \geqq \varphi(x) \geqq S(x)$$

we have (2) also.

III. Monotonicity. The solution $\varphi(x)$ of (1), (2), when it exists, need not to be monotone on the whole real axis. In this section we will first illustrate the above statement, and then give sufficient conditions for the monotonicity of the solution. A lemma that will be of use in the illustration is

LEMMA 2. *Let* $K_a(x)$ *and* $K_b(x)$ *each satisfy* H1–H5, *and in additon suppose that for all* x

$$K_a(x) \leqq K_b(x) .$$

Then if $\varphi_a(x), \varphi_b(x)$ are the corresponding solutions of (1), (2), we have

$$\varphi_a(x) \leqq \varphi_b(x)$$

for all x.

Proof. First,

$$\begin{aligned}\varphi_a(x) &= \int_x^{x+1} K_a(y)\varphi_a(y)dy \\ &\leqq \int_x^{x+1} K_b(y)\varphi_a(y)dy\ .\end{aligned}$$

Now let $\varphi_{a,0}(x) = \varphi_a(x)$, and define

$$\varphi_{a,n+1}(x) = \int_x^{x+1} K_b(y)\varphi_{a,n}(y)dy\ .$$

Then $\{\varphi_{a,n}(x)\}\uparrow_n$ and is bounded above by 1. Hence the sequence converges to a solution of

$$\begin{cases}\varphi(x) = \int_x^{x+1} K_b(y)\varphi(y)dy \\ \lim_{x\to\infty}\varphi(x) = 1\ .\end{cases}$$

The result then follows from Theorem 1 .

Now consider the family

$$K_a(x) = \frac{x^2+a}{x^2+1} \qquad (0 \leqq a \leqq 1)\ .$$

Clearly each $K_a(x)$ satisfies H1–H5. Let $\varphi_0(x)$ satisfy (1), (2) with $K(x) = K_0(x)$. Then

$$\varphi_0'(-1) = -K_0(-1)\varphi_0(-1) = -(1/2)\varphi_0(-1) < 0$$

by Theorem 3. Hence $\varphi_0(x)$ is not monotone. In fact we can invoke Lemma 2 to show that there exists a number $a^*\varepsilon(0,1)$ such that for $a < a^*$ $\varphi_a(x)$ is not monotone. For if not, there exists a sequence $\{a_n\}\downarrow 0$ such that $\varphi_{a_n}(x)$ satisfies (1), (2) with $K(x) = K_{a_n}(x)$ and $\varphi_{a_n}(x)$ is monotone for each n. Since $\{\varphi_{a_n}(x)\}$ decreases to a solution of (1), (2) with $K(x) = K_0(x)$ (by Lemma 2 and Theorem 1) we must have $\varphi_0(x)$ monotone which is a contradiction.

The following theorem, however, gives a sufficient condition for the monotonicity of $\varphi(x)$:

THEOREM 5. *With* H1–H5, *suppose that for almost all* x,

$$K(x+1) \geqq K(x) \int_x^{x+1} K(y)dy \ . \tag{11}$$

Then $\varphi(x)$ is non-decreasing on the real axis.

Proof. Let $S_0(x)$ be the function $S(x)$ of (8). Define

$$S_{n+1}(x) = \int_x^{x+1} K(y) S_n(y) dy \qquad (n = 0, 1, \cdots) \ . \tag{12}$$

Then, for all n,

$$\begin{aligned} &\text{(a)} \quad 0 \leqq S_n(x) \leqq 1 \\ &\text{(b)} \quad \lim_{x \to \infty} S_n(x) = 1 \\ &\text{(c)} \quad S_n(x) \uparrow \varphi(x) \ . \end{aligned} \tag{13}$$

We show next that with (11), the subsequence $\{S_{2n}(x)\}$ is a sequence of non-decreasing functions. Clearly $S_0(x)\uparrow_x$ for all x. Now suppose that for all $k \leqq n$, $S_{2k}(x)\uparrow_x$ for all x. Then

$$S'_{2n+2}(x) = K(x+1)S_{2n+1}(x+1) - K(x)S_{2n+1}(x)$$

a.e.

Now by (13)(c),

$$S_{2n+1}(x+1) \geqq S_{2n}(x+1)$$

and since

$$S_{2n+1}(x) = \int_x^{x+1} K(y) S_{2n}(y) dy \ ,$$

it follows from the inductive hypothesis that

$$S_{2n+1}(x) \leqq S_{2n}(x+1) \int_x^{x+1} K(y) dy \ .$$

Hence

$$\begin{aligned} S'_{2n+2}(x) &\geqq \left[K(x+1) - K(x) \int_x^{x+1} K(y) dy \right] S_{2n}(x+1) \\ &\geqq 0 \quad \text{a.e.} \end{aligned}$$

by (11), which proves the theorem, since $S_{2n+2}(x)$ is absolutely continuous.

IV. Behaviour for large negative values of x. We wish now to explore the limiting behaviour of the solution $\varphi(x)$ as $x \to -\infty$. We have seen that the solution will in general oscillate. We will establish below a sufficient condition for the existence of $\varphi(-\infty)$.

THEOREM 6. *Suppose* $\varphi(x)$ *is a solution of* (1), (2). *Let* $K(x)$ *satisfy* H1–H4, *and further suppose that*

$$\lim_{x\to-\infty}\int_x^{x+1}|K(t+1)-K(t)|\,dt=0\,. \tag{14}$$

Then

$$\lim_{x\to-\infty}\varphi(x)\equiv\varphi(-\infty) \tag{15}$$

exists.

Proof. Let m (resp. M) be the lim inf (resp. lim sup) of $\varphi(x)$ as $x\to-\infty$, and write

$$k=\limsup_{x\to-\infty}\int_x^{x+1}|\varphi'(t)|\,dt\,.$$

Let $\varepsilon>0$ be given. Let $-x_0>0$ be chosen so that $\varphi(x_0)<m+\varepsilon$ and for $x\leqq x_0$, $\int_x^{x+1}|\varphi'(t)|\,dt<k+\varepsilon$. Let x_1 be the first point to the left of x_0 at which $\varphi(x_1)=M-\varepsilon$, so that $\varphi(x)<M-\varepsilon$ on the interval $x_1<x\leqq x_0$. It follows that $x_0<x_1+1$ for otherwise a "proper" maximum for $\varphi(x)$ on $x_1\leqq x\leqq x_1+1$ occurs at x_1, which is impossible. For the same reason there is a point x_2 satisfying $x_1<x_0<x_2\leqq x_1+1$ at which $\varphi(x_2)=M-\varepsilon$. Hence

$$\begin{aligned}k+\varepsilon\geqq\int_{x_1}^{x_1+1}|\varphi'(t)|\,dt&\geqq\int_{x_1}^{x_0}|\varphi'(t)|\,dt+\int_{x_0}^{x_2}|\varphi'(t)|\,dt\\&\geqq\left|\int_{x_1}^{x_0}\varphi'(t)dt\right|+\left|\int_{x_0}^{x_2}\varphi'(t)dt\right|\\&=(M-m-\varepsilon)+(M-m-\varepsilon)\,.\end{aligned}$$

Hence $k\geqq2(M-m)$.
However, since

$$\varphi'(x)=K(x+1)[\varphi(x+1)-\varphi(x)]+\varphi(x)[K(x+1)-K(x)]\,,$$

we find, using (14) $k\leqq M-m$. Thus $M=m$, which proves the theorem, and incidently, $k=0$.

REMARK. $\int_x^{x+1}|K(t+1)-K(t)|\,dt\leqq\int_x^{x+2}|1-K(t)|\,dt$; thus in the above theorem, (14) may be replaced be $1-K(x)\in\mathscr{L}(-\infty,\infty)$, and the conclusion is still valid.

We are now able to prove the following integral relationship.

THEOREM 7. *Suppose* $\varphi(x)$ *is a solution of* (1), (2). *Let* $K(x)$ *satisfy* H1–H4, *and suppose further*

$$1 - K(x) \in \mathscr{L}(-\infty, \infty) . \tag{16}$$

Then

$$\int_{-\infty}^{\infty} [1 - K(y)]\varphi(y)dy = \frac{1 - \varphi(-\infty)}{2} . \tag{17}$$

Proof. Put

$$F(x) = \int_0^1 \varphi(x - y)ydy .$$

Then

$$F'(x) = \int_0^1 \varphi'(x - y)ydy = -\varphi(x - 1) + \int_0^1 \varphi(x - y)dy$$
$$= \int_0^1 \varphi(x - y)[1 - K(x - y)]dy .$$

Since $\varphi(x)$ is bounded and $1 - K(x) \in \mathscr{L}(-\infty, \infty)$, it follows from Fubini's theorem (see reference 4, p. 87) that $F'(x) \in \mathscr{L}(-\infty, \infty)$, and

$$F(\infty) - F(-\infty) = \int_{-\infty}^{\infty} [1 - K(t)]\varphi(t)dt .$$

But since $\varphi(x)$ satisfies (2), $F(\infty) = (1/2)$, and by the remark following Theorem 6, $F(-\infty) = (1/2)\varphi(-\infty)$. This completes the proof.

REFERENCES

1. R. Bellman, *On the existence and boundedness of solutions of non-linear differential-difference equations*, Ann. Math. **50** (1949), 347–355.
2. K. L. Cooke, *The asymptotic behavior of the solutions of linear and nonlinear differential-difference, equations*, Trans. Amer. Math. Soc., **75** (1953), 80–105.
3. G. Plackzek, *On the theory of the slowing down of neutrons in heavy substances*, Physical Review, **69** (1946), 423–438.
4. S. Sáks, *Theory of the Integral*, 1937.
5. E. M. Wright, *The linear difference-differential equation with asymptotically constant coefficients*, Amer. J. Math. **70** (1948), 221–238.
6. B. Yates, *The linear difference-differential equation with linear coefficients*, Trans. Amer. Math. Soc. **80** (1955), 281–298.

THE STRUCTURE OF THREADS

CHARLES R. STOREY

A *thread*, as defined by A. H. Clifford, is a connected topological semigroup in which the topology is the interval topology induced by a total order. A résumé of papers on the subject can be found in the introduction of [1] or in section three of [3].

Briefly, the main classes of threads which have been described are: that of compact threads with an identity and a zero for which the underlying space is a real interval [4]; that of threads defined on the real interval $[0, \infty)$ in which "zero" and "one" play their usual roles [6]; and the class of compact threads with idempotent endpoints, [1] and [2]. Since the separability of the real numbers is not needed for the proofs involved, we will interpret the results of [4] and [6] as applying also to threads in which the underlying space is not real.

The object of this paper is to investigate the structure of more general threads. In the second, third and forth sections we study maximal subgroups, subthreads and the minimal ideal respectively of an arbitrary thread. Theorem 5.5 generalizes the result in [6] by describing all threads S with a zero as an endpoint for which $S^2=S$. In the final section, we are able to describe at least half of any thread satisfying $S^2=S$. More explicitly, if such a thread has no minimal ideal, or if it is itself the minimal ideal, then the entire structure of the thread is determined; while, if there is a proper minimal ideal, then the set of elements larger or the set of elements smaller than the minimal ideal forms a subthread which, satisfying the hypotheses of Theorem 5.5, can be completely described.

It is a pleasure to acknowledge the careful direction by Professor A. H. Clifford of the research leading to this paper.

1. Preliminaries. As defined in [1], a *standard thread* is a compact thread in which the minimal element is a zero and the maximal element an identity. The primary examples are the real interval $[0, 1]$ under the natural order and multiplication and the Rees quotient of $[0, 1]$ by the ideal $[0, \frac{1}{2}]$. The structure of any standard thread can be given as follows [7, Theorem B]: The set of idempotents is closed and thus its complement is a union of disjoint open intervals. If (e, f) is one of these intervals, then $[e, f]$ is a subthread isomorphic with one of the two examples just given. Finally, if e is an idempotent and if $x \leqq e \leqq y$,

Received January 6, 1960. This work was begun while the author received partial support from the Danforth Foundation and also from the National Science Foundation contract with Tulane University, and completed while he held a National Science Foundation Postdoctoral Fellowship.

then $xy = yx = x$.

We say that a thread with a zero and an identity is a *positive thread* if the zero is a least element and if there is no greatest element. The result in [6] is that, in a positive thread, there exists a largest idempotent e less than the identity, $[0, e]$ is a standard thread, $\{t \mid e < t\}$ is isomorphic with the group of positive real numbers, and $xy = yx = x$ whenever $x \leqq e \leqq y$.

Given a thread S which has a zero as a least element, we construct a new thread which we denote by $\mathscr{R}(S)$. Let S' be a copy of $S\backslash\{0\}$, and let x' be the element of S' corresponding to the element x of $S\backslash\{0\}$; put $0' = 0$. Let $\mathscr{R}(S) = S' \cup S$, and extend the order on S to $\mathscr{R}(S)$ by reversing the order in S' and declaring each element of S' to be less than every element of S. Now extend the multiplication in S to $\mathscr{R}(S)$ by defining $x'y = yx' = (xy)'$ and $x'y' = xy$. It is easy to verify that $\mathscr{R}(S)$ is a thread.

We state Lemma 1 of [1] which will be repeatedly used without reference. *If a, b and c are elements in a thread, then* $[ac, bc] \subset [a, b]c$ *and* $[ca, cb] \subset c[a, b]$. *The same holds for open and for half-open intervals.* The proof is a simple application of the fact that a continuous image of a connected set is connected.

If there is a homeomorphism between threads S and T which is also an algebraic homomorphism, S and T are *iseomorphic* and we write $S \approx T$. If the iseomorphism is also order preserving (it must either preserve or reverse the order), then S and T are *isomorphic* and we write $S \cong T$. A *subthread* is, of course, a connected subsemigroup. The *order dual* of a thread is the thread obtained by reversing the order while leaving the multiplication unchanged. As in [8], $H(e)$ is the maximal subgroup containing the idempotent e, $\Gamma(x)$ is the topological closure of the set of powers of x, and $J(x)$ is the ideal generated by x.

The groups of positive and non-zero real numbers will be denoted by $\mathscr{P}$ and $\mathscr{X}$ respectively. Throughout the paper, S will always be a thread.

2. Maximal subgroups. Let e be an idempotent in an arbitrary thread S. We wish to investigate the maximal subgroup $H(e)$ of S having e as its identity. We recall that

$$H(e) = eSe \cap \{x \mid e \in xS \cap Sx\} .$$

Since $H(e)$ is an algebraic group and a topological semigroup, it is homogeneous. Thus, if $H(e)$ contains any open interval of S, it contains an open interval about e. Denoting the component of $H(e)$ containing e by G, either $G = e$ or e is a cut point of G. But G is clearly a cancellative thread, and by a theorem of Acél and Tamari (as stated on page

81 of [1]), every such thread is isomorphic with a subthread of $\mathscr{P}$. Since the only subthread of $\mathscr{P}$ of which the identity is a cut point is $\mathscr{P}$ itself, we see that $G = e$ or $G \cong \mathscr{P}$.

Again, observe that translations of eSe, the set on which e acts as an identity, by elements of $H(e)$ are homeomorphisms. Thus, if any element of $H(e)$ is a cut point of eSe, then e is a cut point. Consequently, if $H(e)$ contains more than two elements, then e is a cut point of eSe.

2.1 LEMMA. *If e is an idempotent in S, then either $e = eSe$, or e is an endpoint of $eS \cup Se$, or e cuts eSe. In the first two cases, $H(e)$ contains at most two elements; while in the last, the identity component of $H(e)$ is isomorphic with $\mathscr{P}$.*

Proof. It will suffice for the proof to show that the following are equivalent: the identity component of $H(e)$ is isomorphic with $\mathscr{P}$; $H(e)$ contains more than two element; e cuts eSe; $e \neq eSe$ and e cuts $eS \cup Se$. Moreover, the first of these obviously implies the second; we have already seen that the second implies the third; and the third clearly implies the forth.

Suppose then that $e \neq eSe$ and that e is a cut point of $eS \cup Se$. Since $eS \cap Se = eSe$, this means that e cuts one of eS and Se, and that $Se \neq e \neq eS$. The two cases being similar, assume that e cuts eS, and choose a and b in eS such that $a < e < b$. Using the continuity of multiplication, there exists an open interval W about e such that $W \subset (a, b)$ and $Wa < e < Wb$. Thus, if x is in W, $e \in (xa, xb) \subset x(a, b)$. Repeating the argument, using W in place of (a, b), we obtain an open interval V about e such that $e \in zW$ for each z in V. Now if $z \in V \cap eSe$, then there exist x in W and s in (a, b) such that $e = zx = xs$. Since $z \in Se$ while $s \in eS$,

$$z = ze = z(xs) = (zx)s = es = s .$$

Hence $V \cap eSe \subset H(e)$. Observing that $V \cap eSe$ is a non-degenerate interval containing e, it follows from the argument of the first paragraph in this section that the identity component of $H(e)$ is isomorphic with $\mathscr{P}$.

2.2 THEOREM. *If e is an idempotent in a thread S and if e cuts eSe, then $H(e) \cong \mathscr{P}$ or $H(e) \approx \mathscr{X}$. Moreover, if the identity component G is not all of S, then the boundary of G in S contains exactly one point f, either $G = (f, \infty)$ or $G = (-\infty, f)$, and f acts as a zero for G.*

Proof. Assuming that e cuts eSe, $G \cong \mathscr{P}$ by 2.1. Certainly, $H(e)$ is a topological group of which G is a normal subgroup. Since the remainder of the theorem is evident otherwise, we assume $G \neq S$.

We claim now that if M and N are cosets of G in $H(e)$ and if $t \in M^*\backslash M$, then tN^* and N^*t contain but one point each. For, since each coset is homeomorphic with G, each coset is open and connected, and thus has at most two boundary points. Since t does not belong to $H(e)$, Nt misses $H(e)$. Thus $Nt \subset (NM)^*\backslash NM$. But Nt is connected and $(NM)^*\backslash NM$, the boundary of some coset, contains at most two points. Hence Nt consists of a single element, and by continuity, the same must be true of N^*t. Likewise, tN^* contains only one element.

Now take f in $G^*\backslash G$, and let C be any coset. If $t \in C^*\backslash C$, then, using the result of the preceding paragraph, $tG^* = te = t$ and $G^*t = et = t$. In particular, f acts as an identity on $C^*\backslash C$. But applying the result again, fC^* and C^*f contain one point each. Thus the coset C has exactly one boundary point. Taking $C = G$, we see that G has only one boundary point f and thus $G = (f, \infty)$ or $G = (-\infty, f)$. Moreover, $fG^* = G^*f = f$ implies that G is iseomorphic (we do not know whether f is the least or the greatest element of G^*) with the thread of non-negative real numbers. If $H(e) = G$, the proof is complete.

Assuming $H(e) \neq G$, it follows from the fact that each coset has only one boundary point in S that there can be only one other coset besides G. Take $b \in H(e)\backslash G$ and observe that the function on G^* which takes g into $b^{-1}gb$ is a continuous automorphism which (since $b^2 \in G$ and G is commutative) is its own inverse. But the only such automorphism of the non-negative real numbers is the identity, and thus $b^{-1}gb = g$ for each g in G. It follows that $H(e)$ is commutative, and from this it is easy to verify that $H(e)$ is iseomorphic with $\mathscr{X}$.

2.3 THEOREM. *Let e be an idempotent in a thread S.*

(1) *If $H(e) = \{d, e\}$ with $d < e$, then $Se = eS = [d, e]$ and there exists a zero for S in (d, e). Denoting the zero by z, $[z, e]$ is a standard thread and $[d, e] \cong \mathscr{R}([z, e])$.*

(2) *If $H(e) \approx \mathscr{X}$ then the complete structure of S is determined. Namely, there exists a positive thread T such that $S \approx \mathscr{R}(T)$.*

Proof. Let $H(e) = \{d, e\}$ with $d < e$, and observe that $[d, e] \subset eSe$. Then $eS \cup Se \leqq e$, for otherwise e cuts eS or Se and 2.1 yields a contradiction. Now since d is in $H(e)$ and $d^2 = e$, left multiplication by d is a strictly decreasing function from eS onto itself. Hence

$$d = de \leqq d(eS) = eS ,$$

so that $[d, e] = eS$. Moreover, there exists a unique element q in eS such that $dq = q$. However, if s is any element of S, then $qs \in eS$ and

$d(qs) = qs$. Since q is unique, q is a left zero for S. Similarly $[d, e] = Se$ and there exists a right zero for S in Se. Evidently these two one sided zeros are equal, and putting $z = q$, z is a zero for S. Now, $[d, e]$ is a subthread with an identity e and a zero z in which $d^2 = e$. Applying part one of Theorem 6.2 in [4], we conclude that $[z, e]$ is a standard thread and that $[d, e] \cong \mathscr{R}([z, e])$.

Turning to the proof of (2), let $H(e) \approx \mathscr{X}$. Since S is iseomorphic with its order dual, we may assume that e is larger than the element u which corresponds to -1. Each coset in $H(e)$ has exactly one boundary point in S and thus $H(e) = (-\infty, h) \cup (f, \infty)$ where $h \leqq f$. Since we have assumed that $u < e$, $(f, \infty) \cong \mathscr{P}$.

One sees easily that $f^2 = h^2 = f$ and that $fh = hf = h$, i.e., $H(f) = \{h, f\}$. If $h = f$ then S is iseomorphic with the multiplicative thread of all real numbers which is certainly $\mathscr{R}(T)$ where T is the thread of non-negative reals. Assuming $h < f$, we may apply the conclusion of (1). Thus S has a zero between h and f, $[z, f]$ is a standard thread, $[h, f]$ is commutative, and $Sf = fS = [h, f]$.

Since f is an identity for $[z, f]$ and a zero for G, each element of G acts as an identity on $[z, f]$. Consequently, $[z, \infty)$ is a positive thread.

If $y \in [z, f]$, then $uy = u(fy) = (uf)y$ and $yu = y(fu)$. Now, f commutes with u, and since $uf \in [h, f]$, uf commutes with y. Thus u commutes with each element of $[z, f]$ as well as with each element of (f, ∞). Armed with these facts, it is a straightforward exercise to show that the function g defined on $\mathscr{R}([z, \infty))$ by $g(t) = t$ and $g(t') = ut$ is an iseomorphism onto S.

2.4 COROLLARY. *If $x^k < x < x^p$ for some x in a thread S and for some positive integers k and p, then $S \approx \mathscr{R}(T)$ for some positive thread T. Moreover, if e is the identity of S, then $x \in H(e)$ and e separates x and x^2.*

Proof. Since x is evidently not an idempotent, we assume that $x < x^2$. The case where $x^2 < x$ is entirely similar. Taking j to be the least positive integer such that $x^{j+1} < x$, we have $2 \leqq j$ and $x < x^j$. Now $x \in (x^{j+1}, x^2)$ and $(x^{j+1}, x^2) \subset x(x, x^j) \cap (x, x^j)x$, so $x = xs = tx$ for some s and t in (x, x^j). It follows that s is a right identity on Sx and that t is a left identity on xS. But $(x, x^j) \subset (x^{j+1}, x^j) \subset xS \cap Sx$, hence $s = ts = t$. Putting $e = s$, $e \in (x, x^j)$ and $(x, x^j) \subset xS \cap Sx = exS \cap Sxe \subset eSe$, so that e is a cut point of eSe. By 2.2, $H(e) \cong \mathscr{P}$ or $H(e) \approx \mathscr{X}$. But $e \in xS \cap Sx$ and $x \in eSe$ imply that $x \in H(e)$, and in view of the hypothesis on the powers of x, $H(e) \cong \mathscr{P}$ is impossible. The result now follows from 2.3.

The following facts concerning the sets eS and Se will be useful later.

2.5 LEMMA. *Let e be an idempotent in a thread S.*

(1) *If $e = eSe$, then either $eS = e$ or $Se = e$; and in either case, SeS is the minimal ideal of S. It is a closed connected set of one sided zeros.*

(2) *Either $eS \subset Se$ or $Se \subset eS$, and thus $SeS = eS \cup Se$.*

Proof. Let $e = eSe$, and recall that $eSe = eS \cap Se$. By way of contradiction suppose that $eS \neq e$ and that $Se \neq e$. Then either $eS \leqq e \leqq Se$ or $Se \leqq e \leqq eS$; and in either case, e is in the interior of $eS \cup Se$. Thus there exists an open interval V about e such that $V^2 \subset eS \cup Se$. Choosing x and y in V such that $x \in eS$, $x \neq e$, $y \in Se$, and $y \neq e$, we have $yx \in eS \cup Se$. But if $yx \in eS$, then

$$e = e(yx)e = (yx)e = (ye)xe = y(exe) = ye = y ,$$

contrary to the choice of y; and if $yx \in Se$, then similarly, $e = x$, contrary to the choice of x.

Now if $eS = e$, $SeS = Se$. Since Se is the image of the connected set S under right translation by e, it is connected; and since it is the set on which right translation by e agrees with the identity mapping, it is closed. Moreover, for each k in SeS,

$$kS = (ke)S = k(eS) = ke = k .$$

Thus, SeS is a closed connected set of left zeros and is clearly the minimal ideal of S. If $Se = e$, then SeS consists of right zeros.

In order to prove (2), consider the three cases of 2.1. If $e = eSe$, then one of eS and Se is just $\{e\}$ and is clearly contained in the other. If e is an endpoint of $eS \cup Se$, then since eS and Se are connected sets extending from e in the same direction, one evidently contains the other. Finally, if e cuts eSe, then the identity component of $H(e)$ extends to one end of the thread. Since $H(e) \subset eS \cap Se$, the result again follows from the connectedness of eS and Se.

3. Subthreads.

3.1 LEMMA. *Let A be a subset of S which contains, with x, all elements larger than x. If A contains no idempotents and if $a < a^2$ for some a in A, then A is a subthread in which* $\max\{x, y\} < xy$ *for each pair of elements x and y in A.*

Proof. Let a be an element in A such that $a < a^2$, and let x be any element of A. If $x^2 < x$, then, since A is evidently connected and since the function mapping each element onto its square is continuous, there is an idempotent between x and a contrary to the assumption that A contains no idempotents. Hence $x < x^2$. If $x^n < x$, for some positive integer n, then there is an idempotent between x and x^2 by 2.4. And again, if $\Gamma(x)$ is bounded, it is a compact semigroup and thus contains an idempotent. Hence $x \in A$ implies that $x \leqq \Gamma(x)$ and that $\Gamma(x)$ is unbounded.

Now suppose that $yz = y$ with y and z in A. For each positive integer n, $yz^n = y$, thus z^n is a right identity for Sy. But both $\Gamma(y)$ and $\Gamma(z)$ are unbounded, so for some n and m, $y^2 < z^n < y^m$. Thus z^n is in Sy and $z^n z^n = z^n$. Since A contains no idempotents, $yz = y$ is impossible.

Finally, if $yz < y$, then, by the continuity of right multiplication by z and the fact that $z < zz$, there exists a t between y and z for which $tz = t$, a contradiction. Hence $y < yz$, and dually $z < yz$.

3.2 LEMMA. *If e is an idempotent, if $eS \cup Se \leqq e$, if C is a connected set containing e as a least element, and if $[e, x) \subset xC \cap Cx$ whenever $x \in C$; then $e \leqq C^2$.*

Proof. Appealing to 2.5 we will lose no generality by assuming that $eS \subset Se$. Thus $t \in eC$ implies $et = te = t$. Moreover, if $t = ex$ with x in C, then $e = sx$ for some s in C, and thus

$$(es)t = (es)(ex) = [(es)e]x = (es)x = e(sx) = e\ .$$

It follows that eC is a subgroup of $H(e)$. But eC is connected and contains e while, by 2.1, $H(e)$ contains at most two elements. Hence $eC = e$.

Now suppose that $xy < e$ for some x and y in C. Clearly $e < x$ and therefore $e < xt$ for some t in C. Now $xy < e < xt$ implies that $e = xw$ for some w between y and t. But if $y < w$, then $xy \in xwC = eC = e$; and if $t < w$, then $xt \in xwC = eC = e$. Since this contradicts $xy < e < xt$, we have $e \leqq xy$. Hence, $e \leqq C^2$.

The following result, which is a generalization of Faucett's Lemma 4 in [5], will be extremely useful in the remainder of the paper.

3.3 THEOREM. *If e and f are idempotents in a thread S and if $eS \cup Se \leqq e < f$, then $[e, f]$ is a standard thread. If, in addition, f cuts fSf, then $[e, \infty)$ is a positive thread.*

Proof. Since $ef \in eS$ and $fe \in Se$, neither ef nor fe is larger than e. But $ef \in Sf$ and $fe \in fS$, and these sets are connected. Thus $e \in Sf \cap fS$, and f acts as an identity on $[e, f]$. Then, for each x in $[e, f]$,

$[e, x] \subset [ex, fx] \cap [xe, xf] \subset [e, f]x \cap x[e, f]$. Consequently, by 3.2, $e \leqq [e, f]^2$, and in particular, e acts as a zero for $[e, f]$.

Now if $fS \cup Sf \leqq f$; then $[e, f] \subset fS$ implies $[e, f]^2 \subset fS$ and the theorem is established. If on the other hand, $fS \cup Sf \not\leqq f$; then, by 2.1, f cuts fSf.

Finally, if f cuts fSf; then, since e cannot be in $H(f)$, it follows from 2.2 that there exists an idempotent h in $[e, f)$ such that (h, ∞) is isomorphic with $\mathscr{P}$. Since $hS \cup Sh \leqq h$, the preceding paragraphs show that $[e, h]$ is a standard thread (it may of course be simply one point if $e = h$). Evidently then, $[e, \infty)$ is a positive thread of which $[e, f]$ is a standard subthread.

3.4 LEMMA. *If $[a, b]$ and $[b, c]$ are subthreads, then so is $[a, c]$.*

Proof. Let $x \in [a, b]$, let $y \in [b, c]$, and suppose that $c < xy$. Then, since $xb \in [a, b]$, $[b, c] \subset [xb, xy] \subset x[b, c]$. Now $\Gamma(x)$ and $[b, c]$ are both compact, and by Wallace's Theorem 1 in [11], we conclude that $[b, c] = x[b, c]$ contrary to $c < xy$. Thus $xy \leqq c$; and similarly, one proves that $a \leqq xy$ and that $a \leqq yx \leqq c$.

3.5 THEOREM. *If e and f are any two idempotents in a thread, then the closed interval between them is a subthread.*

The proof of this result will be postponed until the end of section four. The proof will be much easier then, and we promise not to apply the result in the meanwhile.

4. The minimal ideal.

4.1 THEOREM. *If S has no minimal ideal, then a zero may be adjoined as an endpoint and the resulting semigroup is again a thread.*

Proof. We show first that S has no bounded ideals. Indeed, if M is a bounded ideal, then M^* is a compact ideal. In particular, M^* is a compact topological semigroup, and as such (see Theorem 3 in [10]), there is an idempotent e in M^* such that eM^*e is a group. But M^* is an ideal and thus $eSe = eM^*e$, thus eSe is a compact connected group. It follows from 2.1 and 2.5 that $eSe = e$ and that SeS is the minimal ideal of S. Hence, S has no bounded ideals.

Next observe that every ideal contains a connected ideal. For if x is any element of an ideal J, then SxS is a connected ideal contained in J.

Now fix y in S and let J be an ideal contained in $S\backslash y$. Such an ideal does exist, for if not, then y is in each ideal of S, the intersection of all ideals is not empty, and S has a minimal ideal. Since we may take J to be connected, we lose no generality if we assume that $J < y$.

If $x < y$ then again there is a connected ideal M contained in $S\backslash x$. In fact, $M < x$, for otherwise $M \cap J$ is a bounded ideal. Thus M^* is a connected, closed, unbounded ideal whose elements are all less than or equal to x. Hence, for each x less than y, there exists a c not greater than x such that $(-\infty, c]$ is an ideal. Evidently a zero can be adjoined as a least element.

4.2 THEOREM. *If S has a minimal ideal K, then either $S = K$ and $S \cong \mathscr{P}$, or there exists an idempotent e such that $e = eSe$. In the second case, it follows from* 2.5 *that $K = SeS$ and is a closed connected set of one sided zeros.*

Proof. Let $x \in K$ and consider the subthread xK. We claim that xK contains an idempotent. If not, we may assume without loss of generality that $a < a^2$ for some a in xK. It follows from 3.1 that $a < (xK)a(xK)$. But $K(ax)K$ is an ideal contained in K and must therefore be equal to K. Consequently $(xK)a(xK) = xK$ so that $a \in (xK)a(xK)$. Hence, xK (and by an analogous proof, Kx as well) contains an idempotent for each x in K.

Let e be an idempotent in K and recall that one of eS and Se contains the other by 2.5. Assuming $eS \subset Se$, we have $eSe = eS = eK$. Notice that eSe contains no idempotents other than e. For if $f \in eSe$, then $f = ef = fe$. But also, $f \in K$ so that $e \in SfS = Sf \cup fS$, hence $e = f$.

Now if $x \in eSe$, then xK contains an idempotent. But

$$xK = (xe)K = x(eK) = x(eSe) \subset eSe ,$$

and eSe contains only one idempotent. Hence $x \in eSe$ implies $e \in x(eSe)$, i.e. eSe is a group.

Since eSe is also connected, either $e = eSe$ or $eSe \cong \mathscr{P}$. In the latter case, eSe is both open and closed and hence $eSe = S$. Thus $S \cong \mathscr{P}$ and $S = K$.

We are now in a position to give the overdue proof of Theorem 3.5. We are to show that the closed interval between two idempotents in a thread is a subthread.

Proof of 3.5. Since we can adjoin a zero if not, we assume that S has a minimal K; and since the assertion is vacuously true otherwise, we assume that K consists of one sided zeros. Observe that because of the trivial multiplication within K, any closed interval contained in K is a subthread.

If f is an idempotent larger than each element of K, and if $k = \sup K$, then $[k, f]$ is a standard thread by 3.3. Similarly, if $f < K$ and if

$l = \inf K$, then $[f, l]$ is the order dual of a standard thread. Moreover, the interval between any two idempotents in a standard thread is again a standard thread.

Finally, using these facts along with Lemma 3.4, which allows us to sew the subthreads together, the theorem follows easily.

5. **Threads with a zero.** The principal result of this section is the characterization in 5.5 of all threads which have a zero as an endpoint and for which $S^2 = S$. However, the series of lemmas leading to this result will be used again in the following section; consequently they are more troublesome than is apparently necessary.

It will be convenient to introduce the following partial order whenever S has a zero:

$$x \prec y \text{ if and only if } 0 \leqq x < y \text{ or } y < x \leqq 0 \,.$$

Obviously this does define a partial order on S.

5.1 LEMMA. *Let S be a thread with a zero in which each idempotent e* is an endpoint *of eSe. Then $\Gamma(x)$ is compact for each x in S, $J(x) \leqq x$ when $0 < x$, and $x \leqq J(x)$ when $x < 0$.*

Proof. We show first that $0 < x$ implies $\Gamma(x) \leqq x$. This is clear if $x \in [0, e]$ for some idempotent e, for $[0, e]$ is a standard thread by 3.3. Assume that x is larger than each idempotent, and let e be the largest idempotent. Now if $x < x^2$, then by 3.1, $\max\{y, x\} < xy$ for each y larger than e. By continuity, $x \leqq xe$, and thus, $0 < e < x$ while $x \in Se$. But using 2.1, this implies that e cuts eSe, contrary to hypothesis. Hence $x^2 < x$, and it follows from 2.4 and the assumption that each idempotent e is an endpoint of eSe that $\Gamma(x) \leqq x$. Repeating the argument with all inequalities reversed, $x \leqq \Gamma(x)$ when $x < 0$.

Next we prove that $\Gamma(\mathrm{x})$ is compact for each x larger than 0. This is obvious if $\Gamma(x) \subset [0, x]$. If $\Gamma(x) \not\subset [0, x]$, let x^j be the first power of x which is less than 0. Since $x^j \leqq \Gamma(x^j)$, $x^{jn} \in [x^j, x]$ for each positive integer n. By the choice of j, $x^i \in [x^j, x]$ for each positive integer i less than j as well; therefore $\Gamma(x) \subset [x^j, x]^2 \cup [x^j, x]$, a compact set. Similarly, $\Gamma(x)$ is compact when x is less than 0.

To establish the last statement of the lemma, let $0 < x$ and suppose that $x \leqq sxt$. Then $[0, x] \subset s[0, x]t$, while $[0, x]$, $\Gamma(s)$, and $\Gamma(t)$ are compact. By Corollary 2 in [11], $[0, x] = s[0, x]t$. Therefore $SxS \leqq x$, and using the one sided analogues of the result just used, it can be proved that $Sx \leqq x$ and that $xS \leqq x$. This gives $J(x) \leqq x$, and it follow similarly that $x \leqq J(x)$ when $x < 0$.

5.2 LEMMA. *Let S be a thread with a zero. If $S^2 = S$, then, for* *each x larger than 0, there exist an element u and a compact set A such that $x = uA$ and such that x is in the interior of uV for each open set V which contains A.*

Proof. Given x larger than 0, choose y larger than x; or if x is maximal, put $y = x$. Since $S^2 = S$, we can choose u and v in S so that $y = uv$. Now if $0 < v$, let

$$p = \inf\{t \mid 0 \leqq t \leqq v \text{ and } x \leqq u[t, v]\} ,$$
$$q = \sup\{t \mid p \leqq t \leqq v \text{ and } x = u[p, t]\} ,$$

and let $A = [p, q]$. And if $v < 0$, define p, q, and A analogously. The details are easy to verify in either case. Actually, this proof is just a slight generalization of the usual proof of the intermediate value theorem for continuous functions on the real line.

5.3 LEMMA. *Let S have a zero, let $S^2 = S$, and let $J(x) \leqq x$ for $x > 0$. If T is a connected set containing 0 such that Tu is bounded for each u in S, and if h is defined on $\{x \mid 0 \leqq x\}$ by $h(x) = \sup Tx$, then h is continuous.*

Proof. Since $Tx \subset J(x) \leqq x$, $0 \leqq h(x) \leqq x$ for each x greater than 0, and consequently, h is continuous at 0.

Now let $0 < x$ and let $a < h(x) < b$. Choose c and t so that $t \in T$, $a < tx$, and $h(x) < c < b$, and let u and A be as in 5.2. We have

$$(Tu)^*A \subset (TuA)^* = (Tx)^* \leqq h(x) < c ,$$

and since Tu is bounded by hypothesis, $(Tu)^*$ and A are both compact. Thus (Lemma 2 in [9]) there exists an open set V such that $A \subset V$ and $TuV < c$. If $y \in uV$, then $h(y) = \sup Ty \leqq c < b$; and by 5.2, uV contains an open set about x.

Since $a < tx$, there is another open set W about x such that $a < tW$. Thus, $y \in W$ implies

$$a < ty \leqq \sup Ty = h(y) .$$

Taking the intersection of W and the interior of uV, we have produced a neighborhood of x which is mapped into (a, b). Thus h is continuous.

5.4 LEMMA. *Let S have a zero and let A be a set such that $\Gamma(a)$ is compact for each a in A. If $[0, x) \subset Ax$ for each x greater than 0, then $rt \leqq st$ whenever $0 \leqq r < s$.*

Proof. If 0 lies strictly between rt and st, then there exists c in (r, s) for which $ct = 0$. But then $r \in [0, c)$ so that $rt \in (Ac)t = A(ct) = 0$

which contradicts the assumption that zero lies strictly between rt and st. Hence rt and st are at least comparable with respect to $\prec$.

Since $r \in [0, s)$, we can choose an a in A such that $r = as$. Now if $st \leqq rt$, then

$$\{x \mid 0 \leqq x \leqq st\} \subset \{x \mid 0 \leqq x \leqq ast\} \subset a\{x \mid 0 \leqq x \leqq st\} ;$$

and since both $\Gamma(a)$ and $\{x \mid 0 \leqq x \leqq st\}$ are compact, we have $\{x \mid 0 \leqq x \leqq st\} = a\{x \mid 0 \leqq x \leqq st\}$ (Theorem 1, [11]). Thus $rt = ast \leqq st$.

5.5 THEOREM. *If S is a thread with a zero as a least element and if $S^2 = S$, then S is a standard thread, or S is a standard thread with its identity removed, or S is a positive thread.*

Proof. If there exists an idempotent f in S which cuts fSf, then S is a positive thread by 3.3. Hence, assume that no idempotent e cuts eSe. By 5.1, $\Gamma(x)$ is compact and $J(x) \subset [0, x]$ for each x in S.

If we put $h(x) = \sup Sx$, then h is continuous by 5.3. We claim moreover that h is the identity. For suppose $h(a) \neq a$. Then $a \neq 0$ and $h(a) < a$. Using the continuity of h we choose an element t and an open interval V, containing a, such that $h(V) < t < V$. Since $S^2 = S$, we can write $a = yx$ and thus $h(0) < a \leqq h(x)$. Again using continuity, choose, b so that $a = h(b)$. Now take any c in V such that $c < a$, and observe that $c \in Sb = S(Sb)$. Thus $c \in Sp$ for some p in Sb. But then $c \leqq p \leqq a$ so that $p \in V$, and hence $h(p) < t < c$ contrary to $c \in Sp$.

Since h is the identity, $[0, x) \subset Sx$ for each x; and an analogous argument gives $[0, x) \subset xS$. Thus we conclude from 5.4 and its left-right dual that the multiplication in S is monotone.

If S is compact with w as its largest element, then w is an idempotent and S is a standard thread. Indeed, we can write $w = xy$, and it then follows from $J(x) \leqq x$ and $J(y) \leqq y$ that $w = x = y$.

If S is not compact, then let T be the semigroup obtained by adjoining an identity to S, and extend the order of S to T by declaring that the identity is larger than each element of S. Since S is not compact, T is evidently connected. Finally, the continuity of multiplication in T follows immediately from the continuity and monotonicity is S along with the relation $[0, x) \subset xS \cap Sx$. Thus, T is a thread, and in fact, a standard thread.

5.6 COROLLARY. *If S is a thread with no idempotents, and if $S^2=S$, then S is iseomorphic with the real interval* $(0, 1)$ *under the natural multiplication.*

Proof. Since S has no idempotents, it follows from 4.2 that S has no minimal ideal; and by 4.1, a zero may be adjoined as an endpoint to S. Then either the extended thread or its order dual satisfies the hypotheses of 5.5. Thus, S must be the result of removing both the zero and the identity from a standard thread which has no other idempotents and which has no nilpotent elements. But Faucett proved in Theorem 2 of [5] that any such standard thread is iseomorphic with $[0, 1]$.

6. Threads in which $S^2 = S$.

Let S be a thread satisfying $S^2 = S$. If S has no minimal ideal, then a zero may be adjoined as an endpoint. After taking the order dual, if necessary, the extended thread can then be described by 5.5. Consequently, the structure of S is determined. If S does have a minimal ideal, and if $K = S$, then the structure of S is given by 4.2.

Thus, we have left only the case where S has a proper minimal ideal which consists either of left zeros or of right zeros. We include, of course, the special case in which S has a zero. Throughout this section, when we say that S has a minimal ideal K, it will be tacitly assumed that K is proper and thus consists of zeros.

The following notation will be used when there exists a minimal ideal K:

$$R = \{t \mid k \leqq t \text{ for each } k \text{ in } K\} ,$$
$$L = \{t \mid t \leqq k \text{ for each } k \text{ in } K\} ,$$

If S has a zero, we have, $R = \{t \mid 0 \leqq t\}$ and $L = \{t \mid t \leqq 0\}$.

6.1 LEMMA. *If S has a minimal ideal K, if $S^2 = S$, and if there exists a connected proper ideal of S containing L, then $R^2 = R$.*

Proof. Let J be a connected proper ideal containing L, and let $c = \sup J$. If $J^* = S$, then $S \backslash J = c$; and since $S^2 = S$, c is an idempotent. Thus by 3.3, R is a standard thread, and certainly $R^2 = R$.

Now assume that J^* is a proper ideal, and let $B = \{t \mid c \leqq t\}$. Since J^* is closed and connected, $T = S/J^*$ is a non-degenerate thread with a zero as a least element and with $T^2 = T$. By 5.5, T is a positive thread or T is a standard thread with or without its identity. In any case, $[0, t) \subset tT \cap Tt$ for each t larger than zero in T. Since the natural homomorphism of S onto T is strictly increasing on B and takes J^* onto 0, we conclude that $[c, b) \subset bB \cap Bb$ for each b larger than c in S.

Taking $k = \sup K$, k is the least element of R and $kS \cup Sk \leqq k$. Since bR and Rb are connected sets, $[k, b) \subset bR \cap Rb$ for each b larger than c. Now fix b larger than c and let r be any element of R such that $r \leqq c$. Then there exist s and t in R such that $r = sb = bt$. Thus,

$$[k, r) \subset [sk, sb) \subset s[k, b) \subset sbR = rR ,$$

and similarly, $[k, r) \subset Rr$. Hence, for each r in R, $[k, r) \subset rR \cap Rr$. Applying 3.2, with $C = R$ and $e = k$, we have $R^2 \subset R$. On the other hand, $R^2 \supset R$ follows immediately from the facts that J is a proper ideal containing L and that $S^2 = S$.

6.2 LEMMA. *Let S have a zero, and let $S^2 = S$. If $R \subset LS \cup SL$ and if there exists a set A such that $(d, 0] \subset dA \cap Ad$ for each d less than 0 and such that $\Gamma(a)$ is compact for each a in A, then the multiplication in S is monotone with respect to $\prec$ and 0 is an endpoint of L^2, R^2, LR, and RL.*

Proof. First, notice that the second conclusion follows from the first. Indeed, it suffices to show that if x and y are $\prec$-comparable and if u and v are $\prec$-comparable, then so are xu and yv. But if $x \prec y$ and $u \prec v$; then, assuming that the multiplication is monotone, $xu \leqq yu$ and $yu \leqq yv$, so that $xu \leqq yv$.

To prove monotonicity, observe that (using both order and left-right duality) 5.4 gives $dt \leqq pt$ and $td \leqq tp$ whenever $p < d \leqq 0$. Since $R \subset LS \cup SL$, while each of LS and SL is a connected set containing 0, either $R \subset LS$ or $R \subset SL$; and without loss of generality we assume that $R \subset LS$.

Now if $x > 0$, choose d in L and q in S such that $x = dq$. Then

$$[0, x) = [0q, dq) \subset (d, 0]q \subset Adq = Ax .$$

Thus, again by 5.4, $rt \leqq st$ whenever $0 \leqq r < s$.

The only case left to demonstrate is $tr \leqq ts$ for $0 \leqq r < s$. Again choose d and q with d in L so that $dq = s$. Then $r \in [0q, dq)$ so that $r = pq$ for some p in $(d, 0]$. Since $d < p \leqq 0$, we have $tp \leqq td$, i.e., either $0 \leqq tp \leqq td$ or $td \leqq tp \leqq 0$. In either case we can multiply on the right by q to obtain

$$tr = tpq \leqq tdq = ts .$$

6.3 LEMMA. *If S has a zero, if $S^2 = S$, and if either $L^2 = L$ or $R^2 = R$; then the conclusions of 6.2 hold.*

Proof. The other case being quite similar, let us assume that $L^2 = L$. By 5.5, the order dual of L is a positive thread or a standard thread with or without its identity. In the first case, L has an identity e, $\Gamma(x)$ is compact for each x in $[e, 0]$, and $(d, 0] \subset d[e, 0] \cap [e, 0]d$ for each d less than 0. In the second case, $\Gamma(x)$ is compact for each x in L and $(d, 0] \subset dL \cap Ld$ when $d < 0$.

Hence, if $R \subset LS \cup SL$ as well, then monotonicity follows from 6.2. However, even if $R \not\subset LS \cup SL$, we may still apply 5.4 to conclude that

$dt \leqq pt$ and $td \leqq tp$ for $p < d \leqq 0$. Thus, if we show that $R^2 = R$, then monotonicity follows by dualizing the foregoing argument.

Now assume that $R \not\subset LS \cup SL$; we must show that $R^2 = R$. If R contains an idempotent e which cuts eSe, this is an immediate consequence of 3.3. Assume that each idempotent e in R is an endpoint of eSe.

If each idempotent f in L is also an endpoint of fSf, then by 5.1, $J(x) \leqq x$ whenever $0 < x$. From this it follows that $L \cup SL \cup LS$ is an ideal, and thus a connected proper ideal containing L. If some idempotent f in L cuts fSf, then by 3.3 and 2.5, $fS \cup Sf$ is a connected proper ideal containing L. Thus, in either case, 6.1 yields $R^2 = R$.

6.4 LEMMA. *If S has a zero, if $S^2 = S$, if $J(x) \leqq x$ for $x > 0$, and if $x \leqq J(x)$ for $x < 0$; then either $L \subset L^2$ or $R \subset R^2$.*

Proof. Suppose by way of contradiction that neither $L \subset L^2$ nor $R \subset R^2$. Since $L \subset S^2 = L^2 \cup SR \cup RS$, while each of the three sets on the right is connected and contains 0, L must be contained in one of the three. Consequently $L \subset SR$ or $L \subset RS$.

If $L \subset SR$, then

$$R \subset S^2 = SL \cup SR \subset S(SR) \cup SR = SR = R^2 \cup LR\,,$$

and thus $R \subset LR$. Now

$$R \subset LR \subset L(LR) = (L^2 \cap L)R \cup (L^2 \cap R)R \subset (L^2 \cap L)R \cup R^2\,.$$

Again, $R \subset (L^2 \cap L)R$; and hence $L \subset SR \subset S(L^2 \cap L)R$.

If $L \subset RS$, we obtain similarly, $L \subset R(L^2 \cap L)S$. But then, in either case, $L \subset S(L^2 \cap L)S$; and choosing d less than L^2, $d \in SpS$ for some p in $L^2 \cap L$ contrary to $p \leqq J(p)$.

6.5 LEMMA. *Let S have a zero, let $S^2 = S$, let Sx be bounded for each x, let $J(x) \leqq x$ for $0 < x$, let $x \leqq J(x)$ for $x < 0$, and define a function f on S by:*

$$f(x) = \begin{cases} \sup Sx\,, & \text{if } 0 \leqq x\,, \\ \inf Sx\,, & \text{if } x \leqq 0\,. \end{cases}$$

Then f is continuous. Moreover, if f is the identity on a set B, then f also acts as the identity on BS.

Proof. The continuity of f is immediate from 5.3 and its order dual.

Since Sx is connected and contains 0, $f(x) = x$ if and only if $\{y \mid y \leqq x\} \subset Sx$. Now if $f(b) = b$, and if $t = bs$, then

$$\{y \mid y \leqq t\} = \{y \mid y \leqq bs\} \subset \{y \mid y \leqq b\}s \subset Sbs = St.$$

Thus $f(b) = b$ implies $f(bs) = bs$.

6.6 LEMMA. *Let S be a thread with a zero in which $J(x) \leqq x$ for $0 < x$ and $x \leqq J(x)$ for $x < 0$. Let $\Gamma(x)$ be compact for each x in S, and let $R^2 = S$. Then $R = S$.*

Proof. Since $J(x) \leqq x$ for $0 < x$, $L \cup LS \cup SL$ is an ideal. If $R \not\subset LS \cup SL$, then $L \cup LS \cup SL$ is a proper connected ideal containing L, and by 6.1, $S = R^2 = R$. Hence we assume that $R \subset LS \cup SL$.

If $J(x)$ is unbounded for some x larger than 0, then L is unbounded and $L \subset J(x)$. But then, R is unbounded, and at the same time $R \subset LS \cup SL \subset J(x) \leqq x$. Hence $x \geqq 0$ implies that $J(x)$ is bounded. If $J(x)$ is unbounded for some $x < 0$, then R is unbounded and $R \subset J(x)$. Since $R^2 = S$, $x \in J(r)$ for some r in R. Hence $R \subset J(r) \leqq r$, a contradiction. Thus, Sx and xS are bounded for each x in S.

In the remainder of the proof we will prove that $(d, 0] \subset Sd \cap dS$ for each d less than 0. Actually we only prove that $(d, 0] \subset Sd$; the other case depends on an analogous argument. Then we will be able to apply 6.2 and conclude that 0 is an endpoint of R^2, and thus $S = R$.

Let $a \in S$ and choose h in R such that $a \in Sh$. From $S^2 = S$ it follows that $a \in Sa_1$ for some a_1 in Sh. Continuing inductively, we construct an infinite sequence $\{a_n\}$ such that $a_n \in Sa_{n+1}$ and $a_{n+1} \in Sh$. Replacing $\{a_n\}$ by an infinite subsequence if necessary, we may assume that either $\{a_n\} \subset L$ or $\{a_n\} \subset R$. In either case, it follows from the hypotheses that $a_n \leqq a_{n+1}$.

Since each $a_n \in Sh$ while Sh is bounded, the least upper bound of $\{a_n\}$ with respect to $\prec$ exists. Let b be this least upper bound. Let f be the function defined in Lemma 6.5. Then $a_n \leqq f(a_{n+1}) \leqq a_{n+1}$, and since f is continuous, $f(b) = b$. This means that $\{x \mid 0 \leqq x \prec b\} \subset Sb$. Now if $a_1 = b$ then $a \in Sa_1 = Sb$, and if $a_1 \prec b$ then $a \in Sa_1 \subset S(Sb) = Sb$. We have shown that for each a in S there exists b such that $a \in Sb$ and $f(b) = b$.

Let $B = \{x \mid f(x) = x\}$ and let $A = BS$. We have just proved that $SB = S$ and thus $SA = S$. Moreover, f is the identity on A by 6.5; and since we can write $A = \cup \{bS \mid b \in B\}$, A is a connected right ideal.

Suppose that neither $L \subset A$ nor $R \subset A$. Then choose d in L and r in R such that $d < A < r$. Since $SA = S$, there exist s and t in A such that $d \in Ss$ and $r \in St$. It follows from $d < A$ and $A < r$ that $t < 0 < s$. But then $s \in [0, r]$ and $[0, r] \subset St$, so that $d \in Ss \subset St$ contrary to $t \leqq J(t)$. Hence, either $L \subset A$ or $R \subset A$. Moreover, if $R \subset A$ then $L \subset R^2 \subset AR \subset A$, and thus $L \subset A$ in any case. Finally, f acts as the identity on L and thus $(d, 0] \subset Sd$ for each d less than 0.

6.7 THEOREM. *Let S be a thread with a proper minimal ideal K, and let $S^2 = S$. Then, passing to the order dual if necessary, $R^2 = R$ and is thus completely described by* 5.5. *Moreover, if $L \subset L^2$ then $L = L^2$ as well. Finally, K does not separate R^2, L^2, LR, or RL, and the multiplication in S/K is monotone with respect to $\prec$.*

Proof. Since K is connected and closed, $T = S/K$ is a thread which obviously has a zero and satisfies $T^2 = T$. We show first that, after passing to the order dual if necessary, $R^2 = R$ in T.

If some idempotent f cuts fSf in T then by 3.3 either L or R is a positive thread, and clearly either $L^2 = L$ or $R^2 = R$. Otherwise, each idempotent e in T is an endpoint of eSe and 5.1 can be applied. Thus $J(x) \leqq x$ for $x > 0$, $J(x) \geqq x$ for $x < 0$, and $\Gamma(x)$ is compact for each x. Now by 6.4, either $L \subset L^2$ or $R \subset R^2$, and passing to the order dual if necessary, we assume that $R \subset R^2$. Since $d \leqq J(d)$ for each d less than 0, R^2 is itself a thread. Moreover, it satisfies the hypotheses of 6.6, and thus $R^2 = R$.

Next, applying 6.3 to T, we see that the multiplication in T is monotone with respect to $\prec$, and that 0 is an endpoint of L^2, R^2, LR, and RL. This evidently gives the last assertion of the theorem.

Finally, going back to S itself, we clearly have $R \subset R^2$. Since K does not separate R^2, $K \cup R$ is a thread satisfying 6.1 and thus $R = R^2$. Likewise, if $L \subset L^2$ in S, then $L = L^2$.

REFERENCES

1. A. H. Clifford, *Connected ordered topological semigroups with idempotent endpoints I*, Trans. Amer. Math. Soc., **88** (1958), 80-98.
2. ———, *Connected ordered topological semigroup with idempotent endpoints II*, ibid., **91** (1959), 193-208.
3. ———, *Totally ordered commutative semigroups*, Bull. Amer. Math. Soc., **64** (1958), 305-316.
4. Haskell Cohen and L. I. Wade, *Clans with zero on an interval*, Trans. Amer. Math. Soc., **88** (1958), 523-535.
5. W. M. Faucett, *Compact semigroups irreducibly connected between two idempotents*, Proc. Amer. Math. Soc., **6** (1955), 741-747.
6. P. H. Mostert and A. L. Shields, *On a class of semigroups on E_n*, Proc. Amer. Math. Soc., **7** (1956), 729-734.
7. ———, *On the structure of semigroups on a compact manifold with boundary*, Ann. of Math., **65** (1957), 117-143.
8. A. D. Wallace, *The structure of topological semigroups*, Bull. Amer. Math. Soc., **61** (1955), 95-112.
9. ———, *A note on mobs*, Anis Acad. Brasil. Ci., **24** (1952), 329-334.
10. ———, *A note on mobs II*, ibid., **25** (1953), 335-336.
11. ———, *Inverses in Euclidean mobs*, Math. J. Okayama Univ., **3** (1953), 23-28.

TULANE UNIVERSITY

AN ESTIMATE FOR DIFFERENTIAL POLYNOMIALS IN $\frac{\partial}{\partial z_1}, \cdots, \frac{\partial}{\partial z_n}$

FRANÇOIS TREVES

This article is concerned with polynomials with respect to the Cauchy-Riemann operators

$$\frac{\partial}{\partial z_1} = \frac{1}{2}\left(\frac{\partial}{\partial x_1} + i\frac{\partial}{\partial y_1}\right), \cdots, \frac{\partial}{\partial z_n} = \frac{1}{2}\left(\frac{\partial}{\partial x_n} + i\frac{\partial}{\partial y_n}\right).$$

We establish an L^2-estimate, for such polynomials, and derive from it uniqueness in a class of Cauchy problems. The estimate is quite similar to Hörmander's inequalities and, in fact, can be essentially deduced from them. However, its direct proof is very simple and leads to a constant better than the one in Hörmander's inequalities. We have therefore preferred to present it thoroughly.

The last part of the paper studies a class of Cauchy problems and applies the estimate to obtain uniqueness. There the methods are quite standard (see for instance Nirenberg [1]). The nature of the differential operators considered allows us to remove the strict convexity of the domains in which the solutions are studied, and replace it by a weaker condition.

1. The inequality. We consider a polynomial $P(z)$ on C^n. We set, for $p = (p_1, \cdots, p_n) \in N^n$:

$$P^{(p)}(z) = \left(\frac{\partial}{\partial z_1}\right)^{p_1} \cdots \left(\frac{\partial}{\partial z_n}\right)^{p_n} P(z).$$

We shall denote by $P(D_z)$ the differential polynomial on R^{2n} obtained by substituting $\partial/\partial z_j = 1/2(\partial/\partial x_j + (1/i)(\partial/\partial y_j))$ for z_j $(1 \leqq j \leqq n)$ in $P(z)$.

If S is a subset of R^{2n}, we denote by $\beta_j(S)$ the diameter of S in the complex "direction" $z_j : \beta_j(s) = \sup_{z', z'' \in S} |z'_j - z''_j|$.

THEOREM 1. *Let Ω be an open set in R^{2n}. For all polynomials $P(z)$ on C^n, all functions $H(z)$ defined and holomorphic in Ω, all functions $\phi(x, y) \in C_0^\infty(\Omega)$, all $p = (p_1, \cdots, p_n) \in N^n$:*

$$\| e^{H(z)} P^{(p)}(D_z)\phi \|_{L^2} \leqq \beta_1^{p_1}(\Omega) \cdots \beta_n^{p_n}(\Omega) \| e^{H(z)} P(D_z)\phi \|_{L^2}.$$

It is enough to prove the inequality in Theorem 1 for $p_1 = 1$ and $p_j = 0$ for $j \geqq 2$. We shall denote by $P_1(z)$ the corresponding $P^{(p)}(z)$. On the other hand, we set, for $j = 1, \cdots, n$:

Received January 15, 1960.

$$H_j(z) = \frac{\partial}{\partial z_j} H(z) ,$$

$$A_j = \frac{\partial}{\partial z_j} - H_j(z) .$$

Observe that for all $1 \leqq j,\ k \leqq n$, $(\partial/\partial z_j)H_k(z) = (\partial/\partial z_k)H_j(z)$; it follows from this that the A_j's all commute.

The formal adjoint of A_j is $A_j^* = -\partial/\partial\bar{z}_j - \overline{H_j(z)}$. Observe first that the A_j^*'s all commute, since the A_j's do. But also the A_j^*'s commute with the A_k's, for the $\overline{H_j(z)}$'s are antiholomorphic functions of $\bar{z}$ in Ω.

If $Q(z)$ is a polynomial on C^n, we denote by $Q(A)$ the differential operator on R^{2n} obtained by substituting A_j for z_j $(1 \leqq j \leqq n)$ in $Q(z)$. If $\bar{Q}(z)$ is the polynomial whose coefficients are the complex conjugates of the ones of $Q(z)$, the formal adjoint of the operator $Q(A)$ is $Q^*(A) = \bar{Q}(A^*) = \bar{Q}(A_1^*, \cdots, A_n^*)$. It is easy to check that:

$$(1) \qquad (P_j)^*(A) = -(P^*)_j(A) = -[P^*(A), \bar{z}_j] .$$

Let us denote by $(\, , \,)$ and $\| \quad \|$ the inner product and the norm in $L^2(R^{2n})$. We may as well assume that $\beta_1(\Omega) = 2d$, with $d = \sup_{z\in\Omega} |z_1|$. If $\phi(x, y)$ has its support in Ω, we can write:

$$\begin{aligned}(P^*(A)\phi, z_1(P_1)^*(A)\phi) &= (\phi, P(A)[z_1(P_1)^*(A)\phi]) \\ &= (\bar{z}_1\phi, (P_1)^*(A)P(A)\phi) + (\phi, (P_1)^*(A)P_1(A)\phi) \\ &= (P_1(A)(\bar{z}_1\phi), P(A)\phi) + \| P_1(A)\phi \|^2 \\ &= (P_1(A)\phi, z_1P(A)\phi) + \| P_1(A)\phi \|^2 .\end{aligned}$$

Hence:

$$-\| P_1(A)\phi \|^2 = (P_1(A)\phi, z_1P(A)\phi) + (\bar{z}_1P^*(A)\phi, (P^*)_1(A)\phi) ,$$

by applying (1). We get at once:

$$(2) \quad \| P_1(A)\phi \|^2 \leqq d \| P_1(A)\phi \| \cdot \| P(A)\phi \| + d \| P^*(A)\phi \| \cdot \| (P^*)_1(A)\phi \| .$$

But since the A_j and the A_k^* all commute with each other, $P(A)$ and $P^*(A)$ commute, and $P_1(A)$ and $(P_1)^*(A)$ do. Therefore:

$$\| P^*(A)\phi \| = \| P(A)\phi \| , \qquad \| (P_1)^*(A)\phi \| = \| P_1(A)\phi \| .$$

These relations, together with (2), lead to:

$$(3) \qquad \| P_1(A)\phi \| \leqq (2d) \| P(A)\phi \| .$$

In this inequality (3), let us replace ϕ by $e^{H(z)}\phi$; we have

$$A_j[e^{H(z)}\phi] = e^{H(z)}\frac{\partial\phi}{\partial z_j} ,$$

and hence:

$$Q(A)[e^{H(z)}\phi] = e^{H(z)}Q(D_z)\phi \ ,$$

for any polynomial $Q(z)$ on C^n. Thus, we get, from (3):

$$\| e^{H(z)}P_1(D_z)\phi \| \leqq (2d) \| e^{H(z)}P(D_z)\phi \| \ . \qquad \text{q.e.d.}$$

2. Uniqueness in Cauchy problems. We shall denote by B_a $(a > 0)$ the open ball $|z| < a$ in C^n.

We say that an open set Ω in R^{2n} is *admissible at the point* z_0 if z_0 lies on the boundary of Ω, if the boundary of Ω is, near z_0, a piece of a C^∞ hypersurface and if the following property holds:

(A) *For some $a > 0$, there exists a function $F(z)$, holomorphic in the ball $|z - z_0| < a$, vanishing at z_0 and such that the diameter of the set U_b of those points $z \in \Omega$ which satisfy $|z-z_0|<a$, $-b<Re\,F(z)$ converges to 0 when $b > 0$ does.*

In the sequel, Ω will be an open set in R^{2n} admissible at the origin, a will be a positive number such that (A) holds for $z_0 = 0$ and some function $F(z)$ holomorphic in B_a. Furthermore, we shall assume that the intersection of B_a with the boundary of Ω is a piece S of a hypersurface C^∞ (passing by 0).

Let us clarify a little the geometric situation. Let us denote by W the piece of the hypersurface $Re F(z) = 0$ contained in B_a. Since $0 \in W \cap \bar{\Omega} \subset U_b$ for every $b > 0$, we must have $W \cap \bar{\Omega} = W \cap S = \{0\}$. On the other hand, for any $b > 0$, $U_b \cup C\Omega$ is a neighborhood of 0. For, let $\varepsilon > 0$ be small so that $|z| < \varepsilon$ implies $|Re\,F(z)| < b$. If $z \in B_\varepsilon$, $z \notin U_b$ only if $z \notin \Omega$. The interior of U_b is never empty. For assume it were and let z belong to U_b; z would have a neighborhood N in which $Re\,F$ would still be $> -b$ and since $z \in \bar{\Omega}$, N would intersect Ω; obviously $N \cap \Omega$ is contained in the interior of U_b.

We consider a polynomial $P(z)$ on C^n, of degree $m \geq 1$, and a partial differential operator on R^{2n} with continuous coefficients, Q, of order $\leqq$m $- 1$, satisfying the condition:

$$\| e^{H(z)}Qu \|_{L^2} \leqq K \sum_{p \neq 0} \| e^{H(z)}P^{(p)}(D_z)u \|_{L^2} \ , \tag{1}$$

for all $H(z)$ holomorphic in B_a, all $u(x, y) \in C_0^\infty$ with support in B_a.

THEOREM 2. *Let $U(x, y)$ be a function defined and C^m in $\bar{\Omega}$, with zero Cauchy data on S, satisfying:*

$$|P(D_z)U| \leqq |QU| \text{ in } \bar{\Omega} \ . \tag{2}$$

There exists a neighborhood of 0 in which U vanishes identically.

We keep our previous notations, for a, $F(z)$, etc.

Let us take a function $\beta(z)$, C^∞ in B_a, with the following properties:

$$\beta(z) = 1 \text{ for } z \in B_a \text{ and } -2\varepsilon \leqq Re\, F(z) \leqq 0\ ;$$
$$\beta(z) = 0 \text{ for } z \in B_a \text{ and } -3\varepsilon \leqq Re\, F(z)\ ,$$

where $\varepsilon > 0$ is chosen small enough so that the support of $\beta(z)$ intersects Ω according to a compact set contained in B_a. That is possible because of property (A); notice that the diameter of the compact set in question goes to 0 when $\varepsilon \to 0$.

We define now a function $v(z)$ as being equal to $\beta(z)U$ in Ω and to 0 elsewhere. Notice the following properties of v:

(i) the support of v is compact (and contained in $B_a \cap \bar{\Omega}$);

(ii) $v(z)$ is $m-1$ times continuously differentiable;

(iii) $P(D_z)v = \beta P(D_z)U + RU\varphi$ in Ω, R being a partial differential operator with C^∞ coefficients.

If one extends the definition of RU by 0 outside Ω, it becomes a continuous function in B_a since the order of R is at most $m-1$ and the Cauchy data of U were 0 on S. On the other hand, $P(D_z)U$ vanishes also on S, because of (2) and of the fact that Q is of order $\leqq m-1$. Hence, continuing $\beta P(D_z)U$ by 0 outside Ω leads again to a continuous function in B_a. We see thus that $P(D_z)v$ is a continuous function (in R^{2n}). This fact, together with properties (i) and (ii), allows us to extend to $v(z)$ the inequality of Theorem 1. We see that there exists a constant A such that, for all holomorphic functions $H(z)$ in B_a,

$$\sum_{p\neq 0} \| e^{H(z)} P^{(p)}(D_z)v \|_{L^2} \leqq A\delta \| e^{H(z)} P(D_z)v \|_{L^2}\ , \tag{3}$$

δ being the diameter of the support of v. Remember that $\delta \to 0$ if $\varepsilon \to 0$. Since, on $U_{2\varepsilon}$, $v = U$, by using inequality (1) and (3), we get:

$$\int_{U_{2\varepsilon}} e^{2Re H}(|U|^2 + |QU|^2)dxdy \leqq (2AK\delta)^2 \int_{U_{2\varepsilon}} e^{2Re H} |P(D_z)U|^2 dxdy$$
$$+ (2AK)^2 \int_{CU_{2\varepsilon}} e^{2Re H} |P(D_z)v|^2 dxdy\ .$$

But since $U_{2\varepsilon} \subset \Omega$, we have the right to substitute $|QU|$ for $|P(D_z)U|$ in the first integral of the right hand side; and if we choose ε small enough so that $(2AK\delta)^2 < 1/2$, we obtain finally:

$$\int_{U_\varepsilon} e^{2Re H} |U|^2 dxdy \leqq M \int_{CU_{2\varepsilon}} e^{2Re H} |P(D_z)v|^2 dxdy\ ,$$

M being a constant independent of both $H(z)$ and ε. Observe that the integral on the right hand side is actually performed on $U_{3\varepsilon} \cap CU_{2\varepsilon}$. Let us take $H(z) = (t/2)F(z)$, $t > 0$. The nature of the domains of integration leads us to:

$$e^{-t\varepsilon}\int_{U_\varepsilon} |U|^2 dxdy \leqq Me^{-2t\varepsilon}\int_{CU_{2\varepsilon}} |P(D_z)v|^2 dxdy ,$$

or:

$$\int_{U_\varepsilon} |U|^2 dxdy \leqq M_1 e^{-t\varepsilon} ,$$

where M_1 does not depend on t; we conclude that $U = 0$ in U_ε, q.e.d.

We end now by a few remarks about admissible sets.

1. Any open set Ω, strictly convex at a boundary point z_0 (and bounded near z_0 by a piece of C^∞ hypersurface) is admissible at this point. For simplicity, let us assume that $z_0 = 0$, and let H be an hyperplane passing by 0, such that $\bar{\Omega}$ intersects H only at the origin and lies entirely on one side of H (at least near 0). Let N be the unit vector, orthogonal to H, which lies on the side of H containing Ω. If $N_1, \cdots, N_n$ are the complex components of N, we may choose, as holomorphic function $F(z)$, the hermitian product $\bar{N}_1 z_1 + \cdots + \bar{N}_n z_n$.

2. There are open sets, admissible at a boundary point, which are not strictly convex at this point. For instance, consider an open set Ω whose boundary contains the origin (and is a piece of C^∞ hypersurface near it) and whose complement contains the cylinder $|z_1 - \alpha| < |\alpha|$, α being a complex number $\neq 0$. If the diameter of the intersection of Ω with the cylinder $|z_1 - k\alpha| < e^0 k|\alpha|$ $(k < 1, \varepsilon > 0)$ tends to 0 when $\varepsilon \to 0$, Ω will be admissible at $z = 0$. For then we may take, as holomorphic function $F(z)$, any branch of $-\log(1 - z_1/k\alpha)$. If $n = 1$, any open set whose complement contains the circle $|z_1 - \alpha| < |\alpha|$ (and whose boundary, near 0, is a piece of C^∞ curve passing by 0) is admissible at $z_1 = 0$. If $n > 1$, one may still construct open sets having the desired properties, which are not strictly convex at $z = 0$.

3. *Let $F(z)$ be any holomorphic function of z in a neighborhood U of 0 in C^n, such that $F(0) = 0$. Let U_+ be the set of points $z \in U$ such that $Re\, F(z) > 0$. If $n > 1$, the set U_+ cannot be strictly convex at $z = 0$.*

It U_+ were strictly convex at 0, there should exist an hyperplane H, passing by 0, intersecting $\bar{U}_+$ only at this point 0 and such that U_+ would lie only on one side of H. Let Ω be the other side of H, and $U(b)$ be the set of $z \in U$ such that $Re\, F(z) > -b$, $(b > 0)$. After maybe shrinking U we may say that the diameter of $U(b) \cap \Omega$ converges to 0 when $b \to 0$. For assume that this were not true: there would be pairs of points z'_k, z''_k in $U(1/k)$ such that $|z'_k - z''_k| \geqq c > 0$ for every $k = 1, 2, \cdots$. We could assume that z'_k converges to z', z''_k to z'', and we should have: $|z'-z''| \geqq c$, $z', z'' \in \bar{\Omega}$. But also $Re\, F(z')=0$, $Re\, F(z'')=0$, i.e., $z', z'' \in \bar{U}_+$. But that implies $z' = z'' = 0$, which is absurd. Hence the open set Ω is admissible at $z = 0$. But if Ω is admissible at some boundary point, the same must clearly be true for any open half space in C^n. And this would mean that there is uniqueness in the Cauchy problem for data on an arbitrary hyperplane and for any differential polynomial

$$P\left(\frac{\partial}{\partial z_1}, \cdots, \frac{\partial}{\partial z_n}\right),$$

which is absurd.

Reference

1. L. Nirenberg, *Uniqueness in the Cauchy problem*···, Comm. Pure Appl. Math., **10** (1957), 85–105.

ON THE REPRESENTATION OF OPERATORS BY CONVOLUTION INTEGRALS

J. D. WESTON

1. Introduction. Let $\mathfrak{X}$ be the complex vector space consisting of all complex-valued functions of a non-negative real variable. For each positive number u, let the *shift operator* I_u be the mapping of $\mathfrak{X}$ into itself defined by the formula

$$I_u x(t) = \begin{cases} 0 & (0 \leqq t < u) \\ x(t-u) & (t \geqq u) \end{cases}$$

Evidently, $I_{u+v} = I_u I_v$, for any positive numbers u and v.

A linear operator A which maps a subspace $\mathfrak{D}$ of $\mathfrak{X}$ into itself will here be called a *V-operator* (after Volterra) if

(1.1) for each x in $\mathfrak{D}$, the conjugate function x^* belongs to $\mathfrak{D}$,

(1.2) both $\mathfrak{D}$ and $\mathfrak{X}\backslash\mathfrak{D}$ are invariant under the shift operators,

(1.3) every shift operator commutes with A.

Many operators that occur in mathematical physics are of this type. If $\mathfrak{D}$ is any subspace of $\mathfrak{X}$ having the properties (1.1) and (1.2), the restriction to $\mathfrak{D}$ of each shift operator is an example of a V-operator. All 'perfect operators' (of which a definition may be found in [5][1]) are V-operators, on the space of perfect functions.

In this paper we obtain a representation theorem for V-operators which are continuous in a certain sense. This result leads to characterizations of two related classes of perfect operators, one of which has been considered from a different point of view in [5]. The main representation theorem (Theorem 4) is similar to a result obtained by R. E. Edwards [2] for V-operators which are continuous in another sense; and it closely resembles a theorem given recently by König and Meixner ([3], Satz 3).

2. Elementary properties of V-operators. An important property of V-operators is given by

THEOREM 1. *Let A be a V-operator, and let x_1 and x_2 be two of its operands such that, for some positive number t_0, $x_1(t) = x_2(t)$ whenever $0 \leqq t \leqq t_0$. Then $Ax_1(t) = Ax_2(t)$ whenever $0 \leqq t \leqq t_0$.*

Proof. Let $x = x_1 - x_2$. Then, since $x(t) = 0$ if $0 \leqq t \leqq t_0$, there is

Received January 22, 1960.

[1] And in § 4 below.

a function y such that $x = I_{t_0}y$; and y is an operand of A, by virtue of the property (1.2). Consequently, by virtue of (1.3), $Ax = I_{t_0}Ay$; so that $Ax(t) = 0$ whenever $0 \leqq t \leqq t_0$. But $Ax = Ax_1 - Ax_2$, since A is linear: hence the conclusion of the theorem.

With products and linear combinations defined in the usual way, the V-operators on a given space $\mathfrak{D}$ constitute a linear algebra $\mathfrak{A}(\mathfrak{D})$. If A belongs to $\mathfrak{A}(\mathfrak{D})$ then so does the operator A^* defined by

$$A^*x = (Ax^*)^* ,$$

where x is any function in $\mathfrak{D}$. We therefore have the unique decomposition

$$A = B + iC ,$$

where B and C belong to $\mathfrak{A}(\mathfrak{D})$ and are 'real' in the sense that Bx and Cx are real for every real function x in $\mathfrak{D}$. (The property (1.1) ensures that every function x in $\mathfrak{D}$ can be uniquely expressed as $x_1 + ix_2$, where x_1 and x_2 are real functions in $\mathfrak{D}$.)

If A is a linear combination of shift operators, we have

$$A = \sum_{j=1}^{n}\alpha_j I_{u_j} = I_u\sum_{j=1}^{n}\alpha_j I_{u_j - u} ,$$

where $\alpha_1, \cdots, \alpha_n$ are complex numbers, u is the least of the positive numbers $u_1, \cdots, u_n$, and I_0 is the unit operator (to be denoted henceforth by 'I'). From this it is apparent that A has no reciprocal in the algebra $\mathfrak{A}(\mathfrak{X})$; however, $I - A$ has a reciprocal in $\mathfrak{A}(\mathfrak{X})$, as the following result shows.

THEOREM 2. *Let A be a V-operator on a space $\mathfrak{D}$, and let u be any positive number. Then the formula*

$$Bx(t) = x(t) + \sum_{n=1}^{\infty} I_{nu}A^n x(t) ,$$

where x is any function in $\mathfrak{D}$, and $t \geqq 0$, defines a linear transformation B, of $\mathfrak{D}$ into $\mathfrak{X}$, which commutes with every shift operator and is such that $B(I - I_uA)x = x$ for every x in $\mathfrak{D}$ and $(I - I_uA)Bx = x$ if Bx is in $\mathfrak{D}$.

Proof. The series defining B certainly converges (pointwise): in fact, if $t_0 \geqq 0$ and m is a positive integer such that $mu \geqq t_0$, then, for any x in $\mathfrak{D}$,

$$Bx(t) = x(t) + \sum_{n=1}^{m} I_{nu}A^n x(t)$$

whenever $0 \leqq t \leqq t_0$. Hence if Bx is in $\mathfrak{D}$ then, by Theorem 1,

$$(I - I_u A)Bx(t) = x(t) - I_{(m+1)u}A^{m+1}x(t) = x(t)$$

whenever $0 \leqq t \leqq t_0$; so that $(I - I_u A)Bx = x$, since t_0 is arbitrary. Also, if x is in $\mathfrak{D}$ then $(I - I_u A)x$ is in $\mathfrak{D}$, so that

$$\begin{aligned} B(I - I_u A)x(t) &= (I - I_u A)x(t) + \sum_{n=1}^{m} I_{nu}A^n(I - I_u A)x(t) \\ &= x(t) - I_{(m+1)u}A^{m+1}x(t) = x(t) \end{aligned}$$

whenever $0 \leqq t \leqq t_0$. Thus $B(I - I_u A)x = x$. It can be verified in a similar way that B commutes with the shift operators and is linear.

If the transformation B of Theorem 2 maps $\mathfrak{D}$ into itself, then $I - I_u A$ has a reciprocal in $\mathfrak{A}(\mathfrak{D})$, namely B. This is certainly the case if $\mathfrak{D}$ consists of all the functions x that have some purely local property (for example, continuity, with $x(0) = 0$, or differentiability, with $x(0) = x'(0) = 0$, or local integrability).[2] It is also the case with certain other choices of $\mathfrak{D}$, provided that A is restricted to be a linear combination of shift operators; for example, if $\mathfrak{D}$ consists of the perfect functions, then an operator of the form

$$\alpha_0 I + \alpha_1 I_{u_1} + \cdots + \alpha_n I_{u_n} \tag{2.1}$$

has a reciprocal in $\mathfrak{A}(\mathfrak{D})$ if $\alpha_0 \neq 0$ (this can be seen at once on taking Laplace transforms and using Theorem 6 of [5]).

If $\mathfrak{D}$ contains more than the zero function, it is clear that (2.1) represents the zero operator on $\mathfrak{D}$ only if all the coefficients $\alpha_0, \cdots, \alpha_n$ are zero; and since the product of two operators of this form is another such operator, the reciprocal of (2.1) cannot be expressed in the same form unless it is a scalar multiple of I. Thus it is usual for $\mathfrak{A}(\mathfrak{D})$ to contain operators other than those of the form (2.1). In general it seems to be difficult to decide whether $\mathfrak{A}(\mathfrak{D})$ is commutative or not; but it is shown in § 4 that $\mathfrak{D}$ can be chosen, of moderate size, so that $\mathfrak{A}(\mathfrak{D})$ is not commutative.

The Laplace transformation is naturally associated with the idea of a V-operator, because it converts the shift operators to exponential factors. A locally integrable function x has an absolutely convergent Laplace integral if x is of exponential order at infinity, in the sense that $x(t) = O(e^{ct})$ as $t \to \infty$, for some real number c (depending on x). One can consider V-operators on spaces consisting of such functions, and for some of these spaces the following result is available.

THEOREM 3. *Let A be a V-operator on a space $\mathfrak{D}$ consisting of all*

[2] A property at infinity might be regarded as 'local', but this interpretation is to be excluded here.

the functions in $\mathfrak{X}$ which satisfy some (possibly empty) set of local conditions and are of exponential order at infinity. Then there are positive numbers b, c, and τ such that $|Ax(t)| \leqq be^{ct}$ whenever $t \geqq \tau$ and $|x(t)| \leqq 1$ for all t, with x in $\mathfrak{D}$.

Proof. Assuming the theorem to be false, we shall construct inductively a sequence $\{x_n\}$ in $\mathfrak{D}$, and a sequence $\{t_n\}$ of positive numbers, such that, for each positive integer n,

(i) $|x_n(t)| \leqq 2^{-n}$ for all values of t,

(ii) $t_n \geqq n$,

(iii) $x_n(t) = 0$ if $0 \leqq t \leqq t_{n-1}$, where $t_0 = 0$,

(iv) $|\sum_{j=1}^{n} Ax_j(t_n)| \geqq e^{nt_n}$.

In the first place, if the theorem is false, we can choose x_1 so that $|x_1(t)| \leqq \frac{1}{2}$ for all values of t and $|Ax_1(t)| \geqq e^t$ for some value of t, say t_1, greater than 1. Suppose, then, that the first $m-1$ terms of each sequence have been chosen, where $m > 1$, so that (i)-(iv) hold when $n \leqq m-1$. Let

$$y_m = \sum_{j=1}^{m-1} Ax_j .$$

Since y_m belongs to $\mathfrak{D}$, there is a real number c_m such that $|y_m(t)| \leqq e^{c_m t}$ when t is sufficiently large. We can choose x_m so that $|x_m(t)| \leqq 2^{-m}$ for all t, $x_m(t) = 0$ if $0 \leqq t \leqq t_{m-1}$, and

$$|Ax_m(t_m)| \geqq 2e^{(c_m+m)t_m} ,$$

where t_m is chosen so that $t_m \geqq m$ and $|y_m(t_m)| \leqq e^{c_m t_m}$. Then

$$\left|\sum_{j=1}^{m} Ax_j(t_m)\right| \geqq |Ax_m(t_m)| - |y_m(t_m)| \geqq e^{(c_m+m)t_m} \geqq e^{mt_m} .$$

Thus (i)–(iv) hold when $n = m$.

Now let $x_0 = \sum_{n=1}^{\infty} x_n$. Then $|x_0(t)| \leqq 1$ for all t, by virtue of (i); and x_0 belongs to $\mathfrak{D}$ since, by (iii), it has the appropriate local properties. Hence there is a real number c_0 such that $Ax(t) = O(e^{c_0 t})$ as $t \to \infty$; so that, by (ii), $Ax(t_n) = O(e^{c_0 t_n})$ as $n \to \infty$. But, by (iii) and (iv), and Theorem 1, $|Ax(t_n)| \geqq e^{nt_n}$ for each n. This contradiction proves the theorem.

3. Strong continuity. If the field of complex numbers is given either the discrete topology or the usual topology, the space $\mathfrak{X}$ can be given the corresponding topology of uniform convergence on finite closed intervals. The first of these topologies for $\mathfrak{X}$ has the property that every V-operator is continuous with respect to it, as Theorem 1 shows; but it does not make $\mathfrak{X}$ a topological vector space (it has the defect that $n^{-1}x \to 0$ as $n \to \infty$ only if x is the zero function). The second topology for $\mathfrak{X}$

is more interesting, and will be referred to as the *strong* topology. In fact we shall consider this only in relation to the closed subspace, $\mathfrak{C}_0$, consisting of all the continuous functions x for which $x(0) = 0$. For each x in $\mathfrak{C}_0$, and each non-negative number t, we define $\| x \|_t$ to be the least upper bound of $| x(u) |$ with $0 \leqq u \leqq t$. We can then give $\mathfrak{C}_0$ a metric, which determines the strong topology, by taking the distance between functions x and y to be

$$\sum_{n=1}^{\infty} 2^{-n} \| x - y \|_n / (1 + \| x - y \|_n) .$$

In this way $\mathfrak{C}_0$ becomes a Fréchet space.

In the case of $\mathfrak{C}_0$, which is an example of a space $\mathfrak{D}$ satisfying (1.1) and (1.2), a large class of V-operators, including those of the form (2.1), can be defined in terms of Riemann-Stieltjes convolution integrals. If ν is a function which belongs to $\mathfrak{X}$ and has bounded variation in every finite interval $[0, t]$, then the formula

$$Ax(t) = \int_0^t x(t-u) d\nu(u) \tag{3.1}$$

where x is any function in $\mathfrak{C}_0$, defines a V-operator A on $\mathfrak{C}_0$ (cf. [5], Theorem 3). Moreover, if $0 \leqq v \leqq t$ then

$$| Ax(v) | \leqq \int_0^v | x(v-u) | \, | d\nu(u) | \leqq \int_0^t \| x \|_t \, | d\nu(u) | , \qquad (t \geqq 0) ,$$

so that

$$\| Ax \|_t \leqq \| x \|_t \int_0^t | d\nu(u) | ;$$

whence it follows that A is strongly continuous (continuous with respect to the strong topology). The theorem we are about to prove shows that every strongly continuous V-operator on a sufficiently large space $\mathfrak{D}$ of continuous functions can be represented in this way (and can therefore be extended from $\mathfrak{D}$ to the whole of $\mathfrak{C}_0$).

If A is a linear operator on a subspace $\mathfrak{D}$ of $\mathfrak{C}_0$, and if $t \geqq 0$, we denote by '$\| A \|_t$' the least upper bound of $\| Ax \|_t$ with x in $\mathfrak{D}$ and $\| x \|_t \leqq 1$. It is clear that A is strongly continuous if and only if $\| A \|_t$ is finite for all values of t (or, equivalently, for all sufficiently large values of t).

THEOREM 4. *Let A be a strongly continuous V-operator on a strongly dense subspace $\mathfrak{D}$ of $\mathfrak{C}_0$, and let t be any positive number. Then there is a function ν in $\mathfrak{X}$, with $\nu(0) = 0$ and $\nu(u-) = \nu(u)$ whenever $0 < u \leqq t$, such that $Ax(t)$ is given by* (3.1) *for every x in $\mathfrak{D}$. This function ν is uniquely determined by A, and is independent of t; its total variation* *in the interval* $[0, t]$ *is* $\| A \|_t$.

Proof. For each function x in $\mathfrak{D}$, and for each positive number t, let x_t be the restriction of x to the closed interval $[0, t]$. Then, for a fixed value of t, the mapping $x \to x_t$ is a linear transformation of $\mathfrak{D}$ on to a subspace $\mathfrak{D}_t$ of the complex Banach space $C[0, t]$, consisting of all continuous functions on the interval $[0, t]$; moreover, $\| x_t \| = \| x \|_t$. If $x_t = 0$ then $Ax(t) = 0$, by Theorem 1; we can therefore define a linear functional φ on $\mathfrak{D}_t$ by the formula

$$\varphi(x_t) = Ax(t) .$$

This functional is continuous, with $\| \varphi \| = \| A \|_t$.

An integral representation of φ can be found by adapting a construction used by Banach ([1], 59-60). By a well-known theorem[3], φ can be extended without change of norm to the complex Banach space $M[0, t]$, which contains the characteristic functions of all the subintervals of $[0, t]$. A function ν_t can then be defined on $[0, t]$ so that $\nu_t(0) = 0$ and

(i) $$\int_0^t | d\nu_t(u) | \leqq \| \varphi \| ,$$

(ii) $$\varphi(f) = \int_0^t f(t-u) d\nu_t(u)$$

for every function f in $C[0, t]$.

Without affecting the validity of (i) or (ii), we can adjust ν_t so that it is continuous on the left at each interior point of the interval $[0, t]$. Moreover, if f is a continuous function such that $f(0) = 0$, then the jump of ν_t at the point t makes no contribution to the integral in (ii); therefore, as far as such functions f are concerned, we may suppose ν_t chosen so that $\nu_t(t-) = \nu_t(t)$, giving left-hand continuity throughout the interval $(0, t]$, and retaining (i). Under these conditions, ν_t is uniquely determined by A. For, if $0 < v \leqq t$ and $0 < \delta < v$, there is a function f_δ in $C[0, t]$ such that $\| f_\delta \| = 1$ and

$$f_\delta(u) = \begin{cases} 0 & (0 \leqq u \leqq t - v) \\ 1 & (t - v + \delta \leqq u \leqq t) . \end{cases}$$

Thus

$$\varphi(f_\delta) = \int_0^{v-\delta} d\nu_t(u) + \int_{v-\delta}^{v} f_\delta(t-u) d\nu_t(u) ,$$

and therefore

$$| \varphi(f_\delta) - \nu_t(v - \delta) | \leqq \int_{v-\delta}^{v} | d\nu_t(u) | ,$$

[3] The Hahn-Banach-Bohnenblust-Sobczyk extension theorem: see, for example, [8], 113.

so that $\varphi(f_\delta) \to \nu_t(v)$ as $\delta \to 0$.[4] But since $\mathfrak{D}$ is strongly dense in $\mathfrak{C}_0$, f_δ belongs to the closure of $\mathfrak{D}_t$, in $C[0, t]$; so that, φ being continuous, $\varphi(f_\delta)$ is uniquely determined by A, for each value of δ. This establishes the uniqueness of ν_t.

Now suppose that $t' > t$. By what has been proved, we have, for any x in $\mathfrak{D}$,

$$Ax(t) = \int_0^t x(t-u)d\nu_t(u) .$$

But $Ax(t) = I_{t'-t}Ax(t')$, and $I_{t'-t}A = AI_{t'-t}$; hence

$$Ax(t) = \int_0^{t'} I_{t'-t}x(t'-u)d\nu_{t'}(u) = \int_0^t x(t-u)d\nu_{t'}(u) .$$

It follows that $\nu_t(u) = \nu_{t'}(u)$ whenever $0 \leqq u \leqq t$; in particular, $\nu_t(t) = \nu_{t'}(t)$. Hence if we define the function ν by

$$\nu(t) = \nu_t(t) \qquad (t \geqq 0) ,$$

we obtain the required representation of A.

Finally, (i) shows that

$$\int_0^t | d\nu(u) | \leqq || A ||_t ,$$

and we have previously noted that, for any x in $\mathfrak{D}$,

$$|| Ax ||_t \leqq || x ||_t \int_0^t | d\nu(u) | .$$

Thus $\int_0^t | d\nu(u) | = || A ||_t$, and the proof is complete.[5]

As a corollary, we have

THEOREM 5. *Suppose that the formula*

$$Ax(t) = \int_0^t K(t, u)x(u)du \qquad (t \geqq 0)$$

defines a V-operator A on $\mathfrak{C}_0$, the kernel K being such that $\int_0^t | K(t,u) | du$ exists as a Lebesgue integral which is locally bounded with respect to t. Then there is a function k in $\mathfrak{X}$ such that, for each t, $K(t, u) = k(t-u)$ for almost all values of u.

[4] Here we use the fact that if a function of bounded variation is continuous on the left, then so is its total variation.

[5] In this proof we have not fully used the fact that A maps $\mathfrak{D}$ into itself: it is enough that A maps $\mathfrak{D}$ into C_0.

Proof. For each t, let $\| K \|_t$ be the least upper bound of $\int_0^v | K(v, u) | du$ with $0 \leqq v \leqq t$; this is finite, by hypothesis. Then, for each x in $\mathfrak{C}_0$,

$$\| Ax \|_t \leqq \| K \|_t \| x \|_t ,$$

so that A is strongly continuous. But

$$Ax(t) = \int_0^t K(t, t - u) x(t - u) du ,$$

so that if

$$L_t(u) = \int_0^u K(t, t - v)\, dv$$

then

$$Ax(t) = \int_0^t x(t - u) dL_t(u) .$$

Hence, by Theorem 4, $L_t = \nu$, a function which is independent of t. Since ν has bounded variation, there is a function k such that

$$k(u) = \frac{d}{du} \nu(u)$$

except when u is in a set E whose Lebesgue measure is 0. However, for each value of t,

$$\frac{d}{du} \nu(u) = \frac{d}{du} L_t(u) = K(t, t - u)$$

except when u is in a set E_t of measure 0. Thus

$$K(t, u) = k(t - u)$$

except when u is in the set $t - (E_t \cup E)$, which has measure 0.

The functions in $\mathfrak{C}_0$ which are of exponential order at infinity form a subspace $\mathfrak{E}_0$. The perfect functions form a smaller subspace, $\mathfrak{D}_0$ (in fact $\mathfrak{D}_0$ is the largest subspace of $\mathfrak{E}_0$ which is invariant under the differential operator, D).

THEOREM 6. *$\mathfrak{D}_0$ is strongly dense in $\mathfrak{C}_0$.*

Proof. It is easily seen that $\mathfrak{E}_0$ is strongly dense in $\mathfrak{C}_0$: in fact, if x is in $\mathfrak{C}_0$ and x_n is defined by

$$x_n(t) = \begin{cases} x(t) & (0 \leqq t \leqq n) \\ x(n) & (t \geqq n) , \end{cases}$$

then x_n belongs to $\mathfrak{C}_0$, for each n, and $x_n \to x$ strongly as $n \to \infty$. To show that $\mathfrak{D}_0$ is dense in $\mathfrak{C}_0$, let x be any function in $\mathfrak{C}_0$ and, for each positive number δ, let $g_{(\delta)}$ be a positive perfect function such that if $t \geqq \delta$ then $g_{(\delta)}(t) = 0$ and $\int_0^t g_{(\delta)}(u)du = 1$ (for example, we could take $g_{(\delta)}$ to be $Dh_{(\delta)}$, where $h_{(\delta)}$ is given by Lemma 1 of [5]). Let $x_{(\delta)} = x * g_{(\delta)}$. Then $x_{(\delta)}$ belongs to $\mathfrak{D}_0$ ('$x*$' is a perfect operator), and, if $v \geqq \delta$,

$$x_{(\delta)}(v) - x(v) = \int_0^v x(v-u)g_{(\delta)}(u)du - x(v)$$
$$= \int_0^\delta \{x(v-u) - x(v)\}g_{(\delta)}(u)du .$$

Now let t and ε be any positive numbers. Since x is uniformly continuous in the interval $[0, t]$, with $x(0) = 0$, we can choose δ so that

$$| x(v-u) - x(v) | < \varepsilon$$

whenever $\delta \leqq v \leqq t$, and $| x(v) | < \frac{1}{2}\varepsilon$ whenever $0 \leqq v \leqq \delta$; then

$$| x_{(\delta)}(v) - x(v) | < \varepsilon \int_0^\delta g_{(\delta)}(u)du = \varepsilon$$

if $\delta \leqq v \leqq t$, and if $0 \leqq v \leqq \delta$,

$$| x_{(\delta)}(v) - x(v) | \leqq \int_0^\delta | x(v-u) | g_{(\delta)}(u)du + | x(v) |$$
$$\leqq \tfrac{1}{2}\varepsilon \int_0^\delta g_{(\delta)}(u)du + \tfrac{1}{2}\varepsilon = \varepsilon .$$

Thus $\| x_{(\delta)} - x \|_t < \varepsilon$. It follows that $\mathfrak{D}_0$ is strongly dense in $\mathfrak{C}_0$.

In [5] it is shown that any positive perfect operator has the representation (3.1), with ν a non-decreasing function (in fact this holds for any positive V-operator on a space $\mathfrak{D}$ such that $\mathfrak{D}_0 \subseteqq \mathfrak{D} \subseteqq \mathfrak{C}_0$). It follows that the linear combinations of positive perfect operators, which form a linear algebra $\mathfrak{M}(\mathfrak{D}_0)$[6], are strongly continuous. On the other hand, there are strongly continuous perfect operators which do not belong to $\mathfrak{M}(\mathfrak{D}_0)$: for example, if $\nu(t) = \sin(e^{t^2} - 1)$, and A is defined on $\mathfrak{D}_0$ according to (3.1), then, as is shown in [5], A is a perfect operator which is not in $\mathfrak{M}(\mathfrak{D}_0)$; but of course A is strongly continuous. However, it is possible to characterize $\mathfrak{M}(\mathfrak{D}_0)$ in terms of seminorms, as follows.

THEOREM 7. *A V-operator A on $\mathfrak{D}_0$ is an element of $\mathfrak{M}(\mathfrak{D}_0)$ if and only if there is a real number c such that $\| A \|_t = O(e^{ct})$ as $t \to \infty$.*

Proof. By Theorem 1 of [5], an operator A on $\mathfrak{D}_0$ is in $\mathfrak{M}(\mathfrak{D}_0)$ if

6 $\mathfrak{M}(\mathfrak{D}_0)$ is denoted in [5] by '$\mathfrak{M}$'.

and only if it admits the representation (3.1) with ν a linear combination of positive non-decreasing functions which are of exponential order at infinity. This condition on ν is equivalent to the existence of a real number c such that $\int_0^t | d\nu(u) | = O(e^{ct})$ as $t \to \infty$. Therefore, by Theorems 4 and 6 above, A is in $\mathfrak{M}(\mathfrak{D}_0)$ if and only if $\| A \|_t = O(e^{ct})$ as $t \to \infty$.

Each function y in $\mathfrak{C}_0$ determines a strongly continuous V-operator A on $\mathfrak{C}_0$ according to the formula $Ax = x*y$; for, integration by parts shows that this formula is equivalent to (3.1), with

$$\nu(t) = D^{-1}y(t) = \int_0^t y(u)du \qquad (t \geqq 0) .$$

An important property of convolution in $\mathfrak{C}_0$ is the fact that it obeys the associative law (as well as the commutative law); more generally, we have

THEOREM 8. *Let A and B be strongly continuous V-operators, on $\mathfrak{C}_0$ and on a subspace $\mathfrak{D}$ of $\mathfrak{C}_0$ respectively. If x is any function in $\mathfrak{D}$ then Ax belongs to the strong closure of $\mathfrak{D}$; if Ax is in $\mathfrak{D}$ itself, then $ABx = BAx$. In particular, if y is a function in $\mathfrak{C}_0$ such that $x*y$ is in $\mathfrak{D}$, then $B(x*y) = (Bx)*y$.*

Proof. Let A be represented by a function ν in accordance with Theorem 4. Then for any x in $\mathfrak{D}$, each value $Ax(t)$ can be arbitrarily approximated by sums of the form

$$\sum_{j=1}^{n}\{\nu(u_j) - \nu(u_{j-1})\}x(t - u_j) ,$$

where $0 \leqq u_1 \leqq \cdots \leqq u_n \leqq t$; and this approximation is locally uniform with respect to t. Now the above sum is the value at t of the function

$$\text{(i)} \qquad \sum_{j=1}^{n}\alpha_j I_{u_j} x ,$$

where $\alpha_j = \nu(u_j) - \nu(u_{j-1})$. This function belongs to $\mathfrak{D}$, since $\mathfrak{D}$ satifies (1.2). Thus Ax belongs to the strong closure of $\mathfrak{D}$. Further, the points u_j can be chosen in such a way that, while Ax is strongly approximated by (i), ABx is simultaneously approximated, in the same sense, by

$$\text{(ii)} \qquad \sum_{j=1}^{n}\alpha_j I_{u_j} Bx .$$

But, since B is a V-operator, (ii) is the same as

$$B\sum_{j=1}^{n}\alpha_j I_{u_j} x .$$

Since B is strongly continuous, it follows that $ABx = BAx$ if Ax is an operand of B.

We can now prove a partial converse of Theorem 1, namely.

THEOREM 9. *Let A be a non-zero strongly continuous V-operator on $\mathfrak{C}_0$. Then there is a non-negative number τ such that* (i) *for any function x in C_0, $Ax(t) = 0$ whenever $0 \leqq t \leqq \tau$, and* (ii) *if $Ax(t) = 0$ whenever $0 \leqq t \leqq t_0$, where x belongs to $\mathfrak{C}_0$ and $t_0 \geqq \tau$, then $x(t) = 0$ whenever $0 \leqq t \leqq t_0 - \tau$. In particular, $x = 0$ if $Ax = 0$.*

Proof. Let ν be the function representing A according to Theorem 4, and let τ be the greatest lower bound of the numbers t for which $\nu(t) \neq 0$. Obviously, τ has the property (i) required by the theorem. Suppose that x is a function in $\mathfrak{C}_0$ such that $Ax(t) = 0$ whenever $0 \leqq t \leqq t_0$, where $t_0 \geqq \tau$. Let $g_{(\delta)}$ be defined as in the proof of Theorem 6, and let $x_{(\delta)} = x * g_{(\delta)}$. Then, for each value of δ, $x_{(\delta)}$ has a derivative $x'_{(\delta)}$ in $\mathfrak{C}_0$; in fact $x'_{(\delta)} = x * g'_{(\delta)}$. Also, if $0 \leqq t \leqq t_0$,

$$\int_0^t x'_{(\delta)}(t-u)\nu(u)du = Ax_{(\delta)}(t) = (Ax)*g_{(\delta)}(t)$$
$$= \int_0^t Ax(t-u)g_{(\delta)}(u)du = 0 \ .$$

Therefore, by a theorem of Titchmarsh [4, 327], $x'_{(\delta)}(t) = 0$ whenever $0 \leqq t \leqq t_0 - \tau$ (we cannot have $\nu(t) = 0$ for almost all t in a neighbourhood of τ, since ν is continuous on the left). Hence $x_{(\delta)}(t) = 0$ whenever $0 \leqq t \leqq t_0 - \tau$. Since $x_{(\delta)}(t) \to x(t)$ as $\delta \to 0$, the theorem follows.

It is a consequence of Theorem 8 that every strongly continuous V-operator on $\mathfrak{D}_0$ is a perfect operator (the converse is false; in fact it is easy to see that the differential operator D is not strongly continuous). Thus an operator A represented by (3.1) is a perfect operator if and only if it maps $\mathfrak{D}_0$ into itself. An equivalent condition is given by

THEOREM 10. *The formula* (3.1), *with x in $\mathfrak{D}_0$, represents a perfect operator A if and only if there is a positive integer n such that $D^{-n}\nu$ belongs to $\mathfrak{C}_0$, where*

$$D^{-n}\nu(t) = \int_0^t \cdots \int_0^{u_2} \nu(u_1)du_1 \cdots du_n \qquad (t = u_{n+1} \geqq 0) \ .$$

Proof. For any perfect function x and any positive integer n, we have from (3.1), after integration by parts,

$$Ax(t) = \int_0^t x^{(n+1)}(t-u)D^{-n}\nu(u)du \qquad (t \geqq 0) .$$

Thus if $D^{-n}\nu$ belongs to $\mathfrak{C}_0$ for some value of n, then A is a perfect operator. On the other hand, suppose that A, given by (3.1), is a perfect operator (when restricted to $\mathfrak{D}_0$). By a general representation theorem for perfect operators [6], there is a function y in $\mathfrak{C}_0$ such that, for some positive integer n, and every perfect function x,

$$Ax(t) = \int_0^t x^{(n+1)}(t-u)y(u)du \qquad (t \geqq 0) .$$

Hence $x^{(n+1)}*(y - D^{-n}\nu) = 0$, so that, by Theorem 9, $y = D^{-n}\nu$.

If $\nu(t) = e^{e^t}$, the V-operator A given by (3.1) does not map $\mathfrak{D}_0$ into itself, since ν does not satisfy the condition of Theorem 10.

Every perfect operator A has a Laplace transform, $\bar{A}$: if A is given by (3.1), $\bar{A}$ may or may not be given by

$$\bar{A}(z) = \int_0^\infty e^{-zt}d\nu(t) , \tag{3.2}$$

the integral being convergent when $\Re z$ is sufficiently large. This representation of $\bar{A}$ is certainly valid if A belongs to $\mathfrak{M}(\mathfrak{D}_0)$ (cf. [5], Theorem 4); and also if $\nu(t) = \sin(e^{t^2} - 1)$, for example. But if $D^{-1}\nu(t) = \sin(e^{t^2} - 1)$ the integral in (3.2) does not converge for any value of z (as can be seen on integrating twice by parts). However, (3.2) holds whenever the integral is convergent, as the following result shows.

THEOREM 11. *Let A be any strongly continuous perfect operator, and let ν be a function such that A is represented by* (3.1). *Then the Laplace transform $\bar{A}$ is represented by* (3.2), *with $\Re z$ sufficiently large, if the infinite integral is interpreted in the sense of summability* (C, n), *where n is any non-negative integer such that $D^{-n}\nu$ belongs to $\mathfrak{C}_0$.*

Proof. Let B be the perfect operator obtained on replacing ν by $D^{-1}\nu$ in (3.1). Then, if x is any perfect function, and $t \geqq 0$,

$$DBx(t) = Bx'(t) = \int_0^t x'(t-u)\nu(u)du = \nu(0)x(t) + \int_0^t x(t-u)d\nu(u) .$$

Thus $DB = \nu(0)I + A$. If ν belongs to $\mathfrak{C}_0$ then, since B is determined by the function ν in the sense that $Bx = x*\nu$, B has the same Laplace transform as ν; that is to say, when $\Re z$ is sufficiently large,

$$\bar{B}(z) = \int_0^\infty e^{-zt}\nu(t)dt .$$

Therefore, in this case,

$$\bar{A}(z) = z\bar{B}(z) - \nu(0) = \int_0^\infty ze^{-zt}\{\nu(t) - \nu(0)\}dt = \int_0^\infty e^{-zt}d\nu(t) ,$$

so that (3.2) holds, the integral being convergent.

We now proceed by induction. Suppose that, for some non-negative integer n, (3.2) holds in the sense of summability (C, n) provided that $D^{-n}\nu$ belongs to $\mathfrak{C}_0$ and $\Re z$ is sufficiently large. If $D^{-n-1}\nu$ belongs to $\mathfrak{C}_0$, and $t > 0$, then

$$\int_0^t \left(1 - \frac{u}{t}\right)^{n+1} e^{-zu}d\nu(u) = -\nu(0) + z\int_0^t \left(1 - \frac{u}{t}\right)^{n+1} e^{-zu}dD^{-1}\nu(u)$$
$$+ \frac{n+1}{t}\int_0^t \left(1 - \frac{u}{t}\right)^{n} e^{-zu}dD^{-1}\nu(u) .$$

But, by the induction hypothesis (with $D^{-1}\nu$ in place of ν),

$$\bar{B}(z) = \lim_{t\to\infty} \int_0^t \left(1 - \frac{u}{t}\right)^{n+1} e^{-zu}dD^{-1}\nu(u) = \lim_{t\to\infty} \int_0^t \left(1 - \frac{u}{t}\right)^{n} e^{-zu}dD^{-1}\nu(u)$$

when $\Re z$ is sufficiently large; so that

$$\lim_{t\to\infty} \int_0^t \left(1 - \frac{u}{t}\right)^{n+1} e^{-zu}d\nu(u) = -\nu(0) + z\bar{B}(z) = \bar{A}(z) .$$

Thus

$$\bar{A}(z) = \int_0^\infty e^{-zt}d\nu(t) \quad (C, n+1) ,$$

and the theorem follows.

If $\mathfrak{D}$ is any subspace of $\mathfrak{C}_0$ satisfying (1.1) and (1.2), the strongly continuous V-operators on $\mathfrak{D}$ form a subalgebra of $\mathfrak{A}(\mathfrak{D})$, say $\mathfrak{N}(\mathfrak{D})$. If $\mathfrak{D}$ is strongly dense in $\mathfrak{C}_0$, it follows from Theorem 4 that $\mathfrak{N}(\mathfrak{D})$ effectively consists of those operators in $\mathfrak{N}(\mathfrak{C}_0)$ which leave $\mathfrak{D}$ invariant. In this case, Theorems 8 and 9 show that $\mathfrak{N}(\mathfrak{D})$ is an integral domain (it is commutative, and has no divisors of zero). The full algebra $\mathfrak{N}(\mathfrak{C}_0)$[7] has the further property that any operator which is inverse to an operator in $\mathfrak{N}(\mathfrak{C}_0)$ is itself in $\mathfrak{N}(\mathfrak{C}_0)$: this is special case of

THEOREM 12. *Let A and B be strongly continuous V-operators on a strongly closed subspace $\mathfrak{D}$ of $\mathfrak{C}_0$, and suppose that there is an operator C on $\mathfrak{D}$ such that $A = BC$. Suppose also that $Bx = 0$ only if $x = 0$. Then C is a strongly continuous V-operator.*

[7] $\mathfrak{N}(\mathfrak{C}_0) = \mathfrak{M}(\mathfrak{C}_0)$, consisting of the linear combinations of positive V-operators on $\mathfrak{C}_0$.

Proof. If $u > 0$ and x is any function in $\mathfrak{D}$ then, since A and B are V-operators,

$$B(I_u Cx - CI_u x) = I_u Ax - AI_u x = 0;$$

so that, by the hypothesis concerning B, $I_u Cx = CI_u x$. In a similar way it can be verified that C is linear, and is therefore a V-operator. To show that C is strongly continuous, let $\{x_n\}$ be a strongly convergent sequence in $\mathfrak{D}$ such that the sequence $\{Cx_n\}$ is also strongly convergent. Since A and B are strongly continuous,

$$B(\lim_{n\to\infty} Cx_n - C \lim_{n\to\infty} x_n) = \lim_{n\to\infty} Ax_n - A \lim_{n\to\infty} x_n = 0 ,$$

so that $\lim_{n\to\infty} Cx_n = C \lim_{n\to\infty} x_n$; thus the graph of C is closed. Now $\mathfrak{D}$, being strongly closed, is a Fréchet space relative to the strong topology; hence, by Banach's closed-graph theorem [**1**, 41], C is strongly continuous.

4. Operators that commute with convolution. It is a consequence of Theorem 8 that a subspace $\mathfrak{D}$ of $\mathfrak{C}_0$, satisfying (1.1) and (1.2), is closed under convolution if it is strongly closed. On the other hand, $\mathfrak{D}_0$ is closed under convolution though it is not strongly closed. If $\mathfrak{D}$ is any subspace of $\mathfrak{C}_0$ which is closed under convolution (so forming an integral domain with no unit element), an operator A on $\mathfrak{D}$ will be said to *commute with convolution* if

$$A(x*y) = (Ax)*y$$

for all x and y in $\mathfrak{D}$. Such operators are necessarily linear (cf. [5], § 4), and, for a given choice of $\mathfrak{D}$, they form an integral domain $\mathfrak{D}^\#$ in which $\mathfrak{D}$ is isomorphically embedded (by the correspondence $x \to x*$).

A shift operator belongs to $\mathfrak{D}^\#$ if it maps $\mathfrak{D}$ into itself. Hence if $\mathfrak{D}$ satisfies (1.1) and (1.2), in addition to being closed under convolution, then all the operators in $\mathfrak{D}^\#$ are V-operators; in fact $\mathfrak{D}^\#$ is then a maximal commutative subalgebra of $\mathfrak{A}(\mathfrak{D})$. In this case, Theorem 8 shows that every strongly continuous V-operator commutes with convolution; so that

$$\mathfrak{N}(\mathfrak{D}) \subseteqq \mathfrak{D}^\# \subseteqq \mathfrak{A}(\mathfrak{D}) .$$

If, further, $\mathfrak{D}$ is strongly closed, then $\mathfrak{N}(\mathfrak{D}) = \mathfrak{D}^\#$: for, if B is defined by $Bx = x*y$, with y in $\mathfrak{D}$, and $A = BC$, where C is any operator in $\mathfrak{D}^\#$, then, for any x in $\mathfrak{D}$,

$$Ax = (Cx)*y = C(x*y) = C(y*x) = (Cy)*x;$$

thus the conditions of Theorem 12 are satisfied, so that C belongs to $\mathfrak{N}(\mathfrak{D})$. In particular, the operators on $\mathfrak{C}_0$ that commute with convolution are precisely the strongly continuous V-operators on $\mathfrak{C}_0$ (and can therefore be represented according to Theorem 4).

An operator A on $\mathfrak{F}_0$ which commutes with convolution can be extended to the whole of $\mathfrak{C}_0$ so as to preserve this property. For, if x is any function in $\mathfrak{C}_0$, let x_n be defined, for each positive integer n, as in the proof of Theorem 6: then x_n belongs to $\mathfrak{F}_0$, and Theorem 1 shows that $Ax_n(t)$ is independent of n provided that $n \geqq t$; therefore, if $t \geqq 0$, we can define $Ax(t)$ to be $Ax_n(t)$, where $n \geqq t$, without ambiguity. Since convolution is defined locally this extension of A is an operator on $\mathfrak{C}_0$ which commutes with convolution. It follows that A is strongly continuous, and that its extension to $\mathfrak{C}_0$ is unique (since $\mathfrak{F}_0$ is strongly dense in $\mathfrak{C}_0$).

The integration operator, D^{-1}, is an example of an operator on $\mathfrak{C}_0$ which commutes with convolution. Since $\mathfrak{D}_0$ can be expressed as $\bigcap_{n=1}^{\infty} D^{-n}\mathfrak{F}_0$, any operator on $\mathfrak{C}_0$ which commutes with convolution and leaves $\mathfrak{F}_0$ invariant must leave $\mathfrak{D}_0$ invariant. The converse of this is false: for, if A is defined by (3.1), ν being such that $D^{-2}\nu$ belongs to $\mathfrak{F}_0$ but $D^{-1}\nu$ does not, and $\nu(0) = 0$, then A maps $\mathfrak{D}_0$ into itself, by Theorem 10; however, if $x(t) = t$ then

$$Ax(t) = \int_0^t (t - u) d\nu(u) = D^{-1}\nu(t) ,$$

so that x is in $\mathfrak{F}_0$ but Ax is not.

The operators on $\mathfrak{D}_0$ that commute with convolution are the perfect operators. These can be characterized as those V-operators on $\mathfrak{D}_0$ which are continuous in a sense defined in terms of Laplace transforms [7][8]. The strongly continuous perfect operators are the strongly continuous V-operators on $\mathfrak{D}_0$, constituting the algebra $\mathfrak{N}(\mathfrak{D}_0)$; this algebra, and also its subalgebra $\mathfrak{M}(\mathfrak{D}_0)$, can be characterized in terms of convolution, as follows.

THEOREM 13. *A perfect operator belongs to $\mathfrak{N}(\mathfrak{D}_0)$ if and only if it can be extended to the whole of $\mathfrak{C}_0$ so as to commute with convolution; it belongs to $\mathfrak{M}(\mathfrak{D}_0)$ if and only if this extension (necessary unique) leaves $\mathfrak{F}_0$ invariant.*

Proof. If an operator A on $\mathfrak{D}_0$ can be extended to $\mathfrak{C}_0$ so as to commute with convolution, then its extension belongs to $\mathfrak{N}(\mathfrak{C}_0)$, so that A itself belongs to $\mathfrak{N}(\mathfrak{D}_0)$. On the other hand, any operator A in $\mathfrak{N}(\mathfrak{D}_0)$ admits the representation (3.1), which provides an extension of A to $\mathfrak{C}_0$: this extension, being strongly continuous, commutes with convolution;

[8] It is not at present known whether there are any V-operators on $\mathfrak{D}_0$ which are not perfect; that is to say, it is not known whether $\mathfrak{A}(\mathfrak{D}_0)$ is commutative or not (but there are linear operators on $\mathfrak{D}_0$ which commute with D and are not perfect [**6**]).

it is also unique, since $\mathfrak{D}_0$ is strongly dense in $\mathfrak{C}_0$.

If a perfect operator A has a strongly continuous extension to $\mathfrak{C}_0$ which leaves $\mathfrak{E}_0$ invariant, we can regard A as a V-operator on $\mathfrak{E}_0$; then, by Theorem 3, there is a real number c such that $\| A \|_t = O(e^{ct})$ as $t \to \infty$, and this implies, by Theorem 7, that A belongs to $\mathfrak{M}(\mathfrak{D}_0)$. On the other hand, if A belongs to $\mathfrak{M}(\mathfrak{D}_0)$ then the extension of A to $\mathfrak{C}_0$ given by (3.1) leaves $\mathfrak{E}_0$ invariant, by Theorem 3 of [5].

Finally, we give an example of a V-operator, on a strongly dense subspace of $\mathfrak{C}_0$, which does not commute with convolution. Let h be the Heaviside unit function ($h(t) = 1$ if $t \geqq 0$), and let $\mathfrak{D}_1$ be the class of all functions x given by

$$x = D^{-1}(y + Bh) , \tag{4.1}$$

where y belongs to $\mathfrak{C}_0$ and B is an operator of the type (2.1). Then $\mathfrak{D}_0 \subseteqq \mathfrak{D}_1 \subseteqq \mathfrak{C}_0$, and $\mathfrak{D}_1$ satiyfies (1.1) and (1.2); moreover, $\mathfrak{D}_1$ is closed under convolution. It is clear that y and B in (4.1) are uniquely determined by x, and that the mapping $x \to y$ is a V-operator, say A, on $\mathfrak{D}_1$. The operator D^{-1} maps $\mathfrak{D}_1$ into itself and commutes with convolution. However, $AD^{-1}x = x$ and $D^{-1}Ax = y$, so that $AD^{-1} \neq D^{-1}A$. Hence A does not commute with convolution. It follows that the algebra $\mathfrak{A}(\mathfrak{D}_1)$, of all V-operators on $\mathfrak{D}_1$, is not commutative.

References

1. S. Banach, *Théorie des opérations linéaires*, Warsaw, 1932.
2. R. E. Edwards, *Representation theorems for certain functional operators*, Pacific J. Math., **7** (1957), 1333-1339.
3. H König und J. Meixner, *Lineare Systeme und lineare Transformationen*, Math. Nachrichten, **19** (1958), 265-322.
4. E. C. Titchmarsh, *Theory of Fourier integrals*, Oxford, 1937.
5. J. D. Weston, *Positive perfect operators*, Proc. London. Math. Soc. (3), **10** (1960), 545-565.
6. ———, *Characterizations of Laplace transforms and perfect operators*, Archive for Rational Mechanics and Analysis, **3** (1959), 348-354.
7. ———, *Operational calculus and generalized functions*, Proc. Roy. Soc. A, **250** (1959), 460-471.
8. A. C. Zaanen, *Theory of integration*, Amsterdam, 1958.

UNIVERSITY OF DURHAM, NEWCASTLE UPON TYNE
AND
CALIFORNIA INSTITUTE OF TECHNOLOGY, PASADENA

NORMAL SUBGROUPS OF SOME HOMEOMORPHISM GROUPS*

J. V. WHITTAKER

1. Introduction. The normal subgroups of the group of all homeomorphisms of a space X have been enumerated by Fine and Schweigert [2] when X is a line, by Schreier and Ulam [3] when X is a circle, by Ulam and von Neumann [4] and Anderson [1] when X is a 2-sphere. In each of these cases there are either one or two proper normal subgroups. However, when X is an n-cell ($n > 1$), there are infinitely many. The object of this paper is to investigate the normal subgroups for a class of spaces X which includes the n-cell. Some of these normal subgroups, although not all, can be defined in terms of the family of fixed point sets of their elements, and we investigate this relationship at some length. A smallest normal subgroup is exhibited, and the corresponding quotient group is represented as a group of transformations of a related space.

2. Families of fixed point sets. Let X be a set, $\Pi(X)$ the group of all permutations of X (one-to-one mappings of X onto itself), and G a subgroup of $\Pi(X)$. Suppose that $\mathscr{F}$ is a non-empty family of subsets of X satisfying the following conditions:

(i) If $F_1, F_2 \in \mathscr{F}$, then there exists an $F_3 \in \mathscr{F}$ such that $F_3 \subset F_1 \cap F_2$,

(ii) If $F_1 \in \mathscr{F}$ and $g \in G$, then there exists an $F_2 \in \mathscr{F}$ such that $F_2 \subset g(F_1)$.

We shall call $\mathscr{F}$ ecliptic relative to G. For example, if $\mathscr{F}$ consists of the complements of all finite subsets of X, then $\mathscr{F}$ is ecliptic relative to $\Pi(X)$. If X has a topology, we denote the group of homeomorphisms of X by $H(X)$. Let X be a closed unit ball B_n in euclidean n-space and $\mathscr{F}_0$ consist of the complements in B_n of those balls which are concentric with B_n and have radius less than one. Then $\mathscr{F}_0$ is ecliptic relative to $H(B_n)$. In this connection, we note that for $h \in H(B_n)$, $h(S_{n-1}) = S_{n-1}$, where S_{n-1} is the boundary of B_n.

Let X again be an arbitrary set and G a subgroup of $\Pi(X)$. We introduce a partial ordering among the families of subsets of X as follows: $\mathscr{F} \leqq \mathscr{F}'$ provided that, for every $F \in \mathscr{F}$, there exists an $F' \in \mathscr{F}'$ such that $F' \subset F$. Evidently $\mathscr{F} \subset \mathscr{F}'$ implies $\mathscr{F} \leqq \mathscr{F}'$, where $\mathscr{F} \subset \mathscr{F}'$ means set inclusion, but the converse is false. We define equivalence of $\mathscr{F}$ and $\mathscr{F}'$ to mean $\mathscr{F} \leqq \mathscr{F}'$ and $\mathscr{F}' \leqq \mathscr{F}$, and we write $\mathscr{F} \cong \mathscr{F}'$.

Received October 26, 1959. This paper was written while the author was a Summer Research Institute Fellow of the Canadian Mathematical Congress.

LEMMA 1. *If $\mathscr{F}, \mathscr{F}'$ are families of subsets of X, $\mathscr{F} \cong \mathscr{F}'$, and $\mathscr{F}$ is ecliptic relative to G, then $\mathscr{F}'$ is ecliptic relative to G.*

Proof. If $F_1', F_2' \in \mathscr{F}'$, then there exist sets $F_1, F_2, F_3 \in \mathscr{F}$ and $F_3' \in \mathscr{F}'$ such that $F_1 \subset F_1'$, $F_2 \subset F_2'$, and $F_3' \subset F_3 \subset F_1 \cap F_2 \subset F_1' \cap F_2'$. Second, if $F_1' \in \mathscr{F}'$ and $g \in G$, then we can find $F_1, F_2 \in \mathscr{F}$ and $F_2' \in \mathscr{F}'$ such that $F_1 \subset F_1'$ and $F_2' \subset F_2 \subset g(F_1) \subset g(F_1')$.

To any family $\mathscr{F}$ we can adjoin all subsets of X which contain some element of $\mathscr{F}$ and thus obtain a family $\mathscr{F}^*$ which is clearly equivalent to $\mathscr{F}$ and, by Lemma 1, is ecliptic relative to G if $\mathscr{F}$ is. In fact, $\mathscr{F}^*$ has the property that $F_1^*, F_2^* \in \mathscr{F}^*$ and $g \in G$ implies $F_1^* \cap F_2^*, g(F_1^*) \in \mathscr{F}^*$. In addition, $\mathscr{F}^*$ is an upper bound, with respect to set inclusion, among the families equivalent to $\mathscr{F}$. We shall call $\mathscr{F}$ replete if it is equivalent to no larger family.

If $f \in \Pi(X)$, we set $K(f) = \{x \in X : f(x) = x\}$. For any family $\mathscr{F}$ of subsets of X, we define

$$S(\mathscr{F}, G) = \{g \in G : K(g) \supset F \text{ for some } F \in \mathscr{F}\} .$$

We note that if the empty set $\emptyset \in \mathscr{F}$, then $S(\mathscr{F}, G) = G$.

LEMMA 2. (a) *$\mathscr{F} \cong \mathscr{F}'$ implies $S(\mathscr{F}, G) = S(\mathscr{F}', G)$.*
(b) *If $\mathscr{F}$ satisfies* (i), *then $S(\mathscr{F}, G)$ is a subgroup of G.*
(c) *If $f \in \Pi(X)$, then*

$$f[S(\mathscr{F}, G)]f^{-1} = S(f(\mathscr{F}), fGf^{-1}) .$$

(d) *If $\mathscr{F}$ is ecliptic relative to G, then $S(\mathscr{F}, G)$ is a normal subgroup of G.*

Proof. For (a) we show that $\mathscr{F} \leqq \mathscr{F}'$ implies $S(\mathscr{F}, G) \subset S(\mathscr{F}', G)$. Indeed, if $g \in S(\mathscr{F}, G)$ and $K(g) \supset F$ for some $F \in \mathscr{F}$, we can find $F' \in \mathscr{F}'$ such that $F' \subset F \subset K(g)$, whence $g \in S(\mathscr{F}', G)$. In (b) we need merely observe that, for any $f_1, f_2 \in \Pi(X)$, $K(f_1, f_2) \supset K(f_1) \cap K(f_2)$ and $K(f_1^{-1}) = K(f_1)$. In part (c) we use the relation $K(fgf^{-1}) = f(K(g))$. If $g \in f[S(\mathscr{F}, G)]f^{-1}$, then $g = fg_1f^{-1}$, where $g_1 \in G$ and $K(g_1) \supset F$ for some $F \in \mathscr{F}$. Hence, $g \in fGf^{-1}$, $K(g) \supset f(F)$, and $g \in S(f(\mathscr{F}), fGf^{-1})$. If $g \in S(f(\mathscr{F}), fGf^{-1})$, then $g = fg_1f^{-1}$ for some $g_1 \in G$, and $K(g) \supset f(F)$ for some $F \in \mathscr{F}$. Hence, $K(g_1) \supset F$, and $g \in f[S(\mathscr{F}, G)]f^{-1}$. In part (d), let $f \in G$. From (c), $f[S(\mathscr{F}, G)]f^{-1} = S(f(\mathscr{F}), G)$. Normality will follow from (a) if we can show that $f(\mathscr{F}) \cong \mathscr{F}$. Clearly (ii) implies $f(\mathscr{F}) \leqq \mathscr{F}$. If $F_1 \in \mathscr{F}$, then there is an $F_2 \in \mathscr{F}$ such that $F_2 \subset f^{-1}(F_1)$, whence $f(F_2) \subset F_1$, and $\mathscr{F} \leqq f(\mathscr{F})$.

We shall assume, from now on, that X is a Hausdorff topological space, unless the contrary is explicitly stated. For $S(\mathscr{F}, H(X))$ we shall write $S(\mathscr{F})$, and if $\mathscr{F}$ is ecliptic relative to $H(X)$, we shall simply say that $\mathscr{F}$ is ecliptic. For any family of subsets of X, we introduce a further condition:

(iii) If $F \in \mathscr{F}$ and $U \subset X$ is open ($U \neq \varnothing$), then there exists an $h \in H(X)$ such that $h(cF) \subset U$, where cF is the complement of F in X. An ecliptic family which satisfies (iii) will be called strictly ecliptic. The family $\mathscr{F}_0$ of subsets of B_n defined above is evidently strictly ecliptic. If $\mathscr{F}$ satisfies (iii) and $\mathscr{F} \geqq \mathscr{F}'$, then clearly $\mathscr{F}'$ satisfies (iii). Since $K(h)$ is closed for every $h \in H(X)$, there is no loss of generality in assuming that the elements of any family $\mathscr{F}$ are closed, and this assumption will be made from now on, unless the contrary is stated.

LEMMA 3. *If X admits families $\mathscr{F}, \mathscr{F}'$ which satisfy* (ii) *and* (iii) *and contain more than one element, then $\mathscr{F} \cong \mathscr{F}'$.*

Proof. We may as well assume that $\mathscr{F}, \mathscr{F}'$ are replete in the closed subsets of X. If $F \in \mathscr{F}$, $F \neq X$, and $F' \in \mathscr{F}'$, then we can find $h \in H(X)$ such that $h(cF') \subset cF$. Hence, $h(F') \supset F$, $h(F') \in \mathscr{F}$, and $F' = h^{-1}(h(F')) \in \mathscr{F}$. Thus $\mathscr{F}' \subset \mathscr{F}$ and, similarly $\mathscr{F} \subset \mathscr{F}'$.

Some spaces contain no ecliptic families except $\{X\}$ and the set $\mathscr{C}(X)$ of all closed subsets of X. For, by Lemma 2, such a family defines a normal subgroup of $H(X)$; when X is a 1-sphere, Schreier and Ulam [3] showed that the only proper normal subgroup of $H(X)$ consists of the orientation-preserving elements of which some have no fixed points.

3. **Minimal normal subgroups.** We shall need to know something more about $H(X)$. Rather than make specific and detailed assumptions about the existence of certain homeomorphisms, we shall assume a mildly euclidean structure for X, namely:

(iv) If $U \subset X$ is open ($U \neq \varnothing$), then there exists an open $V \subset U$ which is homeomorphic to an open ball in a euclidean space of positive dimension.

The dimension of the ball may vary for different open sets. We shall refer to V as a euclidean neighborhood in X.

THEOREM 1. *Suppose X satisfies* (iv) *and contains a strictly ecliptic family $\mathscr{F}$. If N is a normal subgroup of $H(X)$, then either $N \supset S(\mathscr{F})$ or N consists of the identity e.*

Proof. Suppose $N \neq \{e\}$ and $g_0 \in N$, $g_0 \neq e$. Then $g_0(x) \neq x$ for some $x \in X$, and we can find a neighborhood U_0 of x and a euclidean neighborhood V_0 such that $g_0(U_0) \cap U_0 = \varnothing$ and $V_0 \subset g_0(U_0)$. Let ω map V_0 homeomorphically onto an open ball in some euclidean space, let $B \subset \omega(V_0)$ be a closed unit ball of the same dimension, and set $W_0 = \omega^{-1}(\operatorname{int} B)$, where int denotes interior. We wish to construct a homeomorphism h_0 of $\bar{W}_0$ in its relative topology with the following properties:

(a) $K(h_0) \supset \bar{W}_0 \cap cW_0$,

(b) there exists an open $V \subset W_0$ such that, for all integers $n > 0$, $h_0^n(\bar{V}) \cap \bar{V} = \emptyset$,

(c) if $A = \bigcup_{n=0}^{\infty} h_0^n(\bar{V})$, then $\bar{A} \cap cA$ is a single point. To do this is evidently equivalent to constructing such a homeomorphism k_0 of B, for then $h_0 = \omega^{-1}k_0\omega$ has the desired properties in $\bar{W}_0$. Let θ be a homeomorphism of $[0, 1]$ such that $K(\theta) = \{0, 1\}$ and $\theta(r) < r$ for $0 < r < 1$. If $p \in B$ lies at a distance r from the center of B, then we define $k_0(p)$ to be the point on the same radial line at a distance $\theta(r)$ from the center. By choosing a sufficiently small open ball in B which does not meet either the center or boundary, we can satisfy (a), (b), and (c).

We now define the function h_1 as follows: $h_1(x) = h_0(x)$ if $x \in W_0$, $h_1(x) = x$ if $x \in cW_0$ Clearly, $h_1 \in H(X)$. Now $g_1 = g_0 h_1^{-1} g_0^{-1} h_1 \in N$ since N is normal, and $g_0 h_1^{-1} g_0^{-1} h_1(x) = h_1(x)$ for $x \in W_0$, since $g_0^{-1}(W_0) \subset cW_0$. Thus $g_1(x) = h_0(x)$ for $x \in W_0$. Let g be any element of $S(\mathscr{F})$. Then there exists an $F \in \mathscr{F}$ and $h_2 \in H(X)$ such that $K(g) \supset F$ and $h_2(cF) \subset V$. Thus $K(h_2 g h_2^{-1}) \supset cV$. If we can construct an $h \in H(X)$ such that

$$g_1^{-1} h g_1 h^{-1} = h_2 g h_2^{-1} = f \,, \tag{1}$$

then we will have shown that $g \in N$ and $S(\mathscr{F}) \subset N$, since the left member of (1) lies in N. Let us rewrite (1) as $hg_1 = g_1 f h$. We set

$$h(x) = \begin{cases} g_1^n f g_1^{-n}(x) & \text{for} \quad x \in g_1^n(V) \,, \quad n = 1, 2, \cdots, \\ x & \text{for} \quad x \in c\left(\bigcup_{n=1}^{\infty} g_1^n(V)\right) . \end{cases}$$

By property (b) above, $m \neq n$ implies $g_1^m(V) \cap g_1^n(V) = \emptyset$, whence h is single-valued. Since $K(f) \supset cV$, the restriction of f to $\bar{V}$ is a homeomorphism of $\bar{V}$, and the same holds for $g_1^n f g_1^{-n}$ and $g_1^n(\bar{V})$, $n = 1, 2, \cdots$. Let $\bar{A} \cap cA$ consist of the point x_0, where $A = \bigcup_{n=0}^{\infty} g_1^n(\bar{V})$. Then each $x \neq x_0$ has a neighborhood which meets at most one of the sets $g_1^n(\bar{V})$, and h, h^{-1} are evidently continuous at such points. By the construction of h_0 and V, every neighborhood of x_0 contains all but a finite number of the sets $g_1^n(\bar{V})$, whence h, h^{-1} are continuous here as well. Hence, $h \in H(X)$. If $x \in c\bar{A}$, then $g_1(x) \in K(h)$ and $hg_1(x) = g_1 f h(x)$. When $x \in V$,

$$hg_1(x) = g_1 f g_1^{-1}(g_1(x)) = g_1 f(x) = g_1 f h(x) \,.$$

Finally, if $n \geqq 1$ and $x \in g_1^n(V)$, we have $g_1^n(V) \subset K(f)$, so that $g_1 f g_1^n(y) = g_1^{n+1}(y)$ when $y \in V$. Hence,

$$hg_1(x) = g_1^{n+1} f g_1^{-n-1}(g_1(x)) = g_1 f g_1^n f g_1^{-n}(x) = g_1 f h(x) \,.$$

This establishes (1) and completes the proof.

We offer the following example of a non-Hausdorff space X without euclidean neighborhoods which admits a strictly ecliptic family $\mathscr{F}$ such that $S(\mathscr{F})$ is not minimal. Let X be an infinite set in which $\mathscr{C}(X)$ consists of the finite subsets of X and X itself. Then $H(X) = \Pi(X)$. For $\mathscr{F}$ we take the collection of non-empty open sets and form $S(\mathscr{F})$. Since X is not Hausdorff, $K(h)$ need not be closed for $h \in H(X)$. Clearly $\mathscr{F}$ is strictly ecliptic, but $S(\mathscr{F})$ contains, as a proper normal subgroup, the set of $h \in H(X)$ such that $cK(h)$ is finite and h is an even permutation of $cK(h)$.

4. **Normal subgroups of $H(B_n)$.** As we remarked in § 2, the family $\mathscr{F}_0$ of complements of smaller, open, concentric balls in B_n is strictly ecliptic. When $\mathscr{F}_0$ is extended to a replete family, it will consist of all closed sets containing a neighborhood of the boundary S_{n-1}. In this section, we will also be concerned with the group $H_0(B_n)$ of those $h \in H(B_n)$ such that $K(h) \supset S_{n-1}$. Evidently $H_0(B_n) \supset S(\mathscr{F}_0)$, and $H_0(B_n)$ is normal in $H(B_n)$.

THEOREM 2. *If N is a normal subgroup of $H(B_n)$ which contains an element not in $H_0(B_n)$, then $N \supset H_0(B_n)$.*

Proof. We will assume, to begin with, than $n \geqq 2$. Suppose $g_0 \in N \cap cH_0(B_n)$, and choose $x \in S_{n-1}$ so that $g_0(x) \neq x$. Let W_0 be the part of an open ball with center x which lies in B_n and is small enough so that $g_0(W_0) \cap W_0 = \emptyset$. We wish to construct a homeomorphism h_0 of $\bar{W}_0$ and an open set $W \subset W_0$ such that $W \cap S_{n-1} \neq \emptyset$ and h_0, W satisfy (a), (b), (c) in the proof of Theorem 1. Let B, k_0, and V be the same as in that proof. If Π is an $(n-1)$-dimensional hyperplane which passes through the center of B and meets V, then Π divides B into two regions (including boundaries) Δ, Δ' such that $\Delta \cap \Delta' = \Pi$. The restriction of k_0 to Δ is evidently a homeomorphism of Δ. Let ψ map Δ homeomorphically onto $\bar{W}_0$ in such a way that $\psi(\Pi) = \bar{W}_0 \cap S_{n-1}$. Then $h_0 = \psi k_0 \psi^{-1}$ and $W = \psi(\Delta \cap V)$ clearly satisfy (a), (b), (c). We define $h_1(x) = h_0(x)$ for $x \in W_0$, $h_1(x) = x$ for $x \in cW_0$, so that $h_1 \in H(B_n)$. Then $g_1 = g_0 h_1^{-1} g_0^{-1} h_1 \in N$, and $g_1(x) = h_1(x)$ for $x \in W_0$, as in the proof of Theorem 1. If g is any element of $H_0(B_n)$ such that $K(g) \supset cW$, it follows from the construction in the same proof that $g \in N$.

Let $p, q \in S_{n-1}$ be antipodal, D the diameter joining them, and Π_1, $\Pi_2 \subset B_n$ two $(n-1)$-dimensional hyperplanes perpendicular to D. Now Π_1, Π_2 divide B_n into three regions (including boundaries) Δ_1, Δ_2, Δ_3 and, correspondingly, S_{n-1} into three zones (including boundaries) Z_1, Z_2, Z_3. We take Δ_2 to be the middle region, so that $\Delta_1 \cap \Delta_2 = \Pi_1$, $\Delta_2 \cap \Delta_3 = \Pi_2$, $p \in \Delta_1$, $q \in \Delta_3$. Let P, Q be arbitrary neighborhoods of p, q, respectively, such that $\bar{P} \subset \Delta_1$, $\bar{Q} \subset \Delta_3$.

Next, we construct $h_2, h_3 \in H(B_n)$ such that $h_2(\bar{P}) \supset cW$, $h_3(\bar{Q}) \supset cW$.

For example, h_2 might first expand $\bar{P}$ until its complement is quite small and then rotate the complement into W. If $g \in H_0(B_n)$ and $K(g) \supset \bar{P}$, then $K(h_2 g h_2^{-1}) \supset cW$, whence $g \in N$. Similarly, $K(g) \supset \bar{Q}$ implies $g \in N$. We now wish to construct a homeomorphic mapping θ of B_n onto $\varDelta_2 \cup \varDelta_3$ such that $\theta(x) = x$ for all $x \in \varDelta_3$. To accomplish this, we introduce spherical coordinates $r, \phi_1, \cdots, \phi_{n-1}$ for the points $x \in B_n$ such that ϕ_1 is the angle between D and the radial line through x. Then Π_i satisfies the equation $r \cos \phi_1 = k_i$, $|k_i| < 1$ $(i = 1, 2)$. Let r, ϕ_1 be regarded as polar coordinates for the closed upper half-plane in euclidean 2-space, and let R be the set of (r, ϕ_1) such that $r \leqq 1$, $0 \leqq \phi_1 \leqq \pi$. The lines $r \cos \phi_1 = k_i$ $(i = 1, 2)$ divide R into three regions (including boundaries) R_1, R_2, R_3. Let ω be a homeomorphic mapping of R onto $R_2 \cup R_3$ such that $\omega(y) = y$ for all $y \in R_3$. We then set

$$\theta(r, \phi_1, \cdots, \phi_{n-1}) = (\omega(r, \phi_1), \phi_2, \cdots, \phi_{n-1}) .$$

Let f be any element of $H_0(B_n)$. Then $\theta f \theta^{-1} \in H(\varDelta_2 \cup \varDelta_3)$, and $K(\theta f \theta^{-1}) \supset \Pi_1 \cup Z_2 \cup Z_3$. We define $g_2(x) = \theta f \theta^{-1}(x)$ if $x \in \varDelta_2 \cup \varDelta_3$, $g_2(x) = x$ if $x \in \varDelta_1$. Clearly, $g_2 \in H(B_n)$, and $K(g_2) \supset \bar{P}$, whence $g_2 \in N$. In addition, $g_2(x) = f(x)$ for $x \in \varDelta_3 \cap f^{-1}(\varDelta_3)$, so that $K(g_2^{-1} f) \supset \bar{Q} = \varDelta_3 \cap f^{-1}(\varDelta_3)$, and $g_3 = g_2^{-1} f \in N$. Hence, $f = g_2 g_3 \in N$, and $H_0(B_n) \subset N$. When $n = 1$, the constructions in the first half of the proof can not be carried out in S_0. The theorem follows, in this case, from the result obtained in [2] that the only proper normal subgroups of $H(B_1)$ are $S(\mathscr{F}_0)$ and $H_0(B_1)$.

If $G \subset \Pi(X)$ and $Y \subset X$, we denote the restrictions of the elements of G to Y by $G \mid Y$. For any orientable space X, we let $E(X)$ denote the group of all orientation-preserving homeomorphisms of X.

LEMMA 4. *If N is a normal subgroup of $H(B_n)$, then $N \mid S_{n-1}$ is a normal subgroup of $H(S_{n-1})$.*

Proof. Clearly $N \mid S_{n-1}$ is a subgroup of $H(S_{n-1})$. If $h_0 \in H(S_{n-1})$, we can extend h_0 to an element h of $H(B_n)$. Let p_0 be the center of B_n, $p \neq p_0$ a point of B_n lying on the sphere S with center p_0, and π the radial projection of S onto S_{n-1}. We define $h(p) = \pi^{-1} h_0 \pi(p)$, $h(p_0) = p_0$. Clearly $h \in H(B_n)$. Then $N \mid S_{n-1} = (hNh^{-1}) \mid S_{n-1} = h_0(N \mid S_{n-1}) h_0^{-1}$.

COROLLARY. *If N is not contained in $H_0(B_n)$ and $n \leqq 3$, then N is either $E(B_n)$ or $H(B_n)$.*

Proof. By Lemma 4, $N \mid S_{n-1}$ is a normal subgroup of $H(S_{n-1})$ different from $\{e\}$. It was proved in [3] for $n = 2$ and in [1] for $n = 3$ that the only normal subgroups of $H(S_{n-1})$ are $\{e\}$, $E(S_{n-1})$, and $H(S_{n-1})$. Hence, if $h \in E(B_n)$, there exists a $g \in N$ such that $h \mid S_{n-1} = g \mid S_{n-1}$.

Then $f = g^{-1}h \in H_0(B_n) \subset N$ by Theorem 2, and $h = gf \in N$. If $N \not\subset E(B_n)$, a similar argument shows that $N = H(B_n)$. We note that $H_0(B_1) = E(B_1)$.

5. The lattice of normal subgroups. In the first part of this section, we revert to the assumption that X is an arbitrary set. The intersection of two ecliptic families may be empty. If, for example, $\mathscr{F}_0$ is the ecliptic family defined above for B_2 and $\mathscr{F}_1$ is the family of complements of interiors of simple polygons lying entirely in the interior of B_2, then $\mathscr{F}_0 \cap \mathscr{F}_1 = \emptyset$, although $\mathscr{F}_0^* = \mathscr{F}_1^*$. However, the intersection of any collection of replete, ecliptic families is also replete, ecliptic, and non-empty, since it always contains $\{X\}$. The smallest ecliptic family (up to equivalence) which contains a given collection $\{\mathscr{F}_\alpha\}$ of ecliptic families consists of all finite intersections of elements in $\cup \mathscr{F}_\alpha$. We denote this set by $\vee \mathscr{F}_\alpha$ and set $\wedge \mathscr{F}_\alpha = \cap \mathscr{F}_\alpha$. If the $\mathscr{F}_\alpha$ are replete, then $\vee \mathscr{F}_\alpha$ is also replete. For if $F_1, \cdots, F_n \in \cup \mathscr{F}_\alpha$, and $F \supset \bigcap_i F_i$, then $F \cup F_i \in \cup \mathscr{F}_\alpha$ $(i = 1, \cdots, n)$, and $F = \bigcap_i (F \cup F_i)$.

For any collection $\{G_\alpha\}$ of subgroups of a group G, we set $\wedge G_\alpha = \cap G_\alpha$ and define $\vee G_\alpha$ as the smallest subgroup of G which contains G_α.

LEMMA 5. *If G is a subgroup of $\Pi(X)$ and $\{\mathscr{F}_\alpha\}_{\alpha \in A}$ is a collection of replete ecliptic families relative to G, then*

$$S(\wedge \mathscr{F}_\alpha, G) = \wedge S(\mathscr{F}_\alpha, G),\ S(\vee \mathscr{F}_\alpha, G) \supset \vee S(\mathscr{F}_\alpha, G)\ .$$

Proof. If $g \in S(\wedge \mathscr{F}_\alpha, G)$, then there is an $F \in \cap \mathscr{F}_\alpha$ such that $K(g) \supset F$, whence $g \in S(\mathscr{F}_\alpha, G)$ for each $\alpha \in A$. If $g \in \wedge S(\mathscr{F}_\alpha, G)$, then, for each $\alpha \in A$, there is an $F_\alpha \in \mathscr{F}_\alpha$ such that $K(g) \supset F_\alpha$. Hence, $K(g) \supset F = \bigcup_{\alpha \in A} F_\alpha$, $F \in \mathscr{F}_\beta$ for each $\beta \in A$ since $\mathscr{F}_\beta$ is replete, and $g \in S(\wedge \mathscr{F}_\alpha, G)$. This proves the first relation. In the second, if $g \in \vee S(\mathscr{F}_\alpha, G)$, then there are sets $F_1, \cdots, F_n \in \bigcup_{\alpha \in A} \mathscr{F}_\alpha$ and elements $g_1, \cdots, g_n \in G$ such that $K(g_i) \supset F_i$ $(i = 1, \cdots, n)$ and $g = g_1 \cdots g_n$. Hence, $K(g) \supset F = \bigcap_i F_i$, $F \in \vee \mathscr{F}_\alpha$, and $g \in S(\vee \mathscr{F}_\alpha, G)$.

We now return to the case $X = B_n$.

LEMMA 6. *Let $\mathscr{G}$ be a family of* (*not necessarily closed*) *subsets of S_{n-1} which*

(a) *satisfies* (i), *or*

(b) *is ecliptic relative to $H(B_n)$. Let $\mathscr{F}$ be the family of closed subsets of B_n which contain a member of $\mathscr{G}$ in their interior* (*in the relative topology of B_n*). *Then*

(a) *$\mathscr{F}$ is ecliptic relative to $H_0(B_n)$, or*

(b) *$\mathscr{F}$ is ecliptic relative to $H(B_n)$. In either case, $\mathscr{F}$ is replete.*

Proof. If $F_0, F_0' \in \mathscr{G}$ and $F_0 \subset \operatorname{int} F$, $F_0' \subset \operatorname{int} F'$, then $F_0 \cap F_0' \subset \operatorname{int}$

$F \cap \operatorname{int} F' = \operatorname{int}(F \cap F')$, and $F \cap F' \in \mathscr{F}$, whence $\mathscr{F}$ satisfies (i). In part (a), if $h \in H_0(B_n)$, then $\operatorname{int} h(F) = h(\operatorname{int} F) \supset h(F_0) = F_0$, and $h(F) \in \mathscr{F}$. In part (b), if $h \in H(B_n)$, then there is an $F_0'' \in \mathscr{G}$ such that $h(F_0) \supset F_0''$, and $\operatorname{int} h(F) \supset F_0''$ as in (a), so that $h(F) \in \mathscr{F}$. Thus (ii) is verified in each case. The repleteness of $\mathscr{F}$ is obvious.

We will indicate the above relationship between $\mathscr{F}$ and $\mathscr{G}$ by saying that $\mathscr{F}$ is derived from $\mathscr{G}$. The simplest example of a derived ecliptic family relative to $H_0(B_n)$ is that in which $\mathscr{G}$ consists of a single subset of S_{n-1}. An ecliptic family relative to $H(B_n)$ is obtained by letting $\mathscr{G}$ consist of the complements in S_{n-1} of finite subsets of S_{n-1}. When $n = 2$, a family equivalent to the latter can be described as the set of complements in B_2 of interiors of simple closed curves which meet S_1 in a finite number of points. The construction can be varied by taking the set of complements of countable or first category subsets in S_{n-1}.

Returning to Lemma 5 and the case $X = B_n, G = H(B_n)$, we have not been able to determine whether equality holds in the second relation even for the case $S(\mathscr{F} \vee \mathscr{F}') \supset S(\mathscr{F}) \vee S(\mathscr{F}')$, $\mathscr{F} \vee \mathscr{F}' = \mathscr{C}(X)$. However, we do have the following result for derived families.

THEOREM 3. *Let* $\mathscr{F}, \mathscr{F}'$ *be derived from* $\mathscr{G}, \mathscr{G}'$, *respectively, where* $\mathscr{G} = \{P_0\}$, $\mathscr{G}' = \{Q_0\}$, *and suppose that* $\bar{P}_0, \bar{Q}_0$ *can be separated in* S_{n-1} *by an* $(n-2)$*-sphere* $\sum \subset S_{n-1}$ *which is tame relative to* $H(B_n)$. *Then*

$$S(\mathscr{F} \vee \mathscr{F}', H_0(B_n)) = S(\mathscr{F}, H_0(B_n)) \vee S(\mathscr{F}', H_0(B_n)) . \tag{2}$$

Proof. Let Π_1 be an $(n-1)$-dimensional hyperplane passing through the center of B_n, and set $\Sigma_1 = \Pi_1 \cap S_{n-1}$. Choose $h \in H(B_n)$ such that $h(\Sigma) = \Sigma_1$. Since Π_1 and $h(\bar{Q}_0)$ are closed and disjoint, we can find a second hyperplane Π_2 parallel to Π_1 and lying between Π_1 and $h(\bar{Q}_0)$. Now Π_1, Π_2 divide B_n into three regions (including boundaries) $\Delta_1, \Delta_2, \Delta_3$ such that $h(\bar{P}_0) \subset \Delta_1$, $h(\bar{Q}_0) \subset \Delta_3$. In fact, $P_0 \subset \operatorname{int} h^{-1}(\Delta_1)$, $Q_0 \subset \operatorname{int} h^{-1}(\Delta_3)$, where int denotes interior in the relative topology of B_n. Hence, $h^{-1}(\Delta_1) \in \mathscr{F}$ and $h^{-1}(\Delta_3) \in \mathscr{F}'$. Since these sets are disjoint, $\emptyset \in \mathscr{F} \vee \mathscr{F}'$ and $S(\mathscr{F} \vee \mathscr{F}', H_0(B_n)) = H_0(B_n)$. By setting $\Delta_1 = P$, $\Delta_3 = Q$, and following the argument in the second half of the proof of Theorem 2, we can show that the group generated by those $g \in H_0(B_n)$ such that $K(g) \supset \Delta_1$ or Δ_3 is precisely $H_0(B_n)$. Since $K(g) \supset \Delta_1$ implies $g \in S(h^{-1}(\mathscr{F}), H_0(B_n))$, and $K(g) \supset \Delta_3$ implies $g \in S(h^{-1}(\mathscr{F}'), H_0(B_n))$, it follows from Lemma 2(*c*) that

$$\begin{aligned} H_0(B_n) &= h[S(h^{-1}(\mathscr{F}), H_0(B_n)) \vee S(h^{-1}(\mathscr{F}'), H_0(B_n))]h^{-1} \\ &= h[S(h^{-1}(\mathscr{F}), H_0(B_n))]h^{-1} \vee h[S(h^{-1}(\mathscr{F}'), H_0(B_n))]h^{-1} \\ &= S(\mathscr{F}, H_0(B_n)) \vee S(\mathscr{F}', H_0(B_n)) . \end{aligned}$$

Hence, (2) is established. When $n = 1$, the hypothesis of the theorem states that P_0 and Q_0 are the two points in S_0. The construction in the second half of the proof of Theorem 2 can evidently be carried through in this case.

6. Quotient spaces. we turn now to the problem of representing the quotient groups $H_0(B_n)/S(\mathscr{F})$, where $\mathscr{F}$ is an ecliptic family, as groups of transformations.

THEOREM 4. *Let $A \subset S_{n-1}$ have the property that the set of its neighborhoods in B_n has a countable base, and let $\mathscr{F}$ be the ecliptic family derived from $\{A\}$. Then $H_0(B_n)/S(\mathscr{F})$ can be represented as a group of order-preserving transformations of a partially ordered set Z onto itself.*

Proof. Let Y be the set of all countable sequences $\{U_k\}$ of open subsets $U_k \subset B_n$ such that $U_1 \supset U_2 \supset \cdots$, and $\{U_k\}$ is a base for the set of neighborhoods of A. We introduce a partial ordering in Y as follows: $\{U_k\} \leqq \{V_k\}$ if there exists a $k_0 > 0$ such that $k > k_0$ implies $U_k \subset V_k$. We call $\{U_k\}$ and $\{V_k\}$ equivalent if $\{U_k\} \leqq \{V_k\}$ and $\{V_k\} \leqq \{U_k\}$, and we write $\{U_k\} \equiv \{V_k\}$. Thus $\{U_k\} \equiv \{V_k\}$ means that $U_k = V_k$ for all but a finite set of k's. If $\{U_k\} \equiv \{V_k\}$ and $\{U_k\} \leqq \{W_k\}$, then clearly $\{V_k\} \leqq \{W_k\}$. Let Z be the set of equivalence classes in Y formed by the relation $\equiv$. If $u, v \in Z$, we define $u \leqq v$ to mean that the same ordering subsists between their respective equivalence classes. Moreover, $u \leqq v$ and $v \leqq u$ implies $u = v$.

If $h \in H_0(B_n)$ and $\{U_k\} \in Y$, then $\{h(U_k)\} \in Y$. Furthermore, $\{V_k\} \in Y$ and $\{U_k\} \leqq \{V_k\}$ implies $\{h(U_k)\} \leqq \{h(V_k)\}$. In particular, $\{U_k\} \equiv \{V_k\}$ implies $\{h(U_k)\} \equiv \{h(V_k)\}$. Thus, corresponding to h there is an element $\omega(h) \in \Pi(Z)$ which is order-preserving, and $g \in H_0(B_n)$ implies $\omega(gh) = \omega(g)\omega(h)$. We now show that $h \in S(\mathscr{F})$ if, and only if, $\omega(h) = i$, where i is the identity in $\Pi(Z)$. If $h \in S(\mathscr{F})$, then there is an $F \subset B_n$ such that $K(h) \supset F$ and $\operatorname{int} F \supset A$. For any $u \in Z$, let $\{U_k\}$ be a representative of u in Y. Since $\{U_k\}$ is a base for the neighborhoods of A, we can find $k_0 > 0$ such that $k > k_0$ implies $U_k \subset \operatorname{int} F$, whence $\omega(h)(u) = u$, and $\omega(h) = i$. Conversely, if $h \notin S(\mathscr{F})$, then for each $\{U_k\} \in Y$, there exists a sequence $\{x_k\}$ of points in B_n such that $x_k \in U_k$ and $h(x_k) \neq x_k$ ($k = 1, 2, \cdots$). Setting $V_k = U_k \cap c\{h(x_k)\}$ for each k, we have $\{V_k\} \in Y$ and $\{h(V_k)\} \not\equiv \{V_k\}$. If $\{V_k\}$ is a respresentative of $v \in Z$, then $\omega(h)(v) \neq v$, and $\omega(h) \neq i$. This proves our assertion. Let θ denote the canonical mapping from $H_0(B_n)$ onto $H_0(B_n)/S(\mathscr{F})$. Then $\theta(g) = \theta(h)$ if, and only if, $\omega(g) = \omega(h)$. Hence, $\omega\theta^{-1}$ is an isomorphism between $H_0(B_n)/S(\mathscr{F})$ and $\omega(H_0(B_n))$.

If A is closed in B_n, then A is compact, and the uniform $(1/k)$-

neighborhoods of A form a base for its set of neighborhoods, so that the hypothesis of the theorem is satisfied in this case. If $A = S_{n-1}$, then $\mathscr{F} = \mathscr{F}_0$, and the construction in the proof allows us to represent $H(B_n)/S(\mathscr{F})$ as a subgroup of order-preserving elements in $\Pi(Z)$ which contains $\omega(H_0(B_n))$.

References

1. R. D. Anderson, *The algebraic simplicity of certain groups of homeomorphisms*, Amer. J. Math., **70** (1958), 955-963.

2. N. J. Fine and G. E. Schweigert, *On the group of homeomorphisms of an arc*, Annals of Math., **62** (1955), 237-253.

3. J. Schreier and S. Ulam, *Eine Bemerkung uber die Gruppe der topologischen Abbildungen der Kreislinie auf sich selbst*, Studia Math., **5** (1934), 155-159.

4. S. Ulam and J. von Neumann, *On the group of homeomorphisms of the surface of the sphere*, (abstract), Bull. Amer. Math. Soc., **53** (1947), 506.

University of British Columbia

Pacific Journal of Mathematics

Vol. 10, No. 4 December, 1960